Handbooks in Finance

HANDBOOK of ASSET and LIABILITY MANAGEMENT Volume 2

APPLICATIONS and CASE STUDIES

Editors: S.A. Zenios and W.T. Ziemba

North-Holland

HANDBOOK OF
ASSET AND LIABILITY MANAGEMENT
VOLUME 2:
APPLICATIONS AND CASE STUDIES

HANDBOOKS IN FINANCE

Series Editor

WILLIAM T. ZIEMBA

Advisory Editors

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

North-Holland is an imprint of Elsevier

N·H

HANDBOOK OF ASSET AND LIABILITY MANAGEMENT

VOLUME 2: APPLICATIONS AND CASE STUDIES

Edited by

S.A. ZENIOS
University of Cyprus, Nicosia, Cyprus
and
The Wharton Financial Institutions Center, Philadelphia, PA, USA

W.T. ZIEMBA
University of British Columbia, Vancouver, Canada

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

North-Holland is an imprint of Elsevier

N·H

North-Holland is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2007

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-444-52802-5
ISSN: 1872-0978

For information on all North-Holland publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

INTRODUCTION TO THE SERIES

Advisory Editors:

Kenneth J. Arrow, Stanford University, George C. Constantinides, University of Chicago, Harry M. Markowitz, University of California, San Diego, Robert C. Merton, Harvard University, Stewart C. Myers, Massachusetts Institute of Technology, Paul A. Samuelson, Massachusetts Institute of Technology, and William F. Sharpe, Stanford University.

The Handbooks in Finance are intended to be a definitive source for comprehensive and accessible information in the field of finance. Each individual volume in the series presents an accurate self-contained survey of a sub-field of finance, suitable for use by finance and economics professors and lecturers, professional researchers, graduate students and as a teaching supplement. The goal is to have a broad group of outstanding volumes in various areas of finance.

William T. Ziemba
University of British Columbia

This page intentionally left blank

CONTENTS OF THE HANDBOOK

Volume 2
Applications and Case Studies

This page intentionally left blank

PREFACE

God may not be playing dice with nature, according to Einstein's famous quip, but chance and spontaneity are unavoidable in the financial markets. This simple but profound fact is driven eloquently in Bernstein's 1996 book. Organizations operating in the financial markets—be they financial institutions, manufacturing or distribution firms, or service providers—cannot achieve their main goals of creating economic value unless they come to terms with the risks associated with the uncertainties of the financial markets (Harker and Zenios, 2000). The alignment of a firm's operations and tactics with its uncertain environment is today perceived as a key strategy for all businesses. It draws the attention and demands heavy investment of resources by chief executives and boards of directors worldwide. The management of risky assets and uncertain liabilities in an integrated, coherent, framework not only remains the core problem of financial institutions today, but it has spilled over to other business enterprises as well.

The foundations for addressing today's challenges were laid more than fifty years ago in the Doctoral dissertation by Harry M. Markowitz (Markowitz, 1952, 1991) at the University of Chicago. This work laid the foundations for modern finance and was recognized by a Nobel prize in Economics in 1990. The early use of Markowitz's optimization theories was in developing normative models for understanding the financial markets, and as theoretical tools in financial economics. Since the 1980s however this line of research also evolved from a theoretical tool of positive analysis to a practical tool for normative analysis (Zenios, 1993). Optimization models are today at the core of decision support systems for financial engineers. The drive to integrate multiple interrelated risk factors of the global enterprise brought to the fore the power of asset and liability management models. At the same time developments of large-scale numerical optimization techniques, advances in optimization models for planning under uncertainty, and the availability of user-friendly modelling languages, put optimization tools in the hands of researchers and practitioners alike.

Thus, the tools of asset and liability management have flourished. And the symbiosis between optimization tools and financial decision-making is becoming more fertile as we enter the 21st Century marked by business globalization, rapid technological changes, financial innovations, and increased volatility in the financial markets.

Needless to say, the optimization models used in asset and liability management have been extended significantly—and in many cases deviated substantially—from the way shown by the pioneers of the fifties. The use of multi-period stochastic programming being perhaps the single most noteworthy generalization of the early works (Ziemba and Vickson, 1975, 2006). Indeed, the proliferation of models for practical asset and liability modeling has been vast and witnessed by the sample of research articles collected in

Ziemba and Mulvey (1998) or Wallace and Ziemba (2005), and the discussion of the practical use of these models in Ziemba (2003), and Ziemba and Ziemba (2007).

It is therefore fitting that the series ***Handbooks in Finance*** devotes a *Handbook to Asset and Liability Management*. What may come as a surprise is that it took two volumes to collect what we perceive as the essential contributions of the last fifty years. Volume 1 contains chapters that lay the theoretical foundations and develop the methodologies essential for the development of asset and liability management models. Volume 2 considers several diverse business settings and a chapter devoted to each discusses problem-specific issues and develops realistic asset and liability management models. While all applications are drawn from financial institutions, readers interested in other business settings will find in both volumes sufficient material to gain deep insights into the asset and liability management modeling of other types of enterprises. The coverage is broad both in methodology and applications with chapters on term and volatility structures, interest rates, risk-return analysis, dynamic asset allocation strategies in discrete and continuous time, the use of stochastic programming models, bond portfolio management and the Kelly capital growth theory and practice in Volume 1. Volume 2 discusses applications of ALM models in banking, insurance, money management, individual investor financial planning, pension funds and social security.

We would like to thank all the authors for contributing chapters that address some aspect of asset and liability modeling that goes beyond the authors' own research contributions to the field. Having asked leading researchers to contribute each chapter we have been able to present the state-of-the-art in the field, while no efforts were spared in making the chapters accessible to a wider audience and not being restricted to the cognoscenti. And when a chapter may err on the side of focusing somewhat more narrowly on a specific research direction—dictated by the authors' preferences—extensive bibliographies at the end of each chapter point readers to the vast fields beyond.

We hope that this collection of chapters and their references will be an invaluable resource for practitioners and the regulators of financial institutions, for researchers in the fields of finance and financial engineering, scholars in optimization and mathematicians, and both doctoral and masters students.

Stavros A. Zenios
University of Cyprus and Wharton Financial Institutions Center

William T. Ziemba
University of British Columbia

References

Bernstein, P.L., 1996. Against the Gods: The Remarkable Story of Risk. John Wiley and Sons, New York, NY.

Harker, P.T., Zenios, S.A. (Eds.), 2000. Performance of Financial Institutions: Efficiency, Innovation, Regulations. Cambridge University Press, Cambridge, UK.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7, 77–91.

Markowitz, H.M., 1991. Portfolio Selection: Efficient Diversification of Investments, second ed. Blackwell, Oxford, UK.

Wallace, S.W., Ziemba, W.T. (Eds.), 2005. Applications of Stochastic Programming. SIAM–MPS, Philadelphia, PA.

Zenios, S.A., 1993. Financial Optimization. Cambridge University Press, Cambridge, England.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset Liability and Wealth Management. AIMR, Charlottesville, VA.

Ziemba, W.T., Mulvey, J.M. (Eds.), 1998. Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK.

Ziemba, W.T., Vickson, R.G. (Eds.), 1975. Stochastic Optimization Models in Finance. Academic Press; second ed. World Scientific, 2006.

Ziemba, R.E.S., Ziemba, W.T., 2007. Scenarios for Risk Management and Global Investment Strategies. Wiley, July.

CONTENTS

Chapter 11

ALM IN BANKING

JEAN DERMINE*

INSEAD, Fontainebleau, France
e-mail: jean.dermine@insead.edu

Contents

* The author is grateful to three anonymous referees for helpful comments, and to J. Cropper for editorial assistance.

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02011-4

Abstract

The main purpose of the chapter is to discuss *Asset & Liability Management*, the control of value creation and risks in a bank. This chapter is innovative in two ways. First, unlike the usual practice of restricting ALM to the control of interest rate and liquidity risks, we propose a framework to analyze both value creation and the control of risks. Second, rather than discuss the ALM issues one by one in an independent manner, the chapter provides a microeconomic-based valuation model of a bank. This allows an integrated discussion of fund transfer pricing, deposit pricing (fixed and undefined maturities), loan pricing, the evaluation of credit risk provisions, the measurement of interest rate risk for fixed and undefined maturities, the diversification of risks, and the allocation of economic capital.

Besides a comprehensive summary of the literature on ALM in Banking, the chapter makes six contributions related to transfer pricing, risk-adjusted pricing of loans, provisioning of credit risk, the relevant maturity to price and hedge deposits with uncertain maturities, the after-tax valuation of equity, and the hedging of economic profit.

Keywords

ALM, bank valuation, credit risk, interest rate risk, liquidity risk

JEL classification: G21, G28

Introduction

The main purpose of this chapter is to discuss Asset & Liability Management, the control of value creation and risks in a bank. The chapter aims to be comprehensive with a large coverage of the ALM literature, and to be innovative in two ways. First, unlike the usual practice of restricting ALM to the control of interest rate and liquidity risks arising from positions on balance sheet (the banking book),[1] we propose a framework to analyze both value creation and the control of risks. Second, rather than discuss the ALM issues one by one in an independent manner, we provide a microeconomic-based valuation model of a bank. It allows us to discuss, in an integrated way, fund transfer pricing, deposit pricing (fixed and undefined maturities), loan pricing, the evaluation of credit risk provisions, the measurement of interest rate risk for fixed and undefined maturities, the diversification of risks, the marginal risk contribution, and the allocation of economic capital. The traditional purpose of ALM, the control of interest rate and liquidity risks, is thus integrated into a richer framework.

With reference to the organizational structure of banks, our integrative approach is closer to the organization of a J.P. Morgan Chase which operates with two major corporate risk committees: Capital and Risk Management. The Capital Committee reviews the adequacy of the firm's capital and liquidity, and recommends the allocation of capital within the firm. The Risk Management Committee provides oversight and direction of risk profile and risk appetite, and reviews and approve corporate policies and risk strategies in a comprehensive way, not restricted to liquidity or interest rate risk on the banking book.

In Section 1, five specific but interrelated functions of banks are discussed in the light of modern banking theory. This permits us to identify the various services provided by banks. The balance sheet and income statement of a representative bank are presented in Section 2. As bank modeling is often concerned with the management of risks, fifteen sources of risk in banking are identified in Section 3, and the economics of risk management are discussed. Two microeconomic models of the banking firm are developed in Section 4. A neoclassical model facilitates discussion of the bank separation theorem and the pricing of deposits with fixed and undefined maturities. A bank valuation model enables us to break the value of the equity of a bank into four components: a liquidation value, a franchise value, a corporate tax penalty, and tax savings due to unrealized capital gains. Specific attention is given to relevant risk-adjusted discount rates to value bank assets and liabilities. Three applications of the model follows. In Section 5, the valuation model is applied to the pricing of risky loans and the fair evaluation of credit risk provisions. The discussion of the measurement of interest rate and liquidity risks in Section 6 will concern the risk on the banking book from both an 'accounting earnings' and an 'economic value' perspectives. Both finite maturity-products, such as term loans or term deposits, or products with undefined maturities, such as demand deposits

[1] For instance, Crouhy, Galai and Mark (2001) and Bessis (2002).

or credit card loans, will be analyzed. We discuss, in Section 7, the aggregation of risks, the concept of marginal risk contribution, and the allocation of economic capital. Finally, in Section 8, we review briefly the rationale for bank regulation and the main types of regulations, as they relate to capital adequacy, interest rate risk, and liquidity risk.

Besides a comprehensive summary of the literature on *ALM in Banking*, this chapter makes six contributions related to transfer pricing, risk-adjusted pricing of loans, provisioning of credit risk, the relevant maturity to price and hedge deposits with uncertain maturities, the after-tax valuation of equity, and the hedging of economic profit.

1. Economics of banking, five main functions

A bank is a firm whose assets include primarily financial claims issued by borrowers, such as households, corporate firms, governments, and other financial intermediaries, and whose liabilities are sold as secondary claims to capital surplus units in various forms, such as demand deposits, savings deposits, term deposits, subordinated debt (loan capital), or equity shares. Keeping up with financial innovations, banks engage in various credit insurance-related activities, such as letters of credit, note-issuance facilities, or credit derivatives. Others types of contingent claims include financial derivatives, such as forwards, options or swaps, the payoffs of which are related to movements in interest rate, exchange rates, equity or commodity prices. With the exception of the transaction cost and the cash premium received or paid, these activities do not create an asset or a liability on the balance sheet. They belong to the *off-balance sheet* activities. Although the services provided by banks are interrelated, it is convenient to distinguish five categories of increasing complexity: underwriting and placement, portfolio management, payment (transmission) services, monitoring or information-related services, and risk sharing.

Underwriting and placement. A first service provided by financial intermediaries is to bring together savers and borrowers. Underwriting and placement of securities is a function which helps borrowers (corporate firms or public institutions) to meet surplus units, and structure or customize the type of securities that meet the risk/return requirements of borrowers and lenders. In this function, the underwriter is involved not only in designing the security, but also in the valuation of assets and the pricing of securities to ensure that the terms of the issue are competitive. Increasingly, rating agencies play a crucial role in providing independent evaluation of the risks incurred on these claims. As investors may wish in the future to transform these claims into cash, consumption or other securities, they need to be exchanged. Brokers/dealers or market makers provide these services to ensure secondary trading and liquidity. In a pure underwriting and placement service, it is assumed that the return and risk of the securities can be properly defined, so that there is no major problem of asymmetric information (agency problem) between lenders and borrowers. In this case, monitoring is not an issue. A pure case is the financing of public debt in countries where the sovereign risk is minimal. With the underwriting and placement service, the end-investor holds directly the claims on deficit units.

Portfolio management. At low cost, investors can acquire a diversified portfolio of securities issued by deficit spending units. The pure case is the mutual fund or unit trust (called SICAV in France and Luxembourg) which supplies a diversified portfolio to the holders of its shares. The income derived from the financial assets is paid to the holders of the shares less a fee paid to the fund manager.[2] The reason for the existence of these funds is two-fold. The first is to reduce the divisional cost incurred in issuing many securities. The second is that investors may wish to delegate the assessment of economic prospects and fund management to specialists.

Payment mechanism. A third function performed by financial markets is the management of the payment system, i.e., to facilitate and keep track of transfers of wealth among individuals. This is the bookkeeping activity of banks realized by debiting and crediting accounts. Although the payment system is limited by regulation to a specific type of deposits (demand deposits), it could be achieved by debiting or crediting any type of liquid assets. The so-called cash management or sweep account which automatically transfers money from mutual funds into demand deposits is a perfect illustration of the possibility of extending the payment system to other assets.

Monitoring and information-related services. Private information held by borrowers leads to contracting problems, because it is costly to assess the solvency of a borrower or to monitor his/her actions after lending has taken place (Stiglitz and Weiss, 1981). Sometimes, it is useful to package these claims in a portfolio, and banks perform a useful function in reducing the costs of screening and monitoring borrowers. The delegation of screening and monitoring to banks has been shown to be an efficient mechanism.[3] This fourth service is linked to the first one, underwriting and placement. It is taken here as a separate service as it corresponds to those cases where significant information asymmetries make it difficult to issue financial claims traded on securities markets. While the second service, portfolio management, refers to the management of liquid assets, this fourth function refers to the management of the credit portfolio, most often the far larger part of a bank's balance sheet.

Risk-sharing service. An increasingly important function of banks is to make the market more complete, i.e., to provide tools to transfer money (consumption) across states

[2] See Black (1970), Fama (1980), and Dermine, Neven and Thisse (1991) for a portfolio view of financial firms.

[3] Diamond (1984) and Fama (1985).

of the world. Several examples will be illustrated. First, banks not only supply diversified assets, but also organize efficiently the distribution of risky income earned on the asset pool. The debt holders receive a fixed payment while the shareholders receive the residual income. Other insurance services include interest rate insurance (floating rate lending with various ceilings on interest rates called caps or floors), inflation insurance with real contract, and liquidity insurance, option for deposit holder or the holder of a line of credit to withdraw quickly at face value (Diamond and Dybvig, 1983; Rajan, 1998). Allen and Santomero (1998, 2001) have emphasized the growing importance of risk management services provided by commercial banks.

Generic economic functions of banks have been presented. The Second Banking directive[4] of the European Commission lists the specific activities that can be authorized by central banks in the European Union:

- Deposit-taking and other forms of borrowing.
- Lending.
- Financial leasing.
- Money transmission services.
- Issuing and administering means of payments (credit cards, travelers' cheques and bankers' drafts).
- Guarantees and commitments.
- Trading for own accounts or the account of customers in:
  Money market instruments,
  Foreign exchange,
  Financial futures and options,
  Exchange and interest rate instruments,
  Securities.
- Participation in share issues and the provision of services related to such issues.
- Money broking.
- Portfolio management and advice.
- Safekeeping of securities.
- Credit reference service.
- Safe custody service.

This complete list describes the activities of a *universal bank*. In some countries, such as the United States or Japan, the list of permissible activities was greatly reduced (Saunders and Walter, 1994). But there has recently been a regulatory convergence towards the *universal banking* model. For instance, the Financial Modernization Act (Gramm–Leach–Bliley) of 1999 in the United States has repealed the Glass–Steagall Act which separated commercial banking and securities underwriting. The banking systems of most countries of Latin America and Central and Eastern Europe were deregulated with the adoption of the universal banking model.

[4] Directive 89/646/EEC.

2. The bank's balance sheet and income statement

Before discussing modeling, it is useful to present the balance sheet and income statement of a representative bank. The consolidated balance sheet presented in Table 1 is that of the Royal Bank of Canada for the year ended October 31, 2000.

The consolidated Income Statement of the Royal Bank of Canada for the year ended October 31, 2000 follows (see Table 2).

With respect to the five functions of banks discussed earlier, one can observe that non-interest revenues, originating from trading of securities, management of the payment system, or fund management, exceeds net interest income, respectively, Can$ 6,680 million and Can$ 5,279 million. Over the years, banks have diversified their sources of revenue, and the share of fee-based services has increased.

To model the bank, we very much simplify the balance sheet and income statement. The balance sheet of a bank used for modeling purpose is as follows:

Assets	**Liabilities and shareholders' equity**
Reserves with Central Banks	Retail deposits
Retail loans	Demand deposits
	Savings deposits
	Term deposits
Corporate loans	Corporate deposits
	Demand deposits
	Term deposits
Interbank loans	Interbank deposits
Government bonds	Subordinated debt
Fixed assets	Equity

A bank's simplified balance sheet.

Banks collect deposits on the retail and corporate markets. Some of these are withdrawable on demand (demand deposits used for the payment system). Others, such as savings deposits, can be transferred into other deposit accounts on demand. The maturities of these first two types of deposits are said to be 'undefined', as deposits could stay in the bank for a few days or a few years. The undefined maturity creates a specific problem to price these deposits, and to measure their interest rate and liquidity risks. These issues will be addressed in Sections 4 and 6. Others deposits, the term deposits, have a fixed contractual maturity. Finally, the pricing could be a 'fixed rate', or a 'floating rate' linked to a short-term benchmark rate, such as the interest rate on government treasury bills or the interbank rate. Besides money raised form the public, banks borrow from one another on the interbank market with interbank deposits. Finally, two sources of long-term funds include subordinated bonds and equity. Subordinated debt plays a special role in banks as they qualify, along with equity and upon some eligibility limits, for the definition of regulatory capital.

On the asset side, banks must hold some reserves at the central bank, reserves which pay a low (often zero) interest rate. As discussed in Section 1, a main function of banks

Table 1
Consolidated balance sheet

Assets (Can$ million)		Liabilities and shareholders' equity (Can$ million)	
Cash resources		**Deposits**	
Cash and due from banks	947	Canada	
Interest-bearing deposits with other banks	18,659	Non-interest bearing	22,011
		Interest bearing	116,113
	19,606	International	
		Non-interest bearing	863
		Interest bearing	67,250
Securities		**Other**	
Trading account	46,366	Acceptances	11,628
Available for sale	13,199	Obligations related to securities sold short	12,873
Held to maturity	698		
	60,263	Obligations related to assets sold repurchase agreements	9,005
		Derivative-related amounts	18,574
		Other liabilities	15,912
			67,992
Assets purchased under repurchase agreements	18,303	**Subordinated debentures**	5,825
Loans		**Non-controlling interest in subsidiaries**	703
Residential mortgage	62,984		
Personal	28,019		
Credit card	4,666		
Business and govern. loans and acceptances	72,143		
	187,812		
Allowance for loan losses	(1,871)		
	165,941		
Other		**Shareholders' equity**	
Derivative-related amount	19,334	Capital stock	
Premises and equipment	1,216	Preferred	2,001
Goodwill	693	Common	3,074
Other intangibles	208	Retained earnings	8,314
Other assets	8,490	Accumulated other comprehensive income	(92)
	29,941		13,297
Total assets	294,054	**Total liabilities and shareholders' equity**	294,054

Source: Royal Bank of Canada, 2000 Annual Report.

is to lend money to individuals or corporations. Excess funds can be lent to other banks (interbank loans) or used to purchase government bonds. Finally, a small proportion of funds is used to purchase fixed assets, such as buildings and computers. The mismatch

Table 2
Consolidated income statement

Interest income (Can$ million)	
Loans	11,538
Trading account securities	1,435
Available for sale and held to maturity securities	1,083
Assets purchased under reverse repurchase agreements	1,078
Deposits with banks	975
	16,109
Interest expense	
Deposits	9,057
Other liabilities	1,429
Subordinated debentures	344
	10,830
Net interest income	5,279
Provisions for credit losses	691
Net interest income after provisions for credit losses	4,588
Non-interest revenue	
Capital market fees	1,810
Trading revenues	1,540
Deposit and payment service charges	756
Investment management and custodial fees	684
Mutual funds revenues	528
Credit card revenues	420
Securitization revenues	104
Gain (loss) on sale of securities	(11)
Other	849
	6,680
Non-interest expenses	
Human resources	4,695
Occupancy	570
Equipment	664
Communications	695
Other	1,004
	7,628
Net income before income taxes	3,640
Income taxes	1,412
Net income before non-controlling interest	2,208
Non-controlling interest in net income of subsidiaries	20
Net income	2,208
Preferred share dividends	(134)
Net income available to common shareholders	2,074

Source: Royal Bank of Canada, 2000 Annual Report.

between the maturities of assets and liabilities can create interest rate and liquidity risks. This is discussed in Section 6.

In addition to balance sheet items, banks are involved into a large set of off-balance sheet activities, such as derivatives (forward rate agreements, options or swaps) or loan commitments and guarantees. They are called off-balance sheet or contingent claims, as, except for incurring a small transaction cost or a premium, they have no impact on the balance sheet at origination. They will create a cash flow (positive or negative) when some contingency occurs.

The simplified income statement used for modeling purpose is as follows:

	Interest income
+	Fee/trading income
–	Interest expense
–	Loan provisions
–	Operating expense
	Earning before tax
–	Tax
	Earnings after tax

A bank's simplified income statement.

Interest income is earned on loans and bonds, while fee/trading income is generated by various services or trading. Interest expenses include the cost of deposits and subordinated debt. A very specific and difficult issue in banking concerns the creation of loan provisions to take into account the loss of value of loans. As loans are usually not traded on capital markets, there is no information readily available on what is the fair value of a loan. Moreover if the borrower goes into difficulties and bankruptcy proceedings, it may take several years to know the exact amount of loan losses. As a consequence, a method must be devised to set the *fair* level of provisions for credit risk. This is discussed in Section 5.

3. Risk management in banking

An identification of the sources of risk in banking is followed by a discussion of the economics of corporate risk management.

3.1. Risks in banking

At least fifteen sources of risk can be identified in banking. They can be grouped into four major categories: credit, market, liquidity, and operational risks. A comprehensive, and possibly joint, management of these sources of risks is referred to as ERM, Enterprise-wide Risk Management (Rosen and Zenios, 2006).

- Credit risk
  - Retail and corporate credit risk

  - Counterparty risk
  - Settlement risk
  - Environmental risk
  - Country risk
- Market risk
  - Interest rate risk
  - Foreign exchange risk
  - Equities
  - Commodities
- Liquidity risk
- Operational risk
  - Execution risk
  - Model risk
  - Fraud
  - Legal risk
  - Regulatory risk

Credit risk refers to the non-payment on time by a retail, corporate, or institutional borrower of interest and/or principal. **Counterparty risk** refers to a particular type of credit risk, in which the borrower is a financial institution. **Settlement risk** (sometimes referred to as *Herstatt* risk)[5] is a particular type of counterparty risk. It refers to the risk involved in selling securities or foreign exchange. A time difference in settlement dates could imply that one party has already delivered the security before the payment is completed. **Environmental risk** is a type of credit risk in which the guarantee on a loan contract may force the bank to hold real assets with some environmental liability. **Country risk** refers to the potential losses that could arise when a country, facing a severe economic crisis, takes actions that are detrimental to the bank (such as nationalization, increases in taxes, or capital controls).[6]

Market risk refers to the loss of revenue due to adverse movement in interest rate, foreign exchange, and prices of securities or commodities. With specific regard to interest rate risk, one makes a distinction between the banking book and the trading book. Due to accounting rules, the banking book is accounted for on an accrual basis (i.e.,

[5] The medium-sized German Bankhaus Herstatt defaulted on 26 June 1974 at the end of the business day. Some of the bank's counterparties had irrevocably paid Deutsche Marks to the bank before the bank's license was withdrawn. These counterparties were expecting to receive US dollars in New York that same day. However, the termination of the bank took place at 10:30 am in New York, prompting Herstatt's correspondents to suspend all outgoing US dollar payments. It left US counterparties with losses exceeding $600 million.

[6] An example of country risk is that of Argentina on January 6, 2002 when the Parliament approved the end of the currency-board system (a one-to-one parity between the peso and the US dollar). A key-measure of the plan was to convert the banks' dollar loans into pesos at a one-to-one exchange rate, while deposits were converted at 1.4 pesos to the dollar. This created a large and unexpected currency mismatch between assets and liabilities, which increased even further in 2003. As the peso plunged to 3.18 pesos to the dollar, 180,000 depositors filed lawsuits again the decree. On March 5th, the Supreme Court ruled that the conversion of deposits to pesos was illegal, allowing depositors to claim dollars (The Economist, March 8th, 2003).

assets and deposits are recorded at acquisition cost). The banking book generates a net interest income (net interest margin) in the profit & loss account (P&L). A first source of interest rate risk concerns therefore the volatility of the net interest income. The trading book is marked-to-market, either at actual market prices when these are available, or at calculated *fair* present values. The change in value of the trading book is recorded in the P&L. A second source of interest rate risk is the impact of interest rate on the value of the trading book.

Liquidity risk refers to the shortage of cash originating from a loss of bank deposits, unexpected draw downs on loan commitments, or margin calls on trading transactions.

Operational risk, in its widest definition, includes every risk other than credit, market and liquidity risks. It is the risk of loss resulting from inadequate or failed internal processes, people and systems or external events. More specifically, **execution risk** refers to losses due to data entry errors[7] or computer failures. **Model risk** refers to losses incurred when the mathematical modeling of financial instruments does not match movements in actual market prices. **Fraud** refers to outright stealing of value by employees or clients. **Legal risk** involves unexpected losses due to legal liabilities. **Regulatory risk** refers to the losses arising from an unexpected change in regulations, such as more stringent capital requirements.

Sometimes, a specific loss (such as trading or credit losses) can generate a much larger fall in the market value of the shares of a bank than the loss itself. This is often explained by a loss of confidence in the management of a bank. The "blowing up" of an initial loss, arising from credit, market, liquidity or operational risks, is referred to as **reputational risk**. It is particularly acute in banking because of the *opacity* of bank operations. Given the very large number of transactions, sometimes in many countries, and the holding of non-tradable instruments, it is very difficult for an outsider to verify the quality of the management of a bank. A loss can therefore be a signal of bad quality, which explains the amplified fall in market value.

3.2. The economics of risk management

In a world with perfect information and complete markets, corporate firms, for instance banks, should not allocate resources to risk management. Perfectly informed shareholders could manage themselves an optimal reduction of risks. Four motivations for corporate risk management have been advanced in the literature (e.g., Santomero, 1995; or Froot and Stein, 1998): managerial self-interest, non-linearity of taxes, cost of financial distress, and capital market imperfections. These are discussed briefly.

First, managerial self-interest refers to the fact that managers, having a significant fraction of their permanent income attached to the firm, cannot diversify risks adequately. Managers' risk aversion will lead to risk mitigation. Second, the non-linearity

[7] On May 14, 2001, a trader of an investment bank in London keyed in the wrong number of shares on his/her trading screen. The mistake meant that a sell order on a basket of shares worth a reported US$ 30m turned into one valued ten times that. The FTSE 100 Index fell by 2.2% when the trade was made (The Economist, May 19, 2001).

of taxes means that losses may not be fully tax-deductible, or that large profits could be taxed at a higher rate. In this case, a reduction of profit variance leads to a reduction of expected tax payments. Third, the cost of financial distress refers to the loss of value due to a state of distress. In banking, this could imply a loss of clientele or a loss of a profitable banking license (the 'charter value'). Fourth, costs may arise from capital market imperfections. Because of asymmetric information, banks may find it costly to raise external funds. In such a context, losses could lead to a lower equity level, and profitable investment opportunities which are missed. Stabilization of profit can reduce the call for expensive external finance, and lead to the realization of profitable investments. An alternative explanation of the resources spent on risk control is linked to reputational risk. Because of opacity, investors cannot evaluate whether a reported loss is due to bad luck or to inferior management quality. In this context, stabilization of profit prevents a loss of value. So, even in the absence of bank regulation, there are several economic motivations for the control of risks in a bank.

4. Asset and liability modeling for banks

The modeling approach presented below is rooted in the microeconomics of banking and in finance theory. A neoclassical economic model of the banking firm is followed by a valuation-based model.

4.1. A neoclassical model of the banking firm and the separation theorem

The simplest model is the neoclassical model of the banking firm developed by Klein (1971) and Monti (1972).[8] The asset side of the balance sheet of the bank consists of reserves with the central bank (R), loans (L), and market-traded assets such as government bonds (B) or interbank loans. The liability side includes deposits (D) and equity (E). The regulatory reserves, yielding no interest, are a fraction r of deposits. The supply of government bonds (yielding an interest rate b) is perfectly elastic in competitive markets. The balance sheet is as follows:

Asset	Liabilities and shareholders' equity
Reserves	
Loan	Deposits
Bonds	Equity

The loan demand by borrowers ($L(.)$) is a decreasing function of the interest rate p and the deposit supply ($D(.)$) is an increasing function of the interest rate d. All these assets and deposits have the same maturity, say one year and, at this stage, all parameters

[8] See extension of the neoclassical model in Baltensperger (1982), Santomero (1984), Dermine (1984, 1986), and Freixas and Rochet (1997).

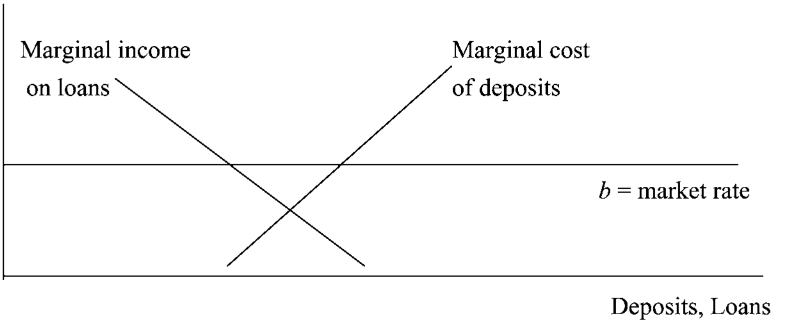


Fig. 1. The separation theorem.

are known with certainty. Operating expenses linked to deposits and loans are left out for simplicity.

The opportunity cost of equity in this certain world is the exogenous government bond rate b. The bank chooses deposit and credit rates to maximize its end-of-period economic profit (EP), that is the accounting profit reduced by an opportunity cost of equity:

$$\text{Max}\, EP = (p \times L + b \times B - d \times D) - b \times E \tag{1}$$
$$\text{s.t.} \quad R + L + B = D + E.$$

Substituting the balance sheet constraint into the objective function, one has:

$$\text{Max}\, EP = \big[(p - b) \times L\big] + \big[\big(b \times (1 - r) - d\big) \times D\big]. \tag{2}$$

The economic profit is the sum of two terms: income on loans net of an opportunity cost (the government bond rate b), and income on deposits invested in securities with a return reduced by the central bank's reserve requirement. For each Euro raised, only $(1 - r) \times$ Euro is available for investment. This simple relation has given rise to the important *separation theorem* in banking, which suggests pricing loans and deposits independently, with reference to the market rate (government bonds or interbank rate). The first order condition for deposit and credit rates are:

$$\partial EP/\partial d = \big(b \times (1 - r) - d\big) \times D' - D = 0, \tag{3}$$
$$\partial EP/\partial p = (p - b) \times L' + L = 0. \tag{4}$$

Denoting by η_D and η_L the interest rate elasticity of deposits and loans, one obtains:

$$d = b \times (1 - r) \times \big(1 + \eta_D^{-1}\big)^{-1}, \qquad p = b \times \big(1 + \eta_L^{-1}\big)^{-1}. \tag{5}$$

This can be represented in the following graph, the reserve requirement being set at zero, for the sake of exposition (Figure 1).

The horizontal line represents the perfectly elastic market rate b. The other two lines represent the marginal income on loan and the marginal cost of deposits. The graph illustrates that the optimal volume of deposits is reached when the marginal cost of deposits is equal to the opportunity market rate. Similarly, the optimal volume of loans is reached when the marginal revenue from loans is equal to the marginal investment return, the market rate b. One notices the separation between the lending and funding decisions. Knowledge of the market rate, the reserve requirement and the price elasticities allows one to choose the optimal interest rates on deposits and loans. The difference between the optimal volumes of deposits and loans is the net position in market assets. In the graph, it is positive with the bank being a net lender in the market. But it could be negative with the bank being a net borrower.

The bank's separation theorem is useful in practice for two reasons. First, it allows us to identify the relevant interest rate to price deposits or loans. Second, it permits the division of the bank into a set of profit centers. Indeed the economic profit in Eq. (2) is shown to be the sum of two terms: the profit (net interest margin) on loans and the profit on deposits. One can therefore break up the bank into different sources of economic profit, or into different value centers, the transfer price needed to evaluate the profit on deposits or loans being identified as the market rate with the same maturity as the product. This rate is known as the MMMVF, the matched-maturity marginal value of funds. The expression 'breaking up the bank' was used to refer to the fact that the deposit-taking and lending activities could be, each, evaluated on its own.

Given the wide application of the MMMVF in banking, a natural question arises as to which economic factors would break the bank separation theorem To break the theorem, one simply needs to introduce a joint cost or revenue function in deposits and loans. A joint revenue function could exist if the volume of deposits received is linked to the volume of loans granted. For instance, the terms of mortgage loans could impose the opening of deposit accounts with the same bank. The joint cost function could have two origins: there could be joint operating expenses in delivering deposits and loans, and, if there is a cost due to a liquidity crisis, it will be linked to the joint amount of deposits and illiquid loans. Indeed, *ceteris paribus*, if there is a large volume of loans, there will be a low volume of liquid government bonds that can be sold to meet deposits withdrawals. One must realize that *joint maximization* of economic profit on loans and deposits will be the rule in the case of a joint cost or revenue function. The simplicity of the separation theorem disappears in this case. Joint maximization is particularly relevant for banking systems experiencing significant liquidity risks. Cases are to be found not only in emerging countries, but also in countries in which the deposit base of banks has been eroded by competing financial products, such as money market funds, mutual funds, or pension funds. An additional argument for joint maximization, risk diversification, will be discussed in Section 7.

4.2. A multi-period neoclassical model of the banking firm

In the one-period model, the relevant transfer price for pricing and for the evaluation of profitability is the matched-maturity market rate. A six-month interbank rate is the relevant rate to price a six-month deposit. However, two additional features can introduce a multi-period consideration (Dermine, 1984; Hannan and Berger, 1991). The first is that the supply of deposits can be a lagged function of past deposit rates. This is likely to be the case in the retail sector in which customers, facing switching costs, will display some form of loyalty. This creates a lag in the deposit supply function. The second reason for dynamic consideration is that, for marketing reasons, it can be costly to continuously adjust the deposit rate. One speaks of deposit rate rigidity, or "stickiness".[9]

Let us consider a two-period model with b_1 the bond rate in Year 1, and b_2 the forward rate in Year 2.

Let us first consider the lagged supply case, i.e.,

$$D_2 = D_2(d_2, d_1).$$

In this case, the maximization of profit on deposit over two periods becomes:

$$\text{Max}(1 + b_2) \times (b_1 - d_1) \times D_1(d_1) + (b_2 - d_2) \times D_2 \times (d_1, d_2). \tag{6}$$

The optimal condition for the deposit rate becomes:

$$d_t \times \left(1 + \eta_t^{-1}\right) = b_1 + \frac{(b_2 - d_2)}{1 + b_2} \times \frac{\partial D_2/\partial d_1}{\partial D_1/\partial d_1}. \tag{7}$$

One keeps the equality between the marginal cost of deposits and the marginal return. The marginal return incorporates not only the current matched-maturity market rate, but also the marginal profit on next-period deposits.

The second dynamic case concerns the situation of interest rate rigidity or stickiness. This is particularly relevant for products such as savings deposits. Assume that the interest rate is fixed over two periods, then:

$$\text{Max}(1 + b_2) \times (b_1 - d_1) \times D_1(d_1) + (b_2 - d_1) \times D_2(d_1). \tag{8}$$

The first-order condition is:

$$d_1\left[D_1'\left(1 + \eta_1^{-1}\right) + D_2'\frac{1 + \eta_2^{-1}}{1 + b_2}\right] = D_1' b_1 + D_2'\frac{b_2}{1 + b_2}. \tag{9}$$

The elasticity (return) of the single period optimum is replaced by a weighted sum of current and future elasticities (return), the weights being the discounted supply derivatives.

An important conclusion for pricing is that, once dynamic features are introduced in the neoclassical model, it becomes clear that the appropriate marginal income on savings deposits is not the short-term rate, but rather a weighted sum of current and future interest rates. On operational grounds, the choice of the relevant maturity for the benchmark becomes important in the case of non-flat yield curves. This is clarified in

[9] Rate stickiness is only discussed in the context of deposit pricing. A similar reasoning applies in the case of other products with rigid pricing, such as credit card loans (Ausubel, 1991).

Appendix A, in which I argue that one should distinguish among three separate applications for the choice of the relevant benchmark maturity: the pricing of products, the choice of the hedge instruments, and the selection of a transfer price to evaluate *ex post* the performance of the branch.

4.3. A valuation model of the banking firm: no tax, no risk, no growth

The neoclassical microeconomic model of the banking firm presented above provides some useful tools to price deposits and loans, and to evaluate *ex post* the profitability on loans or deposits. From a finance perspective however, it is not totally satisfactory, as it is not rooted in valuation. In the light of the fact that management wants to maximize shareholder value, one would need a valuation-based modeling approach for the banking firm. The valuation model of the banking firm will be presented step by step, from simple cases with no tax, no risk, no growth, to more complete models with tax and risk. It is based on Dermine (1985a, 1987), and Dermine and Hillion (1992).

We assume, for expository convenience, that a bank has two more years to survive before it is liquidated. The bank has, on its asset side, a portfolio of two-year-to-maturity loans (L) and bonds (B) which have been acquired in the past and which carry fixed interest rates, respectively, p and b. These assets are financed with deposits (D) and equity (E). Deposits have a maturity of two years, offering a fixed interest rate d. Loans, bonds, and deposits are recorded at their historical book values, L, B, and D, respectively. The current one-year return on similar assets and liabilities are, respectively, p^*, b^*, and d^*, constant over the next two years. Since this section does not focus on risk, all the variables are to be taken as certain or certainty equivalent. As in the neoclassical model outlined earlier, the bonds and equity markets are assumed perfectly elastic, and the (certainty equivalent) cost of equity is the current market bond rate b^*. The balance sheet of the bank is given below:

Loans $L(p)$	Deposits $D(d)$
Bonds $B(b)$	Equity $E(b^*)$

Notes: Two-year-to-maturity fixed rate assets and deposits.
Historical rates on loans, bonds and deposits are p, b, and d.
Current rates on loans, bonds and deposits are p^*, b^*, and d^*.

There can be many reasons as to why we observe an interest rate differential between assets and liabilities. The longer maturity of assets may command a risk premium and, with deposits withdrawable at short notice, the posted deposit rate does not include the extra cost of refinancing the bank in the case of deposit withdrawals. However, this does not explain the differential in the model. We have assumed that the return and cost, p^* and d^*, are net of the price for risk, and we postulate that it is imperfect competition or regulation on some markets which creates the interest rate differentials. Barriers to entry or regulation (such as regulations on interest rate paid on demand deposits) prevent the creation of perfect substitutes which would erase the interest rate differentials. The relevance of imperfect competition can be questioned in a period of global deregulation, but it would seem that market concentration due to bank mergers or asymmetric information can create imperfections in at least some markets. In any case, the model is quite general as perfect competition will appear as a special case.

The growth path of assets and liabilities must be defined to close the model. We assume, for simplicity, that the assets and liabilities are constant in book value terms over the next two years, and that the accounting profit (the net interest income) is paid out as dividends. The subscripts 1, 2 and 3 will indicate, respectively, the beginning of the first and second period, and the end of the second period.

The market value of the equity of the bank (*MV*) is the discounted value of the dividends and liquidation value at the end of Year 2, discounted at the shareholders' opportunity cost of funds, b^* (the cost of equity).

$$MV = \frac{pL + bB - dD}{1 + b^*} + \frac{(1+p)L + (1+b)B - (1+d)D}{(1+b^*)^2}. \tag{10}$$

This relation can be expressed in a more cumbersome but very meaningful way (see the proof in Appendix B).

$$MV = L_1^* + B_1^* - D_1^* + \left[\frac{(p^* - b^*) \times L_1^*}{1 + b^*} + \frac{(p^* - b^*) \times L_2^*}{(1+b^*)^2}\right] + \left[\frac{(b^* - d^*) \times D_1^*}{(1+b^*)} + \frac{(b^* - d^*) \times D_2^*}{(1+b^*)^2}\right], \tag{11}$$

where L_i^*, B_i^*, and D_i^* are the current economic values in Year i of the loans, bonds and deposits evaluated at the current loan, bond and deposit rates. For instance, the economic values of loans at the beginning of their respective periods are, respectively:

$$L_1^* = \frac{pL}{(1+p^*)} + \frac{(1+p)L}{(1+p^*)^2}, \tag{12}$$

$$L_2^* = \frac{(1+p)L}{(1+p^*)}. \tag{13}$$

The market value of the bank is the sum of two terms: the difference between the economic value of assets and liabilities evaluated at their respective current rates and the franchise value. The franchise value represents the ability to pay below market rates on deposits and the ability to charge above market rates on loans. We have assumed implicitly that $p^* > b^* > d^*$. This is to be expected if financial intermediaries are to exist, but it needs not to be observed for each single asset and liability.

The current economic value of deposits and loans D_i^* and L_i^* can be interpreted as their liquidation (reimbursement) value if these liabilities and assets are valued at the depositors's and borrowers's opportunity cost of funds (d^* and p^*, respectively). For instance, a demand deposit with contractually a very short term maturity would always be valued at par, while a fixed-rate term deposit could be valued above par if the current deposit rate has fallen. The economic values of loans represent the values from the borrowers' perspective, that is the amount they would be willing to pay back immediately to the bank if they were able to borrow again at the current rate p^*.[10] This presentation allows us to distinguish between the value of equity on a *liquidation* basis and the value of the bank as a *going concern*. The liquidation value is equal to $(L_1^* + B_1^* - D_1^*)$, while the going concern value entails a second term, the value of the franchise, i.e., the ability of the bank to earn rents in the future. Solvency of banks must be evaluated as the greater of the *liquidation* and *going concern* values. This was already noted by Paul Samuelson many years ago (1945, p. 24): "It should not be necessary to argue before economists that the banking system is a going concern and should be treated as such".

In anticipation of the discussion on interest rate risk in Section 6, one can see that the market value of a bank will be affected not only by changes in value of assets and liabilities, but also by the impact of interest rate on the franchise value.

It can be shown that in a multi-period model, the valuation formula becomes

$$MV = L_1^* + B_1^* - D_1^* + \sum_{t=1}^{\infty} \frac{(p_t^* - b_t^*)L_t^* + (b_t^* - d_t^*)D_t^*}{\prod_{t=1}^{\infty}(1 + b_t^*)}. \tag{14}$$

4.4. *A valuation model of the banking firm with taxes (no risk)*

Corporate taxes are likely to affect the value of the bank for two reasons. The first one is that the cost of equity is not tax deductible, and the second is that capital gains (losses) on assets are taxed only when realized at the corporate level. The importance of the tax treatment of capital gains (losses) appears to have been somewhat ignored, as the accounting rule SFAS 107 on *fair value accounting* does not take them into account.

We assume that all the assets and deposits have been acquired at par, so that only accounting income (net interest margin) is being taxed. Denoting by t the corporate tax rate, we obtain the following valuation formula for the shares of the bank:

$$MV = \frac{(1 - t) \times (pL + bB - dD)}{1 + b^*} + \frac{(1 - t)(pL + bB - dD) + L + B - D}{(1 + b^*)^2}. \tag{15}$$

Following the proof reported in Appendix C, we obtain a quite meaningful formula:

$$MV = L_1^* + B_1^* - D_1^* + \left[\frac{(p^* - b^*) \times L_1^*}{1 + b^*} + \frac{(p^* - b^*) \times L_2^*}{(1 + b^*)^2}\right] + \left[\frac{(b^* - d^*) \times D_1^*}{(1 + b^*)} + \frac{(b^* - d^*) \times D_2^*}{(1 + b^*)^2}\right] + \left[-\frac{tb^* E_1^*}{1 + b^*} - \frac{tb^* E_2^*}{(1 + b^*)^2}\right]$$

[10] These economic values are identical to the *fair* value of financial instruments reported by US banks (SFAS 107 on "Disclosures about Fair Value of Financial Instruments").

$$+\left[\begin{array}{l}\dfrac{t(L_2^*-L_1^*)}{1+b^*}+\dfrac{t(L_3^*-L_2^*)}{(1+b^*)^2}+\dfrac{t(B_2^*-B_1^*)}{1+b^*}+\dfrac{t(B_3^*-B_2^*)}{(1+b^*)^2}\\ \quad -\dfrac{t(D_2^*-D_1^*)}{1+b^*}-\dfrac{t(D_3^*-D_2^*)}{(1+b^*)^2}\end{array}\right]. \tag{16}$$

The market value is the sum of four terms: the current value of assets net of the liabilities (the liquidation or economic value of equity), the after-tax value of the franchise on deposits and loans, the present value of the non-tax deductibility of equity cost (the M&M corporate tax penalty, Modigliani and Miller, 1958), and the present value of the tax savings due to the non-taxation of capital gains/losses on assets (liabilities) over the life of these assets (liabilities).

This valuation formula requires some explanatory comments. The current economic values of assets and liabilities L_i^*, B_i^*, and D_i^*, are the 'true' economic values as defined by Samuelson (1964), that is, the present value of after corporate tax cash flows discounted at the after corporate tax discount rate $(p^*(1-t), b^*(1-t)$, and $d^*(1-t)$, respectively) where taxes are paid on current capital gains and losses. Equivalently, given Samuelson's Invariance Theorem (1964), they are the present value of the before-tax cash flows discounted at the before-tax discount rate. We call these true economic values, S values. The first term in the valuation formula is the S equity, that is the difference between the S value of assets and liabilities. The second term is the after-tax franchise, that is the ability to pay below market rate on deposits, and to charge above market rate on loans. The third term is the non-tax deductibility of the cost of the S equity. One will notice that the economic value of equity is not constant during these two years because the current S value of assets and liabilities changes over time. The last set of terms takes into account the specific tax treatment of capital gains and losses on assets and deposits. They were not taxed, in our example, so that the present value of the tax savings (losses) must be included in the market value. The assumption of not taxing the unrealized capital gains and losses was made for expository convenience. This leaves room for tax management with losses to be taken immediately and gains realized only later to maximize the tax shelter. A very practical implication of this valuation formula is that one must consider the tax status of assets to measure the true value of a fixed-income portfolio (for instance, fixed-rate mortgages) when interest rate goes up. There can be a substantial difference between the discounted value of before-tax cash flows at the before-tax discount rate (i.e., the Samuelson S value) and the value of these assets for the bank's shareholders. These assets are worth much more to shareholders of the bank than to any investor in the market, because only the (low) interest income is taxed, while the capital gain earned over the life of these assets as they reach maturity will be tax-free (unless they are realized).

To summarize, the determinants of the market value of the bank include the S value of economic equity, the after-tax value of the franchise, the tax penalty on equity, and the present value of tax savings on unrealized capital gains. This presentation was chosen to highlight explicitly the determinants of the market value, with reference to SFAS 107 which suggests calculating the fair value of assets and liabilities on a before-tax

Table 3
Beta of banks' shares

BNP-Paribas	0.94	J.P. Morgan Chase	1.57
Société Générale	1.08	Citigroup	1.17
ABN-AMRO	1.25	Wells Fargo	0.13
ING Groep	1.48	Bank of America	0.55

Source: Thomson Analytics (December 2005).

basis (i.e., the S true economic value). This creates a need to consider explicitly the tax benefits on capital gains/losses.[11]

The growth of a bank's assets and deposits has been ignored. As shown in Dermine (1985b,1987) and in Dermine and Hillion (1992), the framework can be extended for real growth and inflation by simply replacing A_i^* by the current value of assets (deposits) in Year i, which may incorporate some old assets at their current value plus new assets booked into that period.

So far, risk and its effect on the choice of a risk-adjusted discount rate for a bank, has been ignored. This is discussed in the next section.

4.5. A valuation model of the banking firm: corporate tax and risk

The framework has ignored risk and the difficulty of choosing a *risk-adjusted* discount rate to value assets and liabilities. Standard corporate finance theory suggests discounting dividends at the cost of equity. This one is calculated as the expected return on the bank share, which can be estimated with a standard CAPM or a discounted dividend model (Brealey, Meyers and Allen, 2006). As an illustration, the *betas* of the shares of several European and American banks are reported in Table 3.

Whenever the risk of specific assets is different from that of the average bank, the standard corporate finance text-book recommends finding, in the stock market, shares of firms with similar risk as the one analyzed. For instance, in the case of a conglomerate firm with businesses in the chemical sector and in other sectors, one recommends using, as the cost of equity to evaluate projects in the chemical sector, the expected return on shares of companies specialized in the chemical sector. In principle, with a bank having assets with different types of risk, from very safe to very risky, one could be tempted to make a similar recommendation. Specialist banks, also called monolines, such as credit card providers (e.g., Capital One), global custodians (e.g., State Street, Bank of New

[11] An alternative way to calculate the market value of the equity of a bank is to value the economic value of assets and liabilities as the present value of after-tax cash flows discounted at the after-tax asset-specific rate, where the after-tax cash flows do not include the tax on capital gains. In this case, the tax benefits are directly incorporated in the current value, and the last term in the valuation formula, the value of the tax savings on unrealized capital gains, disappears.

York), or private banks (Vontobel in Switzerland) can help to estimate a risk premium specific to some activities of a universal bank. However, the standard 'corporate finance' recommendation is very unlikely to work for bank lending, for the reason that, on the stock market, specialized banks lending to just one business sector one are not easy to find. It is for this reason that banks often use one common average cost of equity to evaluate different activities, the expected return on the bank's own shares (Zaik et al., 1996). In this section, we propose a methodology to take into account specific risk-adjusted discount rates.

Let us consider the equity of a bank invested into loans. We focus on one asset for expository convenience and further assumes that it is a perpetual loan. The single asset approach is generalized next.

We define:

$L =$ loan (perpetuity),
$p =$ expected return on loan,
$t =$ corporate tax rate,
$p^* =$ expected return on new loan,
$b^{**} =$ shareholders' opportunity rate.

The balance sheet of this position is as follows:

Loan $L(p)$	Equity (b^{**})

We argue that, rather than searching for banks specialized in lending to a single business sector to recover the *beta* of their shares and the relevant risk premium, an alternative would be to use the *expected return on corporate bonds of similar risk* as an opportunity cost for the banks' shareholders. In other words, to value a loan to a particular business sector with a specific credit grade, one could use information from the corporate bond market.

We are now equipped to value the equity of this bank. It is the present value of the (perpetual) flows of dividends discounted at the shareholders' opportunity rate:

$$MV \text{ of equity} = PV \text{ of dividends} = \frac{(1-t) \times p \times L}{b^*}.$$

It can be shown to be equal to:

$$\begin{aligned}\text{Market Value} &= \frac{(1-t)pL}{(1-t)p^*} + \frac{(1-t)(p^* - b^{**})L^*}{b^{**}} - \frac{tb^{**}L^*}{b^{**}} \\ &= L^* + \frac{(1-t)(p^* - b^{**})L^*}{b^{**}} - \frac{tb^{**}L^*}{b^{**}}. \qquad (17)\end{aligned}$$

This formula is similar to the one discussed earlier. The value of the equity is the sum of three terms: the value of the loan after-tax cash flows discounted at the loan after-tax current rate, the after-tax value of the franchise, and the Modigliani–Miller tax penalty.

Table 4
Excess return (basis points) on corporate bonds (1993–1997) (actual return on corporate bonds–return on Federal bond)

	USA	Germany	United Kingdom
AAA	117 bp	−7	280
AA	260 bp	209	260
A	200 bp	58	280
BAA	360 bp	140	320

Source: Delianedis and Santa-Clara (1999).

The valuation formula highlights that the relevant opportunity rate should be the expected rate on a corporate bond with similar risk as the loan (b^{**}).[12]

Although the availability of data on expected return on corporate bonds is currently not as widely available as data on expected return on shares, one can expect that, with the growth of corporate bonds and asset-backed securities markets, more information will be available on the expected return on corporate bonds. For instance, a study by Delianedis and Santa-Clara (1999) provides the following information (see Table 4).

If one repeats the same approach for bonds and deposits, and applying the principle of value additivity of Modigliani and Miller (1958), one obtains:

$$\begin{aligned} MV = {} & L_1^* + B_1^* - D_1^* \\ & + \left[\frac{(1-t)\times(p^*-b_L^*)\times L_1^*}{1+b_L^*} + \frac{(1-t)\times(p^*-b_L^*)\times L_2^*}{(1+b_L^*)^2}\right] \\ & + \left[\frac{(1-t)\times(b_D^*-d^*)\times D_1^*}{(1+b_D^*)} + \frac{(1-t)\times(b_D^*-d^*)\times D_2^*}{(1+b_D^*)^2}\right] \\ & - \left[\frac{tb_L^*L_1^*}{1+b_L^*} + \frac{tb_L^*L_2^*}{(1+b_L^*)^2}\right] - \left[\frac{tb^*B_1^*}{1+b^*} + \frac{tb^*B_2^*}{(1+b^*)^2}\right] \\ & + \left[\frac{tb_D^*D_1^*}{1+b_D^*} + \frac{tb_D^*D_2^*}{(1+b_D^*)^2}\right]. \end{aligned} \tag{18}$$

The value of the equity is the value of the net assets discounted at the after-tax one-period asset specific discount rate (b_L^*, b_D^*, and b^*) plus the after-tax franchise value on loans and deposits, plus the tax penalty on assets and tax savings on deposits.

This valuation formula can be compared to the results of the neoclassical model discussed earlier. While that model, focused on economic profit, had identified the relevant transfer price as being the matched-maturity marginal value of funds (MMMVF), the market rate on the interbank or bond markets, the valuation formula states that banks

[12] There is no term for the unrealized capital gains/losses, as the value of the perpetual loan is constant.

should attempt to increase their franchise value net of a tax penalty, and that the relevant opportunity rate is the *expected return on bonds* with a similar risk as the asset under review.[13] Transfer prices used for pricing or for profit evaluation could therefore be differentiated according, not just to maturity, but also to credit risk. One can guess that, for assets with very low credit risk, this should not be much of an operational issue, as the beta and risk premium on this type of corporate debt must be small. However, for risky loans, the beta and risk premium could increase significantly (Kaplan and Stein, 1990).

5. Application I. Pricing loan and loan loss provisioning

The bank valuation framework presented above provides an integrated framework to discuss several ALM issues. In this section, we apply the model to the pricing of loans and to the calculation of fair credit risk provisions. In the following sections, it will be used to discuss the measurement of interest rate and liquidity risks, and also the aggregation of risks.

With regard to lending, two separate issues need to be distinguished. The first one occurs at the origination date of the loan at time 0. What is the break-even interest rate on the loan? A second issue occurs one year later. The interest on the loan has been paid, and a question arises as to how much loan provisions one should create or, equivalently, how one should measure the risk-adjusted profit at the end of the first year. This last issue is of great managerial relevance as, if no provisions were created early, there could be a temptation to go into high margin and high risk lending with a view to showing large profit in the early years.

5.1. The pricing of loans

To calculate the break-even interest rate on a risky loan, three types of data are needed: the funding structure (equity vs. debt), the probability of default over time, and the loss given default (LGD). The first one concerns the amount of equity needed to fund the loan. In line with standard practice (Basel Committee, 2004), banks, to ensure their solvency, need to provide enough equity to cover potential loan losses in case of a severe recession. This equity cushion is referred to as *economic* or *risk-based capital*. Potential loan losses can be measured in two ways: change in value of the loan (marked-to-market mode, MTM), or the cost of a default (default mode, DM) over a specific time interval,

[13] As an alternative to the expected return on corporate bonds, one could attempt to estimate a beta with reference to CAPM theory: $\text{beta}_i = \text{Cov}(R_i, R_M)/\sigma_M^2$. In the absence of empirical evidence from bank shares, the covariance can be estimated with the correlation between the accounting income or cash flow of a specific business and the market return. The key difference with our approach is that the expected return on corporate bonds can be estimated directly from market data, while the alternative ‘beta’ approach makes the strong implicit assumption that the CAPM holds.

which is one year in many banks. The change in value is the most comprehensive approach as it recognizes not only the states of default, but also the loss of value resulting from a downgrade or upgrade of the counterparty. The second approach (default mode) recognizes only two states of the world: default or no-default. Potential losses are evaluated with a certain threshold of confidence, such as 99.9%. The second set of data concerns the probability of default over time. The third, and final, parameter is the estimate of the losses in case of default (loss given default), which includes not only data on recovery, but also on the tax shield on losses that will be recognized by tax authorities.

A complete coverage of the, rapidly expanding, credit risk literature is outside the scope of this chapter.[14] In summary, three approaches are used to estimate the risk of default. First, one based on a collection of historical data for different risk grades. The second approach attempts to model the stochastic process of latent variables that drive the default event. The application of option pricing (Merton, 1974) to evaluate credit risk has led to successful commercial applications (e.g., Kealhofer, 1995; KMV, 2002). A third approach estimates directly the probability of default from observed bond prices and volatilities (Duffie and Singleton, 1999).

Let us consider the following example of a loan with two-year-to-maturity:

- €100 M two-year-to-maturity fixed rate loan (interest paid at the end of the year and principal at maturity)
- Corporate tax rate of 40%
- Expected return on one-year and two-year-to-maturity similar risk (zero coupon) corporate bond of 10.1 and 10.2%, respectively
- Fixed interbank rate of 10% for the first year, and 10% for the second year
- Equity (economic capital) funding: 6%
- Interbank funding: 94%
- Probability of default in Year One: 0%
- Probability of default in Year Two: 3%
- Loss Given Default = 40 (i.e., recovery of 60 in case of default)

The break-even loan interest rate, R, is such that the discounted value of expected cash flows is equal to the initial equity investment,

$$\begin{aligned}\text{Equity} = 6 = {} & \frac{R \times (1 - 0.4)}{1.101} \\ & + \frac{0.97 \times [R \times (1 - 0.4) + 100] + 0.03 \times [60 + 0.4 \times 40]}{1.102^2} \\ & - \frac{94 \times 0.10 \times (1 - 0.4)}{1.1} - \frac{94 \times 0.10 \times (1 - 0.4) + 94}{1.1 \times 1.1}. \end{aligned} \tag{19}$$

[14] See the surveys by Santomero (1997), Shimko (1999), Jackson and Perraudin (2000), Crouhy, Galai and Mark (2000, 2001), and Saunders and Allen (2002). The estimation of loan losses-given-default (LGD) is discussed in Dermine and Neto de Carvalho (2006).

The expected cash flows from the loan in Year 2 comprises two parts: The expected revenue in the case of non-default, and the recovery plus the tax shelter created by the losses in the case of default. The cash flows from the loan are discounted at the expected return on corporate bonds, the opportunity return available to shareholders, while the cash flows from the debt are discounted at the current debt rate. The break-even loan rate, $R = 11.45\%$, will capture implicitly the funding structure, the probability of default, the expected losses arising out of default, and the opportunity expected return on corporate bonds.[15] A separate, but related, issue will concern the creation of fair credit risk provisions over the life of the asset.

5.2. *Fair provisioning*

The issue of fair provisioning for credit risk is a very important one in banking for two reasons. The first one concerns the estimate of solvency of a bank and the need to measure properly its equity. Instead of waiting for problems to occur, one would want to recognize earlier the loss of value of assets to force a bank to reduce dividends and/or to increase its equity base. The second issue concerns the creation of proper incentives inside a bank. If provisions are not recognized early, there would be a *myopic* temptation to go into high risk–high margin lending to show very good profit and performance early, especially when rewards and bonuses are linked to performance. So, to reduce this eventual bias, the creation of early provisions would help to reduce the apparent profit. Whatever the reason, solvency or risk-adjusted performance evaluation, there is a need for a proper methodology to measure provisions. We argue that the value-based model allows a level of fair provisions to be created, fully consistent with finance theory.

Consider the example of the previous loan, priced at 11.45%. At the end of Year 1, the interest rate has been collected and the issue concerns the proper calculation of provisions, and the evaluation of the performance of the loan officer.

For the sake of the example, imagine that one year later, the parameters remain unchanged.

- €100 M two-year-to-maturity fixed rate loan (interest, 11.45%, paid at the end of the year, and principal paid at maturity)
- Corporate tax rate of 40%
- Expected return on one-year-to-maturity similar risk (zero coupon) corporate bond of 10.2%
- Fixed interbank rate of 10% for the second year
- Equity funding: 6%
- Interbank funding: 94%
- Probability of default in Year Two: 3%

[15] We refer to this loan pricing approach as EVAL, economic value added in lending. It allows us to calculate the break-even rate on a loan transaction (Dermine, 1995, 1998).

- Loss given default: 40 (i.e., recovery of 60 in case of default)

We propose to follow a marked-to-market approach, recognizing the change in fair value of the loan.

$$\text{Provisions} = \Delta\text{Net Loan Value} = \Delta(\text{value of loan} - \text{value of debt})$$

$$\begin{aligned}\text{Net Loan Value} &= \frac{0.97 \times [100 + 11.45 \times (1 - 0.4)] + 0.03 \times [60 + 0.4 \times 40]}{1.102} \\ &\quad - \frac{94 \times 0.10 \times (1 - 0.4) + 94}{1.1} \\ &= 96.138 - 90.582 = 5.556. \end{aligned} \tag{20}$$

Since the net value of the loan is €5.556 million at the end of Year 1, when it was €6 million at the beginning, one calculates the provisions as the change in net value over the year:

$$\text{Provisions} = 6 - 5.556 = 0.444.$$

The *risk-adjusted* profit on the loan, and the economic profit (*EP*) are calculated as follows:

$$\begin{aligned}\text{Profit} &= \text{after-tax interest margin} - \text{provisions} \\ &= (1 - 0.4) \times (11.45 - 10\% \times 94) - 0.444 = 1.224 - 0.444 = 0.78. \\ \text{Economic Profit} &= \text{Profit} - \text{cost of allocated equity} = 0.78 - (6 \times 13.2\%) = 0.\end{aligned}$$

The economic profit is zero in this example. This is to be expected as the loan rate of 11.45% was chosen as the break-even loan rate that would given shareholders the minimum required return on their investment. One notices that the methodology used to calculate the fair provisions is identical to that used to price loans. On a technical note and contrary to common belief, one observes the creation of provisions in Year 1, although the probability of default remained unchanged (at 3%). The intuition is that the high break-even rate of 11. 45% was needed to cover the expected cost of default in Year 2. Of course, the level of provisions would increase further if, in the event of a recession looming, the estimate of probability of default and/or the estimate of the recovery were being revised. There is a widespread debate on fair provisions (Bank for International Settlements, 2001), and the fear that accounting manipulation could lead to under reporting.[16]

To close this section, one needs to clarify the conceptual difference between the loan provisions proposed above, and the practice of some banks which compute provisions as the present value of expected losses. This last measure is related to our change in net

[16] The official estimate of bad loans held by the top 15 Japanese banks in 2001 rose to ¥20,700 bn (4% of gross domestic product). Analysts say that the real figure could be seven times this amount (Financial Times, 13 December 2001).

value, but it includes only the expected losses, while the change in market value will take pricing into account. We argue in favor of the change in market value as, not only is it consistent with value-based finance, but it also provides a common methodology to evaluate performances across the bank on a marked-to-market basis.

The first application of the bank valuation model was concerned with loan pricing and the creation of fair provisions for credit risk. We next turn to what was traditionally the corner stone of ALM in a bank, that is the measurement of interest rate and liquidity risks.

6. Application II. The measurement of interest rate and liquidity risks

The measurement of interest rate and liquidity risks on the banking book of a bank has had a central place in ALM. Early studies include Stigum and Branch (1983), Platt (1986), Farin (1989), and Fabozzi and Konishi (1991). The consensus is that banks should focus on two approaches to measure interest rate risk (Houpt and Embersit, 1991, 1996; Basel Committee, 1997; Dermine and Bissada, 2002). The first approach to measure interest rate risk is concerned with the impact of a change of interest rate on the Profit & Loss account, that is the impact on the net interest income (NII) of the bank. The second approach focuses on the impact of a change in interest rates on the fair economic value of the equity of a bank. The two alternative approaches will be discussed consecutively.

6.1. Net interest income at risk

The Income Statement of a representative bank, the Royal Bank of Canada, was introduced in Section 2. In the year 2000, it reported a net interest income of Can$ 5,279 million and a total non-interest revenue (fees, trading profit) of Can$ 6,680 million. Although the second source of revenue—fee income—has increased over the years, it remains that the net interest income of most banks is still a very substantial part of revenue. Therefore, a lot of emphasis is placed on the control of the net interest margin.

To evaluate the impact of a change in interest rate on the net interest margin of a bank, banks compute a repricing gap table, also called an interest rate sensitivity table. The time scale is broken into discrete time buckets, for instance quarters:[17] $1, 2, \ldots$. Considering the balance sheet of the bank at a specific date t, one measures the stock of assets and debt that will be repriced at time $t + i$ ($i = 1$ to N), respectively A_{t+i} and D_{t+i}, if there is an instantaneous movement in the yield curve at time $t+\varepsilon$. For instance, a one-quarter Treasury Bill and a floating rate loan repriced every quarter would be slotted into the one-quarter bucket. A five-year-fixed rate bond would be slotted in the five-year bucket. A representative repricing gap table follows (see Table 5).

[17] The length of the time bucket—a quarter—is arbitrary, being chosen for expository convenience. Banks, in a very volatile environment, should work with finer buckets: daily or weekly.

Table 5
Repricing gap table

	Roll-over date or nearest interest rate adjustment date (Can$ million)					
	Sight up to 3 month	3 to 6 months	6 months to 1 year	1 to 5 years	over 5 years	Non-interest sensitive
Assets	121,95	14,377	16,158	70,186	11,48	55,589
Loans						
Bonds						
Other assets						
Liabilities	156,198	13,602	18,852	32,52	8,523	60,045
Deposits						
Bonds						
Equity						
Off-balance sheet	16,656	−1,926	−6,378	−3,284	1,204	−6,272
Repricing gap	−17,592	−1,151	−9,072	34,382	4,161	−10,728
Cumulative gap	−17,592	−18,743	−27,815	6,567	10,728	0

Source: Royal Bank of Canada, Annual Report 2000.

Assuming that the roll-over (reinvestment of the asset) takes place over a quarter,[18] a positive cumulative gap indicates that there will be a net excess of assets to reprice in the coming quarter, while a negative cumulative gap indicates an excess of deposits to reprice. In the case of an increase in the interest rate curve, a positive cumulative gap will help the bank to increase its net interest margin, while a negative cumulative gap would generate a loss of net revenue. With reference to Table 5, the negative gaps run by the Royal Bank of Canada imply that the bank would benefit from a fall in interest rate.

Let us define the current curve of forward rates as $R_1, R_2, R_3, \ldots$. Formally, the change in the net interest margin at time $t+n$ for a change in the forward rate ΔR_{t+n} is equal to:

$$\begin{aligned}\Delta \text{Net Interest Margin}_{t+n} &= \sum_{i=1}^{N} A_{t+i} \times \Delta R_{t+n} - \sum_{i=1}^{N} D_{t+i} \times \Delta R_{t+n} \\ &= \sum_{i=1}^{N} (A_{t+i} - D_{t+i}) \times \Delta R_{t+n} \\ &= \text{Cumulative Gap}_{t+n} \times \Delta R_{t+n}. \end{aligned} \tag{21}$$

[18] Note that the one-quarter roll-over at the one-quarter forward rate is not a restrictive assumption. Indeed, the roll-over over two quarters at a two-quarter rate is, by definition of the six-month forward rate, equivalent in present value terms to a series of one quarter-roll-overs.

Banks have introduced the powerful concept of Earnings-at-Risk (*EAR*) to indicate the potential impact of an adverse change of interest rate on the P&L account of a particular quarter:

$$EAR_{t+n} = \Delta\text{Net Interest Margin}_{t+n} = \text{Cumulative Gap}_{t+n} \times \Delta R_{t+n}. \quad (22)$$

Once, a repricing gap table has been tabulated, an immediate question concerns the relevant change of interest rate (ΔR_{t+n}) that should be chosen to measure earnings-at-risk. Banks often report two measures of risk. A first measure evaluates the risk for a confidence interval of 99%. This measure indicates that the potential loss would be underestimated in 1% of cases. A second measure of risk concerns the measure of risk for rare big shocks (often referred to as 'stress scenario'), that is, an attempt to measure risk for those cases in the one percent interval.[19]

$$EAR_{t+n,99\%} = \text{Cumulative Gap}_{t+n} \times \Delta R_{t+n,99\%}.$$
$$EAR_{t+n,\text{stress}} = \text{Cumulative Gap}_{t+n} \times \Delta R_{t+n,\text{stress}}.$$

The repricing gap table provides a first tool to measure the interest rate risk of a bank. It must be completed with the use of a simulation model for several reasons. First, such a table only gives information on the current structure of assets and liabilities. It ignores the dynamic changes in volumes of business over time. Indeed, the volumes of future loans or deposits could be affected by a movement in the yield curve. For instance, the volume of corporate demand deposits and retail consumer loans are likely to decrease when interest rates tighten up. Second, the earnings-at-risk calculation implicitly assumes that interest rates on assets and liabilities will adjust by the same percentage change as the change in the market yield curve. It ignores the interest rate elasticity which could be very different from 1. This is particularly the case in the retail market with savings deposits or consumer loans. It is, indeed, well known that some deposit rates and credit rates can display a fairly low elasticity. In general, the absence of perfect correlation between two interest rates in the same bucket is referred to as *basis risk*. In order to take into account the impact of a change of interest rates on future volumes of business and the imperfect correlation between some interest rates, simulation models have been developed (Platt, 1986). Although Monte Carlo simulations could in principle generates thousands of scenarios, banks usually consider few scenarios (around a dozen) that take into account several movements in the market yield curve and various responses of volumes or interest rates. Finally, various products can have embedded options. A classical example is the pre-payment option on a fixed-rate long-term mortgage. Although, the loan is unlikely to be pre-paid in the case of a rising rate environment, this would not hold in the case of a decreasing interest rate. As is the case with options, one loses the symmetry between the effect of an increase and a decrease in interest rate.

[19] A review of volatility forecasts can be found in Figlewski (1997). Discussion of stress testing can be found in Longin (2000) or in Committee on the Global Financial System (2001).

Simulation of pre-payment under various interest rate scenarios can help to capture this complexity.

Several weaknesses of repricing gaps or bank simulation models have been identified (Dermine, 1991a, 1993). Repricing gaps most often ignores the payment of interest/coupons and taxes. Fixed income instruments should be treated as a series of zero coupon instruments with different maturity dates. Most often, floating rate assets/deposits are slotted into the first quarter gap. This ignores the fixed spread on the floating rate which creates the equivalent of a fixed rate annuity (Dermine, 1991b). Last, but not least, is the inappropriate treatment of equity. Indeed, equity is most often slotted into the last bucket (non-interest sensitive), as if its cost was not sensitive to interest rate. This is correct from an accounting and net interest income perspective as the cost of equity is not included in a P&L account. However, from a finance perspective, the correct measure of profitability should be an economic profit which takes away from net income the opportunity cost of equity. As the estimate of the cost of equity is based on the current opportunity return available on risk-free government bonds, one concludes that the cost of equity is interest sensitive, and therefore that equity should be included in the first bucket.[20]

The focus on the impact of interest rate on the net interest margin and the P&L account is understandable as bank analysts focus very much on ROE-based measures of performance. But, this measure risks the dangers of myopia if management focuses only on the short-term impact. Following the Savings & Loans Association (S&Ls) crisis in the United States in the early 1990s, increasing attention has been focused on the impact of interest rate on the change in the fair economic value of the equity of a bank. This is the second major approach to the measurement of interest rate risk on the banking book.

6.2. *Economic value at risk*

Following the large maturity mismatch, run by US S&Ls, between short-term retail deposits and long-term fixed rate mortgages, bank supervisors have been increasingly concerned with solvency and the need to ensure that the fair value of assets exceeds the fair value of debt. Defining the economic value of the equity of a bank as the difference between the value of assets and debt, one needs to calculate the impact of a change in interest rate on the economic value (EV) of equity.

$$\text{Economic value of equity} = EV = \text{Value of assets} - \text{Value of debt},$$
$$\Delta\text{Economic value of equity} = \Delta EV = \Delta\text{Value of assets} - \Delta\text{Value of debt}. \quad (23)$$

[20] One observes here an inconsistency in bank practice. Many banks control the impact of interest rate on the net interest margin (ignoring the cost of equity capital), while they evaluate internally the performance of business units on an economic profit basis, that is net of a cost of equity capital.

With A and D referring to the current value of assets and debt, Du_A and Du_D referring to the duration of asset and debt, and applying the MacAulay (1938) duration[21] formula to the change in value of assets and debt, one obtains:

$$\Delta EV = \left[-A \times Du_A/(1+R) \times \Delta R)\right] - \left[-D \times Du_D/(1+R) \times \Delta R)\right],$$

$$\Delta EV = -A \times (Du_A - D/ADu_D)/(1+R) \times \Delta R, \tag{24}$$

$$\Delta EV/EV = -(A/EV) \times \left[(Du_A - D/ADu_D) \times 1/(1+R)\right] \times \Delta R,$$

$$\Delta EV/EV = -\text{Leverage} \times \text{modified duration gap} \times \Delta R. \tag{25}$$

The last expression gives a very useful summary measure of interest rate risk, that is the percentage change in the value of the equity of a bank for a change in interest rates. It is the product of three factors: the leverage (assets over economic value), the duration mismatch between assets and liabilities, and the change in interest rate. Although the MacAulay duration applies only to a parallel shift in the yield curve, it can easily be extended to different twists in the yield curve with a *vector duration* approach (Chambers, Carleton and McEnally, 1988), or a Value-at-Risk approach that will be discussed in Section 7.

If the above approach is adequate for assets traded on perfectly competitive markets, such as government bonds or subordinated debt issued by the bank, it raises the practical issue of the application of the duration concept to special accounts, such as demand deposits, savings deposits, or even credit cards loans. Indeed, if the contractual maturity of a demand deposit is extremely short as these deposits can be withdrawn on demand, the effective maturity is much longer as a core of deposits is likely to be stable. For banks collecting this kind of deposits, a question arises as to the choice of the effective duration of these accounts. Fortunately, the bank valuation model proposed in Section 4 will allow us to answer the question (Dermine, 1985a, 1993).

Let us consider the following valuation formula where, for ease of exposition, we focus on the franchise value of deposits, ignoring taxes and the franchise value on loans. The market value (*MV*) of the equity of the bank is the sum of the liquidation value and the franchise value:

$$MV = MV_A - MV_D + \sum_{t=1}^{\omega} \frac{(b-d)D}{(1+b)^t}. \tag{26}$$

Assuming for simplicity a constant perpetual franchise on deposits, one obtains,

$$MV = MV_A - MV_D + \frac{(b-d)D}{b}. \tag{27}$$

Since we are interested in the response of the franchise or charter value to a change in the market rates, we make the realistic assumption that the volume of deposits is a

[21] The MacAulay duration of a fixed-income asset is its weighted average maturity, the weight, applied to each date of a cash flow receipt, being the present value of that cash flow divided by the value of the asset.

function of the deposit rate d and of the market rate b ($D(d, b)$) and that the deposit rate will respond to a change in the market rate.

The impact on the market value of a bank of a change in interest rate is given by:

$$\frac{\partial MV}{\partial b} = \frac{\partial MV_A}{\partial b} - \frac{\partial MV_D}{\partial d} \times \frac{\partial d}{\partial b} + \frac{\partial (\frac{(b-d)D}{b})}{\partial b}. \tag{28}$$

That is, the effect of a market rate change on the market value of current assets, the effect of a consecutive change in the deposit rate on the market value of current deposits, and finally, the effect on the franchise value or goodwill. This shows explicitly that the effective duration of a demand deposits is a direct function of the sensitivity of the charter value to a change in interest rate. Often, because of the inelasticity of the deposit rate, the margin on deposits increases when the interest rate goes up. The total impact on the charter value is then related to the sensitivity of the volume of deposits. So the less competitive the market, the higher will be the charter value on deposits, and the larger is likely to be the effective duration of demand deposits.

The market value sensitivity can be expressed in a more operational form where the term η_{xy} denotes the elasticity of variable x with respect to y, i.e., the percentage change in the x variable to a percentage change in the y variable:

$$\begin{aligned}\eta_{MV,b} = {} & \left(\eta_{MV_A,b} \times \frac{MV_A}{MV}\right) - \left(\eta_{MV_D,b} \times \frac{MV_D}{MV}\right) \\ & + \frac{(b-d)D/b}{MV} \times \left(\eta_{D,d} \times \eta_{d,b} + \eta_{D,b} + \eta_{(b-d)/b,b}\right)\end{aligned} \tag{29}$$

with $\eta_{MV_A,b} = -Du_A \times b/(1+b)$, Du_A denoting the MacAulay duration of assets, and similarly, $\eta_{MV_D,b} = -Du_D \times b/(1+b)$, Du_D denoting the MacAulay duration of deposits. Eq. (29) states that the elasticity of the market value of equity to a change in the market rate is a weighted sum of elasticities. The weights are the current value of assets, deposits and the charter value as a percentage of total market value. The series of elasticities are as follows: the elasticity of the value of the asset with respect to the asset rate, the elasticity of the value of deposits with respect to the deposit rate times the elasticity of the deposit rate to the market rate, and finally a series of elasticities measuring the sensitivity of the charter value to a change in the market rate. Obviously, the smaller the franchise value or its sensitivity to the interest rate, the smaller is its relevance of the measurement of interest rate risk. This 'imperfect market' approach to the measurement of interest rate risk was applied in the case of deposits. It can be extended to any asset or debt with a franchise value.[22]

[22] Application of this methodology to other assets and stochastic yield curves can be found in O'Brien, Orphanides and Small (1994); O'Brien (2000); Hutchison and Pennachhi (1996); or Jarrow and van Deventer (1998a, 1998b).

6.3. Hedging interest rate risk

Potentially, two approaches can be used to hedge interest rate risk, commercial or financial. A commercial approach involves the choice of maturities or repricing dates for assets and deposits to ensure matching. If this approach can be undertaken, it is often very costly as the restrictions on maturity and repricing terms can reduce profitability. Indeed, in some countries, consumers are used to fixed rate loans and short-term deposits. If a bank wishes to switch from fixed rate lending to variable rate lending, margin might suffer. For this reason, banks prefer to use financial instruments to manage their interest rate exposure. They could run an opposite mismatch on the interbank market or use financial derivatives, such as forward rates agreements (FRAs), financial futures, interest rate options,[23] or swaps. With regard to the use of derivatives, an additional difficulty has arisen recently. In the United States, the accounting rule, FAS 133, enforces the marked-to-market of derivatives.[24] This could create volatility of reported accounting income if the hedge instrument is marked-to-market, while the hedged position is accounted in the banking book at par value. At the international level, a similar debate arises with the International Accounting Standard (IAS) rule 39, which was implemented in 2005.

6.4. The measurement of liquidity risk

In Section 1, we mentioned that one of the five main functions of a bank is to provide insurance, and that one type was liquidity insurance whereby depositors or borrowers are able to withdraw money on demand. This creates a liquidity risk for a bank. Moreover, on a day-to-day basis, banks must have enough liquidity to cover the payments made on the central bank's clearing system, such as Fedwire in the US or Target[25] in the European System of Central banks. This liquidity consists of: (1) balances with the central bank, (2) borrowing from other banks, (3) discount window borrowing from the central bank, and (4) expected incoming transfers from other banks.

A representative statement of cash flows of a bank is first presented. It allows to understand the various sources of cash inflows and outflows. A discussion of the measurement of liquidity risk follows.

The consolidated statement of cash flows presented in Table 6 is that of the Royal Bank of Canada for the year ended October 31, 2000.

The annual consolidated statement of cash flows starts with the net income of Can$ 2,208 million reported in the consolidated income statement discussed in Section 2. It then takes into account all the non-cash items included in the income statement, such as

[23] The management of interest rate options is beyond the scope of this chapter. Useful references include Jarrow (1996) and Rebonato (1996).

[24] Hedge accounting (matching the accounting rule of the hedge to the rule applied to the hedged instruments—marked-to-market or accrual) can be used under very restrictive circumstances.

[25] Target: Trans-European Automated Real-time Gross settlement Express Transfer system.

Table 6
Consolidated statement of cash flows (C$M)

Cash flows from operating activities (Can$ million)	
Net income	2,208
Adjustments to determine net cash provided (used in) operating activities	
Provision for credit losses	691
Depreciation	369
Restructuring	–
Amortization of goodwill and other intangibles	91
Gain on sale of assets	(4)
Change in accrued interest receivable and payable	110
Net loss (gain) on sale of available for sale securities	11
Changes in operating assets and liabilities	
Deferred income tax	(206)
Current income taxes payable	(434)
Unrealized gains and amounts receivable on derivative contracts	(4,183)
Unrealized losses and amounts payable on derivative contracts	3,355
Trading account securities	(11,078)
Securities sold with recourse	(312)
Obligations related to securities sold short	(5,867)
Other	97
Net cash provided by (used in) operating activities	−15,152
Cash flows from investing activities	
Change in loans	(11,728)
Proceeds from the maturity of held to maturity securities	500
Purchases of held to maturity securities	(114)
Proceeds from available for sale securities	10,525
Proceeds from the maturity of available for sale securities	16,269
Purchases of available for sale securities	(23,640)
Change in interest-bearing deposits with other banks	1,927
Net acquisitions of premises and equipment	(293)
Net proceeds from sale of real estate	–
Change in asset purchased under reverse repurchase agreements	1,969
Net cash used in acquisition of subsidiaries	(323)
Net cash used in investing activities	−4,908
Cash flows from financing activities	
Issue of RBC Trust Capital securities (RBC TruCS)	650
Increase in domestic deposits	8,818
Increase in international deposits	9,405
Issue of subordinated debentures	1,200
Subordinated debentures matured	(20)
Issue of preferred shares	–
Preferred shares redeemed for cancellation	–
Issuance costs	(4)
Issue of common shares	59
Common shares redeemed for cancellation	(660)
Dividends paid	(791)
Change in securities sold under repurchase agreements	(391)

(*continued on next page*)

Table 6
(*Continued*)

Cash flows from financing activities	
Change in liabilities of subsidiaries	281
Net cash provided by financing activities	18,547
Net change in cash and due from banks	(1,513)
Cash and due from banks at the beginning of the year	2,460
Cash and due from banks at end of year	947

Source: Royal Bank of Canada, 2000 Annual Report.

depreciation and provision for credit losses. Finally, it takes into account the cash flows linked to investing or financing activities. In banking, two items require a clarification: accrued interest receivable and payable, and the 'float'.

Interest income (expense) is generally recognized (accrued) as income (expense) over time even if it has not yet been paid. The following accounting relations have to be used to compute the actual net cash flows related to interest accruals:

$$\begin{aligned}&\text{Accrued interest receivable (payable)}_{\text{Oct. 2000}}\\&\quad= \text{Accrued interest receivable (payable)}_{\text{Oct. 1999}}\\&\qquad+ \text{interest income (expense) accrual}_{2000}\\&\qquad- \text{interest income (expense) actually paid}_{2000}.\end{aligned}$$

The stock of accrued interest receivable (payable) at a specific date is equal to the stock of interest receivable (payable) a year earlier, plus the interest accruals of the year, minus the interest actually paid during the year. This relation allows the cash inflows (outflows) linked to interest accruals to be calculated:

$$\begin{aligned}\text{Cash inflow (outflow)}_{2000} &= \text{Interest income (expense) accrual}_{2000}\\&\qquad - \Delta\text{ accrued interest receivable (payable).}\end{aligned}$$

$$\begin{aligned}\text{Net cash flows}_{2000} &= \text{net interest accrual margin}_{2000}\\&\quad - \Delta\text{ accrued interest receivable}\\&\quad + \Delta\text{ accrued interest payable.}\end{aligned}$$

In the RBC example, for instance, the net change in accrued interest receivable and payable created a positive cash inflow of Can$ 110 million.

The second cash flow item to discuss is the 'float'. The practice of many banks is to credit (debit) deposit accounts several days after (before) the bank has actually received (paid) the cash. At any time t, the 'float' is a non-interest bearing source of funds, which represents the volume of deposits that have been debited 'early' or paid 'late' to clients

Table 7
Liquidity profile

Cash flow gap (daily, weekly . . .) for 'normal times' (Can$ million)

		Day One	Day Two	Day Three . . .
	Interest accrual income	44.1	46.2	
−	Interest accrual expense	−29.7	−28.3	
−/+	Margin calls	−1,000	−1,000	
−	Operating expense	−20.9	−22	
−	Tax	−3.9	−3.9	
+	Δ Assets ('normally' maturing)	76.71	78.3	
−	Δ Deposits ('normally' maturing)	−15,000	−1,000	
−	Δ Assets (new commitments)	−90.41	−93	
+	Δ Deposits (new flows)	3,000	2,000	
−	Δ Accrued interest receivable	−0.44	−0.44	
+	Δ Accrued interest payable	0.74	0.74	
+	Δ Float	−0.82	0.83	
+	Marketable assets	12,000		
	Net cash flows	−1,024.62	−21,57	
	Cumulative net cash flows	−1,024.62	−1,046,19	

Source: Author's example.

because of the 'value date' system. An increase in the 'float', an item typically not disclosed by banks, represents, therefore, an additional sources of cash available.

An annual consolidated statement of cash flows was presented to identify the various sources of cash inflows and outflows. To control liquidity, i.e., the imperative to have cash available to meet, on a day-to-day basis, various commitments such as, for instance, payments of interest, reimbursement of deposits, or payment of taxes, the banks use an instrument very similar to the repricing gap table, except that the concern now is the amount of cash flowing in or out over a particular, very short-term time interval. One such table (Table 7) is constructed for 'normal' time, in which a large proportion of deposits, withdrawable on demand, remain with the bank. These are referred to as 'core' deposits. A second exercise is done for 'stress' time, during which the deposits outflow is much larger.

If the cumulative cash flow is positive, there is no liquidity problem. If it is negative, the bank will be forced to borrow on the interbank market or at the discount window of the central bank. To avoid a market disruption with the bank coming with too large a call for liquidity, central banks have put caps on the size of the cumulative cash outflows. For instance, the Financial Services Authority (FSA) in the United Kingdom requires all banks to report all cash flows on the maturity ladder for periods of up to six months. Mismatch guidelines are set for cumulative periods of up to eight days and up to one month. Typically, these would be zero and minus 5% of the deposit base, respectively.[26]

[26] Hall (1989) and Financial Stability Review (2000).

6.5. *Cash flow gaps for 'stress' scenario*

A similar exercise is conducted, but one considers a case of a severe liquidity shock, such as a run on bank deposits or the inability of some clients to repay their loans. Under such extreme stress circumstances, a bank needs enough liquidity to survive for a few days, a period over which the banking industry or the national central bank is expected to intervene. Two historic examples of central banks' contingency liquidity plans include the plan for the century date change, *Y2K* (Drossos and Hilton, 2000) and the liquidity provision following the September 11, 2001 attacks on the World Trade Center and the Pentagon (McAndrews and Potter, 2002). The physical disruptions had left some banks unable to execute payments to other banks, resulting in an unexpected shortfall for other banks. To meet this liquidity problem, discount windows loans rose from about US$ 200 million to US$ 45 billion on September 12, 2001.

7. Application III. Portfolio diversification, marginal risk contribution, and the allocation of economic capital

In Section 5, we discussed loan pricing and introduced the concept of economic capital. Economic capital allocated to a particular loan or a business entity is the amount of equity needed to cover potential losses with some degree of confidence, such as 99.9%. Economic capital is the capital needed to ensure the solvency of the bank in bad times. Diamond (1984) and Merton and Perold (1993) have developed a theory of bank capital based on *opaqueness* and on the desire of many customers (such as retail depositors or swaps counterparties) to deal with a very safe institution. Because the asset holdings of a bank are only disclosed with a lag and can be changed very rapidly, it is very difficult to assess the true risk of an institution. Banks will therefore face high "agency" or "information" costs in raising external capital on short notice. Therefore, a need arises to have enough capital to keep the bank solvent or, for a given amount of capital, to be well diversified. These unique characteristics of financial firms create the need for diversification of risks, a need often considered irrelevant for a corporate firm (Brealey, Meyers and Allen, 2006). The usual finance argument is that diversification of risks by firms is not necessary, as home-made diversification by shareholders will lead to the same result. The argument of Merton and Perold is that banks are *special* because information and agency costs are likely to be very high in banking, and because customers want a solvent firm with enough capital or diversification.[27] Under this argument, high solvency generates a larger clientele and franchise value. If the concept of economic capital is readily applicable to a single source of risk, it becomes more complex when one is considering a series of risks. Indeed, it is well understood that diversification

[27] Other authors, who have expressed a similar view, include Santomero (1995), Oldfield and Santomero (1997), and Froot and Stein (1998).

Table 8
Gaps, volatility and EAR

	Repricing gap (US$ million)	Volatility of rates (σ)	EAR (US$ M) (position × volatility)
US$	gap_{US}	σ_{US}	$\text{gap}_{\text{US}} \times \sigma_{\text{US}}$
YEN	gap_{Yen}	σ_{Yen}	$\text{gap}_{\text{Yen}} \times \sigma_{\text{Yen}}$
DM	gap_{DM}	σ_{DM}	$\text{gap}_{\text{DM}} \times \sigma_{\text{DM}}$
FFR	gap_{FFR}	σ_{FFR}	$\text{gap}_{\text{FFR}} \times \sigma_{\text{FFR}}$

Total risk = EAR.

is likely to reduce total risk (Markowitz, 1959) and the amount of required economic capital. We first present an application of portfolio theory to the global measurement of interest rate risk of an international bank, with interest rate exposure in four markets. We discuss next the concept of marginal risk contribution.

7.1. Aggregate interest rate risk, an example

As an illustration, we will consider the real case of an American bank running, in the late 1980s, a money market book (maturities less than a year) in four currencies: the US Dollar, the Yen, the Deutsche Mark, and the French Franc. In each book, the bank is running a mismatch position, given by the gap discussed in Section 6 (the difference between short-term assets and short-term liabilities). The currency exposure is supposed to be hedged, so that the only sources of risk are interest rate movements in the four countries. The gaps and volatilities of interest rates (σ) are reported in Table 8. In the last column, we report the Earnings-at-Risk (EAR) for one standard deviation of interest rate.

In 1989, this bank was measuring the total risk on its banking book with the sum of the absolute risks. However, concerned with a proper allocation of economic capital to its treasury department, and worried that central banks would use this measure of risk to calculate the capital required for interest rate risk, the bank started to apply modern portfolio theory to take into account the diversification of risks.[28]

To apply portfolio theory, one additional piece of information is needed: the correlation between the four interest rates (see Table 9).

The calculations of variance and standard deviation of total interest rate risk follow:

$$
\begin{aligned}
&\text{Variance}_{\text{total risk}}\\
&\quad = (\text{gap}_{\text{US}}\Delta R_{\text{US}} + \text{gap}_{\text{Yen}}\Delta R_{\text{Yen}} + \text{gap}_{\text{DM}}\Delta R_{\text{DM}} + \text{gap}_{\text{FFR}}\Delta R_{\text{FFR}})\\
&\quad = \sigma^2_{\text{total risk}} = \text{gap}^2_{\text{US}}\sigma^2_{\text{US}} + \text{gap}^2_{\text{Yen}}\sigma^2_{\text{Yen}} + \text{gap}^2_{\text{DM}}\sigma^2_{\text{DM}} + \text{gap}^2_{\text{FFR}}\sigma^2_{\text{FFR}}
\end{aligned}
$$

[28] Historians will notice that it took thirty years to transfer the application of portfolio theory from the equity markets, where it was first applied, to overall market risks.

Table 9

Correlation (ρ)	US$	Yen	DM	FFR
US$	$\rho_{\mathrm{US,US}}$			
Yen	$\rho_{\mathrm{Yen,US}}$	$\rho_{\mathrm{Yen,Yen}}$		
DM	$\rho_{\mathrm{DM,US}}$	$\rho_{\mathrm{DM,Yen}}$	$\rho_{\mathrm{DM,DM}}$	
FFR	$\rho_{\mathrm{FFR,US}}$	$\rho_{\mathrm{FFR,Yen}}$	$\rho_{\mathrm{FFR,DM}}$	$\rho_{\mathrm{FFR,FFR}}$

$$\begin{aligned} &+ 2\mathrm{gap}_{\mathrm{US}}\mathrm{gap}_{\mathrm{Yen}}\rho_{\mathrm{US,Yen}}\sigma_{\mathrm{US}}\sigma_{\mathrm{Yen}} + 2\mathrm{gap}_{\mathrm{US}}\mathrm{gap}_{\mathrm{DM}}\rho_{\mathrm{US,DM}}\sigma_{\mathrm{US}}\sigma_{\mathrm{DM}} \\ &+ 2\mathrm{gap}_{\mathrm{US}}\mathrm{gap}_{\mathrm{FFR}}\rho_{\mathrm{US,FFR}}\sigma_{\mathrm{US}}\sigma_{\mathrm{FFR}} + 2\mathrm{gap}_{\mathrm{Yen}}\mathrm{gap}_{\mathrm{DM}}\rho_{\mathrm{Yen,DM}}\sigma_{\mathrm{Yen}}\sigma_{\mathrm{DM}} \\ &+ 2\mathrm{gap}_{\mathrm{Yen}}\mathrm{gap}_{\mathrm{FFR}}\rho_{\mathrm{Yen,FFR}}\sigma_{\mathrm{Yen}}\sigma_{\mathrm{FFR}} + 2\mathrm{gap}_{\mathrm{DM}}\mathrm{gap}_{\mathrm{FFR}}\rho_{\mathrm{DM,FFR}}\sigma_{\mathrm{DMK}}\sigma_{\mathrm{FFR}}. \end{aligned} \tag{30}$$

Total aggregate interest rate risk $= \sigma = \Delta$Variance.

The issue of the aggregation of risks, which takes into account diversification, was applied to the measure of total interest rate risk of a bank running a ‘banking’ book in four different markets. This general approach, which can be applied to any source of risk, has received a lot of attention in the trading area. Value-at-Risk (VAR) attempts to measure the risk arising from a change in the market value of a trading book (Duffie and Pan, 1997; Jorion, 2001). As traders face many sources of risk—such as interest rate, foreign exchange, prices of equities or commodities—a tool, VAR, was needed to measure total aggregate risk on a trading portfolio. This measure is useful for central bankers wanting to impose a capital requirement on trading risks, or for banks wanting to allocate economic capital to business units. Modern Portfolio Theory (Markowitz, 1959) and simulation-based tools are widely used by banks to measure trading risks.[29] The benefits of diversification was recognized explicitly in the BIS accord on capital requirement for trading risks (Basel Committee, 1996). This was a landmark accord as, not only were banks allowed to take into account diversification, but they could also use their own *internal* model for the measurement of risk.

7.2. Marginal risk contribution

The application of portfolio theory and the use of simulation models have allowed banks to measure the aggregate risk of a portfolio involving many sources of risk. The objective is to have a measure of risk which incorporates the diversification benefits of risk reduction. For financial firms with multiple businesses—such as commercial banking,

[29] As an illustration, in its 2001 Annual Report, J.P. Morgan Chase reports a measure for total market risk arising from both its ‘banking’ and ‘trading’ books. A total market risk of US$ 129.2 million is split as follows: VAR of US$ 67.4 million for the trading book, VAR of US$ 107.2 million for the investment portfolio and A&L activities, and a portfolio diversification risk reduction of US$ 45.4 million.

investment banking and insurance (or businesses in several countries), a new issue arises as to how to measure the risk contribution of a specific business unit, that is the *marginal risk contribution*. Although one can calculate the *stand alone* risk of this business (i.e., the volatility of income or the worst potential loss), one can guess that, integrated into a multi-business firms, part of the *marginal risk* will disappear through diversification.

Let us consider a bank with three assets: A, B, and C. It could refer to investment in commercial banking, insurance, and investment banking, or banking businesses in three different countries. We define:

$A =$ asset position (A, B, C)
$\sigma_A =$ standard deviation of return on asset A (A, B, C)
$\sigma_P =$ Aggregate risk = Standard deviation of portfolio income
$\rho_{A,B} =$ correlation between return on asset A and return on asset B
$\rho_{A,P} =$ correlation between return on asset A and return on portfolio P

Then, the variance of profit (σ_P^2) is equal to:

$$\begin{aligned}\text{Variance of profit } \left(\sigma_P^2\right) = {} & A^2 \times \sigma_A^2 + B^2 \times \sigma_B^2 + C^2 \times \sigma_C^2 \\ & + 2 \times A \times B \times \rho_{A,B} \times \sigma_A \times \sigma_B \\ & + 2 \times A \times C \times \rho_{A,C} \times \sigma_A \times \sigma_C \\ & + 2 \times B \times C \times \rho_{B,C} \times \sigma_B \times \sigma_C. \end{aligned} \tag{31}$$

The standard deviation of portfolio risk (σ_P), a measure of the aggregate risk, is equal to:

$$\text{Aggregate risk} = \sigma_P = \text{Variance}.$$

The standard aggregate risk formula is not very tractable, because the chain of cross-products in the variance makes it difficult to analyze the specific risk contribution of each business unit. To calculate the marginal risk contribution, the standard deviation of portfolio risk (σ_P) can also be written as (see proof in Appendix D):

$$\begin{aligned}\text{Aggregate risk} = \sigma_P = {} & [A \times \sigma_A \times \rho_{A,P}] + [B \times \sigma_B \times \rho_{B,P}] \\ & + [C \times \sigma_C \times \rho_{C,P}]. \end{aligned} \tag{32}$$

This expression of the aggregate risk (σ_P) is attractive as the total risk is now the sum of three components: the risks due to businesses A, B, and C.[30] The contribution of each component is the product of the *stand alone* risk $(A \times \sigma_A)$ multiplied by the correlation $(\rho_{A,P})$ between this business and the total bank. The concept of *marginal risk contribution* follows. For instance, the marginal risk contribution of business A is equal to:

[30] This decomposition of the aggregate risk into a sum of components is also referred to as DeltaVAR or DVAR (Crouhy, Galai and Mark, 2001).

$$\text{Marginal risk contribution of business } A = A \times \sigma_A \times \rho_{A,P}. \tag{33}$$

Some have suggested using the *marginal risk contribution* to allocate economic capital to business units, that is a measure of risk which takes into account the stand alone risk of this business and the correlation between this business and the bank. However, several authors[31] have pointed out that this approach would be misleading, because the measure of aggregate risk is not a sum of independent terms. Indeed, one can see that the marginal change in aggregate risk (32) for an increase (decrease) of business A is not just *the marginal risk contribution* ($\Delta A \times \sigma_A \times \rho_{A,P}$), but also the change in correlation between each business and the total bank ($\rho_{A,P}$, $\rho_{B,P}$, $\rho_{C,P}$), as a change in one business will change the total bank portfolio P. Total risk and capital are usually not additive in the risk of each component, with the implication that the separation theorem, which allows the bank to be split into a number of value centers, breaks down. Maximization of economic profit becomes again a joint maximization problem.[32]

8. Bank regulations

In a perfect market (i.e., one with full information and free entry), financial institutions will compete and the outcome will be socially optimal. Black (1970) and Fama (1980) have developed these theories of unregulated financial markets. In Section 3, we listed four private arguments for risk management in banking: managerial risk aversion, nonlinearity of taxes, costs of financial distress, and market imperfections. To justify public intervention and bank regulations, one needs to identify market failures.

With respect to banking services, two main explanations have been advanced for the existence of potential market failures: Imperfect (asymmetric) information which could prevent the proper functioning of unregulated private markets, and the potential for bank runs and the related fear of systemic crises.[33]

8.1. Imperfect (asymmetric) information and investor protection

The first and most important case of asymmetric information concerns the imperfect knowledge about the solvency of a banking firm. Depositors find it costly to evaluate the solvency of their bank. The economics literature (e.g., Kay and Vickers, 1988) recognizes that the inability of consumers to properly evaluate the quality of a product can create a market failure. When depositors are uninformed, there are fewer incentives to limit the riskiness of the assets of a financial institution or its degree of financial leverage (deposit-to-equity ratio). Indeed, finance theory (Merton, 1977) has shown that,

31 Merton and Perold (1993); Artzner et al. (1999); Stoughton and Zechner (1999); Turnbull (2000); Perold (2001); and Rosen and Zenios (2006).

32 In Section 4, we identified two alternative reasons for the breakdown of the separation theorem: joint operating costs and joint demand. Here, we have an additional argument: the externality created by diversification.

33 A complete discussion of the economics of banking regulations is available in Dermine (2000, 2003).

whenever depositors are not properly informed, shareholders of banks benefit from an increase in risk.[34] With perfect information, depositors would react by requesting an interest rate increase to offset the transfer going to shareholders. With imperfect (asymmetric) information, this would be difficult, and would raise a well identified and documented moral hazard problem.[35]

The economics literature has identified a first potential market failure rooted in imperfect information. It is legitimate to let countries draw up prudential regulations to protect the 'uninformed' investors.

8.2. Bank runs and systemic risks

The second main source of market failure is the potential for bank runs and systemic crises. Banks are special because the financial contract that emerges—illiquid loans funded by short-term deposits—creates a potential market failure and a need for public intervention. The financial contract creates the risk that depositors run to withdraw their funds. A run can be triggered by a bad news about the value of bank assets or by any unexplained fear. In either cases, there may be a loss since illiquid assets will be sold at a discount. Moreover, a bank failure could eventually trigger a signal on the solvency of other banks, leading to a systemic crisis. Here, a distinction should be drawn between a "domino" effect and a systemic crisis.

A domino effect exists if the failure of one bank directly endangers the solvency of other banks. This risk is substantially reduced today since banks systematically measure and control their counterparty exposure through, for instance, netting arrangements. A pure case of a *systemic run* could occur if, lacking information, depositors run to withdraw their funds from a significant number of banks.

This market failure explains the establishment of safety nets and banking regulations to guarantee the stability of banking markets.

8.3. Menu of banking regulations

In most countries government interventions and/or regulations have taken various forms:

Limits on entry. To collect bank deposits from the public, one needs to receive the authorization (a license) from the central bank. In order to do this, a bank's management

[34] The underlying intuition is that an increase in risk (variance of asset return) allows the shareholders of a firm to reap potentially large gains, while limiting the downside risk to zero because of the limited liability nature of equity shares.

[35] The potential existence of imperfect information *per se* does not yet justify public intervention. It has to be shown that private mechanisms cannot succeed. Solutions to the imperfect information problem are three-fold: information disclosure, reputation to protect the long-term value of the franchise, and the supply of risk-free deposits.

must show credentials of being ‘fit and proper’. Moreover, there is often a minimum capital requirement.

Rules of conduct. The rules of conduct can concern the opening of branches, deposit and lending rates, investment ratios, liquidity regulation, capital requirement, and limits on permissible activities (such as insurance or underwriting of securities). Prudential control can limit the size of interest rate and foreign exchange exposures. In many countries, central banks have exercised pressure on banks to create an Asset & Liability Committee (ALCO), whose mission is to supervise the profitability of the bank, its control of risks, and its compliance with regulations. Finally, governance pressure has been exercised on the boards of banks to nominate a few members with the specific task of monitoring the risk management process.

As many of these regulations have been found to be ineffective and often prone to regulatory capture, banking markets have been deregulated in most countries around the world and central banks[36] have retained five major forms of control: market entry (banking license), capital regulation based on an assessment of risks, liquidity regulation, regulations on interest rate and foreign exchange risks, and public disclosure of financial information. Capital regulation has been the object of intense international debate. In 1988, the Basel Committee on Banking Supervision, established by central banks of twelve countries,[37] drew up a minimal equity regulation that all international banks should meet. This regulation on capital is commonly referred to as the ‘Capital Adequacy ratio’, the BIS ratio, or the Cooke ratio, named after the first chairman of the Basel Committee. This regulation, applied since January 1993, states that the capital ratio must exceed 8%. A revised BIS ratio will be applicable on 1 January 2007, at the earliest (Basel Committee, 2004). With regard to the control of market risks, the new capital accord, Basel 2, proposes the adoption of a combination of approaches. Under ‘pillar one’, a formal capital requirement will be imposed on market risks originating from the trading book. Under ‘pillar two’, banking supervisors will be invited to control the level of interest rate risk on the banking book. One will observe that no formal capital requirement is imposed on interest rate risk on the banking book. The reason for

[36] In many countries, a debate has arisen as to which institution, central bank or separate public agency, should supervise banks. This has often been the mission of the central bank on the grounds that, as the provider of liquidity, the lender-of-last-resort should have full information on banks. Moreover, as the business of banks, investment banks and insurance companies is increasingly blurred and housed in one corporate financial services group, a question has arisen on the need to merge the functional regulators. In the United Kingdom, for instance, banking supervision has been transferred from the Bank of England to the FSA, the Financial Supervisory Authority.

[37] It consists of senior representative of bank supervisory authorities and central banks from Belgium, Canada, France, Germany, Italy, Japan, Luxembourg, The Netherlands, Sweden, Switzerland, The United Kingdom, and the United States. The secretariat of the committee is located at the Bank for International Settlements (BIS), in Basel Switzerland.

this is that banking regulators have been unable to agree on a unique formula to measure interest rate risk on the banking book. The sources of disagreement are related to our discussion of the control of interest risk in Section 6, that is uncertain maturity on accounts such as saving deposits, demand deposits, or consumer loans. The absence of a capital requirement on interest rate risk on the banking book is a significant shortcoming of the new capital accord as, in most countries, that source of risk is substantially larger than risk on the trading book.

9. Conclusion

Two models of the commercial banking firm have been presented in this chapter. The neoclassical model and the bank separation theorem allow discussion on transfer pricing and the special issue of deposits with undefined maturities. The finance-based valuation model allows the market value of the equity of a bank to be broken down into four components: the liquidation value, the after-tax franchise value, the corporate tax penalty, and the value of the unrealized capital gains/losses. It allows discussion on risk-adjusted transfer price for loans, the pricing of loans, the creation of fair loan loss provisions, and the measurement of interest rate risk. The discussion of the marginal risk contribution shows that risks are not additive, thus providing an additional argument for the cautious use of decentralization into profit centers. Breaking up the bank, and maximizing separately the value of each part, is unlikely to be optimal. Joint maximization of market value should be the objective.

Besides a comprehensive summary of the literature on ALM in Banking, this chapter makes six contributions. (1) It identifies the proper transfer price that should be used to evaluate the margin on risky loans. The very much used matched-maturity interbank market rate should be replaced by the expected return on a matched-maturity corporate bonds with similar risk. (2) It provides a risk-adjusted methodology to price risky loans. (3) It shows that the fair provisioning of credit losses on loans should be based on the change in the fair market value over time. This measure is related to, but different from, the much used value of expected losses. (4) It identifies the relevant maturity of the market rate that should be used to price deposits with undefined maturities, the relevant maturity of the asset that should be used for hedging, and the relevant maturity that should be used to evaluate *ex post* the performance of a deposit gathering department. It is argued that the relevant maturity of these rates is not always the same. (5) It calls the attention to the fact that *fair value accounting* proposes the evaluation of assets and liabilities of banks on a before-tax basis, and that the correct market value of the equity of a bank will also include value for tax savings on unrealized gains on assets and liabilities. (6) Finally, this chapter offers a critical review of the approach of many banks which control the impact of interest rate on accounting profit, when a sounder measure of risk should be the impact of interest rate on economic profit.

Appendix A. The relevant maturity of the transfer price

It is argued that the choice of the relevant maturity for the transfer-benchmark rate depends on each specific application: pricing the product, selecting the hedge instrument, or evaluating *ex post* a branch performance. Let us consider the case of deposit collection, a representative case. Similar reasoning can apply to retail loans.

A.1. Maturity of the transfer price for pricing

It was argued in Section 4, that, although the maturity of a deposit could be quite short, a lag in the supply of deposit and/or interest rate rigidity would require using, as a transfer price, some average of the current and forward market rates. Therefore, with regard to pricing, the choice of a longer term benchmark rate has to be justified either by a lag or by interest rate "stickiness".

A.2. Effective duration of deposits

In Section 6, we showed that the effective duration of a deposit would be longer than the contractual maturity whenever an increase in the interest rate increased the value of the franchise on deposits. In this case, it is imperfectly-competitive markets (the source of the interest margin and the franchise value) that drives the choice of the appropriate duration of the hedge instrument. A simple example is the case of demand deposits with very low or no interest. As one observes that the franchise on these deposits is usually much higher in countries with high interest rates, the duration of demand deposits is much longer than the contractual short-term maturity. It must be noticed that the maturity used for pricing could differ from the maturity used for hedging. Indeed, if repricing occurs frequently and if there is no lag, the duration applicable for *pricing* should be close to the maturity of the product. However, the duration used for *hedging* could be much longer, if the franchise value increases with the level of interest rate.

A.3. Transfer price for ex post *evaluation of performance*

A golden rule in management accounting to evaluate a value center is to choose a transfer price that leads to value creation decisions. Based on Section 4, a branch should be given as transfer price for deposits (or loans) the relevant market rate used for pricing (i.e., an average of current and forward rates whenever there are lags or rigid rates). The transfer price must reflect the marginal profit of new business. However, the use of the current rate (or average of current and forward rates) implies that a low interest rate environment will often lead to very low performance for a branch collecting retail deposits. The reason for this is that interest margins on deposits are usually much lower in a low interest rate environment. It would be somewhat unfair to penalize the branch manager for low performance, especially when the bank, as a whole, can be hedged with the purchase of longer duration assets and capital gains created in a declining rate environment. In short: in a declining interest rate environment, who should get the benefits of the hedge? The retail branch or the bank? Although a formula could be devised to return the benefits of the hedge to the branch, we take the view the value center system should reflect the true current marginal profitability based on the current rate (ignoring the benefits of the hedge). The implication is that the profitability of a deposit-gathering branch is likely to be correlated with the level of interest rates. In our view, branch performance should be evaluated vis-à-vis a 'benchmark performance level', not in absolute terms. In a low interest rate environment, one should adjust the benchmark level of performance downward to recognize that 'normal benchmark' profit will be lower.

Appendix B. Bank valuation, no tax-no growth

We show that a two-period asset A with a historical return a, a current (one-period) return a^* and a discount rate of b^* is equal to:

$$MV = \frac{aA}{1+b^*} + \frac{(1+a)A}{(1+b^*)^2} = A_1^* + \left[\frac{(a^*-b^*)\times A_1^*}{1+b^*} + \frac{(a^*-b^*)\times A_2^*}{(1+b^*)^2}\right]$$

with

$$A_1^* = \frac{aA}{(1+a^*)} + \frac{(1+a)A}{(1+a^*)^2}, \qquad A_2^* = \frac{(1+a)A}{(1+a^*)}.$$

The analysis can be repeated for loans, bonds and deposits to obtain the valuation formula.

Proof.

$$\begin{aligned}
MV &= \frac{aA}{1+b^*} + \frac{(1+a)A}{(1+b^*)^2} = \left(\frac{aA}{1+a^*} + \left(\frac{aA}{1+b^*} - \frac{aA}{1+a^*}\right)\right) \\
&\quad + \left(\frac{(1+a)A}{(1+a^*)^2} + \left(\frac{(1+a)A}{(1+b^*)^2} - \frac{(1+a)A}{(1+a^*)^2}\right)\right) \\
&= \left(\frac{aA}{1+a^*} + \frac{(1+aA)}{(1+a^*)^2}\right) \\
&\quad + \left(\frac{aA(a^*-b^*)}{(1+a^*)(1+b^*)} + \frac{(1+a)A(a^*-b^*)((1+a^*)+(1+b^*))}{(1+a^*)^2(1+b^*)^2}\right) \\
&= A_1^* + \frac{(a^*-b^*)\left(\frac{aA}{1+a^*} + \frac{(1+a)A}{(1+a^*)^2}\right)}{(1+b^*)} + \frac{(a^*-b^*)\frac{(1+a)A}{1+a^*}}{(1+b^*)^2} \\
&= A_1^* + \left[\frac{(a^*-b^*)\times A_1^*}{1+b^*} + \frac{(a^*-b^*)\times A_2^*}{(1+b^*)^2}\right].
\end{aligned}$$

Appendix C. Bank valuation, the corporate tax case

We show that the after-tax value of a two-period asset A issued at par with historical return a, current (one period) return a^*, and discount rate b^* is equal to:

$$\begin{aligned}
MV &= \frac{(1-t)aA}{(1+b^*)} + \frac{(1-t)aA + A}{(1+b^*)^2} \\
&= A_1^* + \left[\frac{(a^*-b^*)\times A_1^*}{1+b^*} + \frac{(a^*-b^*)\times A_2^*}{(1+b^*)^2}\right] + \left[-\frac{tb^*A_1^*}{1+b^*} - \frac{tb^*A_2^*}{(1+b^*)^2}\right] \\
&\quad + \left[\frac{t(A_2^*-A_1^*)}{1+b^*} + \frac{t(A_3^*-A_2^*)}{(1+b^*)^2}\right],
\end{aligned}$$

where

$$A_1^* = \frac{aA}{(1+a^*)} + \frac{(1+a)A}{(1+a^*)^2}, \qquad A_2^* = \frac{(1+a)A}{(1+a^*)}, \qquad A_3^* = A.$$

Proof.

$$\begin{aligned}
MV &= \left[\frac{aA - t(aA + A_2^* - A_1^*)}{1+b^*} + \frac{(1+a)A - t(aA + A_3^* - A_2^*)}{(1+b^*)^2}\right] \\
&\quad + \left[\frac{t(A_2^*-A_1^*)}{1+b^*} + \frac{t(A_3^*-A_2^*)}{(1+b^*)^2}\right]
\end{aligned}$$

denoting by T the last factor (the capital gain tax shelter), we have:

$$\begin{aligned}
MV &= \frac{aA - t(aA + A_2^* - A_1^*)}{1+a^*(1-t)} + \frac{(1+aA) - t(aA + A_3^* - A_2^*)}{(1+a^*(1-t))^2} \\
&\quad + \left[\frac{aA - ta^*A_1^*}{1+b^*} - \frac{aA - ta^*A_1^{*1}}{1+a^*(1-t)}\right] \\
&\quad + \left[\frac{(1+a)A - ta^*A_2^*}{(1+b^*)^2} - \frac{(1+a)A - ta^*A_2^*}{(1+a^*(1-t))^2}\right] + T.
\end{aligned}$$

The Tax Invariance Theorem of Samuelson (1964) allows us to write:

$$\begin{aligned}
MV &= A_1^* + \left[\frac{(aA - ta^*A_1^*)(a^*(1-t)-b^*)}{(1+b^*)(1+a^*(1-t))}\right. \\
&\quad \left. + \frac{((1+a)A - ta^*A_2^*)(a^*(1-t)-b^*)(1+a^*(1-t)+(1+b^*))}{(1+b^*)^2(1+a^*(1-t))^2}\right] + T \\
&= A_1^* + (a^*(1-t)-b^*)\left(\frac{\frac{aA-ta^*A_1^*}{1+a^*(1-t)} + \frac{(1+a)A-ta^*A_2^*}{1+a^*(1-t)^2}}{1+b^*}\right) \\
&\quad + \frac{\frac{(1+a)A-ta^*A_2^*}{1+a^*(1-t)}}{(1+b^*)^2} + T,
\end{aligned}$$

$$MV = A_1^* + \left[\frac{(a^* - b^*) \times A_1^*}{1 + b^*} + \frac{(a^* - b^*) \times A_2^*}{(1 + b^*)^2}\right] + \left[-\frac{tb^* A_1^*}{1 + b^*} - \frac{tb^* A_2^*}{(1 + b^*)^2}\right] + \left[\frac{t(A_2^* - A_1^*)}{1 + b^*} + \frac{t(A_3^* - A_2^*)}{(1 + b^*)^2}\right].$$

This analysis can be repeated for loans, bonds and deposits to obtain the valuation formula.

Appendix D. Proof of marginal contribution formula

$$\sigma_p^2 = \sum_i \sum_j x_i x_j \operatorname{Cov}(R_i, R_j) = \sum_i x_i \operatorname{Cov}(R_i, R_p) = \sum_i x_i \rho_{i,p} \sigma_i \sigma_p,$$

$$\sigma_p = \sum_i x_i \sigma_i \rho_{i,p}.$$

Most often, available correlations concern those between pairs of businesses. As the above formula demands the correlation between each business and the portfolio, here are the formulae to move from pairwise correlations to those between one business and the portfolio.

We define:

Asset position: A, B, C
$\sigma_A =$ Standard deviation of return R_A on asset A (B, C)
$P =$ Portfolio Income $= A \times R_A + B \times R_B + C \times R_C$
$\sigma_P =$ Standard deviation of portfolio income
$\rho_{A,B} =$ correlation between return on asset A return and return on asset B
$\rho_{A,P} =$ correlation between return on asset A return and income on portfolio P
$\rho_{A,P} =$ Covariance (R_A, Portfolio Income)$/(\sigma_A \times \sigma_P)$
with

$$\begin{aligned}
&\text{Covariance}(R_A,\ \text{Portfolio Income}) \\
&\quad = \text{Covariance}(R_A, AR_A + BR_B + CR_C) \\
&\quad = A\,\text{Covariance}(R_A, R_A) + B\,\text{Covariance}(R_A, R_B) + C\,\text{Covariance}(R_A, R_C) \\
&\quad = A \times \sigma_A \times \sigma_A + B \times \rho_{A,B} \times \sigma_A \times \sigma_B + C \times \rho_{A,C} \times \sigma_A \times \sigma_C.
\end{aligned}$$

References

Allen, F., Santomero, A.M., 1998. The theory of financial intermediation. Journal of Banking & Finance 21, 1461–1485.

Allen, F., Santomero, A.M., 2001. What do financial intermediaries do? Journal of Banking & Finance 25, 271–294.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.
Ausubel, L.M., 1991. The failure of competition in the credit card market. American Economic Review 81 (1), 50–81, March.
Baltensperger, E., 1982. Alternative approaches to the theory of the banking firm. Journal of Monetary Economics 6 (1), 1–37.
Basel Committee on Banking Supervision, 1996. Amendments to the capital accord to incorporate market risks, Basel, pp. 1–54.
Basel Committee on Banking Supervision, 1997. Principles for the management of interest rate risk. Bank for International Settlements, Basel, pp. 1–33.
Basel Committee on Banking Supervision, 2004. International Convergence of Capital Measurement and Capital Standards. Bank for International Settlements, Basel.
Bank for International Settlements, 2001. Annual Report, Basel.
Bessis, J., 2002. Risk Management in Banking, second ed. John Wiley & Sons.
Black, F., 1970. Banking and interest rates in a world without money. Journal of Bank Research, 9–20, Autumn.
Brealey, R.A., Meyers, S.C., Allen, F., 2006. Corporate Finance, eighth & International ed. McGraw-Hill.
Chambers, D.M., Carleton, W.T., McEnally, R.W., 1988. Immunizing default-free bond portfolio with a duration vector. Journal of Financial and Quantitative Analysis 23 (1), 89–104.
Committee on the Global Financial System, 2001. A survey of stress tests and current practice at major financial institutions. Bank for International Settlements, Basel.
Crouhy, M., Galai, D., Mark, R., 2000. A comparative analysis of current credit risk models. Journal of Banking & Finance 24, 59–117.
Crouhy, M., Galai, D., Mark, M., 2001. Risk Management. McGraw-Hill.
Delianedis, G., Santa-Clara, P., 1999. The exposure of international corporate bond returns to exchange rate risk. In: Dermine, J., Hillion, P. (Eds.), European Capital Markets with a Single Currency. Oxford University Press, Oxford.
Dermine, J., 1984. Pricing Policies of Financial Intermediaries. Studies in Contemporary Economics, vol. 5. Springer-Verlag, Berlin.
Dermine, J., 1985a. The measurement of interest rate risk by financial intermediaries. Journal of Bank Research 16 (2), 86–90, Summer.
Dermine, J., 1985b. Taxes, inflation and banks' market values. Journal of Business, Finance and Accounting 12 (1), 65–74.
Dermine, J., 1986. Deposit rates, credit rates and bank capital, the Klein–Monti model revisited. Journal of Banking and Finance 10, 99–114.
Dermine, J., 1987. Measuring the market value of a bank, a primer. Finance, Revue de l'Association Française de Finance 8 (2), 91–108.
Dermine, J., 1991a. The BIS proposal for the measurement of interest rate risk, some pitfalls. Journal of International Securities Markets.
Dermine, J., 1991b. Floating rate securities and duration, a note. Mimeo, INSEAD.
Dermine, J., 1993. The evaluation of interest rate risk, some warnings about the Basle proposal. Finanzmarkt und Portfolio Management 7.
Dermine, J., 1995. Loan arbitrage-free pricing. The Financier 2 (2), 64–67.
Dermine, J., 1998. Pitfalls in the application of RAROC, with reference to loan management. The Arbitrageur–The Financier 1 (1), 21–27.
Dermine, J., 2000. The economics of bank mergers in the European Union. Journal of Common Market Studies 38 (3), 409–425.
Dermine, J., 2003. European banking, past, present, and future. In: Second ECB Central Banking Conference—The Transformation of the European Financial System. ECB, Frankfurt.
Dermine, J., Bissada, J.Y., 2002. Asset & Liability Management, A Guide to Value Creation and Risk Control. Financial Times–Prentice-Hall.

Dermine, J., Hillion, P., 1992. Deposit rate ceilings and the market value of banks, the case of France 1971–1981. Journal of Money, Credit and Banking, 184–194.
Dermine, J., Neto de Carvalho, C., 2006. Bank loan losses-given-default, a case study. Journal of Banking & Finance 30.
Dermine, J., Neven, D., Thisse, J., 1991. Towards an equilibrium theory of the mutual fund industry. Journal of Banking and Finance 15, 485–499.
Diamond, D., 1984. Financial intermediation and delegated monitoring. Review of Financial Studies 51, 393–414.
Diamond, D., Dybvig, P., 1983. Bank runs, deposit insurance and liquidity. Journal of Political Economy 91, 401–419.
Drossos, E.S., Hilton, S., 2000. The Federal Reserve's Contingency Financing Plan for the century date change. Current Issues in Economics and Finance 6 (15), 1–6.
Duffie, D., Pan, J., 1997. An overview of value at risk. The Journal of Derivatives, 7–49.
Duffie, D., Singleton, K., 1999. Modeling term structure of defaultable bonds. Review of Financial Studies 12, 687–720.
Fabozzi, F.J., Konishi, A. (Eds.), 1991. Asset/Liability Management. Probus Publishing Company, Chicago, IL.
Fama, E., 1980. Banking in the theory of finance. Journal of Monetary Economics 6 (1), 39–57.
Fama, E., 1985. What's different about banks? Journal of Monetary Economics 15, 29–36.
Farin, T.A., 1989. Asset/Liability Management for Savings Institutions. The Institute of Financial Education.
Figlewski, S., 1997. Forecasting volatility. Journal of Financial Markets, Institutions and Instruments 6 (1), 1–88, January.
Financial Stability Review, 2000. Banking system liquidity, developments and issues. Bank of England, December.
Freixas, X., Rochet, J.C., 1997. Microeconomics of Banking. MIT Press.
Froot, K.A., Stein, J.C., 1998. Risk management, capital budgeting, and capital structure policy for financial institutions, an integrated approach. Journal of Financial Economics 47, 55–82, January.
Hall, M., 1989. Handbook of Banking Regulation and Supervision. Woodhead-Faulkner Ltd., Cambridge.
Hannan, T., Berger, A., 1991. The rigidity of prices: Evidence from the banking industry. American Economic Review 81, 938–945.
Houpt, J.V., Embersit, J., 1991. A method for evaluating interest rate risk in commercial banking. Federal Reserve Bulletin, 625–637, August.
Houpt, J.V., Embersit, J., 1996. An analysis of commercial bank exposure to interest rate risk. Federal Reserve Bulletin, 115–128, February.
Hutchison, D., Pennachhi, G., 1996. Measuring rents and interest rate risk in imperfect financial markets: The case of retail bank deposits. Journal of Financial and Quantitative Analysis 31 (3), 399–417.
Jackson, P., Perraudin, W., 2000. Credit risk modeling and regulatory issues. Journal of Banking and Finance 24 (1–2), Special Issue.
Jarrow, R.A., 1996. Modelling Fixed Income Securities and Interest Rate Options. McGraw-Hill, New York.
Jarrow, R.A., van Deventer, D.R., 1998a. The arbitrage-free valuation and hedging of demand deposits and credit card loans. Journal of Banking and Finance 22, 249–272.
Jarrow, R., van Deventer, D., 1998b. Integrating interest rate risk and credit risk in ALM. In: Kamakura Corporation (Ed.), Asset & Liability Management, A Synthesis of New Methodologies. Risk Books.
Jorion, P., 2001. Value at Risk, second ed. McGraw-Hill.
Kaplan, S., Stein, J., 1990. How risky is the debt in highly leveraged transactions. Journal of Financial Economics 27, 215–245.
Kay, J., Vickers, J., 1988. Regulatory reform in Britain. Economic Policy 7, 286–351.
Kealhofer, S., 1995. Managing default risk in portfolios of derivatives. In: Derivative Credit Risk. Renaissance Risk Publications, pp. 49–66.
Klein, M.A., 1971. A theory of the banking firm. Journal of Money, Credit and Banking 3 (2), 205–218.
KMV, 2002. Modeling default risk. KMV LLC, San Francisco.

Longin, F., 2000. From value at risk to stress testing: The extreme value approach. Journal of Banking & Finance, 1097–1130.

MacAulay, F.R., 1938. The movements of interest rates, bond yields and stock prices in the United States since 1856. National Bureau of Economic Research.

Markowitz, H.M., 1959. Portfolio Selection: Efficient Diversification of Investment. Cowles Foundation Monograph, vol. 16. Yale University Press, New Haven.

McAndrews, J.C., Potter, S.M., 2002. Liquidity effects of the events of September 11, 2001. FRBNY Economic Policy Review, 59–79, November.

Merton, R., 1977. An analytic derivation of the cost of deposit insurance and loan guarantees. Journal of Banking and Finance 1, 3–11.

Merton, R., Perold, A., 1993. Theory of risk capital in financial firms. Journal of Applied Corporate Finance 6, 16–32, Fall.

Modigliani, F., Miller, M., 1958. The cost of capital, corporation finance and the theory of investment. American Economic Review 48, 261–297.

Monti, M., 1972. Deposit, credit and interest rate determination under alternative bank objective functions. In: Shell, K., Szego, G. (Eds.), Mathematical Methods in Investment and Finance, pp. 431–454.

O'Brien, J., 2000. Estimating the value and interest rate risk of interest-bearing transactions deposits. In: Finance & Economics Discussion Series, vol. 53. Federal Reserve Board, pp. 1–44.

O'Brien, J., Orphanides, A., Small, D., 1994. Estimating the interest rate sensitivity of liquid retail deposit values. In: Finance and Economics Discussion Series, vol. 15. Federal Reserve Board, pp. 1–26.

Oldfield, G.S., Santomero, A., 1997. The place of risk management in financial institutions. Sloan Management Review, Summer.

Merton, R., 1974. On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 28, 449–470.

Perold, A.F., 2001. Capital allocation in financial firms. In: Working Paper Series # 98-072. Harvard Business School Competition and Strategy, pp. 1–38.

Platt, R.B., 1986. Controlling Interest Rate Risk. John Wiley & Sons, New York.

Rajan, R., 1998. The past and future of commercial banking viewed through an incomplete contract lens. Journal of Money, Credit and Banking, 524–550.

Rebonato, R., 1996. Interest-Rate Option Models. Wiley, New York.

Rosen, D., Zenios, S., 2006. Enterprise-wide asset and liability management: Issues, institutions and models. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 1. In: North-Holland Handbooks in Finance. Elsevier Science B.V.

Samuelson, P.A., 1945. The effect of interest rate increases on the banking system. American Economic Review 35, 16–27.

Samuelson, P.A., 1964. Tax deductibility of economic depreciation to insure invariant valuation. Journal of Political Economy 72, 604–666.

Santomero, A., 1984. Modelling the banking firm. Journal of Money, Credit and Banking 16 (4), 576–616.

Santomero, A., 1995. Financial risk management, the whys and the hows. Financial Markets, Institutions & Instruments 4, 1–14.

Santomero, A., 1997. Commercial bank risk management: An analysis of the process. Journal of Financial Services Research 12 (2–3), 83–115.

Saunders, A., Allen, L., 2002. Credit Risk Measurement, second ed. John Wiley & Sons, New York.

Saunders, A., Walter, I., 1994. Universal Banking in the United States: What do we Gain? What do we Loose? Oxford University Press, Oxford.

Shimko, D. (Ed.), 1999. Credit Risk, Models and Management. Risk Books.

Stigum, M., Branch, R.O., 1983. Managing Bank Assets and Liabilities. Dow Jones-Irwin.

Stoughton, N.M., Zechner, J., 1999. Optimal capital allocation using RAROC and EVA. Manuscript. UC Irvine and University of Vienna, pp. 1–33.

Stiglitz, J., Weiss, A., 1981. Credit rationing with imperfect information. American Economic Review 71, 393–410.

Turnbull, S.M., 2000. Capital allocation and risk performance measurement in a financial institution. Financial Markets, Institutions & Instruments 9, 325–357.

Zaik, E., Walter, J., Kelling, G., James, C., 1996. RAROC at bank of America: From theory to practice. Journal of Applied Corporate Finance 9 (2), 83–93, Summer.

This page intentionally left blank

Chapter 12

DYNAMIC FINANCIAL ANALYSIS FOR MULTINATIONAL INSURANCE COMPANIES*

JOHN M. MULVEY

Bendheim Center for Finance, Princeton University, Princeton, NJ 08544-0001, USA

BILL PAULING[1]

Hartford Investment Management

STEPHEN BRITT[2]

Insurance Australia Group Asset Management

FRANÇOIS MORIN

Towers Perrin–Tillinghast

Contents

* Research support from the US National Science Foundation (DMI 0323410) and Towers Perrin and its Tillinghast subsidiary are gratefully acknowledged.

[1] Bill Pauling worked at Towers Perrin during the writing of this article.

[2] Steve Britt worked at Towers Perrin during the writing of this article.

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02012-6

Abstract

Global insurance/reinsurance companies can gain significant advantages by implementing an enterprise risk management system. The major goals are: (1) increase long-term profitability, (2) reduce enterprise risks, and (3) identify the firm's optimal capital structure. Profitability depends upon evaluating uncertainties as a function of a set of common factors across the enterprise. The system takes up all major decisions: identifying optimal liability-driven asset strategies, expanding/contracting insurance lines and pricing policies, constructing sound reinsurance treaties, and setting the firm's capital structure. To implement risk management in global companies we present a decentralized system. An automated toolkit eases constructing an enterprise system (Appendix A). Real-world experiences are highlighted via case studies.

Keywords

dynamic financial analysis, enterprise risk management, asset and liability management, stochastic programs, optimizing insurance companies

JEL classification: C61, D81, G11, G22

1. Introduction to dynamic financial analysis

Leading financial companies have begun to integrate their strategic decisions within enterprise-wide planning systems. The underlying modeling concepts are identified by several terms depending upon the context, including total integrated risk management (TIRM^{TM}), dynamic financial analysis (DFA), enterprise risk management (ERM), and asset and liability management (ALM). We employ the DFA name in this paper, given our emphasis on applying the technology to insurance companies. Each title portrays a shift in emphasis and agenda; DFA focuses on financial risks/analysis, while ERM includes operational and other non-financial sources of risks. The overall goals are similar—to improve a company's economic performance by collecting, analyzing, and managing information regarding the primary risk and return factors. Once the risks and rewards are properly identified, company executives can take actions to maximize the firm's shareholder value.

The primary areas of application for DFA in an insurance company are: (1) asset allocation strategies, (2) business decisions regarding the lines of insurance to underwrite, (3) pricing strategies, (4) reinsurance deals, (5) capital allocation, and (6) setting target rates of return. Important firm-wide decisions of the chief executive, chief financial officer and other executives can be evaluated by means of a strategic DFA system. Likewise, operational decisions such as underwriting policy can be improved by means of coordinated risk management. We suggest that lower level decisions should be accommodated within tactical risk management systems, which are linked to the strategic DFA system (see Section 4). In this report, we emphasize the strategic nature of DFA.

What motivates a company to implement an integrated risk management system? First, there have been spectacular failures despite the steep regulations for insurance companies, for example, Equitable Life in the United Kingdom. These failures show that current regulations do not ensure the solvency of an insurance company. Part of the failure can be attributed to organizational and competitive issues, such as periods of industry-wide under-pricing practices. But also, the failure can be attributed to the lack of modern risk management systems. Legacy systems often employ historical data for estimating losses, for example, without regard for the factors that drive the risks. This approach can underestimate the rare events since most economic and physical risks do not fit the normal distribution in the tails. See Section 2 for further details on this point.

Importantly, there is recognition that a modern insurance company can increase its profits by careful analysis of the sources of its risks, intelligent pricing strategies, and by diversifying away from the convergent risks. As an example, by growing an international business, e.g., AIG, or by limiting exposure to the factors giving rise to the rare events, e.g., Renaissance Reinsurance (Lowe and Stanard, 1996), and Smartwriter at St. Paul, a company can reduce the amount of capital needed to support a fixed level of business (Berger, Mulvey and Nish, 1998). Risk-adjusted profits increase through these strategies. In this paper, we differentiate the traditional static valuation methods from dynamic methods for valuing the activities of the firm. Section 4 describes an approach for deriving a dynamic valuation in the context of a decentralized DFA approach.

There have been several key successful applications of the DFA methodology. The Frank Russell Company developed one of the early DFA systems, for the Yasuda PC Company in Japan (Cariño et al., 1994; Cariño and Ziemba, 1998). This implementation has been widely cited in the literature. Other examples include: Tillinghast–Towers Perrin (Mulvey, Gould and Morgan, 2000), Renaissance Reinsurance (Lowe and Stanard, 1996), American Re-insurance (Berger and Madsen, 1999), Swiss Re (Laster and Thorlacius, 2000), the Norwegian life insurance company Gjensidige (Hoyland, 1998), and recently AXA. In the last and other examples, a significant issue involves calculating the target capital to deploy across its worldwide operations. While conducting independent analyses for each of its global subsidiaries, AXA aggregates its businesses in order to calculate the overall diversification benefits. These results provide better information regarding the safety of the organization in conjunction with the desired level of capital to maintain.

Other industries have made progress on improving efficiency through integrated decision-making and optimization. As a prominent example, the area of *integrated logistics* has benefited by real time information flows across the supply chain—from sales source backward to the distributor, the warehouse locator, the transportation system, and the manufacturer. Each part of the organization is aware of its role and especially the costs of delays and inefficient operations. For example, inventory allows a company to service the customer quickly, but it comes at a high price in terms of flexibility to change the product mix. Dell Computer provides a prototypical example of minimizing inventory—building their computers to exact customer specifications. Many process and manufacturing companies now deploy the enterprise technology, including AspenTech, IBM, I2 Technologies, etc. See Geoffrion and Krishnan (2001), Erkan (2006) and Lin et al. (2000). These industries directly benefit by applying enterprise-focused technology to improve efficiency. Employees enhance value by improving the quality of information on which decisions are based.

An insurance company's capital can be treated as inventory—but with a dual purpose. First, the capital serves to reduce risks of adverse consequences; it is a buffer to keep the firm solvent and to maintain a positive credit rating. Second, the capital forms a part of the asset base in which investment returns occur. As interest rates and inflation rates have dropped since 1980 with an accompanying increase in stock and fixed income asset returns, the importance of investment income has increased (Enz and Karl, 2001; Consiglio and Zenios, 2001). However, insurance companies are unlikely to experience the same investment climate over the next few years and, therefore, a modern insurance company will need to improve its underwriting results in order to gain a respectable return on equity. The capital structure plays an important role in determining the enterprise's risks and rewards.

There are several primary goals for an integrated risk management system: (1) to grow the company's capital with the optimal risk adjusted rate, (2) to ensure that the company survives a series of adverse circumstances, and (3) to improve the firm's profitability by increasing efficiency and making comparable decisions regarding risk

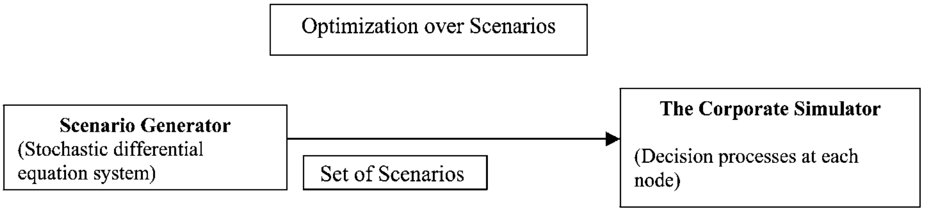


Fig. 1. Primary elements of a DFA system.

and reward. Taken together, these goals are synonymous with "maximizing shareholder value".

In the next section, we describe the major components of an enterprise risk management system. The three primary elements consist of: (1) a stochastic scenario generator, (2) a corporate simulator, and (3) a module for identifying improving decisions and processes. Each element builds on the previous (Figure 1).

A DFA system depends, first of all, on modeling the important economic uncertainties—interest rates, inflation, stock returns, insurance losses, bond returns. This module, the "scenario generator", consists of stochastic differential equations for each of the economic factors, asset returns, and the liability (insurance related) cash flows. The scenario generator provides inputs for the decision simulator. The model evaluates selected policy/decision rules across each of the scenarios—thus simulating the company as it reacts to the uncertainties embedded in the scenarios. Policy rules should reflect a reasonable approximation of the company's business strategy in the face of changing economic and business conditions. Under each scenario, the company will achieve a specified performance. The set of performance indicators across the scenarios completes the decision simulator. As with any stochastic model, the results depict probability distributions of outcomes. Temporal issues complicate the problem by adding performance measures at various time periods.

The remainder of this report is organized as follows. The DFA model is defined in the next section, including a discussion of the scenario generator and the optimization algorithms available for solving the resulting model. The search for an optimal solution is hindered due to the presence of sampling errors, the possibility of a non-convex objective function, and the need for multiple goals depending upon the stakeholders' positions. Section 3 describes a series of case studies in which the DFA system can improve decision-making. Prominent examples include: asset allocation strategies, selection of business activities to increase or decrease, the structure of reinsurance treaties, and the company's degree of leverage (gearing). Since many of these decisions interact with each other, we suggest that the full system be run in an integrated fashion. In Section 4, we show that the DFA system can be implemented in large, decentralized companies by extending existing capital allocation policies. The scenario approach lends itself well to the decentralized management, common in most multinational financial organizations. Last, we make recommendations for future developments in the DFA arena (Section 5).

Appendix A discusses the use of software tools to build and maintain a complex DFA modeling system. Since an integrated system must address the major elements of an insurance company, there are numerous sources of data and estimation, requiring a series of underlying assumptions throughout the process. We have found that the efficient design of a DFA software system requires a DFA toolkit in which the system is easy to use, flexible, and can be customized by the user to accommodate the complex structure of many insurance products. Also, users should be able to quickly pinpoint the underlying assumptions. An example of a DFA toolkit is shown in Appendix A.

2. Basic structure of a DFA system

An expanded version of a strategic DFA system is shown in Figure 2. The first issue involves setting the planning period. Typically, the DFA model projects the company over the next 3 to 5^+ years, with annual or possibly quarterly time periods. The firm's position is evaluated at the end of the planning period—the horizon. Special attention is paid to the next year's profit and surplus, of course, along with the risk of a downgrade in credit quality and other short-term indicators. The company's stock price and future well-being depend upon these performance indicators. The definition of risk measures and related aspects of the objective function are discussed below.

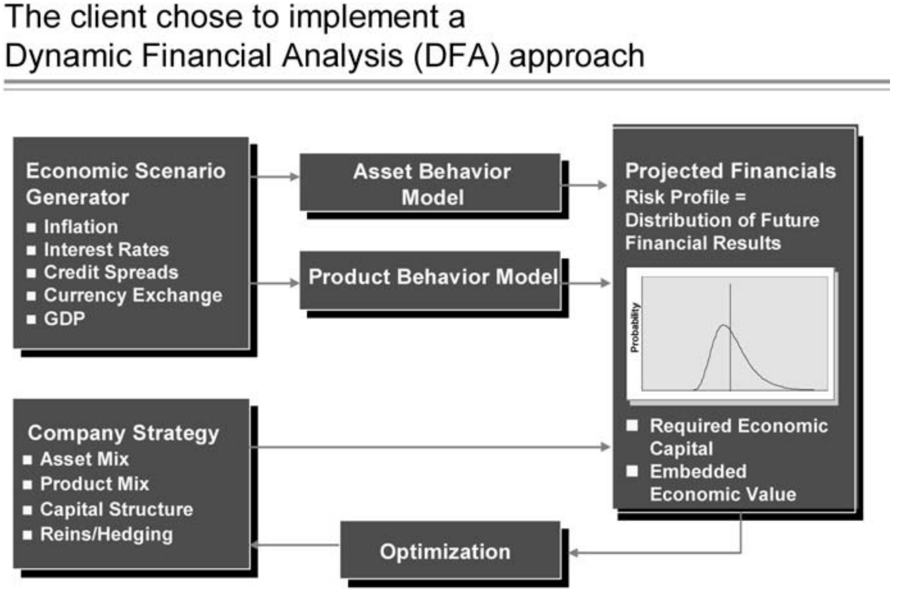


Fig. 2. The basic structure of an integrated risk management system. The planning period ($t = 1, 2, \ldots, T$).

The scenario generator projects a series of plausible paths over the multi-period planning horizon (step 1). Scenarios are consistently modeled across the organization; the behavior of the assets *and* the liabilities comes from the same factors. Given a set of policies, we simulate the corporation over each scenario (step 2). The company's overall financial health is ascertained with regard to the pre-determined goals. Last, an optimal search procedure (step 3) is implemented in order to find a set of policies that optimize the company's shareholder value. As there are numerous choices of goals facing insurance executives, the optimizing search must necessarily engage the judgments of the senior managers. The problem fits the domain of multi-objective optimization. In this section, we describe the DFA model at a relatively high level.

2.1. Scenario generators

Financial simulation requires a consistent method for projecting the random variables over the multi-period planning horizon. The scenario generators are critical for analyzing the insurance company in an integrated fashion, for example, when evaluating the merger of financial and insurance risks. Actuaries, financial planners and insurance executives actively employ stochastic generators. An early economic generator in the UK is the Wilkie model (1987). Mulvey (1989) developed a system for Pacific Mutual and latter for Towers Perrin–Tillinghast (1996, 1998). These generators possess a cascade structure with a nonlinear relationship among the factors. Alternatively, firms have developed linear systems such as vector auto-regressive models.

There are three key steps when building a scenario-generating module: (1) defining the modeling equations, (2) estimating the parameters, and (3) sampling the system. First, we discuss the model equations. Herein, Figure 3 depicts an important issue that arises when generating scenarios—the close connection between the underlying economic factors and both the assets returns and changes in the liability cash flows. Interest rate changes will impact asset returns (for fixed income asset categories) and the discount rate for valuing liabilities. Linking the assets and liabilities in a consistent fashion requires modeling the driving factors, in this case interest rates. The structural approach begins with modeling the economic and monetary factors, especially interest rates, inflation, GDP, corporate earnings, and credit spreads, over the multi-period horizon. As we will see, interest rates are generally modeled by means of mean reverting diffusion equations; bond returns are a direct function of interest rate swings due to changing economic conditions. The calibration of the parameters is complicated, and approaches such as maximum likelihood must be extended. Among others, methods of indirect inference apply to the resulting calibration problem; see Duffie and Singleton (1993), Berger and Madsen (1999), Gourieroux, Monfort and Renault (1993), Mulvey, Rosenbaum and Shetty (1999), and Chen and Bakshi (2001) for differing approaches. Also, see Hoyland and Wallace (2001) for a practical method for generating scenarios over a stochastic programming tree.

We depict uncertainty by a set of discrete realizations $s \in \mathbb{S}$. The projection module is based on a continuous set of stochastic differential equations, as defined by $s \in \Omega$.

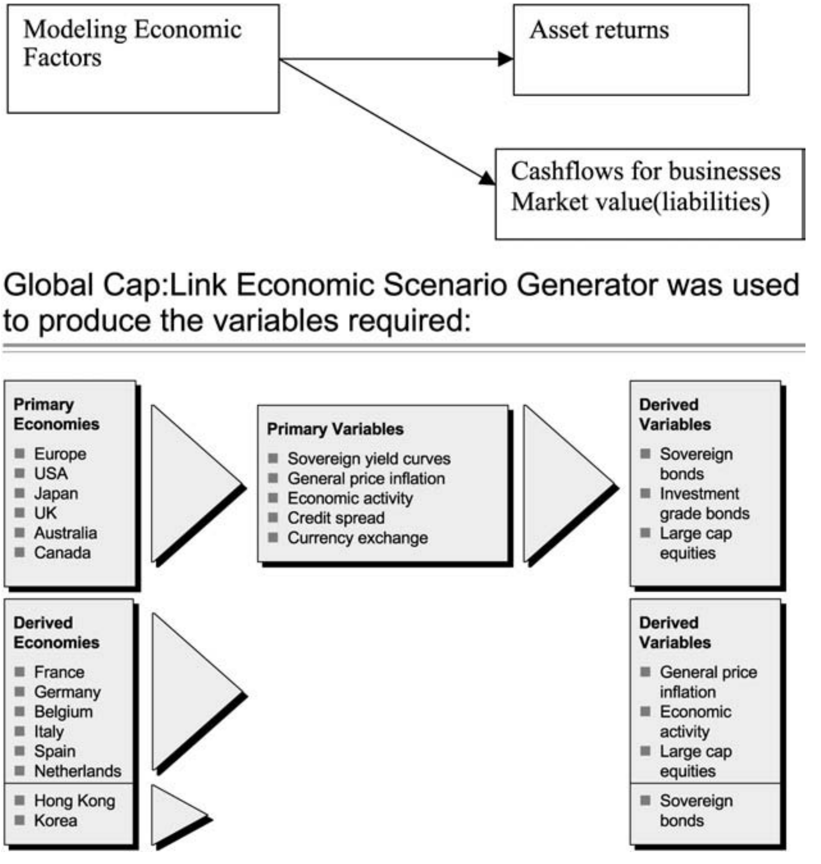


Fig. 3. The relationship of underlying factors to assets and liability cash flows, valuations, and returns.

To give an illustration, we describe the nominal interest rate process within a real-world DFA system—the CAP:Link scenario generator. This scenario generator has been implemented throughout the world for pension plans and insurance companies (Mulvey, Gould and Morgan, 2000).

The CAP:Link interest rate model is a spot rate model that is derived in a three-stage process. The first stage is a pair of linked stochastic processes. We assume that long and short interest rates link together through a correlated white noise term and by means of a stabilizing term that keeps the spread between the short and long rates under control.

Define short and long nominal interest rates (spot rates) as follows:

$$dr_t = \kappa_r(\bar{r} - r_t)\,dt + \gamma_r(s_t - \bar{s})\,dt + \phi_r\sqrt{r_t}\,dZ_r,$$

$$dl_t = \kappa_l(\bar{l} - l_t)\,dt + \gamma_l(s_t - \bar{s})\,dt + \phi_l\sqrt{l_t}\,dZ_l,$$

$$s_t = r_t - l_t,$$

where

r_t	short interest rate
l_t	long interest rate
dr_t, dl_t	the short and long interest rates increments over dt
s_t	the spread between long and short interest rates
dZ_r, dZ_l	correlated standard Brownian motions (white noise terms)
κ_r, κ_l	drift on short and long interest rates
γ_r, γ_l	drift on the spread between long and short interest rates
ϕ_r, ϕ_l	instantaneous volatility
$\bar{r}$	mean reversion level of short rate
$\bar{l}$	mean reversion level of long rate
$\bar{s}$	mean reversion level of the spread between long and short interest rates

Among them, $\bar{r}$, $\bar{l}$, $\bar{s}$, ϕ_r, ϕ_l, κ_r, κ_l, γ_r, γ_l are unknown parameters.

The second stage models a third point on the curve—mid-point. The equation for the mid rate is

$$dm_t = dE(m_t) + \kappa_m (E(m_t) - m_t)\,dt + \phi_m\,dZ_m,$$

where

$$E(m_t) = \theta_m r_t + (1 - \theta_m) l_t$$

and where

m_t	mid interest rate
dm_t	the mid interest rate increments over dt
dZ_m	standard Brownian motion (the white noise term)
κ_m	drift on mid interest rate
ϕ_m	instantaneous volatility
θ_m	exponential weight between long and short interest rates

Among them, $\kappa_m, \phi_m, \theta_m$ are parameters.

The mid-point accounts for curvature shifts in the yield curve.

In stage three we fill out the spot yield curve using a smoothing procedure:

$$Spot(t, T) = a + b\mathrm{e}^{-\omega_1 T} + c\mathrm{e}^{-\omega_2 T},$$

where $0 \leqslant T \leqslant \overline{T}$, $\overline{T} := 30$ yrs and T is the maturity of the long interest rates. Given parameters ω_1 and ω_2, this equation is linear, leading to a quick solution. These parameters do not generally change as a result of a re-calibration of the model. The interest rate submodel is defined in terms of the indicated twelve parameters.

The second element in building a scenario generator involves parameter estimation. When calibrating the model we weigh the competing needs of the users. For instance,

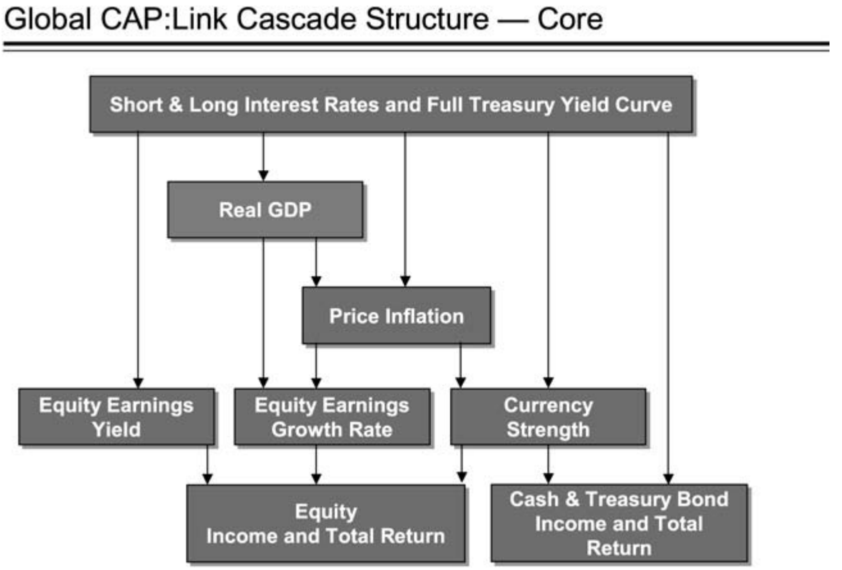


Fig. 4. Cascade structure of CAP:Link system.

there is a need to model accurately the covariance structure of interest rates at relatively high frequency (monthly), as some insurance contracts are sensitive to term spread and volatility of changes in bond yields. Other contracts have longer-term interest rate guarantees and are more sensitive to the range of interest rates over longer (10 to 20 year) periods. The calibration targets relate to both bond yield characteristics and bond return characteristics.

Targets are set for:

- Volatility of monthly change in bond yields;
- Correlation of changes in bond yields;
- Highs and lows (measured as the quantiles of the distribution); and
- Bond return characteristics, over 12 month time frames.

The relationships of other random variables within the CAP:Link system are displayed in Figure 4. For example, price inflation is projected in conjunction with the nominal interest rates, and as a consequence, the real (nominal minus inflation) government spot rate curve is derived as part of the process.

DFA models require a rich set of economic factors in the economic scenario generators, to adequately cover the range of risks encountered by insurance and other financial institutions. The US Society of Actuaries and US Casualty Actuarial Society are currently advocating a publicly available model to include the following variables:

1. Long term government interest rates,
2. Short term government interest rates,
3. Shape of yield curve,

4. Stock market price levels—large cap,
5. Stock market price levels—small cap,
6. General inflation rate,
7. Medical inflation rate,
8. Wage level inflation,
9. Real estate price levels,
10. Unemployment rate (optional), and
11. Economic growth rate (optional).

The third stage is to perform a sampling exercise in order to select scenario set $\mathbb{S}$. There are several alternative sampling procedures; see references: Consigli and Dempster (1998), and Rush et al. (2000). The goal is to generate scenarios that minimize the sampling errors and lead to robust recommendations. Variance reduction methods can be applied as appropriate. The number of scenarios depends on the usage of the DFA system. For example, insurance companies require more precision than pension plans and accordingly a greater number of scenarios. It is not uncommon for a DFA model to employ 10,000 to 100,000 or more scenarios so that capital allocation decisions can be based on the DFA system.

The scenario generator module provides a foundation for the remainder of the DFA system—the corporate simulator and later the optimizer. Thus, company users should pay close attention to the scenarios coming out of the module. Some questions to ask: Are the resulting scenario paths sensible? See Figure 5 for a graph of short and long interest rates and inflation. Do the asset returns link to the underlying economic factors in a consistent fashion? For example, are bond returns derived from interest rate

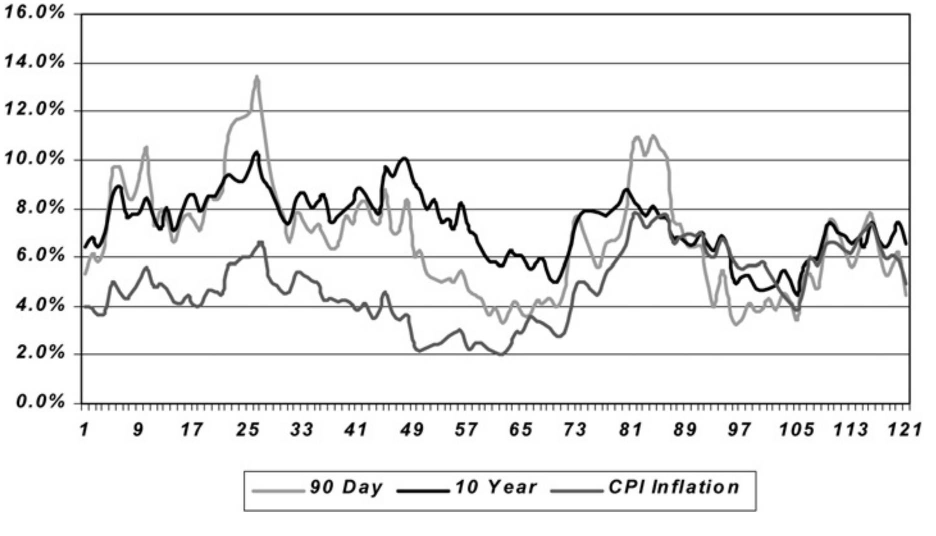


Fig. 5. Examples of interest rates and inflation over 10 year planning period (yield curve inversion occurs at two periods—months 25 and 86).

changes, especially the spot rate and yield curves? What are the summary statistics of the generated scenarios? Does the model include an adequate number of scenarios with tail events, such as recessions or yield curve inversions?

Scenario generators are established in the insurance liability arena. An important example involves estimating the probability of loss for covered properties under projected catastrophic events—mostly hurricanes and earthquakes. These stochastic generators play a critical function when insurance companies and state regulators approve insurance prices. Running these models helps in calculating a company's capital requirements. Insurance companies routinely evaluate their books of business with the catastrophic simulation models (from AIR, RMS, and EQE international, among others). 10,000 to 100,000$^+$ scenarios must be generated due to the large number of possible events and the resulting severe consequences and low probabilities for large losses (Bowers, 2002). Strikingly, the scenarios display losses with highly non-normal tail properties, i.e., high severity and low probabilities with highly skewed left tailed loss distributions (Mulvey and Erkan, 2006).

An emerging area for scenario generation entails weather-related projections and topics such as degree-days above or below average for selected locations and time-periods. This application will expand along with new weather-linked securities and derivatives as financial companies introduce products for hedging energy costs by energy producers and consumers. The growth of this market hinges on well-conceived, trustworthy scenarios for risk management.

Constructing a scenario generator requires attention to three key issues: (1) the realism of the model equations, (2) calibration of the parameters, and (3) procedures to extract the sample set of scenarios, the projection systems should be evaluated with historical data (back-testing), as well as on an ongoing basis. Confidence in these systems will continue only if they display a sufficient degree of accuracy. Research is needed in this area, especially with regard to points 1 and 2 above.

2.2. Enterprise simulators

The simulation module mimics the insurance company's decisions over the planning period for each of the presented scenarios. An insurance company, for example, makes decisions regarding their asset mix, business strategies for growing or shrinking their insurance lines, and the firm's capital structure (leverage, exposure, etc.). An insurance executive sets investment strategies along with policy for capital contributions, business expansion, etc. These decisions involve tax and other regulatory implications. The investment process consists of $t = \{1, 2, 3, \dots, T\}$ time stages. The first decision juncture represents the current date. The end of the period, T, the planning horizon, may depict a point in which the company has some critical planning purpose, such as the repayment date of a substantial liability, or more likely a point in the future consistent with the company's plans. A substantial portion of the simulation involves modeling the regulatory requirements including Statutory and GAAP positions, and the accompanying constraints on capital allocation. Due to the regulatory burden, it is difficult to operate solely on economic grounds. Thus, there is a need for a separate set of constraints and variables for the statutory, accounting and tax rules. See references (Burket, McIntyre and Sonlin, 2001; Cariño and Ziemba, 1998; Hoyland, 1998; Kaufman and Ryan, 2000; and Lowe and Stanard, 1996) for detailed discussions of the regulatory environment for insurance companies and other modeling complications.

The modeling problem becomes complicated when the value of the liability cash flow is a direct function of the asset returns and the company's annual decisions on crediting and payouts. This linkage occurs in life insurance. In this case, the resulting simulation system must be designed to closely link the asset allocation with the company's crediting policies. Asset and liability decisions are made in a coordinated fashion.

Next, we define the DFA model as a multi-stage stochastic program. The underlying approach builds upon the model in Mulvey, Correnti and Lummis (1997), with special attention to managing the asset and liability decisions within an insurance environment.

First, we divide the planning horizon $\mathbb{T}$ into two discrete time intervals $\mathbb{T}_1$ and $\mathbb{T}_2$, where $\mathbb{T}_1 = \{0, 1, \ldots, T\}$ and $\mathbb{T}_2 = \{T + 1, \ldots, \tau\}$. The former corresponds to periods in which investment decisions are made. Period T defines the date of the planning horizon; we focus on the company's position at the beginning of period T. Decisions occur at the beginning of each time stage. Much flexibility exists. At the tactical level, an active bond trader might see his time interval as short as minutes, whereas the CEO will focus on longer planning periods such as the next Board of Director's meeting. It is possible for the time steps to very-short intervals at the beginning of the planning period and longer intervals towards the end. $\mathbb{T}_2$ handles the horizon at time T by calculating economic and other factors beyond period T up to period τ. The investor does not render any active decisions after the end of period T.

Asset investment categories are defined by set $\mathbb{I} = \{1, 2, \ldots, I\}$, with category 1 representing cash. The remaining assets can include broad investment groupings such as stock subindices, fixed income groups, and real estate. The categories should track well-defined market segments. Ideally, the co-movements between pairs of asset returns would be relatively low so that diversification can be done across the asset categories. Most insurance companies require many fixed income asset categories due to the nature of the business and the underlying degree of financial leverage.

The business activities, mostly insurance products, are designated by the set $\mathbb{J} = \{1, 2, \ldots, J\}$. In many cases, we reference the insurance lines as separate activities. Depending upon the level of detail, we might aggregate multiple lines into a single j variable, or the aggregation can project common lines across geographical areas. The main feature is to be able to ascertain the losses with regard to the business activity, as a function of scenarios and time period. For this paper, we place the reinsurance decisions within a distinct activity (i.e., on the liability side of the balance sheet despite the fact that the resulting cash flows resemble asset categories).

A second partition of the enterprise tracks the profits and losses of distinct entities, called business units or alternatively divisions. The business units are identified by the symbol $\mathbb{K} = \{1, 2, \ldots, K\}$; we include a variety of decompositions within this identifier. For example, we might separate the company into two divisions: (1) the US insurance company, and (2) the European insurance company. Accordingly, we associate all of the assets and the liabilities to one of these two divisions, within a mutually exclusive and collectively exhaustive partition. Thereby, we evaluate the company in two distinct ways—by asset and liability (insurance products) categories and by distinct divisions. If a company maintains their assets without regard to the liabilities, we generally employ a basket of assets that represents the current asset mix. Of course, any matching of assets to activities is allowed from the modeling standpoint. It is natural to subdivide the enterprise according to accounting and regulatory boundaries since the divisional context will be employed for calculating profit and losses. For our model, we assume that the income statement can be calculated as a function of the decision variables (x, y, etc.) This part of the simulation can be rather complex due to insurance rules and regulations and therefore outside the domain for this exposition. Nevertheless, we assume that profits and losses can be calculated for each division as a function of the decision variables:

$$\rho_k^s = f(x, y, \mathbb{Z}),$$

where $\mathbb{Z}$ is the vector of key performance indicators and the decisions are set in the asset and liability space (x, y), whereas the profit/loss results, ρ_k^s, are taken in the divisional space $\mathbb{K}$. The $f()$ function encompasses technical issues, such as computing taxes, reserving for expected losses, addressing fixed versus variable costs, etc. We do not wish to minimize this part of the simulation, but the extent of the accounting rules and other regulations would overwhelm the statement of the basic DFA model. In addition at times, we may adjust the profit/loss values in order to improve the performance measures for example by engaging in reinsurance treaties (Section 4). The overall profit for the enterprise comes from summing the divisions' profits:

$$\mathbb{P}_t^s = \sum_k \rho_{k,t}^s.$$

We assume that the company's profit/loss is evaluated at the end of each period, using the enterprise as defined at the beginning of the period, with the results occurring over the reference time span. For convenience, we also assume that the cash flows are reinvested in the generating asset category and all the borrowing is done on a single period basis. This assumption can be readily dropped in an actual implementation.

For each $i \in \mathbb{I}$, $t \in \mathbb{T}_1 = \{0, 1, \ldots, T\}$, and $s \in \mathbb{S}$, we define the following parameters and decision variables. The elemental decision variables, such as asset allocation and business activity, refer to the state of the company at the beginning and during the time period.

Parameters.

$r_{i,t}^s$ Total return for asset i, time period t, under scenario s (projected by the stochastic scenario generator as discussed).

π_s Probability that scenario s occurs, $\sum_{s=1}^S \pi_s = 1$.

C_0^s Economic capital (wealth) in the beginning of time period 0.
σ_i Transaction costs incurred in rebalancing asset i at the beginning of time period t (symmetric transaction costs are assumed, i.e., cost of selling equals cost of buying; comparable parameter for liabilities σ_j).

Decision variables.

$x_{i,t}^s$ Economic value of asset i, at beginning of time period t, under scenario s, after rebalancing (X_t^s equals value of all assets; $xe_{i,t}^s$ equals asset value at end of period t).
$x_{i,t}^{+s}$ Value of asset i purchased at period t, under scenario s to rebalance the portfolio.
$x_{i,t}^{-s}$ Value of asset i sold at period t, under scenario s, to rebalance the portfolio.
$y_{j,t}^s$ Economic value of business activity j, at the beginning of time period t, under scenario s (Y_t^s equals value of all liabilities; $ye_{j,t}^s$ equals comparable value at end of period).
$y_{j,t}^{+s}$ Amount of activity j added for rebalancing in period t, under scenario s.
$y_{j,t}^{-s}$ Amount of activity j subtracted for rebalancing in period t, under scenario s.
C_t^s Firm economic capital (wealth) at the beginning of time period t, under scenario s. (c_k^t is the designated capital for division k.)
$\mathbb{P}_t^s$ Firm profit/loss during period t, under scenario s. (ρ_k^s is the designated profit/loss for division k.)

Given these definitions, we present the deterministic equivalent of the multi-stage DFA model (called the strategic planning model—SPM).

$$\text{(SPM)} \quad \text{Maximize } \mathbb{U}(Z_1, Z_2, \ldots, Z_m), \tag{1}$$

where the key performance indicators are defined by the vector: $[\mathbb{Z}_1, \mathbb{Z}_2, \ldots, \mathbb{Z}_m]$ and the company's utility function depicts the generic form for maximizing shareholder value. More is said about the firm's objective function below and in Section 4. Typically, there is no single objective function since the insurance decision environment encompasses many conflicting goals. For instance, the firm may wish to grow future earnings at the expense of the next period's earnings. Insurance companies have many stakeholders: insurance policyholders, shareholders, company employees, the public in the state, etc. We suggest that the overall problem be put into the framework of a multi-objective optimization model so that company executives can evaluate and compare the alternative business strategies.

The constraints for the asset side of the balance sheet are shown in Figures 6 and 7. A parallel set of constraints is needed for the business/insurance activities in the obvious way. These are omitted to simplify the exposition.

Subject to

$$\sum_i x_{i,0}^s = X_0^s \qquad \forall s \in \mathbb{S}, \tag{2}$$

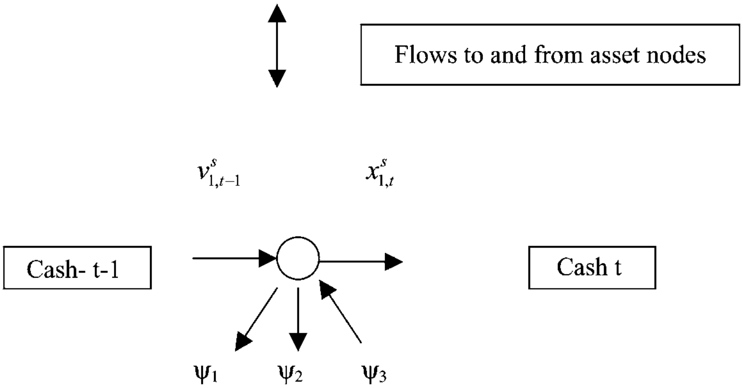


Fig. 6. Cash flow nodes appear at each time period (t) under each scenario(s).

$$\sum_i x^s_{i,t} = X^s_t \qquad \forall s \in \mathbb{S},\ t = 1, \ldots, T, \tag{3}$$

$$xe^s_{i,t-1} = r^s_{i,t-1} x^s_{i,t-1} \qquad \forall s \in \mathbb{S},\ i \in \mathbb{I},\ t = 1, \ldots, T, \tag{4}$$

$$x^s_{i,t} = xe^s_{i,t-1} - x^{+s}_{i,t} + x^{-s}_{i,t}(1-\sigma_i) \qquad \forall s \in \mathbb{S},\ i \in \mathbb{I}\backslash\{1\},\ t = 1, \ldots, T, \tag{5}$$

$$x^s_{1,t} = xe^s_{1,t-1} + \sum_{i \neq 1} x^{-s}_{i,t}(1-\sigma_i) - \sum_{i \neq 1} x^{+s}_{i,t} + \sum_{l=1}^{3} \psi^s_{l,t} \qquad \forall s \in \mathbb{S},\ t = 1, \ldots, T, \tag{6}$$

$$x^{s_1}_{i,t} = x^{s_2}_{i,t} \qquad \forall s_1, s_2 \in \mathbb{S} \text{ with identical past up to time } t. \tag{7}$$

The external cash flows for the firm arise from the following conditions: payments to the policyholders for losses, expenses and premiums, ψ_1; payments for dividends ψ_2; and any infusions of cash to replenish the capital accounts ψ_3, due to an unusual level of losses or in order to build the capital for greater underwriting capacity.

A generalized network investment model is presented in Figure 7. This graph depicts the flows across time for each of the asset categories. While all constraints cannot be put into a network model, the graphical form is easy for managers to comprehend. General linear and nonlinear programs, the preferred model, are now readily available for solving the resulting problem. However, a network may have computational advantages for extremely large problems, such as security level models.

Asset and liabilities appear in two forms. First, we calculate the economic surplus as:

$$C^s_t = X^s_t - Y^s_t. \tag{8}$$

The projected leverage factors under the scenarios will be a prime performance indicator: X^s_t / C^s_t as discussed in Section 4.

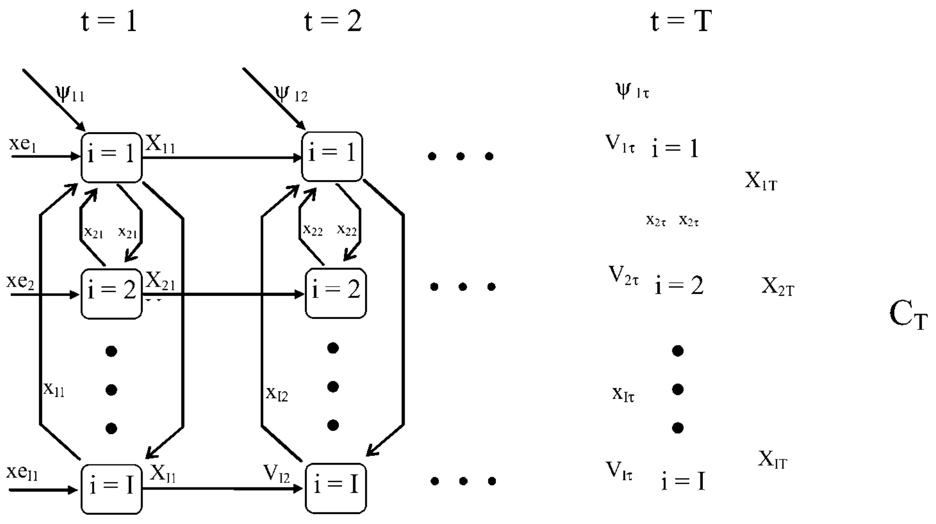


Fig. 7. Network representation of asset decision model (for a single scenario).

As with single-period models, the nonlinear objective function (1) can take several different forms. If the classical return-risk function is employed, then (1) becomes $\text{Max}\, Z = \eta\, \text{Mean}(w_T) - (1 - \eta)\text{Risk}(w_T)$, where $\text{Mean}(w_T)$ is the expected total wealth and $\text{Risk}(w_T)$ is the risk of the total wealth across the scenarios at the end of period T. Parameter η indicates the relative importance of risk as compared with the expected value. This objective leads to an efficient frontier of wealth at period T by allowing alternative values of η in the range [0, 1]. An alternative to mean-risk is the von Neumann–Morgenstern expected utility of wealth at period T.

Constraint (2) guarantees that the total initial investment equals the initial economic surplus. Constraint (3) represents the total surplus in the beginning of period T. As implied, this constraint can be modified to include assets, liabilities, *and* investment goals, in which case, the modified result is called the surplus wealth or net economic capital (Mulvey, 1989). Many investors render investment decisions without reference to their liabilities or especially their investment goals.

Mulvey employs the notion of surplus wealth to the mean-variance and the expected utility models to address liabilities and goals in the context of asset allocation strategies. Constraint (4) depicts the asset wealth $xe^s_{i,t}$ accumulated at the beginning of period $t+1$ (end of period t) before rebalancing in asset i. The flow balance constraint for all assets except cash for all periods is given by constraint (5). This constraint guarantees that the amount invested in period t equals the net wealth for the asset. Constraint (6) represents the flow balancing constraint for cash. Non-anticipativity constraints are represented by (7). These constraints ensure that the scenarios with the same past will have identical decisions up to that period. While these constraints are numerous, solution algorithms take advantage of their simple structure (Birge and Louveaux, 1997).

Model (SPM) depicts a split variable formulation of the stochastic DFA problem. This formulation has proven successful for solving the model using techniques such as the progressive hedging algorithm of Rockafellar and Wets (1991). The split variable formulation can be beneficial for direct solvers that use the interior point method.

By substituting constraint (7) back in constraints (2) to (6), we obtain a standard form of the stochastic allocation problem. Constraints for this formulation exhibit a dual block diagonal structure for two stage stochastic programs and a nested structure for general multi-stage problems. This formulation may be better for some direct solvers. The standard form of the stochastic program possesses fewer decision variables than the split variable model and is the preferred structure by many researchers in the field. This model can be solved by means of decomposition methods, for example, the L-shaped method—a specialization of Benders algorithm. See references (Birge and Louveaux, 1997; Dantzig and Infanger, 1993; Zenios, 1991).

Scenarios may reveal identical values for the uncertain quantities up to a certain period—i.e., they share common information history up to that period. Scenarios that share common information must yield the same decisions up to that period. We address the representation of the information structure through non-anticipativity conditions. These constraints require that any variables sharing a common history, up to time period t, must be set equal to each other. See Eqs. (7).

The multi-stage model can provide superior performance over single period models. The evaluation of the value of a business unit is more accurate when it is considered within a DFA system, as compared with the static valuation using simply discounting cash flows. As an example, asset category possessing relatively low returns and high volatility can still add value as long as the returns occur at the proper time period—when other assets are under performing (e.g., Mulvey, Ural and Zhang, 2007). A static single period evaluation would understate the value of these assets. Fernholz and Shay (1982) provides an example of volatility pumping in the context of a stochastic process.

Next, we develop a special case of SPM possessing a policy rule, called fixed mix or dynamically balanced. Other policy rules can be defined in a similar fashion. First, we set the proportion of wealth to be: $\lambda_{i,t}^s$ for each asset $i \in \mathbb{I}$, time period $t \in \mathbb{T}_1$, under scenario $s \in \mathbb{S}$. A dynamically balanced portfolio enforces the following condition at each time juncture:

$$\lambda_i = \frac{x_{i,t}^s}{w_t^s}, \quad \text{where } \lambda_i = \lambda_{i,t}^s. \tag{9}$$

This constraint ensures that the fraction of wealth in each asset category $i \in \mathbb{I}$ is equal to λ_i at the beginning of every time period. Ideally, we would maintain the target λ fractions at all time periods and under every scenario. Practical considerations, mostly transaction and market impact costs, prevent this simple rule from being implemented. Rather, we define a no trade-zone for modeling the rebalancing decision. The goal of this approach is to minimize trading within the no trade-zone and to rebalance the portfolio whenever the asset proportions fall outside their respective zones. Adding decision rules to model (SPM) gives rise to a non-convex optimization model. Thus, the search for the best solution requires specialized non-convex algorithms (e.g., Maranas et al., 1997).

2.3. Searching for superior recommendations

The complex nature of DFA requires a compromise among the various competing objectives. As part of the search procedure, we strive to identify recommendations that are not dominated by other solutions.[3] Multi-objective optimization is ideally suited to this problem. For instance, the company may accept an incremental increase in short-term risks in order to achieve its long-term profit objectives. Addressing questions of this form is made easier by finding a series of efficient frontiers for pairs of objectives. The use of optimization assists the decision maker by restricting the search to recommendations that lie on the efficient frontiers. Several examples are shown in Section 4.

Optimizing a stochastic model over a multi-period horizon is complicated by an expanding set of future conditional decisions. These decisions will depend upon the evolution of uncertainties that occur between period 1 and any future target date. The second and later stage decisions are conditional on the economic and other environmental conditions. Two basic approaches are available for modeling these conditional decisions, as discussed in the next section.

What are practical objectives for the company? First, it is generally agreed that the company's expected profitability (now and in the future) is a critical measure of shareholder value. Other measures include free cash flow and related metrics such as economic value added. These indicators define the company's return objective, for each of the time periods. For simplicity, the DFA system reports profitability at several key time junctures—for example, at the end of the first year, and at the end of 3 to 5 years.

2.3.1. Risk measures

The next objectives involve evaluating risks. Several alternative metrics have been employed, including downside risks (expected value below target return), selected quantiles of the profit/loss distribution (value at risk), expected policyholder deficit, volatility of earnings over selected time periods, and the probability of a downgrade in the firm's credit quality (e.g., Mango and Mulvey, 2000). See Artzner et al. (1999) for properties of risk measures, and Rockafellar and Uryasev (2000) for a critique of VaR and an alternative measure (Conditional VaR). Also see Zenios (1993) for a discussions of applications of risk management via optimization models. It is impractical to select any single risk measure for large organizations. We suggest presenting alternative definitions of risks to senior management.

As expected, the risk measures focus on the left (losses) tail of the company's profit/loss or surplus distribution. Many US primary insurers maintain adequate capital

[3] Dominated solutions have the usual interpretation—better objectives along one or more dimensions, while keeping all other dimensions at the same level of utility.

to protect the firm at the 99% confidence level, i.e., they expect that the company possesses sufficient capital to weather 99 times out of 100 simulated years. As a modeling issue, the tail of these distributions requires non-normal functions. Normal distributions generally underestimate the extreme events. Insurance companies have several options to reduce their exposure to the rare events—by diversification, by reinsurance, by securitization, and by selective underwriting and pricing decisions.

2.3.2. Comparing optimization approaches

Today, there are two practical approaches for optimizing a multi-period DFA system.[4] The first involves stochastic programs. There have been several implementations of multi-stage stochastic programming models in the insurance arena. See Cariño et al. (1994) and Hoyland (1998). The resulting stochastic programming approach is characterized by a scenario tree representation of the uncertainties across the planning period.[5] The scenario tree begins at the current date with a single node—depicting the state of the company before any decisions have been made. At this juncture, decisions are suggested to improve the company's operations, for instance, by altering the asset allocation, revising the reinsurance agreements, etc. The actual set of plausible decisions depends upon many issues, including past agreements, current market opportunities, company policies, current surplus levels, and turnover constraints. For our purposes, we do not specify the constraints that a company must adhere to. Rather, the DFA modeler must fill in these essential details when constructing his or her mathematical model. See Burket, McIntyre and Sonlin (2001), Hoyland (1998), Kaufman and Ryan (2000), and Lowe and Stanard (1996) for illustrations of constraints employed within the insurance industry. Also, see Boender (1997); Consigli and Dempster (1998); Dert (1995); Frauendorfer and Schurle (2000); Kusy and Ziemba (1986); Worzel, Vassiadou-Zeniou and Zenios (1995); and Zenios (1998) for applications in other financial domains.

Each node in the scenario tree depicts a set of decisions—indexed by scenario, by time period, and by decision variables (asset, liability, etc.). Branches in the tree depict the unfolding spectrum of uncertainties over the planning period. As mentioned, underlying the scenarios is a system of stochastic differential equations. After the initial node, decisions are conditioned upon the economic environment along the specified branch. For example, we will likely render one type of decision for the upper branch and another decision for the lower branch. The stochastic program will render the optimal decisions at each node in the scenario tree.

The stochastic programming process is completely general, with limits due to explicit linear constraints imposed by the modeler. For instance, he might wish to limit turnover

[4] Dynamic stochastic control can be employed in certain simplified circumstances. See Ziemba and Mulvey (1998).

[5] Strictly speaking, uncertainties can be modeled by means of continuous functions. However, a sampling scheme is needed to render the problem tractable. See Birge and Louveaux (1997) and Consigli and Dempster (1998) for discussions of scenario selection.

of assets at any revision point to a fixed percentage of total assets—say 5%. Clearly, an advantage of a stochastic program is its ability to anticipate a problem developing in the future and to plan its way around the difficulty, or even better to avoid the problem entirely. The model will find the best set of decisions, based on the scenario tree, the model's stated objective functions, and the stated constraints.

An alternative to stochastic programming involves developing a set of policies to guide the company across the planning period at each decision node. Rather than allowing complete generality with regard to the range of decisions, we impose a discipline on the decisions through a pre-determined policy (or policies). Optimization proceeds by finding the best set of policies along with the optimal set of parameters.

In selected cases, policy optimization provides a practical alternative to multi-stage stochastic programming. Herein, the decisions at each node of the scenario tree depend upon the "state of the system", rather than any specific scenario. The goal of a policy model is to simulate the organization (a.k.a. Monte Carlo simulation) within an optimization context. A policy rule can be simple or complex, depending upon the company's ability to implement the process with available ongoing information. Perhaps, it is best to describe a few examples of policy rules:

Asset allocation:

- Fixed-mix asset allocation (with or without no-trade-zones).
- Surplus-equity allocation (e.g., constant proportional portfolio insurance). See Perold and Sharpe (1988).
- State dependent allocation (optimistic, neutral, pessimistic on interest rates).

Pricing insurance policies:

- Price as a function of economic environment and competition.
- Risk adjusted profitability with limited capacity—select policies that meet target marginal RAROC.

Reinsurance policy:

- Insure all losses above a specified threshold for the company as a single entity.

In each case, the simulation determines its current state based on the set of information available at the time of the decision. It is relatively straightforward to include path dependent information within the framework by adding accounting and related variables. For instance, we can determine the company's economic surplus by projecting future cash flows, such as liability estimates. The asset allocation can then depend upon the surplus value at node $k \in K$.

Within a policy rule, there are ranges of variations that depend upon the selected parameter settings. A popular example is the fixed-mix asset allocation, where the parameter settings equal the asset proportions. The standard benchmark for many investors is the 60/40 mix—60% stock, and 40% bonds. Policy models have been implemented by a number of organizations, including Towers Perrin–Tillinghast (Mulvey, Gould and Morgan, 2000), American Re (Berger and Madsen, 1999), Swiss Re (Burket, McIntyre and Sonlin, 2001), and Renaissance Reinsurance (Lowe and Stanard, 1996), among others.

Each of the two approaches—stochastic programming and policy optimization—has something to offer. Policy optimization is perhaps the easiest to implement. Of course, the scenario generator must be constructed and validated, along with a dependable policy rule such as the fixed-mix asset allocation. The resulting model becomes a Monte Carlo simulation in which the policy rule and the accompanying policy parameter setting are fixed. In addition, we can readily calculate the sampling errors and related statistical estimates. The recommended solutions can be evaluated by means of sensitivity analysis regarding the assumptions within the scenario generator.

Stochastic programming has the potential for improving the recommendations, as compared with policy optimization. There are several provisos. First, the number of scenarios must be large enough to prevent the model from mis-estimating the full range of uncertainties. Second, the model must be solvable in a reasonable computer run time so that sensitivity analyses can be conducted. Third, the recommendations of the stochastic program should be understandable to the high level executives who are ultimately responsible for the decisions. These issues present challenges, but advances in computer hardware and data accessibility is improving the situation.

3. Applications of DFA

In this section, we present several case studies taken from real-world applications. These examples show the benefits of an integrated framework for decision-making. In addition, we present samples of model output to illustrate its important role in a DFA implementation. One could build the perfect DFA model, but it would be completely useless if its results could not be presented and interpreted in a meaningful way.

3.1. Company A—Optimal capital allocation

The first use of DFA is to examine whether, given the currently employed strategies relating to investment and insurance risk, there is adequate capital to protect the company, and to demonstrate this to third parties such as regulators. Along with other financial intermediaries, property casualty companies are constrained by the amount of capital they have available for investment. Unlike most other industries, though, the capital allocated to a specific insurance operation is not additive. A dollar invested in writing workers compensation policy plus a dollar invested backing a commercial auto policy does not require two dollars of capital.

The basic efficient frontier taken from a multi-period DFA model is illustrated in Figure 8, which shows in simplified form the usual trade-off between opportunities to obtain higher expected profit by accepting greater risk. On the vertical axis is the level of expected return on equity (ROE). Moving up the vertical axis would involve increasing

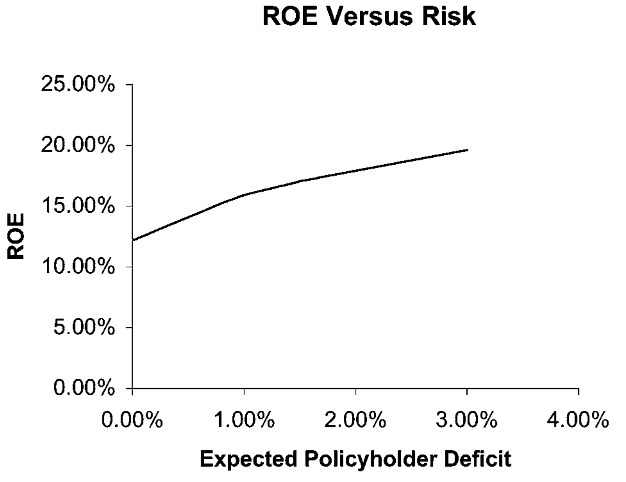


Fig. 8. Multi-period efficient frontier for insurance Company 1 (3 year horizon).

exposure to various risks such as interest rate risk, credit risk, liquidity risk and stock market risk. The horizontal axis represents the amount and type of insurance risk underwritten, including exposure to different lines of business. Given constraints, perhaps imposed by regulators, rating agencies or given economic constraints, and assuming an amount of capital, there is a limit to the amount of investment and insurance risk taken.

Company A is a global insurance company with multiple lines of business, operating in 18 countries. Company A wished to know the amount of capital required to operate its business. To correctly assess the amount of capital required, Company A developed economic and financial models that met the following criteria:

(A) The scenario generator must model economic variables for multiple countries in a globally consistent manner.

(B) The economic variables in the scenario generator must adequately capture the financial and economic risk factors affecting their business.

(C) Asset class returns must be handled consistently within the model. For example, the return on stocks earned by a US investor must be consistent with the return on US stocks earned by a UK investor, with the difference being the simulated currency return.

(D) Financial results must be modeled by line, by country and then aggregated into a single set of results, denominated in Company A's home currency.

By employing DFA, Company A was able to explore the portfolio effect realized from its globally diversified businesses. As a result of this portfolio effect, Company A discovered that it was able to hold 18% less capital, freeing up more capital to invest in expanding its insurance business oversees. By holding less capital and by investing this newly freed capital in its growing businesses, Company A was able to earn a higher

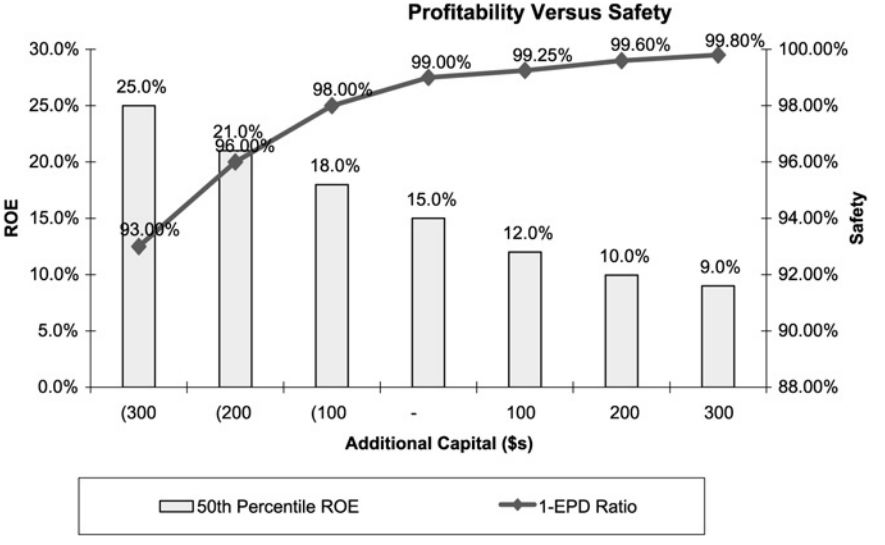


Fig. 9. Tradeoff among capital, expected return, and safety.

expected return on a smaller equity base, thereby increasing the return on shareholders equity from 11.2 to 14.3%.

Figure 9 depicts the impact of adding capital and the commensurate impact on the insurance company's performance. Here, the probability of a rating downgrade is plotted against the expected return on equity, for various levels of capital. For instance, by adding $100 million in capital, the probability of a downgrade is reduced from 1% to 0.75%. The executive management committee must make decisions regarding the proper level of capital. Too much capital and the company's profits will sag. Too little capital and the risks will become unacceptable. The DFA system assists in pinpointing the best compromise values for these tradeoffs.

3.2. Company B—Asset allocation

Within the traditional Markowitz framework, asset allocation is determined by maximizing risk-adjusted asset return over a single period. For an insurance operation, however, this is inadequate for several reasons. It ignores the liabilities, which are influenced by factors that impact market prices. An efficient portfolio in asset return space is unlikely to be efficient in asset–liability space. By using a single time period, the model ignores the illiquid and long-term nature of many insurance liabilities, which creates opportunities to take on liquidity risk, but also creates exposure to reinvestment risk. It requires company management to evaluate stakeholders, quantify the risk aversion levels of each stakeholder, and then to weight them in a way to satisfy all parties. This is impractical, where many parties (e.g., policyholders) will possess differing levels of risk aversion.

DFA helps address each of these issues. Once the model is developed, an optimization routine identifies recommendations to maximize a given risk measure. However, there are several differences. A popular choice for the reward measure is the economic value of the distributable earnings of the company, assuming that it remained open for a number of years, T_1 and then closes to new business. The concerns of other stakeholders are dealt with as constraints on the optimization process. The reward measure is a multi-period one, e.g., present value of all distributable earnings. This can capture the liquidity constraints on the liabilities and reinvestment risk on the asset side. These issues give rise to a non-convex problem; the optimization technique must be able to cope with the non-convex nature of the DFA problem.

Company B operates in Bermuda, selling reinsurance to P&C companies across the globe. It wishes to optimize the present value of distributable earnings assuming the company continues to write new business over five years and then closes to new business, subject to the following constraints. The measures used by regulators and rating agencies to assess the companies (surplus to written premium, for example) should not be breached ‘too often’; the volatility of reported (e.g., GAAP) earnings should not be ‘too high’.

The company’s original asset mix consisted of short-duration fixed income securities so that cash could be readily available for projected payouts. A DFA model was built to analyze asset strategies. As an alternative to the fixed-income portfolio, the optimization model recommended a strategy in which a portion of the assets should be placed in equity. In particular, the model suggested that 65% of the company’s surplus (above economic value of the liabilities) be placed in a US index fund. This strategy improved the company’s surplus at the end of the five-year planning horizon, as compared with the short-duration portfolio (Figure 10). In fact, the surplus equity strategy almost stochastically dominates the so-called safer “cash equivalent” strategy.

The DFA model took up several other issues for Company B, including the pros/cons of hedging the currency risks for insured payouts, i.e., matching liabilities with currency hedging. This currency hedging strategy did not improve the risk profile since the losses occurred at random and rare time periods. In fact, hedging the insured losses increased the volatility of earnings since the company reported earnings in US dollars. The DFA model showed that improving the asset allocation would increase the company’s growth in earnings, without a noticeable increase in enterprise risks.

3.3. Company C—Capital allocation for a large multinational insurance company

An important use of DFA is the allocation of capital to insurance product lines. These decisions affect capital budgeting, and determine the risk-adjusted return on capital to reward management.

A key issue in capital budgeting relates to the appropriate risk adjustment to apply to cash flows. It could be argued that beta to the stock market should be used to calculate

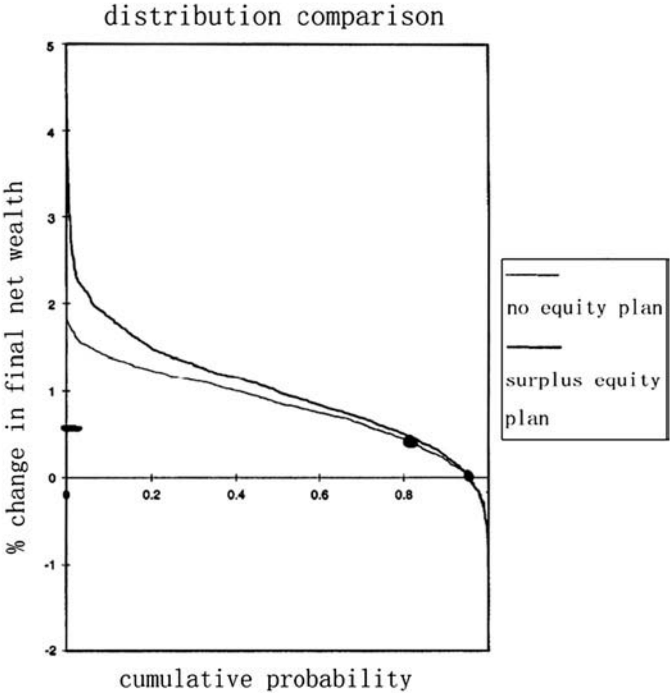


Fig. 10. Improvements to profitability by adding equity to asset portfolio.

the risk adjustment. This simple adjustment does not reflect the skewed distributions of insurance payouts over the time frames used to assess company management. A second issue involves allocating the ‘diversification’ benefit. Adding a relatively independent line to a portfolio of claims does not increase the capital requirements in a linear fashion. The diversification benefit could be allocated back to the various lines in order to improve the company’s competitive position. The allocation should reflect the fact that the degree of association between two lines is not constant. For example, workers compensation claims and personal property claims may be weakly correlated. However, following a large catastrophe property claims will be high. Overtime may be high as rebuilding occurs and the work place may not be as safe. These relationships would impact the severity and frequency of workers compensation claims, leading to increased correlation ‘in the tail’ of the distribution.

Company C is one of the largest insurance companies in the world, with its headquarters in a European capital and offices throughout the world. The company has grown rapidly over the past decade by expanding into Asia and the US, primarily by mergers and acquisitions. Company C has implemented a global DFA system in order to allocate the firm’s capital to product lines and countries in the most cost effective manner. The company must conserve its capital in order to continue to achieve its target growth plans. The DFA system shows the benefits of global diversification. Regulators are given access to the model's recommendations so that they can gain confidence in the company's ability to maintain a safe margin. In addition, company executives are shown the benefits of certain executive decisions, such as shrinking business in selected areas and encouraging growth in other domains. By allocating capital in an optimal fashion, Company C has a distinct advantage over its competitors and has grown accordingly.

3.4. *Company D—Reinsurance decisions*

Reinsurance is sometimes referred to as renting the reinsurer's balance sheet. A company with a competitive advantage in underwriting may write more business than would be available given their own capital. In this case they would cede some of the business to the reinsurer. DFA can identity the optimum way to reinsure the exposure. In practice, there would be a small number of options available to the insurer, and the process would involve evaluating each available opportunity to assess the 'best' one.

Company D is a US based P&C insurer with 30% exposure to non-US insurance business. In the past, the company maintained a sequence of reinsurance treaties, one for each of their five lines of business and a general coverage treaty for US losses between $500 and $750 million. The combined cost for the reinsurance was approximately $85 million per year. A DFA system was built, recommending a series of changes to the company's operations. One set of modifications was to reduce their exposure to commercial lines by roughly 15%, especially in earthquake prone areas, and to combine the reinsurance treaties into a single large coverage. The total benefit to the company was to increase their expected profit from 9.8 to 12.7%, without any increase in the overall corporate risks (as measured by several metrics including EPD and shortfall risks).

At first, the reinsurance recommendation was difficult to implement since the heads of each of the business lines wanted protection for their respective operations. In addition, several state regulators were concerned about the creditworthiness of the reinsurer, despite its triple-A rating. A compromise resulted, whereby the company set up a small captive insurer in Bermuda to allay the division heads and split the reinsurance treaty into two pieces with two reinsurers. The company reduced its risk and made a modest improvement in profits by these adjustments. More importantly, the senior executives became aware of the impact of individual reinsurance decisions on the overall company profit and surplus.

3.5. *Examples of model output*

Next, we present specific examples of how the model's output can be used to evaluate an enterprise. These examples show the benefits of a flexible reporting system for preparing reports and graphs for management discussions and presentations.

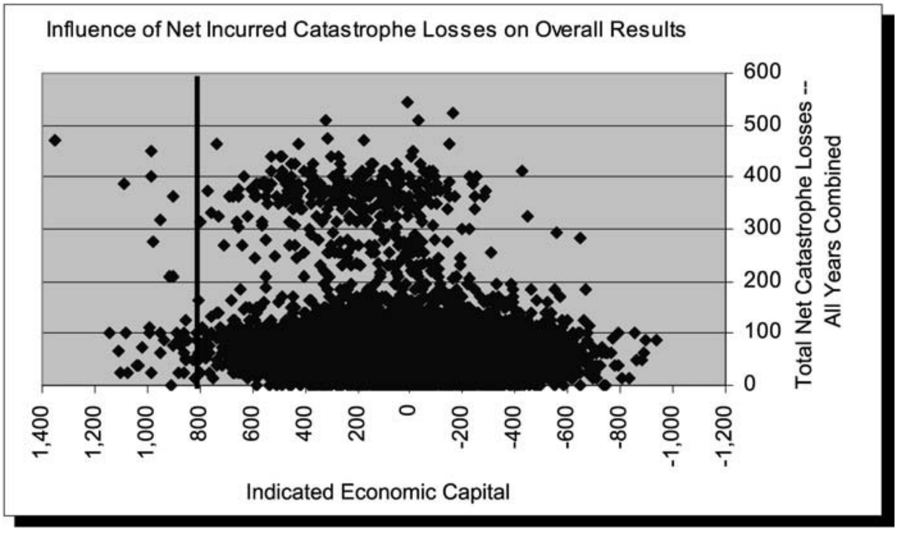


Fig. 11. Impact of CAT losses on economic capital.

3.5.1. *Example 1—Impact of catastrophes on economic capital*

The DFA system can be used to examine the relationship between variables to determine where the risks lie. This example shows the results of an analysis performed for a large property casualty insurance company. The management wanted to determine whether enterprise failures were being caused by catastrophes. Each business unit ran a DFA model based on a common set of scenarios to produce the sources of failure. Figure 11 shows the results of these analyses. The company established a maximum level of economic capital, indicated by the vertical line. Scenarios with indicated capital in excess of this threshold were considered failures. Management expected to see a positive relationship between catastrophes and failures. However, because catastrophe losses rarely exceeded reinsurance limits, there was no relationship between catastrophe losses and failures.

Often, the DFA model will confirm management's expectations. However, sometimes the DFA model will present results that run counter to management's intuition. In these situations, an added value of the model is realized—when unforeseen risks become apparent.

3.5.2. *Example 2—Enterprise diversification benefit*

The DFA system can also facilitate the exploration of the diversification benefit of a multi-line, multi-country insurance enterprise. This example shows the results of an analysis performed for a large multinational, multi-line insurance company. The com-

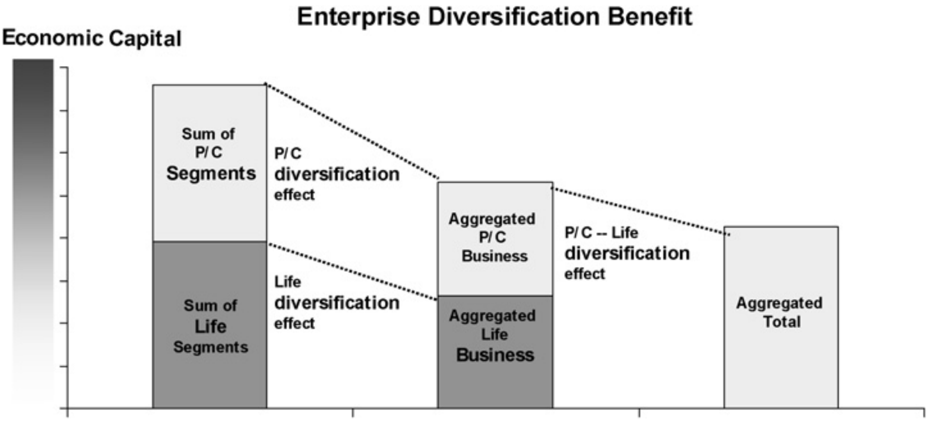


Fig. 12. Computing diversification benefits for multinational insurance company.

pany wanted to quantify the benefit from being diversified across multiple insurance lines with operations in different countries. The analysis began by dividing the enterprise into two business segments—life and property/casualty. The business segments were further divided by insurance line and by country of operation. Each line in each country then produced a DFA model using a set of consistent economic scenarios and the model results were combined. Figure 12 shows the required economic capital for the combined enterprise. The left-hand bar shows the amount of economic capital required when the enterprise is viewed as the sum of all the components. The middle bar shows the amount of economic capital required when the model results are aggregated at the business level. This allows for quantification of the diversification benefit attributable to multiple lines of business operating in multiple countries. The right-hand bar shows the required economic capital when the model results are aggregated at the enterprise level. This allows for the quantification of the diversification benefit attributable to the life and property/casualty business segments.

As a result of this analysis, the enterprise was able to quantify the benefit from operating multiple lines of business across multiple countries. At the present, this benefit is being held at the enterprise level rather than being distributed back down to the various business segments. Naturally, different enterprises may choose to distribute the benefit differently, depending upon their objectives. More is said about this issue in Section 4.

3.5.3. Example 3—Asset allocation

Studies have shown that asset allocation explains about 94% of the variability in portfolio returns (Brinson, Hood and Beebower, 1986). Hence, asset allocation policy is an important determinant that affects the financial health of the organization. In this next example we show how the DFA system can be used to aid in making asset allocation decisions. We begin by optimizing using the approach described in Section 2. Risk and reward can be defined in a number of ways. Reward is usually expressed as a desirable financial outcome. Risk is usually expressed as a function of the various performance indicators, although multiple indicators can define it. At its simplest, risk is equal to the standard deviation of the reward or as a target semi-deviation measure (i.e., down-

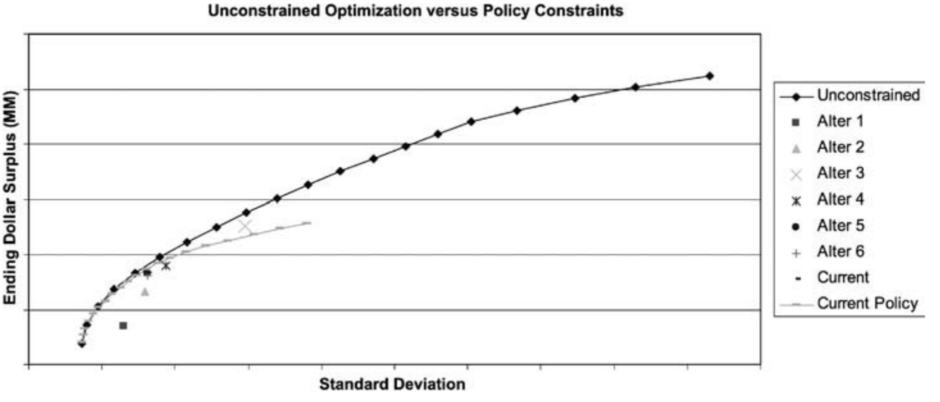


Fig. 13. Efficient frontier for company at end of multi-period DFA model.

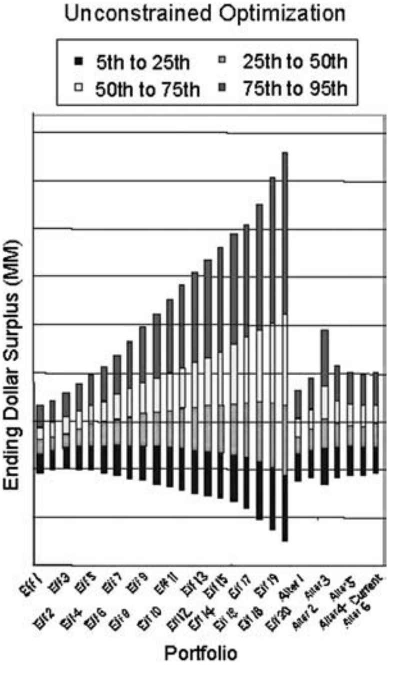


Fig. 14. Probability distribution of ending surplus at various points on efficient frontier.

side risk below a specified return value—profit or surplus). In this example, we specify reward as ending surplus, and risk as the standard deviation of ending surplus.

Perhaps the most relevant report for assessing asset allocation is an efficient frontier. Figure 13 presents an example of an asset/liability efficient frontier, where the reward axis (y-axis) shows the expected surplus and the risk axis (x-axis) shows the volatility of surplus. Current and alternative policies can be shown together on the efficient frontier to demonstrate the risk/reward characteristics of each. A floating bar graph allows us to examine the percentile distribution of various alternatives. Figure 14 depicts a floating bar graph of the efficient asset mixes, current mix and alternative mixes.

These three examples and the preceding ones show the benefits of an integrated approach to evaluation and decision-making. The impact of any major decision can be evaluated on the entire organization. The value of each activity can be ascertained with regard to its benefit as a standalone business or as part of a portfolio of insurance products and services. The next section takes up these issues in more detail.

4. Capital allocation and decentralized risk management

The goal of a large multinational insurance company is to maximize its shareholder value, by means of growing the stock price with minimum volatility, by increasing the company's economic surplus and by improving related performance measures. Operational inefficiencies may arise if a central committee must approve major decisions. Large financial companies operate on a global basis; it seems unlikely that a centralized decision making structure to be workable in a worldwide setting. (See Stein (2002) for a discussion of alternative organizational structures.) Thus, we turn to methods for improving a large organization by means of decentralized processes. A standard approach in risk finance is to allocate the firm's capital to divisions or business units.[6] Here, each unit operates as a standalone basis (at least in form) with its profits shown as a proportion of the company's capital allocation.

As before, we assume that the corporate entity has been separated into distinct business units. There is much generality to this structure; for example, a unit can represent a new activity such as adding a new service to its businesses—selling life insurance in addition to auto insurance. This activity might be deemed a project, but for our purposes it is called a business unit. A unit can be the designated corporation within a country, aligned with regulatory oversight and accounting requirements. A large multinational financial company will often possess a relatively large number of units (divisions), depending upon the organizational set up and the regulatory constraints. The capital allocation constraint is as follows:

$$\sum_k \alpha_k \leqslant C, \tag{10}$$

[6] We allow all parts of the organization to be business units—including independent companies, etc.

where the assigned capital for business unit k is α_k and the total corporate capital is equal to the variable C within the DFA model.

For the first two subsections, we assume that headquarters can bring together the basic information for the corporation from each of the business units and can run a full DFA system. We relax this assumption in Section 4.3.

4.1. Bottom up capital allocation

Capital allocation can be carried out in two basic ways. First, a bottom up analysis can be performed as shown in Figure 15. Here each unit is evaluated in turn, considering the various risk metrics and the associated amount of required capital. In order to insure consistency, each unit should employ the same set of scenarios and probabilities, coming out of the scenario-generating module. By employing this approach, we develop a DFA model for business unit k so that the optimal decisions are made with reference to the situation of the unit on a standalone basis. In this context, the level of capital is an endogenous variable within the DFA model for each business unit. We assume that the unit will maximize a VM utility function or equivalent risk/reward function.

The capital for each unit is equal to α_k for division k. The total required capital for the enterprise is then the sum of the required amounts—$C = \sum_k \alpha_k$.

There are several advantages. It fits traditional regulatory procedures since it aligns with standard processes. Local decision makers who are able to address local concerns can readily manage the procedure. Compensation can be linked to the outcomes of the

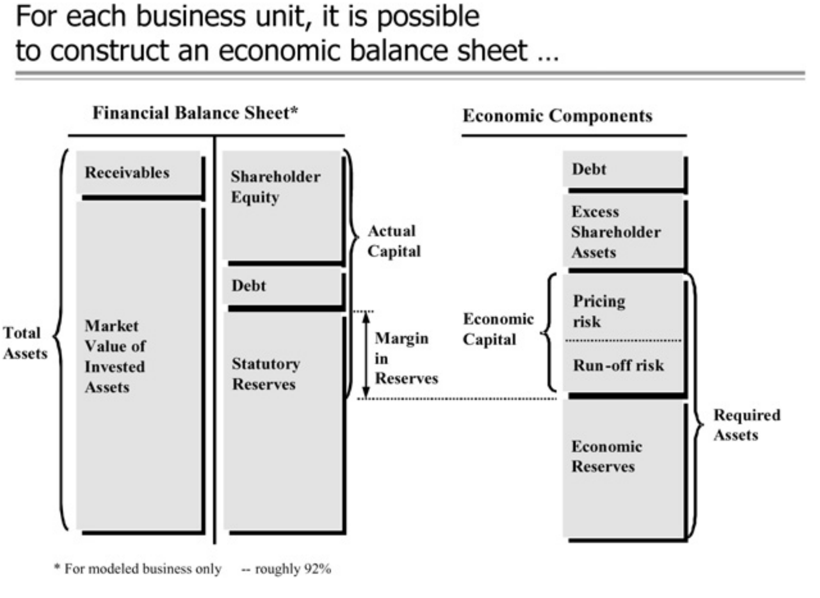


Fig. 15. Bottom up capital allocation.

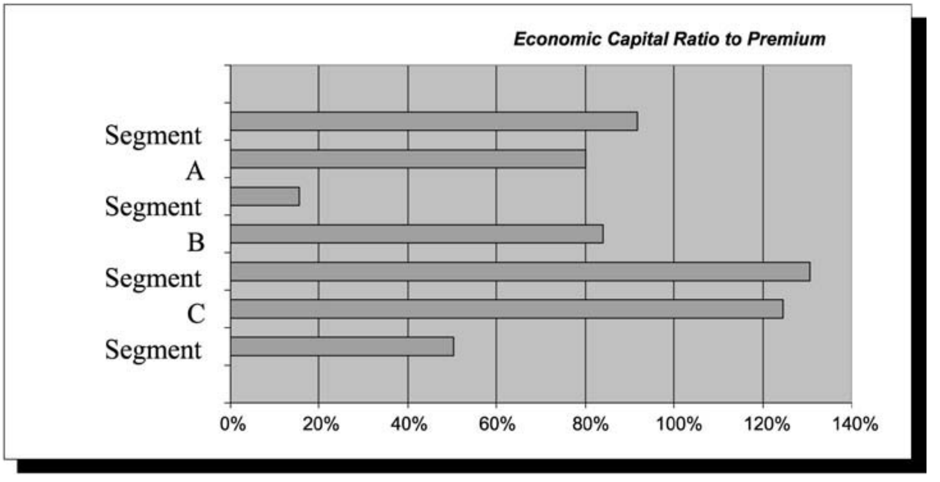


Fig. 16. Varying levels of capital for business units.

divisions, without the need for any intervention from corporate headquarters. Implementation is relatively straightforward once the scenario generator is made available for each unit.

As a disadvantage, the capital allocation may lead to inefficient operations by requiring excess capital. The method may underestimate diversification benefits or ignore concentration exposures since it is difficult to coordinate the capital requirements across units. As a variant, capital allocation can be made using expert opinion regarding the business prospects and other aspects of the kth division. Given the specified capital, the division operates in a way to maximize its value. The corresponding decision variables are rendered by means of the local DFA system.

The final step of the top–down approach is to assemble the various plans for the business units into a single DFA system. Under the combined plan, the decision variables are fixed as determined previously. Information flows from the business units up to the headquarters. At this point, the company's capital requirements can be calculated based on the enterprise risk-reward measures. Depending upon the risk measure, the overall company will typically find that there are at least some diversification benefits for operating a multi-product or geographically separated company. These benefits can be transferred back to the business units or they might be held at the headquarters for use in future mergers or acquisitions. As a technical point, certain risk measures such as the standard Value-at-Risk will not always lead to lower capital requirements due to concentration of risks. Subadditive risk measures (Artzner et al., 1999) will avoid this problem. However, these measures may be too conservative, requiring excess capital for the overall enterprise. The top–down strategy, proposed next, avoids this potential pitfall.

The bottom up approach has been implemented in one of the largest global insurance companies. The DFA models were developed and implemented for 15 business units,

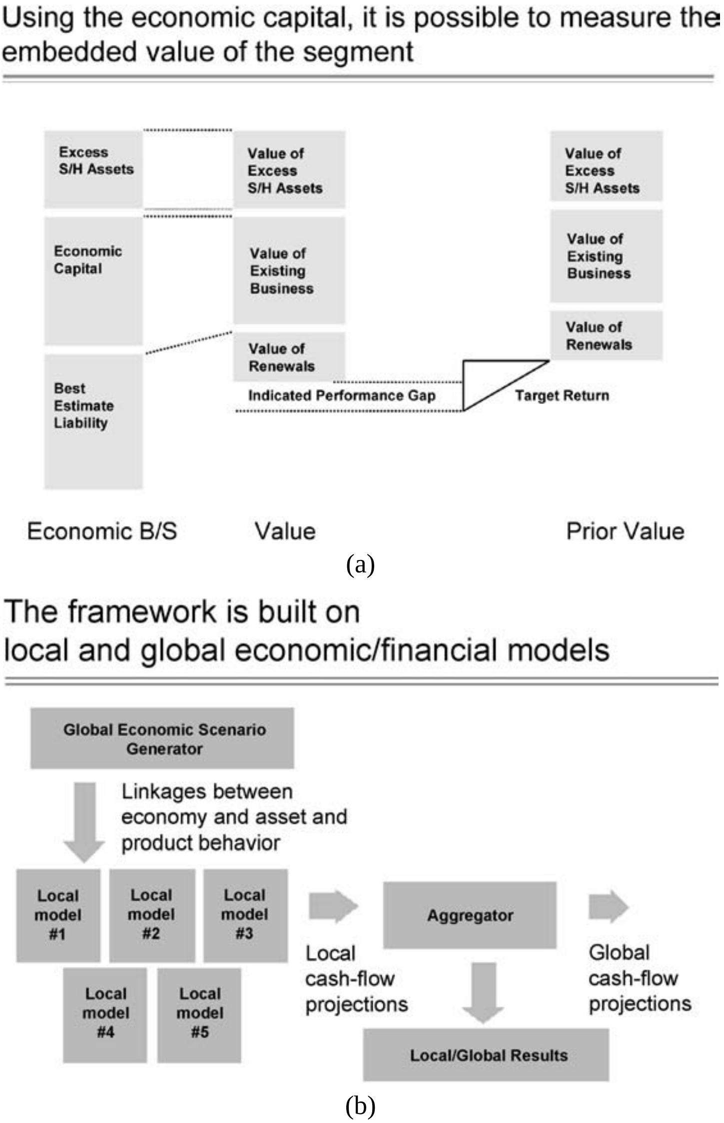


Fig. 17. (a) Measuring value of activities via bottom-up DFA. (b) Measuring value of activities via bottom-up DFA.

operating in 8 countries, 7 Life units and 8 P/C units. Typically, the models incorporated 90^{+}% of the business in each unit. In addition, client teams in each unit were trained in the economic capital approach, and built their DFA model. Each entity presented its results to a central management group, with the overall results presented to client's board of director. In this context, the bottom up approach has shown itself to be a practical procedure.

A few comments are relevant regarding this implementation. First, the different types of business require markedly different levels of economic capital (Figure 16). These differences are sometimes greater than conventional wisdom or rules of thumb would advise. Also, the differences from regulatory requirements are also marked—and slice in both directions. In some business segments the economic capital requirements are entirely met by margins in reserves, whereas in other business units the DFA strategies materially affect capital requirements. The economic capital assists in the measurement of the value of the units and forms the basis for performance evaluation measures (Figure 17). A potential disadvantage of the approach is the need to re-run the DFA system when significant changes in the corporation occur.

4.2. *Top–down capital allocation*

Under this approach, the firm's capital is determined by means of the integrated risk management system at the headquarter level, aimed at modeling the entire organization. The resulting DFA system is run for the entire organization at a relatively detailed level. Then, given these results that include diversification benefits and concentration penalties, the projected capital allocation values for each division are computed and communicated to the division heads. See Figure 18 for a flow chart of the resulting information flows.

Running the DFA system at headquarters accomplishes several aims. First, the required capital for the enterprise is calculated as a single entity. The decision variables can be set so that diversification benefits are optimizing along with expected profits and risk measures. In addition, the projections of rare events can be done without assuming a multivariate normal distribution for losses. The enterprise is optimized in systematic fashion, based on the best global information.

Unfortunately, several difficulties may arise when a centralized DFA system is run in a large multinational insurance company. Mostly, the information required and the or-

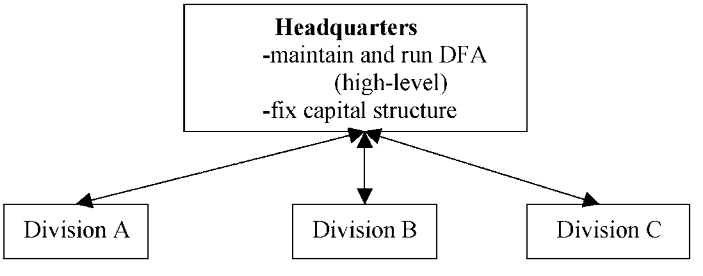


Fig. 18. Top–down capital allocation. (Information flows down from headquarters to divisions, e.g., capital allotments, with divisions supplying projected performance measures based on headquarters allocations and/or prices.)

ganizational support structure is often inadequate at headquarters. There are simply too many constraints and relationships needed to find the optimal strategies for the business units. Also, the units can be hesitant to follow the dictates of a central administrator. As a consequence, the top–down approach, while theoretically attractive, poses a number of implementation barriers. The next section takes up the topic of finding a practical approach.

4.3. Decentralized DFA

This section takes up the issue of managing a company within a decentralized setting. Rather than assuming that headquarters reviews the major decisions, we desire to evaluate decisions on an ongoing basis without running the DFA system. The proposed approach is based on a well-formulated theory in optimization—called large-scale optimization (see Lasdon (1970) for an early reference). The goal is to combine the best of the top–down and the bottom-up approaches, by sending information between headquarters and the business units in a systematic fashion. To develop the approach, we first define the standard valuation formula for evaluating new activities, using risk adjusted discounting. We call this a static evaluation formula:

Static valuation.

$$v_j = \sum_t c_{t,j}/(1 + r_j)^t \quad \text{for } j \in \mathbb{J}, \tag{11}$$

where the cashflows (both positive and negative) are defined for division j at time t as $c_{j,t}$, and where the discount factor r_j depends upon the calculated risks for division j (called the risk adjusted discount factor).

Next, we show that this standard valuation calculation can be expanded to address the marginal benefits (if any) when evaluating new projects/activities (see below). Beforehand, we extend the static evaluation to consider the cost of capital within the valuation formula:

Risk-adjusted profit (relative version).

$$ROP_{j,t} = \frac{\rho_{j,t}}{\alpha_{j,t}}, \tag{12}$$

where the profit for division j at time t is $\rho_{j,t}$, and its capital is $\alpha_{j,t}$ and where the risk adjusted profit for division j is $ROP_{j,t}$.

There is a large body of literature on the calculation of risk-adjusted returns on capital, largely coming out of banking applications. See Doherty (2000), Matten (1996), Shimpi et al. (2000), Zaik, Walter and Kelling (1996). An alternative to the ROP formula is one that charges the units an amount based on the capital allocation and the risk engaged in:

Fig. 19. Information flows for decentralized DFA system.

Risk-adjusted profit (value added version).

$$VA_j = \rho_j - \alpha_j, \tag{13}$$

where the profit for division j is adjusted by subtracting its cost of capital, α_j.

The value added approach provides an alternative to the relative calculation; these concepts are known as Economic Valued Added (EVA) and promoted by firms such as Stern Stewart (Ferguson and Leistikow, 1998; Uyemura, Kantor and Pettit, 1996).

Capital allocation decisions have a substantial impact on relative profitability and "value" for each division and as a consequence there is controversy regarding calculating capital allocation. Importantly, everyone agrees that the amount of capital should depend upon the risks inherent in the division; activities possessing high risks must be backed with more capital as compared with other activities. However, current methods do not employ the DFA framework, grounded in a dynamic optimization context. See Froot and Stein (1998), Mulvey, Correnti and Lummis (1997), and Mulvey, Simsek and Pauling (2003) for discussions of optimizing a financial intermediary.

As an alternative, we can manage an insurance company by means of a decentralized DFA approach. The ideas build on the literature on large-scale optimization. Under this theory, a single optimization problem is broken into a set of smaller, more manageable subproblems, to be solved in a sequential fashion with coordination from a headquarters. Figure 19 shows the process. Note the similarity with the goals of capital allocation. Information is sent to the divisions, typically prices of joint resources under each of the scenarios—the state prices. In turn, the divisions optimize their affairs given these prices.

4.3.1. Corporate objectives in a decentralized setting

There are several goals for capital allocation to business units:

(A) Protect the company and the public against adverse losses,
(B) Improve the company's operations by building on current expertise and activities,
(C) Provide diversification benefits, and
(D) Grow the enterprise by mergers and acquisitions, as appropriate.

These goals must be part of any decentralized DFA system. An integrated planning system requires a systematic objective function in order to search for a set of recommendations for improving the organization. This issue can become complicated when local requirements and constraints exist for individual entities. A global insurance company operates within a large number of regulatory boundaries, including countries, states, cities, and other geographical regions, as well as offering a wide range of products (each with its own rules, traditions, etc.).

Generally, the objective of a business unit is to maximize shareholder value—as measured by the long-term growth of the company's stock (and/or surplus). There may be few commonly agreed definitions of this growth on an anticipatory basis for a business unit, since units may not be trades in a market. Still, the risk-adjusted profits as developed above can serve to define the unit's objectives.

4.3.2. Constraining the feasible region

Regulatory issues can be addressed by means of additional constraints on the integrated system. For instance, there are often accounting and statutory rules to be followed in order for the entity to continue to operate within its existing territories. These rules are often complicated due to the continual evolution of regulations in the financial industry. In addition, the divisions may have constraints imposed by the headquarters, such as the type and location of business it can engage in.

4.3.3. Decentralized DFA algorithm

The DFA system provides an ideal framework for analyzing the benefits of undertaking new activities and for changing the level of an existing activity. For instance, we can extend the static valuation calculation by employing scenarios and a scenario dependent discount factor. The result is a

Dynamic valuation formula.

$$v_j = \sum_s \sum_t c^s_{j,t} / \left(1 + r^s_{j,t}\right)^t \quad \text{for } j \in J,\ t \in T,\ s \in S. \tag{14}$$

The risk adjusted discount factors $r^s_{j,t}$ are derived from the DFA system as dual variables for the joint capacity constraints (Eqs. (2) and (3)). These dual variables approximate the Arrow–Debreu state prices. A new activity can be evaluated *on the margin* as to its benefit or penalty to the organization. Diversification aspects are addressed in a direct fashion.

For decisions with large impacts, the headquarters model must be rerun to evaluate the contribution of the decision. A marginal analysis is inadequate due to the changing nature of the state prices for substantial changes in the organization's position. This result comes directly from linear programming theory.

In a similar fashion, we can employ the state prices for calculating risk-adjusted profits:

$$ROP_{j,t} = \sum_{s} \frac{\rho^{s}_{j,t}}{\alpha^{s}_{j,t}} \qquad \text{for the relative formula,} \tag{15}$$

and

$$VA_{j,t} = \sum_{s} \left(\rho^{s}_{j,t} - \alpha^{s}_{j,t}\right) \quad \text{for the valued-added formula.} \tag{16}$$

Again, these calculations more accurately reflect the impact of the divisional decision on the total firm than the simple formula presented earlier.

The full-decentralized DFA system employs a series of iterations—between headquarters and the business units. At each step, the headquarters takes in the solutions of the divisional DFA results in a manner similar to the bottom-up approach. Headquarters will then solve its own DFA system in order to determine the total capital for the organization. In this context, the optimal reinsurance treaties will be calculated. A new set of state prices (dual variables as state prices) for each scenario $s \in \mathbb{S}$ will emerge from this step and be transmitted back to the divisions. The decentralized DFA system combines elements of both the top–down and the bottom-up strategies. The business units optimize their affairs, given state prices for the scenarios. Headquarters optimizes its decisions by combining the decisions proposed by the business units; it determines the optimal capital structure, the shape of reinsurance treaties and other firm-wide issues. Periodically, a new set of state prices will be issued to the units for their decision-making. The approach is an extension and adaptation of the Dantzig–Wolfe decomposition method, with modifications for the stochastic nature of the DFA problem and the fact that headquarters will render decisions as discussed. In addition, the Dantzig–Wolfe method requires that headquarters finds the best convex combination of all decision proposals; this requirement would be difficult to implement in practice.

Also, from a practical standpoint, the number of iterations should be relatively few to reduce the costs and time to carry out the procedure. The state prices should not change in a radical way between iterations, if possible. The nature of the optimization should take into account these issues by adding constraints and/or penalties to the dual variables in the DFA system. A presentation of decentralized DFA appears in Erkan (2006), and Mulvey and Erkan (2006).

In summary, decentralized DFA provides a practical framework for managing the risks of a large organization while improving the firm's profitability. A company that understands its risks can decrease overall risks and increase profits—*at the same time*. The decentralized organizational structure can be preserved so that individual can operate in a semi-autonomous fashion and more importantly can respond to the strict regulatory constraints and operating procedures.

5. Conclusions and future work

This paper describes the modeling of insurance companies within the enterprise framework. We present several prominent decisions, including asset allocation, evaluating new business activities, and other areas in which DFA improve long-run performance. A prominent example involves the optimal design of reinsurance treaties; a firm's performance is directly affected by these agreements. In addition, we describe capital allocation issues from the standpoint of managing a large global company. The overriding goal is to optimize the enterprise value in a decentralized fashion.

We focus on significant issues that arise when constructing, simulating and ultimately optimizing a DFA system. We discuss convergent risks and the related matter of capital allocation based on risk measures affected by the tails of the profit/loss distributions. Not only must we estimate the expected loss payouts and other cash flows, but also we must determine the co-movements of these values, especially during periods of moderate to extreme volatility. The firm's capital allocation decision is critical: capital protects the organization from credit downgrades and even bankruptcy. A nonlinear factor model provides the best approach, in our opinion, for modeling uncertainties. The DFA system helps develop the company's responses to stressful scenarios in an anticipatory fashion. Optimization can play a significant role by pinpointing the best responses to the difficult scenarios, as well as balancing these responses when favorable scenarios occur.

What are barriers to wider implementation, similar to the success of enterprise management systems common in logistics and the process industry? First, the regulatory bodies, rating agencies, and company executives will need to become familiar with the advantages and disadvantages of the approach. Second, DFA promises to improve the profitability of a company while decreasing enterprise risks. As leading companies fulfill this promise, others will follow. It took considerable time before supply management became the norm in the process industry, even though examples were evident for many years beforehand. Also, regulation in the financial industry will have a direct impact on the use of DFA models. As an example, banking regulations have largely driven the development of risk management systems for banks vis-à-vis the Basel accords.

There are several directions for future research. First, study is needed of the processes causing the tails of the profit/loss distribution, especially factors that give rise to contagion. As an example, some hedge funds employ convergent trades based on differential amounts of liquidity. For these trades, there should be a factor in the hedge-fund risk model relating to liquidity and credit risks. Insurance companies work hard to minimize adverse selection and correlated losses. Still, improved risk transfer mechanisms will provide better liquidity and enhance information.

As novel insurance/financial products and services become available, the regulatory framework should be aimed at improving the industry's risk-reward profile. It is in the best interest of the public to reduce the overall risks to insurance organizations. In the US, state regulators will be ultimately held responsible for losses that fall outside the company's capital position. There are many mechanisms for reducing the company's risks, including reinsurance, securitization, selective underwriting, and greater levels of capital. Each alternative can be tested via the DFA system to determine their respective costs and benefits. There may be companies who will try to understate their low probability, high-severity risks, and this coupled with the asymmetric information leads to interesting game theoretical issues. Just as in bidding, however, governmental regulators should strive to find solutions that are in the long-term interest of the parties. Aligning these interests will be an important topic for future research.

Last, financial companies will continue to merge diverse operations, such as insurance, banking, and security firms. Managing these financial conglomerates will require not only new managerial skills but also risk management systems that link together capital across diverse areas. Decentralized approaches will be needed for coordinating large global organizations.

Appendix A. Toolkit for constructing a DFA system

This appendix describes the use of software tools for constructing a DFA system. As evident, a DFA system can become complex since it must include a large number of interacting variables, data sources, and processes. Simplifying the construction of a DFA system provides several benefits. First, a well-conceived toolkit can reduce the cost of development. Building a C++ system can take several years and cost over $1 million just for the software developers! The computer developer costs can be much reduced by employing a toolkit.

Second and importantly, the developer's toolkit provides a blueprint for the construction of the DFA system. The assumptions are readily shown by reference to the construction diagrams, as evident in the figures listed in this appendix. Each of the modeling assumptions and the choices are apparent. Accessibility has several purposes. First, the DFA model can be easily changed as new information is collected or as mergers and acquisitions modify the organization. The DFA system has the ability to anticipate possible changes as well, for example, the impact of adding a new line of insurance can be evaluated. Thus, sensitivity analysis can be directly performed without the need for costly rewriting the DFA system.

Figure 20 provides the main elements of a sophisticated DFA toolkit. The heading—economic scenarios, investment operations, insurance operations, non-insurance operations, and corporate and financing operations—identify each aspect of the model. Likewise, a number of example models are depicted, such as alternative asset allocation strategies. These can be compared with each other across multiple dimensions. The goal of the software toolkit is to manage a set of DFA models for a large, complex organization. Simplicity and ease of access are the primary subgoals.

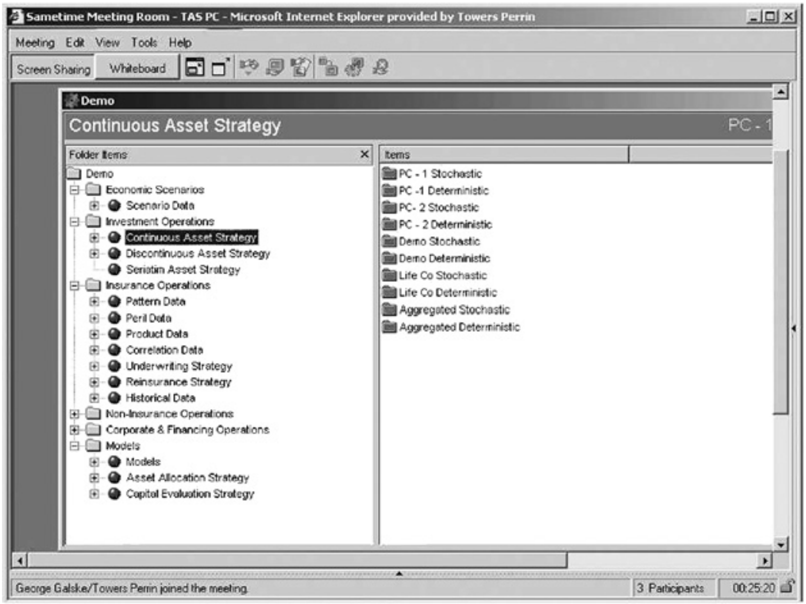


Fig. 20. Main toolbar for DFA developers.

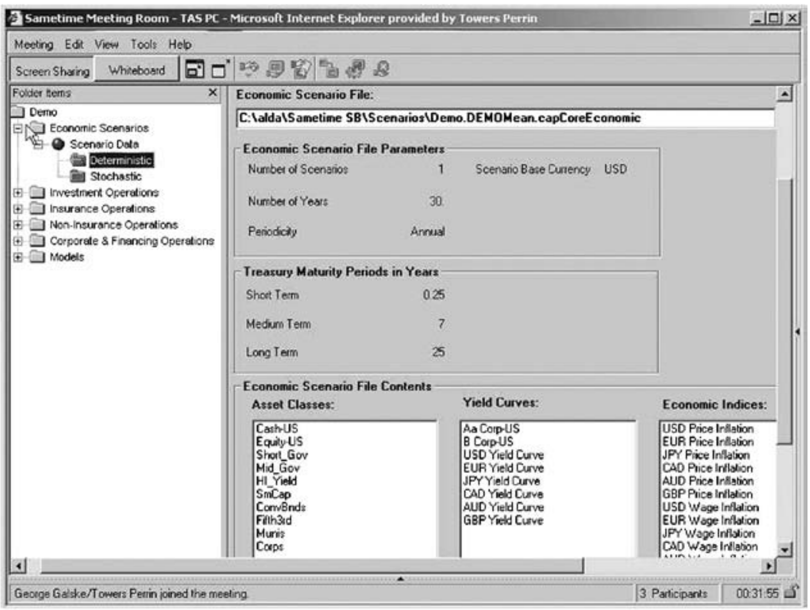
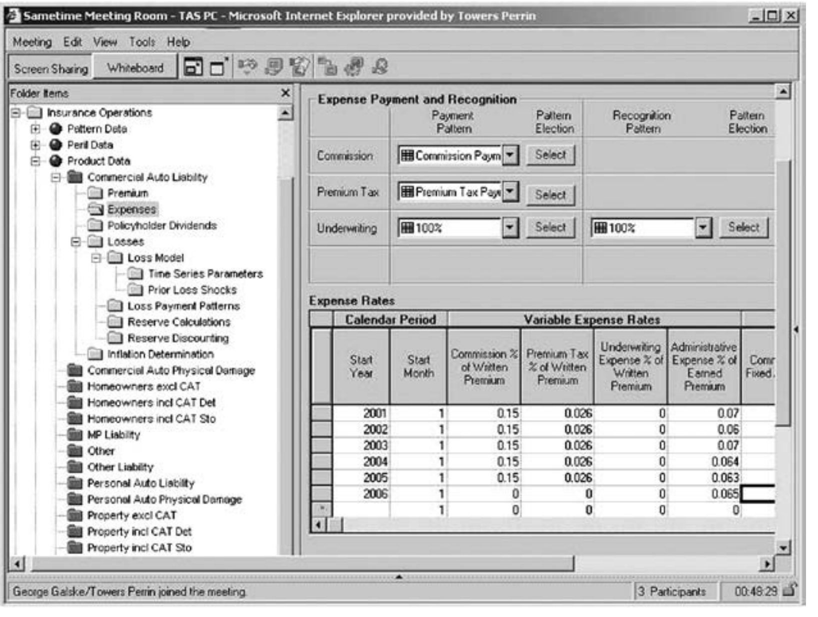


Fig. 21. Illustration of DFA construction for asset allocation.



Start Year	Start Month	Commission % of Written Premium	Premium Tax % of Written Premium	Underwriting Expense % of Written Premium	Administrative Expense % of Earned Premium
2001	1	0.15	0.026	0	0.07
2002	1	0.15	0.026	0	0.06
2003	1	0.15	0.026	0	0.07
2004	1	0.15	0.026	0	0.064
2005	1	0.15	0.026	0	0.063
2006	1	0	0	0	0.065
	1	0	0	0	0

Fig. 22. Illustration of constructing insurance lines.

For each of the main sections, we can drill down to the special requirements for that topic. As an example, we define the economic scenarios in two ways—as deterministic time series values or as a group of stochastic scenarios. Figure 21 gives an idea of the former. Herein, a single scenario is chosen for an analysis. Economic scenarios fall into three classes—underlying economic factors, asset classes, and yield curves. The basic data is stored in files for ready access.

The same notions apply to the generation of the business lines as shown in Figure 22. There are many interacting issues in constructing the projected losses for each insurance line. Again, the goal is to build up from previously stored generic patterns. The figure involves the development of commercial automobile insurance, in this example, the expense patterns for the next six years. Commissions and other variable expenses are defined as a function of the time in the future. We fix these values as shown. Other issues, such as reserving and the loss patterns are defined in a similar fashion. The DFA modeling exercise begins at the highest level and works it way down to the specific elements for the business lines.

A second insurance product is depicted in Figure 23. Here, we define the cumulative loss payments for homeowner insurance—70.32% of the losses on average occur during the first year, 93.11% by the end of the second year, 97.63% by the end of the third year, and 100% by the end of the fourth year. This pattern is one of many possibilities, depending upon the types of products available to the insurance company. Each business activity is defined in turn.

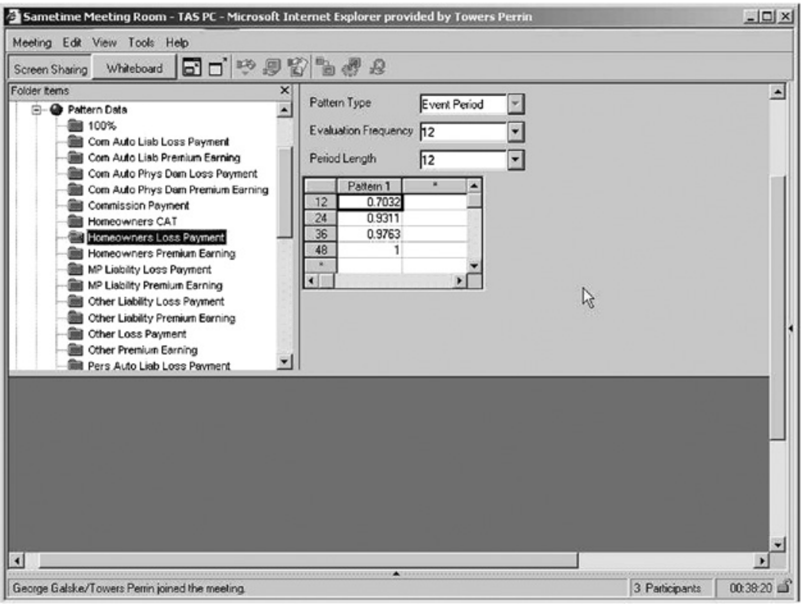


Fig. 23. Illustration of second line of insurance.

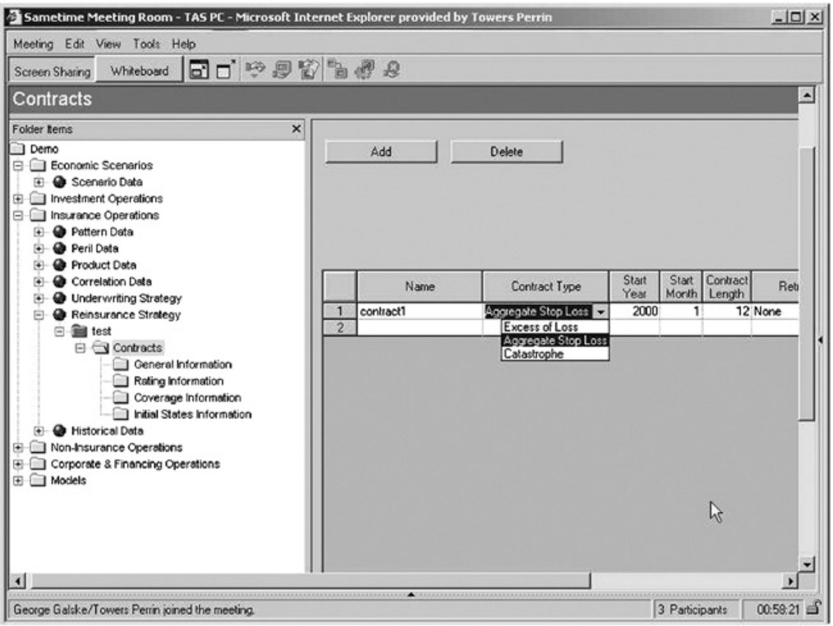


Fig. 24. Illustration of reinsurance treaty.

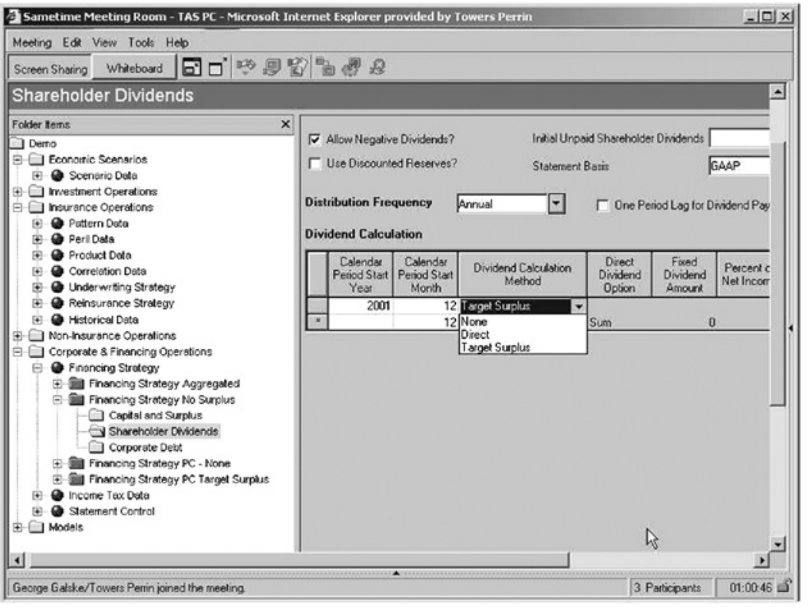


Fig. 25. Illustration of dividend policy.

A significant application of DFA involves the definition of reinsurance treaties. Each contract must be carefully constructed by reference to the primary losses and the resulting impact on the overall losses. Figure 24 depicts a reinsurance example with aggregate stop losses. The main forms of Reinsurance are saved for future analyses.

Last, we build the dividend policy for the company in Figure 25. In this case, the dividends can be positive or negative depending upon the target surplus. If a deficit occurs under this regime, there will be a capital infusion. Otherwise, a dividend will be made to the stockholders or the policyholders. The dividends flow from these assumptions.

While there are many other aspects, it should be evident that the toolkit approach greatly simplifies the construction of the DFA model. The assumptions are clearly delineated and easily found by the insurance users and DFA developers. Implementation of complex mathematical models is greatly improved and made more efficient by these software tools.

References

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9.

Berger, A., Madsen, C., 1999. A comprehensive system for selecting and evaluating DFA model parameters, in: Casualty Actuarial Society Forum, Summer.

Berger, A., Mulvey, J.M., Nish, K., 1998. A portfolio management system for catastrophe property liabilities. in: Casualty Actuarial Society Forum, pp. 1–14.

Birge, J.R., Louveaux, F., 1997. Introduction to Stochastic Programming. Springer-Verlag, New York.
Boender, G.C.E., 1997. A hybrid simulation/optimization scenario model for asset/liability management. European Journal of Operations Research 99.
Bowers, B., 2002. Ready for anything. Best's Review, February.
Brinson, G., Hood, R., Beebower, G., 1986. Determinants of portfolio performance. Financial Analysts Journal, July–August.
Burket, J., McIntyre, T., Sonlin, S., 2001. Dynamic financial analysis: DFA insurance company case study. in: Casualty Actuarial Society Forum, Spring.
Cariño, D.R., Ziemba, W.T., 1998. Formulation of the Russell–Yasuda Kasai financial planning model. Operations Research, July–August.
Cariño, T.K., Myers, D., Stacy, C., Sylvanus, M., Turner, A., Watanabe, K., Ziemba, W., 1994. The Russell–Yasuda Kasai model: An asset liability model for a Japanese insurance company using multi-stage stochastic programming. Interfaces 24, 29–49.
Chen, Z., Bakshi, G., 2001. Stock valuation in dynamic economics. Yale University Report, ICF working paper, May 29.
Consigli, G., Dempster, M., 1998. The CALM stochastic programming model for dynamic asset–liability management. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press.
Consiglio, A., Zenios, S., 2001. Integrated simulation and optimization models for tracking international fixed income indices. Mathematical Programming 89 (2), 311–339.
Dantzig, G., Infanger, G., 1993. Multi-stage stochastic linear programs for portfolio optimization. Annals of Operations Research 45, 59–76.
Dert, C.L., 1995. Asset liability management for pension plans, PhD thesis. Erasmus University, Rotterdam.
Doherty, N., 2000. Integrated Risk Management: Techniques and Strategies for Managing Corporate Risk. McGraw-Hill, March.
Duffie, D., Singleton, K., 1993. Simulated moment estimation of Markov models of asset prices. Econometrica 61, 929–952.
Enz, R., Karl, K., 2001. Profitability of the non-life insurance industry: its back-to-basics time. Swiss Re Sigma 5.
Erkan, H.G., 2006. Decentralized enterprise risk management for global companies. PhD thesis. Princeton University, NJ.
Ferguson, R., Leistikow, D., 1998. Search for the best financial performance measures: basics are better. Financial Analysts Journal, January–February.
Fernholz, R., Shay, B., 1982. Stochastic portfolio theory and stock market equilibrium. Journal of Finance 37, 615–624.
Frauendorfer, K., Schurle, M., 2000. Stochastic optimization in asset and liability management: a model for non-maturing accounts. In: Uryasev, S. (Ed.), Probabilistic Constrained Optimization: Methodology and Applications. Kluwer.
Froot, K., Stein, J., 1998. Risk management, capital budgeting and capital structure policy for financial institutions: an integrated approach. Journal of Financial Economics 47, 55–82, January.
Geoffrion, A., Krishnan, R., 2001. Prospects for operations research in the e-business era. Interfaces 31, 6–36, March.
Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8.
Hoyland, K., 1998. Asset liability management for a life insurance company: a stochastic programming approach. PhD Dissertation. Norwegian University of Science and Technology.
Hoyland, K., Wallace, S., 2001. Generating scenario trees for multistage problems. Management Science, 295–307, February.
Kaufman, A., Ryan, T., 2000. Strategic asset allocation for multi-line insurers using dynamic financial analysis. in: Casualty Actuarial Society Forum, Summer.
Kusy, M.I., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34, 356–376.

Lasdon, L., 1970. Optimization Theory for Large Scale Systems. Macmillan, London.

Laster, D., Thorlacius, E., 2000. Asset-liability management for insurers. Swiss Re Sigma 6.

Lin, G., Ettl, M., Buckley, S., Bagchi, S., Yao, D., Naccarato, B., Allan, R., Kim, K., Loenig, L., 2000. Extended enterprise supply chain management at IBM Personal Systems Group and other divisions. Interfaces 30, January–February.

Lowe, S.P., Stanard, J., 1996. An integrated dynamic financial analysis and decision support system for a property catastrophe reinsurer. In: CAS Seminar on Dynamic Financial Analysis, Spring CAS Forum.

Mango, D., Mulvey, J.M., 2000. Capital adequacy and allocation using dynamic financial analysis. In: Casualty Actuarial Society Forum, Summer.

Maranas, C., Androulakis, I., Berger, A., Floudas, C.A., Mulvey, J.M., 1997. Solving stochastic control problems in finance via global optimization. J. Econ. Dyn. Control 21, 1405–1425.

Matten, C., 1996. Managing Bank Capital: Capital Allocation and Performance Measurement. John Wiley.

Mulvey, J.M., 1989. A surplus optimization perspective. Investment Management Review 3.

Mulvey, J.M., 1996. Generating scenarios for the Towers Perrin investment system. Interfaces 26.

Mulvey, J.M., Correnti, S., Lummis, J., 1997. Total integrated risk management: insurance elements. Princeton University Report SOR-97-2.

Mulvey, J.M., Erkan, H.G., 2006. Applying CVaR for decentralized risk management of financial companies. Journal of Banking and Finance 30 (2), 627–644.

Mulvey, J.M., Gould, G., Morgan, C., 2000. An asset and liability management system for Towers Perrin–Tillinghast. Interfaces 30, 96–114.

Mulvey, J.M., Rosenbaum, D., Shetty, B., 1999. Parameter estimation in stochastic scenario generation systems. European Journal of Operations Research 118, 563–577.

Mulvey, J.M., Simsek, K., Pauling, W., 2003. A stochastic network approach for integrating pension and corporate financial planning. Innovations in Financial and Economic Networks.

Mulvey, J.M., Thorlacius, A., 1998. The Towers Perrin global capital market scenario generation system: CAP:Link. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press.

Mulvey, J.M., Ural, C., Zhang, Z., 2007. Improving performance for long-term investors: wide diversification, leverage, and overlay strategies. Quantitative Finance 7 (2), 1–13.

Perold, A., Sharpe, W., 1988. Dynamic strategies for asset allocation. Financial Analysts Journal, 16–27, January.

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional Value-At-Risk. The Journal of Risk 2 (3), 21–41.

Rockafellar, R.T., Wets, R.J.-B., 1991. Scenarios and policy aggregation in optimization under uncertainty. Mathematics of Operations Research 16, 119–147.

Rush, R., Mulvey, J.M., Mitchell, J., Willemain, T., 2000. Stratified filtered sampling in stochastic optimization. J. Appl. Math. Decision Sci. 4, 17–38.

Shimpi, P.A., Helbling, C., Durbin, D., Laster, D. (Eds.), 2000. Integrating Corporate Risk Management. Texere, December.

Stein, J., 2002. Information production and capital allocation: decentralized vs. hierarchical firms. J. of Finance, October.

Uyemura, D., Kantor, C., Pettit, J., 1996. EVA for banks: value creation, risk management and profitability measurement. J. of Applied Corporate Finance 9, Summer.

Wilkie, A., 1987. Stochastic investment models: theory and applications. Insurance: Mathematics and Economics 6, 65–83.

Worzel, K., Vassiadou-Zeniou, C., Zenios, S.A., 1995. Integrated simulation and optimization models for tracking fixed-income indices. Operations Research 42, 223–233.

Zaik, E., Walter, J., Kelling, G., 1996. RAROC at Bank of America: from theory to practice. J. of Applied Corporate Finance 9, Summer.

Zenios, S.A., 1991. Massively parallel computations for financial planning under uncertainty. In: Mesirov, J.P. (Ed.), Very Large Scale Computation in the 21st Century. Society for Industrial and Applied Mathematics, Philadelphia, pp. 273–294.

Zenios, S.A., 1993. Financial Optimization. Cambridge University Press, Cambridge.

Zenios, S.A., 1998. Asset and liability management under uncertainty for fixed income securities. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press.

Ziemba, W., Mulvey, J.M. (Eds.), 1998. Worldwide Asset and Liability Modeling. Cambridge University Press.

Chapter 13

STOCHASTIC PROGRAMMING MODELS FOR STRATEGIC AND TACTICAL ASSET ALLOCATION—A STUDY FROM NORWEGIAN LIFE INSURANCE

KJETIL HØYLAND

Gjensidige-NOR Asset Management, P.O. Box 276, NO-1326 Lysaker, Norway
e-mail: kjetil.hoyland@spk.no

STEIN W. WALLACE

Molde University College, P.O. Box 2110, NO-6402 Molde, Norway
e-mail: stein.w.wallace@himolde.no

Contents

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02013-8

Abstract

In this chapter we describe the development and use of two decision support models for asset allocation used within the Gjensidige-NOR Group,[1] one of Norway's three largest financial groups. For strategic, long-term, asset liability management, the life insurance company within the Group uses an ALM-model. GN Asset Management, the asset management company within the group, uses a different model for shorter term tactical asset allocation in a hedge fund. Both models are based on stochastic programming. The models have been developed in close cooperation between GN Asset Management and the Norwegian University of Science and Technology.

We describe the institutional setting, with an emphasis on structure, relevant parts of the legal framework and the competitive situation. Based on this, and an outline of assets and liabilities, we outline the models and motivate the chosen modeling frameworks. We discuss our experiences regarding interactions with users during the developing process, the time spent in different phases of the project, possible pitfalls in the developing phase, and how the models have changed the organization.

A model is no better than its input. Hence, we provide a detailed description of the scenario generation procedure, from data collection and the formation of market expectations, to the final scenario tree needed by the stochastic programming models.

We provide a case study which examines the whole investment process, from establishing market expectations and generating scenarios, to constructing, implementing, and managing the portfolio. We conclude the paper by presenting the track record of the hedge fund.

[1] On 4 December, 2003, Gjensidige-NOR and DnB merged to create DnB NOR which became Norway's largest financial group.

1. Introduction

Gjensidige-NOR is one of Norway's three biggest financial groups. It consists of Union Bank of Norway, Gjensidige NOR Life Insurance and Gjensidige NOR Non-Life. The asset management company is a separate legal entity owned by the bank. This company manages a large part of the assets of the Life and Non-Life companies, in addition to funds outside of Gjensidige-NOR.

When the authors made contact with the asset management group, the structure of the asset management was different from today. The asset management company was an integrated part of the life company. When the contact was established in 1995, the focus for the life insurance company was to improve and structure the asset liability management process. A challenge was recognized in addressing and quantifying different kinds of risks in a consistent manner. More specifically, it was recognized that the life company had difficulties in transforming its views on the different risks and its views on the financial markets into a portfolio consistent with these views. It was identified that quantifying the different types of risks and addressing the question of constructing an asset allocation mix consistent with these risks and the asset managers' market expectations, was a key issue for the success of the asset liability management.

The asset allocation decision was a result of a mixture of the asset managers' market expectations, the balance sheet risk, and the competitive risk (the risk of achieving a lower return on the assets than the competitors and hence potentially losing business). Potential conflicting goals between absolute return, balance risk and return relative to the competitors made the asset allocation decision complex. Within Norwegian life insurance there has been several examples of life insurance companies that have been forced to sell equities after a major down turn in the equity markets. This did not occur because the asset managers believed that equities would continue to perform badly after major falls, but because risk capacity was no longer there.

Conflicting goals made applying appropriate performance measurements difficult, and it also complicated the incentive structure. In 1999 the asset management group was established as a separate legal entity, partly because of the above. In addition, the asset allocation decision-making process in the life company was divided in two: Strategic asset allocation (SAA) and tactical asset allocation (TAA). At the strategic level, the main focus is on balance risk (ALM) and competitor risk. At the tactical level the focus is on generating excess return relative to a benchmark that is set by the SAA-team. The SAA is run by the Finance group within the life insurance company, while the TAA is run by GN Asset Management and other external asset managers.

The original stochastic programming model was developed for the strategic level. A modified version of this model is still in use as decision support for planning of the ALM-structure in the life insurance company. Later, we developed a tactical asset allocation model that has been in use since 1999. We will present both models and compare them with the familiar Markowitz' mean-variance model.

Both models require similar model input, though the strategic model is multi-period whereas the tactical model is single period. One key element for the success of these models, in terms of whether they are utilized or not, is that the decision-makers are comfortable with and completely understand how the input to the model is generated. In the developing phase, much emphasis was therefore put on the scenario generation procedure. The process of collecting data, expressing (judgmental) market expectations, and transforming these expectations into a format applicable to the stochastic programming model, will be illustrated in detail.

The TAA-model is in use on a continuous basis and has had the biggest impact on the organization. We will explain how the TAA-group has adapted as the model to an increasing degree has become the core of its operations.

The experiences the TAA-group has gained from using the model will be described. The modeling framework has been useful for introducing new employees to the investment philosophy of the TAA-team. The framework is flexible enough to allow for differences in analytic styles among team members, and it has proven to be well suited for generating consensus decisions. Further, we will explain how the quantitative framework creates a basis for an efficient learning process. The modeling framework also reinforces and disciplines the team to engage in a structured and transparent investment process. An internal model is also by itself valuable for improving the employees' understanding of the link between market data used as input and the investment policy recommendations resulting from the model.

There are several challenges in using the modeling framework, for example, in calibrating the market expectations of different members in the group. In particular, members of a team most likely have differences in their personal risk preferences, which can skew decisions regarding optimal portfolio allocation. Further, we will explain the challenge in using the modeling framework in dynamic and fast changing financial markets.

Section 2 motivates the modeling framework and describes the main characteristics of the two models. In particular we discuss why a single period model was chosen for the tactical asset allocation problem, and a multi-period model for the strategic problem. The modeling framework is compared with the familiar Markowitz mean-variance model. Section 3 discusses data collection and scenario generation. Section 4 illustrates how the TAA-model has influenced the organization of the TAA-group. Section 5 elaborates experiences from the developing phase and from using the TAA-model on a continuous basis since mid-1999. We conclude the paper in Section 6.

2. Motivation and model description

An asset allocation model can be distinguished by the following characteristics:

- Risk measure,
- Description of uncertainty,
- Single versus multi-period,
- Constraints,
- Investment vehicles.

The TAA and the SAA models are both stochastic programs (see Kall and Wallace (1994) and Birge and Louveaux (1997) for details) and use the same framework regarding the description of uncertainty and definition of risk. In Section 2.1 we will explain the framework for modeling risk, while Section 2.2 explains the specific risks that the decision-makers face on the two decision levels. To understand these risks we need to understand the legal framework and the competitive situation under which the decision-makers operate. The two models differ regarding the investment vehicles, which are described in Section 2.3, the constraints, which are described in Section 2.4, and the dynamics, which is addressed in Section 2.5. In Section 2.6 we compare the framework with the mean-variance approach, first suggested by Markowitz (1952), see also Markowitz and van Dijk (2006). Finally, Section 2.7 explains the link between the SAA-model and the TAA-model.

2.1. Framework for modeling risk

The objectives of both models reflect a tradeoff between expected return and some measure of risk. As a portfolio manager who's performance is measured relative to a benchmark, the risk is to under-perform the benchmark. As financial director responsible for the financial health of a life insurance company, the risks typically involve falling below thresholds for capital adequacy, solvency, or other legal requirements or to achieve lower returns on assets relative to competitors.

In order to use a quantitative approach to the asset allocation problem the various risks must be quantified. To do this, we introduce targets from which shortfalls will be associated with a cost (in excess of the actual cost in terms of reduced value of the assets). These targets can be an absolute return target (for instance, 0% return), relative return targets (the return on some benchmark) or financial targets like capital adequacy and solvency requirements. A shortfall cost function measures the (subjective) cost of shortfalls of different magnitudes. We assume a convex shortfall cost function, i.e., increasing marginal penalty. A simple quadratic version of such a shortfall cost function is illustrated in Figure 1.

The objective of the optimization is to maximize the expected return net of expected shortfall costs. The objective function will be linear for returns above the target return

Fig. 1. A shortfall cost function with quadratic penalty.

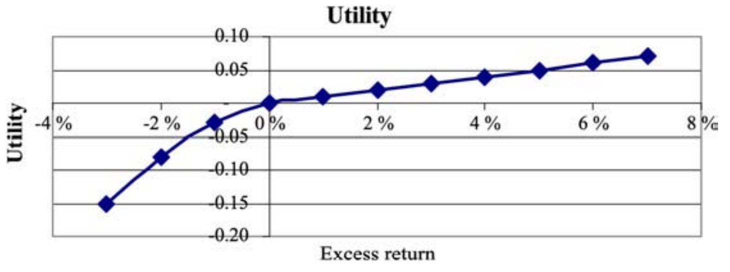


Fig. 2. The utility function consistent with the shortfall cost function in Figure 1.

and strictly concave for returns below the target return. The utility function consistent with the shortfall cost function in Figure 1 is shown in Figure 2. For a comprehensive discussion of risk and utility functions, see Bell (1995).

To identify risk preferences there are at least two major approaches. One is to define risk based on theoretical considerations, studying the theory of market risk and individual risk attitudes. The other is to focus on how the user sees his own risks, and model accordingly. The approach using shortfall cost functions as described, is an example of the latter way of thinking. The choice of which risks to include and what parameters to use, is done in close cooperation with the user. It is useful to observe that the result can be interpreted as a utility function, but our goal has not been to produce a utility function.

A utility function, like the one in Figure 2, has several parameters describing its shape. In our case, these parameters were defined in close interaction with the user. Initially, we ran the model with one set, discussed the model results with the user, and the user concluded such as "We are not that risk averse", or "This is too risky relative to so-and-so legal constraint". Eventually, a utility function reflecting the user's view on his own risks is derived. We believe that this is the most reasonable way of setting up an objective function reflecting risk. For a discussion, see Kallberg and Ziemba (1983).

Of course, there are many other risks involved, the major one being whether or not a model makes sense in the first place. When discussing the track record of the model in Section 5, we shall return to this issue. A variety of tests can be done, but they will all depend on certain assumptions, and there will always be an aspect of belief or hope in the decision to use a model.

2.2. Risk measures

At the strategic level, the risks can be split into balance risk and competitor risk. The asset liability management is influenced by the following legal framework issues:

- Capital adequacy requirement,
- Solvency requirement,
- The minimum guaranteed return,
- Regulations regarding how much of the profit that must be allocated to the customers,

- The customers' right to move from one life insurance company to another, free of charge.

The capital adequacy requirement states that the buffer capital (mainly equity and subordinated debt) must exceed 8% of the value of risk-adjusted assets. The assets are assigned risk weights, with for instance equity counting 100% and Norwegian money market investments counting 0%. If, for example, all assets were invested in equities, then the buffer capital of the company had to be 8% of the total balance. The solvency requirement states that the buffer capital of the company must be larger than a percentage of the pension liabilities. Both requirements need to be modeled because both can be binding dependent on the asset mix.

The minimum guaranteed return is currently about 3.8% per annum as an average for existing customers. The guarantee applies for every single year. If the return on assets is lower than the guaranteed return, buffer capital must be used in order to fulfill the guarantee. If the company achieves a profit after costs, and after the guaranteed rate of return is paid, the regulations state that a minimum of 70% must be returned to the customers. These two regulations have severe impact on the long-term planning of the insurance company. If return on assets is well below the guarantee in a given year, the buffer capital must be used to cover the guarantee, and rebuilding the buffer capital is a slow process since a maximum of 30% of a year's profit (after costs and guaranteed rate of return) can be used for this purpose. As a result one bad year can have severe long-term effects on the risk capacity and hence on future returns.

The customers' right to move their policies, free of charge, creates a very competitive environment and makes it crucial not to under-perform the competitors in any year. The goal of not under-performing the competitors will in some situations conflict with the goal of securing the long-term financial health of the company.

The strategic ALM model incorporates three shortfall variables: Capital adequacy, solvency requirement and return relative to the competitors. Setting targets from which to measure the shortfalls and estimating the cost function are challenging tasks. Targets are typically set higher than the legal minimums, and the marginal costs are increasing. At the point where the company is close to bankruptcy, the marginal costs clearly increase dramatically.

The shortfall costs are difficult to estimate accurately, but they are "real" costs in the sense that they should reflect the business costs of under-performing the targets. It is straightforward to estimate their limit values; they will vary between zero and the value of the company. The shortfall cost functions can be estimated. The cost of under-performing the competitors, for example, can be estimated by finding the average cost of losing one customer, and estimating the relationship between under-performing the competitors and the expected number of customers lost. In the multi-period setting, we have added shortfall costs over periods. From a theoretical point of view, additivity in utility is a very strong assumption. However, the user is satisfied with the implications of such an assumption, particularly taking into account that the decision-maker is a company and not an individual.

The legal regulations and their effect on the decision-making in a life insurance company is elaborated, and a mathematical description of the model given, in Høyland and Wallace (2001b). Similar studies have been undertaken for a casualty insurance company in Gaivoronski and De Lange (2000). de Lange, Fleten and Gaivoronski (2004) and Gaivoronski, Høyland and De Lange (2000) also address ALM-planning in the casualty insurance business.

For the TAA-problem, the legal framework is, to a large extent, defined by the mandate given by the client. For the funds that the TAA-group manages for the life insurance company, the mandate is usually given by a benchmark, risk limits relative to the benchmark, and constraints defining upper and lower limits for each asset class. The task of the TAA-team is to achieve excess return relative to the benchmark, while operating within its risk limits and constraints. As a manager evaluated on a relative basis, the risk is to under-perform the benchmark. The shortfall costs in this case should reflect the expected future loss by potentially losing the client in case of a shortfall, and also the subjective personal utility functions of the portfolio managers. Note that there are less conflicting goals with different time horizons in this problem, and hence the need for a multi-period setting is less than for the strategic problem, see more on this in Section 2.5.

2.3. *Investment vehicles*

For the strategic ALM problem, the model is typically applied for the planning of the following financial year, and two problems are solved simultaneously: The allocation between major asset classes and the structure of the liability side of the balance sheet. The investment universe is split into a limited number of broad asset classes: typically, domestic and international stocks, domestic and international bonds, domestic money market, and real estate. The task of the SAA-team is to establish a benchmark allocation between these asset classes. Most of the currency risk is hedged. By law, the maximum currency exposure is 15% of the liabilities.

For the TAA-problem, active market positions are implemented via derivatives. The TAA-model applies for both linear, e.g., futures and non-linear, e.g., options, derivatives. The asset classes defined within the investment universe are money market, bonds, equities, currencies, and commodities. USA, Japan, UK, EMU-area, Norway, and Sweden are potential regions for investments.

For the money market, the investment vehicles are futures on the LIBOR interest rates at different expiry dates. For bonds, they are two-, five-, and ten-year bonds. For equities, the investment vehicles are the most liquid futures in each region. For currencies, the number of potential asset classes equals the number of regions less one, since we define all currencies relative to the US dollar. All together, the investment universe consists of more than 40 assets. The number of assets included in each run will, however, be much smaller, typically 15–25. One part of the investment process is to select the asset classes on which we have the strongest views.

Note that the currency risk is small when dealing with derivatives. Only the margin accounts in foreign currency (which for equities are less than 10% of the exposure, and for bonds even less) are exposed to currency risk. In the TAA-problem currencies are treated as any other asset class. Currency exposure is only accepted if it is the result of an active view on the return from investing in a currency cross. With no active view on currencies, significant currency exposures will be hedged. As the performance of the fund is measured in Norwegian kroner (NOK), NOK is the base currency.

There are several reasons for using derivatives instead of cash instruments:

- Liquidity,
- Transaction costs,
- Mobility of the strategy,
- Tailor-made positioning.

Most of the standardized derivatives used are highly liquid, often more liquid than the underlying cash instrument. The transaction costs are lower due to lower execution fees and less market impact. Further, a derivatives strategy can be put on top of any underlying cash portfolio. This means that a strategy can be implemented efficiently across different clients with different underlying portfolios. Finally, the use of options allows for a much larger degree of tailor-made portfolios relative to the portfolio managers' view on the market.

The modeling implications of using derivatives are minor. The market expectations are given on the underlying assets and it is usually straightforward to calculate the future value of the derivatives given the future value of the underlying.

2.4. Constraints

Both models have both soft and hard constraints. The soft constraints are defined by the shortfall variables, and the shape of the shortfall cost function defines the softness.

The SAA-model has the following constraints:

- **Legal hard constraints**: Examples are maximum 35% exposure to equities (in percent of the pension liabilities) and maximum 15% (of the pension liabilities) in unhedged currency exposure.
- **Legal soft constraints**: Capital adequacy and solvency ratio.
- **Balance sheet constraints**: These constraints define the profit and loss equations, and the relationship between the result variables and the assets and liabilities.
  Høyland and Wallace (2001b) describe the legal framework in more detail.

The TAA-model has three types of constraints:

- **Liquidity constraints.** As a relatively large player in futures and options, liquidity is a constraint in certain markets, in particular in the domestic markets. We have chosen to model these as hard constraints by simply limiting the maximum exposure to each investment vehicle.

- **Subjective constraints.** When solving the model, we often find that there are many different portfolios that have fairly equal risk reward profiles. Subjective constraints are sometimes used in order to push the solution to a portfolio that better fits the portfolio managers' judgmental views on the portfolio composition. Examples are constraints that limit the allocations directly, or constraints that limits the amount of risk taken on a single asset class, relative to the total risk.
- **Risk constraints.** A key decision is to decide on the degree of utilization of the risk limit. This constraint is actually "outside" the model, because of its non-linearity and an iterative procedure secures that the desired risk level is achieved.

For both models, all constraints are linear. A mathematical description of a simple version of the model is given in the equation below. To simplify the presentation we have presented a single period TAA-model that assumes cash investing, where short selling is allowed, and we have left options out. We have assumed a target return of zero.

Define the following parameters:

P_s	probability of scenario s,
$R_{i,s}$	return on asset i in scenario s,
A, B	shortfall cost parameters,
LL_i, LU_i	lower and upper liquidity bound,
SL_i, SU_i	lower and upper subjective bound,

and the variables

x_i	investment in asset i,
z_s	shortfall in scenario s.

The mathematical model is then:

$$\text{Max} \sum_s \bigl(\alpha \cdot P_s \cdot R_{i,s} \cdot x_i - (1-\alpha) \cdot P_s \cdot (A \cdot z_s + B \cdot z_s^2)\bigr)$$

subject to

$$z_s \geqslant -\sum_i R_{i,s} \cdot x_i,$$
$$LL_i \leqslant x_i \leqslant LU_i,$$
$$SL_i \leqslant x_i \leqslant SU_i,$$
$$z_s \geqslant 0.$$

A linear plus quadratic shortfall cost function is chosen. The linear term is included in order not to underestimate the penalty cost for small shortfalls. An iterative procedure is applied where α is tuned until the target risk level is achieved.

2.5. *The single versus multi-period framework*

In the literature many papers argue for dynamic, multi-period models, see, for example, Cariño and Ziemba (1998), Consigli and Dempster (1998), and Consiglio, Cocco and Zenios (2001); Kouwenberg and Zenios (2006) and Dert (1995). We agree with the arguments given by these authors, but we argue that the value of a multi-period framework depends on the problem to be solved, and that there might also exist good reasons for choosing a single period framework.

In Section 2.2 risk measures for our two models were discussed and for the SAA-problem we concluded there were more conflicting goals with different time horizon than in the TAA-model. For the SAA-model, it would be impossible to model these conflicting goals in a single period framework, whereas for the TAA-problem, the risks are realistically described in a single period model.

Furthermore, there are more pragmatic issues to be considered when deciding upon the dynamics of a model. One needs to consider the frequency of model runs, and the construction of input data. For the TAA-problem, the estimations of the input data are extensive and time consuming, and the model is run on a monthly or more frequent basis. A multi-period framework radically increases the number of input parameters that must be estimated or specified, so moving to a multi-period setting would be a severe challenge from a practical point of view. Clearly, if the necessary distributions are too extensive, the quality of the input data will decrease. See Section 3 for a discussion on generating the input data, and see the case study in Section 5 for a detailed description of how the input data is specified for the TAA-model.

The SAA-problem is solved for fewer asset classes and it is also solved less frequently. For the SAA-model, a multi-period framework is chosen; the model currently in use has two periods (technically a three-stage stochastic programming model). Despite the potential advantages of a multi-period framework, the TAA-team uses a single period model (two-stages) for the tactical model.

In the future we will consider moving into a multi-period framework also for the tactical model, and investigate further the advantages and disadvantages. In particular, we believe that the multi-period setting is useful when options are allowed as investment vehicles. In addition we hope to achieve a better understanding of the dynamics of portfolio management by analyzing the solutions from a multi-period decision model. The challenge will be to secure high quality input data, in particular, the inter-period dependencies of the uncertain variables need to be investigated.

2.6. Comparison with the mean-variance model

The standard mean-variance model has a single period whereas the SAA-model has multiple periods. See the previous section for a discussion of advantages and disadvantages regarding the choice of a static versus a dynamic model. In addition, our models differ from the standard mean-variance framework in at least four aspects:

- transaction costs,
- treatment of risk,
- asymmetric and fat tailed distributions,
- sensitivity to changes in input parameters.

We saw in Section 2.3 that derivatives are used as investment vehicles for the TAA-model. When trading futures, the transaction costs in terms of broker fees are relatively low. The transaction costs are, however, larger if we take into account the bid offer

Table 1
Marginal central moments

	Money market	Bonds	Equities
E (return)	5.0%	6.6%	9.0%
Standard dev	0.0%	8.0%	20.0%
Skewness	0	0	0
Kurtosis	3	3	3

Table 2
Correlation estimates

	Money market	Bonds	Equities
Money market	1	0	0
Bonds	0	1	0
Equities	0	0	1

spread cost and the potential market impact cost. With cash instruments as investment vehicles the total transaction costs are usually larger. With all costs included, the transaction costs influence the optimal asset allocation mix. While in the standard mean-variance model, transaction costs are not taken into account, both the TAA- and the SAA-model incorporate transaction costs.

Section 2.2 described how risk is defined in both models. Only downside risk is penalized in the objective function. This is in contrast to the mean-variance framework, where deviations from the mean, both down- and upward, are treated equally. With symmetric return distributions (as the mean-variance framework assumes) this is only a conceptual problem, since using semi- (downside) variance or variance as risk measures will lead to the same solution. However, with asymmetrical distributions this is not the case. We allow for asymmetrical distributions in both the TAA- and SAA-model.

Let us illustrate and quantify the effect of asymmetrical distributions with an example. Consider a case with three asset classes; cash, bonds and equities. Expectations for the returns on these asset classes are given in Tables 1 and 2. Note that the expectations in Tables 1 and 2 are consistent with a multi-variate normal distribution. We now run the scenario generation model and create 1000 outcomes with statistical properties as in Tables 1 and 2. We then solve with the mean-variance model and the model proposed in this paper, while targeting equal risk levels (in terms of the variance of the return of the portfolio). With a reasonable shortfall cost function, the optimal solutions obtained from the two approaches are equal for all practical purposes. Now, let us assume that the return distribution for equities is skewed to the right as in Figure 3. Solving the mean-variance model leads to exactly the same solution as before, since this model only reacts to the two first moments of the distribution. With the asymmetrical risk measure however, the optimal holding of equities rises from 17.5 to 26.4%. (In addition to the

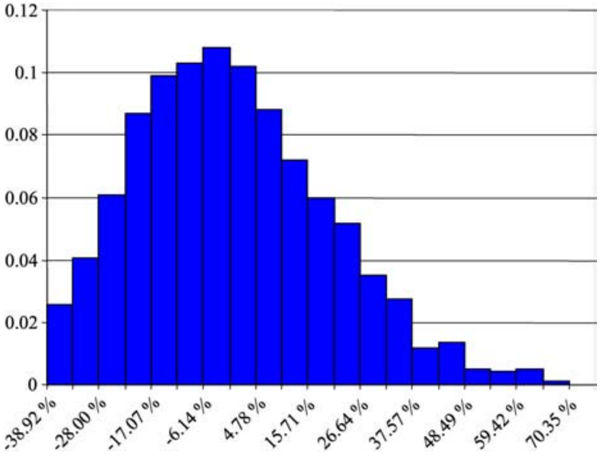


Fig. 3. A distribution with skewness of +1.

market expectations, the only other inputs are the parameters defining the shape of the shortfall cost functions. Changing these parameters within reasonable bounds does not significantly alter the results.)

The equity holding increased because the (positive) asymmetry in the return distribution of equities fits the asymmetry of the utility function. As explained in Section 2.1, the objective is to maximize the expected return net of expected shortfall costs, where the marginal shortfall cost is increasing with respect to the shortfall. Since outcomes with relatively large negative excess returns give large negative contributions to the objective function value, the model will seek to avoid such outcomes. Since a right skewed return distribution has less probability mass in the left tail of the distribution relative to a symmetrical return distribution, the right skewed return distribution will be preferred relative to a symmetric return distribution (all other factors being equal).

The above example leads us to the final difference between the two approaches. The mean-variance model has been criticized for being over-sensitive to changes to the input parameters, in particular to the expected returns. Chopra and Ziemba (1993) find in their investigations that the relative importance between means, variances and covariances is about 20:2:1, but such that the importance of the mean can be even higher depending on the level of risk aversion. With our framework, the sensitivity is lower. This is because there are other properties than the mean, variance and the correlations that stabilize the optimal solution with respect to changes in these properties. As argued above, the model will for example prefer asset classes that are positively skewed as they fit the utility function. At this point however, this is only an empirical and intuitive result, which can be of interest for future research. See also Dupacova (2001) for a discussion on sensitivity analysis of financial models.

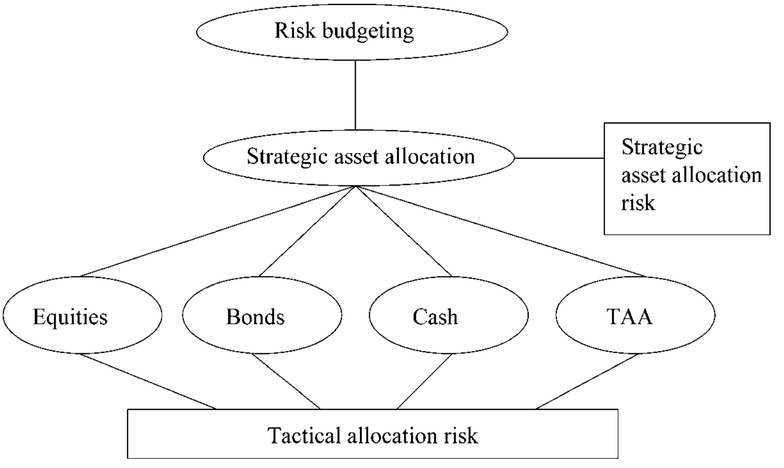


Fig. 4. Risk budgeting.

2.7. *The linkage between the SAA-problem and the TAA-problem*

In Section 2.5, the conflicting goals of the ALM-problem were described. The key is to find an asset allocation mix that takes into account the balance sheet risk, the competitor risk and the SAA-team's longer term market expectations. Such an asset allocation is found with the help of the SAA-model and the cash is allocated to various asset managers who manage funds within each asset class. These mandates can be quite broad (for example, foreign equities) or they can be more specialized (for example, European Telecom). For each of these mandates a benchmark and risk limits relative to the benchmark are established. On top of the cash allocations within each asset class, the SAA group allocates risk to TAA-mandates, see Figure 4. This process of allocation risk at different levels and to different decision makers is in the literature called *risk budgeting*, see, for example, Rahl (2000) and the references therein for a discussion.

For the TAA mandate described in this paper the SAA group has allocated a limited amount of cash that is necessary to cover margin requirements on the TAA derivative positions. With little capital used the cost of capital in terms of alternative return is close to zero. Therefore, the benchmark return for the TAA-mandate has been set to zero.[2] The only constraint in the TAA-mandate is a risk constraint given by a Value at Risk limit, which is equivalent to a tracking error limit relative to a benchmark position vector that consist of zero's.

[2] Theoretically the zero benchmark return is too low because the TAA group will use risk capacity of the life insurance company, which should be compensated. Hence, the benchmark return should be risk adjusted to a positive number. However, this issue is of more importance when discussing the incentive structure between the TAA group and the life insurance company, since adjusting the benchmark return would not change the important characteristics of the utility function.

3. Data collection, market expectations and scenario generation

This section discusses the generation of input data to a stochastic programming model, an area we believe the academia has not paid enough attention to. Much of the research within stochastic programming has focused on the decision models themselves. Different models have been compared, without too many questions on assumptions regarding the treatment of the input data. In particular, there is rarely discussion regarding whether or not the scenario tree used in the optimization represents the data appropriately. This section discusses different approaches for generating the input data and explains how the input generation is done for the TAA- and the SAA-models discussed in this paper.

The goal of the process is to generate a set of scenarios that represents an adequate description of the uncertain variables. We divide this process into three major steps, collecting data and information, analyzing the data and specifying market expectations, and finally, generating the scenarios used as input to the model. Section 3.1 briefly explains the data and information process. In Section 3.2 we motivate the methodology we have chosen for expressing (judgmental) market expectations. Section 3.3 discusses our view on scenario generation, in addition to details about what we actually do. Section 3.4 briefly outlines the generation of the multi-stage scenario tree.

3.1. Data collection

At a relatively low cost, almost unlimited amounts of data can be collected, whether empirical price data, macro economic data, micro economic data, or analyses from central banks and organizations like IMF and OECD. The data required to support a given decision model depends on the philosophy behind the modeling. The input to the decision models that have been reported in the stochastic programming literature, typically only requires empirical price data (and sometimes also empirical macro economic data). We will explain in the next section how the asset managers of Gjensidige-NOR use subjective judgments when estimating the input data of the decision models. To support that judgment, the key is to extract and structure the data that is relevant. It is however, out of the scope of this paper to explain this process in detail.

3.2. Formation of market expectations

In order to run a portfolio optimization model, one must elicit and quantify the market expectations. The number of asset classes in the TAA-model is typically between 15 and 25. It is a comprehensive task to specify, say, a 20-dimensional joint distribution. Several methodologies have been developed over the last few years. Many rely on regression procedures to reduce the number of random variables that needs to be identified and analyzed explicitly. This modeling is based on sampling or the development of stochastic processes (see, for example, Mulvey, 1996). The choice will depend on analytical preferences as well as the availability of data and understanding of the stochastic processes.

The asset managers of Gjensidige-NOR prefer to express market views by quantifying their return expectations for the asset classes directly. They preferred to describe the return of an asset class by its marginal distribution, and its interrelation with the other asset classes. This raises the question of which properties of a distribution are in fact important to capture correctly in order to have a good decision support model. Høyland and Wallace (2001a) discuss why the selection of important statistical properties is problem dependent. The Markowitz mean-variance model, for example, reacts to only the first two moments of a distribution (including the cross moments). Any efforts to obtain correct descriptions of higher moments are a waste of time since the model does not react to them.

To find a portfolio that has the best risk reward balance, it is necessary to estimate the expected portfolio return and some measure of uncertainty. Hence, expected value and standard deviation are obvious statistical properties that need to be specified. However, as we saw in Section 2.6, the *shape* of the probability distribution can influence the optimal portfolio choice to a large degree.[3] Skewness and kurtosis give us the ability to express our subjective views on how the uncertainty is distributed around the mean for a given asset. The motivation for introducing the skewness and the kurtosis comes from two basic facts: Firstly, asset class returns are not normal, at least not with reasonably short time horizons, see, for example, Jackwerth and Rubinstein (1996), and secondly, the optimal portfolio is highly influenced by these distribution properties (for all reasonable utility functions).

For the asset allocation problems at hand, we concluded that for the marginal distributions, the first four moments (i.e., expected value, standard deviation, skewness and kurtosis) are needed to create a stable decision support model (see the next section for a discussion). For describing the relationship among asset classes, correlations were found to be enough. The first four moments of a marginal distribution can either be derived from the marginal return distributions, or they can be specified explicitly. As the experience of the asset managers has increased, the latter approach has become standard. Hence, describing market expectations is equivalent to describing the necessary moments and correlations.

It is a well-known fact that in stressed market conditions the longer-term average correlation structures tend to break down. This was the main reason for the failure of Long Term Capital Management, see Jorion (2000). Crash scenarios influence the optimal portfolio structure, and to capture the crash scenarios in an appropriate way, we split the future into two states, one "normal" state and one "crash" state, and then specify the marginal distributions and correlations separately for the two states. The two are then combined using an estimated probability of the two states occurring. This idea is later followed up by others. Geyer et al. (2002) use three sets of covariance matrices: 70% regular, 20% volatile, 10% crash. In crash, bonds and stocks are negatively correlated while in the other two they are positively correlated.

[3] This is under the assumption of an asymmetrical utility function for which we argued in Section 2.1.

The portfolio managers find the possibility of expressing views on the asymmetry and the degree of fat tails on the return distributions very useful. In many cases, return distributions are obviously asymmetric. The zero level in interest rates, for instance, naturally creates asymmetry. There is less obvious asymmetry for equities, but the pricing of the market, relative to future expected earnings and the macro economic environment, often indicate whether potential large, longer-term moves are to the upside or to the downside.

Investors are always faced with the possibility of a stock market crash. The ability to express crash scenarios, and assign them probabilities, is considered to substantially increase the quality of the input data. Both the expected magnitude of a potential crash, and the probability for it to happen, are changing over time and the portfolio managers seek to have an active view on these factors.

3.3. Scenario generation

From a methodological point of view, there is no difference between how we generate scenarios for the two models. In this section we will explain our approach with reference to the single period tactical model. In Section 3.4, we will comment on the multi-period case.

Before turning to the technicalities of the scenario tree generation, let us explain our main philosophy. There is no general agreement among either scientists or practitioners about best practice. First, there is the issue of understanding the random variables. For most, the obvious staring point is data. However, from there we can go in many different directions. Some believe in estimating stochastic processes from the data. Others, and we belong to that group, prefer to add subjective information to the data. One issue is whether or not the past describes the future well. The decision-makers in Gjensidige Nor are not willing to invest on that assumption. Further, theory tells us how to extract implicit information from market data. Jackwerth and Rubinstein (1996) show how to derive the risk neutral return distributions from option prices. For a general overview over scenario generation methods, see Dupacova, Consigli and Wallace (2000).

At some point, the user will be satisfied with his description of the possible futures. He may have empirical data that he believes describes the future well, he may have some stochastic processes (estimated or guessed), he may have the possible futures extracted from market data, or he may simply have subjective views. But it is unlikely that the data is in a form suitable for stochastic programming. We think it is important to point out that the problem of making a scenario tree from specifications, whatever they are and wherever they come from, is a separate issue from understanding the random variables.

Our methodology is based on constructing a scenario tree from the specifications. We create the outcomes of the tree by using an iterative procedure that combines simulation, Cholesky decomposition and various transformations, see Høyland, Kaut and Wallace (2003) for details. The scenario generation process is illustrated in Figure 5: The scenario generation model constructs a limited set of outcomes, in the academic literature known as a *scenario tree*, which is consistent with the specified moments and correlations, i.e., the tree has exactly the required properties.

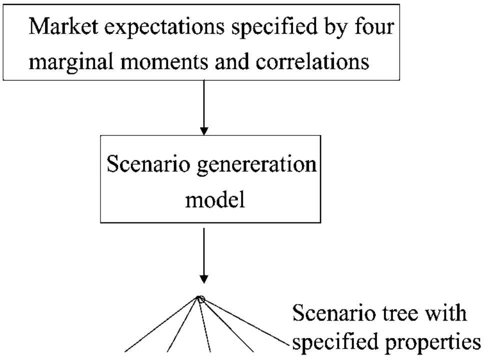


Fig. 5. Scenario generation process.

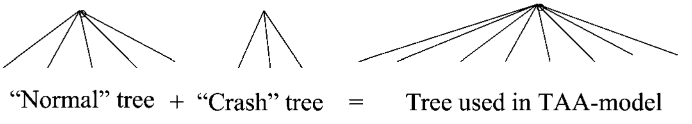


Fig. 6. Aggregation of scenario trees.

In order to capture the extreme events of a market crash, two scenario trees are generated this way (one for normal market conditions and one for crashes) and these are aggregated to a larger tree. The large tree is used as input to the TAA-model, see Figure 6.

A final issue is whether specifying four marginal moments and correlations is enough (or too much). Or more generally, how do we know that the tree we have is a good one? Ultimately, this is an impossible question to answer. Except in special situations, like that of throwing a (fair) dice, there is no way of knowing what are the correct descriptions of (future) random events. So although we know, in principle, what we are looking for, we cannot know if we have found it. We can check, ex post, if we make money or not, and if we do, we can choose to be happy, but we may simply have been lucky. And we may of course lose money even if we happen to have a perfect description of the possible future events. So losing money is not a proof of a bad tree. We must always allow for bad luck.

So we have to go for something less to test the quality of what we have. The methodology described above takes care of one potential problem—that the discretization itself has introduced unwanted properties. It has moments and correlations according to specifications (in Høyland and Wallace (2001a) we present a methodology allowing for more involved sets of properties). This may sound obvious, but many methods used today do not have this control, in particular, sampling will not provide this control for trees small enough to be used with a stochastic program. Sampling will only give us what we need in the limit, but in the limit we cannot solve our optimization problem. So knowing what we have is a first step.

Our next step is to perform an in-sample test. We generate a large number of trees all having these properties. Each of these trees is used to solve the portfolio model, and we verify whether the expected objective function value is the same (within an epsilon) in all cases. If not, our tree generation has not passed its in-sample test. What we then do, in practice, is to increase the size of the trees, to stabilize the properties of the trees that we have not specified. The effect is that, eventually, the in-sample test is passed. In effect, whichever tree we use, as long as it comes from our generation procedure, the expected objective function value of the portfolio model will be the same. (We use expected objective function value as measure and not solution, since we do not wish to declare two investments policies, which happen to be equivalent in terms of behavior, to be different.) Empirical testing has shown that for a TAA problem with 20 asset classes, a few thousand scenarios are necessary to pass this in-sample test.

Normally, we know (or think we know) more about the random variables than we could put into the tree. This could be expressed in terms of a simulation model or maybe simply an event tree far too large for a stochastic program (but such that we, for some reason, believe it to be better than the relatively small tree we used for optimization). If so, we do an out-of-sample test, defining this big tree or simulation model to be "the truth". To pass the out-of-sample test, all investments coming from the different trees in the in-sample test must produce the same expected objective function value also in the out-of-sample test.

It is hard to get any further than this. What remains of possibilities is to test using historical data. When using partly subjective data, as we do, historical back testing is difficult, as we must estimate how the subjective data would have been, had we used the model in the past. In addition, we use historical results. This is presented in Section 5. The results are positive, and we believe there is a reason for this, but maybe we are simply lucky?

On a regular basis we collect (historical) data, extract data from financial instruments, add subjective views, and then generate scenario trees that will pass the in-sample and out-of-sample tests. The portfolio model is then run, and the results used as a basis for investments.

3.4. From a single to a multi-period scenario tree

From a technical point of view, generating multi-period scenario trees is not so different. What we do in practice, is to start at the root level of the tree, corresponding to the circle in the simplified tree of Figure 7, and generate the outcomes for the first period, as described. We then move to one square at the time, in order to calculate the tree below that square. If the properties of such a second-period tree depend on the first period outcomes, they can be calculated before the tree is made. When all squares are completed, we progress to the triangles. Although this is just a heuristic, it has always

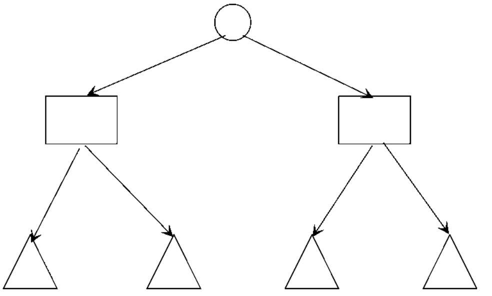


Fig. 7. Illustration of the generation of multi-period scenario trees.

worked in practice. A major challenge in this procedure is to estimate the inter-period dependencies.

4. The impact of the TAA-model on the organization

The TAA-model has had most organizational impact, and this section will focus on how the TAA-model has influenced the setup of the TAA-team. Section 4.1 discusses in detail how the TAA-team is set up relative to the required input to the model. The processes related to the TAA-model represent the core of the TAA-group's activities, and in Sections 4.2 and 4.3 we explain how these processes fit into the whole investment process and investment philosophy of the TAA-team.

4.1. Distribution of responsibilities

Section 2.3 should clarify that the investment universe is large for a small team. In order to cover the investment universe and obtain market expectations of high quality, we need to delegate responsibilities, make use of other internal and external resources, and use information technology and modeling systems in an efficient way. It is not within the scope of this paper to explain these processes in detail. We focus on how the TAA-team is organized relative to the required model input and illustrate how the organization of the team influences modeling choices in the TAA-process.

The TAA-model is run on a team consensus basis. In other words, the team as a whole discusses and agrees on the expectations about the markets. The model requires much input data and it is seen as crucial that each number has its "owner". This ensures that no model input is overlooked and should hopefully ensure high quality input.

The asset class universe naturally is divided into five regions; Japan, US, UK & Sweden, EMU-Europe and Norway, and one analyst is assigned to each region. Further, one *global* analyst has his main focus on the *global* expectations for cash, bonds, equities, commodities and currencies.

Expectations need to be quantified both for the marginal distributions and for correlations. The distribution of responsibilities is clear-cut for the marginals. For correlations, however, the process is naturally more complex.

As regards marginal distributions, the regional analyst is responsible for selecting the appropriate asset classes within the regional asset universe and to quantify the expectations. Typically, this involves specifying the four central moments for a broad equity index, ten-year bonds and one or two money market futures. The global analyst's main responsibility is input to the expectations for the "global returns" in cash, bonds, equities, and commodities. Clearly, the global analyst plays an important role, and is typically the most experienced analyst.

Naturally, expectations within different regions, expectations for global cash, bonds, commodities and equities, and expectations for sectors, all interrelate due to common factors driving the financial markets across regions. Further, input from the different regions influences the views of the global analyst and vice versa. It is clear that the TAA-team is facing a challenge when all expectations are to be calibrated, see Section 5.3 for further discussion.

Specifying the marginal distributions is a comprehensive task, but specifying the correlation matrix can be even more of a challenge. With, say, 20 assets there are 190 correlations to estimate. It is widely accepted that correlations can dramatically change over time, not only in size but also in sign. Just as the expectations for the marginal return distributions are judgmental, so are the expectations on the correlations. The TAA-team is not comfortable with uncritically using empirical correlation matrices as input to the model and seeks to gain an active view also on correlations. Empirical analyses are run, but the results are not mechanically used as input to the model.

In addition to the marginal distributions, the analyst responsible for a region must also specify his expectations for the correlations within the region. The hardest correlations to understand are the cross asset class—cross region correlations, such as the correlation between US stocks and UK bonds. To specify these correlations simple heuristics are applied. For instance, given the correlation between US stocks and US bonds, and US bonds and UK bonds, it is possible to make an intelligent guesstimate of the correlation between US stocks and UK bonds.

In addition to dividing responsibilities between regions, responsibilities are also split between asset classes and model development. To improve the quality of the input data, each member of the team has a special responsibility to develop decision support models for an asset class. These are then used across regions.

Out of a total of six analysts, two to three spend some time each day closely following the information flow and the market development. There is always one portfolio manager who is sitting close to the market, and whose responsibility is to inform the team if there are significant changes in the market conditions and to execute the trading strategies.

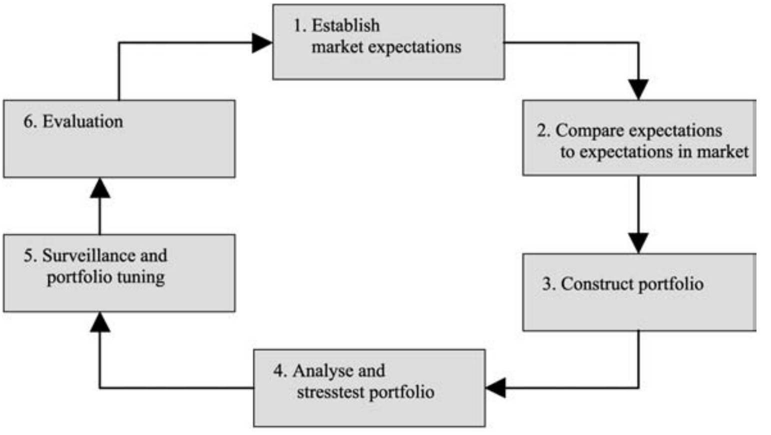


Fig. 8. The investment process.

4.2. The model and the investment process—a case study

This section describes the investment process and explains the role of the TAA-model. We will particularly emphasize the parts of the process that are influenced by the modeling framework. We have divided the investment process in six steps as shown in Figure 8.

1. The previous section explained the required model input. At least once a month each analyst must quantify his market expectations. To secure the quality of these inputs, a structured information collection process and a systematic market analysis approach is required. Section 5 will describe our experiences and discuss reasons why we believe that the modeling framework has created a structured and transparent analysis process. An example of the output from the market expectation process is shown Table 3.
2. Having established market expectations, these expectations are compared and evaluated relative to what is implied by the option markets.
3. The portfolio construction is an iterative process as illustrated in Figure 9. The starting point is our market expectations. A large number of scenarios consistent with these expectations, typically a few thousand, are generated and visualized by histograms for each marginal distribution. See Figure 10 for an example of a return distribution. The team discusses the visualization, and more often than not, some expectations are changed when the team is confronted with the graphical representation of the expectations. When the generated scenarios are accepted, they are input to the TAA optimization model.

   The TAA-model is run, and the optimal asset allocation mix is analyzed. As discussed in Section 2, experience has demonstrated that there are many asset allocation mixes that are almost equally as good in the sense that they provide approximately

Table 3

Market expectations as of February 8th, 2002. In this particular run, 15 asset classes were used. The matrix illustrates the market expectations for each asset class in the "normal" state of the world. In addition to this matrix, the correlation matrix must be specified. Specifications for the "crash" state of the world are specified in the same way, in addition to the probability of the crash occurring. The expectations are given with a three month time horizon. Note that for money market and bond futures, the expectations are given on the interest rate level not returns

	EV	Stdev	Skew	Kurtosis	Spot rate	Future rate
3 months money market, LIBOR						
September 2002 USA	2.75%	0.50%	0.30	4.00	1.91	2.67
September 2003 USA	4.51%	0.60%	−0.30	4.00	1.91	4.61
September 2002 Europe	3.59%	0.40%	0.00	4.00	3.35	3.68
10 year government bonds						
USA	5.10%	0.45%	0.50	3.50	4.887	4.99%
Japan	1.60%	0.35%	0.40	3.50	1.49	1.53%
UK	5.10%	0.55%	0.40	3.50	4.99	4.98%
Germany	5.10%	0.50%	0.20	3.50	4.98	5.05%
Sweden	5.48%	0.60%	0.40	3.50	5.42	5.45%
Equities						
USA	5.5%	9.00%	0.60	3.80		
Japan	4.0%	15.00%	0.60	3.80		
UK	6.0%	10.00%	0.60	3.80		
Germany	5.9%	12.00%	0.60	3.80		
Norway	6.8%	9.00%	0.60	3.80		
Sweden	6.0%	14.00%	0.60	3.80		
Commodities						
Oil (WTI)	−0.50%	20.00%	−0.15	4.50		
Goldman Sachs Commodity Index	1.0%	12.50%	0.00	4.00		

the same risk reward tradeoff. If we are not comfortable with the allocation mix, this is usually due to the market expectation input. We would therefore reevaluate some of the market expectations. However, if we believe market expectations are appropriate and we still dislike certain minor aspects of the optimal asset allocation, we will constrain the portfolio construction problem based on experience, even though the consequence is that the constrained optimization problem will find a portfolio with slightly worse risk reward tradeoff. An example can be to constrain the position in an asset class for which we lack conviction in our estimate for the expected return distribution (but where the estimate is still our best guess). It is important to notice, however, that the model decides the main structure of the portfolio, and that the subjective constraints only to a small extent influence that structure. From a modeling perspective, it is our belief that the error introduced by the subjective constraints is well within the error bound of the model itself. The (unconstrained) optimal portfolio given the market expectations partly described in Table 3 is shown in Table 4.

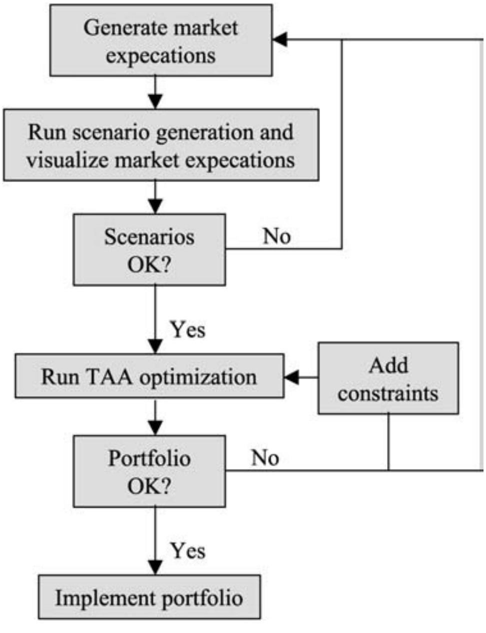


Fig. 9. The portfolio construction process.

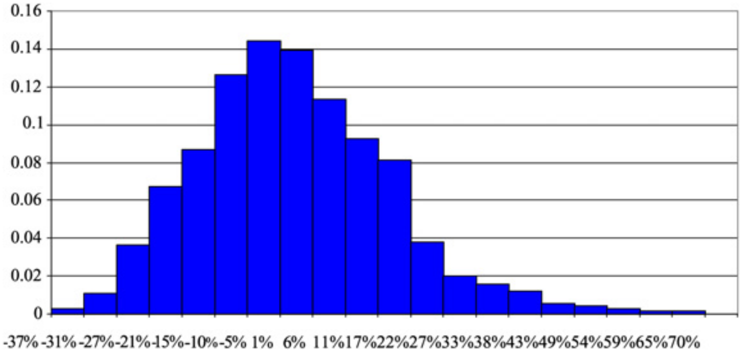


Fig. 10. Visualization of the expected return distribution of Japanese equities. The histogram includes all scenarios, both from the "normal" and the "crash" state of the world. The skew of the distribution is 0.55, whereas the kurtosis is 3.61 (compared to 3 of the normal distribution).

4. Return and risk attribution analyses, in addition to stress testing, are applied to the optimal portfolio. The risk and return attribution allocates the expected return and expected risk to the individual positions in the portfolio. For stress testing the portfolio is tested by historical simulation and on subjective stress scenarios (which of

Table 4

Optimal portfolio (in mill USD). Notice that the volatility of the return of each asset class varies substantially, see Table 3 for estimates. For instance, the volatility of equity futures can typically be 100 times larger than the volatility of the money market futures. That explains the large differences in magnitude of the positions

	Mill USD
3 months money market, LIBOR	
September 2002 USA	−1500.0
September 2003 USA	2029.4
September 2002 Europe	1029.4
10 year government bonds	
USA	−39.7
Japan	−31.1
UK	−27.6
Germany	12.6
Sweden	14.1
Equities	
USA	9.0
Japan	3.2
UK	3.1
Germany	−0.4
Norway	5.9
Sweden	−3.7
Commodities	
Oil (WTI)	0.0
Goldman Sachs Commodity Index	2.2

course are different from the ones used in the optimization). The historical simulations show the performance and volatility of the return of the portfolio during different economic regimes. Such information is useful for estimating the risk, and analyzing the characteristics of the portfolio. The historical simulation will, for example, provide information about whether the portfolio is a market trend follower or, to the contrary, is seeking for a turning point in the financial markets.

5. New information arrives and prices change continuously. In such an environment, the TAA-team finds it important to respond quickly to new information or changes in market conditions. The TAA-optimization creates a core portfolio. As a response to new information, this portfolio can be tuned without a new run of the TAA-optimization model. But the main structure of the portfolio will not change without a re-run. Adjustment of the size of the positions, however, and even putting on or taking off positions which are fairly independent, i.e., have a low correlation with the rest of the portfolio, is done on a continuous basis. As each person in the team is responsible for a subset of the market expectations, an initiative to make changes in the portfolio can come from anyone in the team. Often it will come from one of the port-

folio managers who closely monitors the markets and the news flow. To reallocate, the portfolio manager in charge and an additional team member, either a regional analyst or the global analyst, must agree. If new information or large price moves triggers the need for major revisions, a complete review of the market expectations and a new TAA-run will be done.

6. At the monthly (or more frequent) TAA-meetings, the previous portfolio is evaluated. There are several challenges in the evaluation process. First, simply looking at the profit and loss of each position is inadequate. Some positions will be so-called profit generators, while others will be hedges. Often, hedge positions have an expected loss. So we need to compare expectations with realizations. Evaluating the estimates for the mean is straight forward, as the realized return can be compared with the estimate directly. Evaluating the estimates for correlations, standard deviations, skewness and kurtosis is, however, more difficult. For the period over which the expectations are given, there is of course only a single observation. Therefore, we need to split the period into shorter time units, typically days, and use these intra period observations to estimate the statistical properties. Since the statistical properties are not time-scale invariant, this approach is not correct, but it is the best one can do.

   Clearly, it is possible to look at the profit or loss for the fund as a whole in order to measure it's success. Evaluating the contribution from each member of the team is more difficult. But, since the market expectations are quantified and documented, and since they can, to some extent, be compared to what actually happened, we believe the evaluation process is of high quality and adds value. We are very careful not to base our evaluation on hindsight.

   The incentive structure is related to the evaluation process. In building the incentive structure, the main goal has been to make sure that the incentives of the asset managers are in line with the incentives of the clients and also reflect the longer-term goals of the asset management company. The TAA-team as a whole has a performance related bonus and a bonus based on qualitative aspects. The challenge is to motivate the individuals by giving them credit for good work, but at the same time ensuring that the incentive system does not have a negative influence on the group decision process and creates internal conflicts.

4.3. Diversification in time

The TAA-model described in this paper represents the core of the TAA-group's activities. However, the group also makes investment decisions without employing the model. This section describes how the TAA-model fits into the broader picture of the group's activities.

Financial markets have trends and tend to be mean reverting with different time horizons. Some market participants seek to explore such trends on an intra-day basis; others explore long-term trends related to the business cycle.

The TAA-team seeks to exploit the price movements within the different time horizons. The market analysis processes seek to identify on what time horizon the different factors will influence the markets. The tactical asset allocation is divided into three groups, which have different investment style characteristics:

- Long-term—3 to 12 months.
- Medium-term—1 to 3 months.
- Short-term—less than one month.

The model is used for decision support on all three horizons, but in a varying degree. The medium-term process is described in this paper and captures the core of the activities in the TAA group and is where the model is used most extensively.

The long-term TAA positioning seeks to exploit what we see as major long-term trends during an investment year. Typically, the longer-term positioning is done via options. The long-term positions will be evaluated on a continuous basis, as well as in each TAA-meeting. However, as the horizon is longer for these positions, larger price moves or more significant new information is needed to change them.

The short-term TAA style includes the activities discussed earlier in this section regarding tuning of the medium-term portfolio as prices change and new information arrives. This tuning activity seeks to exploit shorter-term market fluctuations. The TAA-team also includes discretionary traders who are active in the market on a daily basis. The time horizon for the positions taken varies from intra day to intra month. The traders sit close to the market in terms of news flow and price action. They are not only there to generate excess return, but also to function as an information provider and "noise-filter" for the rest of the group.

The primary goal for all activities is to generate excess return. If each investment style generates positive expected excess return, and is not perfectly correlated with the others, then adding them together clearly improves the risk reward trade-off compared to a single strategy fund. We are aiming at creating excess return from all three investment styles, and believe we achieve what we call *diversification in time*.

We also believe that the different activities provide synergies. For example, the medium- and long-term styles depend on the input from the traders and the traders depend on the input from the investment processes done in the medium-term style.

The goal is to allocate risk between the three styles so that there is a balance between this risk and competence, expected excess return and resources at each level.

5. Experiences

In the introduction, we explained how the strategic and tactical asset allocation decisions now are separated. The life insurance company is responsible for the strategic decisions. The horizon of these decisions is typically a year or more, and the strategic asset allocation will typically only be changed more frequently if there are large movements in the financial markets that lead to significant changes in the balance sheet risk or there are macro economic events that are considered to have significant influence on the expected return of one or more asset classes. The modeling framework has proved useful in analyzing the ALM problem in a consistent manner, taking into account the conflicting goals that the decision-makers face. The model has been used on an annual basis to support the construction of both the asset allocation and the liability structure. It has also proven useful in stressed market conditions when it has been run on an ad hoc basis. However, it is almost impossible to evaluate the success of the model because of the very low number of historical performance numbers and because of the conflicting goals. On which criteria should the performance be evaluated? Evaluating the success of the model is simpler for the TAA-problem. It is run on a monthly basis, giving us more historical performance data, and the success criterion is simpler.

Section 5.1 focuses on the development phase for both the strategic and the tactical model. Section 5.2 and onwards focus on the experiences obtained from using the TAA-model actively, with real money behind it since July 1999. We discuss our experience with respect to integrating new employees into the investment style/philosophy of the team. As described in Section 4, running the portfolio requires input from six analysts. Section 5.3 discusses the decision-making process in the team, including how the consensus expectations are calibrated. Section 5.4 puts some of these observations into the picture of organizational learning, while Section 5.5 presents the track record of the fund.

5.1. Developing the models

Since contact was established late 1995 between the asset management team of Gjensidige-NOR and the authors of this paper, the focus for the project has evolved as follows:

- 1995–96: Established the conceptual modeling framework for the ALM-problem of the life insurance company, including the definition of risk measures and the dynamics of the model.
- 1996–98: Developed the framework for scenario generation.
- 1997–98: Developed the TAA-model.
- 1998–00: Improved the algorithms for the scenario generation.
- 2000–01: Improved user friendliness of the TAA-model.
- Until spring 2002: Model used as described in this article.
- Until summer 2005: Used for tactical allocation, but in a different organizational setting.

After the initial work with the conceptual modeling framework for the ALM-problem, see Høyland and Wallace (2001a), we realized that the biggest challenge in terms of developing a model that would potentially be used for practical decision-making, was to construct adequate input to the model. This input had to be consistent with the judgmental views of the asset managers. In the financial industry there was, and still is, widespread skepticism regarding using asset allocation models. This is partly due to the fact that if managers have tried quantitative tools they have applied a mean-variance framework with the weaknesses discussed in Section 2.6. So, the first step was to agree upon how the judgmental views could be expressed. The conclusion to this step was that the asset managers wanted to express views on the return distribution of the asset classes directly, not on some common underlying factors that drove asset prices. Secondly, it was considered crucial to be able to express asymmetrical and potentially fat tailed distributions. After starting with a system where the users specified a marginal distribution by specifying a certain number of its percentiles, and derived the first four moments from these marginal distributions, the asset managers got used to, and became comfortable with, specifying the moments directly. See Høyland and Wallace (2001a) for more details on the work undertaken on the scenario generation methodology. From late 1997 and onwards, the strategic model was used as decision support for the ALM-planning of the life insurance company. The tactical asset allocation group was first established in 1998, and the TAA-model was then developed. As the TAA-group wanted to use the methodology on a more continuous basis, it became crucial to speed up the solution times. In particular, the scenario generation process was slow and a lot of effort was put into making efficient algorithms for solving the scenario generation problem, see Høyland, Kaut and Wallace (2003). Having speeded up the scenario generation process dramatically, the model could be used in a truly interactive way. The next focus was then to make the modeling system more user friendly.

The resources allocated to the project have been one full time PhD-student, plus additional resources from academia and Gjensidige-NOR. We believe one reason for the success of the project has been the ability to focus on the right issues at the right stages of the project. It is easy to get stuck on details and lose the “big picture” view in a project like this. An obvious reason why we have been able to avoid this is the very tight link between model developers and users. Involvement and commitment from the users have also secured that they have developed their thinking during the project and that they are familiar and comfortable with the concepts of the model.

5.2. Training and integrating employees

There are many different investment styles, or philosophies, for managing equity, bonds, and other kinds of assets. In one extreme, there are funds that are 100% dependent on one or a few star portfolio managers. In the other extreme, there are funds that can be run almost without any investment/market competence, for instance by making use of mechanical trading rules based on empirical prices. Clearly, funds that completely rely on one or a few star portfolio managers and their personal investment styles are vulnerable.

Asset managers compete for skilled labor, and will experience a certain turnover. For a fund it is therefore essential to have the ability to replace human competence smoothly, i.e., to be efficient in the process of educating and integrating new employees. Our experience in building the TAA-team is that the modeling framework has been very valuable in this respect. New employees are introduced to an established structured process. They will hopefully learn from the investment process, and in particular from the processes of generating market expectations. After observing from the sideline for a period, a new employee is allocated a responsibility and is forced to quantify market expectations. This normally creates motivation to develop a structured information collection and analyses process.

Different analysts will have different approaches in analyzing the markets. One extreme is the pure technician who bases expectations purely on historical price movements and overlooks economic fundamentals. Another extreme does not consider technical analysis at all, and bases expectations purely on fundamental analysis. In principle, the modeling framework we are describing does not exclude any analytical style. However, as a macro hedge fund, we are based on fundamental analysis. The group has developed a common platform that all the analysts apply, but within this platform, there are many different approaches to analyzing the markets, with differences in emphasis on different factors. The group believes that a variety of analytical styles is desirable.

5.3. The decision-making process

Consensus decision-making is generally a difficult task. The decision-making is not only about reaching a conclusion, but also a process where responsibilities have to be delegated and changed in a dynamic world. To ensure high quality decisions, it is important that consensus decision-making does not dilute responsibilities and remove possibilities for taking actions based on individual knowledge in special situations.

For consensus decision-making in portfolio management, an asset management group needs to agree both on market expectations and on the portfolio composition. As regards market expectations, the first step is to agree on which factors influence the portfolio composition. Should, for instance, asymmetry or fat tales in the return distributions be considered, and should a possible stock market crash be taken into account? Having reached agreement, consensus expectations can be established. Given consensus expectations, the next step is to agree on the portfolio composition.

For the problems we are discussing in this paper, the portfolio composition itself is a high dimensional and continuous optimization problem. Finding the potential profit positions and potential hedges can be difficult, and judging the magnitude of each position is even harder. In general, the group consensus process is made more complicated by differences among team members in both *experience* and their *personal risk attitude*. Due to these difficulties it is very common that investment funds are not based on consensus decision-making. Instead, the funds are often split into several sub-funds, which are run by individual fund managers.

We believe that the team consensus process is substantially simplified by using the modeling framework described in this paper. Its very existence forces the team members to see the full picture, and they all will have to be concerned, not only about their own input, but also that of the others. The portfolio construction itself is done by the TAA-model. Hence, the only consensus decisions are the market expectations. If there are disagreements regarding the quality of the portfolio, discussions will always have to revert to the data that produced the result.

There are, however, still challenges in the group consensus market expectation process. In particular, differences in experience and personal risk attitude can create imbalances in the expectation process. By the risk reward trade-off in the objective function of the TAA-model, a common risk attitude function for the group as a whole is established. However, all analysts have a feel for how the portfolio is constructed relative to the expectations. Hence, a risk averse analyst will be careful in presenting expectations that are likely to lead to large positions within "his" asset classes, while less risk averse analysts are more likely to seek such positions. The personal risk aversion levels for the members of the group are functions of both personal attitudes and experience. Due to such differences, the group has put a lot of emphasis on the process of calibrating the expectations.

When calibrating the expectations, much attention is given to the key expectations, which are the expectations for the return on global cash, bonds, equity and commodities, in addition to the ranking of the different regions within each asset class. A potential danger in this calibration process is that the senior group members dominate. However, the group accepts that there are differences in experience, and seeks to obtain a balance between influence and experience.

5.4. Organizational learning

The evaluation process was briefly discussed in the introduction to Section 5. Many traders use so-called trading plotters, where the reasoning behind a trade and the profit and loss taking levels are documented. With its quantitative approach, the TAA-team is naturally forced to do this documentation and all material from each TAA-meeting, including the analyses and the actual expectations, are collected. This is used to evaluate the performance from month to month and also to undertake a more comprehensive evaluation at the end of the year. Our belief is that systematic evaluation sharpens and motivates the members of the team and that this improves the quality of the investment process over time.

It may be useful to relate, briefly, the above discussions to organizational learning, see Huber (1991) for an overview. First, the knowledge acquisition here is mostly through performance monitoring. The organization keeps track of all its decisions with respect to investments, including the basis for the decisions, and the financial results. That way, it is possible to understand, ex post, why a certain decision was made, and hopefully learn, irrespective of the outcome. In particular, this way of keeping records will help avoid hindsight taking the front seat in the learning process. It is a well-known phenomenon that, ex post, all humans seek to explain what happened. Fischhoff (1982) calls this *creeping determinism*. It is hard to accept that an outcome was random, and that something else could have happened as well. Notice that a good decision can lead to a negative result, simply because of bad luck. Good records will not automatically overcome this problem, but it is a necessary basis for proper learning about random phenomena.

5.5. Track record

In July 1999, a separate portfolio, with positions based on the TAA-model, was established. Figure 11 shows the performance of the fund since the start up. The accumulated

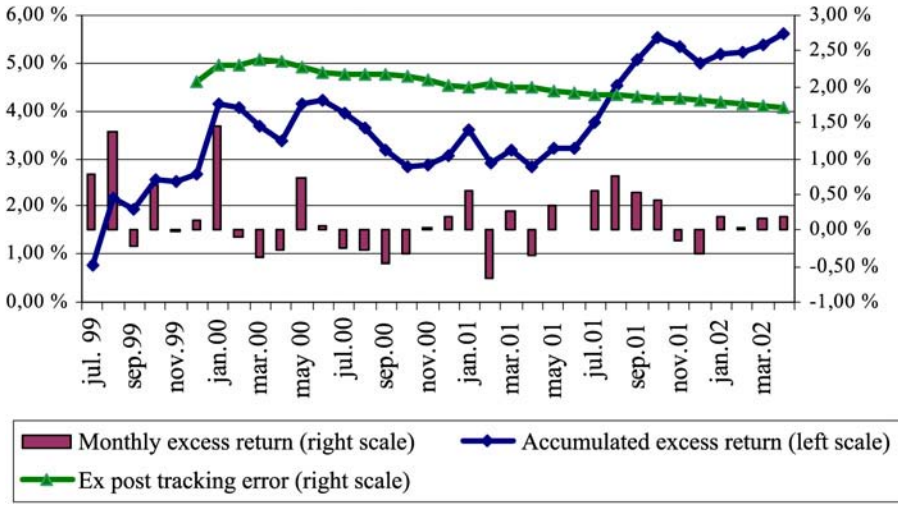


Fig. 11. Track record.

excess return from startup to the end of April 2002 is 5.60%. The risk, measured by annualized ex post tracking error (one standard deviation of excess monthly returns), is 1.71%. This gives an average information ratio[4] of approximately 1.19 each year.

The risk limit for the fund has been 5% in tracking error (one standard deviation of annual excess returns). The fund has been run for Gjensidige-NOR Spareforsiking and Gjensidige-NOR Forsikring (the Life and Non-Life insurance companies), and the capital base has been 7.5 billion NOK, equivalent to more than 800 million USD, since May 2000.

The fund applies a pure overlay strategy, meaning that only derivatives are used to construct the bet structure. This implies that the underlying funds can be invested in any benchmark, for instance, S&P 500 or Norwegian money market. Hence, for measuring the performance of the fund, the choice of benchmark is not important. The derivative activities generate an *excess* return. The total return for the client will be the return on the benchmark (whatever that is) plus the excess return generated from the derivative strategies.[5]

The risk profile of the fund has been modest relative to other funds with a similar macro hedge fund style. Since the fund has been used to run the tactical asset allocation for the Life and Non-Life companies (both running a moderated balance risk), we have

[4] The information ratio is the excess return divided by the standard deviation of the excess return.

[5] The model implication is that the shortfalls (outcomes that generates a negative contribution to the objective function value) will be measured relative to zero. For a cash portfolio, where the cash is allocated to the different asset classes, the shortfalls would be measured relative to the return on a specific benchmark.

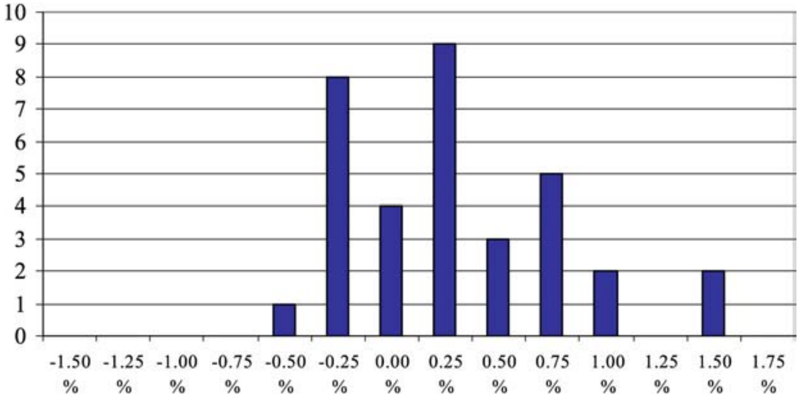


Fig. 12. Histogram of monthly excess returns. The lowest monthly return is −0.67%, whereas the highest is 1.46%.

chosen to run the fund with a large capital base and low risk. It is the relative risk multiplied by the capital base that will be relevant for the results. For example, if the capital base had been defined to be 750 million NOK instead of 7.5 billion, the excess return and the tracking error would have been 56 and 17.1%, respectively, instead of 5.6 and 1.71%. To evaluate the result we need to consider the information ratio and the shape of the actual return distribution. Figure 12 shows a histogram of the monthly returns since startup. Notice that the histogram is skewed to the right. The five highest monthly returns in absolute value are positive, and there are nine observations with a higher return than 0.5%, whereas only one observation with a lower return than −0.5%. The number of observations is low, but it appears that the group has been able to construct an asset allocation—which is consistent with the goal of optimization—to create a right skewed return distribution.

Our track record indicates ability to generated excess return even though the number of observations is too low to claim this with any statistical significance. Of course, it is possible the results are due to pure luck—it is hard to judge. If it is not pure luck we need to justify why we are able to generate this excess return—what market inefficiencies are we exploiting?

The motivation for the modeling framework comes from what the TAA-team believes to be the two key sources of potential excess return (i.e., return in excess of a benchmark return):

- Our own market expectations and
- Portfolio construction.[6]

[6] With this motivation for the modeling framework, the efficient market theory is abandoned. For a comprehensive discussion on market efficiency, see Malkiel (2003) and the references therein.

There are of course other sources of excess return, such as market timing, arbitrage opportunities and dynamic utilization of risk limits (as well as luck and chance), but we believe that our own market expectations and the portfolio construction are the two most important elements. Most portfolio managers are subjective/judgmental on both elements. Hence, the decisions made are a mixture of judgmental views on the market and a judgmental view on how to create a portfolio consistent with the market expectations. Within the modeling framework we have described, we are still judgmental on market expectations, but we run a tactical asset allocation model to create a portfolio that is consistent with our own market expectations.

We believe that there are particularly large inefficiencies across asset classes, whilst there may be lesser inefficiencies within asset classes. This is partly caused by the prevailing industry standard of having sector specialists who do not look at bets mixing asset classes. As have been highlighted in this paper, the TAA-group is mostly betting on the relationships among classes and not within classes.

To remove inefficiencies among classes, there must be some players in the market filling the gaps. But this requires appropriate tools, as it is impossible for a human brain to get a proper overview over the effect of bets in light of all correlations and lacks of normality. And few have these tools. This is where we believe we have an advantage.

6. Conclusions

The paper describes the use of two stochastic programming based decision support models, one which is the core of a macro hedge fund run by GN Asset Management, and another which is utilized on a strategic level for Gjensidige NOR Life Insurance company. The models are the result of a close cooperation between academic resources and GN Asset Management. We have described the model, with an emphasis on the definition of risk, and the treatment of data. In particular, we have described the process from data collection to a scenario tree, so that the tree expresses the view of the decision-maker. A small case has been presented, and the track record discussed.

References

Bell, D., 1995. Risk, Return and utility. Management Science 41, 23–30.

Birge, J., Louveaux, F., 1997. Introduction to Stochastic Programming. Springer, Heidelberg.

Cariño, D.R., Ziemba, W.T., 1998. Formulation of the Russell–Yasuda financial planning model. Operations Research 46 (4), 443–449.

Chopra, V.K., Ziemba, W.T., 1993. The effect of errors in means, variances and covariances on optimal portfolio choice. Journal of Portfolio Management 19, 6–11.

Consigli, G., Dempster, M.A.H., 1998. Dynamic stochastic programming for asset–liability management. Annals of Operations Research 81, 131–161.

Consiglio, A., Cocco, F., Zenios, S.A., 2001. The value of integrative risk management for insurance products with guarantees. The Journal of Risk Finance, 6–16, Spring.

de Lange, P.E., Fleten, S.-E., Gaivoronski, A.A., 2004. Modeling financial reinsurance in the casualty insurance business via stochastic programming. Journal of Economic Dynamics and Control 28 (5), 991–1012.

Dert, C.L., 1995. Asset liability management for pension funds, a multistage chance constrained programming approach, PhD thesis. Erasmus University, Rotterdam, The Netherlands.

Dupacova, J., 2001. Output analysis for approximated stochastic programs. In: Uryasev, S., Pardalos, P.M. (Eds.), Stochastic Optimization: Algorithms and Applications. Kluwer, pp. 1–29.

Dupacova, J., Consigli, G., Wallace, S.W., 2000. Generating scenarios for multistage stochastic programs. Annals of Operations Research 100, 25–53.

Fischhoff, B., 1982. For those condemned to study the past: heuristics and biases in hindsight. In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, pp. 335–351, Chapter 23.

Gaivoronski, A.A., De Lange, P., 2000. An asset liability management model for casualty insurance companies: Complexity reduction versus parameterized decision rules. Annals of Operations Research 99, 227–250.

Gaivoronski, A.A., Høyland, K., De Lange, P., 2000. Statutory regulations of casualty insurance companies: An example from Norway with stochastic programming analysis. In: Uryasev, S., Pardalos, P.M. (Eds.), Stochastic Optimization: Algorithms and Applications, pp. 53–83.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2002. The Innovest Austrian Pension Fund Financial Planning Model InnoALM, Working paper. University of British Columbia.

Huber, G.P., 1991. Organizational learning: The contributing process and the literatures. Organizational Science 2 (1), 88–115.

Høyland, K., Kaut, M., Wallace, S.W., 2003. A heuristic for moment-matching scenario generation. Computational Optimization and Applications 24, 169–185.

Høyland, K., Wallace, S.W., 2001a. Generating scenario trees for multistage decision problems. Management Science 47, 295–307.

Høyland, K., Wallace, S.W., 2001b. Analyzing legal regulations in the Norwegian life insurance business using a multistage asset liability management model. European Journal of Operational Research 134, 293–308.

Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51 (5), 1611–1631.

Jorion, P., 2000. Risk management lessons from long-term capital management. European Financial Management 6, 277–300.

Kall, P., Wallace, S.W., 1994. Stochastic Programming. Wiley, Chichester.

Kallberg, J.G., Ziemba, W.T., 1983. Comparison of alternative utility functions in portfolio selection problems. Management Science XXIX, 1257–1276.

Kouwenberg, R., Zenios, S.A., 2006. Stochastic programming models for asset liability management. In: Handbook of Asset Liability Management, Vol. A: Theory and Methodology. In: Handbooks in Finance. North-Holland, pp. 253–303.

Malkiel, B.G., 2003. A Random Walk Down Wall Street, eighth ed. W.W. Norton & Company, New York.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

Markowitz, H., van Dijk, E., 2006. Risk-return analysis. In: Handbook of Asset Liability Management, Vol. A: Theory and Methodology. In: Handbooks in Finance. North-Holland.

Mulvey, J.M., 1996. Generating scenarios for the Towers Perrin investment system. Interfaces 26, 1–13.

Rahl, L. (Ed.), 2000. Risk Budgeting, A New Approach to Investing. Risk Books, London.

Chapter 14

DESIGN AND MANAGEMENT OF UNIT-LINKED LIFE INSURANCE CONTRACTS WITH GUARANTEES

RONALD HOCHREITER and GEORG PFLUG

Department of Statistics and Decision Support Systems, University of Vienna, Universitätsstraße 5/3, A-1010 Vienna, Austria
e-mails: ronald.hochreiter@univie.ac.at; georg.pflug@univie.ac.at

VOLKERT PAULSEN

Department of Mathematics, Christian-Albrechts-University Kiel, Ludewig-Meyn Straße 4, D-24098 Kiel, Germany
e-mail: paulsen@math.uni-kiel.de

Contents

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02014-X

Abstract

In this chapter, modeling and optimal management of unit-linked life insurance contracts with guarantee (ULLIG) is considered. The insurance premium inflow is invested to build a customer-specific portfolio. The minimal guarantee obliges the insurance company to pay either some pre-specified guaranteed sum or the actual accumulated portfolio value. Such contracts require a careful hedging against possible shortfalls. Both a discrete-time as well as a continuous-time version of this management problem is discussed. In the discrete-time case, the problem is formulated as a multi-stage stochastic optimization model. This specific modeling instance aims at maximizing expected portfolio value, and penalizes possible losses. Such classes of models can conveniently be utilized to assess shortfall risk and to optimally design new contracts. In the latter case, main parameters of the contract are determined, such that they meet the customers needs, and include constraints to meet regulatory as well as organizational constraints of the insurance company. Simulation studies can be applied to achieve an optimal contract design. The continuous-time modeling approach of optimal portfolio allocation in a guarantee setting with regular premium inflows and random benefit outflows provides further theoretical insights.

Keywords

unit-linked life insurance, financial guarantees, multi-stage stochastic programming, optimal investment strategies

JEL classification: C61, G11, G22

1. Introduction

1.1. Unit-linked life insurance contracts with guarantee (ULLIG)

Unit-linked life insurance contracts with guarantee (ULLIG) provide a combination of insurance and investment. By sacrificing some of the upside gain, the client obtains some guarantee to cover possible downside losses. The feature of a guarantee as well as the range of possible adoptions and specializations of underlying contracts to meet specific needs of costumers attracts a wide range of prospective and mostly risk averse policyholders.

Typical conditions of an ULLIG contract can be summarized as follows: The time horizon is split into discrete time-stages $t = 0, \dots, T$. The client pays a deposit b at the beginning ($t = 0$) of the contract and an annual premium B at the beginning of subsequent years $t = 1, \dots, T$. The premium may also be paid in monthly or quarterly installments. Throughout this chapter, yearly payments are assumed. Setting either b or B to zero, one gets the special cases of a regular annual payment, or a single installment, respectively. The premium inflow is split into an insurance part and an investment part. The latter part is used to build a customer-specific portfolio Y_t. This portfolio can be composed from various asset categories. In our investigation it consists of an (externally) managed reference fund as well as a zero-coupon bond. This portfolio is rebalanced at regular times.

If the client dies in year $t < T$, before the maturity date of the contract, her legal successors receive a death benefit D_t and the contract expires. If the client survives the maturity date, she gets a survival benefit S, which is most commonly the maximum of the actual portfolio value Y_T and some guaranteed sum G_T.

The death benefit as well as the guaranteed survival benefit may depend on the performance of the reference fund Z_t within the contract period. These values are determined by the death benefit formula D_t

$$D_t = D_t(b, B, Z_0, Z_1, \dots, Z_t), \tag{1}$$

and the minimal guarantee formula G_T

$$G_T = G(b, B, Z_0, \dots, Z_T). \tag{2}$$

Both formulas as well as the specification of the reference fund are an essential part of the insurance contract. As mentioned above, in most cases the survival benefit S is defined as the maximum of the underlying reference fund and the guarantee, i.e.,

$$S = \max(Y_T, G_T).$$

Examples for different survival, death, as well as minimal guarantee formulas are presented in Section 1.2 below.

The cash-flows originating from such a contract are shown in Figures 1 and 2. A bar in the upward direction symbolizes a positive cash-flow for the insurance company, a downward bar denotes a negative cash-flow.

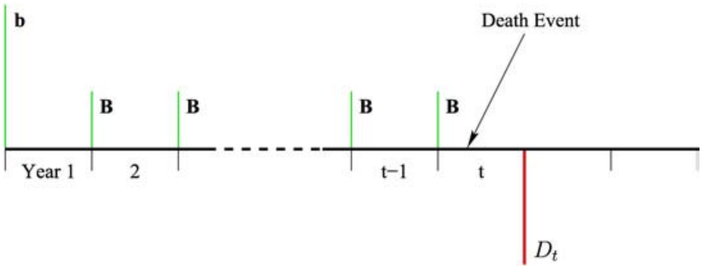


Fig. 1. Cash-flows, if customer dies in year t.

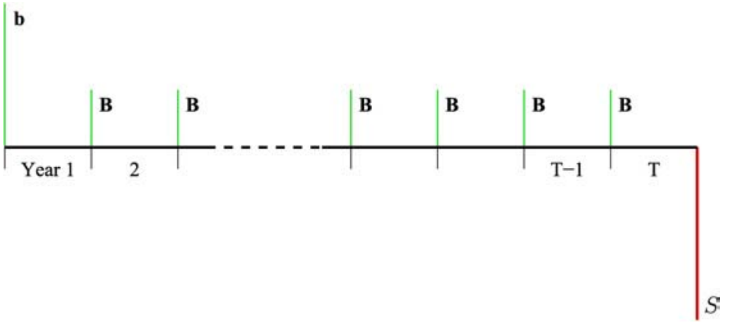


Fig. 2. Cash-flows, if customer survives year T.

In this chapter, the methodology of optimally managing ULLIG contracts on the customer level is discussed. The goal is to find an optimal allocation of the insurance part and the investment part, as well as the optimal composition of the investment portfolio. As a byproduct, the model allows for calculation of profit margins for the insurance company. Furthermore, parameter studies can be conducted and applied to optimally design new ULLIG products.

It should be noted that some insurance contracts permit to lapse the contract, and specify lapse times as well as surrender values. In this chapter, lapse options are not considered.

1.2. ULLIG contract specification

ULLIG contracts differ in the way, how death benefits and guarantees are calculated. Some examples for death benefit formulas are, e.g.,

Fixed death benefit. $D_t = f_1 \cdot b + f_2 \cdot B$, where f_1 and f_2 are factors, depending on age and gender of the customer.

Death benefit depending on total contribution. $D_t = b + B \cdot (\tau - 1)$, where τ is the number of stages until the death event.

Portfolio dependent death benefit. The benefit is the maximum of a fixed sum and the actual portfolio value.

Fund value dependent death benefit. The benefit is the maximum of fixed sum and a percentage of the fund value increase.

There are numerous ways to define guarantee formulas. Consider, e.g., a guaranteed annual increase and a guaranteed yield to maturity, i.e.,

Guaranteed annual increase.

$$G = b(1+g)^T + B \cdot \sum_{t=1}^{T-1} (1+g)^{T-t}.$$

Guaranteed yield to maturity.

$$G = f \cdot \left(b \max\bigl(Z_T/Z_0, (1+g)^T\bigr) + B \cdot \sum_{t=1}^{T-1} \max\bigl(Z_T/Z_t, (1+g)^{T-t}\bigr) \right),$$

where $0 < f < 1$ is a predefined factor (see, e.g., the SU 2001 example below).

A valuable example for an ULLIG contract with complex benefit formulas is the SU 2001 (Safe Unit 2001) contract, issued by the Italian company IntesaVITA. SU 2001 was placed from January 15th, 2001 to April 10th, 2001. The total volume was 215 million Euro. The conditions of this contract can be summarized as follows:

The client pays a fixed sum b (multiples of 2500 Euro) at the initial date and makes no more payments until maturity of the contract. SU 2001 is based on the fund SUG 2001, whose value is denoted by Z_t. The quota of fund ownership is defined as $Q = b \cdot 0.98/Z_0$, i.e., an underwriting fee of 2% is deducted at the beginning. The contract matures at time T. The minimal guarantee is defined as shown in Eq. (3).

$$G = Q \cdot \max\Bigl(\max_{0 \leqslant t \leqslant T_1} Z_t, 0.8 \cdot \max_{T_1 \leqslant t \leqslant T} Z_t \Bigr). \tag{3}$$

$T_1 < T$ is an intermediate observation date. The death benefit D_t at time t is defined by

$$D_t = Q\bigl(Z_t + \min(f \cdot Z_t, 10)\bigr), \tag{4}$$

where f is a factor which depends on the gender and age at time t of death the customer, as shown in Table 1.

1.3. Guarantee products and option pricing

In the beginning, pricing of guarantee-based financial products was accomplished in embedded American option pricing frameworks, see Brennan and Schwartz (1976), Boyle and Schwartz (1977), and Brennan and Schwartz (1979), or more recently Boyle and Hardy (1997, 2003), and Wilkie, Waters and Yang (2003).

Further applications of this approach can be found in Miltersen and Persson (1999), Grosen and Jørgensen (2000), Jensen, Jørgensen and Grosen (2001). Specific features

Table 1
Age- and gender-related factors f for the SU 2001 ULLIG contract

Age at time t	Male	Female
18–40	1.09	1.24
41–50	1.04	1.08
51–60	1.014	1.03
61–75	1.004	1.008
76–	1.0	1.0

of contracts due to legal peculiarities were described by, e.g., Bacinello (2001), Susinno and Giraldi (2000), Siglienti (2000) for the Italian market, and Chadburn (1997) for the UK market.

1.4. Pricing of contingent claims and stochastic optimization

Pricing of contingent claims and dynamic management of portfolios are two sides of the same coin. Since survival and death benefits are contingent on the tradable fund value in our guarantee setting, we may consider the problem as a pricing problem for contingent claims, which depends on both an underlying fund as well as mortality.

Suppose that a claim S_T contingent to a stock price Z_T has to be paid by some financial agent at time T. The pricing problem for S_T is given by calculating the minimal initial capital v needed to super-replicate S_T with a portfolio of stocks and bonds. A mathematical programming formulation is shown in model (5).

$$
\begin{aligned}
&\text{minimize}(x_s, y_s)\ v \\
&\text{subject to} \quad && x_0 + y_0 Z_0 \leqslant v, \\
& && x_{s-1}(1+r) + y_{s-1} Z_s \geqslant x_s + y_s Z_s, \quad 1 \leqslant s < T, \\
& && x_T + y_T Z_T \geqslant S_T.
\end{aligned}
\tag{5}
$$

In this formulation x_s denotes the amount invested in a risk-free asset (with return r) and y_s is the number of shares bought from the stock in year s.

The correct price of a contingent claim does not apply the principle: *Price equals the expected, discounted cash-flow*. Whenever replication (or super-replication) with pre-priced financial instruments is possible, the price has to be determined by solving an optimal management problem of type (5).

When the number of time steps between 0 and the maturity time T tends to infinity, and the process Z_s converges to the Geometric Brownian Motion in distribution, then the limit of the optimal values of optimization problems of type (5) is the well-known Black–Scholes formula (Black and Scholes, 1973).

If the initial capital v_0 is given and is larger than the minimal capital v required by model (5), the usual way of formulating an optimal management problem is to optimize the expected utility of excess over the claim, which is shown in model (6).

$$\begin{aligned}
&\text{maximize}(x_s, y_s) \ \mathrm{E}\big(U(x_T + y_T Z_T - S_T)\big) \\
&\text{subject to} \quad x_0 + y_0 Z_0 \leqslant v_0, \\
&\qquad x_{s-1}(1+r) + y_{s-1} Z_s \geqslant x_s + y_s Z_s, \quad 1 \leqslant s < T, \\
&\qquad x_T + y_T Z_T \geqslant S_T,
\end{aligned} \tag{6}$$

where U is an utility function. See Kallberg and Ziemba (1983) for a thorough discussion of utility functions.

In the discrete-time case, one can apply the stochastic programming approach. This approach provides a convenient modeling tool for solving long-term Asset Liability Management models, see Section 2.1 for more details. The optimization problem model formulations (5) and (6) above illustrate the link between discrete- and continuous-time modeling approaches. For more details, see King (2002) and King, Koivu and Pennanen (2005). In addition, Consiglio and De Giovanni (2006) provide a direct application of this stochastic programming contingent claim approach to American Option-based guaranteed product pricing.

Another chapter in this Volume (Consiglio, Cocco and Zenios (2006)) describes the PROMETEIA model for the Italian market. In this model, the initial capital, which consists of the initial equity capital plus premium payments of the customers, is invested in a portfolio, which consists of stocks and bonds. The initial portfolio composition is kept constant for the whole holding period. Single payment contracts are considered. Neither a premium inflow from customers to the company, nor a rebalancing due to bad performance takes place at later stages. Liabilities grow at market rate, but not slower than the minimal guaranteed rate. Lapse probabilities may depend on the market performance (variable lapse), mortality is considered through its expectation. If the funds are not sufficient, new equity capital has to be put into the company to cover liabilities. The objective is to maximize the Certainty Equivalent Excess Return on Equity (CEexROE) for shareholders using the Log Utility or Power Utility function. The excess return is defined as

$$\frac{\text{Assets} - \text{Liabilities}}{\text{Equity capital}}$$

at the terminal stage.

In contrast, the model presented in this chapter is based on the following design principles:

- It is based on the level of single contracts. Specific features of single contracts (age and gender of the client, single installment or periodic installments, ...) are taken into account.
- Mortality risk is a separate risk factor independent of market risk and is modeled by coupling the mortality event risk model with market risk model.
- A dynamic portfolio strategy is considered. The portfolio is rebalanced at predetermined stages.

- The portfolio decision, i.e., the allocation of assets, is extended by adding more complex structures. In the modeling instance below, a short term conventional life insurance based (re-)insurance is considered.

This chapter is organized as follows. Section 2 describes the discrete-time approach, which is inspired by the discrete-time structure of ULLIG contracts. Section 3 is devoted to a continuous-time modeling approach to provide further theoretical insights, and the link to the classical mathematical finance approach. Section 4 concludes this chapter.

2. Discrete-time modeling

2.1. Multi-stage stochastic programming

Let decisions in the multi-stage stochastic optimization model be taken at discrete time stages $t = 0, \dots, T$, whereby the root stage ($t = 0$) is chosen to be deterministic. The underlying multi-stage stochastic decision process is defined by a sequence of decisions: at each strategy switching stage $t > 0$ (in the future) the realization of a random variable ξ_t is observed and a decision x_t based on all observed values until stage t, i.e., $\xi_0, \dots, \xi_t$ is taken. At the terminal stage T a sequence of decisions $x = (x_0, \dots, x_T)$ with realizations $\xi = (\xi_0, \dots, \xi_T)$ leads to some cost. The goal is to find a sequence of decisions $x(\xi)$, which minimizes a functional F (commonly the expectation or some risk functional) of this stochastic cost function $f(x(\xi), \xi)$. A multi-stage stochastic optimization program can be notated as shown in program (7):

$$\begin{aligned} &\underset{x}{\text{minimize}}\ \mathrm{F}\big(f\big(x(\xi), \xi\big)\big) \\ &\text{subject to}\ \big(x(\xi), \xi\big) \in \mathcal{X}, \\ &\qquad\qquad x \in \mathcal{Z}. \end{aligned} \tag{7}$$

The main component of (7) is the (multi-dimensional) stochastic process ξ, which describes the future uncertainty. The stochastic model is commonly chosen independently from the optimization model. Furthermore, a constraint-set $\mathcal{X}$, which specifies all feasible (x, ξ) and captures organizational, regulatory, and physical constraints, has to be defined. Additionally, a set $\mathcal{Z}$ of functions $\xi \mapsto x$, such that x_t is based on realizations up to stage t $(\xi_0, \dots, \xi_t)$ only, is necessary. These special constraints $\mathcal{Z}$ are called non-anticipativity constraints.

See Ruszczyński and Shapiro (2003) for a recent overview of the area of stochastic programming, and Wallace and Ziemba (2005) for stochastic programming languages, environments, and applications.

Financial applications are often modeled using multi-stage stochastic programming techniques, because it is possible to incorporate multiple correlated sources of risk for both assets and liabilities, support long-term time horizons at discrete time stages, and accommodate risk aversion. A variety of risk measures is supported, and dynamic portfolio rebalancing while satisfying operational or regulatory restrictions and policy requirements is the main modeling strategy. Furthermore, an integration of realistic constraints, e.g., transaction costs, is possible without changing the underlying model completely. See Ziemba and Mulvey (1998), Ziemba (2003), as well as Kouwenberg and Zenios (2006) for further discussion on stochastic programming and financial engineering.

This technique was also successfully applied to handle guarantee-based products, see, e.g., Consiglio, Cocco and Zenios (2000, 2001). Applications to the UK market are discussed in Consiglio, Saunders and Zenios (2006) and to the Italian market in Consiglio, Saunders and Zenios (2003).

2.2. Modeling discrete-time ULLIG models

Pricing and management of ULLIG contracts follows the general pattern presented in the introduction. In this section, a management model for ULLIG contracts is developed, which consists of the following building parts:

- a mortality model,
- a stochastic fund value model,
- a stochastic interest rate model.

Mortality depends on gender and age of the customer. Let τ be the residual lifetime variable of the customer at the beginning of the contract. If death occurs in year t, then $\tau = t$.

Let $\pi_t^{(D)} = \mathrm{P}\{\tau = t\}$ denote death probabilities and $\pi_t^{(S)} = \mathrm{P}\{\tau > t\}$ survival probabilities. Both are conditional on the fact that the customer is alive at the beginning of the contract.

Death probabilities can be calculated using mortality tables, which are published by, e.g., governmental statistical agencies. Let $q_a^{(s)}$ be the yearly hazard rates in a mortality table, where a is the age of the customer at the beginning of the contract and s is the respective gender. Then $q_a^{(s)}$ and $\pi_t^{(D)}$, respectively, $\pi_t^{(S)}$ are related by

$$\pi_t^{(D)} = \bigl(1 - q_{a+1}^{(s)}\bigr)\bigl(1 - q_{a+2}^{(s)}\bigr) \cdots \bigl(1 - q_{a+t-1}^{(s)}\bigr) q_{a+t}^{(s)}, \qquad 1 \leqslant t \leqslant T,$$
$$\pi_t^{(S)} = \bigl(1 - q_{a+1}^{(s)}\bigr)\bigl(1 - q_{a+2}^{(s)}\bigr) \cdots \bigl(1 - q_{a+t-1}^{(s)}\bigr)\bigl(1 - q_{a+t}^{(s)}\bigr), \quad 1 \leqslant t \leqslant T.$$

The fund value model as well as the interest rate model are typically estimated using a balanced mixture of historical data and expert opinion. In the past years different classes of models have been developed to model financial time series. Among those are simple random walk models, ARMA models, GARCH models with all its variants, diffusion processes, jump-diffusion processes and specialized models, like the Wilkie model for actuarial use Wilkie (1986). The advantage of the multi-stage stochastic programming approach is, that there are no limitations for the type of model for both the fund process and the interest rate process. Once a model for Z_t is determined and estimated, contingent values for death benefits D_t and the minimal guarantee G_T can be calculated using the death benefit formula (1) as well as the minimal guarantee formula (2).

The portfolio management problem is given as follows. At every decision stage, the insurance company may decide to restructure the customers portfolio. The total capital at time t may be invested in three investment classes:

- reference fund Z_t,
- bonds accruing a random interest R_t,
- conventional life insurance contracts with a duration of the time between the respective stages, i.e., one year in our model. In case of death, this insurance pays a sum of α_t for each unit of premium in the subsequent stage.

The formula $\alpha_t = 0.95/q_{a+t-1}^{(s)}$, where a is the age and s is the gender of the customer at beginning of the contract, will be used for the insurance asset class in the example below. An extension to more complicated formulas is straightforward.

First, consider the cash-flow process of the company in the survival case. This process consists of income of size b at the beginning of the contract and of B every subsequent year, as well as a payment of S at the end of year T. The insurance company aims at building an optimal portfolio consisting of an amount x_t invested in bonds and y_t invested into the fund Z_t. An amount of w_t is invested into insurance. Following its portfolio strategy, suppose that the insurance company has built an individual portfolio value of Y_T. The objective is to maximize

$$\mathrm{E}\big([Y_T - G_T]^+ - \delta[Y_T - G_T]^-\big)$$

under the financing constraints for the portfolio. $\delta > 1$ denotes a penalty for shortfall. The optimization problem is shown in model (8):

$$\begin{aligned}
&\text{maximize}(x_s, y_s, w_s) && \mathrm{E}\big([Y_T - G_T]^+ - \delta[Y_T - G_T]^-\big)\\
&\text{subject to} && Y_0 = \gamma_1 \cdot b,\\
& && Y_0 = x_0 + y_0 Z_0 + w_0,\\
&\forall 1 \leqslant s < T: && Y_s = \big(x_{s-1}(1 + R_{s-1}) + y_{s-1} Z_s + B\big)\gamma_2,\\
&\forall 1 \leqslant s < T: && Y_s \geqslant x_s + y_s Z_s + w_s,\\
& && Y_T = x_{T-1}(1 + R_{T-1}) + y_{T-1} Z_T,\\
& && x_t, y_t, w_t \geqslant 0.
\end{aligned} \tag{8}$$

The constants $0 < \gamma_1, \gamma_2 \leqslant 1$ represent net factors, after deduction of management fees, taxes, etc. In model (8), the insurance part w_t is not considered and the optimal choice is $w_t = 0$ for all t.

A complete model considers death events as well. If the residual lifetime of the customer is $\tau = t$, then the payment is D_t at the end of year t and the whole contract process expires. A technical discount rate r is used to compare payments at different times. This extended model is given in (9):

$$\begin{aligned}
\text{maximize}(x_s, y_s, w_s) \sum_{t=1}^{T} (1+r)^{-t} \pi_t^{(D)} \mathrm{E}\big(&[Y_t + \alpha_t w_{t-1} - D_t]^+\\
&- \delta[Y_t + \alpha_t w_{t-1} - D_t]^- | \tau = t\big)\\
&+ (1+r)^{-T} \pi_T^{(S)} \mathrm{E}\big([Y_T - G_T]^+ - \delta[Y_T - G_T]^- | \tau > T\big)
\end{aligned}$$

$$
\begin{aligned}
\text{subject to} \quad & Y_0 = \gamma_1 \cdot b, \\
& Y_0 = x_0 + y_0 Z_0 + w_0, \\
\forall 1 \leqslant s < T: \quad & Y_s = \big(x_{s-1}(1 + R_{s-1}) + y_{s-1} Z_s + B\big)\gamma_2, \\
\forall 1 \leqslant s < T: \quad & Y_s \geqslant x_s + y_s Z_s + w_s, \\
& Y_T = x_{T-1}(1 + R_{T-1}) + y_{T-1} Z_T, \\
& x_t, y_t, w_t \geqslant 0.
\end{aligned} \tag{9}
$$

Alternatively, one could use the interest rate process R_t itself to discount future payments.

$$
\begin{aligned}
\text{maximize}(x_s, y_s, w_s) \quad & \sum_{t=1}^{T} \pi_t^{(D)} \mathrm{E}\big((1 + R_{t-1})^{-1} \cdots (1 + R_0)^{-1} \\
& \big([Y_t + \alpha_t w_{t-1} - D_t]^+ - \delta[Y_t + \alpha_t w_{t-1} - D_t]^-\big) | \tau = t\big) \\
& + (1 + r)^{-T} \pi_T^{(S)} \mathrm{E}\big((1 + R_{T-1})^{-1} \cdots (1 + R_0)^{-1} \\
& \big([Y_T - G_T]^+ - \delta[Y_T - G_T]^-\big) | \tau > T\big) \\
\text{subject to} \quad & Y_0 = \gamma_1 \cdot b, \\
& Y_0 = x_0 + y_0 Z_0 + w_0, \\
\forall 1 \leqslant s < T: \quad & Y_s = \big(x_{s-1}(1 + R_{s-1}) + y_{s-1} Z_s + B\big)\gamma_2, \\
\forall 1 \leqslant s < T: \quad & Y_s \geqslant x_s + y_s Z_s + w_s, \\
& Y_T = x_{T-1}(1 + R_{T-1}) + y_{T-1} Z_T, \\
& x_t, y_t, w_t \geqslant 0.
\end{aligned} \tag{10}
$$

Both model (9) and its variant (10) are linear optimization problems in the decision functions x_s, y_s, w_s. In order to solve them numerically, one has to use approximations in order to convert decision functions x_t, y_t, w_t into decision vectors. This is done by approximating the processes Z_t and R_t by tree processes, which may only take a finite number of values.

2.3. Multi-stage scenario tree models

Discrete-time, non-recombining trees constitute a simple but flexible model for approximating various stochastic processes. If the process is Markovian, a lattice model (recombining tree) may suffice for the representation. However, since decisions may depend on the whole history of observed financial processes, it is defined on the complete history tree, where the decision variables have to be placed. Hence, we assume that a tree was estimated for the financial processes.

Moreover, trees can be used to represent processes, which are not first order Markovian, for example, higher-order Markovian or non-Markovian processes. We assume that

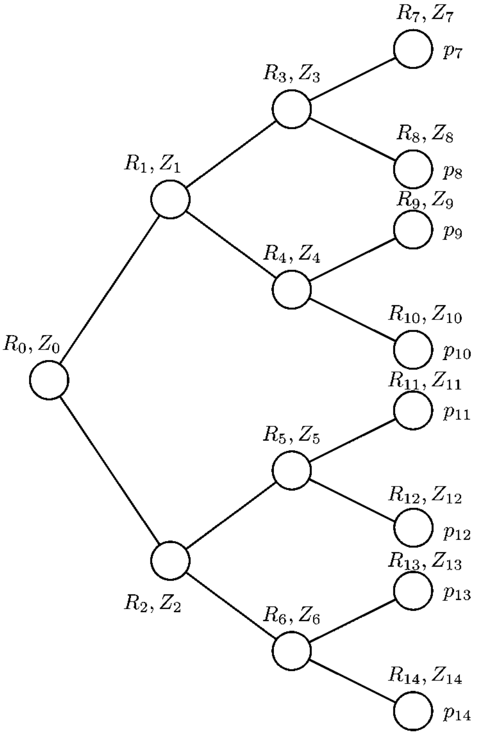


Fig. 3. Tree of height 3, representing the processes R_t, Z_t.

lifetime is independent of the finance processes (R_t, Z_t). However, a dependency between the fund process (Z_t) and the interest rate process (R_t) is allowed, as shown in Figure 3.

First, a tree process representing fund values and interest rates is estimated using historical data and expert projections. A more detailed description of how to get from data to trees is contained in Section 2.4.

The mortality process is a binary process consisting of survival nodes, which branch to the next stage and death nodes, which are terminal nodes. Path probabilities are given by $\pi_1^{(D)}, \dots, \pi_T^{(D)}$ for the death paths and $\pi_T^{(S)}$ for the survival path.

Since statistical independence of the financial processes and the mortality process is assumed, both trees can be combined by constructing the product tree, which is called the tensor product of trees.

The size of the product tree clearly depends on the tree structure of the value tree. Consider the value tree and let n_t be the total number of nodes in stage t and N_t be the number of total nodes up to stage t, i.e., $\sum_{i=1}^{t} n_i$ (without root node). The total number of nodes of the product tree is shown in Table 2. E.g., coupling a binary value

Table 2
Number of nodes in the value, event and product tree

Stage	0	1	2	...	$T-1$	T
Value tree	1	n_1	n_2		n_{T-1}	n_T
Event tree	1	2	3		$T-1$	$T-1$
Product tree	1	N_1+n_1	N_2+n_2		$N_{T-1}+n_{T-1}$	$N_{T-1}+n_T=N_T$

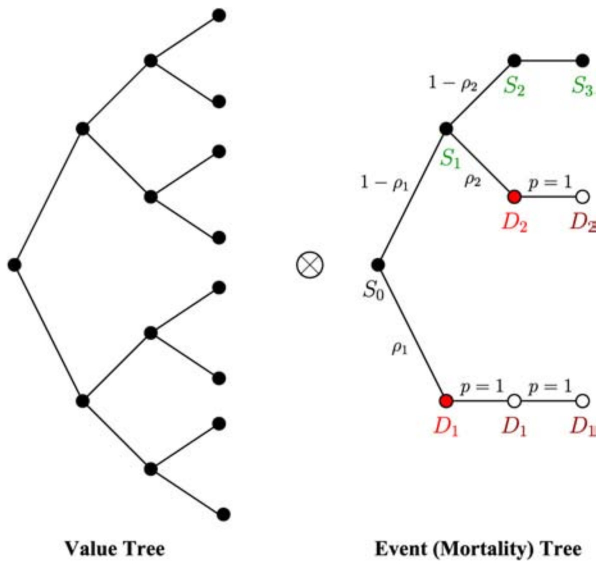


Fig. 4. Building a tensor tree of value (market) and mortality (event) tree.

tree of height $T=3$ with a mortality tree of the same height (see Figure 4) results in a $n=(1,4,10,14)$ product tree, as shown in Figure 5.

Let $\mathcal{N}$ be the set of the nodes of the tensor product tree (except the root), let $\mathcal{N}_t = \mathcal{N}_t^s \cup \mathcal{N}_t^d$ be the set of all nodes at time t, where $\mathcal{N}_t^s$ is the set of all survival nodes and $\mathcal{N}_t^d$ is the set of all death nodes. Let $\mathcal{T} = \mathcal{N}_T$ the set of all terminal nodes and $\mathcal{T}^s$ the subset of all terminal survival nodes. If n is a node, then $n-$ denotes the predecessor of n (this excludes the root node). The set of successors of n is denoted by $n+$. The probability to reach node n is p_n. $t(n)$ denotes the stage of node n.

Both the fund values and the interest rate process are defined on the value tree, denoted by $Z_n, n \in \mathcal{N}$, and $R_n, n \in \mathcal{N}$, respectively. Using the contracted formulas for the death benefit (1) and the minimal guarantee (2), one may calculate death benefits D_n for all death nodes and guarantees G_n for the terminal survival nodes.

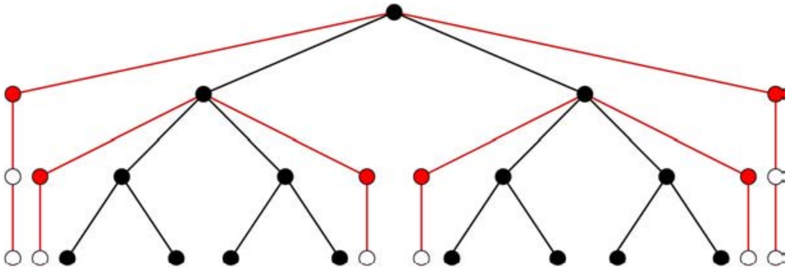


Fig. 5. The coupled product tree.

The optimization problem (9) may now be rewritten in a tree formulation as shown in model (11).

$$\begin{aligned}
&\text{maximize}(x_n, y_n, w_n) && \sum_{t=1}^{T}(1+r)^{-t}\pi_t^{(D)} \sum_{n\in\mathcal{N}_t^d} p_n\big([Y_n+\alpha_t w_n - D_n]^+ \\
& && \quad - \delta[Y_n+\alpha_t w_n - D_n]^-\big) \\
& && \quad + (1+r)^{-T}\pi_T^{(S)} \sum_{n\in\mathcal{T}^s} p_n\big([Y_n - G_n]^+ - \delta[Y_n - G_n]^-\big) \\
&\text{subject to} && Y_0 = x_0 + y_0 Z_0 \leqslant \gamma_1 \cdot b, \\
&\forall n \in \mathcal{N}\setminus\mathcal{T}: && \big(x_{n-}(1+R_{n-}) + y_{n-}Z_n + B\big)\gamma_2 \geqslant x_n + y_n Z_n + w_n, \\
&\forall n \in \mathcal{T}^s: && Y_n = x_{n-}(1+R_{n-}) + y_{n-}Z_n, \\
& && x_n, y_n, w_n \geqslant 0.
\end{aligned} \tag{11}$$

The set of inputs consists of characteristic features of the contract (survival benefit, death benefit, guarantee, stochastic fund model, ...). Solving this model computes the optimal risk management strategy (x_n, y_n, w_n) for the specific contract. In addition, the probability of shortfall, i.e., the probability that $Y_n + \alpha_{t(n)} < D_n$ or $Y_n < G_n$ is also obtained. The primary use of the multi-stage stochastic optimization model (11) is to transform characteristics of the contract into characteristics of the optimal management and associated risk. The model may also be used for the design of new products. Choosing a constraint for risk exposure (expressed, e.g., in terms of the shortfall probability for instance), the maximal guarantee which does not lead to a violation of the risk constraint may be calculated.

The model can be extended easily. A variant of this model would include a lapse model. In that case, it is slightly more difficult to integrate lapse events. Empirical evidence shows that the frequency of lapses depends on the performance of the fund and on the credit rating of the insurance company (see Consiglio, Cocco and Zenios (2000)). If lapse occurs, surrender values are calculated using common formulas employed by insurance companies. One way to extend the model above is to include an additional dependent risk factor into the tree. This would enlarge the tree, but would allow for more flexibility.

2.4. Numerical example

The multi-stage stochastic optimization models (8) and (11) were modeled with the AMPL modeling language (Fourer, Gay and Kernighan (2002)) and solved using the MOSEK solver. Scenario simulation, and optimization, as well as the tree coupling methods described above have been implemented using the MatLab programming environment.

Coupling a value tree and a mortality tree might result in huge (tensor product) scenario trees, even when only a moderate number of stages and number of succeeding nodes at each node of the tree is considered, see Table 2. Hence, the scenario generation methodology for generating the value tree must be chosen carefully. The multi-stage scenario generator used for calculating the nodes, is based on a stage-wise fixed number of nodes, i.e., the number of nodes per stage can be specified in advance. The scenario generator calculates the optimal values of the respective number of nodes per stage as well as the optimal links between stages, and finally assigns correct (optimal) probabilities to these arcs. The scenario generation is based on a minimization of probability metrics. See Pflug (2001) for theoretical details and Hochreiter and Pflug (2006) for an implementation and further numerical studies.

The multi-stage stochastic ULLIG model offers a rich set of parameters, which can be used to achieve an optimal adoption of the model to the needs of the insurance company, i.e., the integration of contract specific details.

To represent the future development of the underlying fund, the Standard and Poors 500 Index was taken as a reference for the scenario estimation, simulation and generation. Daily closing values of 8 years (January 1996 to January 2004) have been used to fit an ARMA(1, 1)/GJR(1, 1) time series model, from which 1000 paths have been simulated describing fund development of the next 5 years. It is obvious, that the choice of the underlying fund development process has a significant impact on the results. For practical purposes, the selection and calibration of the underlying model is one of the most important steps in the optimal management process.

A scenario tree with a stage-wise fixed structure (25, 50, 75, 100, 200 nodes per stage) was generated. This scenario tree as well as the probabilities of each scenario in the final stage are shown in Figure 6.

Austrian mortality tables from the years 2000/2001 have been used to calculate survival and death probabilities for different age and gender classes. Table 3 summarizes death probabilities for selected age and gender classes, which are used below. If long-term models are considered, it is noteworthy, that the use of cohort specific projected mortality tables should be considered.

The investment problem is to decide on the investment into three asset categories described above: the underlying fund, a risk-free bond, and re-insurance. The time horizon is $T = 5$ yearly stages, and there is a single installment $B = 1000$ at the beginning of the contract.

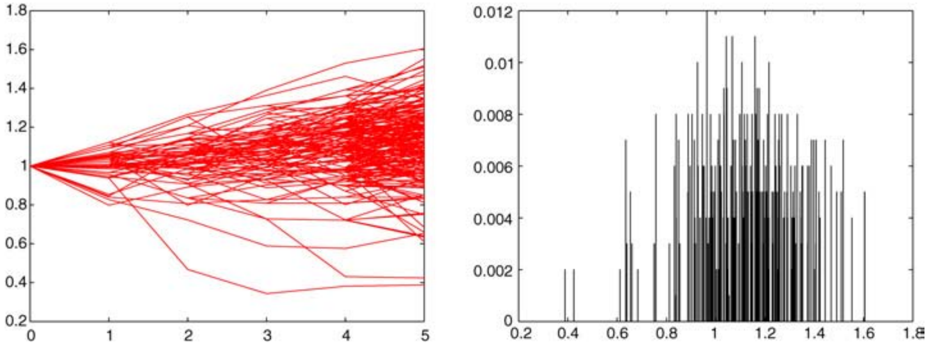


Fig. 6. Scenario tree: Values (left), probabilities of scenarios in final stage (right).

Table 3
Example: Death probabilities $q_{x+t}^{(s)}$ $(\cdot 10^{-3})$ for five subsequent years

Age	Gender	Year 1	2	3	4	5
30	Female	0.29	0.37	0.37	0.47	0.53
30	Male	0.88	0.81	1	0.99	1.12
50	Female	2.60	2.75	3.15	3.31	3.49
50	Male	4.73	5.59	6.04	6.66	7.32

The underlying fund tree and one possible wealth development is shown in Figure 7. In this case, the calculation was done for a 30-year old woman, with a fixed Euro 80 000 death benefit, an annual guaranteed survival benefit of 2% on the initial installment. A risk-free rate of 4% per year was used.

Out of the variety of possible parameter studies with the ULLIG management model, two selected examples are outlined below.

The first example shows the shortfall probability, as well as the expected shortfall over a range of guaranteed annual survival benefit rates. This visualization supports a decision, which level of survival benefit rate should be granted. The second example suggests the optimal amount of wealth invested in the re-insurance asset per stage.

Shortfall probability and expected shortfall. Figure 8 depicts the shortfall probability of a 30-year old woman (left) and a 50-year old man (right). A fixed Euro 80 000 death benefit was used. Figure 9 shows the expected shortfall in these two cases. The calculations have been conducted for 4 different risk free rates

$$r = 1.04, 1.06, 1.07, 1.08$$

over a range of guaranteed annual survival benefit rates

$$s = 1.02, 1.025, 1.03, 1.035, 1.04, 1.045, 1.05.$$

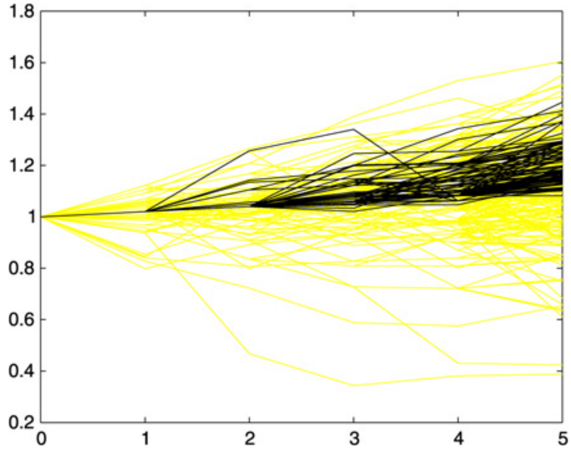


Fig. 7. Example: Underlying fund tree and wealth development.

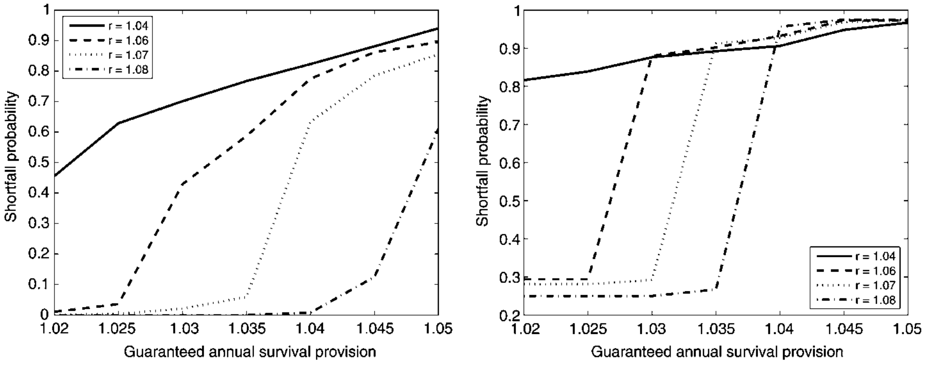


Fig. 8. Shortfall probability for different levels of guaranteed survival benefit rates.

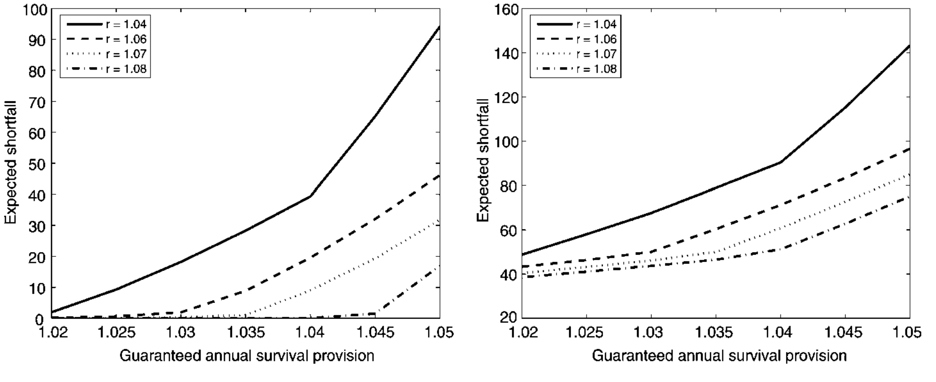


Fig. 9. Expected shortfall for different levels of guaranteed survival benefit rates.

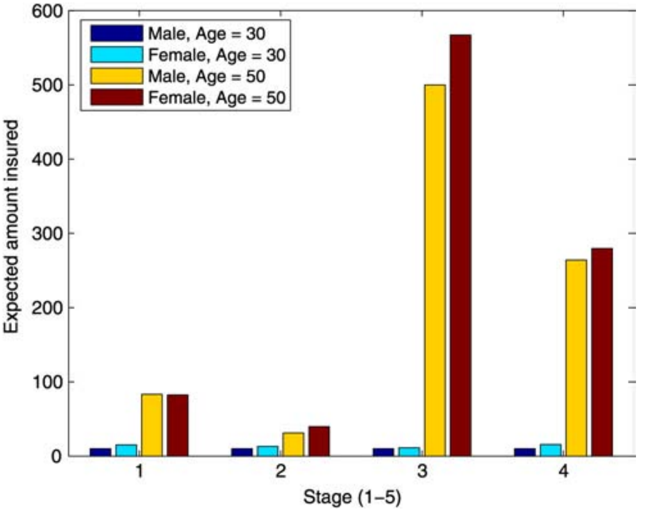


Fig. 10. Insurance per stage.

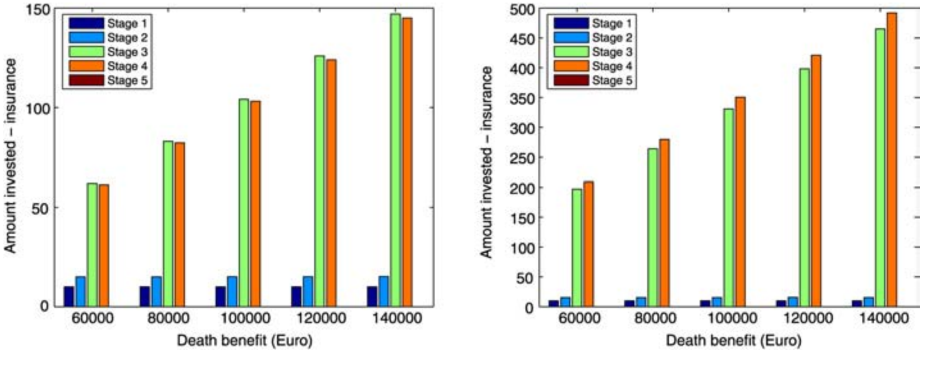


Fig. 11. Insurance per stage with different guaranteed death benefits.

Amount of insurance per stage. Figure 10 displays the amount invested into the conventional life insurance per stage for four person classes (Female/Male, Age 30/50). Nothing will be invested in the last stage, as the current model is designed to pay the survival benefit, even if the client dies in the last stage. As above a fixed death benefit of Euro 80 000, and an annual guaranteed survival benefit of 2 percent on the initial installment was assumed. The risk-free rate was set to 3 percent per year.

Figure 11 summarizes the effects of higher guaranteed death benefits on the amount invested in insurance. The two examples were calculated for a 30-year old man (left) and a 50-year old woman (right) with the same assumptions on the survival benefit and the risk-free rate as above.

3. Continuous-time modeling

In this section, we present a continuous time variant of our decision model and derive some of its properties on a pure analytical basis. We assume that the insurance company does investment and trading on a continuous time basis, while the premium inflow as well as the outflow of death as well as survival benefits occurs only at discrete times $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = T$.

To allow for slightly more flexibility, we allow that these fixed dates may also be fractions of years. Thus the customer pays an initial installment b and at times t_i; $i = 1, \dots, n$, the premium B. In case of death in the interval $(t_i, t_{i+1}]$, the death benefit D is paid. In case of survival to maturity T, the survival benefit S is paid, and this sum is guaranteed. As before, denote by τ the residual lifetime variable of the customer.

The fundamental difference between the continuous time model and the discrete model introduced in Section 2 is that the guarantee—given that it is feasible by pure bond investments—may be reached with probability 1. Therefore, the objective in this section is no longer to penalize the shortfall, but to maximize the utility of the surplus, under the constraint that the shortfall is zero. In case of a linear utility, the previous discrete model is just a penalty cost formulation of the continuous time model.

3.1. *Market model*

We suppose that the market consists of $d + 1$ assets. One asset is a standard bond and has no systematic risk. Its price process is $(\beta(t))_{0 \leqslant t \leqslant T}$. The other d assets are risky and their prices are

$$Z(t) = \big(Z_1(t), \dots, Z_d(t)\big), \quad t \in [0, T].$$

We assume a Black–Scholes (Geometric Brownian Motion) model with random coefficients. Thus we have the dynamics

$$\mathrm{d}\beta(t) = \beta(t) r(t)\,\mathrm{d}t, \quad \beta(0) = 1$$

for the bond process and

$$\mathrm{d}Z_i(t) = Z_i(t)\left[\mu_i(t)\,\mathrm{d}t + \sum_{j=1}^{d} \sigma_{i,j}(t)\,\mathrm{d}W_j(t)\right], \quad Z_i(0) = z_i \in (0, \infty), \tag{12}$$

$i = 1, \dots, d$, for the equities. Here $W(t)$ is a d-dimensional Wiener process $W(t) = (W_1(t), \dots, W_d(t))$, $0 \leqslant t \leqslant T$, which generates a complete filtered probability space $(\Omega, (\mathcal{F}_t)_{0\leqslant t\leqslant T}, \mathrm{P})$ after augmentation. The interest rate $(r(t))_{0\leqslant t\leqslant T}$, drift vector $\mu(t) = (\mu_1(t), \dots, \mu_d(t))$ and volatility matrix processes $\sigma_{i,j}(t)_{1\leqslant i,j\leqslant d}$, $0 \leqslant t \leqslant T$, are progressively measurable with respect to the filtration $(\mathcal{F}_t)_{0\leqslant t\leqslant T}$ and satisfy the mild integrability condition

$$\int_0^T \left(\left|r(t)\right| + \left\|\mu(t)\right\| + \left\|\sigma(t)\right\|^2\right) \mathrm{d}t < \infty, \quad \text{a.s.} \tag{13}$$

For an arbitrage-free model of a complete financial market we assume that the volatility matrix $\sigma(t)$ is invertible at each $0 \leqslant t \leqslant T$. Furthermore the *market price of risk* process

$$\theta(t) = \sigma(t)^{-1}\big(\mu(t) - r(t)\mathbb{I}\big), \quad 0 \leqslant t \leqslant T,\ \mathbb{I} = (1, \dots, 1), \tag{14}$$

is supposed to fulfill

$$\int_0^T \big\|\theta(t)\big\|^2 \,\mathrm{d}t < \infty, \tag{15}$$

and

$$\mathrm{E}\left[\exp\left(-\int_0^T \theta(t)\,\mathrm{d}W_t - \frac{1}{2}\int_0^T \big\|\theta(t)\big\|^2 \,\mathrm{d}t\right)\right] = 1. \tag{16}$$

For this market model the process

$$L(t) = \exp\left(-\int_0^t \theta(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t \big\|\theta(s)\big\|^2 \,\mathrm{d}s\right), \quad 0 \leqslant t \leqslant T, \tag{17}$$

is a martingale and the unique martingale measure P^* equivalent to P can be defined by

$$\mathrm{P}^*(A) = \mathrm{E}L(T)\mathbb{I}_A, \quad A \in \mathcal{F}_T. \tag{18}$$

Applying Girsanov's Theorem provides that

$$\overline{W}(t) = W(t) + \int_0^t \theta(s)\,\mathrm{d}s, \quad 0 \leqslant t \leqslant T \tag{19}$$

is a Wiener process with respect to P^* and the stock price processes fulfill with respect to $\overline{W}$ the dynamics

$$\mathrm{d}Z_i(t) = Z_i(t)\left[r(t)\,\mathrm{d}t + \sum_{j=1}^{d} \sigma_{i,j}(t)\,\mathrm{d}\overline{W}_j(t)\right], \quad i = 1, \dots, d. \tag{20}$$

The martingale measure P^* can be used for pricing contingent t-claims, which are $\mathcal{F}_t$ measurable random variables S payable at time t. The unique arbitrage-free initial price $V_0(S)$ of such a contract in our complete financial market is the expected discounted payoff with respect to the martingale measure P^*:

$$V_0(S) = \mathrm{E}^*\beta(t)^{-1}S. \tag{21}$$

Trading strategies are progressively measurable R^d-valued processes

$$\xi(t) = \big(\xi_1(t), \dots, \xi_d(t)\big), \quad 0 \leqslant t \leqslant T,$$

with

$$\int_0^T \big\|\xi(t)\sigma(t)\big\|^2 \,\mathrm{d}t + \int_0^T \big|\xi(t)^\top\big(\mu(t) - r(t)\mathbb{I}\big)\big| \,\mathrm{d}t < \infty, \quad \text{a.s.} \tag{22}$$

The real number $\xi_i(t)$ denotes the amount of money the investor holds in the ith stock at time t. If we denote by $Y^{b,\xi}(t)$ the wealth of the investor at t, assuming that the initial budget was b and the trading strategy ξ is followed, then $Y(t) - \sum_{i=1}^{d} \xi_i(t)$ is the amount invested in bonds. If no additional investment or consumption is allowed we are in a self-financing setting and the wealth process $Y^{b,\xi}(t)$ satisfies the dynamics

$$\begin{aligned} \mathrm{d}Y^{b,\xi}(t) &= \left(Y^{b,\xi}(t) - \xi^\top(t)\mathbb{I}\right)r(t)\,\mathrm{d}t + \xi(t)'\left[\mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t)\right], \\ Y^{b,\xi}(0) &= b. \end{aligned} \tag{23}$$

A solution of the above equation for the discounted wealth can be given in the following form

$$\beta(t)^{-1}Y^{b,\xi}(t) = b + \int_0^t \beta(s)^{-1}\xi(s)^\top\left[\sigma(s)\,\mathrm{d}W(s) + \left(\mu(s) - r(s)\mathbb{I}\right)\mathrm{d}s\right] \tag{24}$$

for $0 \leqslant t \leqslant T$.

We emphasize that each contingent t-claim S can be replicated from initial wealth $b = V_0(S)$ with a self-financing trading strategy ξ followed up to time t, i.e., the associated wealth process satisfies

$$Y^{V_0(S),\xi}(t) = S, \tag{25}$$

i.e., coincides with S at maturity t.

The optimal management problem may now formulated as follows: The insurance company receives initial premiums and regular installments B, given that the customer is alive. On the other hand, it has to pay the death benefit in case of death and the survival benefit in case of survival. Such strategies are no longer self-financing strategies, but allow money inflow until a stopping criterion (the death event or the maturity) is met. In addition, some constraints may be put on the set of feasible strategies.

The goal is to maximize the expected final utility under the given constraints. In general, this problem is a dynamic optimization problem. However, the powerful martingale method suited for the Black–Scholes model allows to reduce the problem to a static variational problem. Given the solution of the static problem, the optimal strategy may be derived in a second step.

We review this technique briefly here. The books of Karatzas (1997) or Korn (1997) serve as a standard reference.

3.2. Maximizing expected utility

Let U be strictly increasing, continuously differentiable and concave utility function defined on the positive reals fulfilling the Inada condition (see Inada (1963))

$$\lim_{v\to\infty} U'(v) = 0, \qquad \lim_{v\to 0} U'(v) = +\infty. \tag{26}$$

The most important examples are $U(v) = \log(v)$ and $U(v) = \frac{1}{\alpha}v^\alpha$, $\alpha \in (0, 1)$.

The problem of an investor who would like to maximize the expected utility of terminal wealth by investing in a self-financing trading strategy ξ and who uses initial capital $b > 0$, which will never be negative is called Merton problem, i.e.,

$$\max_{\xi \in \mathcal{A}(b)} U\big(Y^{b,\xi}(T)\big), \tag{27}$$

where $\mathcal{A}(b)$ the set of all self-financing trading strategies ξ such that its associated wealth process $Y^{b,\xi}(t)$ never becomes negative and satisfies

$$\mathrm{E}U\big(Y^{b,\xi}(T)\big)^{-} < \infty.$$

This dynamic optimization problem was firstly solved by Merton (1971) using standard tools of control theory. The Martingale Method was introduced, e.g., by Pliska (1986), Cox and Huang (1989), and Karatzas et al. (1991). We would like to recall the main ideas in the following. Because of the fundamental relationship between strategies and correctly prices contingent claims (25), every attainable terminal wealth from an initial capital not exceeding b can be seen as a contingent T-claim with initial price not larger than b and vice versa. To determine an optimal terminal wealth, one has to look for (A) an optimal T-claim financeable with an initial capital not exceeding b and (B) a trading strategy replicating this optimal claim.

The static problem (A) can be easily solved by a pointwise Lagrange approach. Denote by $\mathcal{B}(b)$ the set of contingent non-negative T-claims S such that $\mathrm{E}U(S)^{-} < \infty$ and $V_0(S) \leqslant b$. Then the static problem is given by

$$\max_{S \in \mathcal{B}(b)} \mathrm{E}U(S). \tag{28}$$

To solve this problem by a Lagrange approach, it is noteworthy that the constraint can be expressed as

$$V_0(S) = \mathrm{E}^{*}\beta(T)^{-1}S = \mathrm{E}\beta(T)^{-1}L(T)S = \mathrm{E}H(T)S \tag{29}$$

with $H(t) = \beta(t)^{-1}L(t)$, $0 \leqslant t \leqslant T$, denoting the discounted martingale process. Introducing a Lagrange multiplier λ, the pertaining unconstrained static problem is

$$\max_{S \geqslant 0} \mathrm{E}\big(U(S) - \lambda H(T)S\big). \tag{30}$$

If the utility function satisfies the Inada conditions (26), the solution is given by

$$S = I\big(\lambda H(T)\big), \tag{31}$$

where I is the inverse of U'.

To determine finally a solution to (28), we calculate a Lagrange multiplier λ satisfying the constraint, hence

$$\mathrm{E}H(T)I\big(\lambda H(T)\big) = b. \tag{32}$$

If U' does not vanish at infinity, then there is no finite solution, since $H(T)$ is unbounded. This fact distinguishes the continuous-time Geometric Brownian Motion model from the discrete-time model discussed in the previous section. In fact, maximizing expected wealth is a reasonable objective in discrete models, but not in the Black–Scholes situation.

The above mentioned arguments can be made rigorous and we refer to Karatzas (1997, Theorem 2.3.2):

Theorem 3.2.1. *In order to solve the dynamic problem (27) it is sufficient to solve the static problem (28) and then find the strategy which replicates this solution. Suppose the static problem (28) has a finite optimal value for each initial capital $b > 0$, then the optimal attainable terminal wealth is given by (31) with λ the unique solution of (32).*

The second step (B) is to determine the optimal trading strategy ξ that replicates the optimal terminal wealth $I(\lambda H(T))$. Note that its associated wealth process $Y(t), 0 \leqslant t \leqslant T$, fulfills

$$Y(t) = \frac{1}{H(t)} \mathrm{E}\big[H(T) I\big(\lambda H(T)\big) \big| \mathcal{F}_t\big], \quad 0 \leqslant t \leqslant T. \tag{33}$$

Due the martingale representation theorem there exists a progressively measurable process ψ such that

$$H(t)Y(t) = b + \int_0^T \psi(s)\,\mathrm{d}W_s, \quad 0 \leqslant t \leqslant T. \tag{34}$$

Together with (24) and integration by parts, we obtain the optimal trading strategy ξ by solving

$$\sigma'(t)\xi(t) = \frac{\psi(t)}{H(t)} + Y(t)\theta(t), \quad 0 \leqslant t \leqslant T. \tag{35}$$

For $U(v) = \log(v)$ the previous approach leads to the optimal terminal wealth

$$S = \frac{b}{H(T)}. \tag{36}$$

Its associated wealth process fulfills $Y(t) = \frac{b}{H(t)}$, $0 \leqslant t \leqslant T$, and can be obtained by using the Merton strategy, i.e.,

$$\xi(t) = \big(\sigma(t)\sigma^\top(t)\big)^{-1}\big(\mu(t) - r(t)\mathbb{I}\big)Y(t), \quad 0 \leqslant t \leqslant T, \tag{37}$$

which is an optimal trading strategy. For power utility, an explicit solution can be obtained in a model of deterministic coefficients, see Karatzas (1997, Example 2.2.5) for further details.

3.3. *Guarantee constraints*

For guaranteed products, we now add the constraint that the terminal wealth must exceed an initially set benchmark $G > 0$. For simplicity we assume that G is a positive constant but our arguments will also work for a contingent T-claim. We only need to adapt the previous approach to obtain a solution.

Let $\mathcal{A}_1(b)$ be the set of trading strategies $\xi \in \mathcal{A}(b)$ such that its terminal wealth a.s. exceeds G and $\mathcal{B}_1(b)$ the set of those contingent claims $S \in \mathcal{B}(b)$ such that $S \geqslant G$ almost surely. Then the optimization problem with guaranteed payoff G at terminal time T is

$$\max_{\xi \in \mathcal{A}_1(b)} \mathrm{E}U\big(Y^{b,\xi}(T)\big). \tag{38}$$

Its static counterpart is

$$\max_{S \in \mathcal{B}_1(b)} \mathrm{E}U(S). \tag{39}$$

To ensure a guaranteed payoff G at terminal time at least an initial capital $G \cdot P(0, T)$ is necessary. Here we denote with $P(t, T)$ the price at t of a zero coupon bond which pays 1 at maturity T. We have that

$$P(t, T) = \mathrm{E}^*\left(\frac{\beta_t}{\beta_T}\middle|\mathcal{F}_t\right). \tag{40}$$

We assume $b > G \cdot P(0, T)$ for our initial capital b and a pointwise Lagrange maximization yields an optimal terminal wealth $S = S(\lambda)$ given by

$$S(\lambda) = \max\big\{G, I\big(\lambda H(T)\big)\big\}. \tag{41}$$

The Lagrange multiplier λ can be computed by solving

$$V_0\big(S(\lambda)\big) = G \cdot P(0, T) + \mathcal{C}\big(I\big(\lambda H(T)\big), S\big) = b \tag{42}$$

with $\mathcal{C}(I(\lambda H(T)), S)$ denoting the initial price of a call with strike G on an asset with payoff $I(\lambda H(T))$ at T. Since the call price tends to ∞ for $\lambda \to 0$, respectively, 0 for $\lambda \to \infty$ the above equation can be solved for each $b > G \cdot P(0, T)$ and we get an analogous statement to Theorem 3.2.1.

Theorem 3.3.1. *Let the assumptions of Theorem 3.2.1 be fulfilled and let $b > G \cdot P(0, T)$. Then a solution of (39) is given by (41) with λ being the unique solution of (42). The optimal trading strategy, a solution of (38), can be determined by*

(1) *buying G zero coupon bonds with maturity T for a price of $P(0, T)$ each,*
(2) *investing the remaining initial capital of $b - G \cdot P(0, T)$ in a strategy that replicates a call on $I(\lambda H(T))$, which is the optimal wealth in the unconstrained portfolio optimization problem with respect to a modified initial capital.*

Consider the Black–Scholes model for a bond with constant interest rate r, one stock with constant coefficients (μ, σ) and log-utility. In this model, the Black–Scholes call price has volatility $|\theta| = |\frac{\mu-r}{\sigma}|$. Thus Eq. (42) for a $G > 0$ such that $Ge^{-rT} < b$ is

$$Ge^{-rT} + \mathcal{C}\left(\frac{1}{\lambda}H(T), S\right) = b, \tag{43}$$

where

$$\mathcal{C}\left(\frac{1}{\lambda}H(T), S\right) = \frac{1}{\lambda}\phi\left(g_1\left(\frac{1}{\lambda}, T\right)\right) - S\mathrm{e}^{-rT}\phi\left(g_2\left(\frac{1}{\lambda}, T\right)\right) \tag{44}$$

with

$$g_1(y, T) = \frac{\log(\frac{y}{K}) + (r + \frac{1}{2}\theta^2)T}{\sqrt{\theta^2 T}}, \tag{45}$$

$$g_2(y, T) = \frac{\log(\frac{y}{K}) + (r - \frac{1}{2}\theta^2)T}{\sqrt{\theta^2 T}}. \tag{46}$$

We may solve (43) to obtain the optimal expected utility of terminal wealth

$$w(\lambda) = \mathrm{E}\log\left(\max\left\{G, \frac{1}{\lambda}H(T)\right\}\right).$$

A straightforward calculation yields

$$w(\lambda) = \log\frac{1}{\lambda} + \log(\lambda G)\Phi\left(h_1(\lambda)\right) + \left(r + \frac{1}{2}\theta^2\right)T\Phi\left(h_2(\lambda)\right) + |\theta|\sqrt{T}\varphi\left(h_1(\lambda)\right) \tag{47}$$

with functions

$$h_1(\lambda) = \frac{\log(\lambda G) - (r + \frac{1}{2}\theta^2)T}{\sqrt{\theta^2 T}}, \tag{48}$$

$$h_2(\lambda) = \frac{\log(\lambda G) + (r + \frac{1}{2}\theta^2)T}{\sqrt{\theta^2 T}}, \tag{49}$$

where φ denotes the density, and Φ the distribution function of a standard normal distribution. The optimal expected return is

$$\overline{R}(\lambda) = \frac{1}{T}\left(w(\lambda) - \log b\right) \tag{50}$$

with λ depending on the initial capital $G\mathrm{e}^{-rT}$ needed to ensure the terminal wealth guarantee.

We plotted the above defined function $\overline{R}$ when $r = \log(1 + 0.03)$, initial capital $b = 1000$, and $|\theta| = 0.1$, respectively, $|\theta| = 0.3$ in Figure 12. The more money we need to reserve for the terminal wealth guarantee, the less is the optimal expected return. This effect is more intensive for higher $|\theta|$. At the limit we obtain a return from a pure Bond, i.e., the Merton strategy. In the plot we included bounds on $\overline{R}(\lambda)$ given by

$$\max\left\{r, \frac{1}{T}\left(\log\left(\frac{1}{\lambda}\right) - \log b\right) + \left(r + \frac{1}{2}\theta^2\right)\right\} \leqslant \overline{R}(\lambda) \leqslant r + \frac{1}{2}\theta^2. \tag{51}$$

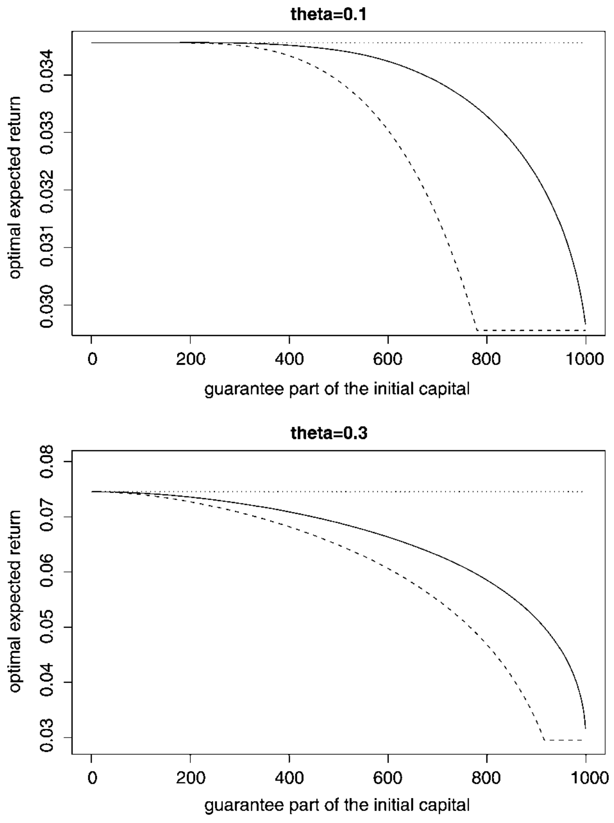


Fig. 12. Optimal expected return in a Black–Scholes model.

The upper bound follows from the fact that investing money without a constraint on terminal wealth must lead to a higher return than trading subject to a terminal constraint. The lower bound is true since

$$\mathrm{E}\log\left(\max\left\{G, \frac{1}{\lambda}H(T)\right\}\right) \geqslant \mathrm{E}\log\left(\frac{1}{\lambda}H(T)\right) = \log\left(\frac{1}{\lambda}\right) + \left(r + \frac{1}{2}\theta^2\right)T. \quad (52)$$

3.4. Periodical investment and consumption

The next step is to introduce periodical inflows and outflows. Assume that in additional to its initial capital b the investor receives non-negative amounts $B(t_1), \ldots, B(t_n)$, and has to pay non-negative amounts $C(t_1), \ldots, C(T_n)$ at time points $0 < t_1 < \cdots < t_n < t_{n+1} = T$. Introduce the net values $A(t_i) = B(t_i) - C(t_i)$. These values may be random.

We assume that each $A(t_i)$ is an $\mathcal{F}_{t_i}$ measurable contingent t_i-claim which is uniformly bounded. Thus it admits a unique price process given by

$$V_t\big(A(t_i)\big) = \beta(t)\mathrm{E}^*\big(\beta(t_i)^{-1}A(t_i)|\mathcal{F}_t\big), \quad 0 \leqslant t \leqslant t_i, \tag{53}$$

and the capitalized initial value of the income–outcome stream A is

$$V_0(A) = \sum_{i=1}^{n} V_0\big(A(t_i)\big). \tag{54}$$

We might have initial negative wealth but assume that $V_0(A)+b > 0$. The investor may choose a trading strategy ξ that is financed by the net income stream A. This means that its associated wealth process $Y^\xi(t)$ fulfills (23) with initial condition b and additionally

$$Y^\xi(t_i) - Y^\xi(t_i-) = A(t_i) \tag{55}$$

for each $1 \leqslant i \leqslant n$. Starting from an initial wealth b, the trading strategy is self-financing between two t_i, and income $B(t_i)$ at t_i provides a jump for the wealth process which will be invested immediately according to $\xi(t_i)$. In this setting the maximization of expected utility of terminal wealth has been treated by, e.g., Karatzas et al. (1991) and El Karoui and Jeanblanc-Picque (1998).

Denote by $\bar{\mathcal{A}}(b, A)$ the set of all trading strategies ξ financed by A such that its associated wealth process Y^ξ has initial value b, is allowed to become negative but the terminal wealth must be non-negative and fulfills $\mathrm{E}U(Y^\xi(T))^- < \infty$. Then the following borrow strategy becomes optimal (compare El Karoui and Jeanblanc-Picque (1998)):

- sell for each t_i a contract that delivers a payoff $B(t_i)$ at t_i,
- buy for each t_i a contract that delivers $C(t_i)$ at t_i,
- invest $b' = b + V_0(A)$ in an optimal trading strategy ξ as defined in Theorem 3.2.1 and trade until the end,
- at each t_i use income $B(t_i)$ to deliver the payoff with respect to our initially sold contract,
- at each t_i take the payoff from our initially bought contract to satisfy our consumption $C(t_i)$.

If we include the constraint that terminal wealth must exceed a predetermined benchmark $G > 0$ we have to assume that the initial net capitalized value $b + V_0(A) > G \cdot P(0, T)$. Then an optimal trading strategy can be defined as above by replacing the optimization step with its guaranteed terminal wealth counterpart, see Theorem 3.3.1, i.e.,

- invest $b' = b + V_0(A)$ in an optimal trading strategy ξ as defined in Theorem 3.3.1 and trade until the end.

The strategy above is optimal, see Theorem 3.4.1.

Theorem 3.4.1. *Let $V_0(A) = \sum_{i=1}^{n} V_0(A_{t_i})$ be the capitalized value of the net income stream $A = (A(t_1), \ldots, A(t_n))$ such that $b' = b + V_0(A) > 0$. Then for each trading* *strategy* $\phi \in \bar{\mathcal{A}}(b, A)$ *there exists a trading strategy* $\xi \in \mathcal{A}(b')$ *such that their terminal wealth coincide, hence* $Y^{\xi}(T) = Y^{\phi}(T)$ *and vice versa. In particular, the optimal value of both optimization problems coincide, i.e.,*

$$\max_{\phi \in \bar{\mathcal{A}}(b,A)} \mathrm{E}U\big(Y^{\phi}(T)\big) = \max_{\xi \in \mathcal{A}(b')} \mathrm{E}U\big(Y^{\xi}(T)\big). \tag{56}$$

Proof. Let $\phi \in \bar{\mathcal{A}}(b, A)$. Since the financial market is complete we may consider for each t_i a B-contract that delivers payoffs of $B(t_i)$ and a C-contract that delivers payoffs of $C(t_i)$ at t_i. These contracts have unique fair price processes $V_0(B(t_i))$ and $V_0(C(t_i))$ with $V_0(A(t_i)) = V_0(B(t_i)) - V_0(C(t_i))$. The corresponding trading strategy ξ can be defined in the following way: Use the initial wealth $b' > 0$

(1) to go long in the B-contract for each t_i, $1 \leqslant i \leqslant n$,
(2) to go short in the C-contracts for each t_i, $1 \leqslant i \leqslant n$,
(3) to invest b with respect to the trading strategy ϕ and trade according to ϕ until end,
(4) use at each t_i the payoff $A(t_i) = B(t_i) - C(t_i)$ of the contracts to invest and trade until the end.

Then the associated wealth process of ξ fulfills for each $0 \leqslant k \leqslant n$

$$Y^{\phi}(t) + \sum_{i=k+1}^{n} V_t\big(A(t_i)\big) = Y^{\xi}(t) \quad \text{for } t_k \leqslant t \leqslant t_{k+1}. \tag{57}$$

In the last trading interval from t_n to T both wealth processes coincide, hence also at terminal time. Due to our requirements, $Y^{\phi}(t)$ and therefore $Y^{\xi}(t)$ always stay above a fixed lower bound and have non-negative terminal wealth. Since ξ is self-financing, arbitrage arguments provide that $Y^{\xi}(t)$ is non-negative for all $0 \leqslant t \leqslant T$. Hence $\xi \in \mathcal{A}(b')$.

To prove the other direction, we define according to $\xi \in \mathcal{A}(b')$ the trading strategy $\phi \in \bar{\mathcal{A}}(b, A)$ in the following way:

- go short in the B-contracts for each t_i,
- go long in the C-contracts for each t_i,
- invest the obtained capital b' to invest with respect to the trading strategy ξ and trade until end,
- use the net income $B(t_i)$ at each t_i to deliver the payoff at t_i from our short position in the B-contract and pay $C(t_i)$ from our long position in C-contract.

Then we get the same evolution of wealth as in (57), hence the terminal wealth of both strategies coincide. That ϕ is indeed contained in $\bar{\mathcal{A}}(b, A)$ can be seen from (57) due to the fact that $Y^{\xi}(t) \geqslant 0$, $\sum_{i=k+1}^{n} V_t(A(t_i))$ stays uniformly bounded for all $0 \leqslant t \leqslant T$. □

Remark. Suppose that there is only inflow B and no outflow ($C = 0$). The following trading strategy seems to be intuitively reasonable: Invest each received income $B(t_i)$ at t_i in an optimal trading strategy for the remaining trading interval $[t_i, T]$. It turns out, that the optimal strategy described above has a better performance. We illustrate this in a Black–Scholes model with constant coefficients and log-utility. Consider the case where we have initial capital b and one additional investment the same amount of b at $t_1 \in (0, T)$. A strategy without borrowing leads to an evolution of wealth given by

$$\frac{b}{H(t)} \quad \text{for } 0 \leqslant t < t_1 \quad \text{and} \quad \frac{b}{H(t)} + \frac{bH(t_1)}{H(t)} \quad \text{for } t_1 \leqslant t < T \tag{58}$$

with $H(t) = \mathrm{e}^{-rt} \exp(-\theta W_t - \frac{1}{2}\theta^2 t)$, $\theta = \frac{\mu - r}{\sigma}$. Its terminal wealth Y has the expected utility

$$\begin{aligned} \mathrm{E} \log Y &= \mathrm{E} \log\left(\frac{b}{H(T)}\big(1 + H(t_1)\big)\right) \\ &= \log b + \mathrm{E} \log \frac{1}{H(T)} + \mathrm{E} \log\big(1 + H(t_1)\big) \\ &= \log b + \left(r + \frac{1}{2}\theta^2\right)T + \mathrm{E} \log\big(1 + H(t_1)\big) \\ &< \log b + \left(r + \frac{1}{2}\theta^2\right)T + \log\big(1 + \mathrm{E} H(t_1)\big) \\ &= \log(b + b\mathrm{e}^{-rt_1}) + \left(r + \frac{1}{2}\theta^2\right)T, \end{aligned} \tag{59}$$

where the last term on the right is the expected utility of terminal wealth from the optimal borrow strategy described above, i.e., which invests the enlarged initial capital $b + b \exp(-rt_1)$ on the whole trading interval $[0, T]$ into the optimal Merton strategy.

If we include the guarantee constraint $Y(T) \geqslant G$ a.s., we have to assume that the initial net capitalized value is large enough, i.e., $b + V_0(A) > G \cdot P(0, T)$, to make the problem feasible. An optimal trading strategy can be defined as above only replacing the optimal investment step with its guaranteed terminal wealth counterpart, as shown in Theorem 3.3.1,

- invest $b' = b + V_0(A)$ in an optimal trading strategy ξ as defined in Theorem 3.3.1 and trade until end.

3.5. Mortality risk

So far we have investigated how to optimally invest, if periodical in- and outflows from the portfolio may happen. In this section we clarify how the preceding notions can be applied to an insurance setting.

We take the view of an insurance company that has to manage a large portfolio of insurance contracts. In such a case the fluctuations average out and one might work with expected flows instead of random flows. Alternatively, one could consider an insurer who strictly separates his insurance portfolio from his investment portfolio. While the market risk is modeled in the Black–Scholes model, the event risk (mortality risk) is replaced by expectations only. Additional costs in risk capital provision for mortality fluctuations are not taken into consideration.

First we investigate what amount of money an insurance company must periodically consume to cover its obligation from mortality risk. As before we consider a contract running time $[0, T]$ which is divided into $n+1$ periods with endpoints

$$0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = T,$$

and investigate the obligation from a death insurance contract from an actuarial point of view. Ingredients of such a contract are the

(1) distribution of the random residual life time τ of an individual of age a, and

(2) death benefit $D > 0$, which has to be delivered at the end of that time period at that a death would occur.

At the beginning of each time period, insurance managers have to announce, which amount of money is needed for each contract to cover its mortality risk for the next period. If the injured individual is alive at t_{i-1} the risk sum D is payable at t_i in the case of death during $(t_{i-1}, t_i]$. Hence the insurance company has to reserve or consume for each individual alive the expected payoff

$$D \cdot P(t_{i-1}, t_i)\mathrm{P}(\tau \leqslant t_i | \tau > t_{i-1}) \tag{60}$$

at t_{i-1}. Only with probability $\mathrm{P}(\tau > t_{i-1})$, such a contract is alive at our portfolio of contracts at t_{i-1}. The manager has to announce a consumption of

$$C(t_{i-1}) = D \cdot P(t_{i-1}, t_i)\mathrm{P}(t_{i-1} < \tau \leqslant t_i) \tag{61}$$

at t_{i-1}. This leads to an initial consumption of $C(0) = D \cdot P(0, t_1)\mathrm{P}(\tau \leqslant t_1)$ for covering the mortality risk of the first time period and a future consumption stream $C = (C(t_1), \ldots, C(t_n))$ for the remaining future time periods. The initial consumption together with the capitalized value of the consumption stream C can be seen from an actuarial point of view as the initial price of such a death insurance contract. Its value is

$$C(0) + V_0(C) = \sum_{k=0}^{n} D \cdot P(0, t_{k+1})\mathrm{P}(t_k < \tau \leqslant t_{k+1}). \tag{62}$$

A death insurance contract will be financed by a premium income stream. At each t_i the insurer demands a constant premium $b > 0$ from each injured individual alive at t_i. Thus a portfolio manager in the insurance company may initially calculate the expected income at t_i

$$B(t_i) = b\mathrm{P}(\tau > t_i) \quad \text{for } 0 \leqslant i \leqslant n.$$

We receive an initial income $x = b$ and a future income stream

$$B = \big(B(t_1), \ldots, B(t_n)\big).$$

The capitalized value of all incomes are

$$b + V_0(B) = b + \sum_{k=1}^{n} B \cdot P(0, t_k)\mathrm{P}(\tau > t_k). \tag{63}$$

If this value coincides with the initial price determined by the consumption stream, i.e.,

$$b + V_0(B) = C(0) + V_0(C), \tag{64}$$

then the equivalence principle holds. The premium income stream just suffices to cover the mortality risks. No additional capital can be invested in financial markets.

3.6. *Managing unit-linked life insurance contracts*

If $b + V_0(B) > C(0) + V_0(C)$ the insurance company does not only cover its obligations from mortality risk, but also uses some capital to invest into the market. Hence it can be seen as an investor that would like to maximize terminal wealth by taking his income–outcome stream defined by B, respectively, C, and his initial wealth given by $b - C(0)$ into consideration.

If there is no guaranteed survival payoff at terminal time, then it is a death insurance contract which allows the insurance company to invest a part of the received premium in risky assets of the market. Ingredients of such a contract are

(1) a running time interval $[0, T]$ divided into periods with end-points

$$0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = T,$$

(2) a mortality risk sum D payable at the end of that time period at that a death will occur,

(3) constant premium income B at each t_i from each injured individual alive at t_i,

(4) investment of a part of the premium in financial markets.

The portfolio manager of an insurance company may see the above investment problem as a portfolio optimization problem with income–outcome stream $A = (A(t_1), \ldots, A(t_n))$ defined by

$$A(t_i) = B(t_i) - C(t_i) = B\mathrm{P}(\tau > t_i) - DP(t_i, t_{i+1})\mathrm{P}(t_i < \tau \leqslant t_{i+1}) \tag{65}$$

for each $1 \leqslant i \leqslant n$. Furthermore the initial wealth of such a contract is

$$b' = b - DP(0, t_1)\mathrm{P}(\tau \leqslant t_1), \tag{66}$$

i.e., the first premium b minus the insurance premium for mortality risk during the first period. In this setting we may apply the results of our preceding sections and refer to the optimal strategy defined in continuation to Theorem 3.4.1.

If a guaranteed survival payoff $S \geqslant G$ payable at terminal time T (if the insured person is alive at maturity) is included in addition to the preceding unit-linked life insurance contract, then we obtain a portfolio optimization problem with income–outcome stream A, initial wealth b, and a constraint on terminal wealth given by $S\mathrm{P}(\tau > T)$. To finance this constraint, our capitalized net value $b + V_0(A)$ must exceed the initial capital $G\mathrm{P}(\tau > T)P(0, T)$ required for ensuring the terminal wealth constraint. By applying the modified Theorem 3.4.1, we obtain the following optimal strategy

(1) for each t_i sell a contract that delivers a payoff $B(t_i) = B\mathrm{P}(\tau > t_i)$ at t_i,

(2) for each t_i buy a contract that delivers at t_i

$$C(t_i) = DP(t_i, t_{i+1})\mathrm{P}(t_i < \tau \leqslant t_{i+1}),$$

(3) buy $G\mathrm{P}(\tau > T)$ zero coupon bonds with maturity T,

(4) invest $b + V_0(B) - GP(0, T)\mathrm{P}(\tau > T)$ in a call on $I(\lambda H(T))$ with strike $G\mathrm{P}(\tau > T)$ and maturity T with λ solving the equation

$$b + V_0(A) = G\mathrm{P}(\tau > T)P(0, T) + V_0\Big(\big(I\big(\lambda H(T)\big) - G\mathrm{P}(\tau > T)\big)^+\Big),$$

(5) at each t_i use the income $B(t_i)$ to deliver the payoff with respect to our initially sold contract,

(6) at each t_i take the payoff from our initially bought contract to satisfy our consumption $C(t_i)$.

With this strategy we obtain the terminal wealth

$$G\mathrm{P}(\tau > T) + \big(I\big(\lambda H(T)\big) - G\mathrm{P}(\tau > T)\big)^+.$$

3.7. Numerical example

Consider the log-utility case in a one stock Black–Scholes model with constant coefficients (as shown in the example above with $\theta = 0.3$). We consider a 30 year old male individual who pays a yearly premium of 1500 Euro for a time period of 30 years. This income stream is initially valued via

$$b + V_0(B) = 29\,290 \text{ Euro}.$$

He is injured against mortality risk with a risk sum of 100 000 Euro which leads to the initial price of the consumption stream

$$C(0) + V_0(C) = 7826 \text{ Euro}.$$

Dependent on the survival payoff we have calculated the optimal expected return from managing a unit-linked contract with respect to

$$b' = b + V_0(B) - C(0) + V_0(C) = 21\,463 \text{ Euro},$$

the initial net capital that can be invested. We may obtain the plot like in the example above and observe that a survival payoff less than 60 000 Euro can be financed with the income stream, see Figure 13. Furthermore the higher the survival payoff the less is the optimal expected return. In the limits we get the return from a pure Bond, respectively, the optimal Merton strategy. As in the previous plot we have included the theoretical lower and upper bound on the optimal expected return.

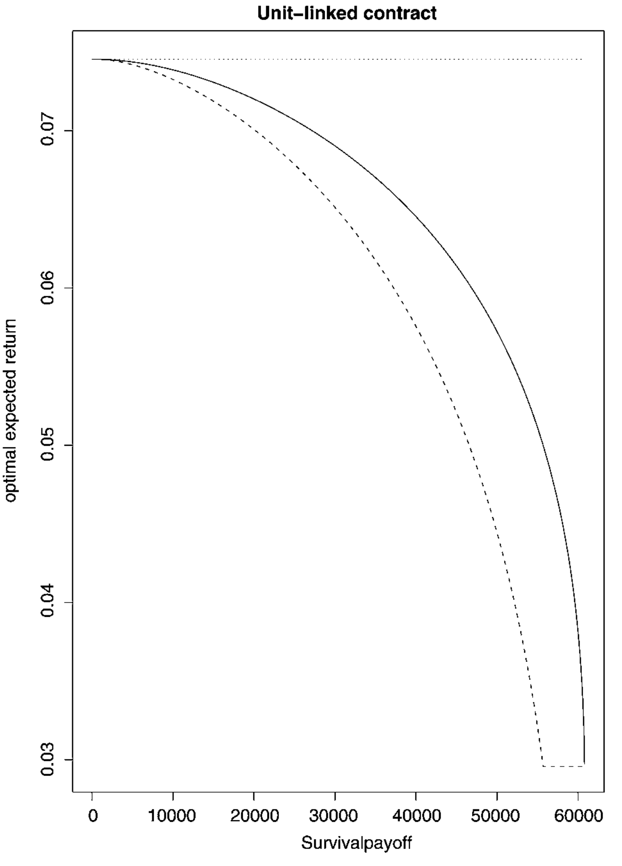


Fig. 13. Optimal expected return for a unit-linked contract with survival payoff.

4. Conclusion

In this chapter, both discrete-time and continuous-time versions of the problem to optimally manage unit-linked life insurance contracts with guarantee (ULLIG) have been discussed.

In the discrete-time case, we presented a model, which allows for pricing, managing and designing ULLIG products subject to various legal circumstances. This multi-stage stochastic optimization model allows for managing assets to balance liabilities emerging from such a contract. The model may be used to find the fair price of such a product, as well as to calculate risks, in particular, shortfall risk associated with a specific contract. Various types of insurance risk can be limited and controlled. Extensions such as transaction costs, taxes, legal constraints, lapse risk, etc. may be added to the model easily. The availability of robust and fast mathematical programming solvers allows for numerical solutions of realistically large models. An additional non-standard asset class (short-term life insurance) is considered for building contract specific portfolios. The return of such an investment is contingent to the death event and is sharply distinct from the return of market investments. It turns out that the fraction invested in insurance depends on the mortality risk as well as the fund performance. The rule is that the higher the funds and the lower the mortality risk, the smaller is the amount invested in insurance. Numerical results were presented to substantiate the usability of this model for management purposes.

In the continuous-time case, an analytic model for optimal investment under guarantee was studied. It is based on the Geometric Brownian Motion model for prices of all investment categories. Therefore, in this model the investment in insurance cannot be optimized and we considered the situation of a fixed pre-determined insurance part. For such a model, we were able to demonstrate, that the optimal management under guarantee adopting the log utility can be found by rules of going short or long in term contracts, buying options and using delta hedging. This result is in accordance with the findings of Brennan and Schwartz (1976). However, it turns out that the optimal dynamic strategy is to capitalize the expected premium inflow right from the beginning, instead of waiting until inflows occur.

While the multi-stage stochastic optimization model allows for considering various types costs and constraints, the analytic model shows how conventional instruments have to be composed for hedging the contract. A comprehensive management strategy should consider an application of a balanced mixture of both modeling views to support an optimal decision management process.

References

Bacinello, A., 2001. Fair pricing of life insurance participating policies with a minimum interest rate guaranteed. ASTIN Bulletin 31 (2), 275–297.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Boyle, P., Hardy, M., 1997. Reserving for maturity guarantees: two approaches. Insurance: Mathematics and Economics 21, 113–127.

Boyle, P., Hardy, M., 2003. Guaranteed annuity options. ASTIN Bulletin 33 (2), 125–152.

Boyle, P., Schwartz, E., 1977. Equilibrium prices of guarantees under equity-linked contracts. The Journal of Risk and Insurance 44, 639–660.

Brennan, M., Schwartz, E., 1976. The pricing of equity-linked life insurance policies with an asset value guarantee. Journal of Financial Economics 3, 195–213.

Brennan, M., Schwartz, E., 1979. Alternative investment strategies for the issuers of equity linked life insurance policies with an asset value guarantee. Journal of Business 52, 52–63.

Chadburn, R., 1997. The use of capital bonus policy and investment policy in the control of solvency for with-profits life insurance companies in the UK. Actuarial Research Paper 95. City University, London.

Consiglio, A., Cocco, F., Zenios, S.A., 2000. Asset and liability modeling for participating policies with guarantees, Working Paper 00-41-C. The Financial Institutions Center, The Wharton School.

Consiglio, A., Cocco, F., Zenios, S.A., 2001. The value of integrative risk management for insurance products with guarantees. Journal of Risk Finance, 6–16.

Consiglio, A., Cocco, F., Zenios, S.A., 2005. Scenario optimization asset and liability modeling for endowments with guarantees. Annals of Operations Research.

Consiglio, A., Cocco, F., Zenios, S.A., 2006. The PROMETEIA model for managing insurance policies with guarantees. In: Zenios, S., Ziemba, W. (Eds.), Handbook of Asset and Liability Management. In: Handbooks in Finance. Elsevier.

Consiglio, A., De Giovanni, D., 2006. Evaluation of insurance products with guarantee in incomplete markets, Working Paper 06-02. Department of Statistics and Mathematics, University of Palermo.

Consiglio, A., Saunders, D., Zenios, S.A., 2003. Insurance league: Italy vs. UK case. Journal of Risk Finance 2003, 1–8, Summer.

Consiglio, A., Saunders, D., Zenios, S.A., 2006. Asset and liability management models for insurance products with guarantees: The UK case. Journal of Banking and Finance 30, 645–667.

Cox, J., Huang, C., 1989. Optimum consumption and portfolio policies when asset prices follow a diffusion process. Journal of Economic Theory 49 (1), 33–83.

El Karoui, N., Jeanblanc-Picque, M., 1998. Optimization of consumption with labor income. Finance and Stochastics 2, 409–440.

Fourer, R., Gay, D., Kernighan, B., 2002. AMPL: A Modeling Language for Mathematical Programming. Duxbury Press and Brooks/Cole Publishing Company.

Grosen, A., Jørgensen, P., 2000. Fair valuation of life insurance liabilities: the impact of interest rate guarantees, surrender options, and bonus policies. Insurance: Mathematics and Economics 26, 37–57.

Hochreiter, R., Pflug, G.C., 2006. Financial scenario generation for stochastic multi-stage decision processes as facility location problems, Annals of Operations Research.

Inada, K., 1963. On a two-sector model of economic growth: Comments and a generalization. Review of Economic Studies 30 (2), 119–127.

Jensen, B., Jørgensen, P., Grosen, A., 2001. A finite difference approach to the valuation of path dependent life insurance liabilities. The Geneva Papers on Risk and Insurance Theory 26, 57–84.

Kallberg, J.G., Ziemba, W.T., 1983. Comparison of alternative utility functions in portfolio selection problems. Management Science 29 (11), 1257–1276.

Karatzas, I., 1997. Lectures on Mathematical Finance. CRM Monograph Series, vol. 8. American Mathematical Society.

Karatzas, I., Lakner, P., Lehoczky, J., Shreve, S., 1991. Equilibrium in a simplified dynamics economy with heterogeneous agents. In: Meyer-Wolf, E., Merzbach, E., Zeitouni, O. (Eds.), Stochastic Analysis Liber Amicorum for Moshe Zakai. Academic Press, New York, pp. 245–272.

King, A.J., 2002. Duality and martingales: A stochastic programming perspective on contingent claims. Mathematical Programming, Series B 91, 543–562.

King, A.J., Koivu, M., Pennanen, T., 2005. Calibrated option bounds. International Journal of Theoretical and Applied Finance 8, 141–159.

Korn, R., 1997. Optimal Portfolios—Stochastic Models for Optimal Investment and Risk Management in Continuous Time. World Scientific.

Kouwenberg, R., Zenios, S., 2006. Stochastic programming models. In: Zenios, S., Ziemba, W. (Eds.), Handbook of Asset and Liability Management. In: Handbooks in Finance, vol. 1. Elsevier, Ch. 6.

Merton, R., 1971. Optimum consumption and portfolio rules in a continuous time model. Journal of Economic Theory 3, 373–413.

Miltersen, K., Persson, S., 1999. Pricing rate of return guarantees in a Heath–Jarrow–Morton framework. Insurance: Mathematics and Economics 25, 307–325.

Pflug, G.C., 2001. Optimal scenario tree generation for multiperiod financial planning. Mathematical Programming, Series B 89, 251–271.

Pliska, S., 1986. A stochastic calculus model of continuous trading: Optimal portfolios. Mathematics of Operations Research 11, 371–382.

Ruszczyński, A., Shapiro, A. (Eds.), 2003. Stochastic Programming. Handbooks in Operations Research and Management Science, vol. 10. Elsevier.

Siglienti, S., 2000. Consequences of the reduction of interest rates on insurance. The Geneva Papers on Risk and Insurance 25 (1), 63–77.

Susinno, G., Giraldi, C., 2000. Insurance optional. Risk 13 (4), April.

Wallace, S.W., Ziemba, W.T. (Eds.), 2005. Applications of stochastic programming. MPS/SIAM Series on Optimization, vol. 5. Society for Industrial and Applied Mathematics (SIAM).

Wilkie, A., 1986. A stochastic investment model for actuarial use. Transactions of the Faculty of Actuaries 39, 341–381.

Wilkie, A.D., Waters, H.R., Yang, S.Y., 2003. Reserving, pricing and hedging for policies with guaranteed annuity options. British Actuarial Journal 9 (2), 263–391.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset Liability and Wealth Management. AIMR, Charlottesville, VA.

Ziemba, W.T., Mulvey, J.M. (Eds.), 1998. Worldwide Asset and Liability Modeling. Cambridge University Press.

Chapter 15

THE PROMETEIA MODEL FOR MANAGING INSURANCE POLICIES WITH GUARANTEES

ANDREA CONSIGLIO

University of Palermo, Italy
e-mail: consiglio@unipa.it

FLAVIO COCCO

Prometeia, Bologna, Italy
e-mail: flavio.cocco@prometeia.it

STAVROS A. ZENIOS

Department of Public and Business Administration, University of Cyprus, 1678 Nicosia, Cyprus
and
Senior Fellow, The Wharton School, University of Pennsylvania, USA
e-mail: zenioss@ucy.ac.cy

Contents

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02015-1

Abstract

Insurance products become increasingly more innovative in order to face competitive pressures. Insurance policies today come with guarantees on the minimum rate of return, bonus provisions, and surrender options. These features make them attractive for investors who seek not only insurance but also investment vehicles. However, new policies are much more complex to price and fund than traditional insurance products. In this chapter we discuss the development of a scenario-based optimization model for asset and liability management for the participating policies with guarantees and bonus provisions offered by Italian insurers. The changing landscape of the financial services in Italy sets the backdrop for the development of this system which was the result of a multi-year collaborative effort between academic researchers, the research staff at Prometeia in Bologna, and end-users from diverse Italian insurers. The model is presented, its key features are discussed in detail, and several extensions are briefly introduced. The resulting system allows the analysis of the tradeoffs facing an insurance firm in structuring its policies as well as the choices in covering their cost. It is applied to the analysis of policies offered by Italian insurance firms. While the optimized model results are in general agreement with current industry practices, inefficiencies are still identified and potential improvements are suggested. Extensive numerical experiments provide significant insights on features of the participating guaranteed policies.

Keywords

risk management, asset–liability management, insurance products with guarantee

JEL classification: C61, G22, G32

1. Introduction

The last decade brought about a phenomenal increase of consumer sophistication in terms of the financial products they buy. This trend is universal among developed economies, from the advanced and traditionally liberal economies of North America to the increasingly deregulated economies of the European Union and pre-accession States, and the post-Communist countries.

The numbers are telling: In the 1980s almost 40% of the US consumer financial assets were in Bank deposits. By 1996 bank deposits accounted for less than 20% of consumers' financial assets with mutual funds and insurance/pension funds absorbing the difference (Harker and Zenios, 2000, Ch. 1). Similar trends are observed in Italy. The traded financial assets of Italian households more than doubled in the 5-year period from 1997, and the bulk of the increase was absorbed by mutual funds and asset management; see Table 1.

The increase in traded financial assets comes with increased diversification of the Italian household portfolio, similar to the one witnessed in the US a decade earlier. Figure 1 shows a strong growth of mutual funds and equity shares at the expense of liquid assets and bonds. Today one third of the total revenues of the Italian banking industry is originated by asset management services.

These statistics reveal the *outcome* of a changing behavior on the part of consumers. What are the changing characteristics of the consumers, however, that bring about this new pattern of investment? The annual *Household Savings Outlook* carried out by Prometeia—a Bologna based company established in 1981 to carry out economic research and analysis, and provide consulting services to major financial institutions and government agencies in Italy—in collaboration with Eurisko—a Milan based company conducting research on consumption, communications, and social transformation—provides important insights. First, the traditional distinction between *delegation* of asset management to a pension fund or an insurance firm by the majority of consumers, and *autonomy* in the management of assets by wealthy investors, no longer appears to be valid. Both attitudes are present in the behavioral patterns of private savers.

Table 1
Traded financial assets by Italian households during 1997–2002 in billions of ITL

	1997	1998	1999	2000	2001	2002
Household total	944.853	1427.999	1781.996	2124.102	2488.154	2877.773
% of household's assets	23.6	31.4	34.6	38.3	41.9	44.8
Mutual funds	368.432	720.823	920.304	1077.360	1237.964	1386.519
Asset management	375.465	542.205	673.500	781.300	880.450	956.970
Life and general insurance	165.000	202.300	257.400	329.600	433.400	574.000

Source: ISVAP, the board of regulators for Italian insurers.

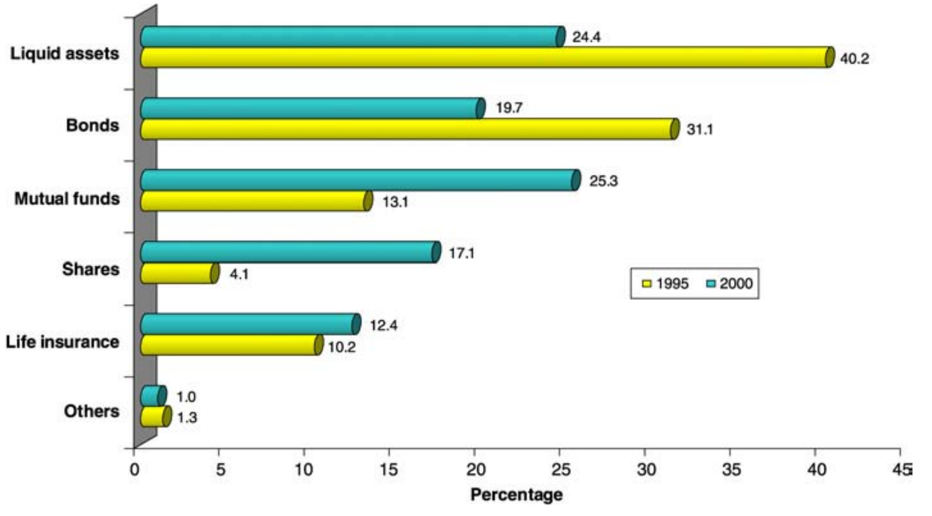


Fig. 1. The evolution of Italian household portfolios.

Second, the trend in behavioral profiles is towards higher levels of autonomy, and there is an increased propensity towards innovative instruments as manifested in the data of Figure 1. The group of Italian households classified as "innovators" grew steadily from 6.7% in 1991 to 22.6% by 2001. Each percentage point increase added a further 200,000 households to this category. Today this segment numbers 4.3 million Italian households. Households in this category adopt a very professional approach to questions of finance. They are able—or at least they feel so—to manage their financial affairs, and they rely on integrated delivery channels for doing so, using on-line information and conducting business by phone.

Third, an analysis of the influence of quantitative variables on the savings habits of households shows that awareness of financial indicators, and in particular the performance of managed asset returns, is influencing household behavior. Investors in older age groups are more aware of such indicators than the younger generations. The survey also reveals that the trend towards increased diversification of assets under management will continue unabated during the next three years. The investors' favorites are insurance and portfolio management. (The survey was conducted just prior to the stalling of the world-wide bull markets so the projection of a continued favor towards portfolio management can be questioned.)

In this environment the Italian insurance industry has come under increasing pressure. The statistics of Table 1 reveal that assets invested in life and general insurance increased by 99% in the period of interest while assets in mutual funds increased by 190%, and those under asset management by 110%. Insurance companies trail the competition in claiming a share of the household's wallet. The industry expects to reverse this trend by 2002. By that time Italian households are expected to increase their traded assets by 200%, with the insurance policies increasing their share by 250%, mutual funds by 280%, and asset managers by 150%. The main competitive weapon in the arsenal of the insurance firms are innovative policies that offer both traditional insurance and participation in the company's profits. These policies combine features of traditional insurance from actuarial risks and of investment vehicles such as mutual funds.

Insurance products with minimum guaranteed rate of return and bonus provisions play today a key role in the insurers' business portfolio. Such products were first offered by insurance companies in the inflationary seventies. In order to compete with the high yields of Treasury bonds of that time, insurance policies were enhanced with both a minimum guaranteed rate of return and a bonus provision when asset fund returns exceed the minimum guarantee. The right to surrender the product at any time before maturity is also often given to policyholders. Such policies, known as unit-linked or index-linked, are prevalent among continental European insurance companies, but they are also encountered in the UK, United States and Canada. In the low-inflation 1990s insurance companies still could not abandon these products due to the competitive pressures outlined above.

With the historically low interest rates prevailing currently the management of such policies is becoming more challenging. Reliance on fixed-income assets is unlikely to yield the guaranteed rate of return. For instance, Italian guaranteed rates after 1998 are at 3%. The difference between the guaranteed rate and the ten-year yield is only 1%, which is inadequate for covering the firm's costs. In Germany the guaranteed rates after 1998 are at 3.5% differing from the ten-year yield only by 0.5%. Danish products offered guarantees of 3% until 1999, which were reduced to 2% afterwards. In Japan Nissan Mutual Life failed on a $2.56 billion liability arising from a 4.7% guaranteed policy.

In response to the challenges facing Italian insurers, Prometeia developed an asset and liability management system for participating insurance policies with guarantees (Consiglio, Cocco and Zenios, 2000, 2001). The system utilizes recent advances in financial engineering, with the use of scenario-based optimization models, to integrate the insurer's asset allocation problem with that of designing competitive policies. The competing interests of shareholders, policyholders and regulators are cast in a common framework so that efficient tradeoffs can be reached. In this chapter we discuss the model and illustrate its performance. In particular, it is shown that traditional methods are inadequate and innovative models are needed to address the complexities of these products. The resulting model allows the insurer to address asset allocation issues both locally and internationally in a way that is consistent with the offering of competitive products and the shareholders' interests, while satisfying the regulators. Section 2 discusses the Italian insurance industry and describes the characteristics of modern insurance products. Section 3 describes the model and Section 4 reports on model performance from the perspective of the shareholders, the policyholders and the regulators.

2. The Italian insurance industry

The Italian insurance industry is regulated and supervised by ISVAP, Istituto per la Vigilanza sulle Assicurazioni, established by law in 1982. The supervisory framework aims at the stability of the market of insurance undertakings, and at the solvency and efficiency of insurance market participants. ISVAP ensures that the technical, financial and accounting management of institutions under its supervision complies with the laws, regulations and administrative provisions in force.

In the performance of its duties ISVAP may require supervised undertakings to disclose data, management practices and other related information. This supervision monitors the undertaking's financial position, with particular regard to the existence of sufficient solvency margins and adequate technical provisions to ensure that adequate assets are available to cover the entire business.

Progress of the Italian legal framework over the last twenty years—the ISVAP web page lists 51 regulatory provisions—lead supervisors to devote increasing attention to data processing and real-time analysis of data. With solid preventive supervision in place ISVAP can intervene on a timely fashion in any risky situation. The availability of sophisticated safeguards, and the increased financial activity of the last decade driven by the changing nature of the Italian consumer described above, brought the creation of numerous and complex groups of insurance undertakings. These undertakings offer more innovative products, in response to market pressures, and they also take a more active role in the management of their assets and market risks in delivering quality products to clients. The average composition of the portfolios for life insurance, for instance, has been evolving towards more aggressive positions with increasing holdings in equity and high-quality corporate bonds as shown in Table 2. During the same period the industry has been promoting novel insurance policies with guarantees and participation in the profits.

2.1. Guaranteed products with bonus provisions

Financial products with guarantees on the minimum rate of return come in two distinct flavors: *maturity guarantees* and *multi-period guarantees*. In the former case the guarantee applies only to maturity of the contract, and returns above the guarantee occurring before maturity offset shortfalls at other periods. In the later case the time to maturity is divided into subperiods—quarterly or biannually—and the guarantee applies at the end of each period. Hence, excess returns in one subperiod cannot be used to finance shortfalls in other subperiods. Such guaranteed products appear in insurance policies, guaranteed investment contracts, and some pension plans, see, e.g., Hansen and Miltersen (2002).

Policyholders *participate* in the firm's profits, receiving a *bonus* whenever the return of the firm's portfolio exceeds the guarantee, creating a *surplus* for the firm. Bonuses may be distributed only at maturity, at multiple periods until maturity, or using a combination of distribution plans. Another important distinction is made according to the

Table 2
The structure of portfolios of Italian life insurers in percentage of total assets held in the major asset categories

Year	Titoli di Stato (Govt. bonds)	Azioni (stocks)	Obbligazioni (bonds)	Titoli in valuta (intnl. investment)
1995	65.2	7.8	14.8	11.4
1996	65.2	7.6	13.6	13.2
1997	60.2	9.1	14.7	15.0
1998	55.2	10.0	16.6	16.4

Source: ISVAP, the board of regulators for Italian insurers.

bonus distribution mechanism. In particular, some products distribute bonuses using a *smoothing* formula such as the average portfolio value or portfolio return over some time period, while others distribute a pre-specified fraction of the portfolio return or portfolio value net any liabilities. The earlier *unit-linked* policies would pay a benefit—upon death or maturity—which was the greater of the guaranteed amount and the value of the insurer's reference portfolio. These were simple maturity guarantees with bonus paid at maturity as well. At the other extreme of complexity we have the modern UK insurance policies. These policies declare at each subperiod a fraction of the surplus, estimated using a smoothing function, as *reversionary* bonus which is then guaranteed. The remaining surplus is managed as an *investment reserve*, and is returned to customers as terminal bonus if it is positive at maturity or upon death; see Ross (1989) and Chadburn (1997). These policies are multi-period guarantees with bonuses paid in part at intermediate times and in part at maturity. Further discussion on the characteristics of products with guarantees is found in Kat (2001) and the papers cited below.

The Prometeia model described here considers multi-period guarantees with bonuses that are paid at each subperiod and are subsequently guaranteed. The bonus is contractually determined as a fraction of the portfolio excess return above the guaranteed rate during each subperiod. The guaranteed rate is also contractually specified. To illustrate the nature of this product, we graph in Figure 2 the growth of a liability that participates by 85% in a given portfolio while it guarantees a return of at least 3% in each period. The liability is *lifted* every time a bonus is paid and the minimum guarantee applies to the increased liability: what is given cannot be taken away. This feature creates a complex nonlinear interaction between the rate of return of the portfolio and the total return of the liability.

2.2. Current asset and liability management practices

The shift from actuarial to financial pricing of insurance liabilities (Embrechts, 2000; Babbel, 2001) and widely perceived problems (highlighted by the Nissan bankruptcy

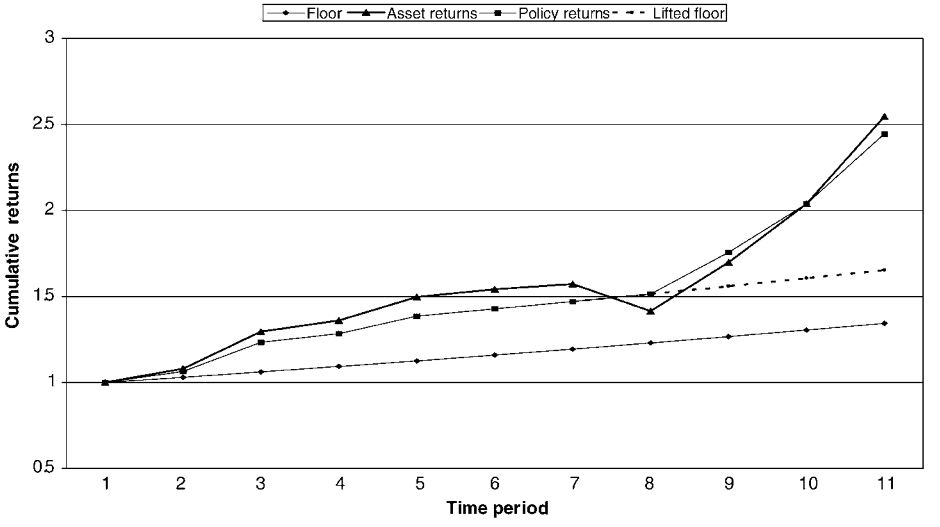


Fig. 2. Typical returns of the asset portfolio and a participating policy with a multi-period guaranteed return of 3% and participation rate of 85%. The guarantee applies to a liability that is lifted every time a bonus is paid as illustrated at period seven. The asset portfolio experienced substantial losses at period seven while the liability grew at the 3% guaranteed rate. Subsequent superior returns of the assets allowed the firm to recover its losses by the tenth period and achieve a positive net return at maturity.

case) brought about an interest in applying the theory of financial asset pricing to the analysis of insurance policies with guarantees and bonus provisions; see, e.g., Giraldi et al. (2003). Single period guarantees have payoffs that resemble those of a European-type option, as the policyholder receives at maturity the maximum between the guaranteed amount and the value of the bonus. Multi-period guarantees may have features, such as a surrender options, that makes their payoff identical to American type options. Hence, option pricing could be applied to the pricing of these policies.

The pricing of the option embedded in the early products with guarantees was addressed in the seminal papers of Brennan and Schwartz (1976) and Boyle and Schwartz (1977). They analyzed unit-linked maturity guarantee policies. Perhaps the most complete analysis of modern life insurance contracts—complete in the sense that it prices in an integrated framework several components of the policy—is due to Grosen and Jørgensen (2000). They decompose the liability of modern participating policies with guarantees into a risk-free bond (the minimum guarantee), a bonus option, and a surrender option. The first two taken together are a European contract and all three together are an American contract, and the authors develop numerical techniques for pricing both. Hansen and Miltersen (2002) extend this model to the pricing of contracts with a smoothing surplus distribution mechanism of the form used by most Danish life-insurance companies and pension plans. They use the model to study different methods for funding these products, either by charging the customers directly or by keeping a share of the surplus. Similarly, Bacinello (2001) develops pricing models that permit her to study the interplay between the volatility of the underlying asset portfolio, the participation level for determining bonuses, and the guaranteed rate. Boyle and Hardy (1997) take this line of inquiry in a different direction by analyzing alternative reserving methods for satisfying the guarantee. More practical aspects of the problem are studied by Giraldi et al. (2003) and Siglienti (2000).

It is worth noting that current literature assumes the asset side is given *a priori* as a well-diversified portfolio which evolves according to a given stochastic process. For instance, Brennan–Schwartz, Grosen–Jørgensen and Bacinello assume a geometric Browning motion, while Miltersen and Persson (1999) rely on the Heath–Jarrow–Morton framework and price multi-period guaranteed contracts linked either to a stock investment or the short-term interest rate. There is nothing wrong with these approaches, of course, except that part of the problem of the insurance companies is precisely to determine the structure of the asset portfolio. Indeed, all of the above references carry out simulations for different values of the volatility of the assets. Brennan and Schwartz (1979) devote a section to the analysis of "misspecification of the stochastic process". Bacinello goes on to suggest that the insurance company should structure several reference portfolios according to their volatility and offer its customers choices among different triplets of guaranteed rate, bonus provision, and asset portfolio volatility. To this suggestion of endogenizing the asset decision we subscribe. It is a prime example of *integrated financial product management* advocated by Holmer and Zenios (1995).

Independently of the literature that prices the option embedded in the liabilities we have seen an interest in the use of portfolio optimization models for asset and liability management for insurance companies. The most prominent example is for a Japanese insurance firm—not too surprising given what has transpired in the Japanese financial markets—the Yasuda Kasai model developed by the Frank Russel Company. This model received coverage not only in the academic literature but also in the press, see Carinõ and Ziemba (1998). Other successful examples include the Towers Perrin model of Mulvey and Thorlacius (1998), the CALM model of Consigli and Dempster (1998) and the Gjensidige Liv model of Høyland (1998). These models have been successful in practical settings but their application does not cover participating policies with guarantees. One reason is that insurance firms pursued integrated asset and liability management strategies for those products they understood well. This has been the case for policies that encompass mostly actuarial risk such as the fire and property insurance of the Yasuda Kasai model. Another reason is that the technology of scenario optimization through large-scale stochastic programming has only recently been developed into computable models, see, e.g., Censor and Zenios (1997).

Finally, the combination of a guarantee with a bonus provision introduces nonlinearities which complicate the model. Traditional approaches such as the mean-variance analysis are inadequate as they fails to capture some important characteristics of the problem. There is nothing efficient about efficient portfolios when the nonlinearity of the embedded options is properly accounted for. Novel models are needed to integrate the asset management problem with the characteristics of liabilities with minimum guarantee. Such a model was developed through a multi-year collaborative effort between academic researchers, the research staff at Prometeia in Bologna, and end-users from diverse Italian insurers. It is presented next.

3. The scenario optimization model

We develop in this section the model for asset and liability management for multi-period participating policies with guarantees. It is a mathematical program that models stochastic variables using discrete scenarios. All portfolio decisions are made at $t = 0$ in anticipation of an uncertain future. At the end of the planning horizon the impact of these portfolio decisions in different scenarios is evaluated and risk aversion is introduced through a utility function. Portfolio decisions optimize the expected utility over the specified horizon.

3.1. Features of the model

In the model we consider three accounts: (i) a liability account that grows according to the contractual guaranteed rate and bonus provision, (ii) an asset account that grows according to the portfolio returns, net any payments due to death or policy surrenders, and (iii) a shortfall account that monitors lags of the portfolio return against the guarantee. In the base model shortfall is funded by equity but later we introduce alternative reserving methods.

The multi-period dynamics of these accounts are conditioned on discrete scenarios of realized asset returns and the composition of the asset portfolio. Within this framework a regulatory constraint on leverage is imposed. At maturity the difference between the asset and the liability accounts is the surplus realized by the firm after it has fulfilled its contractual obligations. In the policies considered here this surplus remains with the shareholders. This surplus is a random variable, and a utility function is introduced to incorporate risk aversion.

3.2. Notation

We let Ω denote the index set of scenarios $l = 1, 2, \dots, N$, indicating realizations of random variables, $\mathcal{U}$ the universe of available asset instruments, and $t = 1, 2, \dots, T$, discrete points in time from today ($t = 0$) until maturity T. The data of the problem are as follows:

- r_{it}^l rate of return of asset i during the period $t - 1$ to t in scenario l.
- r_{ft}^l risk free rate during the period $t - 1$ to t in scenario l.
- g minimum guaranteed rate of return.
- β participation rate indicating the percentage of portfolio return paid to policyholders.
- ρ regulatory equity to debt ratio.
- Λ_t^l probability of abandon of the policy due to lapse or death at period t in scenario l.

The variables of the model are defined as follows:

x_i percentage of initial capital invested in the ith asset.
y^l_{At} expenses due to lapse or death at time t in scenario l.
z^l_t shortfall below the guaranteed rate at time t in scenario l.
A^l_t asset value at time t in scenario l.
E^l_t total equity at time t in scenario l.
L^l_t liability value at time t in scenario l.
R^l_{Pt} portfolio rate of return during the period $t-1$ to t in scenario l.
y^{+l}_t excess return over g at time t in scenario l.
y^{-l}_t shortfall return under g at time t in scenario l.

3.3. Variable dynamics and constraints

We invest the premium collected (L_0) and the equity required by the regulators ($E_0 = \rho L_0$) in the asset portfolio. Our initial endowment $A_0 = L_0(1+\rho)$ is allocated to assets in proportion x_i such that $\sum_{i\in\mathcal{U}} x_i = 1$, and the dynamics of the portfolio return are given by

$$R^l_{Pt} = \sum_{i\in\mathcal{U}} x_i r^l_{it}, \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{1}$$

The investment variables are nonnegative so that short sales are not allowed.

We now turn to the modeling of the liability account. Liabilities will grow at a rate which is at least equal to the guarantee. Excess returns over g are returned to the policyholders according to the participation rate β. The dynamics of the liability account are given by

$$L^l_t = \left(1 - \Lambda^l_t\right) L^l_{t-1}\left(1 + \max\left[\beta R^l_{Pt}, g\right]\right),$$
$$\text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{2}$$

The max operator introduces a discontinuity in the model. To circumvent this difficulty we introduce variables y^{+l}_t and y^{-l}_t to measure the portfolio excess return over the guaranteed rate, and the shortfall below the guarantee, respectively. They satisfy

$$\beta R^l_{Pt} - g = y^{+l}_t - y^{-l}_t, \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega, \tag{3}$$

$$y^{+l}_t \geqslant 0, \quad y^{-l}_t \geqslant 0, \quad y^{+l}_t y^{-l}_t = 0,$$
$$\text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{4}$$

Only one of these variables can be nonzero at any given time and in a given scenario.

The dynamics for the value of the liability are rewritten as

$$L^l_t = \left(1 - \Lambda^l_t\right) L^l_{t-1}\left(1 + g + y^{+l}_t\right), \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{5}$$

Liabilities grow at least at the rate of g. Any excess return is added to the liabilities and the guarantee applies to the lifted liabilities.

At each period the insurance company makes payments due to policyholders abandoning their policies because of death or lapse. Payments are equal to the value of the liability times the probability of abandonment, i.e.,

$$y^l_{At} = \Lambda^l_t L^l_{t-1}\left(1 + g + y^{+l}_t\right), \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{6}$$

Whenever the portfolio return is below the guaranteed rate we need to infuse cash into the asset portfolio in order to meet the final liabilities. The shortfall account is modeled by the dynamics

$$z^l_t = y^{-l}_t L^l_{t-1}, \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{7}$$

In the base model shortfalls are funded through equity. We assume that equity is reinvested at the risk-free rate and is returned to the shareholders at the end of the planning horizon. (This is not all the shareholders get; they also receive dividends.) The dynamics of the equity are given by

$$E^l_t = E^l_{t-1}\left(1 + r^l_{ft}\right) + z^l_t, \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{8}$$

By assuming the risk free rate as the alternative rate at which the shareholders could invest their money, we analyze the *excess* return offered to shareholders by the participating contract modeled here, over the benchmark risk free investment. In principle, one could use the firm's internal rate of return as the alternative rate, and analyze the excess return offered by the policy modeled here, over the firm's other lines of business. In this setting, however, the problem would not be to optimize the asset allocation to maximize shareholder value, since this would already be endogenous in the internal rate of return calculations. Instead we could determine the most attractive features for the policyholders—g and β—that will make the firm indifferent in offering the new policy or maintaining its current line of business. This approach deserves further investigation. For the purpose of optimizing alternative policies for the shareholders, while satisfying the contractual obligations to the policyholders, the estimation of excess return over the risk free rate is a reasonable benchmark. In Sections 4.3.1 and 4.3.3 we consider other alternatives for funding the shortfalls through long-term debt or short-term borrowing.

We now have the components needed to model the asset dynamics, taking into account the cash infusion that funds shortfalls, z^l_t, and the outflows due to actuarial events y^l_{At}, i.e.,

$$A^l_t = A^l_{t-1}\left(1 + R^l_{Pt}\right) + z^l_t - y^l_{At}, \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{9}$$

In order to satisfy the regulatory constraint the ratio between the equity value and liabilities must exceed ρ. That is,

$$\frac{V^l_{ET}}{L^l_T} \geqslant \rho, \quad \text{for all } l \in \Omega, \tag{10}$$

where V_{ET}^l is the value of equity at the end of the planning horizon T. If the company sells only a single policy the value of its equity will be equal to the final asset value—which includes the equity needed to fund shortfall—minus the final liability due to the policyholders, and we have

$$V_{ET}^l = A_T^l - L_T^l. \tag{11}$$

Having described the assets and liability accounts in a way that the key features of the policy—guaranteed rate and bonus provisions—are accounted for, we turn to the choice of an appropriate objective function. We model the goal of a for-profit institution to maximize the return on its equity, and, more precisely in this case, to maximize any excess return on equity after all liabilities are paid for. Since return on equity is scenario dependent we maximize the expected value of the utility of excess return. This expected value is converted into a certainty equivalent for easy reference. The objective function of the model is to compute the maximal Certainty Equivalent Excess Return on Equity (CEexROE) given by

$$\text{CEexROE} \doteq U^{-1}\left\{\operatorname*{Max}_x \frac{1}{N}\sum_{l\in\Omega} U\left\{\frac{A_T^l - L_T^l}{E_T^l}\right\}\right\}, \tag{12}$$

where $U\{\cdot\}$ denotes the decision maker's utility function and $A_T^l - L_T^l$ is the shareholder's reward in scenario l. We assume a power utility function with constant relative risk aversion of the form $U(V) = \frac{1}{\gamma}V^\gamma$, where $V \geqslant 0$, and $\gamma < 1$. In the base model we assume $\gamma = 0$ in which case the utility function is the logarithm corresponding to growth-optimal policies for the firm. In Section 4.5.1 we study the effect of changing the risk aversion parameter.

As a byproduct of our model we calculate the cost of funding the guaranteed product. Every time the portfolio return drops below the guaranteed rate, we counterbalance the erosion of our assets by infusing cash. This cost can be charged either to the policyholders, as soon as they enter the insurance contract, or covered through shareholder's equity or by issuing debt. These choices entail a tradeoff between the return to shareholders and return to policyholders. We study in the next section this tradeoff.

The cost of the guarantee is the expected present value of reserves required to fund shortfalls due to portfolio performances below the guarantee. The dynamic variable E_t^l models precisely the total funds required up to time t, valued at the risk-free rate. However, E_t^l also embeds the initial amount of equity required by the regulators. This is not a cost and it must be deducted from E_t^l. Thus, the cost of the guarantee is given as the expected present value of the final equity E_T^l adjusted by the regulatory equity, that is,

$$\overline{O}_G = \frac{1}{N}\sum_{l=1}^{N}\left(\frac{E_T^l}{\prod_{t=1}^{T}(1 + r_{ft}^l)} - \rho L_0\right). \tag{13}$$

$\overline{O}_G$ is the expected present value of the reserves required to fund this product. This can be interpreted as the cost to be paid by shareholders in order to benefit from the upside potential of the surplus. A more precise interpretation of $\overline{O}_G$ is as the *expected downside risk* of the policy. This is not the risk-neutral price of the participating policies with guarantees that would be obtained under an assumption of complete markets for trading the liabilities arising from such contracts. This is the question addressed through an options pricing approach in the literature cited above, Brennan–Schwartz, Boyle–Schwartz, Bacinello, Grosen–Jørgensen, Hansen–Miltersen, Miltersen–Persson.

3.4. Linearly constrained optimization model

The model defined in the previous section is a nonlinearly constrained optimization model and is computationally intractable for large scale applications. However, the nonlinear constraints (5)–(9) are definitional constraints which determine the value of the respective variables at the end of the horizon. We solve these dynamic equations analytically (see Appendix A) to obtain end-of-horizon analytic expressions for A_T^l, L_T^l, and E_T^l. These expressions are substituted in the objective function to obtain the equivalent linearly constrained nonlinear program below. The regulatory constraint (10), however, cannot be linearized. For solution purposes the regulatory constraint is relaxed and its validity is tested *ex post*. Empirical results later on demonstrate that the regulatory constraint is not binding for the policies considered here and for the generated scenarios of asset returns. However, there is no assurance that this will always be the case, and we may need to resort to nonlinearly constrained optimization for solving this model.

$$\begin{aligned}\underset{x\geqslant 0}{\text{Maximize}}\ \frac{1}{N}\sum_{l\in\Omega} U\Bigg\{\Bigg[&(1+\rho)\prod_{t=1}^{T}\big(1+R_{Pt}^l\big)+\sum_{t=1}^{T}\big(y_t^{-l}-\Lambda_t^l\big(1+g+y_t^{+l}\big)\big)\\ &\times\prod_{\tau=t+1}^{T}\big(1+R_{P\tau}^l\big)\prod_{\tau=1}^{t-1}\big(1+g+y_\tau^{+l}\big)\big(1-\Lambda_\tau^l\big)\\ &-\prod_{t=1}^{T}\big(1-\Lambda_t^l\big)\big(1+g+y_t^{+l}\big)\Bigg]\\ &\Big/\Bigg[\rho\prod_{t=1}^{T}\big(1+r_{ft}^l\big)+\sum_{t=1}^{T}y_t^{-l}\phi(t,T)\prod_{\tau=1}^{t-1}\big(1-\Lambda_\tau^l\big)\big(1+g+y_\tau^{+l}\big)\Bigg]\Bigg\}\end{aligned} \tag{14}$$

$$\text{s.t.}\quad \sum_{i\in\mathcal{U}} x_i = 1, \tag{15}$$

$$\beta R_{Pt}^l - g = y_t^{+l} - y_t^{-l},\quad \text{for } t=1,2,\ldots,T,\ \text{and for all } l\in\Omega, \tag{16}$$

$$R_{Pt}^l = \sum_{i\in\mathcal{U}} x_i r_{it}^l,\quad \text{for } t=1,2,\ldots,T,\ \text{and for all } l\in\Omega. \tag{17}$$

The inverse of the utility function, U^{-1}, of the optimal objective value of this problem is the CEexROE.

3.5. Surrender option

The probability of abandon Λ_t^l is determined from both actuarial events (death) and economic considerations (surrendering the policy). The actuarial component is readily obtained from mortality tables. However, the lapse behavior of policyholders needs to be modeled taking into account the economic incentive to surrender the policy and invest into competing products. This dimension is modeled here.

Modeling the lapse behavior serves as a sensitivity analysis of the model for studying errors introduced due to various sources of model risk. For instance, in recent years many actuaries have pointed out that the aging of the population has introduced a modeling risk in the actuarial framework. The *longevity risk* affects the probability of survival for sectors of the population in their retirement years. Pension fund managers will then face higher liabilities than those planned. On the contrary, life insurance products benefit from longevity risk since the payments due to death are reduced. The modeling of lapse undertaken here is but one example of the additional sources of uncertainty that could be incorporated in the model if data are available.

We discuss here two assumptions about policy lapse which can be embedded into the model.

Fixed lapse: Under this assumption the probability of surrendering the policy (Λ_t) is constant throughout the life of the contract. This assumption is quite realistic. For instance, an analysis of a panel of British households shows that the percentage of lapse is constant over the period 1994–1997 and it averages to 1.4% (see the Personal Investment Authority report, 1999). An estimate from rough data available to us for Italian households indicate a modest lapse rate of the order of 2%.

Variable lapse: Under this assumption the policyholders' decision to surrender their policy is affected by economic factors. For instance, in the analysis of mortgage backed securities (see Kang and Zenios, 1992) prepayment models are calibrated to describe household attitude towards market factors, such as the prevailing mortgage refinancing rates, and social factors such as age of the household and demographics. Similarly we can link the dynamics of Λ_t^l to economic variables. If we assume that lapse is driven by the minimum guarantee level g, then the lapse probability is a function of the spread between g and the rate on other investments offered in the capital markets

$$\Lambda_t^l = f\left(r_{It}^l - g\right), \tag{18}$$

where r_{It}^l is a suitable benchmark of the return offered by competing products; this can be, for instance, the return on the 10-year Government bond index. The surrender probability is now indexed by scenario as it depends on the competitors' rate r_{It}^l. We expect policyholders to surrender their policies when alternative investments provide a return higher than the guarantee g.

Perhaps the most significant factor affecting lapse is the bonus policy followed by the company. Evidence to this is provided for some similar products—single premium deferred annuities—by Asay, Bouyoucos and Marciano (1993). If the insurance company's crediting rate is significantly lower than that of the competition then lapse rates will be high. In participating policies the credit rate is determined by the performance of the portfolio. Thus, an integrative asset and liability management approach is essential in accurately capturing the lapse rates of these products.

Assuming that the competitors offer rates equal to the relevant market benchmark, we express lapse rates in the form

$$\Lambda_t^l = f\left(r_{It}^l - \left(g + \varepsilon_t^{+l}\right)\right). \tag{19}$$

(Recall that $g + \varepsilon_t^{+l}$ is the rate credited to policyholders and it reflects both the guarantee and the bonus policy.) This formula embodies the complex games facing the insurer: large minimum guarantees subdue the effects of the competition but come at a large cost or low CEexROE. This will also be demonstrated in Section 4 where the model is validated.

A convenient general form for function $f(\cdot)$ governing the surrender behavior has been studied by Asay, Bouyoucos and Marciano (1993). In this study the lapse probability is given by

$$\Lambda_t^l = a + b \tan^{-1}\left[m\left(r_{It}^l - i_t^l - y\right) - n\right]. \tag{20}$$

The variable i_t^l is the company's credit rate which can be modeled as a constant (Eq. (18)) or as a variable determined by policy and market performance (Eq. (19)), r_{It}^l is the rate offered by the competitors, and y is a measure of policyholders' inertia in exercising the surrender option. The parameters a, b, m, n are chosen to give lapse rates that fit historically observed data. For instance, the model should fit the lowest and highest lapse rates that have been observed under extremely favorable and unfavorable conditions, and the lapse rates observed when the insurance product was offering the same credit rates as the benchmark.

Figure 3 shows different lapse curves when varying the parameter to fit maximum and minimum values and different average lapse rates. Lapse rates will be, on the average, lower when there are large penalties for early surrender of the policy. The different curves shown in the figure could fit, for instance, the historically observed lapse rates of policies with different surrender charges. We observe that lapse rates may differ substantially when the company is offering a credit rate which is less than the competitors' rate. This situation occurs when assets perform poorly with respect to the rest of the industry. Careful modeling of the lapse behavior is needed in these cases to avoid igniting a vicious circle which could lead to bankruptcy.

3.6. Model extensions

We point out possible extensions of this model. Periodic premia can readily be incorporated into Eq. (9). Bonus policies based on averaging portfolio performance can also be included in the model. The liability equation (2) must be modified to include average

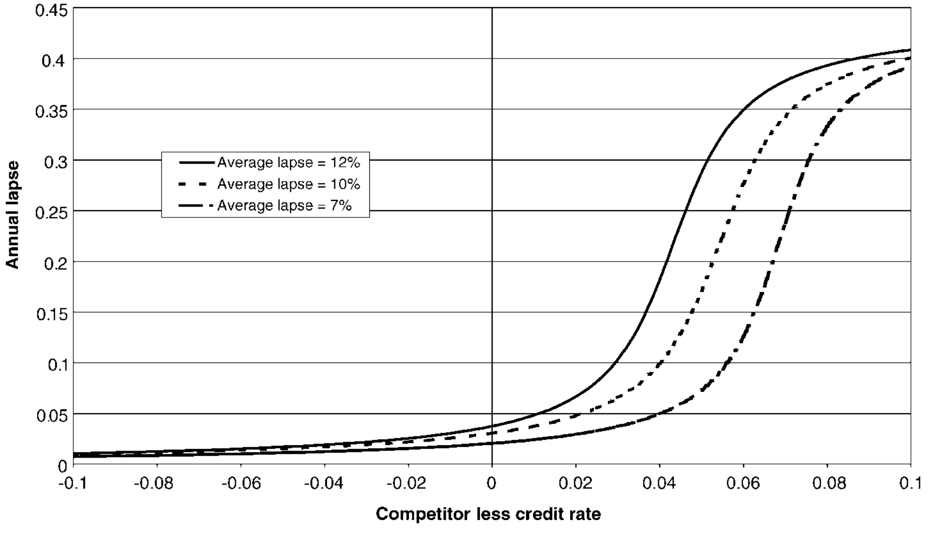


Fig. 3. Typical lapse functions with average lapse rates ranging from 7 to 12%.

portfolio performance over the history of interest (say the last t_h periods) as follows

$$L_t^l = \left(1 - \Lambda_t^l\right) L_{t-1}^l \left(1 + \max\left[\beta \sum_{\tau=t-t_h}^{t} R_{P\tau}^l, g\right]\right), \tag{21}$$

for $t = 1, 2, \ldots, T$, and for all $l \in \Omega$.

Guaranteed rates and bonus rates that are exogenously given functions of time, g_t and β_t, are easy to incorporate. Similarly, we can incorporate liabilities due to lapse, although a lapse model must first be built and calibrated as discussed above. Incorporating participation rates that are functions of the asset returns—as is the case with the UK insurance policies—complicates the model since the participation rate β_t is a variable; see, e.g., Consiglio, Saunders and Zenios (2003, 2006). The split of bonus into reversionary bonus, which is guaranteed, and an investment reserve which is returned as a bonus at maturity, if nonnegative, introduces significant modifications to the model. These issues are discussed in Section 3.7.

The base model developed here funds shortfalls through equity. Extensions to deal with the funding of shortfalls through long- or short-term debt are given in Sections 4.3.1 and 4.3.3, respectively. Furthermore, unlimited access to equity for funding shortfalls is assumed in the base model. We could do away with this assumption by imposing additional constraints, but this would complicate the model rendering it computationally intractable. The probability of insolvency is analyzed through post-optimality analysis in Section 4.3.2, and is used to guide the debt structure in funding shortfalls using a combination of equity and debt.

3.7. Reversionary and terminal bonuses

Some policies use a smoothing mechanism to estimate bonuses, disbursing higher bonuses when market conditions are favorable, and decreasing bonuses when the insurer's portfolio is under-performing. Changes are autoregressive so that big swings are avoided, as those are viewed unfavorably by policyholders. The policies offered in the UK are the best known example with these characteristics. The bonus philosophy of the UK insurers is based on regulatory requirements that bonus distribution should accord with the policyholders' reasonable expectations of the company's behavior (Ross, 1989; Chadburn, 1997).

To satisfy the policyholders' expectations UK insurers offer *reversionary bonuses* that, once announced, are subsequently guaranteed. In addition, they deliver a *terminal bonus*, that is, a function of the excess asset value upon maturity. In general, the reversionary bonus and the guaranteed rates of the UK insurers are lower than those offered by their Italian colleagues. However, policyholders receive the lion's share of any excess asset value, while in the Italian case the insurer's shareholders benefit from the terminal excess asset value. Italian insurers offer a big bird at hand, but nothing in the bush; UK insurers offer a small bird at hand, and ten in the bush.

To model these policies we introduce variable RB_t^l to denote the reversionary bonus disbursed at period t in scenario l. This variable evolves according to the autoregressive equation

$$RB_t^l = 0.5 RB_{t-1}^l + \Delta B_t^l, \tag{22}$$

where the constant 0.5 ensures that policyholders are not too unpleasantly surprised by downward swings of their bonuses, and ΔB_t^l is the change in the bonus. This may be positive or negative and is computed as follows

$$\Delta B_t^l = 0.5 \max\left[\frac{r_{It}^l - g}{1+g}, 0\right] - 0.25 \max\left[\frac{L_t^l - A_t^l}{A_t^l}, 0\right]. \tag{23}$$

The first term on the right of this equation is positive whenever some benchmark return r_{It}^l exceeds the guarantee, otherwise it is zero. The benchmark return is taken in the UK to be the yield on long risk free securities. The second term is positive whenever the asset value is less than the liability value, otherwise it is zero. With this formula the bonus rate is increased whenever the market rates increase, but it is decreased whenever the insurer faces the prospect of insolvency.

The variable dynamics and constraints of policies with smoothed reversionary and terminal bonuses can now be formulated, building on the base model of Section 3.3.

The dynamics of the liability account are given by

$$L_t^l = \left(1 - \Lambda_t^l\right) L_{t-1}^l (1+g)\left(1 + \max\left[RB_{t-1}^l, 0\right]\right), \tag{24}$$

for $t = 1, 2, \ldots, T$, and for all $l \in \Omega$.

Liability payments are exactly as in the base model

$$y_{At}^l = \Lambda_t^l L_{t-1}^l \left(1 + g + y_t^{+l}\right), \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{25}$$

The asset dynamics take into account outflows due to actuarial events but, unlike the base model, there is no cash infusion. The asset value is allowed to go below the liability value and this will have an effect on the reversionary bonus.

$$A_t^l = A_{t-1}^l\bigl(1 + R_{Pt}^l\bigr) - y_{At}^l, \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{26}$$

The equity equation from the base model (8) is split into two equations: One that models the dynamics of the shareholder equity growing at the risk free rate, and one that models shortfalls so that the total shortfall (if any) at maturity can be assessed. The shareholder equity follows the dynamics

$$E_t^l = E_{t-1}^l\bigl(1 + r_{ft}^l\bigr), \quad \text{for } t = 1, 2, \ldots, T, \text{ and for all } l \in \Omega. \tag{27}$$

The lag of assets against the liabilities is given by

$$E_t^{\prime l} = \max\bigl[(1+\rho)L_t^l - A_t^l, 0\bigr], \tag{28}$$

and whenever the lag increases the total shortfall, z_t^l, increases according to the dynamics

$$z_t^l = z_{t-1}^l\bigl(1 + r_{ft}^l\bigr) + \max\bigl[\bigl(E_t^{\prime l} - E_{t-1}^{\prime l}\bigr), 0\bigr], \tag{29}$$

for $t = 1, 2, \ldots, T$, and for all $l \in \Omega$.

With these dynamics the terminal bonus paid to policyholders at maturity T is given by

$$TB_T^l = \gamma \max\bigl[A_T^l - L_T^l, 0\bigr], \tag{30}$$

and the return on equity to shareholders is given by

$$\text{ROE}^l = \frac{A_T^l - L_T^l - TB_T^l}{E_T^l}. \tag{31}$$

4. Model testing and validation

We now turn to the testing of the model. We start first with the application of the traditional portfolio diversification approach based on the mean-variance optimization. We show that the standard application of the mean-variance optimization fails to capture some important characteristics of the problem. There is nothing efficient about efficient portfolios when the nonlinearity of the embedded options is properly accounted for. We show that the novel model based on scenario optimization adds value to the risk management process for these complex insurance products. The value of integrated financial product management is extensively argued in practice, see, e.g., Stulz (1996) but case studies showing that an integrative perspective adds value are scant; see Holmer and Zenios (1995) for some examples.

Second, we show that the model quantifies the tradeoffs between the different targets of the insurance firm: providing the best products for its policyholders, providing the highest excess return to its shareholders, satisfying the guarantee at the lowest possible cost and with high probability. Some interesting insights are obtained on the structure of the optimal portfolios as the tradeoffs vary across the spectrum.

Third, we analyze alternative debt structures whereby the cost of the guarantee is funded through equity or through debt with either long or short maturities.

Fourth, we study some additional features of the model: the effects of the choice of a utility function, the effects of using international asset classes and corporate bonds and the effects of policy surrender options (lapse).

Finally, we will see from the empirical results that the Italian insurance industry operates at levels which are close to optimal but not quite so. There is room for improvement either by offering more competitive products or by generating higher excess returns for the benefit of the shareholders. How are the improvements possible? The answer is found in the comparison of the optimal portfolios generated by our model with benchmark portfolios. We will see that the benchmark portfolios generate tradeoffs in the space of cost of guarantee vs. net excess return on equity that are inefficient. The optimized portfolios lead to policies with the same cost but higher excess return on equity.

The basic asset classes considered in our study are 23 stock indexes of the Milano Stock Exchange, and three Salomon Brother indexes of Italian Government bonds (Appendix B). Italian insurers are also allowed to invest up to 10% of the value of their portfolio in international assets. We report results with the inclusion of international asset classes: the Morgan Stanley stock indices for USA, UK and Japan and the J.P. Morgan Government bond indices for the same countries.

We employ a simple approach for generating scenarios using only the available data without any mathematical modeling, by bootstrapping a set of historical records. Each scenario is a sample of returns of the assets obtained by sampling returns that were observed in the past. Dates from the available historical records are selected randomly, and for each date in the sample we read the returns of all assets classes realized during the previous month. These samples are scenarios of *monthly* returns. To generate scenarios of returns for a long horizon—say 10 years—we sample 120 monthly returns from different points in time. The compounded return of the sampled series is one scenario of the 10-year return. The process is repeated to generate the desired number of scenarios for the 10-year period. With this approach the correlations among asset classes are preserved.

Additional scenarios could also be included, although methods for generating them should be specified. Model-based scenario generation methods for asset returns are popular in the insurance industry—e.g., the Wilkie (1995) model or the Towers Perrin model (Mulvey and Thorlacius, 1998)—and could be readily incorporated into the scenario optimization model. Alternatively, one could use expert opinion or “scenario proxies” as discussed in Dembo et al. (2000).

For the numerical experiments we bootstrap monthly records from the ten year period January 1990 to February 2000. The monthly returns are compounded to yearly returns. For each asset class we generate 500 scenarios of returns during a 10 year horizon ($T =$

120 months). We consider an initial liability $L_0 = 1$ for a contract with participation rate $\beta = 85\%$ and equity to liability ratio $\rho = 4\%$. The model is tested for guarantees ranging from 1 to 15%.

In our experiments we set lapse probabilities to zero and the probability that a policy-holder abandons the policy is the mortality rate which we obtain from the Italian mortality tables. For each model run we determine the net annualized after-tax CEexROE

$$\left(\sqrt[T]{\text{CEexROE}} - 1\right)(1 - \kappa), \tag{32}$$

where κ is the tax rate set at 51%.

4.1. *The value of integrative asset and liability management*

In this section we compare the scenario optimization asset and liability management model with traditional asset allocation using the mean-variance analysis. We demonstrate that the integrative approach adds value to the asset and liability management process.

4.1.1. *Traditional approach using mean-variance asset allocation*

Diversified portfolios of stocks and bonds for an Italian insurance firm are built using the mean-variance optimization. Using the asset classes of the Italian stock and bond markets, we obtain the efficient frontier of expected return vs. standard deviation illustrated in Figure 4. Should an insurance firm offering a minimum guarantee product choose portfolios—based on its appetite for risk—from the set of efficient portfolios? On the same figure we plot each one of the efficient portfolios in the space of shareholder's reward (CEexROE) versus the firm's risk (cost of the guarantee). There is nothing efficient about efficient portfolios when the liability created by the minimum guarantee policy is accounted for. Portfolios from A to G are on the mean-variance frontier that lies below the capital market line. It is not surprising that they are not efficient in the CEexROE vs. cost-of-guarantee space. However, the tangent portfolio G is also inefficient. A more aggressive portfolio strategy is needed in order to achieve the minimum guaranteed return and deliver excess return to shareholder. And still this increasing appetite for higher but risky returns is not monotonic. As we move away from portfolio G towards the most risky portfolio B, we see at first the cost of the guarantee declining and CEexROE improving. But as we approach B the shareholder's value erodes, just as Siglienti found out from his simulations. For these very volatile portfolios the embedded option is deep in-the-money, and shareholders' money are used to compensate for the shortfalls without realizing any excess returns.

This first step of our analysis has shown that it is important to take an integrative view of the asset allocation problem of firms issuing products with guarantees. Properly accounting for the cost of the guarantee is important, if the firm is to avoid unnecessary risk exposures and destroy the shareholder's value.

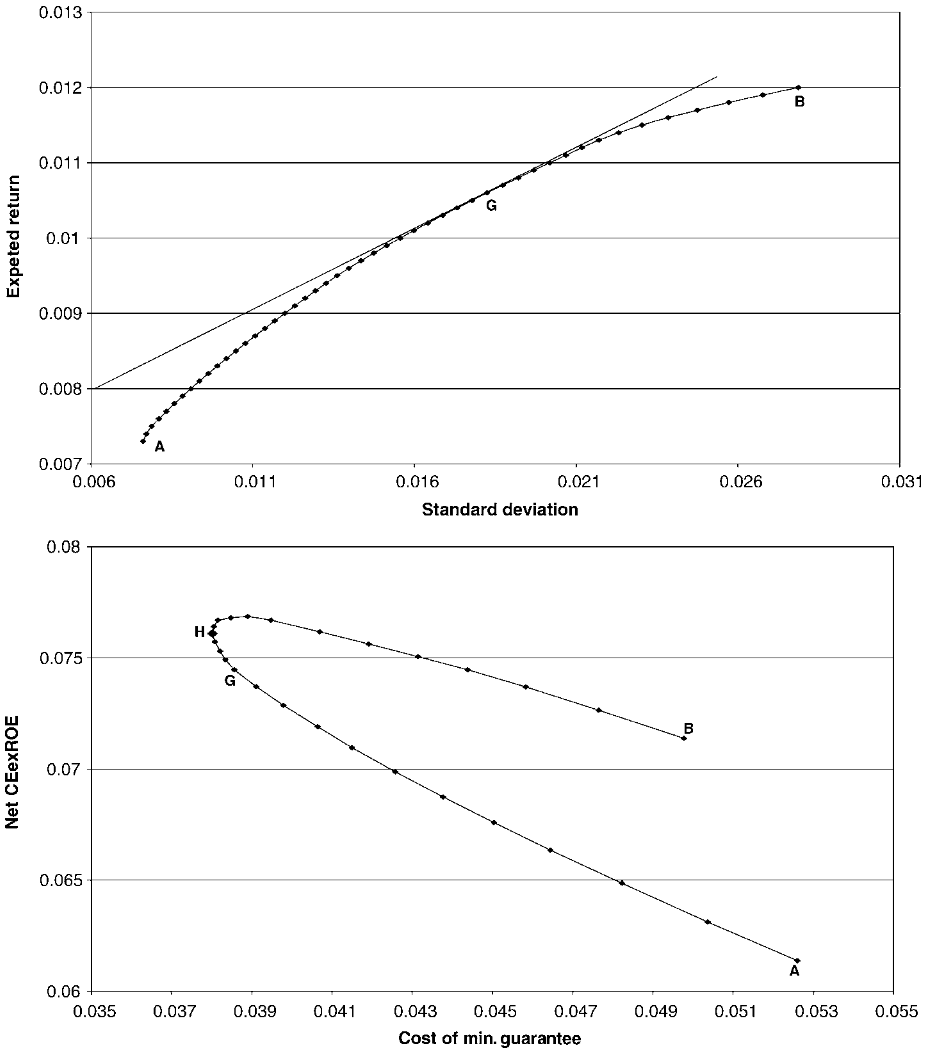


Fig. 4. Mean-variance efficient portfolios of Italian stocks and bonds and the capital market line (top) and the corresponding certainty equivalent excess return of equity (CEexROE) to shareholders vs. cost of the minimum guarantee for each portfolio (bottom).

In a nutshell the management of minimum guarantee products is a balancing act. Too much reliance on bonds and the guarantee is not met. Excessive reliance on stocks and shareholder's value is destroyed.

Is it possible to incorporate the random liability in a mean-variance model, and develop efficient portfolios in the CEexROE vs. cost-of-guarantee space? Unfortunately, the return of the liability depends on the return of the asset portfolio and this is not known without determining simultaneously the structure of the asset portfolio. The return of the liability is endogenous to the portfolio selection model. Furthermore, the liability return has a floor—the minimum guarantee. This creates nonlinearities in the model, and highly asymmetric returns that are not conducive to mean-variance type of modeling. While semi-variance or other risk measures could be used to handle the asymmetric returns, the problem that the return of the liability is endogenous to the portfolio selection model remains. The integrative model developed earlier is essential.

4.1.2. Integrative asset and liability modeling

The results in Figure 5 show the tradeoff between upside potential versus the downside risk achieved when using the models of this paper. Each point on this figure corresponds to an optimal asset portfolio for each level of minimum guarantee. On the same figure we plot the tradeoff between CEexROE and cost of the guarantee from the portfolios of Figure 4. We see that even portfolio H is dominated by the portfolios obtained by an integrative model. The traditional approach of portfolio diversification—Figure 4 (top)—followed by a post optimality analysis to incorporate the minimum guarantee liability and its cost—Figure 4 (bottom)—yields suboptimal results. The integrative approach adds value. The analysis carried out here with market data for a real policy shows that the added value can be substantial.

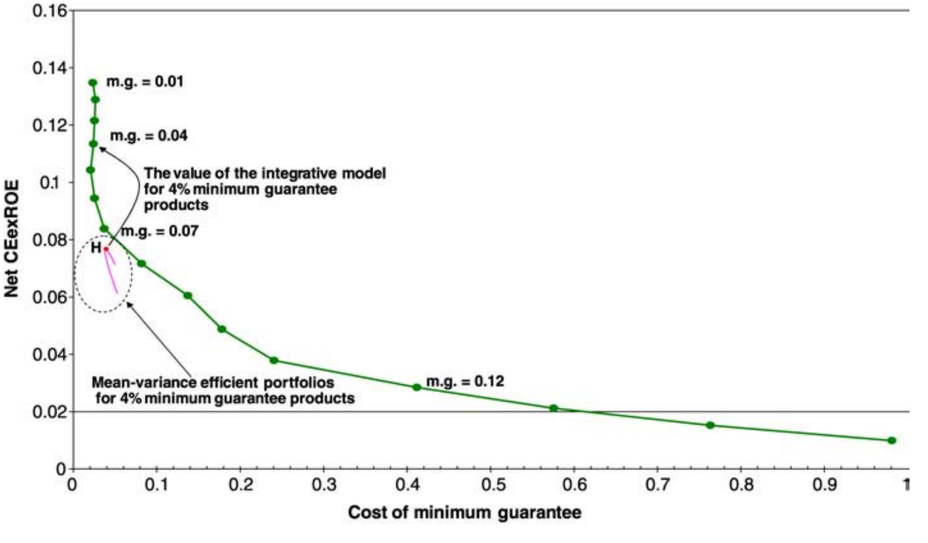


Fig. 5. Certainty equivalent excess return of equity (CEexROE) to shareholders vs. cost of the minimum guarantee for the integrated portfolios at different levels of minimum guarantee, and for the mean-variance efficient portfolios (insert).

4.2. Analysis of the tradeoffs

We now turn to the analysis of the tradeoffs between the guaranteed rate of return offered to policyholders and the net CEexROE on shareholders' equity. This is shown in Figure 6, where the optimal asset allocation among the broad classes of bonds and stocks is also shown for the different guaranteed returns.

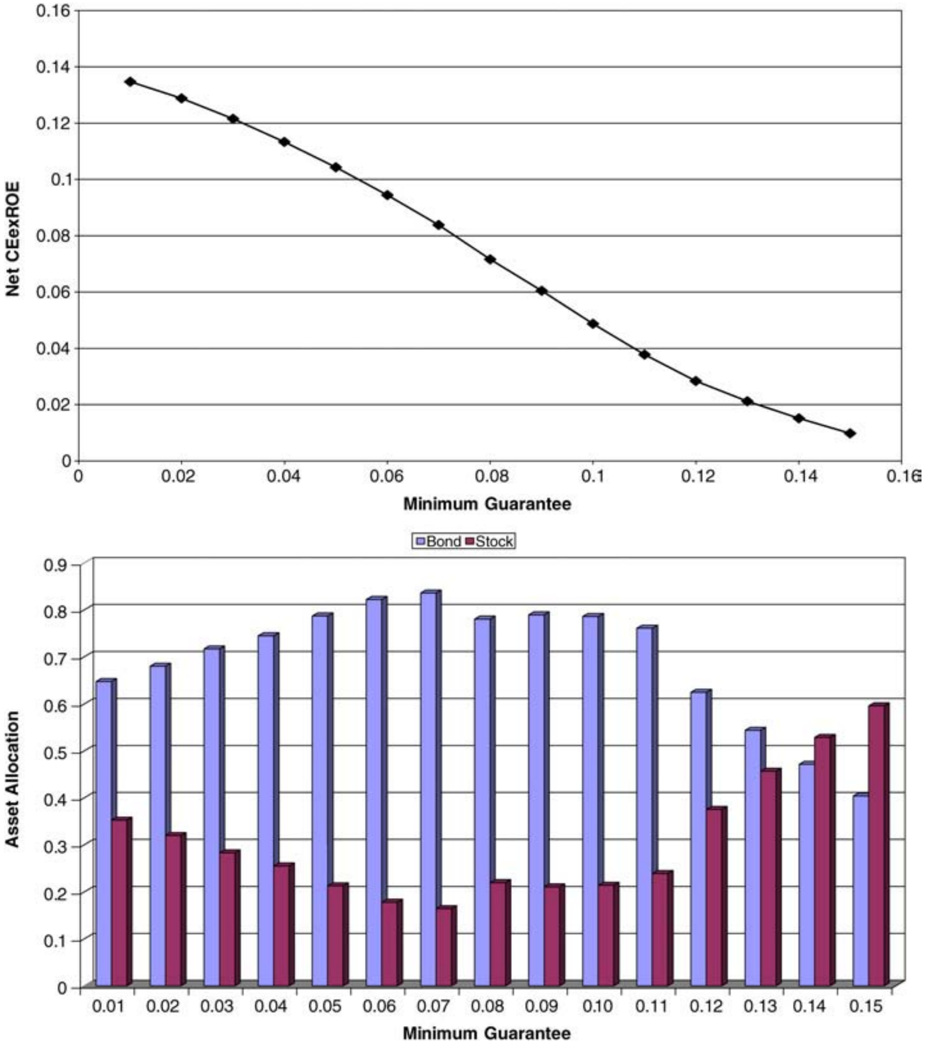


Fig. 6. Net CEexROE (annualized) for different levels of the guarantee (top) and the corresponding broad asset allocations (bottom).

At first glance the portfolio structures appear puzzling. One expects that as the guarantee increases the amount of stock holdings should grow. However, we observe that for low guarantees (less than 7% for the market sectors we consider) the holdings in stock increase with lower guarantees. For low g the embedded option is far out of the money, even when the asset portfolio is mostly equity and very volatile. The asset allocation strategy maximizes CEexROE by taking higher risks in the equities market. A marginal increase of the shortfall cost allows higher CEexROE. This is further clarified in Figure 5, showing the tradeoff between cost of the guarantee and net annualized CEexROE. At values of g less than 7% the option embedded in the liability is out-of-the-money and any excess return is passed on to the shareholders thus improving CEexROE. As the guarantee increases above 7%, the option goes deeper into the money, the cost of the guarantee increases significantly and CEexROE erodes. Note from Figure 6 that higher values of the guarantee must be backed by aggressive portfolios with high equity content, but in this case the portfolio volatility is not translated into high CEexROE for the shareholders but into higher returns for the policyholders. This is consistent with the conclusion of Siglienti (2000) that excessive investments in equity destroy shareholder's value. However, for the guaranteed rates of 3 to 4% offered by Italian insurers it appears that the optimal portfolios consist of 20 to 25% in equities, as opposed to 15% that was obtained by Siglienti using simulations. This discrepancy could be, in part, due to the data of scenario returns used in his study and ours. However, it may also be due to the fact that with the scenario model developed here the portfolio composition is optimized.

4.3. Analysis of alternative debt structures

So far we have assumed that the cost of the guarantee is covered by shareholders. It is possible, however, that such costs are charged to policyholders or funded by issuing debt. (Note that for mutual insurance firms the policyholders are the shareholders so the point of who pays for the cost is mute. However, the issue of raising debt remains.) In either case there are advantages and disadvantages. In particular, if we let the policyholder assume the total cost, we run the risk of not being competitive, loose market share, and experience increased lapse. If we issue debt, we are liable for interest payments at the end of the planning horizon which could reduce our final return. Furthermore, companies face leverage restrictions. It may not be possible to cover all the cost of the guarantee by issuing debt because it will increase the leverage of the company beyond what is allowed by the regulators or accepted by the market.

Another important point in pursuing this question concerns the maturity of the issued debt. We start by considering long-term debt.

4.3.1. Long-term financing of shortfalls

To issue long-term debt we determine the amount of cash that we need to borrow in order to cover, with a certain probability, future expenditures due to shortfalls over all scenarios. If we indicate by α a confidence level we are searching for the α-percentile,

O_G^α, such that the cost of the guarantee O_G^l in scenario l satisfies

$$P\left(O_G^l \geqslant O_G^\alpha \mid l \in \Omega\right) = \alpha. \tag{33}$$

The cost of the guarantee in scenario l is given by Eq. (13) as

$$O_G^l = \frac{E_T^l}{\prod_{t=1}^T (1 + r_{ft}^l)} - \rho L_0. \tag{34}$$

Note that O_G^α need not to be raised through the issue of debt only. It is just the reserves needed to fund shortfalls. Strategic considerations will subdivide O_G^α among policyholder charges, C_G, issue of debt or direct borrowing from money markets, D_G, and/or equity supplement, E_S. Thus, we have

$$O_G^\alpha = C_G + D_G + E_S. \tag{35}$$

Given the debt structure implied in (35) we determine the final income I_T^l, for each scenario $l \in \Omega$, as

$$I_T^l = A_T^l - L_T^l - D_G(1 + r_f + \delta)^T + (C_G - J_S) \prod_{t=1}^{T} \left(1 + r_{ft}^l\right), \tag{36}$$

where J_S are the fixed costs (in percentage of the initial liability) and δ is a spread over the risk free rate so that $r_f + \delta$ is the borrowing interest rate. Debt structures for which at $I_T^l < 0$ for some scenario $l \in \Omega$ should be discarded as leading the firm into insolvency, even if the probability of such events is very low.

The net Return-on-Equity (ROE) corresponding to a given debt structure in each scenario is given by

$$\mathrm{ROE}^l = \frac{I_T^l (1 - \kappa)}{\rho L_0 + E_S}. \tag{37}$$

This is not the *ex ante* excess return on equity optimized with the base model, but the *ex post* realized total return on equity achieved when the structure of debt has also been specified. This measure can be used to analyze the probability of insolvency when the cost of the guarantee is funded by shareholders instead of being charged, at least in part, to policyholders.

We report some results with the analysis described here. Tables are generated to study the tradeoffs between leverage, policyholder charges, and shareholder returns. Table 3 summarizes data that assist the decision maker to take a position according to her strategic views and constraints. If no entries are displayed these choices cannot be implemented, either because some I_T^l are negative (this occurs when charges to policyholders are very low and high debt levels yield a negative final income), or because the amount of money necessary to cover shortfalls is fully covered by the policyholder charges. This implies a negative debt level at maturity of the product.

For example, by choosing a leverage level equal to 0.5, the highest yearly net CEROE is 0.183. Note that, if the firm wishes to achieve higher performance level, the leverage

Table 3

Net CEROE for different combinations of leverage and policyholder charges. The table is built for a guarantee $g = 4\%$ at a confidence level $\alpha = 1\%$

Leverage levels	Policyholder charges									
	0	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0.121125	0.124595	0.128295	0.132256	0.136515	0.141118	0.146123	0.151602	0.15765	0.164391
0.125	0.123946	0.127684	0.131656	0.135891	0.14043	0.145317	0.150612	0.156387	0.16274	0.169795
0.25	0.126654	0.13064	0.13486	0.139346	0.144137	0.14928	0.154834	0.160873	0.167495	0.174827
0.375	0.12926	0.133474	0.137923	0.142638	0.147659	0.153033	0.158821	0.165097	0.17196	0.179538
0.5	0.13177	0.136197	0.140857	0.145783	0.151014	0.156599	0.162599	0.169089	0.176169	0.183968
0.625	0.134193	0.138817	0.143673	0.148794	0.154219	0.159997	0.16619	0.172875	0.180151	0.188151
0.75	0.136533	0.141343	0.146381	0.151682	0.157285	0.163242	0.169612	0.176475	0.183932	0.192114
0.875	0.138798	0.143781	0.148989	0.154458	0.160227	0.166348	0.172882	0.179909	0.18753	0.195879
1	0.140991	0.146137	0.151505	0.15713	0.163053	0.169327	0.176013	0.183191	0.190964	0.199468
1.125	0.143118	0.148417	0.153935	0.159706	0.165774	0.172189	0.179016	0.186335	0.19425	0.202896
1.25	0.145182	0.150626	0.156285	0.162194	0.168396	0.174944	0.181903	0.189353	0.197399	0.206177
1.375	0.147188	0.152769	0.15856	0.164599	0.170928	0.177601	0.184682	0.192255	0.200423	
1.5	0.149138	0.154849	0.160766	0.166927	0.173375	0.180165	0.187362	0.19505	0.203333	
1.625	0.151037	0.156871	0.162907	0.169183	0.175744	0.182644	0.18995	0.197745		
1.75	0.152886	0.158837	0.164986	0.171371	0.178039	0.185044	0.192452	0.200349		
1.875	0.154688	0.160751	0.167007	0.173497	0.180265	0.187369	0.194875			
2	0.156446	0.162616	0.168974	0.175562	0.182427	0.189624	0.197222			
2.125		0.164433	0.17089	0.177572	0.184528	0.191814				
2.25		0.166207	0.172757	0.179529	0.186571	0.193942				
2.375		0.167938	0.174577	0.181435	0.188561					
2.5		0.16963	0.176354	0.183294	0.190499					
2.625			0.178089	0.185108						
2.75			0.179785	0.186879						
2.875			0.181443							
3			0.183064							
3.125										
3.25										
3.375										

Table 4
The relation between net CEROE, policyholder charges and guarantee

Policyholder charges	Minimum guarantee							
	0.01	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**
0	0.144564	0.139163	0.135832	0.13177	0.130433	0.120909	0.110348	0.099402
0.01	0.148057	0.142648	0.139726	0.136197	0.136011	0.126397	0.115193	0.102442
0.02	0.151703	0.146281	0.143803	0.140857	0.141965	0.132226	0.120278	0.105562
0.03	0.155517	0.150077	0.148086	0.145783	0.148361	0.138457	0.125641	0.10877
0.04	0.15952	0.154056	0.152599	0.151014	0.155289	0.145166	0.131326	0.112075
0.05	0.163732	0.158239	0.157375	0.156599	0.162863	0.152453	0.13739	0.115487
0.06	0.168182	0.162651	0.162452	0.162599	0.171238	0.16045	0.143903	0.119017
0.07	0.1729	0.167323	0.167876	0.169089	0.180626	0.169337	0.150957	0.122676
0.08	0.177925	0.172291	0.173703	0.176169	0.191338	0.17937	0.15867	0.126479
0.09	0.183304	0.177599	0.180007	0.183968		0.190924	0.167202	0.130443
0.1	0.189093	0.183304	0.18688	0.192664			0.176778	0.134586

The table is built with confidence level $\alpha = 1\%$ and leverage (debt-to-equity ratio) equal to 0.5.

should also increase. Also, observe the inverse relation between leverage and policyholder charges. The greater the amount we charge to the policyholder, the lower is the leverage required to achieve a given annualized net CEROE.

The model can generate similar tables to study the many interactions of endowment with guarantee. For example, we could be interested in investigating the effect of different guarantee levels to the policyholder charges and yearly returns. We first estimate, at a given confidence level α, the cost of the guarantee O_G^α, and then apportion this cost to policyholders (C_G in Eq. (35)) and fund the rest through debt or equity surcharge. Depending on C_G we observe a change in the CEROE to shareholders. Table 4 shows this relationship. We observe the same behavior we had seen between $\overline{O}_G$ and net CEexROE in Figure 5. The model chooses more aggressive strategies for low g because it is then possible to achieve higher levels of CEexROE at little cost. Recall that we are working with percentiles and the impact of aggressive strategies is much more evident on the tails. When the guarantee is low at $g = 0.01$ we need higher policyholder charges to reach the highest return, while for $g = 0.05$ lower charges are required.

The results in Table 3 should be examined taking into account a measure of risk associated with the CEROE of every combination of policyholder charges and leverage level. The probability that excess value per share, P_{EVS}^-, will become negative is a measure of risk of the CEROE, and these probabilities corresponding to Table 3 are shown in Table 5. Observe that the upper-left entry has a P_{EVS}^- equal to 0.58. This means that there is a 58% chance that the present value of the final equity is less than the amount invested today by the shareholders, even though the net CEROE is acceptable (12%). This position is risky. The reason why this position is quite risky is due to the fact that we are asking our shareholders to fund the total α-percentile cost of the guarantee. No charges are passed on to policyholders.

Table 5
Relationship between P^{-}_{EVS}—the probability that excess value per share will fall below zero—leverage and policyholder charges

Leverage levels	Policyholder charges									
	0	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0.58	0.522	0.462	0.4	0.344	0.278	0.208	0.148	0.096	0.042
0.125	0.534	0.478	0.416	0.366	0.302	0.242	0.172	0.112	0.072	0.02
0.25	0.508	0.444	0.394	0.338	0.274	0.212	0.15	0.1	0.06	0.012
0.375	0.476	0.416	0.368	0.306	0.252	0.188	0.134	0.092	0.042	0.012
0.5	0.444	0.396	0.346	0.284	0.226	0.162	0.118	0.076	0.032	0.006
0.625	0.418	0.374	0.322	0.266	0.212	0.152	0.106	0.068	0.022	0.004
0.75	0.404	0.366	0.304	0.258	0.198	0.144	0.098	0.056	0.016	0.002
0.875	0.4	0.354	0.286	0.234	0.184	0.136	0.092	0.05	0.012	0.002
1	0.378	0.33	0.28	0.224	0.162	0.124	0.088	0.04	0.012	0.002
1.125	0.37	0.318	0.266	0.216	0.156	0.114	0.078	0.036	0.008	0.002
1.25	0.364	0.31	0.264	0.208	0.146	0.108	0.074	0.032	0.008	0.002
1.375	0.356	0.296	0.254	0.2	0.146	0.104	0.07	0.026	0.004	
1.5	0.35	0.286	0.24	0.196	0.142	0.098	0.062	0.026	0.004	
1.625	0.332	0.282	0.234	0.188	0.136	0.096	0.06	0.02		
1.75	0.322	0.276	0.224	0.178	0.132	0.094	0.054	0.016		
1.875	0.316	0.266	0.22	0.162	0.126	0.092	0.052			
2	0.314	0.266	0.214	0.156	0.122	0.086	0.05			
2.125		0.264	0.214	0.15	0.118	0.084				
2.25		0.264	0.208	0.148	0.116	0.08				
2.375		0.26	0.202	0.146	0.112					
2.5		0.244	0.202	0.146	0.11					
2.625			0.198	0.146						
2.75			0.196	0.144						
2.875			0.196							
3			0.19							
3.125										
3.25										
3.375										

The table is built for a guarantee $g = 4\%$ and confidence level $\alpha = 1\%$.

4.3.2. Insolvency risks

So far we analyzed alternative decisions based only on the net CEexROE and market constraints (policyholder charges, leverage, etc.). Our analysis is missing a measure of risk of the ROE. It is not yet clear how alternative guarantees and debt allocations according to Eq. (35) affect the risk of ROE in Eq. (37). One could argue that the risk aversion of the decision maker is embedded in the utility function of the optimization model. This is true, but the utility function was used only to guide decisions on the asset side, and the estimation of net total CEROE from (37) does not incorporate risk aversion when choosing a debt structure. Furthermore, the utility function ensures the solvency of the fund by covering shortfalls with infusion of equity. However, under certain conditions no external sources of equity will be available. The analysis we carry out here compensates for these omissions. It considers the risk of insolvency when structuring the issue of debt, thus incorporating risk aversion in structuring the debt in addition to structuring the asset portfolio.

Define $\overline{R}_I$ as the expected excess return over the risk free rate for this line of business and $\bar{r}_f$ as the expected risk free rate. The rate at which we must discount the final income I_G^l is given by $R_\mu = \bar{r}_f + \overline{R}_I$. For our shareholders I_G^l represents the value of the equity at the end of the planning period and they are willing to stay in this business if the discounted value of this equity is not less than the initial capital invested. The shareholders will keep their shares if the *Excess Value per Share* (*EVS*) is greater than zero with a high probability. Recalling that the initial amount of equity is $\rho L_0 + E_S$ (E_S could be equal to zero) the *EVS* in each scenario is given by

$$EVS^l = \frac{I_G^l\,(1-\kappa)}{(1+R_\mu)^T} - (\rho L_0 + E_S). \tag{38}$$

The risk related to a specific debt allocation is given by the probability that *EVS* is less than zero, i.e., $P_{EVS}^- = P(EVS^l < 0 \mid l \in \Omega)$. This is the probability of insolvency and can be determined by calculating the EVS^l for each $l \in \Omega$, order from the lowest to the highest, and look for the rank of the first EVS^l that is negative, i.e.,

$$P_{EVS}^- = \frac{\text{rank}(EVS^l < 0)}{N}. \tag{39}$$

The *EVS* can be used to determine the amount of policyholder charges required to make P_{EVS}^- equal to a given confidence level. Recall that I_G^l, and consequently EVS^l, is a function of C_G, E_S, and D_G. If we fix E_G then I_G is a function of C_G (D_G is determined from Eq. (35)). Through a linesearch we can determine C_G^* such that

$$P\big[EVS(C_G^*) < 0\big] = \alpha. \tag{40}$$

In our experiments we set $\overline{R}_I = 6\%$ and the probability of insolvency $\alpha = 1\%$. Figure 7 shows the results of the linesearch which solves Eq. (40) for different values of equity supplement E_S. We observe that for guarantees higher than 6% the CEROE increases. How is it possible that higher guarantees can yield higher returns? The puzzle is resolved if we note that the increase in returns is accompanied by a significant increase of policyholder charges. The increases in the policyholder charges fund the guarantee and preserve equity from falling below its present value.

In practice significant increase in policyholder charges would be unacceptable, and would lead to increased lapses. Our analysis can be used as a demarcation criterion between "good" and "bad" levels of the guarantee. For instance, the Italian insurance industry offers products with guarantees in the range 3 to 4%. Our analysis shows that they could consider increasing the guarantee up to 6% without significant increase of

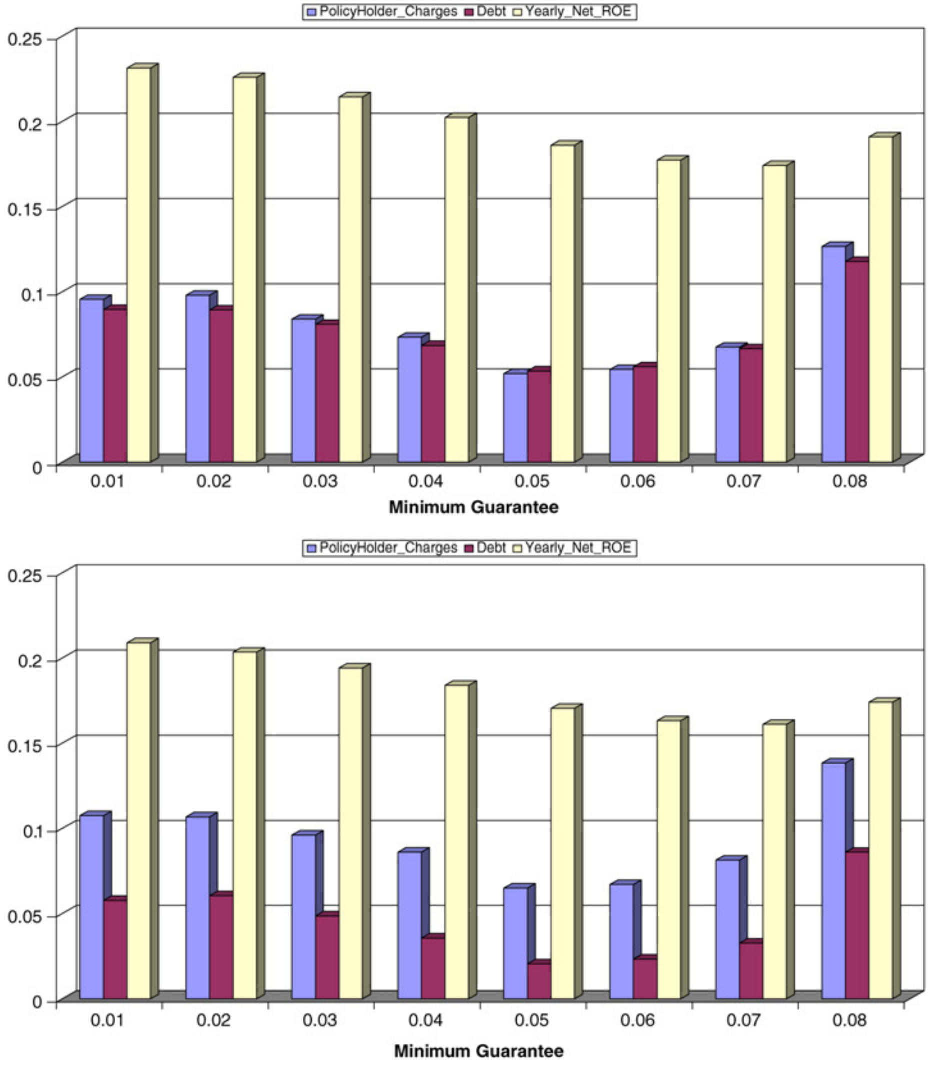


Fig. 7. The levels of policyholder charge, debt and net CEROE such that the probability of insolvency is $P[EVS(C_G^*) < 0] = 1\%$, for equity supplement $E_S = 0$ (top) and $E_S = 0.02$ (bottom).

charges to policyholders or reduction of CEROE. (One may justify the difference from the operating guarantee of 4% to the peak optimized value of 6% as the cost of running the business. If so this cost is high.) For guarantees above 6% we note a substantial increase to policyholder charges at a marginal improvement in CEROE, and this is clearly unacceptable to both policyholders and shareholders.

4.3.3. Short term financing of shortfalls

To this point the analysis has determined the cost of the shortfalls O_G^{α} and funded it through a combination of debt D_G, charges to policyholders C_G, and equity E_S. Now, let us fix policyholder charges and equity and let the debt fluctuate according to the shortfall O_G^l realized in each scenario. Thus we consider funding part of the shortfall through short-term financing. Instead of issuing a bond for a notional equal to D_G and maturity T, we will borrow money when a shortfall occurs. The debt for each scenario is given by

$$D_G^l = O_G^l - C_G - E_S. \tag{41}$$

We assume that it is possible to borrow money at a spread δ over the risk-free rate. The definition of the final income becomes

$$I_T^l = A_T^l - L_T^l - D_G^l \prod_{t=1}^{T} \left(1 + r_{ft}^l + \delta\right) + (C_G - J_S) \prod_{t=1}^{T} \left(1 + r_{ft}^l\right). \tag{42}$$

We can apply the analysis of the previous section to determine policyholder charges C_G, and estimate the distribution of D_G^l. We solve Eq. (40) and display in Figure 8 the C_G^* for different levels of the guarantee and for $\delta = 2\%$. Note that policyholder charges C_G^* are substantially lower than those obtained by solving (40) in the previous section as reported in Figure 7. This is expected as short-term financing of the cost in a dynamic strategy, as opposed to the fixed strategy of issuing long-term debt. These findings are

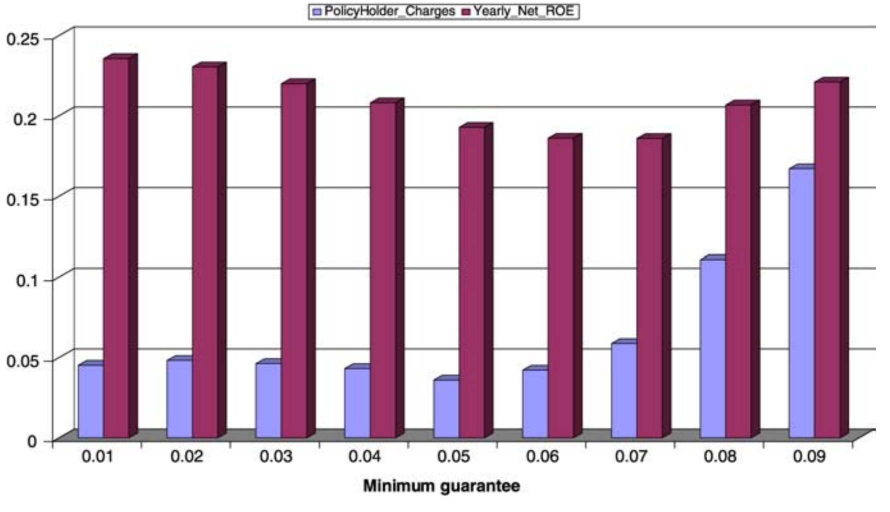


Fig. 8. The levels of policyholder charges and net CEROE for different guarantee such that $P[EVS(C_G^*) < 0] = 1\%$.

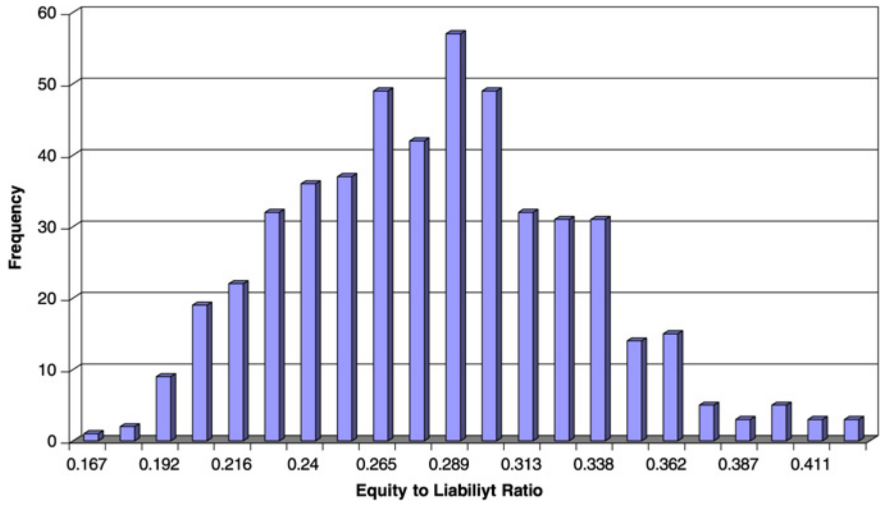


Fig. 9. Distribution of equity-to-liability ratio at the end of the planning horizon for a guarantee of 5%.

consistent with the comparison of the two reserving methods in Boyle–Hardy. Since D_G^l is scenario-dependent, it compensates for those scenarios with high shortfalls, while it is low (or null) for those scenarios with low shortfalls.

4.4. The view from the regulator's desk

We show in Figure 9 the distribution of the equity to liability ratio (cf. Eq. (10)) for a guarantee of 5%. Similar figures were obtained for guarantees ranging from 1 to 10%. This figure shows that for different values of the guarantee the minimum ratio of equity to liability is greater than the regulatory requirement. For the type of policies analyzed here using a logarithmic utility function, and for the scenarios sampled from the past ten bullish years, the regulatory constraint is satisfied without explicitly including it in the model.

4.5. Additional model features

We study now some additional features of the model, namely the effects of the choice of a utility function, the effects of international diversification and investments in corporate bonds, and the effects of the policy surrender option.

4.5.1. Choice of utility function

The decision maker's risk aversion specifies unique asset portfolio to back each guaranteed policy. Clearly increased risk aversion will lead to more conservative portfolios

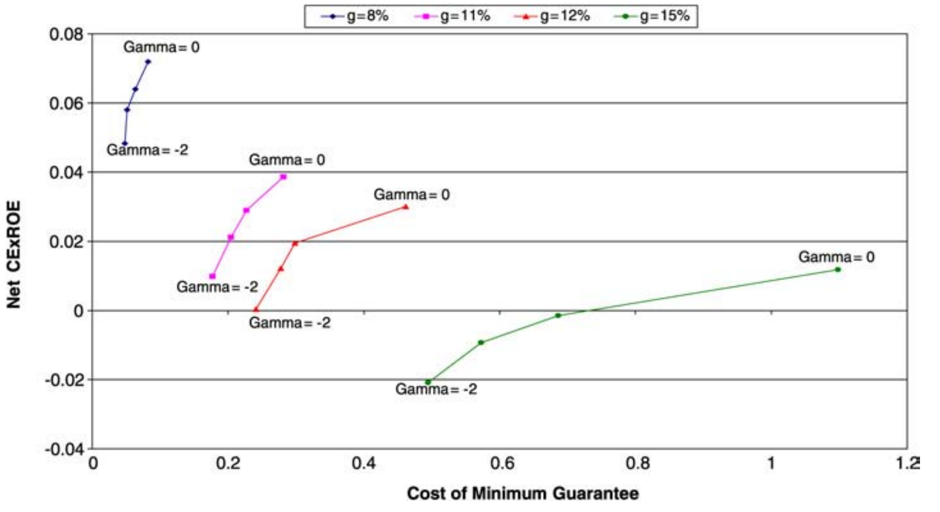


Fig. 10. Tradeoff of CEexROE against cost of the guarantee with varying risk aversion for target guarantees 8 (left), 11, 12 and 15% (right).

with higher contents of fixed income. The result will be a simultaneous reduction in both the CEexROE to shareholders and the cost of shortfalls required to fund the policy. Figure 10 illustrates the tradeoff as the risk aversion parameter γ varies from 0 (base case) to -2 (increased risk aversion) for five different target guarantees.

For low target guarantees we note that an increased appetite for risk results in higher CEexROE, with only a marginal increase in cost of the guarantee. For higher target guarantees (e.g., 15%) we note a substantial increase in the cost of the guarantee as the embedded option goes deep in the money when we increase the risk tolerance and invest into volatile assets. These results confirm our expectations on model performance and are consistent with the results of Figure 5. The model allows users to generate efficient tradeoffs that are consistent with the contractual obligations and the firm's risk tolerance.

4.5.2. *International diversification and credit risk exposures*

We extended the analysis to incorporate other assets permitted by regulations such as corporate bonds and international sovereign debt. Italian insurers are allowed to invest up to 10% of the value of their portfolio in international assets. We run the base model for a guarantee of 4%, and allowing investments in the Morgan Stanley stock indices for USA, UK and Japan and the J.P. Morgan Government bond indices for the same countries. Figure 11 illustrates the tradeoff of CEexROE against cost of the guarantee for international portfolios and portfolios with credit risky securities. The internationally diversified portfolio achieves CEexROE of 0.14 at a cost of the guarantee of 0.02. By

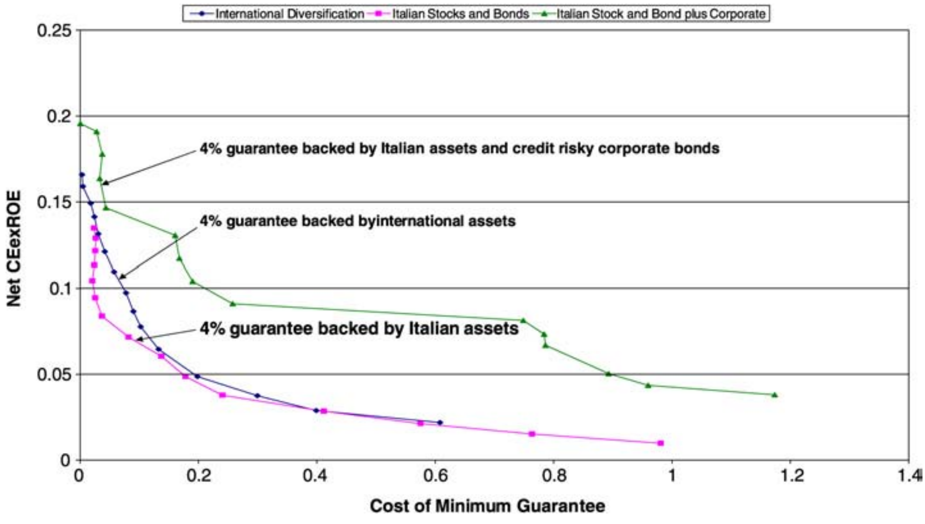


Fig. 11. Tradeoff of CEexROE against cost of the guarantee for internationally diversified portfolios and portfolios with exposure to the corporate bond markets.

contrast, we note that domestic investments in the Italian markets fund the guarantee at the same cost but yield a CEexROE of only 0.11. Similarly, investments in the US Corporate bond market improve further the CEexROE to 0.16, but at an increase of the cost to 0.033.

4.5.3. *Impact of the surrender option*

In our testing so far we took into account only the actuarial risk of the liabilities. Using the various assumptions on lapse behavior discussed in Section 3.5 we study the effect of the surrender option on the cost of the guarantee.

Figure 12 (top) illustrates the effect of lapse on the cost of the guarantee for different levels of the minimum guarantee. It is worth noting that the difference between no lapse at all and fixed lapse is significant for high levels of the minimum guarantee ($g \geqslant 7\%$). This difference is less evident when lapse is modeled as in Eq. (20).

Figure 12 (bottom) illustrates effects of lapse on the net CEexROE. Again, differences are more evident when we switch from no lapse to fixed lapse. It is worth commenting on the effect of lapse rates on the cost of the guarantee and the net CEexROE, over a range of minimum guarantees. Differences in the net CEexROE are observed for low minimum guarantees, say $g \leqslant 6\%$. On the contrary, the alternative lapse assumptions yield substantially different costs for the guarantee for $g > 6\%$. This effect can be explained in view of the option embedded in the policy (see Grosen and Jørgensen, 2000). For low levels of the minimum guarantee the option is almost always out of the money and fixed lapse will depress the net CEexROE through a constant surrender

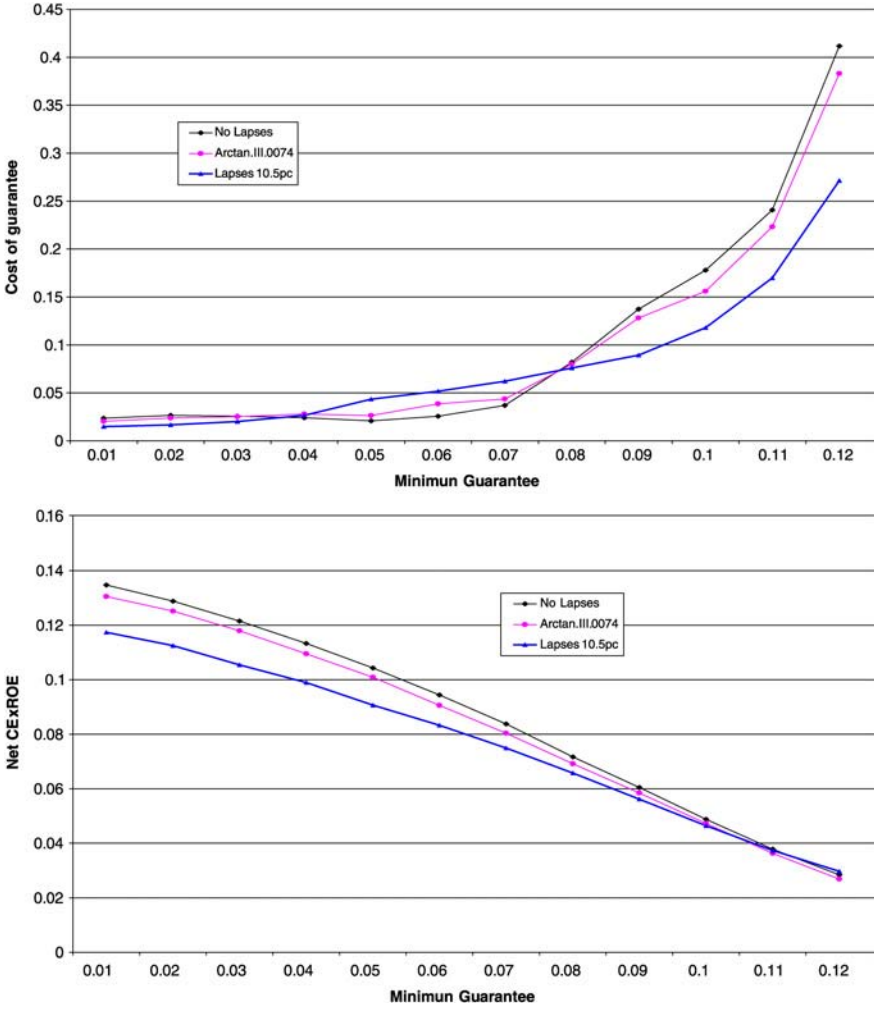


Fig. 12. The effect of lapse on cost of the minimum guarantee (top). The effect of lapse on net CEexROE (bottom).

of policies (see Eq. (6)). For larger values of the minimum guarantee, the insurance company will benefit from lapses since shortfalls are more likely and any lapsed policies relieve the company, in part, from shortfall.

4.6. *Benchmarks of Italian insurance policies*

In order to assess the effectiveness of the model in practice we compare the optimal portfolios with industry benchmarks. We take as benchmark a set of portfolios with a

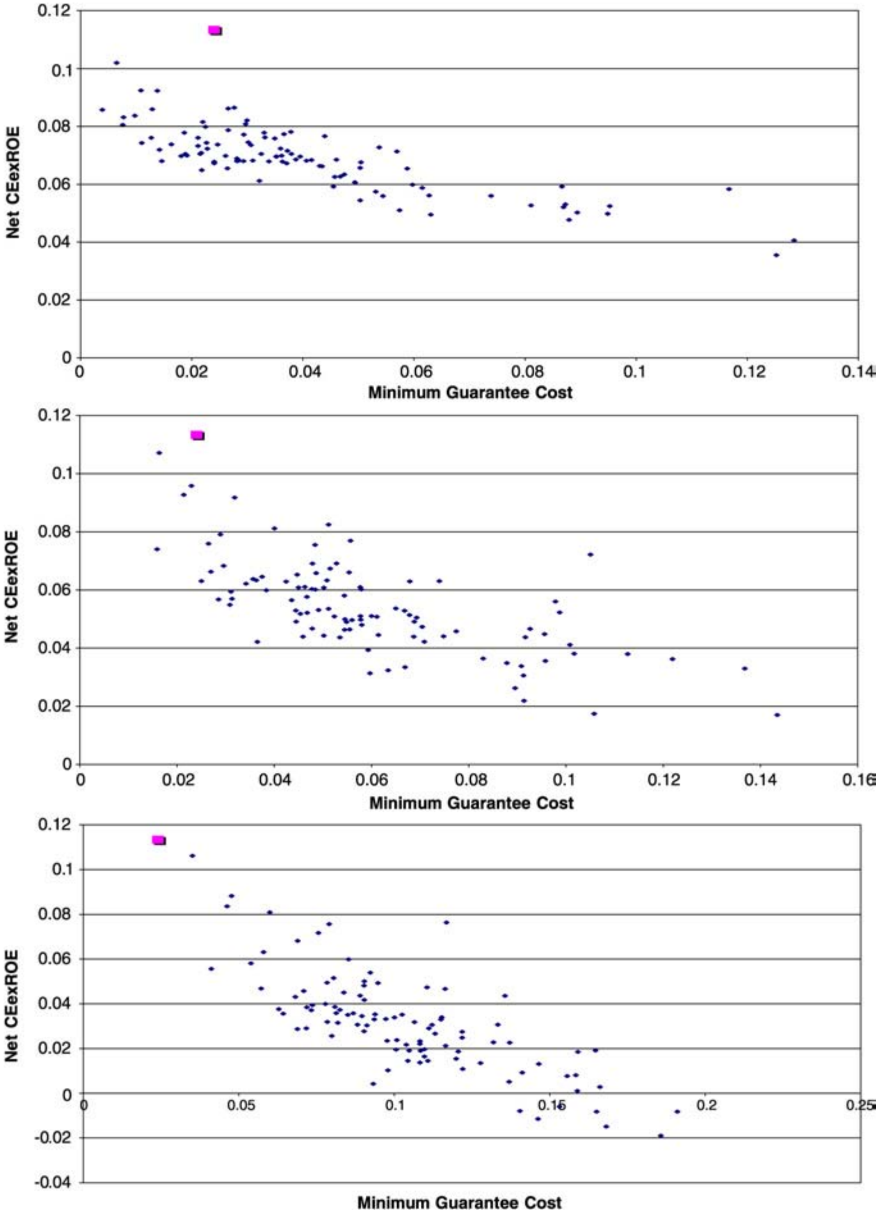


Fig. 13. Performance of benchmark portfolios (diamonds) against the optimized portfolio (square) for $g = 4\%$. Asset allocation for the benchmark portfolios is set to 90/10 (bonds/stocks), 80/20, and 70/30, respectively, from top to bottom.

fixed broad asset allocation between bonds and stocks, and random allocation among specific assets. In order to be consistent with the usual fixed-mix strategies followed by the Italian industry, we set the broad asset allocation between bonds and stocks to 90/10, 80/20, and 70/30. The results of this experiment are reported in Figure 13. Note that the optimized portfolios always dominate the benchmark portfolios in the space of cost-of-guarantee vs. CEexROE. These results justify further the integrative approach taken in this paper, whereby the insurance policy is analyzed *jointly* with the asset allocation decision instead of being analyzed for an *a priori* fixed asset portfolio. Further improvements are possible with an internationally diversified portfolio and with some exposure to credit risky securities, as analyzed in Section 4.5.2.

The results of this section are in general agreement with the current practices of Italian insurers. However, the optimized results suggest that improved policies and associated asset strategies are still possible. In particular, the findings show that the Italian insurers could increase the equity exposure of their portfolio from 20%, which is the current practice, up to 25% to 30%—see the optimal asset allocation corresponding to minimum guaranteed return $g = 3\%$ in Figure 5. This is also evident from Figure 13 where we observe that some random portfolios from the 70/30 asset allocation are closer to the optimized portfolios.

5. Conclusions

This chapter has, first-most, demonstrated that an integrative approach to the management of assets and liabilities for insurance products with guarantees and bonuses adds value. Asset structures generated with an integrative approach for specific insurance policies are efficient, as opposed to asset strategies developed in a non-integrated model.

Several interesting conclusions can be drawn from the use of the model on data from the Italian insurance industry. First, we have quantified the tradeoffs between the different targets of the insurance firm: providing the best products for its policyholders, providing the highest excess return to its shareholders, satisfying the guarantee at the lowest possible cost and with high probability. Some interesting insights are obtained on the structure of the optimal portfolios. In particular, we observe that too little equity in the portfolio and the insurer cannot meet the guarantee, while too much equity destroys shareholder value.

Second, we have analyzed different debt structures whereby the cost of the guarantee is funded through equity or through debt with either long or short maturities. The effects of these choices on the cost of the guarantee and on the probability of insolvency can be quantified, thus providing guidance to management for the selection of policies.

Third, we have seen from the empirical analysis that Italian insurers operate at levels which are close to optimal but not quite so. There is room for improvement either by offering more competitive products or by generating higher excess returns for the benefit of the shareholders and/or the policyholders.

As a caveat we add that the increase in the equity exposure suggested from the use of this model should come with an increased sophistication in the technology used to manage these assets vis-à-vis the liabilities. In particular, the asset portfolios must be carefully fine tuned with models such as the one presented here. Further analysis is needed in developing the scenarios of asset returns to be in agreement with future expectations, and to rely less on historical performance.

A significant extension for the long time horizons of the products considered would be to a multi-stage model where decisions are revised at time instances after $t = 0$ until maturity. Such dynamic stochastic programs with recourse have been developed for asset and liability management by the references given in the introduction. However, for the highly nonlinear problem we are addressing here such models are difficult to develop. The linearization of the single-stage model developed in Appendix A does not apply directly to multistage formulations. Specialized algorithms for geometric programming must be employed for the solution of multistage extensions of this model.

Acknowledgements

The work reported here was funded in part by Prometeia and by the HERMES Center of Excellence in Computational Finance and Economics under European Commission grant ICA1-CT-2000-70015.

Appendix A. Solving the nonlinear dynamic equations

In this section we show how to solve the nonlinear equations (5)–(9) in order to obtain the objective function (12). At time $t = 0$, the liability is the pure premium L_0. At $t = 1$ (to simplify the notation we drop the scenario superscript) we have

$$L_1 = L_0(1 - \Lambda_1)\bigl(1 + g + y_1^+\bigr). \tag{A.1}$$

At $t = 2$ we use the value of L_1 from (A.1) to obtain

$$\begin{aligned} L_2 &= L_1(1 - \Lambda_2)\bigl(1 + g + y_2^+\bigr) \\ &= L_0(1 - \Lambda_2)(1 - \Lambda_1)\bigl(1 + g + y_1^+\bigr)\bigl(1 + g + y_2^+\bigr). \end{aligned} \tag{A.2}$$

Applying this process recursively for each t we obtain the final liability as

$$L_T = L_0 \prod_{t=1}^{T} (1 - \Lambda_t)\bigl(1 + g + y_t^+\bigr). \tag{A.3}$$

For the equity dynamics we have that $E_0 = \rho L_0$. At $t = 1$

$$E_1 = \rho L_0(1 + r_{f1}) + y_1^- L_0. \tag{A.4}$$

At $t = 2$ and substituting for E_1 and L_1 from (A.4) and (A.1) we obtain

$$\begin{aligned} E_2 &= E_1(1 + r_{f2}) + y_2^- L_1 \\ &= \rho L_0(1 + r_{f1})(1 + r_{f2}) + L_0 y_1^-(1 + r_{f2}) + L_0 y_2^-(1 - \Lambda_1)\big(1 + g + y_1^+\big). \end{aligned} \tag{A.5}$$

At $t = 3$ we have

$$\begin{aligned} E_3 &= E_2(1 + r_{f3}) + y_3^- L_2 \\ &= \rho L_0(1 + r_{f1})(1 + r_{f2})(1 + r_{f3}) + L_0 y_3^-(1 + r_{f2})(1 + r_{f3}) \\ &\quad + L_0 y_3^-(1 + r_{f3})(1 - \Lambda_1)\big(1 + g + y_1^-\big) \\ &\quad + L_0 y_3^-(1 - \Lambda_2)(1 - \Lambda_1)\big(1 + g + y_1^+\big)\big(1 + g + y_2^+\big). \end{aligned} \tag{A.6}$$

Applying this process recursively for each t we obtain after some simple algebra

$$E_T = L_0\left[\rho \prod_{t=1}^{T}(1 + r_{ft}) + \sum_{t=1}^{T}\left(y_t^- \phi(t, T) \prod_{\tau=1}^{t-1}(1 - \Lambda_\tau)\big(1 + g + y_\tau^+\big)\right)\right], \tag{A.7}$$

where $\phi(t, T) = \prod_{\tau=t+1}^{T}(1 + r_{f\tau})$ is the cumulative return of the short rate from t to T.

With the same arguments it is possible to show that

$$y_{At} = L_0 \Lambda_t\big(1 + g + y_t^+\big) \prod_{\tau=1}^{t-1}(1 - \Lambda_\tau)\big(1 + g + y_\tau^+\big). \tag{A.8}$$

For the asset dynamics we have that $A_0 = L_0(1 + \rho)$. At $t = 1$

$$\begin{aligned} A_1 &= A_0(1 + R_{P1}) + y_1^- L_0 - y_{A1} \\ &= L_0(1 + \rho)(1 + R_{P1}) + y_1^- L_0 - y_{A1}. \end{aligned} \tag{A.9}$$

At $t = 2$ substituting L_1 from (A.1) we obtain

$$\begin{aligned} A_2 &= A_1(1 + R_{P2}) + y_2^- L_1 - y_{A2} \\ &= L_0(1 + \rho)(1 + R_{P1})(1 + R_{P2}) + y_1^- L_0(1 + R_{P2}) \\ &\quad - y_{A1}(1 + R_{P2}) + y_2^- L_1 - y_{A2}. \end{aligned} \tag{A.10}$$

The value of the assets at maturity is given by

$$\begin{aligned} A_T = {}& L_0(1 + \rho) \prod_{t=1}^{T}(1 + R_{Pt}) L_0 \sum_{t=1}^{T} y_t^- \prod_{\tau=t+1}^{T}(1 + R_{P\tau}) \prod_{\tau=1}^{t-1}\big(1 + g + y_\tau^+\big) \\ & \times \prod_{\tau=1}^{t-1}(1 - \Lambda_\tau) - \sum_{t=1}^{T} y_{At} \prod_{\tau=t+1}^{T}(1 + R_{P\tau}). \end{aligned} \tag{A.11}$$

By substituting y_{At} with the expression in (A.8), we obtain

$$
\begin{aligned}
A_T = {} & L_0(1+\rho)\prod_{t=1}^{T}(1+R_{Pt}) \\
& + L_0\sum_{t=1}^{T} y_t^- \prod_{\tau=t+1}^{T}(1+R_{P\tau})\prod_{\tau=1}^{t-1}\bigl(1+g+y_\tau^+\bigr)(1-\Lambda_\tau) \\
& - L_0\sum_{t=1}^{T}\Lambda_t\left(1+g+y_t^+\right)\prod_{\tau=t+1}^{T}(1+R_{P\tau})\prod_{\tau=1}^{t-1}\bigl(1+g+y_\tau^+\bigr)(1-\Lambda_\tau).
\end{aligned}
\tag{A.12}
$$

Collecting terms we obtain

$$
\begin{aligned}
A_T = {} & L_0(1+\rho)\prod_{t=1}^{T}(1+R_{Pt}) \\
& + L_0\sum_{t=1}^{T}\bigl(y_t^- - \Lambda_t(1+g+y_t^+)\bigr)\prod_{\tau=t+1}^{T}(1+R_{P\tau}) \\
& \times\prod_{\tau=1}^{t-1}\bigl(1+g+y_\tau^+\bigr)(1-\Lambda_\tau).
\end{aligned}
\tag{A.13}
$$

Appendix B. Asset classes

The asset classes used in testing the base model are given in Table 6. They consist of bond indices for short-, medium-, and long-term debt of the Italian government, and stock indices of the major industrial sectors traded in the Milano stock exchange.

Table 6
Asset classes used in testing the base model

Code	Description
SBGVNIT.1-3	Salomon Brother Italian Government Bond 1–3 years
SBGVNIT.3-7	Salomon Brother Italian Government Bond 3–7 years
SBGVNIT.7-10	Salomon Brother Italian Government Bond 7–10 years
ITMSBNK	Milan Mib Historic Banks
ITMSAUT	Milan Mib Historic Cars
ITMSCEM	Milan Mib Historic Chemicals
ITMSCST	Milan Mib Historic Construction
ITMSDST	Milan Mib Historic Distribution
ITMSELT	Milan Mib Historic Electronics

(*continued on next page*)

Table 6
(*continued*)

Code	Description
ITMSFIN	Milan Mib Historic Finance
ITMSFPA	Milan Mib Historic Finance Holdings
ITMSFMS	Milan Mib Historic Finance Misc
ITMSFNS	Milan Mib Historic Finance Services
ITMSFOD	Milan Mib Historic Food
ITMSIND	Milan Mib Historic Industrials
ITMSINM	Milan Mib Historic Industrials Misc
ITMSINS	Milan Mib Historic Insurance
ITMSPUB	Milan Mib Historic Media
ITMSMAM	Milan Mib Historic MineralsMetals
ITMSPAP	Milan Mib Historic Paper
ITMSMAC	Milan Mib Historic Plants & Machine
ITMSPSU	Milan Mib Historic Pub. Util. Serv.
ITMSRES	Milan Mib Historic Real Estate
ITMSSER	Milan Mib Historic Services
ITMSTEX	Milan Mib Historic TextileClothing
ITMST&T	Milan Mib Historic Transportation & Tourism

References

Asay, M.R., Bouyoucos, P.J., Marciano, A.M., 1993. An economic approach to valuation of single premium deferred annuities. In: Zenios, S.A. (Ed.), Financial Optimization. Cambridge University Press, Cambridge, UK, pp. 100–135.

Babbel, D.F., 2001. Asset/liability management for insurers in the new era: Focus on value. Journal of Risk Finance 3 (1), 9–17, Fall.

Bacinello, A.R., 2001. Fair pricing of life insurance participating policies with a minimum guarantee. Astin Bulletin 31 (2), 275–297.

Boyle, P.P., Hardy, M.R., 1997. Reserving for maturity guarantees: Two approaches. Insurance: Mathematics & Economics 21 (2), 113–127.

Boyle, P.P., Schwartz, E.S., 1977. Equilibrium prices of guarantees under equity-linked contracts. Journal of Risk and Insurance 44, 639–660.

Brennan, M.J., Schwartz, E.S., 1976. The pricing of equity-linked life insurance policies with an asset value guarantee. Journal of Financial Economics 3, 195–213.

Brennan, M.J., Schwartz, E.S., 1979. Alternative investment strategies for the issuers of equity linked life insurance policies with an asset value guarantee. Journal of Business 52, 63–93.

Carinõ, D.R., Ziemba, W.T., 1998. Formulation of the Russel–Yasuda Kasai financial planning model. Operations Research 46 (4), 433–449.

Censor, Y., Zenios, S.A., 1997. Parallel Optimization: Theory, Algorithms, and Applications. Numerical Mathematics and Scientific Computation. Oxford University Press, New York.

Chadburn, R.G., 1997. The use of capital, bonus policy and investment policy in the control of solvency for with-profits life insurance companies in the UK, Technical report. City University, London.

Consigli, G., Dempster, M.A.H., 1998. The CALM stochastic programming model for dynamic asset and liability management. In: Ziemba, W., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 464–500.

Consiglio, A., Cocco, F., Zenios, S.A., 2000. Asset and liability modeling for participating policies with guarantees, Working Paper 00-41-c. The Wharton Financial Institutions Center, University of Pennsylvania.
Consiglio, A., Cocco, F., Zenios, S.A., 2001. The value of integrative risk management for insurance products with minimum guarantees. Journal of Risk Finance 2 (3), 1–11.
Consiglio, A., Saunders, D., Zenios, S.A., 2003. Insurance league: Italy vs. UK. Journal of Risk Finance 4 (4), 47–54.
Consiglio, A., Saunders, D., Zenios, S.A., 2006. Asset and liability management for insurance products with minimum guarantees: The UK case. Journal of Banking and Finance 30, 645–667.
Dembo, R., Aziz, A., Rosen, D., Zerbs, M., 2000. Mark To Future—A Framework for Measuring Risk and Reward. Algorithmics Publications, May.
Embrechts, P., 2000. Actuarial versus financial pricing of insurance. Journal of Risk Finance 1 (4), 17–26.
Giraldi, C., Susinno, G., Berti, G., Brunello, J., Buttarazzi, S., Cenciarelli, G., Daroda, C., Stamegna, G., 2003. Insurance optional. In: Lipton, A. (Ed.), Exotic Options: The Cutting-Edge Collection Technical Papers Published in Risk 1999–2003. Risk Books (Chapter 35).
Grosen, A., Jørgensen, P.L., 2000. Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies. Insurance: Mathematics & Economics 26, 37–57.
Hansen, M., Miltersen, K.R., 2002. Minimum rate of return guarantees: The Danish case. Scandinavian Actuarial Journal 4, 230–318.
Harker, P.T., Zenios, S.A. (Eds.), 2000. Performance of Financial Institutions: Efficiency, Innovation, Regulations. Cambridge University Press, Cambridge, England.
Holmer, M.R., Zenios, S.A., 1995. The productivity of financial intermediation and the technology of financial product management. Operations Research 43 (6), 970–982.
Høyland, K., 1998. Asset liability management for a life insurance company: A stochastic programming approach, PhD thesis. Norwegian University of Science and Technology, Trondheim, Norway.
Kang, P., Zenios, S.A., 1992. Complete prepayment models for mortgage-backed securities. Management Science 38 (11), 1665–1685, November.
Kat, H.M., 2001. Structured Equity Derivatives: The Definitive Guide to Exotic Options and Structured Notes. John Wiley & Son.
Miltersen, K.R., Persson, S., 1999. Pricing rate of return guarantees in a Heath–Jarrow–Morton framework. Insurance: Mathematics & Economics 25, 307–325.
Mulvey, J.M., Thorlacius, A.E., 1998. The Towers Perrin global capital market scenario generation system. In: Ziemba, W., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 286–312.
Ross, M.D., 1989. Modelling a with-profit life office. Journal of the Institute of Actuaries 116, 691–716.
Siglienti, S., 2000. Consequences of the reduction of interest rates on insurance. The Geneva Papers on Risk and Insurance 25 (1), 63–77.
Stulz, R., 1996. Rethinking risk management. Journal of Applied Corporate Finance 9 (3), 8–24.
Wilkie, A.D., 1995. More on a stochastic asset model for actuarial use. British Actuarial Journal 1 (5), 777–964.

This page intentionally left blank

Chapter 16

INTEGRATED RISK CONTROL USING STOCHASTIC PROGRAMMING ALM MODELS FOR MONEY MANAGEMENT

N.C.P. EDIRISINGHE

Department of Statistics, Operations, and Management Science,
College of Business Administration, University of Tennessee,
Knoxville, TN 37996,
USA
e-mail: chanaka@utk.edu

Contents

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02016-3

Abstract

A multistage stochastic programming modeling based approach is developed for asset and liability management for fund managers. Managing a fund of investments in a large pool of possible instruments, such as stocks, requires sophisticated analytical capability both in terms of selecting the pool and also in maintaining the performance of the fund within acceptable levels. Given the uncertainty of the future performance of the underlying stocks, the portfolio must be managed or rebalanced temporally as the market and economic conditions change. Such portfolio rebalancing at various points in time allows the fund manager to manage the riskiness of the fund from both the fund managers and the individual clients viewpoints. In doing so, a fund manager must resort to more-advanced analytical risk-control techniques. I apply a multi-prong risk metric system that is designed for the portfolio to achieve desired performance characteristics. The model incorporates important issues such as market impact costs in trading, fund drawdown, market neutrality, and catastrophic risk, within an integrative framework, modeled via stochastic programming. The model is applied to stock fund management involving a large number of securities and its performance is demonstrated with various strategies for portfolio rebalancing. The integrated dynamic multistage stochastic programming model easily outperforms the standard static mean-variance approach (i.e., the Markowitz model) for portfolio management. Sharpe ratios, percent worst draw downs, recovery periods from drawdown, portfolio rate of returns, etc. are used for performance comparisons.

Keywords

portfolio optimization, stochastic multistage programming, risk management, integrated risk control, money management

JEL classification: D81, G11, C61

1. Introduction

There is a tremendous growth in the number of individuals who are engaged in money management in the twenty-first century, yet only a few implement disciplined, professional money management strategies. During the stock market bubble of the late 1990s, limiting risk was an afterthought, but given the recent stock market action, more managers are considering more sophisticated approaches to portfolio risk management. Typically, only a few have the ability to view their portfolios from a risk/return integrated perspective. Instead, many individuals take a defensive or reactive view of risk in which risk is measured to avoid losses. What is essential for successful money management is to have an offensive or proactive posture in which risks are actively managed. This paper proposes such an integrated framework for fund management using multistage stochastic programming modeling.

Managing a fund of investments in a large pool of possible instruments (i.e., stocks) requires sophisticated analytical capability both in terms of selecting the pool and also in maintaining the performance of the fund within acceptable levels. Given the uncertainty of the future performance of the underlying stocks, it is imperative that the portfolio of stocks (in the fund) be managed or rebalanced temporally as the market and economic conditions change. Such portfolio rebalancing at various points in time allows the fund manager to control the *riskiness* of the fund from both the fund managers and the individual clients viewpoints. In doing so, a fund manager must attempt to remove the guesswork based on emotions (or gutt feeling) from the decision making and resort to the more-advanced analytical risk-management techniques.

Reading material on money management is quite abundant—a search on *Amazon.com* on the Internet returns more than 2100 books. Various professional money management services are also available that offer diverse asset allocation strategies based on common risk and time horizon parameters. These allocations are typically optimized using a static mean-variance framework, following the early work on portfolio optimization by Markowitz (1952, 1987), where a static, deterministic model for trading off portfolio expected return with portfolio variance was proposed. See the chapter by Markowitz and Van Dijk in this volume for a current survey of this approach. Variants of this approach that utilize a mean absolute deviation (MAD) functional, rather than portfolio variance, have been proposed, see, e.g., Konno and Yamazaki (1991) and Konno and Kobayashi (1997).

Asset allocation is the practice of dividing resources among different categories such as stocks, bonds, mutual funds, investment partnerships, real estate, cash equivalents, and private equity. Such models are expected to lessen risk exposure since each asset class has a different correlation to the others; when stocks rise, for example, bonds often fall. At a time when the stock market begins to fall, real estate may begin generating above average returns. Consequently, various allocation styles may be specified for the purposes of, say, preservation of capital, generating income, long-term aggressive growth, or striking a balance between income and growth. However, merely diversifying assets to the prescribed allocation model is not going to alleviate the manager from the need to choose individual issues. Indeed, the asset allocation and the choice of individual securities must occur integratively, consistent with the intended risk specification. A further complication in this case is the need for temporal rebalancing in an effort to bring the portfolio back into balance with the original prescription, or a variation thereof consistent with market evolution and changes in risk preferences.

However, static (myopic) models of portfolio optimization fail to capture two important aspects of portfolio rebalancing: (1) trade-off between short-term and long-term consequences of investment strategy, based on the evolution of stochastic factors, and (2) presence of transactions or market impact costs, taxes, etc. that affect portfolio holdings over time, see Ziemba and Mulvey (1998).

In contrast, a sequential decision theoretic optimization framework for portfolio rebalancing is available via multistage stochastic programming (MSP) modeling. This approach allows modeling various future time periods of portfolio revisions explicitly where stochastic dynamic evolution of random parameters (e.g., security returns) can be incorporated via a so-called scenario tree. Each scenario is a timed-sequence of events through the end of the planning horizon with an associated probability of occurrence. Multistage stochastic programming models have been richly applied in a variety of financial applications; see Kusy and Ziemba (1986) for a bank asset/liability management, Cariño and Ziemba (1998) for an asset/liability management problem of a Japanese insurance company, Mulvey and Vladimirou (1992) for a multiperiod stochastic network model for the purpose of asset allocation, and Golub et al. (1997) for fixed income securities management. For applications of stochastic multistage linear programming with a binomial scenario tree of price uncertainty for financial option pricing, see Edirisinghe, Naik and Uppal (1993), and the extensions in Edirisinghe (2004). This article provides a multistage stochastic programming model that incorporates sequential rebalancing of the portfolio in order to maintain the portfolio performance with respect to a multi-prong risk measurement scheme.

Stochastic programming models have found many applications within asset and liability management (ALM) problems, which are generally of a long term nature. ALM attempts to find the optimal investment strategy under uncertainty in both the asset and liability streams. Simultaneous consideration of assets and liabilities can be very advantageous in terms of increasing returns and reducing risk, especially when they have common risk factors leading to high correlations. ALM-MSP models allow for dynamic portfolio rebalancing while satisfying operational or regulatory restrictions and policy requirements, see, e.g., Holmer and Zenios (1995). Early work on application of stochastic programming for financial planning is by Bradley and Crane (1972) and Ziemba and Vickson (1975). Since 1990, there is an increased momentum in such applications, driven partly by the globalization and innovations in the financial markets, but largely due to tremendous improvement in the solution algorithms and computing hardware advances. See Zenios (1995) and Nielsen and Zenios (1996) for fixed income portfolio management; Cariño et al. (1994), Consigli and Dempster (1998), Høyland (1998), and Mulvey, Gould and Morgan (2000) for insurance companies, Dert (1995) for pension funds, Consiglio, Cocco and Zenios (2001) for minimum guarantee products, as well as Zenios (1993) and Ziemba and Mulvey (1998). Also, see Ziemba (2003) and other chapters in this volume.

I present a stochastic programming model for money management where frequent short-term portfolio rebalancing is commonplace and transactions costs and market impact costs play a significant role in determining trade sizes. Stochastic multiperiod models (of ALM type) are utilized to trade-off returns and transactions costs under various metrics of risk aversion. Indeed, such an approach must be adapted to market evolutionary parameters and optimized for possible errors in parameter forecasts via a specific investment strategy. The model is applied to historical time series of a large stock base and out-of-sample portfolio performance is tracked over daily rollover operation. The main focus is in developing the model and discussing its investment performance, rather than efficient solution algorithms for multistage stochastic programs of this type. The reader interested in the latter aspect is referred to the recent paper by Edirisinghe and Patterson (2007), in which a very efficient solution methodology is developed for real-time solution. Given that high frequency data are readily available on a global basis and computing hardware are becoming more powerful by orders of magnitude, sophisticated modeling techniques as presented here are expected to find broader appeal among the professional investment community.

Section 2 presents an introduction to formulating a multistage stochastic programming (MSP) model. This section also collects the necessary tree notation that will be required for understanding a MSP model. Section 3 provides the basic constructs of the MSP model for the investment optimization problem and several key policy constraints. Section 4 covers various issues with risk metrics and it introduces new and known methodologies for risk control. The multistage stochastic program for trade rebalancing problem is presented in Section 5, and the model application and results analyses are in Section 6. Concluding remarks are in Section 7. The required notation is introduced as it becomes necessary.

2. Multistage stochastic programming (MSP)

Multiperiod stochastic programs can be used to model a sequence of *decision-observation* processes in which decisions at any point in time are based upon historical observations coupled with beliefs or expectations concerning the uncertainty of future events. The term *period* is used to refer to the time starting with one decision or set of decisions and lasting until the next decision or set of decisions. See Birge, Edirisinghe and Ziemba (2001) for some applications of stochastic programming models and Birge and Louveaux (1997) for certain terms and definitions concerning stochastic linear programs. For a complete bibliography on stochastic programming, see van der Vlerk (2003).

Decision trees, often referred to as event or scenario trees, are a useful tool for visualizing the resulting model. Decision trees demonstrate the *nonanticipative* requirement of stochastic programs—decisions in a given period must be made without the hindsight

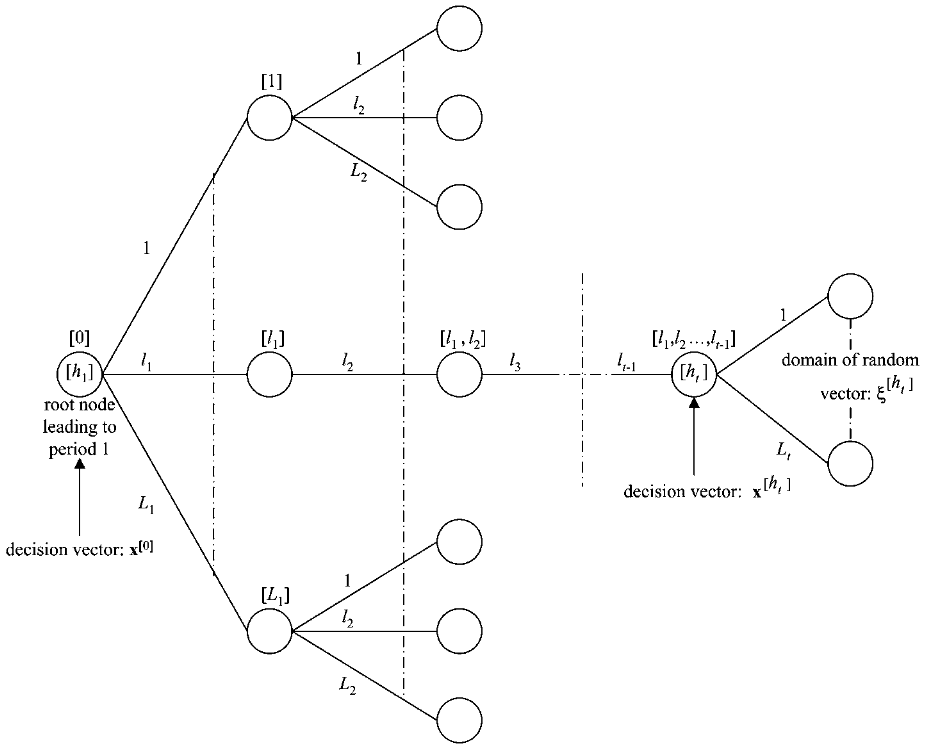


Fig. 1. Decision/scenario tree conventions.

(or the anticipation) of a specific future outcome. The notation used in the development and discussion of multistage stochastic programs is not unique in the literature. The following notation that makes explicit the dependence of decisions and random events on historical scenario paths is used. Assume the number of outcomes L_t possible at any node in period $t \in \{1, \dots, T\}$ for arbitrary but finite T is the same for all nodes in that period. There are then $H_t = \prod_{j=1}^{t-1} L_j$ total nodes in period t and $\sum_{j=1}^{t} H_j$ cumulative nodes in periods 1 through t. Nodes are labeled using a *path vector* format. The path vector, say $[l_1, l_2, \dots, l_{t-1}]$, to a node in period t is a row vector of $(t-1)$ elements where element j, $1 \leqslant j < t$, is the index, l_j, of the outcome in period j along the path of outcomes to the applicable node. The path to a node in period t is called a *t-period scenario* and each T-period scenario is usually simply called a *scenario* of the stochastic program. Let $[h_t] = [l_1, l_2, \dots, l_{t-1}]$ represent a *t-period scenario,* i.e., the (historical) path vector to a node at the beginning of period t, see Figure 1. The single first period *root* node is labeled using the notation $[h_1]$, or simply by [0] for brevity, which represents the absence of any historical scenario path. By convention, $[h_{t-1}]$ and $[h_t]$ appearing in the same expression implies that the former is the parent node of the latter while $[h_{t+1}]$ is a child node of $[h_t]$.

A decision (vector) taken at node $[h_t]$ is denoted by $\mathbf{x}^{[h_t]}$. It is a nonanticipative decision that depends only on the specific history $[h_t] \equiv [l_1, l_2, \ldots, l_{t-1}]$, which corresponds to the sequence of observed outcomes of the random vectors $\boldsymbol{\xi}^{[1]}, \boldsymbol{\xi}^{[h_2]}, \ldots, \boldsymbol{\xi}^{[h_{t-1}]}$, respectively. Having made the decision $\mathbf{x}^{[h_t]}$, the random vector $\boldsymbol{\xi}^{[h_t]}$ will be observed whose domain of realizations are indicated by $\{\boldsymbol{\xi}^{[h_t],1}, \ldots, \boldsymbol{\xi}^{[h_t],L_t}\}$, which form the set of child nodes emanating from node $[h_t]$.

The path-vector notation is also used to implicitly indicate the dependence of a stochastic array at a node in period t on the occurrence of the specific outcomes listed by the *path vector*. Let $\omega_{l_t}^{(t)}$ represent the period t outcome with index l_t, $1 \leqslant l_t \leqslant L_t$, and let $[h_t] = [\tilde{l}_1, \ldots, \tilde{l}_{t-1}]$ represent a node in period t, $1 \leqslant t \leqslant T$. Then,

$$\mathbf{W}^{[h_t]} = \mathbf{W}\big(\omega_{\tilde{l}_{t-1}}^{(t-1)} \big| \omega_{\tilde{l}_{t-2}}^{(t-2)}, \ldots, \omega_{\tilde{l}_2}^{(2)}, \omega_{\tilde{l}_1}^{(1)}\big),$$

represents a conditional stochastic matrix at node $[h_t]$. The conditional probability that the outcome with index l_t, $1 \leqslant l_t \leqslant L_t$, $1 \leqslant t \leqslant T$, is observed at the t-period node $[h_t]$ given the indicated outcomes $\tilde{l}_1, \ldots, \tilde{l}_{t-1}$ in periods $1, \ldots, t-1$ is $p_{l_t}^{[h_t]}$, i.e.,

$$p_{l_t}^{[h_t]} = \mathcal{P}\big(\omega_{l_t}^{(t)} \big| \omega_{\tilde{l}_{t-1}}^{(t-1)}, \omega_{\tilde{l}_{t-2}}^{(t-2)}, \ldots, \omega_{\tilde{l}_2}^{(2)}, \omega_{\tilde{l}_1}^{(1)}\big),$$

where $\mathcal{P}(.)$ is the probability operator and $p_{l_1}^{[h_1]} = \mathcal{P}(\omega_{l_1}^{(1)} | \omega_0^{(0)}) = \mathcal{P}(\omega_{l_1}^{(1)})$. The *joint* probability, $\hat{p}^{[h_t]}$, that the process enters the period t node $[h_t] = [\tilde{l}_1, \ldots, \tilde{l}_{t-1}]$, $1 \leqslant t \leqslant T$, is the product of the conditional probabilities of the outcomes along the indicated path to that node,

$$\hat{p}^{[h_t]} = \hat{p}^{[\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_{t-2}, \tilde{l}_{t-1}]} = p_{\tilde{l}_1}^{[\,]} p_{\tilde{l}_2}^{[\tilde{l}_1]} \cdots p_{\tilde{l}_{t-1}}^{[\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_{t-2}]} = \prod_{j=1}^{t-1} p_{\tilde{l}_j}^{[h_j]},$$

where $\hat{p}^{[h_1]} = 1$, i.e., the single first period (root) node is always entered.

2.1. MSP model formulation

Using this notation for decision trees, a multiperiod stochastic program can be formulated recursively by considering the *nodal* decision problems. A linear stochastic programming model is presented for clarity; extensions to a general nonlinear framework is quite straightforward, see, e.g., formulations in Edirisinghe and Ziemba (1992) and Frauendorfer (1996). The decision process up to (tree) node $[h_t]$ has observed the realization sequence indexed as $[l_1, l_2, \ldots, l_{t-1}]$, and corresponds to the decision vector sequence $\mathbf{x}^{[h_1]}, \mathbf{x}^{[h_2]}, \ldots, \mathbf{x}^{[h_{t-1}]}$. The nodal value function $\mathcal{Q}^{[h_t]}(\mathbf{x}^{[h_{t-1}]})$ for a period t, $2 \leqslant t \leqslant T-1$, is the optimal profits to be realized after observing the event indexed by l_{t-1} and by choosing the decision vector $\mathbf{x}^{[h_t]}$ to optimally trade off current profits $\mathbf{c}^{[h_t]}\mathbf{x}^{[h_t]}$ and the expected value of optimal profits at the child nodes, i.e., $\mathcal{Q}^{[h_{t+1}]}(\mathbf{x}^{[h_t]})$ for each of $l_t = 1, \ldots, L_t$. This nodal decision problem is formulated as follows, for

$2 \leqslant t \leqslant T-1$.

$$\begin{aligned} \mathcal{Q}^{[h_t]}(\mathbf{x}^{[h_{t-1}]}) = {} & \max \mathbf{c}^{[h_t]}\mathbf{x}^{[h_t]} + \sum_{l_t=1}^{L_t} p_{l_t}^{[h_t]} \mathcal{Q}^{[h_{t+1}]}(\mathbf{x}^{[h_t]}) \\ \text{s.t.} \quad & \mathbf{W}^{[h_t]}\mathbf{x}^{[h_t]} = \mathbf{b}^{[h_t]} - \mathbf{B}^{[h_t]}\mathbf{x}^{[h_{t-1}]}, \\ & \mathbf{x}^{[h_t]} \geqslant 0. \end{aligned} \tag{1}$$

The nodal value function, $\mathcal{Q}^{[h_T]}(\mathbf{x}^{[h_{T-1}]})$, for a node at terminal period T, is

$$\begin{aligned} \mathcal{Q}^{[h_T]}(\mathbf{x}^{[h_{T-1}]}) = {} & \max \mathbf{c}^{[h_T]}\mathbf{x}^{[h_T]} \\ \text{s.t.} \quad & \mathbf{W}^{[h_T]}\mathbf{x}^{[h_T]} = \mathbf{b}^{[h_T]} - \mathbf{B}^{[h_T]}\mathbf{x}^{[h_{T-1}]}, \\ & \mathbf{x}^{[h_T]} \geqslant 0. \end{aligned} \tag{2}$$

Array dimensions are not explicitly listed for reasons of brevity and all arrays are assumed to have shapes and sizes compatible with the depicted operation. Then, the multistage stochastic program can be equivalently represented as the following compact recursive value function formulation.

$$\begin{aligned} Z^* = {} & \max \mathbf{c}^{[h_1]}\mathbf{x}^{[h_1]} + \sum_{l_1=1}^{L_1} p_{l_1}^{[h_1]} \mathcal{Q}^{[h_2]}(\mathbf{x}^{[h_1]}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x}^{[h_1]} = \mathbf{b}^{[h_1]}, \\ & \mathbf{x}^{[h_1]} \geqslant 0. \end{aligned} \tag{3}$$

3. Investment optimization model

There are N securities that an investor wishes to trade at time $t=0$, i.e., at the beginning of period 1. The investor's initial position (i.e., the number of shares in each security) is $\mathbf{x}^0$ ($\in \Re^N$) and the initial cash position is C^0. The price of security j at the first trading node, i.e., the root node $[h_1]$, is $\$P_j^{[h_1]}$ per share, see Figure 1 for the decision tree. The investor considers trading in multiple periods of time, denoted by the time index $t = 1, \ldots, T$. Given the historical price vectors $\mathbf{P}^{[h_1]}, \ldots, \mathbf{P}^{[h_t]}$, up until the beginning of trading for period t, the (conditional) rate of return vector for period t is $\mathbf{r}^{[h_t]}$. See Figure 2 for an illustration of the nodal trading information. Thus, price of security j changes during the trading period t to $(1 + r_j^{[h_t]})P_j^{[h_t]}$. Note that $r_j^{[h_t]} \geqslant -1$ since the security prices are nonnegative, i.e., $\mathbf{P}^{[h_t]} \geqslant 0$. Moreover, $\mathbf{r}^{[h_t]}$ is random and it is observed only at the end of the current period t; however, trade decisions $\mathbf{x}^{[h_t]}$ must be made at the beginning of period t, i.e., revision of portfolio positions from $\mathbf{x}^{[h_{t-1}]}$ to $\mathbf{x}^{[h_t]}$. Therefore, $x_j^{[h_t]} - x_j^{[h_{t-1}]}$ is the amount of shares purchased if it is positive; and if it is negative, it is the amount of shares sold in security j. This trade vector is $\mathbf{y}^{[h_t]}$ and equals $|\mathbf{x}^{[h_t]} - \mathbf{x}^{[h_{t-1}]}|$ where $|.|$ is the absolute value.

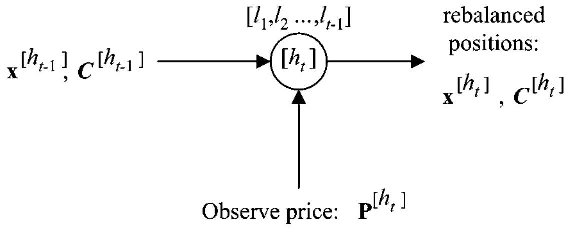


Fig. 2. Trading at a decision node at time t.

The investment optimization problem is concerned with determining the trade vectors $\mathbf{y}^{[h_t]}$ such that various risk specifications for the portfolio are met whilst maximizing the portfolio total expected return. Capturing this problem within a multistage framework is advantageous due to at least two important considerations. First, trading (for portfolio rebalancing) is generally not *costless*. Typically, with larger trade sizes, there is an increasingly diluting effect on the profitability of the trades due to market impact. Thus, it may be beneficial to create an intended position over multiple periods as the trading costs increase. Secondly, due to policy on risk control, portfolio wealth trajectory may have to be monitored, such as the case when portfolio drawdown is an important issue in fund management. Such drawdowns cannot be managed effectively via single period modeling.

3.1. Transactions and slippage costs

Usually, fund managers face transactions costs in executing the trade vector, $\mathbf{y}^{[h_t]}$, which leads to reducing the portfolio net return. Placing a trade with a broker for execution entails a direct cost per share traded, as well as a fixed cost independent of the trade size. For instance, a $10 fixed cost and a 2-cent per-share cost may be levied. In addition, there is also a significant cost due to the size of trading volume $\mathbf{y}^{[h_t]}$, as well as the broker's ability to place the trading volume on the market. If a significant volume of shares is traded (relative to the market daily traded volume in the security), then the trade execution price may be adversely affected. A large buy order usually lead to trade execution at a price higher than intended and a large sell order leads to an average execution price that is lower than desired. This dilution of the profits of the trade is termed the *market impact loss*, or slippage. This slippage loss generally depends on the price at which the trade is desired, trade size relative to the market daily volume in the security, and other company specifics such as market capitalization, and the beta of the security. See Keim and Madhavan (1996, 1997, 1998), Loeb (1983), and Torre and Ferrari (1999), for instance, for a discussion on market impact costs.

In our model, a quadratic simplification for the market impact cost is utilized. That is, the slippage costs are proportional to the square of the volume traded in the market, and the constant of proportionality depends directly on the intended execution price, and inversely on the (daily) share volume in that security. For trading $y_j^{[h_t]}$ shares of security j (with the intended trade price $P_j^{[h_t]}$ when the expected total daily volume is $V_j^{[h_t]}$ shares), thus, leads to a market impact cost of

$$\alpha_{1j}\frac{P_j^{[h_t]}}{V_j^{[h_t]}}\big(y_j^{[h_t]}\big)^2,$$

where α_{1j} is a constant calibrated to the market data. The direct cost of trading is represented by $\alpha_{0j}y_j^{[h_t]}$ where α_{0j} is the per-share transactions cost. The fixed costs of trading are ignored here, and then, the total transactions and slippage loss function $f_j(y_j^{[h_t]})$ is

$$f_j\big(y_j^{[h_t]}\big) := y_j^{[h_t]}\left(\alpha_{0j} + \alpha_{1j}\frac{P_j^{[h_t]}y_j^{[h_t]}}{V_j^{[h_t]}}\right). \tag{4}$$

Therefore, the total loss due to portfolio rebalancing at time t, given the historical scenario h_t of security price evolution, is

$$F(\mathbf{y}^{[h_t]}) := \sum_{j=1}^{N} f_j\big(y_j^{[h_t]}\big). \tag{5}$$

3.2. *Budget and portfolio wealth*

Portfolio rebalancing from period-to-period is assumed to carry forward any cash that is accumulated (i.e., the riskless security) and the total cost of trading at the beginning of period t is thus limited by the cash carried forward from period $t-1$, denoted by the nonnegative variable $C^{[h_{t-1}]}$, and any exogenous funds available at the beginning of period t, denoted by B_t. Then, the following budget constraint must hold.

$$\mathbf{P}^{[h_t]}(\mathbf{x}^{[h_t]} - \mathbf{x}^{[h_{t-1}]}) + F(\mathbf{y}^{[h_t]}) + C^{[h_t]} = B_t + (1+\kappa_{t-1})C^{[h_{t-1}]}. \tag{6}$$

The transposition of the price vector $\mathbf{P}^{[h_t]}$ is suppressed for clarity of exposition in the above expression, as well as throughout the remainder of the chapter. When applying the budget constraint (6) for trading at the root node $[h_1]$, by convention, $\mathbf{x}^{[h_0]} = \mathbf{x}^0$, the initial security positions, and $C^{[h_0]} = C^0$, the initial cash position. Note that "cash" carried forward yields a deterministic monetary return from period to period given by the rate κ_t ($\geqslant 1$), which is already adjusted to the length of the period t. Thus, $(1+\kappa_t)C^{[h_t]}$ is the excess cash carried forward and available for portfolio rebalancing at node $[h_{t+1}]$ of period $t+1$. In the event that trading is required to be *self-financing*, one must set $B_t = 0$. Alternatively, if a certain dollar amount, say b_t, must be taken off the stock market (to satisfy a certain known liability at the current time period), then, one must set $B_t = -b_t$.

Given portfolio rebalancing at the current decision node $[h_t]$ to obtain revised positions $\mathbf{x}^{[h_t]}$, the portfolio gain under (future) return scenario vector $\mathbf{r}^{[h_t]}$, is

$$G^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) := \mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} + \kappa_t C^{[h_t]} - F(\mathbf{y}^{[h_t]}), \tag{7}$$

where $\mathbf{D}(\mathbf{a})$ represents the diagonal matrix in which the jth diagonal element is a_j. Observe that the uncertainty of the return $\mathbf{r}^{[h_t]}$, conditional upon the history h_t being followed, endows the portfolio gain variable $G^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})$ with uncertainty. Consequently, the *projected portfolio wealth*, denoted by $w^{[h_t]}$, under return scenario $\mathbf{r}^{[h_t]}$ satisfies

$$w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) = w^{[h_{t-1}]} + G^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) + B_t, \tag{8}$$

where $w^{[h_{t-1}]}$ denotes the portfolio wealth (consisting of cash and risky securities) at the end of period $t-1$, evaluated before trading at the current decision node $[h_t]$ using the state price vector $\mathbf{P}^{[h_t]}$. When applying (8) for $t = 1$, the initial portfolio wealth prior to trading at the root node $[h_1]$ is $w^{[h_0]} \equiv w^0 = \mathbf{P}^{[h_1]}\mathbf{x}^0 + C^0$. It must be noted that, at some period t under some return scenario $\mathbf{r}^{[h_t]}$, it is possible that the portfolio projected wealth variable $w^{[h_t]} < 0$, which thus indicates the termination of the portfolio. Consequently,

$$w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant 0 \tag{9}$$

is specified as a required feasibility condition under all realizations of the scenario tree.

3.3. *Limiting positions*

The nonnegative trade-size in security j is $y_j^{[h_t]} = |x_j^{[h_t]} - x_j^{[h_{t-1}]}|$. Since the investment objective is to maximize the total portfolio (expected) return, less the convex quadratic slippage loss $F(.)$ in (5), the following linear constraints can be used to replace the nonlinear (absolute-value) expressions for the trade-size vector:

$$\begin{aligned} &\mathbf{x}^{[h_t]} - \mathbf{y}^{[h_t]} \leqslant \mathbf{x}^{[h_{t-1}]}, \\ &\mathbf{x}^{[h_t]} + \mathbf{y}^{[h_t]} \geqslant \mathbf{x}^{[h_{t-1}]}. \end{aligned} \tag{10}$$

Positions $\mathbf{x}^{[h_t]}$, after portfolio revision, must satisfy certain minimum/maximum limits due to policy requirements. This often arises from requiring that no more than (or no less than for the purposes of short positions) a certain dollar allocation be made in a particular security. In share volume terms, these maximum and minimum position limits are $\mathbf{x}^{[h_t],\max}$ and $\mathbf{x}^{[h_t],\min}$, and thus the resulting constraints are

$$\mathbf{x}^{[h_t],\min} \leqslant \mathbf{x}^{[h_t]} \leqslant \mathbf{x}^{[h_t],\max}. \tag{11}$$

Alternatively, long and short positions in the portfolio may be controlled in aggregate by using the *long leverage* and *short leverage* restrictions. That is, a prespecified nonnegative \$ limit M_t^L may be applied for the total long position and M_t^S for the total short position, when trading for period t, which yields

$$\sum_{j=1}^{N} P_j^{[h_t]} \max\{0, x^{[h_t]}\} \leqslant M_t^L \quad \text{and} \quad \sum_{j=1}^{N} P_j^{[h_t]} \max\{0, -x^{[h_t]}\} \leqslant M_t^S.$$

A linear representation of the above is obtained by defining pairs of nonnegative (long/short) variables $u_j^{[h_t]}$ and $v_j^{[h_t]}$, for security j, such that

$$x_j^{[h_t]} = u_j^{[h_t]} - v_j^{[h_t]}. \tag{12}$$

Under the variable description above, the following pair of linear constraints can be used to represent the leverage constraints.

$$\mathbf{P}^{[h_t]}\mathbf{u}^{[h_t]} \leqslant M_t^L, \qquad \mathbf{P}^{[h_t]}\mathbf{v}^{[h_t]} \leqslant M_t^S. \tag{13}$$

Furthermore, the total short position may be controlled relative to the total long position, and vice versa. The primary reason for such joint control is to limit portfolio risk exposure and it is often dependent on fund manager's investment style or policy. Consequently, we identify two major forms of portfolio risk control; those arising from direct position control due to policy and those that control investments due to portfolio risk measurement. We consider several forms of direct position control first; in Section 4, several risk metrics for portfolio risk control are discussed.

3.4. *Excessive shortsale risk*

While position limits or leverage constraints themselves are forms of policy-oriented risk controls, the strategies that belong to this category control relative positions of long and short exposure. The first of this type considered here is *excessive shortsale risk*, or ESR. In this type of control, the total short position is controlled not to exceed a certain fraction of the total long position, within a pre-identified group of securities. For instance, in order to limit the short-selling risk within a volatile sector, such as the Internet stocks, the managerial policy may be that no more than 30% of the long position in Internet stocks may be tied in short positions within the same sector. Such a policy may also be interpreted as some sort of margin requirement within a certain group of securities.

In a more generalized setting, such margin requirements may apply to different groups of stocks differently. Suppose there are K groups of stocks whose index set is denoted by G_k, for $k = 1, \ldots, K$, such that $G_k \subseteq \{1, \ldots, N\}$. Given the value of the allowable shortsale fraction for group G_k of stocks as ρ_k, these constraints are

$$\sum_{j \in G_k} P_j^{[h_t]} \max\{0, -x_j^{[h_t]}\} \leqslant \rho_k \sum_{j \in G_k} P_j^{[h_t]} \max\{0, x_j^{[h_t]}\}, \quad k = 1, \ldots, K. \tag{14}$$

Although the allowable fractions ρ_k may be specified such that they are dependent on the time period or the historical scenario being followed, such dependencies are ignored here for the ease of exposition. Noting that (14) is nonlinear, an alternative linear formulation is available under the condition that $\rho_k \leqslant 1$.

Proposition 3.1. *If* $\rho_k \leqslant 1$ *for* $k = 1, \ldots, K$, *then the nonlinear margin constraints in* (14) *have the following equivalent linear representation*

$$\sum_{j \in G_k} P_j^{[h_t]} v_j^{[h_t]} - \rho_k \sum_{j \in G_k} P_j^{[h_t]} u_j^{[h_t]} \leqslant 0, \quad k = 1, \ldots, K. \tag{15}$$

Proof. Let $\bar{\mathbf{x}}^{[h_t]}$ $(\in \Re^N)$ be a feasible solution in (14). Then, construct the nonnegative pair of vectors $(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \in \Re^N$ such that, for $j = 1, \ldots, N$, $\bar{u}_j := \max\{0, \overline{X}_j\}$ and $\bar{v}_j := \max\{0, -\overline{X}_j\}$. Observe that the pair $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ satisfies the inequality in (15); furthermore, it satisfies the equality in (12) since

$$\begin{aligned}\bar{u}_j - \bar{v}_j &= \max\{0, \bar{x}_j^{[h_t]}\} - \max\{0, -\bar{x}_j^{[h_t]}\} \\ &= \max\{0, \bar{x}_j^{[h_t]}\} - \max\{\bar{x}_j^{[h_t]}, \bar{x}_j^{[h_t]} - \bar{x}_j^{[h_t]}\} + \bar{x}_j^{[h_t]} \\ &= \bar{x}_j^{[h_t]},\end{aligned}$$

hence, a feasible solution to the linear representation of the shortsale constraints.

Conversely, it will be shown that given a pair $(\hat{\mathbf{u}}^{[h_t]}, \hat{\mathbf{v}}^{[h_t]})$ that is feasible in (15), one obtains a feasible solution $\hat{\mathbf{x}}$ in (14) using the construction:

$$\hat{\mathbf{x}} := \hat{\mathbf{u}}^{[h_t]} - \hat{\mathbf{v}}^{[h_t]}.$$

Defining, $\hat{\delta}_j := \min\{\hat{\mathbf{u}}^{[h_t]}, \hat{\mathbf{v}}^{[h_t]}\}$, it follows that $\hat{\delta}_j \geqslant 0$ and

$$\left.\begin{aligned}\max\{0, \hat{x}_j\} &= \max\{0, \hat{u}_j^{[h_t]} - \hat{v}_j^{[h_t]}\} = \hat{u}_j^{[h_t]} + \max\{-\hat{u}_j^{[h_t]}, -\hat{v}_j^{[h_t]}\} \\ &= \hat{u}_j^{[h_t]} - \hat{\delta}_j, \\ \max\{0, -\hat{x}_j\} &= \max\{0, \hat{v}_j^{[h_t]} - \hat{u}_j^{[h_t]}\} = \hat{v}_j^{[h_t]} + \max\{-\hat{v}_j^{[h_t]}, -\hat{u}_j^{[h_t]}\} \\ &= \hat{v}_j^{[h_t]} - \hat{\delta}_j.\end{aligned}\right\} \tag{16}$$

Using this "hat" solution and substituting from (16) in (14):

$$\begin{aligned}&\sum_{j \in G_k} P_j^{[h_t]} \max\{0, -\hat{x}_j\} - \rho_k \sum_{j \in G_k} P_j^{[h_t - 1]} \max\{0, \hat{x}_j\} \\ &= \sum_{j \in G_k} P_j^{[h_t]} \big(\hat{v}_j^{[h_t]} - \hat{\delta}_j\big) - \rho_k \sum_{j \in G_k} P_j^{[h_t]} \big(\hat{u}_j^{[h_t]} - \hat{\delta}_j\big) \\ &= \left[\sum_{j \in G_k} P_j^{[h_t]} \hat{v}_j^{[h_t]} - \rho_k \sum_{j \in G_k} P_j^{[h_t]} \hat{u}_j^{[h_t]}\right] - (1 - \rho_k)\left[\sum_{j \in G_k} P_j^{[h_t]} \hat{\delta}_j\right] \\ &\leqslant \sum_{j \in G_k} P_j^{[h_t]} \hat{v}_j^{[h_t]} - \rho_k \sum_{j \in G_k} P_j^{[h_t]} \hat{u}_j^{[h_t]} \quad (\text{since } \rho_k \leqslant 1 \text{ and } \hat{\delta}_j \geqslant 0) \\ &\leqslant 0 \quad (\text{since } (\hat{\mathbf{u}}^{[h_t]}, \hat{\mathbf{v}}^{[h_t]}) \text{ is feasible in (15)}).\end{aligned}$$

This completes the proof. □

The condition that $\rho_k \leqslant 1$ implies that the total short position in the group cannot exceed the total long position. Thus, the use of ESR constraints allows the fund manager to keep the short investment within the total long investment with respect to a given group of securities. However, often a portfolio is required to be *neutral* in its long and short investments so as to immune the portfolio from any particular directional risk. Two types of portfolio neutral strategies will be discussed in the next section: beta neutrality and dollar neutrality.

3.5. Degree of portfolio neutrality

Portfolio neutrality is provided by hedging strategies that balance investments among carefully chosen long and short positions. Fund managers use such strategies to balance the portfolio so as to buffer it from severe market swings, for instance, see Nicholas (2000) and Jacobs and Levy (2004). Alternatively, a prescribed level of imbalance or nonneutrality may be specified in order for the portfolio to maintain a given bias with respect to market swings. An important metric of portfolio bias relative to the broader market is the *portfolio beta*. A balanced investment such that portfolio beta is zero is considered a perfectly beta neutral portfolio and such a strategy is uncorrelated with the market return. *Beta* is the measurement of a stock's volatility relative to the market. A stock with a beta of 1 moves historically in sync with the market, while a stock with a higher beta tends to be more volatile than the market and a stock with a lower beta can be expected to rise and fall more slowly than the market. Many practitioners of beta neutral long/short trading balance their positions in the same sector or industry. By being sector neutral, the risk of market swings affecting some industries or sectors differently than others may be avoided.

The degree of market-neutrality of the portfolio measures the level of correlation of performance of the portfolio with an underlying broad-market index. Typically, the S&P 500 index may be used as the market barometer; alternatively, if the portfolio is constructed out of the S&P 100 stocks, the S&P 100 index may serve as the underlying market performance metric. Let $\beta_j^{[h_t]}$ be the beta of the security j given the sequence of historical prices observed, as indicated by scenario h_t. Then, $\beta_j^{[h_t]}$ is the (estimated conditional) covariance of the rates of return between the security j and the chosen market barometer (index), scaled by the variance of the market rate of return. Since $r_j^{[h_t]}$ is the random variable representing the rate of return of security j, by denoting the market index rate of return by the random variable $R^{[h_t]}$, it follows that

$$\beta_j^{[h_t]} := \frac{\mathrm{Cov}(r_j^{[h_t]}, R^{[h_t]})}{\mathrm{Var}(R^{[h_t]})}. \tag{17}$$

At a portfolio rebalancing decision node h_t, the resulting portfolio beta $\beta_P^{[h_t]}$ can be obtained by

Proposition 3.2. *At some trading node $[h_t]$, having observed the price vector $\mathbf{P}^{[h_t]}$ at the beginning of period t on the portfolio positions $\mathbf{x}^{[h_{t-1}]}$, let the portfolio value before trading be $w^{[h_{t-1}]}$. Suppose the portfolio is rebalanced at the current node to establish the new security positions $\mathbf{x}^{[h_t]}$. Then, the resulting portfolio beta, $\beta_P^{[h_t]}$, is*

$$\beta_P^{[h_t]} = \frac{1}{w^{[h_{t-1}]}}\left(\sum_{j=1}^{N} \beta_j^{[h_t]} P_j^{[h_t]} x_j^{[h_t]}\right). \tag{18}$$

Proof. Since the projected portfolio value at period t is given by (8), the rate of return of portfolio for period t is the random variable $r_P^{[h_t]} := (w^{[h_t]} - w^{[h_{t-1}]})/w^{[h_{t-1}]}$ and thus, $r_P^{[h_t]} = [G^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) + B_t]/w^{[h_{t-1}]}$. Referring to (7), the conditional covariance given $\mathbf{r}^{[h_{t-1}]}$ and $R^{[h_{t-1}]}$, is

$$\begin{aligned}\operatorname{Cov}\bigl(r_P^{[h_t]}, R^{[h_t]}\bigr) &= \frac{1}{w^{[h_{t-1}]}}\operatorname{Cov}\bigl(G^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}), R^{[h_t]}\bigr)\\ &= \frac{1}{w^{[h_{t-1}]}}\sum_j P_j^{[h_t]} x_j^{[h_t]} \operatorname{Cov}\bigl(r_j^{[h_t]}, R^{[h_t]}\bigr),\end{aligned}$$

since the cash rate κ_t is assumed fixed (and thus it has zero correlation with the market). The result in the proposition then follows. □

To control the portfolio beta at a level $\gamma_0 \pm \gamma_1$, the constraints $\gamma_0 - \gamma_1 \leqslant \beta_P^{[h_t]} \leqslant \gamma_0 + \gamma_1$ must be imposed at each trade rebalancing decision node $[h_t]$. By the nonnegativity in (9), the required portfolio beta constraints are

$$(\gamma_0 - \gamma_1) w^{[h_{t-1}]} \leqslant \mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\beta}^{[h_t]})\mathbf{x}^{[h_t]} \leqslant (\gamma_0 + \gamma_1) w^{[h_{t-1}]}. \tag{19}$$

Portfolio beta is set at γ_0 with a tolerance level of γ_1. With $\gamma_1 \approx 0$, rebalancing strives for a portfolio beta of γ_0. In particular, with $\gamma_0 = \gamma_1 \approx 0$, the model attempts for a perfectly beta-neutral dynamic investment in the portfolio.

In addition to the (degree of) market neutrality afforded by the above beta constraints, fund managers often seek additional neutrality through *dollar neutrality*. In order to create a portfolio with dollar neutrality, equal amounts of long and short investments are made, or alternatively, a specific *dollar imbalance* may be prescribed as policy. Given a targeted maximum dollar imbalance of I_t^D, the "degree of dollar neutrality" constraint is

$$\left|\sum_{j=1}^{N} P_j^{[h_t]} \max\bigl\{0, x_j^{[h_t]}\bigr\} - \sum_{j=1}^{N} P_j^{[h_t]} \max\bigl\{0, -x_j^{[h_t]}\bigr\}\right| \leqslant I_D.$$

Under the variable description in (12), the equivalent linear constraints are:

$$-I_t^D \leqslant \mathbf{P}^{[h_t]}(\mathbf{u}^{[h_t]} - \mathbf{v}^{[h_t]}) \leqslant I_t^D. \tag{20}$$

If portfolio risk control is solely via market neutral long/short trading as an act of portfolio balancing, then it would generally involve a large amount of buying and selling. Such a strategy naturally leads to additional risks as the fund manager now needs to have the ability to execute trades efficiently as well as to keep brokerage costs from severely affecting trading profits. Furthermore, fund managers must also trade in very liquid stocks, indicating a high level of daily volume, in order to ensure they can get quickly in and out of positions, as happens in frequent trading. Therefore, portfolio risk management should not be viewed solely through such policy or balancing strategies. Furthermore, in order to keep trading costs relatively low, relative to short term rates of returns of securities, a multiperiod trading view must be taken when portfolio rebalancing and risk controls must be applied within such a setting.

4. Portfolio risk metrics

While the foregoing discussion alluded to risk control via policy statements, portfolio risk management is closely tied with the ability to describe future uncertainty of stock returns. This is already apparent in the beta neutrality constraints where a vector of stock (conditional) beta, $\boldsymbol{\beta}^{[h_t]}$, must be estimated at each decision node, given a historical scenario h_t. In addition, there may be other parameters of stock returns that need to be estimated, most notably, the conditional expectation of stock returns, $\boldsymbol{\mu}^{[h_t]}$, the conditional standard deviation of stock returns, $\boldsymbol{\sigma}^{[h_t]}$, or the full variance–covariance matrix of returns, $\boldsymbol{\Sigma}^{[h_t]}$.

Once the required parameters have been forecasted at node $[h_t]$, the conditional (descendant) outcomes of security returns for period $t+1$ must be generated, denoted by $\mathbf{r}^{[h_{t+1}]} \in \{\mathbf{r}^{[h_t,1]}, \ldots, \mathbf{r}^{[h_t,L_t]}\}$. That is, a specific *scenario generator* must also be utilized at decision node $[h_t]$ to generate the sample of L_t outcomes. It is this scenario generator that endows the decision model with the multistage scenario tree structure. In this section, we assume that all of the required forecasted parameters, as well as a scenario tree depicting an *approximated view of decision epochs* in the future are available.

Portfolio risk control at a current decision node, as well as along any anticipated sequence of decision epochs, has been one of the major subjects of discussion in multistage financial stochastic optimization. Portfolio wealth variation around its expectation is often used as a way of measuring portfolio risk, such as the case in the classical Markowitz mean-variance analysis, Markowitz (1959). An alternative to mean-variance risk trade-off is the expected utility of wealth at a current decision node. The first formal axiomatic treatment of utility was given by von Neumann and Morgenstern (1944). Other objective functions are possible, such as the one proposed by Zhao and Ziemba (2001). Artzner et al. (1999) introduced the concept of coherent risk measures. This landmark paper initiated a wealth of literature to follow on coherent risk measures with several interesting extensions, see, for instance, Jarrow (2002), Delbaen (2000), Roorda, Engwerda and Schumacher (2002), and Follmer and Schied (2002). Coherent risk measures scale linearly if the underlying uncertainty is changed, and due to this linearity, coherency alone does not lead to risk measures that are useful in applications. As discussed in Warachka, Zhao and Ziemba (2004), one important limitation of coherent risk measures is its inability to yield sufficient diversification to reduce portfolio risk. Alternatively, they propose a methodology that defines risk on the domain of portfolio holdings and utilizes quadratic programming to "measure" portfolio risk.

An alternative method of risk measurement is to use the conditional value-at-risk (CVaR), see, e.g., Rockafellar and Uryasev (2000) and Ogryczak and Ruszczynski (2002). Risk measures based on mean and CVaR are coherent, see Rockafellar, Uryasev and Zabarankin (2002). Such risk measures evaluate portfolio risk according to its value in the worst possible scenario or under the probability measure that produces the largest negative outcome. CVaR can also be used, just as in moment-based approximations, see Edirisinghe (1999), for approximating a given sample of scenarios. Such an approach is taken in Bychkov and Edirisinghe (2004) where CVaR is used as a metric for approximating scenarios for financial investment problems.

In the sequel, we view a risk metric at a decision node h_t as either *Static Risk Control* (SRC), or *Dynamic Risk Control* (DRC). In SRC, the computation of risk, for taking a rebalancing decision $\mathbf{x}^{[h_t]}$, requires only the forecasted parameters of a distribution (of uncertainty). That is, such a risk metric does not require a distributional assumption or a specific approximated random sample from the distribution, but risk can be specified through a closed-form expression. As an example, one may consider variance of the portfolio as a static risk metric, as in mean-variance analysis. One advantage of SRC is that risk computation remains independent of the scenario samples being generated. Then, scenario trees can be generated to guide the rebalancing decisions carefully, without imparting a sample bias in risk computation. With small sample scenario tress, the resulting multistage stochastic optimization models become computationally more attractive. In contrast, with DRC risk controls, computing the risk associated with a rebalancing decision $\mathbf{x}^{[h_t]}$ would require further distributional knowledge of the random vector $\mathbf{r}^{[h_t]}$, and also possibly of the random vectors $\mathbf{r}^{[h_{t+1}]}, \ldots, \mathbf{r}^{[h_T]}$. As an example of DRC, consider CVaR which would require a sample of outcomes for $\mathbf{r}^{[h_t]}$. If one hopes to eliminate possible sampling bias, a sufficiently large number of outcomes (i.e., L_t) must be generated at each decision node for computation of CVaR. In multistage stochastic programming, such a practice would lead to enormous computational difficulties. Therefore, from a pure computational standpoint, SRC risk metrics may appear appealing. However, as it turns out, each risk metric endows the portfolio with different characteristics, as will be discussed later.

Several static controls and dynamic controls will be developed in the remaining sections and applied in the investment optimization model.

4.1. *Static risk control (SRC)*

Two types of risk metrics are considered in the context of SRC, where risk expressions can be obtained in closed-form based on the forecasted parameters of uncertainty at a given decision node $[h_t]$. Such SRC may be applied at all rebalancing nodes in the decision (scenario) tree. The first is a variant of portfolio variance, and the second is what is termed the portfolio *catastrophic risk*.

The use of portfolio variance as a risk expression and using it within a static mean variance trade-off optimization model date back to Markowitz (1959). Its multistage extensions, where mean-variance trade-off is specified at each of the decision nodes in a scenario tree, too are considered in the literature, see, e.g., Steinbach (2001) and the many references therein, and also Gulpinar, Rustem and Settergren (2003). The mean-variance trade-off is a quadratic programming model, and thus, the application of quadratic programming on a multistage stochastic programming model is computationally tedious.

In our development here, a generalized approach to portfolio (quadratic) tracking penalty is considered and portfolio variance is obtained as a special case. Consider a target return that portfolio wishes to track. Such targets can be either deterministic or stochastic, and also, they can be exogenously or endogenously specified, see Geyer et al. (2003). For instance, tracking the portfolio mean is a target specified endogenously, but deterministic. An example of a stochastic (and exogenous) target is the return on a broad market index, which may be correlated with the portfolio securities. Index target is considered exogenous because it is assumed that portfolio rebalancing does not affect the broader market index to any appreciable extent. Benchmark tracking for portfolio management has been well-studied in the literature, see, for example, Dembo and King (1992), Frino and Gallagher (2001), Jansen and van Dijk (2002), Roll (1992), and El-Hassan and Kofman (2003). However, the approach taken here is slightly different, as discussed below.

4.1.1. *Tracking risk control*

Consider a benchmark (stochastic) target, such as the return on a broad market index, say the S&P 500 index. Let the conditional rate of return (RoR) on the benchmark index, at time t at the rebalancing decision node $[h_t]$, be $R^{[h_t]}$, which is a univariate random variable. In a pure investment in the benchmark index, the total investment in the portfolio, given by $\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]}$, yields the target total \$ return $\mathcal{T}^{[h_t]}$ as

$$\mathcal{T}^{[h_t]}(\mathbf{x}^{[h_t]}) = \mathbf{P}^{[h_{t-1}]}\mathbf{x}^{[h_t]}R^{[h_t]}. \tag{21}$$

$\mathcal{T}^{[h_t]}(.)$ is a univariate random variable and it serves as the target return for the portfolio for period t. The conditional expectation of the quadratic tracking penalty for portfolio return on risky investments not following the above target return is

$$E_{\mathbf{r}^{[h_t]}}\left\{\left[\mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} - \mathcal{T}^{[h_t]}\right]^2 \mid h_t\right\}. \tag{22}$$

Denoting the portfolio standard deviation by $\sigma_P^{[h_t]}$, its variance is

$$\left[\sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})\right]^2 = (\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]})'\boldsymbol{\Sigma}^{[h_t]}(\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]}), \tag{23}$$

where $\boldsymbol{\Sigma}^{[h_t]}$ is the (conditional) variance–covariance matrix of the random vector $\mathbf{r}^{[h_t]}$.

Proposition 4.1. *An upper bounding risk metric on the tracking penalty in (22) is*

$$\mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]}) := \big[\sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})\big]^2 + \big[\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\mu}^{[h_t]})\mathbf{x}^{[h_t]} - \mu_R^{[h_t]}\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]}\big]^2 + \big[\sigma_R^{[h_t]}\big(\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\beta}^{[h_t]})\mathbf{x}^{[h_t]} - \mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]}\big)\big]^2, \tag{24}$$

where $\mu_R^{[h_t]}$ is the index rate of return and $\sigma_R^{[h_t]}$ is the index standard deviation.

Proof. Note that

$$\begin{aligned}
&E_{\mathbf{r}^{[h_t]}}\big[\mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} - \mathcal{T}^{[h_t]}\big]^2 \\
&\quad = E_{\mathbf{r}^{[h_t]}}\big[\mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]}\big]^2 - 2E_{\mathbf{r}^{[h_t]}}\big[\mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]}\mathcal{T}^{[h_t]}\big] + E_{\mathbf{r}^{[h_t]}}[\mathcal{T}^{[h_t]}]^2 \\
&\quad = \big[\sigma_P^{[h_t]}\big]^2 + \big[\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\mu}^{[h_t]})\mathbf{x}^{[h_t]}\big]^2 - 2(\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]})\mathbf{P}^{[h_t]}\mathbf{D}\big(E[R^{[h_t]}\mathbf{r}^{[h_t]}]\big)\mathbf{x}^{[h_t]} \\
&\qquad + (\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]})^2 E_{\mathbf{r}^{[h_t]}}[R^{[h_t]}]^2.
\end{aligned}$$

Using the expression

$$\begin{aligned}
E\big[R^{[h_t]}r_j^{[h_t]}\big] &= E[R^{[h_t]}]E\big[r_j^{[h_t]}\big] + \mathrm{Cov}\big(R^{[h_t]}, r_j^{[h_t]}\big) \\
&= \mu_R^{[h_t]}\mu_j^{[h_t]} + \beta_j^{[h_t]}\big[\sigma_R^{[h_t]}\big]^2,
\end{aligned}$$

and upon algebraic manipulation, it follows that

$$\begin{aligned}
&E_{\mathbf{r}^{[h_t]}}\big[\mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} - \mathcal{T}^{[h_t]}\big]^2 \\
&\quad = \mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]}) - \big[\sigma_R^{[h_t]}\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\beta}^{[h_t]})\mathbf{x}^{[h_t]}\big]^2 \leqslant \mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]}),
\end{aligned}$$

due to the nonnegativity of the second term in the above equality expression. □

Observe that the risk metric $\mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]})$ can be interpreted as the sum of the three components,

$$\begin{bmatrix}\text{inherent risk due to}\\ \text{security correlations}\end{bmatrix} + \begin{bmatrix}\text{risk due to not tracking}\\ \text{benchmark mean return}\end{bmatrix} + \begin{bmatrix}\text{risk due to portfolio}\\ \textit{beta}\text{ not being 1.0}\end{bmatrix}.$$

In particular, when the target being tracked is simply the portfolio mean itself, then the above risk metric is

Corollary 4.2. *If the target is specified as $\mathcal{T}^{[h_t]} = \mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\mu}^{[h_t]})\mathbf{x}^{[h_t]}$, then the risk metric becomes $\mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]}) = [\sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})]^2$, i.e., the usual portfolio variance.*

The risk metric $\mathcal{R}_{\mathcal{T}}^{[h_t]}(.)$ in (24) is of *static-type* as it is computable under the forecasted parameters, and also it is convex quadratic in the portfolio positions $\mathbf{x}^{[h_t]}$. The use of such a risk metric in the portfolio model is to seek a trade-off between the risk metric and the portfolio expected net return after rebalancing. This objective, at the current rebalancing node, is

$$\text{maximize}\; G^{[h_t]}(\boldsymbol{\mu}^{[h_t]}, \mathbf{x}^{[h_t]}) - \lambda_{\mathcal{T}}^t\mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]}), \tag{25}$$

where the function $G^{[h_t]}$ is defined in (7), and $\lambda_{\mathcal{T}}^t$ is herein termed a (static) *tracking risk aversion* (TRA) coefficient, which is nonnegative.

4.1.2. Catastrophic risk control

A second form of significant portfolio risk is next considered, largely motivated by practitioners. This form of risk is concerned with the direction of (the future) price of a security being opposite to the sign of the established position in the portfolio. That is, securities in a *long* portfolio fall in price while the securities in a *short* portfolio rise in price. Such risk is often the result of error in forecasting the direction of stock price movement. This would entail observing a drop in price for long securities and an increase in price for shorted securities. Generally, there is no formal mechanism to safeguard against the risk posed by such a *catastrophic* event. Controlling the portfolio variance, or more generally the tracking risk metric $\mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]})$ discussed above, does not necessarily counter the effects of *catastrophic risk*, abbreviated herein as *Cat risk*. The effects of Cat risk control in a portfolio is demonstrated under the section on model application.

We define Cat risk as the anticipated total dollar wealth loss in the event stock returns move against the portfolio positions by one standard deviation, denoted by $\mathcal{R}_{\mathcal{C}}^{[h_t]}(\mathbf{x}^{[h_t]})$, where the positions $\mathbf{x}^{[h_t]}$ are established at event node $[h_t]$ for period t. That is,

$$\mathcal{R}_{\mathcal{C}}^{[h_t]}(\mathbf{x}^{[h_t]}) := \sum_{j=1}^{N} P_j^{[h_t]} \sigma_j^{[h_t]} \left| x_j^{[h_t]} \right|, \tag{26}$$

which is thus a form of static risk control. While the Cat risk expression is free of correlations among securities, it can be shown that it provides an upper bound on the portfolio standard deviation.

Proposition 4.3. *The portfolio standard deviation is a lower bound on the catastrophic risk, i.e.,* $\mathcal{R}_{\mathcal{C}}^{[h_t]}(\mathbf{x}^{[h_t]}) \geqslant \sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})$ *where* $\sigma_P^{[h_t]}(.)$ *is given by* (23).

Proof. Defining the random variable $\xi_j := P_j^{[h_t]} x_j^{[h_t]} r_j^{[h_t]}$, variance of the portfolio is $\mathrm{Var}(\sum_{j=1}^{N} \xi_j)$. Denoting the standard deviation of ξ_j by $\hat{\sigma}_j$ and the correlation between ξ_i and ξ_j by $\hat{\rho}_{ij}$,

$$\begin{aligned} \mathrm{Var}\left(\sum_{j=1}^{N} \xi_j\right) &= \sum_{j=1}^{N} \hat{\sigma}_j^2 + 2 \sum_{(i,j), i \neq j} \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij} \\ &\leqslant \sum_{j=1}^{N} \hat{\sigma}_j^2 + 2 \sum_{(i,j), i \neq j} \hat{\sigma}_i \hat{\sigma}_j = \left(\sum_{j=1}^{N} \hat{\sigma}_j\right)^2. \end{aligned}$$

Noting that $\mathrm{Var}(\xi_j) = [P_j^{[h_t]} x_j^{[h_t]} \sigma_j^{[h_t]}]^2$, and since the prices are nonnegative, we have $\hat{\sigma}_j = P_j^{[h_t]} \sigma_j^{[h_t]} |x_j^{[h_t]}|$. Therefore, it follows that the portfolio variance is bounded from above by $[\mathcal{R}_C^{[h_t]}(\mathbf{x}^{[h_t]})]^2$. □

Since portfolio standard deviation is only a lower bound on Cat risk, controlling portfolio variance, as in Markowitz-style, is not guaranteed to provide adequate protection against catastrophic risk. In contrast, controlling Cat risk will certainly control the portfolio variance directly, and thus, Cat risk metric plays a dual role in portfolio rebalancing. The two risk metrics, $\mathcal{R}_C^{[h_t]}(\mathbf{x}^{[h_t]})$ and $\sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})$, however, have distinct characteristics in shaping portfolio positions, as will be demonstrated numerically later. Geometrically, Cat risk (as a function of portfolio positions) bounds the portfolio standard deviation by a polyhedral convex cone with apex at the origin.

While it is possible to incorporate $\mathcal{R}_C^{[h_t]}(\mathbf{x}^{[h_t]})$ in to the objective function with a risk aversion coefficient, similar to (25), it is easier to specify the maximum allowable $ loss in the portfolio for a catastrophic move (of, say, one standard deviation) as a constraint. The Cat risk constraint is then

$$\mathcal{R}_C^{[h_t]}(\mathbf{x}^{[h_t]}) \leqslant \Gamma_t, \tag{27}$$

where $\$\Gamma_t$ is the pre-specified dollar loss. Since $\mathcal{R}_C^{[h_t]}(.)$ is nonlinear in its arguments, a linear representation is obtained by appealing to the pairs of nonnegative (long/short) variables given in (12). Then, the following linear constraint replaces the Cat risk constraint in (27),

$$\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\sigma}^{[h_t]})(\mathbf{u}^{[h_t]} + \mathbf{v}^{[h_t]}) \leqslant \Gamma_t. \tag{28}$$

Since it is possible to easily argue that (27) and (28) provide equivalent expressions for risk control, details are skipped here. Note that (28) limits positions (long or short) in securities with high volatility (as measured by $\sigma_j^{[h_t]}$), and thus provides a form of *indirect* position control.

4.2. Dynamic risk control (DRC)

As defined earlier, in DRC, risk metric computation requires the availability of a scenario sample (conditional upon the history h_t), in addition to the forecasted parameters, at a rebalancing decision node $[h_t]$. In the *tracking* risk metric, portfolio wealth variation around a target is penalized in either direction. In the case when the target is the portfolio mean, the result is the Markowitz mean-variance trade-off. Mean-variance optimal portfolios are shown to be (stochastically) dominated by carefully constructed portfolios. Such is the case when one penalizes only the variation of portfolio value on the downside, i.e., the risk metrics based on *downside* deviation.

The relative merits of using Markowitz mean-variance type models and those that trade off mean with downside semi-deviation are examined in Ogryczak and Ruszczynski (1999). The semi-deviation risk trade-off approach yields superior portfolios that are efficient with respect to the standard stochastic dominance rules, see Whitmore and Findlay (1978). When quadratic penalty is applied on the downside deviations, with target defined at the portfolio mean, it is called the downside semi-variance risk metric. The concept of semi-variance was described by Markowitz (1959, Chapter IX). Semi-variance fails to satisfy the positive homogeneity property required for a coherent risk measure, see Artzner et al. (1999) and Rockafellar, Uryasev and Zabarankin (2002).

4.2.1. Quadratic downside risk control

The approach of downside deviation, as applied in this chapter, is slightly more general than the downside semi-variance. A target is specified by translating the portfolio mean by an amount determined by the relative volatilities of the underlying securities. The translation could be on the upside or the downside of the mean. Define the *volatility-adjusted-mean* (VAM), denoted by $\nu_j^{[h_t]}$, as

$$\nu_j^{[h_t]} := \frac{|\mu_j^{[h_t]}|}{\sigma_j^{[h_t]}} \tag{29}$$

and specify the wealth target by

$$\mathcal{T}_\theta^{[h_t]}(\mathbf{x}^{[h_t]}) = \sum_{j=1}^{N} P_j^{[h_t]}\big(\mu_j^{[h_t]} + \theta\nu_j^{[h_t]}\big)x_j^{[h_t]}, \tag{30}$$

where θ is a (positive or negative) scalar. For $\theta > 0$, the target is specified above the mean, and for $\theta < 0$, target is below the portfolio mean; however, the resulting target being dependent on the relative risk of an individual security, as measured by $\nu_j^{[h_t]}$. Then, the quadratic risk metric for downside deviation from this target is

$$\mathcal{R}_{\mathcal{Q},\theta}^{[h_t]}(\mathbf{x}^{[h_t]}) := E_{\mathbf{r}^{[h_t]}}\big(\big[\mathcal{T}_\theta^{[h_t]}(\mathbf{x}^{[h_t]}) - \mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]}\big]^+\big)^2. \tag{31}$$

For $\theta = 0$, $\mathcal{R}_{\mathcal{Q},0}^{[h_t]}(\mathbf{x}^{[h_t]})$ is the usual downside semi-variance risk metric. Note that $\theta > 0$ yields more risk aversion ($\theta < 0$ yields less risk aversion) than the standard semi-variance metric. The computation of (31) requires distributional information of the (conditional) security returns, or at least to have a discrete scenario sample to compute it as a finite summation. In the sequel, $\mathcal{R}_{\mathcal{Q},0}^{[h_t]}(\mathbf{x}^{[h_t]})$ is computed based on the descendant outcome sample of scenarios $\mathbf{r}^{[h_{t-1},1]}, \ldots, \mathbf{r}^{[h_{t-1},L_t]}$ at node $[h_t]$. Therefore, choosing a relatively small sample size (L_t) would certainly lead to significant sample bias in the risk metric, while increasing the size of L_t leads to a heavy computational burden. The reason for the latter computational difficulty becomes clear when incorporating $\mathcal{R}_{\mathcal{Q},\theta}^{[h_t]}(\mathbf{x}^{[h_t]})$ in to the trade optimization at node $[h_t]$, i.e.,

$$\text{maximize}\; G^{[h_t]}(\boldsymbol{\mu}^{[h_t]}, \mathbf{x}^{[h_t]}) - \lambda_{\mathcal{Q}}^t \mathcal{R}_{\mathcal{Q},\theta}^{[h_t]}(\mathbf{x}^{[h_t]}), \tag{32}$$

where the function $G^{[h_t]}$ is defined in (7), and $\lambda^t_{\mathcal{Q}}$ is herein termed the (dynamic) *quadratic risk aversion* (QRA) coefficient, which is nonnegative. To evaluate the risk term in (32), the expectation in (31) must be specified in computable terms, by defining downside deviation variables, $d^{[h_t]}$, over the domain of $\mathbf{r}^{[h_t]}$ as

$$d^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) := \max\{0, \mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} - \mathcal{T}^{[h_t]}_{\theta}(\mathbf{x}^{[h_t]})\}. \tag{33}$$

Then, the optimization in (32) becomes

$$\text{maximize}\, G^{[h_t]}(\boldsymbol{\mu}^{[h_t]}, \mathbf{x}^{[h_t]}) - \lambda^t_{\mathcal{Q}} E_{\mathbf{r}^{[h_t]}}\left[d^{[h_t]}(\mathbf{x}^{[h_t]})\right]^2, \tag{34}$$

where the deviation variables satisfy the constraints

$$\left.\begin{aligned} &d^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant \mathbf{P}^{[h_t]}, \mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} - \mathcal{T}^{[h_t]}_{\theta}(\mathbf{x}^{[h_t]}),\\ &d^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant 0, \end{aligned}\right\} \tag{35}$$

and these constraints will be imposed along each sample outcome $\mathbf{r}^{[h_t],l_t}$, for $l_t = 1, \ldots, L_t$. Thus, as the sample size is increased at node $[h_t]$, the size of the resulting *nodal* optimization model will also increase (unlike in the case of SRC).

4.2.2. Drawdown risk control (DDR)

A second type of dynamic risk control of paramount importance to fund managers is the so-called *portfolio drawdown* control. Investors and fund managers do not wish to see the value of the portfolio decline considerably over time. Such drastic declines in portfolio value may lead to perceptions that the fund is *too risky*; it may even lead to losing important client accounts from the fund. Portfolio drawdown is defined as the relative equity loss from the highest peak to the lowest valley of a portfolio value decline within a given window of observation. Mathematically, given the current rebalancing node $[h_t]$ and the portfolio wealth trajectory $w^{[h_\tau]}$, for $\tau \in \{t - \hat{t}, t-1\}$ where $1 \leqslant \hat{t} \leqslant t-1$, the maximum portfolio value in the $\hat{t}$ time window is

$$w^{[h_t]}_{\max} := \max_{\tau = t-\hat{t}, \ldots, t-1} w^{[h_\tau]}. \tag{36}$$

The portfolio, under the revised portfolio positions $\mathbf{x}^{[h_t]}$, is said to have $\hat{t}$-period drawdown under return vector $\mathbf{r}^{[h_t]}$ if $w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) < w^{[h_t]}_{\max}$ holds, and in this case, the relative drawdown is measured by

$$\frac{w^{[h_t]}_{\max} - w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})}{w^{[h_t]}_{\max}}.$$

For instance, suppose a fund manager has logged the portfolio wealth at the beginning of a year at \$5 million, which reaches a peak in June to \$8 million, and then loses its value to \$6 million by the end of the year. Thus, for the period of one year, the fund had a relative drawdown of $(8-6)/8$, or 25%, while the fund has an annual RoR of $(6-5)/5$, or 20%. Portfolio performance, as measured by the reward-to-drawdown (RTD) ratio, is $20/25 = 0.8$, and in this example, it is indicative of poor portfolio performance. Successful money managers strive for ratios of 2 or better. Another important issue in portfolio drawdown is the time it takes to recover from the (drawdown) losses, termed as the drawdown *recovery time*, relative to the drawdown *duration time*. Indeed, shorter recovery times are preferable in money management. Aiming for the highest portfolio return, usually, does not translate to a high RTD ratio. While individuals may have the intention for fast portfolio growth with a "stomach" for large and repeated drawdowns, a low RTD ratio is not a trait of a successful money manager.

Regulatory control in portfolio drawdown makes fund manager's task quite challenging. For instance, according to Chekhlov, Uryasev and Zabarankin (2003), "it is highly uncommon, for a Commodity Trading Advisor (CTA) to still hold a client whose account was in a drawdown, even of small size, for longer than 2 years. By the same token, it is unlikely that a particular client will tolerate a 50% drawdown in an account with an average- or small-risk CTA. Similarly, in an investment bank setup, a proprietary system trader will be expected to make money in 1 year at the longest, i.e., he cannot be in a drawdown for longer than a year. Also, he/she may be shut down if a certain maximal drawdown condition will be breached, which, normally, is around 20% of his backing equity. Additionally, he will be given a warning drawdown level at which he will be reviewed for letting him keep running the system (around 15%). Obviously, these issues make managed accounts practitioners very concerned about both the magnitude and duration of their clients' accounts drawdowns."

Despite this apparent significance of drawdown, theoretical development on a drawdown risk metric is relatively sparse. Grossman and Zhou (1993) consider an exact analytical solution of a maximal drawdown problem for a one-dimensional case under lognormality assumptions. Subsequently, Cvitanic and Karatzas (1995) generalized this work for multiple dimensions. In contrast, Chekhlov, Uryasev and Zabarankin (2003) developed a metric, termed the conditional drawdown-at-risk (CDaR), and applied it along sample paths of portfolio returns with no assumptions on underlying distributions. Nevertheless, all these deal with either a static model that is rolled over in time, or holding constant portfolio weights over the duration of the sample paths.

Since the portfolio drawdowns in multiple periods in the future are affected by the current rebalancing decisions (in the presence of trading costs), applying drawdown risk control within a multistage framework is expected to improve portfolio reward-to-drawdown (RTD) ratio considerably. This is highlighted in the numerical experiments reported in Section 6. To develop our drawdown risk (DDR) metric, it is assumed, without loss of generality, that the drawdown window begins from the root node $[h_1]$ at time $t = 1$, i.e., $\hat{t} = t - 1$ holds in (36) for maximum portfolio value. Define the drawdown random variable at node $[h_t]$ by $z^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})$, i.e.,

$$z^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) := \max\left\{0, w_{\max}^{[h_t]} - w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})\right\}. \tag{37}$$

Our drawdown risk metric is determined with respect to a prescribed level of *acceptable relative drawdown* in the portfolio, herein denoted by the (nonnegative) fraction π. Thus, the DDR metric, denoted by $\mathcal{R}_{\mathcal{D},\pi}^{[h_t]}(\mathbf{x}^{[h_t]})$, takes on value *zero* if

$z^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \leqslant \pi w_{\max}^{[h_t]}$, and it is positive otherwise. Therefore, defining the ($) violations from the acceptable level of drawdown by $q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})$,

$$q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) := \max\bigl\{0, z^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) - \pi w_{\max}^{[h_t]}\bigr\}. \tag{38}$$

The drawdown (quadratic) risk metric is then

$$\mathcal{R}_{\mathcal{D},\pi}^{[h_t]}(\mathbf{x}^{[h_t]}) := E_{\mathbf{r}^{[h_t]}}\bigl[q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})\bigr]^2. \tag{39}$$

Therefore, the nodal trade optimization problem with drawdown risk control at node $[h_t]$ has the objective

$$\text{maximize}\; G^{[h_t]}(\boldsymbol{\mu}^{[h_t]}, \mathbf{x}^{[h_t]}) - \lambda_{\mathcal{D}}^t \mathcal{R}_{\mathcal{D},\pi}^{[h_t]}(\mathbf{x}^{[h_t]}), \tag{40}$$

where the function $G^{[h_t]}$ is defined in (7), and $\lambda_{\mathcal{D}}^t$ is herein termed a (dynamic) *drawdown risk aversion* (DRA) coefficient, which is nonnegative. The constraints that must be imposed to determine the drawdown risk correctly are

$$\left.\begin{aligned} &w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) + q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant (1-\pi) w_{\max}^{[h_t]},\\ &q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant 0, \end{aligned}\right\} \tag{41}$$

where $w_{\max}^{[h_t]}$ must satisfy the following linear constraints along the scenario path up to node $[h_t]$

$$w_{\max}^{[h_t]} - w^{[h_\tau]} \geqslant 0, \quad \forall \tau = 1, \ldots, t-1. \tag{42}$$

The drawdown constraints in (41)–(42) are free of the drawdown variables $z^{[h_t]}(.)$ in (37). Moreover, the constraints (41) must be imposed along each scenario outcome descendant from node $[h_t]$, i.e., for all $\mathbf{r}^{[h_t],l_t}$, $l_t = 1, \ldots, L_t$, which thus makes the computation of the risk metric $\mathcal{R}_{\mathcal{D},\pi}^{[h_t]}(.)$ directly dependent on the generated scenario tree of security returns.

5. Multistage portfolio rebalancing model

In the preceding two sections, the basic constructs of the model for the portfolio rebalancing problem were discussed, as viewed from a given rebalancing decision node $[h_t]$ of the scenario tree. In this section, these model components are collected to construct the multistage dynamic rebalancing stochastic programming model, which allows the manager to determine the optimized portfolio positions at the root node $[h_1]$. Indeed, it is the optimal security positions $\mathbf{x}^{[h_1]}$, revised from the initial positions $\mathbf{x}^0$, that will be implemented, considering the anticipated future of T periods as depicted by the generated scenario tree of security returns.

At each rebalancing node $[h_t]$, define the value function $\phi^{[h_t]}$, which is the optimal objective value after undertaking the optimized portfolio rebalancing. $\phi^{[h_t]}$ depends on the information available prior to initiating portfolio rebalancing, i.e., portfolio holdings $\mathbf{x}^{[h_{t-1}]}$, cash position $C^{[h_{t-1}]}$, and the portfolio wealth trajectory $\mathbf{w}^{[h_{t-1}]} \equiv \{w^{[1]},$

$\ldots, w^{[h_{t-1}]}\}$. Thus, the value function at node $[h_t]$ is $\phi^{[h_t]}(\mathbf{x}^{[h_{t-1}]}, C^{[h_{t-1}]}, \mathbf{w}^{[h_{t-1}]})$. The value function one period later at the descendant scenario node $[h_{t+1}]$, under random return vector $\mathbf{r}^{[h_t]}$, is given by $\phi^{[h_{t+1}]}(\mathbf{x}^{[h_t]}, C^{[h_t]}, \mathbf{w}^{[h_t]})$ where $\mathbf{r}^{[h_t]}$ takes on a (vector) value from the domain $\{\mathbf{r}^{[h_t],l_1}, \ldots, \mathbf{r}^{[h_t],L_t}\}$, the set of outcomes at node $[h_t]$. The value functions $\phi^{[h_t]}$ and $\phi^{[h_{t+1}]}$ are linked by the dynamic recursive trade optimization model at node $[h_t]$, for $t = 1, \ldots, T$, as

$$\begin{aligned}
&\phi^{[h_t]}(\mathbf{x}^{[h_{t-1}]}, C^{[h_{t-1}]}, \mathbf{w}^{[h_{t-1}]}) \\
&\quad := \text{maximize}\, G^{[h_t]}(\boldsymbol{\mu}^{[h_t]}, \mathbf{x}^{[h_t]}) - \lambda^t_{\mathcal{T}} \mathcal{R}^{[h_t]}_{\mathcal{T}}(\mathbf{x}^{[h_t]}) - \lambda^t_{\mathcal{Q}} \mathcal{R}^{[h_t]}_{\mathcal{Q},\theta}(\mathbf{x}^{[h_t]}) \\
&\qquad\qquad - \lambda^t_{\mathcal{D}} \mathcal{R}^{[h_t]}_{\mathcal{D},\pi}(\mathbf{x}^{[h_t]}) + E_{\mathbf{r}^{[h_t]}}\big[\phi^{[h_{t+1}]}(\mathbf{x}^{[h_t]}, C^{[h_t]}, \mathbf{w}^{[h_t]})\big]
\end{aligned} \tag{43}$$

subject to the following sets of constraints on flow balance, policy, and risk control, where the flow balance constraints are

$$\begin{aligned}
&\texttt{(BUDGET):} && \mathbf{P}^{[h_t]}(\mathbf{x}^{[h_t]} - \mathbf{x}^{[h_{t-1}]}) + F(\mathbf{y}^{[h_t]}) + C^{[h_t]} = B_t + (1+\kappa_{t-1})C^{[h_{t-1}]}, \\
& && C^{[h_t]} \geqslant 0, \\
&\texttt{(WEALTH):} && w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) - G^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) = w^{[h_{t-1}]} + B_t, \\
& && w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant 0, \\
&\texttt{(TRADE):} && \mathbf{x}^{[h_t]} - \mathbf{y}^{[h_t]} \leqslant \mathbf{x}^{[h_{t-1}]}, \\
& && \mathbf{x}^{[h_t]} + \mathbf{y}^{[h_t]} \geqslant \mathbf{x}^{[h_{t-1}]}, \\
& && \mathbf{y}^{[h_t]} \geqslant 0, \\
&\texttt{(LNGSHT):} && \mathbf{x}^{[h_t]} - \mathbf{u}^{[h_t]} + \mathbf{v}^{[h_t]} = 0, \\
& && \mathbf{u}^{[h_t]}, \mathbf{v}^{[h_t]} \geqslant 0, \\
&\texttt{(SLIPP):} && F(\mathbf{y}^{[h_t]}) - \sum_{j=1}^{N} \alpha_{0j} y_j^{[h_t]} + \sum_{j=1}^{N} \alpha_{1j} \big(P_j^{[h_t]}/V_j^{[h_t]}\big)\big(y_j^{[h_t]}\big)^2 = 0,
\end{aligned}$$

the policy-type constraints are

$$\begin{aligned}
&\texttt{(LIMITS):} && \mathbf{x}^{[h_t],\min} \leqslant \mathbf{x}^{[h_t]} \leqslant \mathbf{x}^{[h_t],\max}, \\
&\texttt{(LEVERG):} && \mathbf{P}^{[h_t]}\mathbf{u}^{[h_t]} \leqslant M_t^L, \\
& && \mathbf{P}^{[h_{t-1}]}\mathbf{v}^{[h_t]} \leqslant M_t^S, \\
&\texttt{(SHRISK):} && \mathbf{P}^{[h_t]}\mathbf{v}^{[h_t]} - \rho \mathbf{P}^{[h_t]}\mathbf{u}^{[h_t]} \leqslant 0, \\
&\texttt{(PFBETA):} && \mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\beta}^{[h_t]})\mathbf{x}^{[h_t]} \geqslant (\gamma_0 - \gamma_1) w^{[h_{t-1}]}, \\
& && \mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\beta}^{[h_t]})\mathbf{x}^{[h_t]} \leqslant (\gamma_0 + \gamma_1) w^{[h_{t-1}]}, \\
&\texttt{(DOLNEU):} && -I_t^D \leqslant \mathbf{P}^{[h_t]}(\mathbf{u}^{[h_t]} - \mathbf{v}^{[h_t]}) \leqslant I_t^D,
\end{aligned}$$

the *static* risk-type constraints are

$$
\begin{aligned}
\texttt{(TRARISK)}:\quad & \mathcal{R}_{\mathcal{T}}^{[h_t]}(\mathbf{x}^{[h_t]}) - \big[\sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})\big]^2 \\
& \quad - \big[\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\mu}^{[h_t]})\mathbf{x}^{[h_t]} - \mu_R^{[h_t]}\mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]}\big]^2 \\
& \quad - \big[\sigma_R^{[h_t]}\big(\mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\beta}^{[h_t]})\mathbf{x}^{[h_t]} - \mathbf{P}^{[h_t]}\mathbf{x}^{[h_t]}\big)\big]^2 = 0, \\
\texttt{(CATRISK)}:\quad & \mathbf{P}^{[h_t]}\mathbf{D}(\boldsymbol{\sigma}^{[h_t]})(\mathbf{u}^{[h_t]} + \mathbf{v}^{[h_t]}) \leqslant \Gamma_t,
\end{aligned}
$$

and the *dynamic* risk-type constraints are

$$
\begin{aligned}
\texttt{(QSDRISK)}:\quad & \mathcal{R}_{\mathcal{Q},\theta}^{[h_t]}(\mathbf{x}^{[h_t]}) - E_{\mathbf{r}^{[h_t]}}\big[d^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})\big]^2 = 0, \\
& d^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) - \mathbf{P}^{[h_t]}\mathbf{D}(\mathbf{r}^{[h_t]})\mathbf{x}^{[h_t]} - \mathcal{T}_\theta^{[h_t]}(\mathbf{x}^{[h_t]}) \geqslant 0, \\
& d^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant 0, \\
\texttt{(DDRISK)}:\quad & \mathcal{R}_{\mathcal{D},\pi}^{[h_t]}(\mathbf{x}^{[h_t]}) - E_{\mathbf{r}^{[h_t]}}\big[q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]})\big]^2 = 0, \\
& w^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) + q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) - (1-\pi) w_{\max}^{[h_t]} \geqslant 0, \\
& w_{\max}^{[h_t]} \geqslant w^{[h_\tau]}, \quad \forall \tau = 1, \ldots, t-1, \\
& q^{[h_t]}(\mathbf{r}^{[h_t]}, \mathbf{x}^{[h_t]}) \geqslant 0.
\end{aligned}
$$

For ease of exposition, the shortsale risk constraint block `(SHRISK)` represents only one security group that contains all N securities. Also observe that the dynamic risk constraint blocks `(QSDRISK)` and `(DDRISK)` require that these constraints are applied along each sample outcome $\mathbf{r}^{[h_t],l_t}$, for $l_t = 1, \ldots, L_t$, which thus increases the size of the above nodal rebalancing model significantly. If the quadratic risk aversion (QRA) coefficient $\lambda_{\mathcal{Q}}^t = 0$ or the drawdown risk aversion (DRA) coefficient $\lambda_{\mathcal{D}}^t = 0$, then the constraint block `(QSDRISK)` or `(DDRISK)` can be eliminated, respectively, from the model. However, it is neither implied nor asserted that absence of the constraint blocks `(QSDRISK)` and `(DDRISK)` allows one to collapse the scenario tree to a single (mean) scenario for the future. Quite to the contrary, our computational experience supports the view that even when the latter constraint blocks are dropped, applying the model with static risk control along the underlying scenario tree provides a rather effective approach for portfolio risk management, as opposed to using a single scenario *mean model*. The static risk metrics prove very effective in controlling the portfolio risk profile when applied within a scenario tree framework, rather than on a single mean scenario-based deterministic model.

For $t = 1$, to compute the value function at the root node $[h_1]$, we have set the convention that $\mathbf{x}^{[h_0]} \equiv \mathbf{x}^0$ is the initial position vector in risky securities, $C^{[h_0]} \equiv C^0$ is the initial cash position in the fund, and $w^{[h_0]} \equiv w^0 := \mathbf{P}^{[h_1]}\mathbf{x}^0 + C^0$ is the portfolio wealth prior to trading at the root node. By solving the root node rebalancing problem, one obtains the optimal objective value $\phi^{[h_1]}(\mathbf{x}^0, C^0, w^0)$ along with the optimal revised positions $\mathbf{x}^{*[h_1]} \equiv \mathbf{x}^{*[0]}$. These are the positions that the fund manager will establish in the portfolio when rebalancing at the beginning of period 1.

Solution of the model to obtain $\phi^{[h_1]}(\mathbf{x}^0, C^0, w^0)$ or $\mathbf{x}^{*[0]}$ can be quite complicated depending on the number of securities N, the size of the underlying scenario tree and the

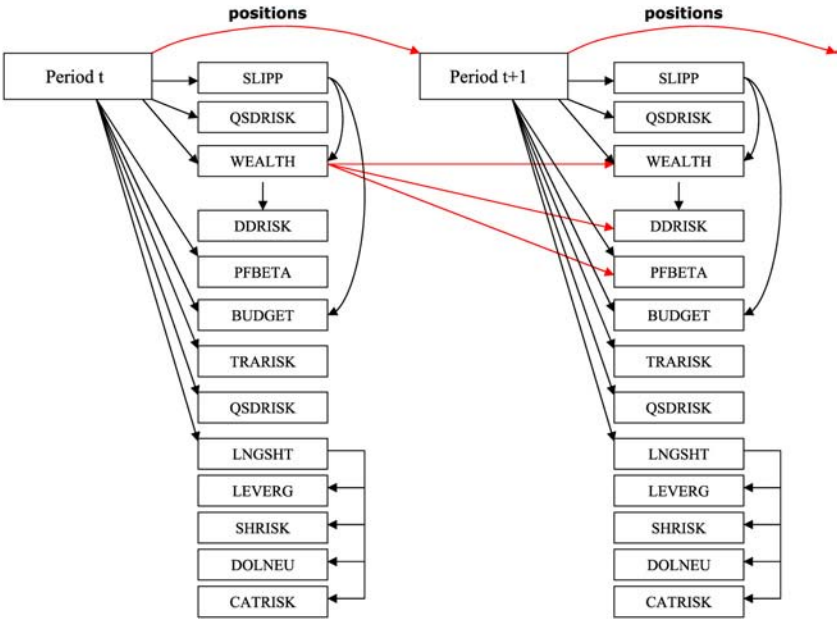


Fig. 3. Inter/intra-period linkages in the model.

number of time stages T, as well as, the particular choice of constraint blocks. Discussion on efficient solution methodology of the model is outside the scope of this chapter. Instead, the reader is referred to Edirisinghe and Patterson (2007) who develop very efficient multistage solution methodology by exploiting the block separable structure in the model, a technique termed the *Block Separable Decomposition* (BSD). Also, any discussion on implementational details of the model is avoided, except to the extent of discussing the numerical experimentation using actual stock data. The period-by-period linkages of the model components are depicted in Figure 3.

5.1. *Key issues in model rollover*

When the model in (43) is applied for portfolio rebalancing at some point in time, say $\tau = \tau_1$, then the investor must first choose a suitable number of trading model periods T as well as respective period lengths (say, a day) for each of the T periods. Then, the model is initiated with root node at time $\tau = \tau_1$. For ease of exposition, suppose the money manager is interested in (actual) rebalancing of the portfolio at pre-specified time epochs $\tau = \tau_i$ $(i = 1, 2, \ldots)$. To assure optimized rebalancing at each of these time epochs, the model in (43) is applied at each time τ_i, $i = 1, 2, \ldots,$ with its own multistage scenario tree. We will assume that each of these trees will have T periods, see Figure 4. Let the sequence of optimized rebalanced positions determined for each trading time τ_i, using the model in (43), is herein denoted by $\mathbf{x}^{*[0]}(\tau_i)$, $i = 1, 2, \ldots,$

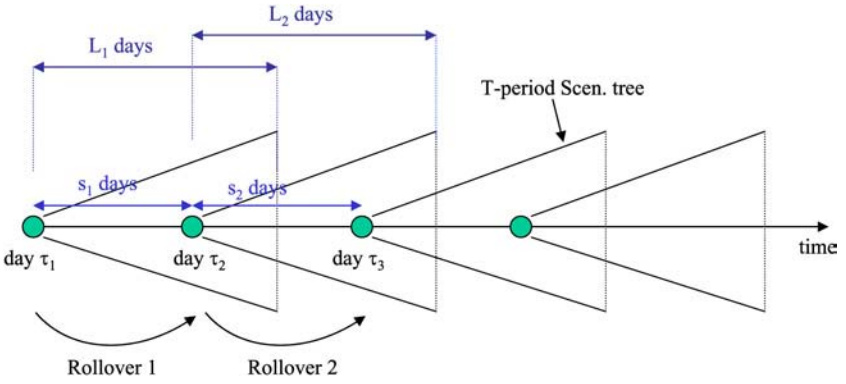


Fig. 4. Rolling horizon implementation of the model.

which are established in the market for the length of time $s_i := \tau_{i+1} - \tau_i$, referred to as the ith trading segment. Thus, it is only appropriate that the model in (43), applied at trading time τ_i, has specified a first period of length s_i. Period lengths are indirectly incorporated into the model via the required forecasted parameters such as the mean vector $\boldsymbol{\mu}^{[h_t]}$, variance–covariance matrix $\boldsymbol{\Sigma}^{[h_t]}$, security beta vector $\boldsymbol{\beta}^{[h_t]}$, and the market index parameters $\mu_R^{[h_t]}$ and $\sigma_R^{[h_t]}$. In addition, a scenario generator is required that is adapted to the period lengths in the model.

There are several pertinent issues that must be addressed in a successful implementation. The treatment here is cursory, as the focus in this chapter is model development and its performance analyses. First, while the first period length is set to s_i (at trading time τ_i), the remaining period lengths (of the $T-1$ periods) need not necessarily correspond to the actual trading epochs. One reason for this is that the computational solution of the model (43) is realistic only for a modest number of periods, for instance, T being not more than 3 or 4. But in order to capture the price dynamics in the future, periods of increasing length may be specified. For infinite horizon problems modeled as finite stage models, Grinold (1986) describes constructing an end effects period to capture the dynamics of the future after a finite number of model periods. For instance, see Cariño, Myers and Ziemba (1998) for an application in financial planning that incorporate end effects periods. In the implementation and the computational results reported in the next section, we pick the period lengths quite arbitrarily, and these lengths remain fixed at all subsequent model rollovers.

The second important issue is the estimation of parameters (of random vectors) for the chosen periods, conditional upon the observed history of actual price series. Suppose the actual price (vector) is denoted by $\widetilde{\mathbf{P}}_\tau$ at time τ. Thus, when forecasting the parameters for trading segment 1 at time $\tau = \tau_1$, one must use historical prices (only) from the *realized* history $\{\widetilde{\mathbf{P}}_\tau\colon\ \tau = 1, \ldots, \tau_1\}$, which will be applied at node $[h_1]$ of the model. Thereafter, for all forecasts at an arbitrary node $[h_t]$ of the model, only the actual observations $\{\widetilde{\mathbf{P}}_\tau \colon \tau = 1, \ldots, \tau_1\}$ must be used; however, they may be adapted to the particular (generated) scenario $[h_t]$ being followed in the scenario tree.

For the implementation in Section 6, second moment forecasts (variances and correlations) remain fixed at all nodes $[h_t]$ and set equal to those at the root node $[h_1]$. However, the first moment forecasts computed at the root node $[h_1]$ are progressively adapted to the scenario outcomes being followed up to a given node $[h_t]$. This adaptation procedure or the computation of the root node forecasts themselves are highly specialized and are omitted from further discussion in this chapter. The subject of parameter estimation in the context of financial market data has received considerable attention and there is a wealth of literature on various methods. The interested reader is referred to, for instance, Andreou and Ghysels (2002), Bai, Russell and Tiao (2001), Cohen et al. (1983), Foster and Nelson (1996), Ledoit and Wolf (2003), Merton (1980), and Scholes and Williams (1977).

Between two model rollovers, say from $\tau = \tau_1$ to $\tau = \tau_2$, the optimized portfolio must be evaluated based on the actual price series, herein termed portfolio (out-of-sample) *simulation* and it represents the actual performance of the portfolio under model-specified trade-sizes. Therefore, the actual (initial) portfolio wealth for a model-run beginning at the trading epoch at $\tau = \tau_2$, denoted by $w^0(\tau_2)$, is given by $w^0(\tau_2) = \widetilde{\mathbf{P}}_{\tau_2}\mathbf{x}^{*[0]}(\tau_1) + \kappa(s_1)C^{*[0]}(\tau_1)$, where $\kappa(s_1)$ is the discount rate adjusted to the trading segment length s_1. $w^0(\tau_1)$ is the wealth at the inception of the portfolio at the beginning of the first trading epoch $\tau = \tau_1$. Then, the portfolio value series, under the model-based continuous rollover rebalancing, is the wealth sequence $\{w^0(\tau)\} := w^0(\tau_1), w^0(\tau_2), \ldots$, which is evaluated for performance.

5.2. *Portfolio performance metrics*

The rebalancing model in (43) is evaluated by computing performance metrics for the (simulated) wealth series $\{w^0(\tau)\}$ of the managed portfolio. These performance metrics are

- ARoR (annualized rate of return): the portfolio daily average rate of return, net of trading costs, annualized over 250 days of trading.
- AStD (annualized standard deviation): the standard deviation of the daily portfolio net rate of return series, annualized over 250 days of trading.
- AShR (annualized Sharpe ratio): ARoR, less the annualized riskfree return rate, divided by AStD, see Sharpe (1994).
- DBeta (portfolio daily beta): computed according to (18), using optimal stock positions for the particular day.
- maxDD (portfolio maximum drawdown): for trading segment s_i is $[w^0_{\max}(s_i) - w^0(\tau_i)]^+$, scaled by $w^0_{\max}(s_i)$, where $w^0_{\max}(s_i) := \max\{w^0(\tau_k)\colon k < i\}$.
- ARTD (adjusted reward-to-drawdown ratio): the ARoR, less the riskfree rate, divided by the maxDD.

- DDDR (drawdown duration to drawdown recovery ratio): the ratio between the number of days during which the maxDD occurred and the time it takes to recover completely from the lowest portfolio value represented by the maxDD.

The last two drawdown metrics have not been considered in the literature, to the best of the author's knowledge. However, they provide key metrics of performance evaluation that are critical for successful money management.

6. Model application

SPDR Trust, which is an exchange-traded fund that holds all of the S&P 500 index stocks, is used as the market barometer in portfolio rebalancing. SPDR trades under the ticker symbol SPY. The model is applied for managing a portfolio of 95 stocks chosen from the Standard and Poors 100 stocks. These stocks are listed in Table 1. The full data set for the experimentation covers the period from January 02, 1996 to August 16, 2002. A period of 200 days of history, immediately preceding the beginning day (i.e., root node) of the multiperiod decision model, is used to estimate required parameters for the RoR random vectors. That is, the first trading segment begins on day 202, $\tau_1 = 202$, on October 16, 1996. Thus, when applying the model for the trading segment s_i, the historical data period used is from day $(\tau_i - 200)$ to day $(\tau_i - 1)$. The model is used for actual *daily rebalancing* of the portfolio, which thus results in 1450 rollovers of the model. That is, trade segment lengths are each a day long, $s_i = 1$, for $i = 1, \ldots, 1450$. The model is run in three versions, $T = 1, 2$, and 3. The model period lengths are first period length $= 1$ day, second period length $= 4$ days, and the third period length $= 5$ days. So, in a 3-stage model (i.e., $T = 3$), the total time horizon is 10 days into the future.

Techniques for generating scenarios for multistage stochastic programs have been investigated at great length, however, there is no one procedure that is best suited for varied applications. The reader interested in scenario generation methods for financial planning problems is referred to, for instance, Gulpinar, Rustem and Settergren (2004), Høyland and Wallace (2001), Kouwenberg (2001), Mulvey (1996), Mulvey, Morin and

Table 1
95 stocks used in portfolio rebalancing

AA	AEP	AES	AIG	AMGN	AOL	AVP	AXP	BA	BAC	BAX	BCC
BDK	BHI	BMY	BNI	C	CCU	CI	CL	CPB	CSC	CSCO	DAL
DD	DELL	DIS	DOW	EK	EMC	ETR	EXC	F	FDX	FTN	G
GD	GE	GM	HAL	HCA	HD	HET	HIG	HNZ	HON	IBM	INTC
IP	JNJ	JPM	KO	LEH	LTD	MAY	MCD	MEDI	MER	MMM	MIK
MOT	MRK	MSFT	MWD	NSC	NSM	NT	NXTL	ONE	ORCL	PEP	PFE
PG	PHA	PSFT	ROK	RSH	S	SLB	SLE	SO	T	TOY	TXN
TYC	UIS	UTX	VIA	VZ	WFC	WMB	WMT	WY	XOM	XRX	

Table 2
Market (SPY) performance metrics from October 16, 1996 to August 16, 2002

No. of days index is		Up: 703	Down: 659	Flat: 88
ARoR: 5.778%	AStD: 35.942%	AShR: 0.049	maxDD: 47.936%	RTD: 0.121
maxDD duration: 1067–1633 (566 days)			maxDD recovery: never	

Fig. 5. Market (SPY) performance from October 16, 1996 to August 16, 2002.

Pauling (1999), Mulvey, Rosenbaum and Shetty (1999), Pflug (2000), Yu, Ji and Wang (2003) and the many references therein. The generation method implemented with our rebalancing system is based on using the Mahalanobis distance-metric, applied on the historical return series and calibrated to market conditions, see Edirisinghe and Patterson (2003). Each scenario outcome at a given node $[h_t]$ is 95-dimensional. To generate a small enough sample that achieves the intended risk management properties, the scenario generation schemes utilized are highly specialized. The details of these schemes are outside the scope of this chapter.

For the experimentation period indicated above, the SPY index performance is used as a surrogate for the market performance, and its performance metrics are presented in Table 2, also see Figure 5 for its daily performance. Cash positions are applied an annualized money market rate of $\kappa = 4\%$, which thus serves as the riskfree rate. All experiments begin with an initial wealth of \$1 million available on $\tau = \tau_1 =$ day 202, and no further exogenous cash flow is available at any time. The initial positions in all risky securities at the beginning of segment s_1 (on October 15, 1996) are set to zero. All model runs allow both long and short trades with no shortsale constraints. Transactions cost (TC) parameters, see (4), are $\alpha_{0j} = 2\%$ and $\alpha_{1j} = 1$ for all stocks. The effects of increased cost rates will be illustrated later.

Table 3
Performance for one period model rollover with Pure SRC

	Pure Markowitz risk metric $[\sigma_P(.)]^2$					
ARoR	2.99%	4.08%	8.85%	15.37%	21.78%	48.99%
AStD	13.89%	20.15%	42.24%	59.95%	86.61%	154.81%
maxDD	28.01%	37.27%	54.32%	59.06%	67.64%	72.75%
ARTD	−0.036	0.002	0.089	0.193	0.263	0.619
	Pure Tracking risk metric $\mathcal{R}_{\mathcal{T}}(.)$					
ARoR	17.54%	22.94%	34.94%	60.93%	99.31%	248.78%
AStD	15.23%	19.26%	26.14%	40.53%	57.33%	102.19%
maxDD	19.17%	20.93%	26.27%	31.22%	35.04%	42.41%
ARTD	0.706	0.905	1.178	1.824	2.720	5.771
	Pure Cat risk metric $\mathcal{R}_{\mathcal{C}}(.)$					
ARoR	17.77%	40.94%	70.93%	104.25%	140.17%	215.44%
AStD	14.60%	28.70%	42.68%	56.14%	69.43%	95.26%
maxDD	9.89%	16.86%	20.88%	23.16%	26.53%	31.59%
ARTD	1.392	2.191	3.205	4.329	5.132	6.694

6.1. Single stage models

The first set of experiments is concerned with analysing the one period model version with varied risk metrics to highlight the relative effects of risk controls. With only SRC applied on one period models, scenarios of uncertainty are not required. These experiments compare SRC risk metrics: TRARISK $\mathcal{R}_{\mathcal{T}}^{[h_t]}(.)$ in (24), CATRISK $\mathcal{R}_{\mathcal{C}}^{[h_t]}(\mathbf{x}^{[h_t]})$ in (26), and the (variance) risk metric $[\sigma_P^{[h_t]}(\mathbf{x}^{[h_t]})]^2$, MARKOW. *Pure* SRC control means only one SRC metric is applied at a time. Table 3 provides a representative summary of portfolio performance for this case, where each column corresponds to a particular value of the objective risk aversion coefficient. See Figures 6 and 7 for a comparison of main portfolio performance measures across pure SRC metrics.

Pure (Markowitz) variance risk metric performs extremely poorly at all attempted coefficients of risk aversion, and none of the performance metrics appear to be within acceptable ranges for fund management. The strategy that achieves a Reward-to-Drawdown ratio of at least 2 with the least maxDD (of 16.86%) is the Pure CatRisk strategy. Even this level of drawdown may be unacceptable to a fund manager, even though it has ARTD > 2. By following a *mixed* strategy in which two SRC metrics are applied simultaneously, one may obtain improved portfolio performance. This is demonstrated next using tracking control and Cat risk control together. Tracking risk aversion coefficient is applied at five different levels: Very Low, Low, Moderate, High, and Very High. For each of these levels, Cat risk is varied to plot portfolio performance, as given in Figures 8 and 9. In each of these two plots, the upper envelope may be

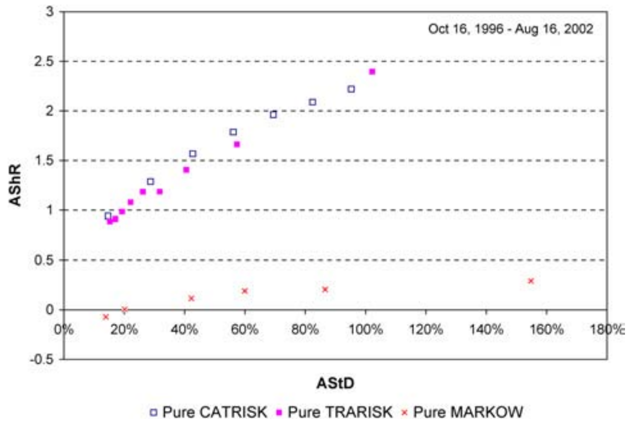


Fig. 6. Pure SRC risk (1 day horizon): Sharpe vs. Std Deviation.

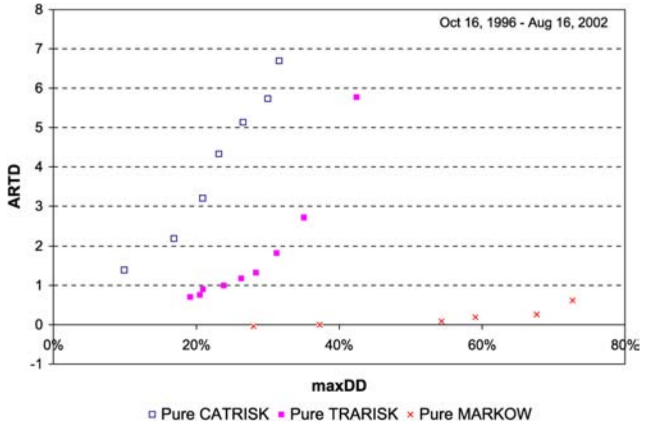


Fig. 7. Pure SRC risk (1 day horizon): Reward-to-Drawdown vs. max Drawdown.

thought of as the appropriate frontier in the two-dimensional risk metric space. Furthermore, Drawdown Duration to Drawdown Recovery time ratio, or DDDR ratio, for the mixed strategy (with *moderate* tracking aversion coefficient) remained between 0.33 and 0.55. A DDDR ratio of greater than 1 is desirable. DDDR for the pure (Markowitz) variance risk metric was zero because the portfolio *never* recovered from its maximum drawdown, for all ranges of variance risk aversion coefficients, during the invested period.

Under two-dimensional integrated SRC, the rebalancing strategy that achieves ARTD just over 2.0, with the lowest maxDD%, corresponds to a run of “Low” tracking risk aversion with CatRisk of \$10,000 a day. This run has Return-to-Drawdown ARTD =

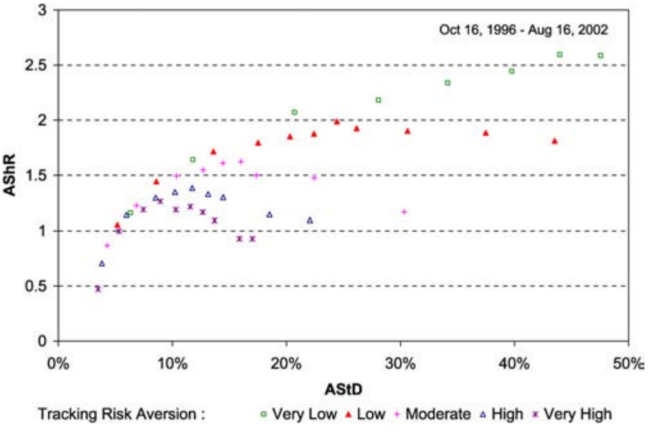


Fig. 8. Mixed SRC risk (1 day horizon): Sharpe vs. Std Deviation.

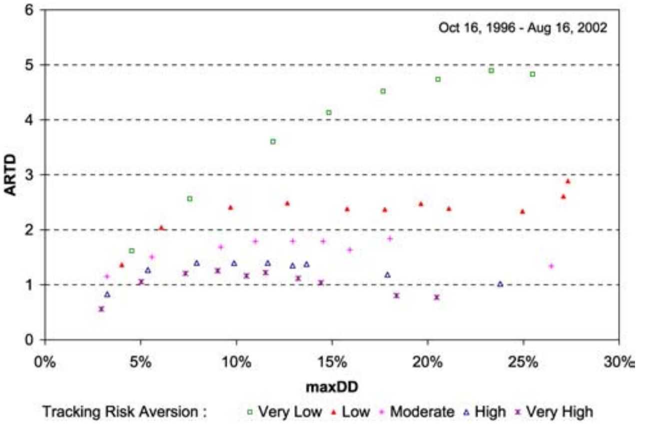


Fig. 9. Mixed SRC risk (1 day horizon): Reward-to-Drawdown vs. max Drawdown.

2.04 with maxDD = 6.07%, AShR = 1.44, and ARoR = 16.40%. The portfolio value over time for this run is depicted in Figure 10. In the same figure, two other graphs are plotted: the run that corresponds to the maximum ARTD of 4.89 (with a maxDD of 23.32%, AshR = 2.59, and ARoR = 118.14% for the pair "Very Low" tracking risk aversion and daily CatRisk = \$60,000) and the run under "moderate" tracking risk aversion and daily CatRisk = \$50,000 which yields the highest Sharpe ratio (AShR = 1.63, with maxDD = 14.53%, ARTD = 1.79 and ARoR = 30.01%). For these three cases, portfolio daily standard deviations are in Figure 11 and the daily portfolio betas are in Figure 12.

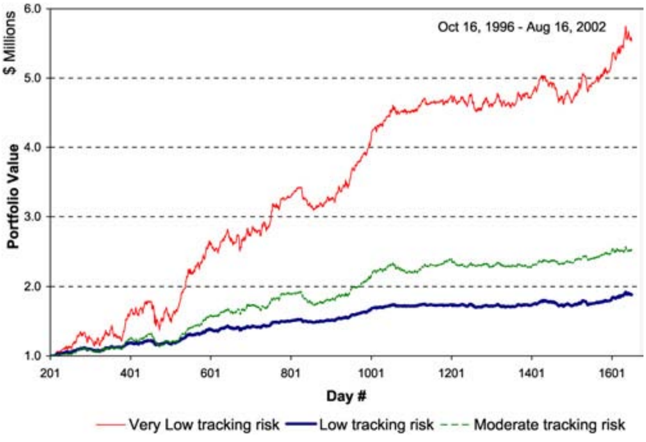


Fig. 10. Portfolio trajectories for selected mixed SRC strategies.

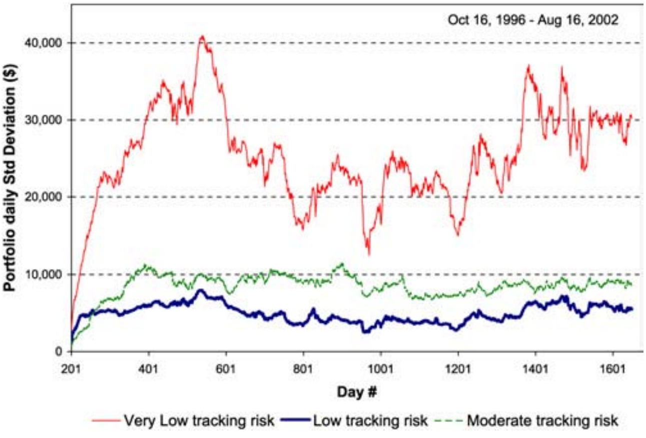


Fig. 11. Portfolio standard deviations for selected mixed SRC strategies.

6.2. Comparison with multistage models

Two-stage models are specified with a day-long first stage and a second stage of 4 days, for a future horizon of 5 days. Three stage models have a third stage of 5 days, for a 10 day-horizon. Multistage models, generally, are better able to trade-off transaction costs and control drawdown. Multistage models too are rolled over the 1450 trading days. Given that both Cat risk and Tracking risk in combination provided better risk control in single stage models, the multistage runs have this dual risk control settings. Portfolio variance as a risk metric (i.e., Markowitz metric) performs poorly even

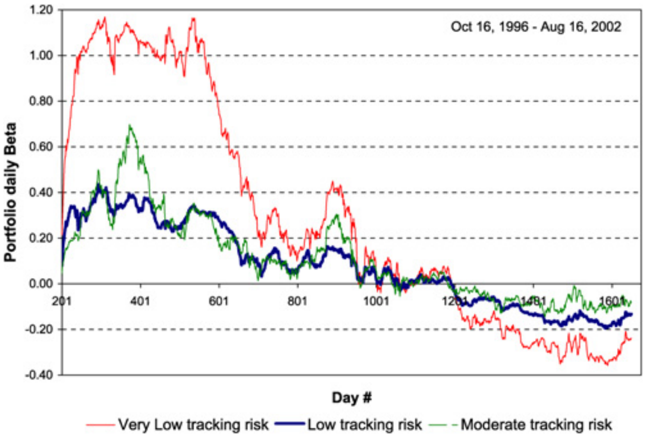


Fig. 12. Portfolio daily betas for selected mixed SRC strategies.

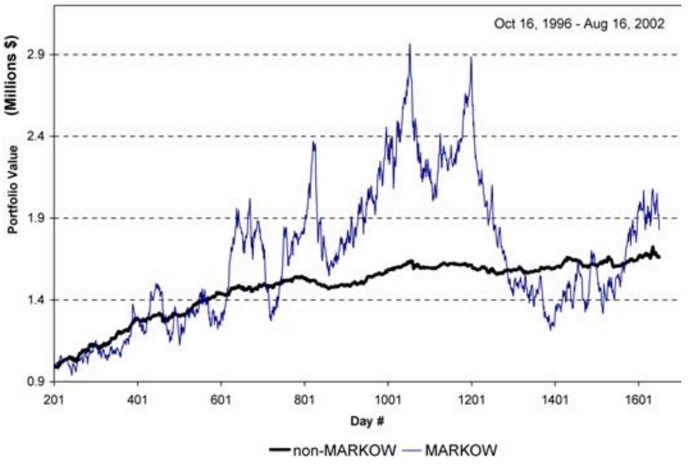


Fig. 13. Comparison of proposed risk metrics and variance risk metric—portfolio values.

in the case of 2-stage models. To illustrate, the 2-stage model is run with pure variance (MARKOW) risk metric, and it is compared with a non-MARKOW model with SRC/DRC metrics, but with the same value for objective (quadratic) risk aversion coefficients. The non-MARKOW model had CatRisk = \$40,000 and drawdown risk metric with $\pi = 10\%$. Portfolio trajectories and portfolio standard deviations are shown in Figures 13 and 14, respectively.

A comparison of risk-return trade-off between 1- and 2-stage models is presented next, where the effects of transactions/market impact costs are also pursued. The per-

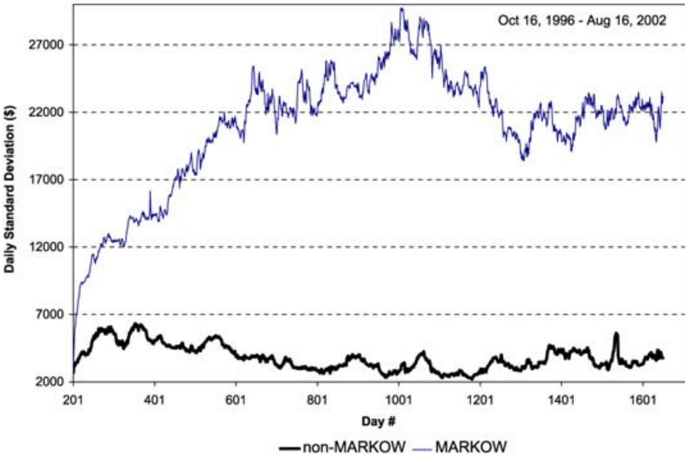


Fig. 14. Comparison of proposed risk metrics and variance risk metric—Standard deviations.

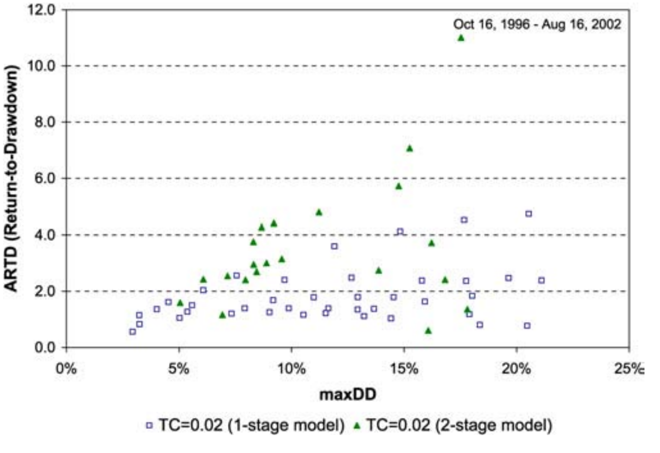


Fig. 15. Comparison of 1- and 2-stage models at trading cost $\alpha_{0j} = 2\%$.

share cost rate α_{0j}, set at 2% in the experiments so far, is increased to 5%. The two model versions have the exact same settings, except for the number of model periods, where the 2-period model is specified with $15 \times 15 = 225$ (95-dimensional) return scenarios for the 5 day period. This analysis reveals two important properties in rebalancing. First, at a fixed transactions cost rate, 2-stage models perform better than 1-stage models. Second, as the transactions costs increase, the advantages of multistage models over single period models become increasingly prominent. These effects are captured in Figures 15–18. As the last figure indicates, 2-stage model performance is quite robust in the event trading costs increase.

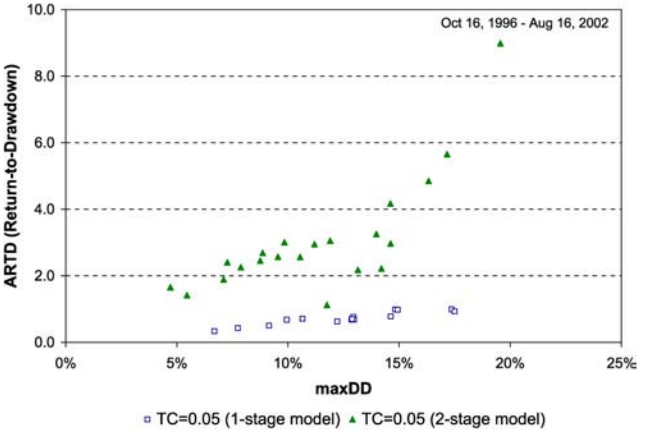


Fig. 16. Comparison of 1- and 2-stage models at trading cost $\alpha_{0j} = 5\%$.

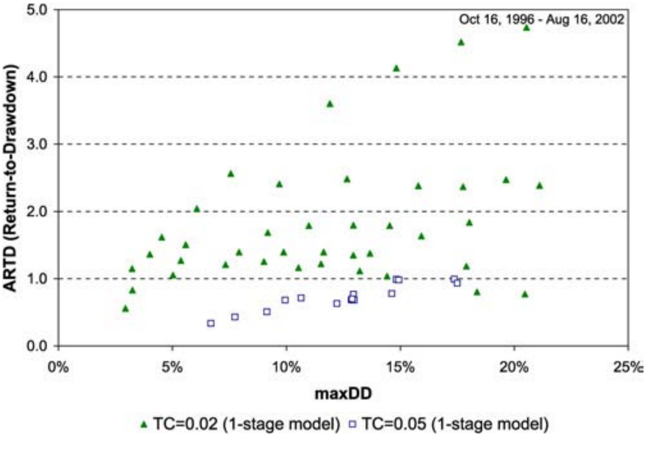


Fig. 17. Performance of 1-stage model with increasing trading costs.

Portfolio rebalancing using models with 3 periods yields significantly better performance with regard to drawdown characteristics. This is due to applying the drawdown risk (DDR) control, see (40), over 3 period-scenarios of the model. The drawdown duration to recovery ratios (DDDR) are compared across 1-, 2-, and 3-stage models with allowable relative drawdown set at 10% and specifying a *very high* drawdown risk aversion coefficient. These runs correspond to a "moderate" tracking risk aversion, for various levels of CatRisk control. As can be seen from Figure 19, these multistage models have rather impressive improvements over the single period counterpart.

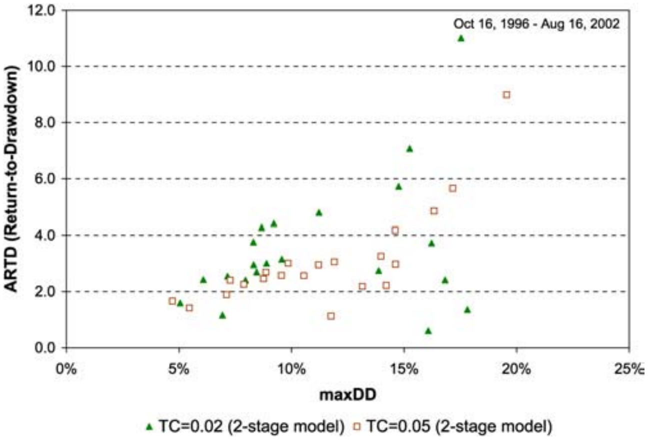


Fig. 18. Performance of 2-stage model with increasing trading costs.

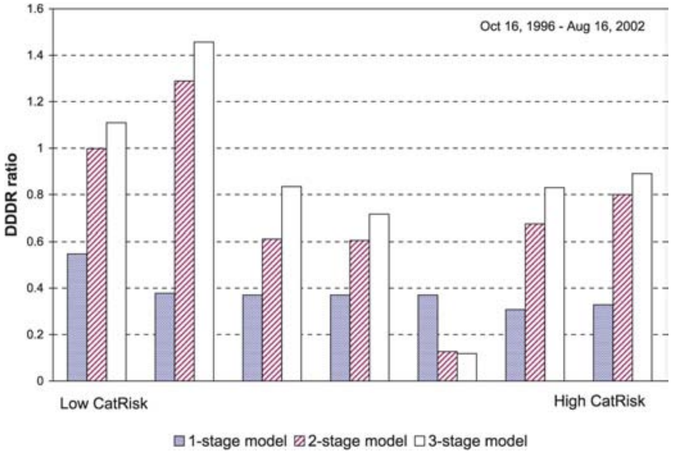


Fig. 19. Drawdown recovery comparison of 1-, 2-, and 3-stage models.

7. Concluding remarks

This article discussed an integrated risk management paradigm for frequent rebalancing of a portfolio of securities, using multistage stochastic programming modeling. By mathematically representing many practically useful risk metrics for fund management, the model is able to provide very effective rebalancing strategies. In particular, with impressive improvements in portfolio drawdown characteristics, such a stochastic programming based model can become an invaluable tool for money managers.

There are various other related issues that have escaped discussion in this article—for instance, market parameter estimations, methods for scenario projections to describe future uncertainty, as well as various other practical issues faced by money managers. In particular, trade execution costs are treated within simple slippage models in this paper, but the actual costs can be quite substantial and modeling those costs can be quite complicated. Also, the stochastic programming model considered here specifies trades as continuous variables, while in practice, trades must be integral, and they are typically lot-sized for execution. The latter requirement would lead to integer stochastic programming models and the solution of such models is generally tedious for real-time implementations of portfolio problems of practical size.

Finally, it was assumed that the universe of securities is pre-specified for the model to determine optimal trades. In a more successful implementation, such a universe must be determined and modified at regular intervals of time, based on considerations other than portfolio risk management. There are no standard techniques for such security selections, although practitioners use either fundamental analysis (of individual issues or market sectors/industries), tools from technical analysis (see, for example, Achelis (2000)), or a combination of both. Edirisinghe and Zhang (2004) describe a procedure based on fundamental analysis (of company income and balance sheets) in which optimization techniques are applied for identifying candidates for the universe of securities for portfolio risk management.

Acknowledgement

The assistance received from Maksym Bychkov and Ike Patterson in implementing the trade rebalancing system is thankfully acknowledged. Also, thanks are due to the SRGP program, College of Business, University of Tennessee, for the financial support received during the initial stages of the research in 2001.

References

Achelis, S.B., 2000. Technical Analysis from A to Z, second ed. McGraw-Hill.

Andreou, E., Ghysels, E., 2002. Rolling-sample volatility estimators: some new theoretical, simulation and empirical results. Journal of Business and Economic Statistics 20, 363–376.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Bai, X., Russell, J.R., Tiao, G.C., 2001. Beyond Merton's utopia (I): Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data, Working Paper. University of Chicago.

Birge, J.R., Edirisinghe, N.C.P., Ziemba, W.T. (Eds.), 2001. Research in Stochastic Programming. Annals of Operations Research, vol. 100. Baltzer Science Publishers.

Birge, J.R., Louveaux, F., 1997. Introduction to Stochastic Programming. Springer-Verlag, New York.

Bradley, S.P., Crane, D.B., 1972. A dynamic model for bond portfolio management. Management Science 19, 139–151.

Bychkov, M., Edirisinghe, N.C.P., 2004. Approximating scenarios for financial investment models using CVaR information, Working Paper. College of Business, University of Tennessee.

Cariño, D.R., Myers, D., Ziemba, W.T., 1998. Concepts, technical issues and uses of the Russell–Yasuda-Kasai financial planning model. Operations Research 46 (4), 450–462.

Cariño, D.R., Ziemba, W.T., 1998. Formulation of the Russell–Yasuda-Kasai financial planning model. Operations Research 46 (4), 433–449.

Cariño, D.R., Kent, T., Myers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., Ziemba, W.T., 1994. The Russell–Yasuda Kasai model: An asset/liability model for a Japanese insurance company using multistage stochastic programming. Interfaces 24 (1), 29–49.

Chekhlov, S., Uryasev, S., Zabarankin, M., 2003. Portfolio optimization with drawdown constraints. In: Scherer, B. (Ed.), Asset and Liability Management Tools. Risk Books, London, pp. 263–278.

Cohen, K.J., Hawawini, G.A., Maier, S.F., Schwartz, R.A., Whitcomb, D.K., 1983. Estimating and adjusting for the intervalling-effect bias in beta. Management Science 29, 135–148.

Consigli, G., Dempster, M.A.H., 1998. The CALM stochastic programming model for dynamic asset and liability management. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, pp. 464–500.

Consiglio, A., Cocco, F., Zenios, S.A., 2001. The value of integrative risk management for insurance products with guarantees. Journal of Risk Finance, 1–11.

Cvitanic, J., Karatzas, I., 1995. On portfolio optimization under drawdown constraints. IMA Lecture Notes in Mathematics and Applications 65, 77–88.

Delbaen, F., 2000. Coherent risk measures on general probability spaces, Working Paper. ETH, Zurich.

Dembo, R.S., King, A.J., 1992. Tracking models and the optimal regret distribution in asset allocation. Applied Stochastic Models and Data Analysis 8, 151–157.

Dert, C., 1995. Asset liability management for pension funds, PhD thesis. Erasmus University, Rotterdam, Netherlands.

Edirisinghe, N.C.P., 1999. Bound-based approximations in multistage stochastic programming: nonanticipativity aggregation. Annals of Operations Research 85, 103–127.

Edirisinghe, N.C.P., 2004. Multiperiod portfolio optimization with terminal liability: Bounds for the convex case, Computational Optimization and Applications.

Edirisinghe, N.C.P., Naik, V., Uppal, R., 1993. Optimal replication of options with transactions costs and trading restrictions. Journal of Financial and Quantitative Analysis 28, 117–138.

Edirisinghe, N.C.P., Patterson, E.I., 2003. Scenario generation for financial stochastic programs using Mahalanobis distance metric, Presentation at INFORMS, Atlanta, Oct. 2003 and Working Paper. College of Business, Univ. of Tennessee.

Edirisinghe, N.C.P., Patterson, E.I., 2007. Multiperiod stochastic portfolio optimization: Block-separable decomposition. Annals of Operations Research, in press.

Edirisinghe, N.C.P., Zhang, X., 2004. Stock selection for portfolio optimization via data envelopment analysis of financial statements: Empirical evidence, Working Paper. College of Business, Univ. of Tennessee.

Edirisinghe, N.C.P., Ziemba, W.T., 1992. Tight bounds for stochastic convex programs. Operations Research 40 (4), 660–677.

El-Hassan, N., Kofman, P., 2003. Tracking error and active portfolio management. Australian Journal of Management 28 (2), 183–208.

Follmer, H., Schied, A., 2002. Convex measures of risk and trading constraints. Finance and Stochastics 6, 429–447.

Foster, D.P., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139–174.

Frauendorfer, K., 1996. Barycentric scenario trees in convex multistage stochastic programming. Mathematical Programming 75 (2), 277–294.

Frino, A., Gallagher, D.R., 2001. Tracking S&P 500 Index funds. Journal of Portfolio Management 28, 44–55.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2003. The Innovest Austrian Pension Fund financial planning model InnoALM, Working Paper. Sauder School of Business, University of British Columbia.

Golub, B., Holmer, M., McKendall, R., Pohlman, L., Zenios, S.A., 1997. Stochastic programming models for money management. European Journal of Operational Research 85, 282–296.
Grinold, R.C., 1986. Infinite horizon stochastic programs. SIAM Journal of Control and Optimization 24, 1246–1260.
Grossman, S.J., Zhou, Z., 1993. Optimal Investment strategies for controlling drawdowns. Mathematical Finance 3 (3), 241–276.
Gulpinar, N., Rustem, B., Settergren, R., 2003. Multistage stochastic mean-variance portfolio analysis with transaction costs. In: Nagurney, A. (Ed.), Innovations in Financial and Economic Networks. Edward Elgar Publishing, pp. 46–66.
Gulpinar, N., Rustem, B., Settergren, R., 2004. Optimisation and simulation approaches to scenario tree generation. Journal of Economic Dynamics and Control 28, 1291–1315.
Holmer, M.R., Zenios, S.A., 1995. The productivity of financial intermediation and the technology of financial product management. Operations Research 43 (6), 970–982.
Høyland, K., 1998. Asset liability management for a life insurance company: A stochastic programming approach, PhD thesis. Norwegian University of Science and Technology, Trondheim, Norway.
Høyland, K., Wallace, S.W., 2001. Generating scenario trees for multi stage problem. Management Science 47 (2), 295–307.
Jacobs, B.I., Levy, K.N., 2004. Market Neutral Strategies. John Wiley and Sons Inc.
Jansen, R., van Dijk, R., 2002. Optimal benchmark tracking with small portfolios. Journal of Portfolio Management 29, 33–49.
Jarrow, R., 2002. Put option premiums and coherent risk measures. Mathematical Finance 12, 125–134.
Keim, D.B., Madhavan, A., 1996. The upstairs market for large-block transactions: analysis and measurement of price effects. Review of Financial Studies 9, 1–36.
Keim, D.B., Madhavan, A., 1997. Transaction costs and investment style: an interexchange analysis of institutional equity trades. Journal of Financial Economics 46, 265–292.
Keim, D.B., Madhavan, A., 1998. The cost of institutional equity trades. Financial Analysts Journal 54 (4), 50–69.
Konno, H., Kobayashi, K., 1997. An integrated stock-bond portfolio optimization model. Journal of Economic Dynamics and Control 21, 1427–1444.
Konno, H., Yamazaki, H., 1991. Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. Management Science 37 (5), 519–531.
Kouwenberg, R., 2001. Scenario generation and stochastic programming models for asset liability management. European Journal of Operational Research 134, 279–292.
Kusy, M.I., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34 (3), 356–376.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10 (5), 603–621.
Loeb, T.F., 1983. Trading cost: the critical link between investment information and results. Financial Analysts Journal 39, 39–43.
Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7, 77–91.
Markowitz, H.M., 1959. Portfolio Selection—Efficient Diversification of Investments. John Wiley, New York.
Markowitz, H.M., 1987. Mean-Variance Analysis in Portfolio Choice and Capital Markets. Blackwell, Oxford.
Merton, R.C., 1980. On estimating the expected return on the market: An exploratory investigation. Journal of Financial Economics 8, 323–361.
Mulvey, J.M., 1996. Generating scenarios for the Towers Perrin investment system. Interfaces 26 (2), 1–15.
Mulvey, J.M., Gould, G., Morgan, C., 2000. An asset and liability management system for Towers Perrin–Tillinghast. Interfaces 30 (1), 96–114.
Mulvey, J.M., Morin, F., Pauling, B., 1999. Calibration of stochastic scenario generators for DFA. Annals of Operations Research 85, 249–266.
Mulvey, J.M., Rosenbaum, D.P., Shetty, B., 1999. Parameter estimation in stochastic scenario generation systems. European Journal of Operations Research 118, 563–577.

Mulvey, J., Vladimirou, H., 1992. Stochastic network programming for financial planning under uncertainty. Management Science 18, 1642–1664.

Nicholas, J.G., 2000. Market-Neutral Investing: Long/Short Hedge Fund Strategies. Bloomberg Press.

Nielsen, S., Zenios, S.A., 1996. A stochastic model for funding single premium deferred annuities. Mathematical Programming 75, 177–200.

Ogryczak, W., Ruszczynski, A., 1999. From stochastic dominance to mean-risk models: Semi-deviations as risk measures. European Journal of Operational Research 116, 33–50.

Ogryczak, W., Ruszczynski, A., 2002. Dual stochastic dominance and related mean-risk models. SIAM Journal of Optimization 13 (1), 60–78.

Pflug, G.C., 2000. Scenario tree generation for multiperiod financial optimization by optimal discretization. Mathematical Programming 89, 251–271.

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. The Journal of Risk 2 (3), 21–41.

Rockafellar, R.T., Uryasev, S., Zabarankin, M., 2002. Deviation measures in risk analysis and optimization, Research Report 2002-7. Department of Industrial and Systems Engineering, University of Florida.

Roll, R., 1992. A mean-variance analysis of tracking error. Journal of Portfolio Management 18, 13–22.

Roorda, B., Engwerda, J., Schumacher, H., 2002. Coherent acceptability measures in multiperiod models, Working Paper. University of Twente.

Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. Journal of Financial Economics 5, 309–327.

Sharpe, W.F., 1994. The Sharpe ratio. Journal of Portfolio Management, 49–58.

Steinbach, M.C., 2001. Markowitz revisited: Mean-variance models in financial portfolio analysis. SIAM Review 43 (1), 31–85.

Torre, N.G., Ferrari, M.J., 1999. The market impact model. http://mim.barra.com/research/misc/market-impact-model.pdf. BARRA Research Insights.

van der Vlerk, M.H., 2003. Stochastic programming bibliography. World Wide Web, http://mally.eco.rug.nl/spbib.html.

von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press.

Warachka, M., Zhao, Y., Ziemba, W.T., 2004. Incorporating Diversification into Risk Management. Research Paper Series.

Whitmore, G.A., Findlay, M.C., 1978. Stochastic Dominance: An Approach to Decision Making under Risk. Heath, Lexington, MA.

Yu, L.Y., Ji, X.D., Wang, S.Y., 2003. Stochastic programming models in financial optimization: a survey. Advanced Modeling and Optimization 5 (1), 1–26.

Zenios, S.A., 1993. Financial Optimization. Cambridge University Press, Cambridge, England.

Zenios, S.A., 1995. Asset/liability management under uncertainty for fixed-income securities. Annals of Operations Research 59, 77–98.

Zhao, Y., Ziemba, W.T., 2001. A stochastic programming model using an endogenously determined worst case risk measure for dynamic asset allocation. Mathematical Programming 89, 293–309.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset, Liability, and Wealth Management. AIMR/Blackwell Publisher.

Ziemba, W.T., Mulvey, J., 1998. World Wide Asset and Liability Modeling. Cambridge University Press.

Ziemba, W.T., Vickson, R.G. (Eds.), 1975. Stochastic Optimization Models in Finance. Academic Press.

Chapter 17

ASSET–LIABILITY MANAGEMENT FOR INDIVIDUAL INVESTORS

GIORGIO CONSIGLI[1]

University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy

Contents

[1] Personal website: www.giorgioconsigli.it, DMSIA, University of Bergamo, giorgio.consigli@unibg.it, Research Associate of the CFR University of Cambridge, www.jims.cam.ac.uk, and Operational Research Systems, www.ors.it.

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02017-5

Abstract

This chapter considers theoretical and practical developments that are currently driving the remarkable growth of individual asset liability management (ALM) applications as part of the fund management industry worldwide. A personal financial planning is a decision problem faced by an individual whose aim is to manage his consumption and investment decisions to achieve a set of real or financial targets, given his current and expected income, over a long-term horizon. ALM has emerged as an ideal framework to address this type of decision problem under uncertainty, in which the achievement of a strategic objective is made conditional on the effective management of assets and liabilities over time.

From the mathematical and financial viewpoint, the three characterizing elements of this type of problem can be briefly summarized.

It is a constrained optimal decision problem. In implicit or explicit form, the individual seeks the maximization of a possibly multi-attribute objective function under a typically large set of institutional and specific constraints. The required inclusion of a risk preference parameter naturally leads to an expected utility problem.

The problem is dynamic. The achievement of intermediate targets as well as changing working conditions over time (active versus retirement, etc.), and the time distribution of liabilities and income changes (salary growth, etc.) induce by definition a dynamic decision problem. The sequence of actions taken in the face of uncertainty and their random consequences need to be taken into account within the given time frame.

The problem is stochastic. The effectiveness of any adopted strategy and the achievement of the individuals objectives depend on a sequence of random events, such as the evolution of the random processes modeling, for instance, the evolution of financial markets. The long-term nature of the decision problem, furthermore, imposes a specific requirement on the model of uncertainty, and specifically on the properties of the generated economic and financial scenarios.

Different mathematical methods are in theory able to accommodate those three features in part common to other classes of ALM problems.

The individual problem can be regarded as an extension of a personal investment-consumption model with a limited number of investment opportunities and a rich set of individual and regulatory constraints with a long-term objective. The peculiarity of the individual ALM problem comes from the extent and implications of a modeling approach, which in principle is expected to capture the different features of the management of a financial position with a typically long-term horizon, up to and sometimes beyond retirement for an investor whose preferences may very well change over the planning horizon. The stochastic programming approach to asset–liability management has thus emerged as an effective and appropriate way to address and analyze the personal financial planning problem. The generality of the individual problem from the financial point of view is considered here looking at several application areas, such as private banking, pension fund management, personal financial planning, and wealth management.

Keywords

asset–liability management, scenario generation, stochastic programming, individual financial planning, individual risk preference, optimal portfolios, scenario optimization

JEL classification: C54, C61, D14, D81, G11

1. Introduction

A personal financial planning (pfp) problem is a decision problem faced by an individual whose aim is to allocate his current wealth over a typically long-term horizon to achieve a series of real or financial objectives, given an expected consumption rate and a flow of life expenses related to his family composition. The area of asset and liability management (ALM), as it is clear from the different contributions in this volume (and others in the past, see, for instance, Mulvey and Ziemba, 1998), has emerged as the proper framework to solve this type of decision problems under uncertainty in which the achievement of a strategic objective does depend on the effective management of assets and liabilities over time at the individual level, with all financial and fiscal implications in principle taken into account.

Depending on the decision problem both the objective, the set of assets and liabilities and the constraints can differ significantly. Here we focus on the asset–liability problem relevant for a single investor, who, typically under an extended set of financial constraints, aims at achieving a number of intermediate and final goals. The problem typically depends on a number of random factors affecting real and financial returns. The investor's constraints depend on its activity, its age, the extension of its family, the fiscal regime in which he operates and other elements.

The individual planning problem is an extension of the classical optimal investment-consumption problem (Merton, 1969; Karatzas, Lehoczky and Shreve, 1987) characterized by a limited number of investment opportunities and a rich set of individual and regulatory constraints with an objective focusing on the individual targets. The peculiarity of individual AL management today and its current relevance in the financial industry, comes from the extent and implications of a modeling framework that in principle must be able to capture the complexity of a financial management problem over a long-term horizon, up to and sometimes beyond retirement (see other contributions in this volume, and Ziemba, 2003).

The solution of the described problem calls for the adoption of several theoretical assumptions and the development of modeling tools, often in an integrated fashion, which provide the core of this chapter.

Individual financial planning has been modeled over time considering a range of possible approaches. From the use of canonical assumptions on financial markets behavior and the solution at individual levels of static portfolio problems (from Markowitz (1952) onwards) to the inclusion in the decision space of financial and insurance contracts within dynamic optimization schemes (Berger and Mulvey, 1998) to continuous time type of models with stochastic control and a limited number of investment opportunities (Purcal, 2003). The generality of the individual asset–liability management problem puts this area of developments well inside the advanced research on stochastic optimization models and solution methods.

The life cycle theory provides a general theoretical framework in which the individual ALM problem is often addressed: over the years individuals tend to balance decreasing consumption rates with increasing saving rates depending on their life expectations. More recently, the central role of individuals psychology in long-term investment processes has been pointed out specifically with respect to investors irrational behaviors in wealth management applications (Kahneman and Riepe, 1998). Incorporating behavioral finance in individual ALM requires from a modeling viewpoint at least the introduction of a dependence of the investment strategy on the current wealth: we consider in the chapter the possible implications of a behavioral approach to the recommended course of actions as targets do approach and the individual wealth reaches high or low levels influencing individuals risk preference.

The interaction between elements of the theory of individuals' financial equilibrium and the challenge posed by the solution of a complex mathematical programming problem (Berger and Mulvey, 1998; Consigli and Dempster, 1998a; Dempster et al., 2002; Consiglio, Cocco and Zenios, 2004; Ziemba, 2003, others in this volume) is beyond the main focus of this chapter, in which we rather intend to analyse the state-of-the-art and possible ways forward in the modeling and computational solution of the described decision problem.

From a mathematical viewpoint, the three constituent elements of the ALM individual problem are:

- It is a constrained optimal decision problem: in implicit or explicit form, the individual seeks the maximization of a possibly multi-attribute objective function under typically a large set of institutional and specific constraints and with a given preference function. All fiscal and financial considerations must in theory be accurately considered. The inclusion of a risk preference parameter, furthermore, naturally leads to an expected utility and nonlinear optimization problem;
- The problem is dynamic: the achievement of intermediate targets as well as changing employment conditions over time (work versus retirement, etc.) and the time distribution of liabilities and income variations (salary growth, etc.) forces a dynamic representation of the problem. The sequence of actions taken in face of uncertainty and their random consequences need to be taken into account within a given time frame. The achievement of an intermediate target may affect individual preferences and result in a shift of its utility function that may have to be considered.
- The problem is stochastic: the effectiveness of any adopted strategy and the achievement of the targets do depend on a sequence of random events, such as the evolution of the random processes modeling a set of relevant financial markets. The long-term nature of the decision problem, furthermore, imposes a specific effort in the development of the model of uncertainty.

Different mathematical models are in theory able to accommodate those three features. They can be distinguished in general by focusing on the probability space assumed to capture the problem uncertainty: continuous time, discrete or continuous state processes (CTDS or CTCS) do generate continuous information flows in a given decision space and typically characterize stochastic control problems (Fleming and Stein, 2004; Purcal, 2003; Rudolf and Ziemba, 2004; Karatzas, Lehoczky and Shreve, 1987; Merton, 1969). The constraint region in this class of problems is often limited and the decision vector, to preserve the computational manageability of the problem, contains only few elements.

A discretization of the time space and more generally the problem uncertainty is instead peculiar of stochastic programming approaches, in which the accuracy of the problem description is maximal and the probability space approximates somehow (Chen et al., 1997; Dupačová, Consigli and Wallace, 2000; Pflug, 2001) an underlying typically continuous distribution, while decision revisions are allowed only at specific time points over the planning horizon. The reference model is in this case a dynamic stochastic programming (DSP) problem as the one considered in Consigli and Dempster (1998b). As such has been treated in one of the first software systems addressing the individual ALM problem (Berger and Mulvey, 1998. See also Mulvey et al. (2007) in this volume). More generally stochastic programs comprehend both static and dynamic problems, with uncertainty generated by multivariate continuous or discrete, parametric or historic probability distributions. A viable and practical alternative to stochastic programming is provided by policy optimization where, given a set of possible economic and financial scenarios, policy rules are directly associated with individual scenarios and the resulting simulated policy distributions can be directly evaluated (see Section 3.4 on this point).

This chapter is structured as follows. In the second section an overview of currently adopted techniques for the solution of individual financial planning problems is proposed with sections dedicated to the description of a set of relevant applications areas and the institutional setting. In section three the general structure of ALM models for individuals is presented with an emphasis on the key role played by the definition of the individual objectives and the adopted model of uncertainty.

In the fourth and fifth sections the key steps from the description of a specific individual investor problem, its mathematical representation, its generation consistently with mathematical programming standards, and the numerical methods available for its solution are discussed in detail. In Appendix A the individual ALM benchmark model is described in detail together with a specific software development for individual financial planning with a rich set of model outlines and graphical evidences.

The state-of-the-art and the steps to be followed in order to take full advantage within the industry of relevant scientific and theoretical results will inspire the content of the different sections.

2. Individual ALM in practice

The area of ALM applications has grown dramatically over the last decade with software developments that have benefited several sectors of the financial industry dedicated to wealth management (Dempster, Scott and Thompson, 2002; Pirbhai, Mitra and Kyriakis, 2003; Ziemba, 2003; Wallace and Ziemba, 2005). We have assisted to a widespread adoption of IT developments to support the solution of optimization problems arising in several segments of the banking and insurance activity. At a global level the relevance of the mutual fund industry and the relative increase within that industry of high net

Table 1
Aggregate data on the Mutual fund industry in Europe and Italy

		2000	2001	2002	2003	2004E
Europe 6						
Bln €	GDP	7198	7538	7703	7794	8096
Households	Financial assets	16 606	16 249	15 603	16 408	17 003
	FinAss % GDP	231%	216%	203%	211%	210%
	Mutual funds	1615	1546	1383	1511	1619
	Mutual funds % GDP	22%	21%	18%	19%	20%
	Mutual funds % FinAss	9.73%	9.52%	8.86%	9.21%	9.52%
Italy						
Bln €	GDP	1167	1218	1261	1302	1349
Households	Financial assets	2774	2629	2791	2914	3167
	FinAss % GDP	238%	216%	221%	224%	235%
	Mutual funds	462	397	337	360	348
	Mutual funds % GDP	39.57%	32.55%	26.74%	27.68%	25.77%
	Mutual funds % FinAss	16.66%	15.08%	12.08%	12.37%	10.98%

Sources: ECB, MF European association.

worth and private portfolios is a phenomenon observed over the last decade in both US and Europe (Consiglio, Cocco and Zenios, 2004). The latter define a predominant share of the demand for individual ALM services.

Table 1 provides a basic set of statistics to infer the macroeconomic relevance of the financial markets and the mutual fund industry in Europe (Europe 6 includes Italy, Germany, Spain, France, UK, and the Netherlands) and Italy from year 2000 to 2004.

At European level households financial portfolios amount to more than twice the GDP, roughly one fifth of which is managed by the mutual fund (MF) industry. In 2004 an amount of more than 1.6 trillion euros was estimated to be under MF management. A growing proportion of the investments come from wealthy and professionally skilled individuals, able to manage their portfolios relying on advanced techniques and open to the use of alternative, specifically Internet-based (see below) distribution channels. In 2003 and 2004 almost 8.8 million of Italian investors—16.5% of the total population, whose saving rate is historically above the European average—held roughly 25% of Italian total financial wealth with individual positions from 40 000 to several million euros. Around 7% of Italian investors—roughly 3.5 million people—manage portfolios worth more than 75 000 euros (sources: Bank of Italy, 2004; Borsa Italiana, 2004). Relatively smaller percentage ratios can be attributed to investors in the other five countries. Those are the macro numbers behind dedicated individual financial management developments in Europe: almost the size of the fund management industry as a whole.

The market growth helps to explain the effort of the MR industry to attract an increasing share of wealthy investors, and the role that can in this setting be played by ALM developments dedicated to individuals. As clarified in the dedicated Sections 2.1–2.5, a common methodological framework is indeed behind the following rather extended class of individual ALM problems:

- The canonical personal financial planning problem: an individual, given an estimated net income (almost all individual revenues and expenses are considered), seeks the achievement of one or more goals over an extended time horizon. Those goals typically take the form of the purchase of a holiday house, a chalet in the mountains, a boat, or any other goal depending on individual wishes.
- The private banking industry has emerged in the last few years as one of the driving areas of the international banking activity: the integrated financial management of high net worth individual portfolios is at the core of this sector of activity. The objective of the fund manager is here usually defined by the maximization of the investor expected wealth over time. The portfolio can in principle include any form of personal commitment: university fees, pension provisions, individual health care, etc. The amount of savings can itself enter the individual objective and thus lead to a subsequent personal financial planning problem.
- Due to population aging, advanced economies are increasingly facing a deterioration of public pension systems based on defined benefit schemes. As a consequence the weight of pension funds among institutional investors has increased over the years with a growing number of individuals depositing a regular part of their savings in the form of defined contribution pension plans. Here again the problem is conveniently addressed as an ALM problem in which the financial institution is required to manage a given amount of incoming flows so to maximize the terminal value of the cumulative plan. Occasional injections and subtractions, or reinvestment of extra returns can be part of the scheme, see in Section 2.3.
- Strategic asset allocation has originally taken the form of lump sum investments over long-term horizons managed by mutual fund representatives often with a minimum guaranteed return. In this case the problem can be seen as an extension of a canonical portfolio selection problem. Quite recently, due to computational advances and the failure of previously adopted portfolio management techniques, the inclusion of structured scenario generation techniques and the explicit modeling of individual risk preferences has put this area also well within the ALM framework described in the sequel.
- Finally, an area which is increasingly attracting the interest of the financial industry and individual investors has to do with the construction, within any of the above areas, of specific portfolio payoffs through engineering techniques so to reshape individual portfolios in a desirable way: here again asset and liability flows do enter the financial problem and are modeled as random processes and the objective of the optimal problem refers to the transformation of the current portfolio payoff profile into something different.

These five examples all imply a demand by the individual to a financial intermediary with a mandate to manage a given share of his income so to achieve one or more strategic objectives under a specific set of constraints (the family size, the average family annual income net of living expenses, the cost of medical and pension provisions, as well as university fees, education costs and, last but not least, fiscal costs and taxes), and an expected net income (generated by the working activity and enhanced by the management offered by the financial intermediary, also a function of the tax regime). These constraints, depending on the problem, can be implicitly or explicitly included in the suggested decision framework. The definition of a corresponding set of market instruments together with the activity of financial engineering put in place by the intermediaries does explain a relevant part of financial innovation regularly observed in the financial markets and the core of the current trends in the expanding industry of wealth management worldwide (Consiglio, Cocco and Zenios, 2004). The competition among intermediaries should make sure that under perfect market conditions, the individuals demand is met at the best possible financial conditions.

The methodologies available to address any of the above decision problems can be roughly classified in increasing order of accuracy depending on the solution output and the adopted problem formulation:

- An optimal buy and hold strategy over a long-term horizon coming from the solution of a static optimization problem with a limited number of asset classes, a given individual risk aversion function, certain assumptions on the financial scenarios and typically an underlying simulation tool to analyse and stress-test the suggested optimal policy.
- The same but with a more extended investment universe including real goods, pension and insurance instruments in addition to canonical fixed-income and equity instruments and a more accurate and structured model of economic and financial uncertainty, plus the above mentioned set of specific problem inputs.
- With yet a comprehensive universe of individual assets and liabilities, but now within a dynamic formulation with decisions including borrowing and investment in financial and insurance instruments whose possible evolution over time is captured through advanced modeling techniques, associated with a possible set of scenario-dependent intermediate and final targets.

This last degree of assistance has recently become standard in mature financial systems and at this point in time can be regarded as the most advanced currently available (see Berger and Mulvey, 1998; Dempster et al., 2002; Pirbhai, Mitra and Kyriakis, 2003; Ziemba, 2003). Advanced ALM modeling approaches (see the other contributions in this volume and Ziemba and Mulvey (Eds.), 1998) share a common framework: the mathematical structure and the relevance of the underlying probability space has been studied thoroughly over the last three decades. The solution approaches described in Section 5 have been developed and tested over the years as well.

What makes the formulation and solution of the individual investor ALM problem difficult is the complexity of the mathematical program that arises from the joint consideration of the above features and the required inclusion of all variables with an impact on the individual financial equilibrium.

The individual is expected to undertake over time a given investment strategy, while at the same time earn a certain salary and consume: the dynamics of the purchasing

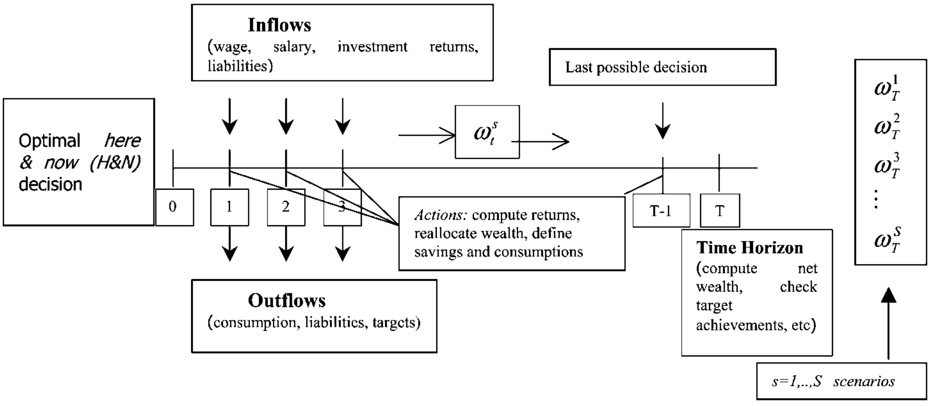


Fig. 1. Individual financial planning problem—time structure.

power of the individual, for instance, requires typically the inclusion of an inflation variable whose dynamic can severely condition the probability of reaching the different targets.

Figure 1 captures the main problem features.

Figure 1 from left to right makes explicit several relevant elements of this dynamic framework: at time 0, current time, a first decision, consistent with the individual long-term objectives, takes place, whose consequences are not known with certainty by the decision maker.

Then after the individual, in view of the observed results of previous decisions and current inflows and outflows—due to investment returns and borrowings, ongoing working activity and family requirements—can take other decisions, so to maximize the probability to achieve one or more targets. At the horizon, time T, no final actions are allowed, so, that the achievement of a target will in any case depend on a sequence of decisions all taken in the face of uncertainty—this implies that a condition of *non-anticipativity* must be satisfied by the decision sequence (Dempster, 1993; Dupačová, 1995; Birge and Louveaux, 1997; Consigli and Dempster, 1998a, 1998b; Mulvey and Ziemba, 1998).

The cash flows, as well as the market value of the targets will in general not be known in advance and are accounted for through a possibly high-dimensional random process $\omega \in (\Omega, \Xi, P)$, defined in an appropriate probability space. For $t = 0, 1, \ldots, T$, the sequence of decisions and realizations of the random process can be described in discrete time as

$$x_0, \omega_1, x_1(x_0, \omega_1), \omega_2, \ldots, x_s(x_0, \omega_1, x_1, \ldots, \omega_s), \omega_{s+1}, \ldots,$$
$$x_{T-1}(x_0, \omega_1, \ldots, x_{T-2}, \omega_{T-1}), \omega_T.$$

At time 0, current time, a first stage decision is taken whose effect is known at time 1, after the first realization ω_1 of the random process. At this point the sequence of deci-

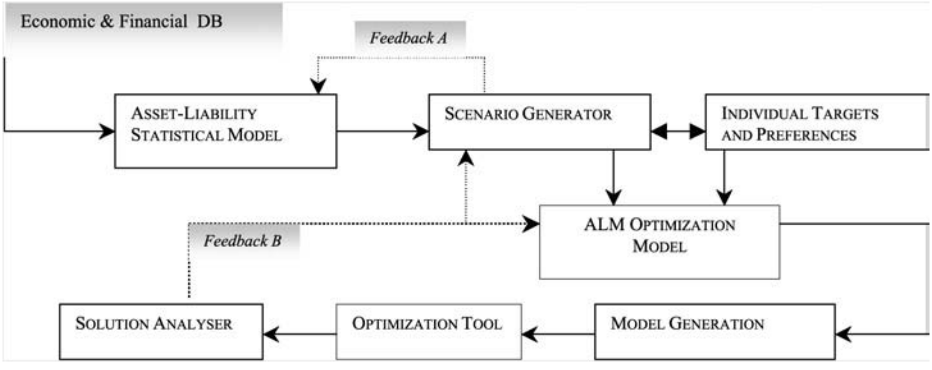


Fig. 2. Architecture for individual ALM problems.

sions from stage two onwards can be revised in order to take into account the revelation of uncertainty. Then a second realization of the random process is observed, leading to a strategy revision, and so on as we go further into the future the decision sequence is updated forward in time. At the end of the planning horizon, the last decision is contingent on the realization of ω_{T-1} and produces its effects at time T.

The individual endowment W_0 at time 0 can in general be distinguished in a disposable or *liquid* part and an *illiquid* part. What is relevant in general ALM problems is that share of wealth that is expected to enter the financial planning problem, the liquid component.

The evolution of wealth over time, denoted by W_t, $0 \leqslant t \leqslant T$, is central to the definition of the individual ALM problem in either cases of a discrete time—stochastic programming—or continuous time—stochastic control problems—model of uncertainty. The wealth is defined at any point in time as the financial net position—assets value minus liability value—and determines the feasibility of the adopted targets.

In summary, a dynamic individual ALM system is based on several modules/components and building blocks summarized in Figure 2.

The time structure, planning horizon, number of time periods and stages, with initial and terminal conditions, are necessary for the problem formulation as the asset and liability universe, the initial individual endowment, the targets and the individual aversion to risk. These concur to define a dynamic but not necessarily stochastic problem. The inclusion of the uncertainty implies the adoption of a statistical model for assets and liabilities, which in turn implies an underlying database of economic and financial data. The scenario generator has as input the statistical model coefficients and a model of evolution for the individual targets, if required.

Jointly the individual targets, the risk aversion estimate and the economic and financial scenarios for the problem define the individual ALM optimization model. The generation and solution of the scenario-dependent optimization problem finally require the adoption of a computationally efficient model generator and optimization tool. The analysis of the generated solution completes the above loop.

The two feedbacks reported in the figure deserve some comments. Feedback A takes into account the possibility, that once economic and financial scenarios have been generated from the given statistical model and historical data, due to inconsistencies with observed market dynamics, a refinement of the original model is required. The validation of the statistical model through out-of-sample back testing is becoming increasingly a required feature of modern decision systems.

Feedback B considers another highly desirable feature of decision support systems: once the solution of the decision problem is obtained, rather then implementing this solution immediately, it is advisable to test the output against a possibly more extended scenario input and/or a modification of the optimization model in order to take into account new evidences collected from such solution. This exercise can lead to a revision of the adopted statistical model or a more realistic set of individual targets and generate a more acceptable set of decisions.

Targets can be included in this framework either directly in the objective function or as liability streams in the constraints: in this case, according to model (3) below, targets are achieved almost surely or with probability one. Observe that in this case a borrowing decision can enforce the problem feasibility, otherwise not necessarily achieved. Consiglio, Cocco and Zenios (2004, 2007) in this spirit treat targets as liabilities and introduce a set of surplus and deficit variables to be then included in the objective function. Berger and Mulvey (1998) consider the targets as arguments of the objective function. This is preferable to allow model flexibility but typically leads to a less intuitive solution analysis.

In summary the mathematical machinery to address and solve the above types of ALM problems must include several specific features:

- An explicit model of individual risk attitude in the optimal decision problem represents a key ingredient of almost all advanced marketed decision tools;
- The objective function adopted to describe the individual optimal problems is expected to capture the complexity of the risk-reward relationship over a long-term planning horizon;
- The long-term nature of the decision problem also requires the adoption of a model able to generate likely evolutions over long-term horizons of the financial and economic scenarios upon which the effectiveness of the adopted decision plan will depend;
- The economic and financial scenarios must be explicitly taken into account in the process of determining an optimal individual policy—contingent on those scenarios—with respect to the predefined targets;
- Scenario generation as further explained below is strictly related to the set of decision variables and financial instruments included in the model: modern individual ALM models include not only classical investment classes such as mutual funds, equities and fixed income securities but also pension contracts, subject to defined contribution schemes, inflation linked instruments and so on. Targets are often scenario dependent themselves and require an associated model of evolution: one such example is the value of a house.

- Individual targets need to be accommodated within the optimization problem: either directly in the objective or as constraints, as further explained in the following sections.

Within this general framework a variety of different decision problems can be considered. The objective functions similarly can vary significantly.

The improvement of decision tools and a more accurate computational treatment of individual ALM problems is today conditioned not only by legal and fiscal uncertainties but also by the problems encountered by the financial industry to move from currently adopted techniques to methods fully capturing the dynamics of the underlying long-term problem.

An effective market interaction between the individuals' demands for asset–liability management and the service suppliers—financial institutions and fund managers' in the first place—is a condition for market efficiency. The issue is related to the accessibility of ALM products and technology by individual investors and the market network supplying those products. This market segment—which includes essentially personal financial planning as well as private banking applications—has developed over the last decade within a limited and flexible regulatory environment and has benefited enormously from the expansion of Internet and web-based technologies (Valente and Mitra, 2003). Both in the US and Europe we assisted to a progressive switch of individuals financial management from traditional channels such as banks dedicated portfolio services to Internet-based systems (Consiglio, Cocco and Zenios, 2004). In all Europe (before the enlargement to 25 countries) as at the end of 2004 around 64% of investors relied on direct services at banks branches and 13% on individual financial advisors (FERI Fund Management Information, 2005). In Italy roughly 13% of wealthy investors declare to rely on online management techniques (Borsa Italiana, 2004).

As at today web-based techniques are primarily benefiting large multinational institutions that supply ALM products and services relying on a central sophisticated ALM engine and a domestic and foreign distribution network.

The actors of this market are thus represented essentially by:

- The individuals with a protected and direct access to web-dedicated page and in principle no need to be at the financial service desk,
- The ALM units of financial institutions with their clients portfolios,
- The Head Quarter (HQ) of the financial institution that implements sophisticated risk and asset–liability management systems, developed by
- Software houses in charge of the decision support systems development to be installed at the financial institution directly or via a consulting firm acting on behalf of the financial institution, and
- Consulting firms can also directly develop and market pfp systems independently.

When the service is offered by a financial institution rather than an independent financial advisor, the optimal wealth allocation is controlled by the institution central unit and the financial service is tailored to enhance the marketing of the institution financial products (insurance, mutual funds and so on). The largest market share in individual ALM support is still concentrated in this form. The central unit system is tuned to pro-

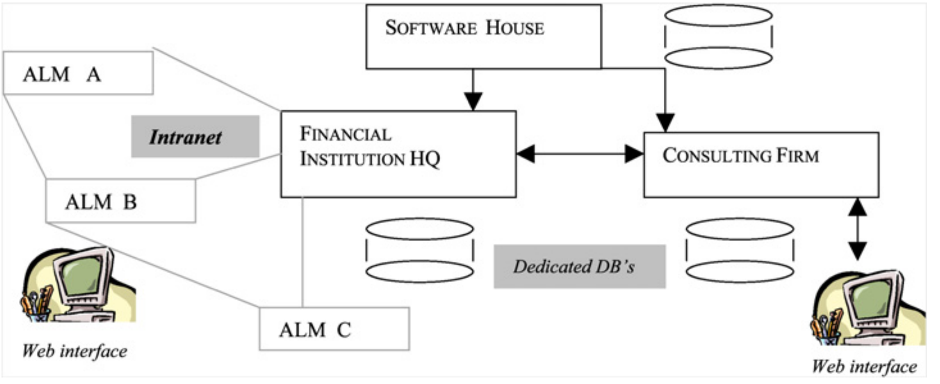


Fig. 3. Individual ALM standard market support.

vide an integrated service to different ALM units and matches individual portfolios with the institution market making activity.

Alternatively the service is directly developed and provided by consulting firms with own development capability and the support of a financial institution to actually implement the required investment strategies.

This framework is reproduced in Figure 3 (see also Section 5 on the individual ALM system implementation).

Under any of the reported applications of individual AL management, the fund manager does not put at stake any financial resource and the financial risk—market and credit—of his activity are entirely passed on to the individual investors.

In presence of a financial crash and inadequate portfolio management by the ALM service provider the loss will go directly to the investor, whose unique option will be to change the intermediary. As such Regulators have essentially imposed behavioral rules and safety conditions to the activity of the financial institutions. The only exception being represented by pension intermediaries expected to follow strict conservative rules.

Small software houses joining their effort with security houses of large financial institutions can to a certain extent reach via world-wide-web investors all over the world and provide an individual ALM service. In this case the system interface is inevitably a web application as a stand-alone Windows application. The competition between ALM service providers is primarily directed to expand the market share of managed high net worth individuals to be supported by private bankers and personal financial planners. The parallel competition between software houses implies the adoption of advanced optimization techniques and the progressive refinement of dedicated ALM developments. The following discussion describes in detail the methodologies, which are currently supporting the growth of this market segment.

We now go back to the mentioned case problems whose adaptability to the described framework we intend to qualify further, considering in practice what type of decision tools are currently available.

2.1. Personal financial planning

Individuals continuously analyse their budgets and take decisions with uncertain implications. Successful individuals generally make the right choice at the right time and take advantage of this compared with unsuccessful ones. Financial planning has to do with some forecasting of future incomes and liabilities and actions to be taken over time and while these incomes do occur. As clearly reported by Consiglio, Cocco and Zenios (2004), the generality of this problem has attracted interest for many years, but from a specific viewpoint: the literature on household portfolios is extended (Guiso, Haliassos and Japelli, 2002), but has been primarily dedicated to analyse household behaviors rather than addressing the key additional issue of how they should tackle their financial and wealth objectives. The key elements of this more comprehensive approach to personal financial planning can be identified rather easily, independently from the adopted underlying ALM technique.

Any high-net worth individual entering a personal adviser office in a large financial institution will be asked to clarify his income sources and current portfolio, his targets and degree of risk aversion and will be offered an investment portfolio whose expected return is sufficient to reach those targets.

What typically supports this exchange are: an investment model for the financial and real instruments representing the assets for the problem, an individual module for individual wealth description and liabilities and often an interview to assess risk attitude, and an optimization tool to find the optimal solution. This solution, when provided by a financial intermediary, will typically include only products marketed by that intermediary.

Liabilities are treated here as simple deterministic constraints to be considered when assessing the individual expected wealth. In several cases the future uncertainty on individual incomes and portfolio returns is captured by the decision tool by simply considering an expected average scenario—generated by some type of parametric or simulated scenario distributions—and then proposing as optimal decision plan the solution of the (deterministic) expected scenario problem. This type of solution does not consider the problem uncertainty in full (see, for instance, Ziemba and Mulvey (Eds.), 1998) and should be discarded by the financial planner.

Figure 4 shows that the structure in Figure 2 includes this as a simple sub case. In the last few years, thanks to theoretical and practical developments in the area of AL

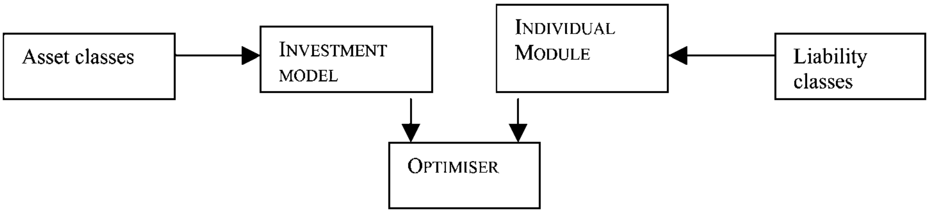


Fig. 4. Basic individual financial planning.

management, this simple framework is properly refined and included in efficient decision tools. A stream we expect to continue and that, as discussed in the rest of this chapter we expect to attract a growing part of software developments dedicates to asset management.

Several systems have been recently put forward in the market to take advantage of the investor's increasing interest towards personal financial planning tools. With a market impact, that not surprisingly has been rather independent from the adopted underlying decision framework. Recent examples come from the Home Account™ of Berger and Mulvey (1998), the FinancialEngines.Com of Sharpe & Associates (1998, updated 2003), with, in particular, the personal asset manager and online advisor modules, and more recently the Personal Financial Tools™ of Prometeia S.r.l. (Consiglio, Cocco and Zenios, 2004), the Personal Financial Planner™ of ORS S.r.l. (see Appendix B) and the two interesting tools of ORTEC International, with the individual risk analyser RAMI and the pension asset–liability manager PALM (2003).

Figure 1 describes in general terms the dynamics of an individual financial planning problem. The systems just mentioned do all accommodate the typical individual goal-oriented decision approach. They can differ, however, on the techniques which are adopted to solve the problem and the assumptions on the probability distribution of the underlying uncertainty. These will affect as explained in the following sections the quality of the suggested optimal strategy.

2.2. *Private banking*

Personal financial planning modules are increasingly incorporated in decision tools used in the banking industry to support the activity of the Private Banking (PB) units. This inclusion is essentially due, rather than to methodological aspects to marketing purposes: the individual financial planning service is directed to individuals with a sufficiently high individual income. The private banking sector does assist individuals with even higher income. The private banker mandate essentially includes the management of the individual wealth from several perspectives: fiscal and financial assistance, insurance protection, pension provision, targets achievement, if any, and in general a sustainable long-term wealth growth. The service is somehow at 360° degrees. As for financial planning problems, here again a key aspect of the PB service is represented by the identification of the individual risk aversion and furthermore implies the adoption of sound investment strategies that take into account the defined set of constraints, fiscal constraints in particular. Asset and liability classes again will require the adoption of a given model of uncertainty and the problem is conveniently formulated as a long-term stochastic programming problem.

The objective of the private banker is often described in the form of the minimization of some deviation from a target wealth with a risk aversion parameter that may lead—see below Section 4—to a trade off with the expected wealth maximization objective.

The common distinction in PB systems between managed portfolio—actively traded by the banker in accordance with the investor, fund portfolio—inheriting the performance of a mutual fund or fund of funds, insurance protection—which may include several insurance contracts—and pension provision—with possibly several life insurance instruments—introduces more complexity in the model formulation and the data input for the description of the uncertainty, but it is still appropriately captured in the stochastic optimization framework underlying the above ALM architecture. More recently PB management tools, even if implementing rather simple mathematical and decision models, have been interfaced or delivered through increasingly powerful browsers reaching individual mobile phones with real-time updated records of individual financial conditions. This stream is rather general but it is nowadays offered specifically for PB applications.

2.3. *Pension fund management*

Dempster et al. (2002) describe a dynamic financial analysis (DFA) system built by a large Fund Manager to support the activity of Pension Funds (PF) worldwide. The dependence of individual pensions from defined contribution type of schemes and the possibility for an individual to switch between different private organizations, on one side puts pension contracts in the domain of financial assets with specific contractual features and constraints (see the case study in Section 4), on the other requires the adoption of adequate decision tools by those Organizations, competing to offer a sufficient performance to the underwriters of the contracts. In this volume Mulvey et al. (2007), describe a similar framework applied to Multinational Insurance Companies. Geyer et al. (2002) describe an application to the Austrian pension fund system in InnoALM (see also Ziemba, 2007). In either case the individual problem coincides with the pension fund problem or the insurance company problem, when analyzed properly.

Consider an investor whose target is the achievement of a sufficient wealth at retirement and assume that this wealth comes from a public pension and a private pension contract with a defined contribution (DC) scheme (see on this the case study in Section 4 and the considerations developed therein). In order to reach his goal the individual needs to put a regular inflow in the Pension Fund and must cope the uncertainty affecting the value of the DC contract.

This uncertainty will depend on the adopted investment policy of the fund, whose interest is to achieve a sufficient return and increase the number of contributors. The pension fund in order to achieve a sufficient return on individual contracts will also face a complex liquidity constraint related to the presence of ongoing defined benefit schemes and uncertain mortality rates. It is clear that possibly modifying the problem structure and adopting a sufficiently general asset universe, this problem falls in the range of reported ALM problems as clarified by Dempster et al. (2002).

2.4. *Long-term asset allocation*

It will become very clear in Section 4, that long-term investment problems represent the core part of decision systems such as the one depicted in Figure 2. The history of asset management essentially starts with the booming of the mutual fund industry in the 1980s. For several years and still now a large part of the activity of fund managers is supported by the adoption of static portfolio tools implementing mean-variance type of techniques (Markowitz, 1952, also in this volume) or more advanced selection rules with less controversial assumption on asset returns distributions. The required extension of planning horizons and the demand for financial models able to capture long-term markets behavior has increased the effort to device asset allocation rules consistent with the definition of dynamic strategies and supported by capital market models (Mulvey and Thorlacius, 1998). In this respect long-term asset allocation systems can be straightaway accommodated in a structure as the one outlined in Figure 2, just eliminating the liability part.

2.5. *Individual financial engineering*

We finally recall an underestimated possible application of individual ALM models: the definition of optimal payoff structures for individual financial protection. ALM techniques have already been fruitfully applied to large Institutions Liability management and can easily be extended to personal planning. Consider, in particular, an individual investor with a given set of targets with random behavior: his optimal hedging strategy (against the excessive appreciation of target values) can be described as a replicating portfolio of a sequence of cap contracts. This can be valued within stochastic programming modules by minimizing the cost of the replicating portfolio and be accommodated in the above framework. This application stream has been investigated starting with Dahl, Meeraus and Zenios (1993) with the case of a collateral mortgage obligation structure. A comprehensive overview of pricing problems in incomplete markets within a stochastic programming framework can be found in King and Korf (2001) and Pennanen and King (2004).

3. ALM modeling for individual investors theory

The generic formulation of a stochastic optimization problem is

$$\min_{x \in X} Ef(x,\omega) \quad \text{s.t.} \quad g(x,\omega) \in Q \text{ a.s.} \tag{1}$$

In (1) $\omega := \{\omega_t\}_{t\in[0,T]}$ is a random process for asset and liability returns, $x := \{x_t\}_{t\in[0,T]}$ is a decision vector in the space of possible strategies $X \subseteq R^n$, while $f : X \times \Omega \to R$ and $g : X \times \Omega \to Q \subseteq R^m$ are the objective function and a convex constraint operator defined as mappings from the decision and sample spaces X, Ω. In the objective E denotes expectation. There is a finite horizon and the time set in full generality can be assumed either continuous or discrete, resulting in a finite number of stages. All constraints in (1) must be satisfied *almost surely*, e.g., with probability 1.

The objective function can take several forms: what, as reported above, is specific of individual ALM models is the inclusion of a utility function over the wealth process which is expressed as a function of the individual risk attitude. This can be estimated through direct methods (interviews, questionnaires) or indirect methods (implied by general behavior, age, etc.).

In Berger and Mulvey (1998) the objective is constructed as a utility over a *financial goal index* (FGI). In Consiglio, Cocco and Zenios (2002) as a linear utility over financial deficits and surpluses with respect to the targets and a risk aversion parameter introducing a choice between the surplus maximization problem and the deficit minimization problem. In Cariño et al. (1994), Cariño, Myers and Ziemba (1998) a penalty function is introduced for deviations from a target wealth, following Kusy and Ziemba (1986). In Consigli and Dempster (1998a, 1998b) and Villaverde (2003) a quadratic utility with respect to a final wealth is considered. Barro and Canestrelli (2005) show how to reduce any nonlinear convex program in the objective to a stochastic quadratic programming problem.

The optimal sequence of decisions does depend on the adopted profile of the individual utility function and is contingent on the adopted model of uncertainty: this is introduced in the problem through a generic probability space (Ω, Ξ, P). In ALM applications, particularly over long-term horizons, the sample space Ω is generated endogenously as an output of the economic and capital market model. The sigma-algebra Ξ includes all possible subsets of the sample space, which are P-measurable. The non-anticipativity constraint is enforced in the model by requiring the measurability of each decision with respect to the current sigma-field or information set Ξ_t: $x_t = \{x_t \mid \Xi_t\}$, $t \in [0, T]$. The probability measure P is also typically generated by ω, whose sample paths are given equal probability of occurrence. The probability P refers to actual scenario probabilities, or real-world probabilities, to be distinguished from the so-called risk neutral probability measure. Several recent stochastic programming applications in finance, and ALM applications, in particular (Pflug, 2004; Pennanen and King, 2004; Villaverde, 2003) are required to fulfill the classical no-arbitrage condition or risk-neutral condition of price dynamics in the market. This is regarded as a medium, long-term equilibrium condition that any financial market should satisfy. We consider in the sequel what are the implications of what we believe a generally overstated principle.

The solution of the optimal problem in (1) implies the adoption of efficient solution algorithms—see Section 5.3—for large scale linearly constrained convex programs. The program convexity is typically induced by the introduction of a risk-averse concave utility function over the terminal wealth.

The terminal wealth equation represents a key element of problem (1). It is denoted by $W_t(x^t, \omega^t)$, where the superscripts denote histories up to time t. The wealth equation is explicitly or implicitly included in almost all dynamic formulations of individual ALM problems: targets achievements are contingent on the currently available wealth.

Implemented ALM individual tools generally assume a set of targets all equally desirable and defined at the end of the time horizon. This simplification must be avoided in many real-world applications and we will consider in what follows a possible way around. As discussed in Section 4, the accurate inclusion of targets as final and intermediate objectives has a certain impact on the problem representation and the solvability, relying on currently available solution techniques. Targets are typically captured within stochastic optimization frameworks either as constraints to be satisfied almost surely or as scenario dependent objectives.

In the first case the problem feasibility requires the introduction of a free borrowing variable: given the market evolution any individual is in theory able to achieve its targets relying on the needed borrowing capacity.

In the second case the targets enter the objective function: the probability to achieve a target will then be given by the percentage number of scenarios under which the final optimal wealth exceeds the final target value, that can also be scenario dependent.

Problem (1) can be formulated and solved following several approaches. For dynamic models essentially two options are available: the stochastic programming and the stochastic control methods. Both techniques have been proposed to address and solve individual ALM problems (Purcal, 2003; Ziemba, 2003; MacLean, Zhao and Ziemba, 2005) and we analyse some of their specific properties in view of their popularity as possible methods to solve dynamic financial problems in general and individual ALM problems in particular.

Figure 1 in Section 2 describes the interaction between alternative scenarios and the decision problem: in presence of $s_t = 1, \ldots, S_t$, $t = 1, \ldots, T$, possible scenarios in each time period, the wealth process itself will evolve over time as a scenario-dependent process $W_t(\omega^{s_t})$. Here scenarios are assumed to evolve with discrete time increments and the number of scenarios is finite. In general financial and wealth scenarios are modeled as continuous time processes and a finite and countable state space is generated from an approximation of a continuous probability distribution. Stochastic control problems (Merton, 1969; Karatzas, Lehoczky and Shreve, 1987) try to solve the decision problem with an extremely accurate description of the probability space and a very fine time discretization at the cost of a less accurate description of the asset and liability model and the constraints of the decision problem.

Both stochastic programming and stochastic control frameworks can handle individual ALM problems with targets, but with a set of relevant differences as briefly outlined here next. The assumptions underlying each approach are quite different, making the dedicated methods alternative to each other. The key distinctive elements refer to:

- The planning horizon: typically discrete with a finite number of time stages for stochastic programming problems and based on a discrete approximation of a continuous time space for stochastic control models;
- The uncertainty: parametric and typically defined by a continuous probability measure in the stochastic control problem and discrete—resulting in event trees—for stochastic programming problems, with arbitrary assumptions on the probability measure;
- The problem dimensionality: the decision process can hardly include more than three–four variables and the associated stochastic model is built on a very limited number of underlying risk factors in stochastic programming formulations, while essentially only numerical tractability limits the accuracy of the problem description in stochastic programming models;

- The constraint region: again very concise and limited for stochastic control models and pretty extended and detailed in stochastic programming approaches.

As a result stochastic control models do consider representative skeleton problems and infer relying on statistical analysis actual policies. Or, do focus on specific sub-problems, whose solution can be accurately embedded in larger problem formulations.

Stochastic programming models, instead, provide an extremely accurate and comprehensive problem formulation and rely on numerical methods to identify an optimal solution to be implemented over time.

Several methodological steps in this case lead from the problem formulation to its generation and solution (see Sections 4 and 5).

3.1. *Mathematical properties of individual ALM stochastic programming problems*

Problem (1) can be described in more extended form by making explicit reference to the dependence of the decision strategy from the sequence of realizations of the random process $\omega \in (\Omega, \Xi, P)$. The triple (Ω, Ξ, P) defines a probability space characterized by a sample space $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_T$, an increasing sequence of *sigma-fields* $\Xi_1 := (0, \Omega) \subset \Xi_2 \subset \cdots \subset \Xi_T = \Xi$ and a probability measure P. In a discrete framework, leaving aside the problem of optimal approximation of a possibly continuous underlying source of uncertainty, it is possible to formulate the decision problem as

$$\begin{aligned} &\max_{x \in X} E_{\Xi_1}\big\{u_1(W_1, \lambda_1) + E_{\Xi_2|\Xi_1}\big[u_2(W_2, \lambda_2) + \cdots \\ &\qquad\qquad + E_{\Xi_T|\Xi_1,\ldots,\Xi_{T-1}}\big(u_T(W_T, \lambda_T)\big)\big]\big\} \\ &\text{s.t.} \quad B_t(\omega_t)x_{t-1} + A_t(\omega_t)x_t = b_t(\omega_t) \text{ a.s.}, \quad t = 1, 2, \ldots, T. \end{aligned} \tag{2}$$

In (2) $x \in X, x_t := \{x_t \mid \Xi_t, t = 0, \ldots, T\}, x_T \equiv 0$ is a decision vector defined in the appropriate decision space. The solution of problem (2) implies the maximization of a nested sequence of decision problems along the scenario structure—see Section 3.3, in which at each stage, an expected utility function is maximized.

From previous remarks, a key role is played by the risk attitude of the investor: this is captured in (2) by the risk aversion parameters λ_t, which are assumed to vary over the given planning horizon. In each time period the utility function defines a period utility that depends on λ_t: as shown below this is a possible approach to handle wealth-dependent strategies and accommodate alternative assumptions on individual risk preferences. Individuals are pretty flexible with regard to setting and achieving their targets.

From (2) at each stage the decision-maker faces a set of linear constraints described by the matrices and RHS vectors B_t, A_t, b_t that are also scenario dependent and define a sequence of conditions to be satisfied *almost surely* (a.s.). As customary for Markovian type of decision problems at each stage the last decision is the only one that matters in the current stage and the full set of constraints can be described recursively and originates over the horizon a block-diagonal constraint region, useful for numerical purposes.

The definition of the objective in (2) requires the additivity of the utility functions $u_t : \Re \times \Re \mapsto \Re$ of the wealth process, typically defined in the form of a sequence of risk-averse concave utility functions. The concavity of u and the convexity of the expectation operator and the constraints allow a two-stage decomposition of problem (2) resulting in a sequence of two-stage problems: at each stage the value function, defined over the remaining time to horizon is approximated by a sequence of supporting hyperplanes (see Consigli and Dempster, 1998a, 1998b).

The dimension of problem (2), thus its computational tractability does depend on several elements:

- The dimension of the decision vector whose elements are typically defined by asset and liability classes.
- The number of stages: as mentioned above personal financial planning problems can reach horizons up to several years. The horizon must be partitioned over time stages whose extension is function of the imposed constraints and the required modeling of cash inflows and outflows.
- The number of scenarios introduced in the problem to capture its core uncertainty: these can amount to several thousands and include economic and demographic scenarios as well as financial scenarios, whose evolution over time is to be modeled in order to assess the targets feasibility.

All these elements are handled in a stochastic programming formulation and make possible a detailed and robust mathematical description of the decision problem. In Sections 4 and 5 this approach is further considered in order to describe in more detail its implementation in real world problems and discussed the steps leading from its generation to its solution from an algorithmic viewpoint. Before that we want to consider the implications of a stochastic control approach, which is also being proposed to model and solve individual ALM problems.

3.2. *The objective function*

Given the initial wealth, the long-term personal net income, the predictable future costs, the personal plan takes the form of a dynamic portfolio strategy, which maximizes a given objective function under several financial and institutional constraints over a certain time horizon. It can be assumed that even if a target is typically associated with a time horizon, it is possible to satisfy a target with some tolerable delay, along a given number of scenarios. Furthermore, once a target is reached—say buy a country house—then the portfolio strategy must be reset in view of achieving a second target.

The decision maker risk attitude can be included introducing an Arrow–Pratt risk aversion measure over possible deviations of the current wealth W_t from a target wealth $\widetilde{W}_t$. The model extends to the case of targets distributed over time as a discrete control problem, in which the utility functions must be additive and, consistently with the adopted tree representation, a nested sequence of optimal problems need to be solved. Following (2), we allow in the objective function a sequence of period utilities that may depend on a risk aversion function λ_t:

$$\begin{aligned}
&\max_{x \in X} E \sum_{t=1,2,\ldots,T} \left[u_t(W^t, \widetilde{W}^t, \lambda_t)\right] \\
&\text{s.t.} \quad \text{for } t = 1, 2, \ldots, T, \\
&\qquad W_t = w(x^t, \omega^t), \\
&\qquad \widetilde{W}_t = w(\omega^t), \\
&\qquad A_t(\omega^t)x_{t-1} + B_t(\omega^t)x_t = b_t(\omega^t) \quad \text{a.s.}
\end{aligned} \tag{3}$$

In (3) $x \in X, x := \{x_t\}_{t=0}^{T-1}$ takes the form of a rebalancing portfolio strategy. The disposable and the target wealth—needed to achieve the desired goals—are both modeled in (3) by a pair of random processes, whose evolution over time is determined by a set of underlying financial and real returns. Targets can in general be accommodated in either the objective functions or the constraints: in the latter case the constraint will have to be satisfied with probability 1—*almost surely* (a.s.)—influencing directly the feasibility region of the problem. The estimation of the success probability is in this case possible but difficult to derive (Consiglio, Cocco and Zenios, 2004). In the former we can directly assess the success probability and identify the scenarios along which the likelihood of reaching the targets decreases.

Popular formulations of the utility functions in (2) and (3) can be derived by either maintaining a canonical risk-reward trade-off framework or by explicitly including a concave period utility resulting in a convex programming problem. In the first case the problem will result in a stochastic linear program (SLP) or a stochastic quadratic program (SQP). In the second we will remain in the general domain of stochastic nonlinear programming (SNLP).

As clarified in Section 3.2.1, we have in general:

- For stochastic linear programming problems:
  - The minimization of the mean absolute deviation from the target values.
  - The minimization of the mean absolute semi deviation from the targets.
  - The maximization of the expected excess wealth with a linear penalty.
  - The minimization of the expected shortfall with respect to the targets.
- For stochastic quadratic programming problems:
  - The minimization of the variance around the targets.
  - The minimization of the semi-variance with respect to a target.
  - The optimization of a trade-off function with a linear growth component and a quadratic downside component.
  - The minimization of a quadratic function of a risk measure in general.
- For general stochastic concave nonlinear programming problems:
  - The maximization of the excess utility over the target values.
  - The minimization of a nonlinear shortfall function of the targets.

Under the different formulations targets are typically assumed to be independent.

The mathematical program above can include different specifications of the utility function as further explained in Section 3.2.1.

The inclusion of targets, such as purchasing a house may require the introduction of integer variables leading to the formulation of mixed-integer stochastic programming (MISP) problems. Integer variables complicate the solution of the problem inducing nonconvex discontinuous objective functions, for which approximation methods are needed. In presence of expected utility problems the associated stochastic programs become easily unmanageable. For this reason we prefer not to formulate the individual ALM problem as a MISP.

The solution of a stochastic dynamic problem such as (3) is intended precisely to identify the sequence of decisions that optimize the given objective in face of the uncertainty actually introduced in the model. As the model of uncertainty, here embodied by the random behavior of market returns and other factors affecting both the objective and the constraints in (3), changes so will the optimal strategy for given targets and risk attitude.

The steps from the formulation of problem (3) to its solution imply the exact specification of the objective function, the definition of the scenario model and the generation of the standard input, in the form of the deterministic equivalent of problem (3), for its numerical solution.

3.2.1. Objective function and targets

Different objective functions can be accommodated in (3) and will result in alternative problems formulations and solution approaches. The risk-reward trade-off framework provides a good general reference for the specification of linear and quadratic programming problems that as mentioned above, can be considered special cases of problem (3). We have in this case, denoting with R and ρ a reward and risk function, respectively, and a trade off based on the coefficients $(\lambda_t, \gamma_t) \geqslant 0$, the problem

$$\begin{aligned}
&\max_{x \in X} E \sum_t \big[\lambda_t R(W_t) - \gamma_t \rho(W_t, \widetilde{W}_t)\big] \\
&\text{s.t.} \quad \text{for } t = 1, 2, \ldots, T, \\
&\qquad A_t(\omega^t)x_{t-1} + B_t(\omega^t)x_t = b_t(\omega^t) \quad \text{a.s.}
\end{aligned} \tag{4}$$

Several formulations fall in this class. Linear programming problems do generally imply the minimization of a deviation measure—adopted as a risk measure—with respect to the defined goals.

For $\lambda_t = 0$, $\gamma_t = 1$ we would have $\min_{x \in X} E \sum_t |W_t - \widetilde{W}_t|$, which extends to a sequence of wealth differences the *Mean-Absolute-Deviation* (MAD) measure of Konno and Yamazaki (1991), see also Zenios (1993, 1995). From the MAD model, by constraining only the downside, a second type of shortfall model, the *Mean Semi-Deviation* (MSD) can be generated: $\min_{x \in X} E \sum_t [W_t - \widetilde{W}_t \wedge 0]$, considered in Ogryczak and Ruszczynski (1999).

In the MSD problem, the decision maker seeks the minimization of the negative deviations from the target wealth $\widetilde{W}$, without penalization of the upside. In the case of a discrete number of scenarios for the random variable W, the MSD problem is a convex piecewise linear function of the scenario realizations, and is also LP computable. Consiglio, Cocco and Zenios (2004) consider an objective function in the form of (4) with $\lambda_t \equiv 1$, $\gamma_t = \gamma \geqslant 0$ and a reward function given by the expected terminal surplus and a risk function defined as expected terminal deficit against the targets. This also leads to a SLP formulation that can be easily solved relying on current techniques.

With random target values, it is easy to show that problem (4) also includes the class of tracking error models (Zenios, 1993) common in fund management applications.

Quadratic programming problems do follow under the formulation (4) by considering a quadratic risk measure such as the variance or the semi-variance around the targets. An interesting set of problem specifications come up in this case and have recently been proposed for the solution of dynamic ALM problems (Dempster et al., 2002). With specific reference to a long-term financial planning problem, indicating with $\widetilde{W}_t$, $t = 1, \ldots, T$, the wealth necessary to achieve the defined targets, a pure quadratic programming problem will result for $\lambda_t \equiv 0$, $\gamma_t = \gamma = 1$, $\rho(W, \widetilde{W}) := (W_T - \widetilde{W}_T)^2$. A mean semi variance model is again derived for the case $\lambda_t \equiv 0$, $\gamma_t \equiv 1$, $\rho(W, \widetilde{W}) := (W_T - \widetilde{W}_T \wedge 0)^2$.

Villaverde (2003) considers the downside quadratic objective in a global fund management problem $(\lambda_t, \gamma_t) \geqslant 0$, $R(W_t) := W_t$, $\rho(W_t, \widetilde{W}_t) := (W_t - \widetilde{W}_t \wedge 0)^2$. Barro and Canestrelli (2004) propose for the solution of a multistage asset allocation problem with an explicit model of individual risk attitude, a rather general objective function with a quadratic risk function plus a linear penalty, to limit portfolio turnover $\min_{x \in X} E \sum_t [\gamma_t (W_t - \widetilde{W}_t)^2 + \lambda(1' x_t - 1' x_{t-1})]$. The adoption of a penalty function on deviations from the necessary wealth dynamics is very much in the spirit of stochastic control problems, and can be accommodated in convex programming problems (see also Kusy and Ziemba, 1986; Cariño et al., 1994; Ziemba, 2003) of the type proposed in the generic individual ALM model instance in Section 4:

$$
\begin{aligned}
&\max_{x \in X} E\left[\frac{1}{\lambda}(W_T)^{\lambda} - \gamma \sum_t \left|W_t - \widetilde{W}_t\right|\right] \\
&\text{s.t.} \quad \text{for } t = 1, 2, \ldots, T, \\
&\qquad A_t(\omega^t) x_{t-1} + B_t(\omega^t) x_t = b_t(\omega^t) \quad \text{a.s.}
\end{aligned} \tag{5}
$$

The following elements characterize problem (5):

- A convex objective based on a concave power utility function of terminal wealth and a penalty function for absolute period deviations from the target wealth process.
- A constant Arrow–Pratt relative risk aversion (CRRA) measure and penalty coefficient: the utility is thus of iso-elastic type within the CRRA class.
- A set of scenario dependent constraints: both sides of the constraints are random. At each stage borrowing and investment decisions must satisfy a set of constraints and will maximize the expected utility of the period excess wealth, while penalizing any deviation from the required wealth path, that thus needs to be computed explicitly at each stage.

In presence of additive utility, problem (5) can be conveniently decomposed in a sequence of nested optimization problems as in formulation (2). Ignoring the penalty term an extension of (5) would be given by the maximization of the expected utility of terminal wealth, conditional on this being higher than the target wealth. In the terminology of utility theory this is equivalent to allow a positive utility only in presence of excess with respect to the wealth required for goal satisfaction. Penalty functions can be conveniently imposed in the form of symmetric penalties on MAD models or as asymmetric functions on general tracking models (see recently Jobst and Zenios (2002) on credit portfolios; Ziemba, 2003).

3.2.2. Risk attitude: taking preferences into account

Problems (3) and then (4) and (5) describe different approaches to the treatment of individual's risk attitude with respect to target values. In general down side linear and quadratic objective functions do not discriminate between individuals eagerness towards those targets: whatever the risk, that's welcome if the targets are reached. A possible inclusion of individual preferences can be achieved in the stochastic quadratic programming problem, along the lines indicated by Villaverde (2003), with an objective function explicitly defined in terms of period utilities $u_t(W_t)$, $t = 1, 2, \ldots, T$, with $\lambda_t, \gamma_t, \widetilde{W}_t \geqslant 0$

$$\max_{x \in X} E \sum_t \left\{ \lambda_t W_t - \gamma_t \left[W_t - \widetilde{W}_t \wedge 0 \right]^2 \right\}. \tag{6}$$

At each stage it is possible in (6) to refine the coefficients λ_t and γ_t according to a predefined updating rule. This is in the spirit of a commonly adopted approach to implicitly taking into account individual preferences via a weighting scheme between a linear and a quadratic objective in the very spirit of the original Markowitz mean-variance model. Consider, in particular, the case in which $\lambda \in [0, 1]$ and $\gamma = 1 - \lambda$, with a terminal wealth objective of the linear-quadratic type:

$$\max_{x \in X} E \left\{ \lambda W_T - (1 - \lambda) \left[W_T - \widetilde{W}_T \right]^2 \right\}. \tag{7}$$

The risk-reward trade off is captured in (7), allowing λ varying from 0 to 1 from pure stochastic quadratic programming or variance problem to pure stochastic linear programming or growth problem, respectively. The switch between growth and variance can be calibrated over time in (6) in order to limit risky strategies near target horizons. In either cases, the trade off coefficients are user-defined and do not imply any estimation of individual risk aversion as customary in utility theory and described in Section 4.3.

On the contrary expected utility models do imply an explicit estimation of the individual relative and absolute risk aversion. This is represented by the absolute and relative risk aversion measures: $-u''(W_T)/u'(W_T)$ and $(-u''(W_T)/u'(W_T))W_T$, respectively. $u : \Re \rightarrow \Re$ is a concave risk-averse utility function of the terminal wealth such as the power utility in (5). It can be assumed in this case that a positive utility is associated with a terminal wealth exceeding the target wealth. Dempster et al. (2002) discuss the implications of three alternative utility functions—the negative exponential, the power and the mean-downside variance—in a global asset and liability management problem. In presence of a concave utility, the issue of correctly specifying the curvature of the utility function and his positive slope are central to the model construction and must precede the set-up of the optimization problem. In Section 4 a classical procedure to capture the individual risk attitude is described for completeness.

An interesting peculiarity of ALM modeling arises in presence of several targets distributed over time: it is assumed that no preference among the different objectives is present and their time distribution is not fixed. These are rather weak and generally accepted assumptions. The scenario structure of the stochastic programming model simplifies the treatment of this type of problem in which the individual will achieve its targets depending on the realized scenario at possibly different time stages. Consider the SQP problem:

$$\max_{x \in X} E \sum_{t=1,2,\ldots,T} \left\{ \lambda_t W_t - \gamma_t \left(W_t - \widetilde{W}_t \right)^2 \right\}. \tag{8}$$

In (8) the set of coefficients γ_t will be positive and different from 0 only in the stages with associated targets. The coefficients λ_t may decrease before target periods and determine the switch from the risk neutral linear problem to the quadratic problem accordingly. A stage dependent trade off is considered, for instance, in Barro and Canestrelli (2004). In this setting the preference partial order over states of nature is enriched within a dynamic framework. In general the choice of a sequence of period utility functions or a utility function directly on the horizon terminal wealth, can be used to force the trajectory of the wealth process over scenarios of the stochastic programming problem. Different utility functions can be used to perform this task with different risk properties (Hakansson, 1974; Kallberg and Ziemba, 1979, 1983).

Figure 5 reports a set of utility functions that have been adopted in ALM applications to capture the decision maker risk attitude: the quadratic utility, the log utility, the power utility and the exponential utility functions.

The exponential utility function is also referred to as the constant absolute risk aversion (CARA) utility because its Arrow–Pratt absolute risk measure is constant and independent of Wealth, a feature that quite often leads to the preference of the power utility, also referred to as (CRRA) utility, because the relative risk aversion is constant (Hakansson, 1974). The log utility is indeed a limiting case of the power utility and has similar risk properties (see Thorp (2006) and MacLean and Ziemba, (2006) in volume A of this Handbook).

The quadratic utility, as reported, has several variations, in particular, to penalize only downside risk as in Villaverde (2003), or in tracking models.

In any instance a module for risk aversion estimation needs to be included in the model and the dependence on the terminal wealth or the excess wealth requires the specification of an underlying model of uncertainty. In Section 4 we describe the canonical procedure commonly used to determine the risk preference of an individual. It has been

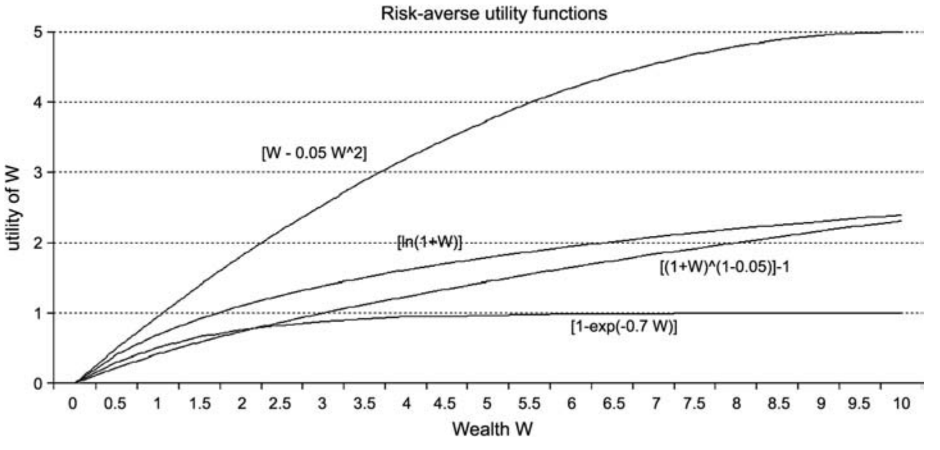


Fig. 5. Examples of concave utility functions of terminal wealth.

mentioned above that when modeling long-term decision problems, which depend crucially on human preferences, on the other hand, we may want to allow a certain degree of flexibility in the model and relax the assumption of a given constant risk-aversion of the decision maker. Already now, in the interview that typically anticipates the definition of a personal financial plan (see Appendix B), the wealth manager tries to estimate the individual reaction to alternative future economic scenarios, under which some or all the agreed targets can become unreachable. Again the individual psychology comes into the picture and the decision tools currently used in the industry are far from being able to use to full extent the richness of the provided information. The developments described here try to a certain extent to go in that direction, particularly when considering time dependent utility functions and scenario dependent targets and policies. Further efforts in this direction from the scientific community are in any case desirable. The definition of a flexible approach to individual financial planning is to a certain extent behind the adoption of simulation methods for alternative policy rules evaluation. Policy simulation as briefly explained in Section 3.4 is increasingly suggested to support more complex and structured stochastic programming developments. The approach also requires the introduction of an appropriate model of uncertainty.

3.3. The model of uncertainty

In this section, we address an issue quite often understated when modeling individual rather than institutional ALM problems: the relevance and general implications of the adopted model of uncertainty (Dupačová, Consigli and Wallace, 2000; Høyland and Wallace, 2001; Dempster, Scott and Thompson, 2002). This takes the form of a random vector process whose evolution over time will determine the likely achievement of a target and the financial soundness of the family budget. Quite contrary to general beliefs, the scenario model for the individual portfolio is rather complex and requires extended and careful statistical testing.

As further clarified in Section 4.1, a key aspect of the individual ALM problem is represented by the required estimation of the random behavior of both the individual wealth and the target values. Their driving financial processes can be explicitly considered in this introductory section to clarify the relevance of the scenario model in dynamic stochastic programming problems of this type.

In the case problem in Section 4, an individual pursues a double objective, buy a country house and retire with a sufficient amount of wealth.

The following set of dependencies can be postulated:

- The individual wealth dynamics is a function of the initial wealth and the cumulated portfolio gains/losses, the salary, the pension contributions, the family general costs, the fiscal treatment, others,
- The target values do depend on the appreciation rate of the real estate market and the projected terminal pension value.

In this simple framework, the following dynamic equations must be considered for $t = 1, 2, \ldots, T$:

$$W(t) = W(t-1) + I(t) + R(t) - C(t), \quad \text{given } W(0), \tag{9}$$

$$I(t) = G(t) - \mathrm{T}(t), \tag{10}$$

$$R(t) = x(t-1)' r(t) + F(t-1)' \varphi(t), \tag{11}$$

$$C(t) = L(t) + \Theta(t) + D(t), \tag{12}$$

$$\widetilde{W}(t) = H(t) + \delta(t, T) F(T). \tag{13}$$

Eq. (9) decomposes the current wealth $W(t)$ at time t, in the previous period wealth plus the individual income, plus the portfolio return, minus current costs. The income $I(t)$ is defined in Eq. (11) by the period salary $G(t)$, minus tax payments $\mathrm{T}(t)$.

The portfolio return $R(t)$ is also decomposed in Eq. (12) into the return of the financial managed portfolio $x(t-1)'r(t)$ and the revaluation of the pension fund denoted by $F(t-1)'\varphi(t)$.

The costs $C(t)$ are generated in (12) by living costs, pension contributions and fiscal costs.

Finally, the current targets value is computed in Eq. (13) as the sum of the house value and the discounted value of the pension fund terminal target value. The discount factor $\delta(t, T)$ in (13) can be assumed to be the risk free rate for that maturity, in order to enforce a sufficiently conservative provisioning policy.

Let us consider more in detail the stochastics' behind the random evolution of the processes (9)–(13).

3.3.1. Risk factors: stochastic and deterministic variables

The wealth and target wealth evolution can be modeled in discrete time approximating their continuous distribution through a finite number of states, resulting in a scenario tree over the planning horizon. The wealth scenarios from Eq. (9) will depend on the random behavior of several factors that represent the core part of the stochastic model for the problem.

Consider again the decision sequence described in Figure 2 in Section 2. Time evolves from $t = 0$, current time in which the first decision is taken, and $t = 1, 2, \ldots, T$ subsequent points in time in which the investor will analyse the results of previous decisions according to the currently revealed uncertainty and take corrective actions. Uncertainty is modeled through a discrete and finite number of scenarios $\omega := \{\omega_1, \omega_2, \ldots, \omega_T\}$. The realizations of this random process need not be regularly distributed over time, but in general we allow inhomogeneous time periods, or stages of the stochastic programming problem.

At the horizon T, the decision maker can only observe the result of its strategy in terms of goal achievements and $x_T \equiv 0$ by definition. Assume that the targets are incorporated in the objective as scenario dependent wealth values with a penalization over time below such targets.

An estimated risk aversion parameter will instead define the utility over the terminal wealth.

It is convenient to analyse explicitly the time evolution of the dynamic problem and the role played by previous decisions on subsequent dynamics. For this is convenient to distinguish between points in time and time periods in the labeling of the relevant problem variables.

The first decision, also called the *here-and-now (H&N)* decision, is the only one taken facing complete uncertainty over the future. At any point in time the investor knows the values $\nu_0, \nu_1, \nu_2, \ldots$ determining his wealth, such as prices or exchange rates, or other random variables. These values, after $t = 0$, define the wealth evolution after the strategy has been revised. At the beginning of each period, the decision maker can rebalance his portfolio with an effect revealed by the sequence of returns $r_1, r_2, \ldots$ in period 1, 2, etc., where $r_t = \nu_t/\nu_{t-1} - 1$, $t = 1, 2, \ldots, T$.

Already at time 0 we do model a revision of the previously available set of allocations, defined by $\vec{x}_0$, to account for the uncertainty, which is faced by the investor at that point in time. As further clarified by the model instance in Section 4.1, the signs $+$ and $-$ attached at the sequence of decision periods refer to investments, additional resources for future commitments, and disinvestments decided by the individual.

The decisions following the first one, all depend on previous decisions, the assumed realization of the return process and its likely conditional future behavior. Observe in Figure 6 that a wealth value will thus be defined at each time point *before* and *after* the strategy revision along the different scenarios $x = \{x_s; s = 0, 1, \ldots, t; t = 1, 2, \ldots, T\}$.

In principle the investor is expected to formulate and solve a new instance of the pfp problem as the uncertainty resolves and the feedback from previous decisions is known.

The solution output is represented explicitly by a sequence of optimal decisions along a scenario tree, whose structure is introduced in Figure 7 (see Chen et al., 1997).

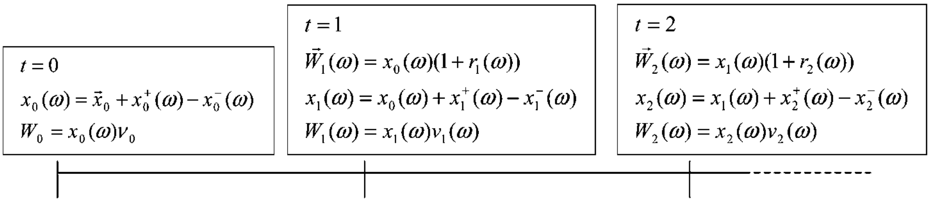


Fig. 6. Time points and periods in the DSP problem.

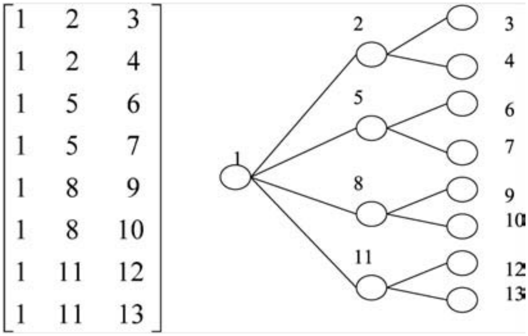


Fig. 7. Example of scenario tree and nodal partition matrix.

Each node in the tree corresponds to a different state of the vector process ω_t, whose elements are returns and prices, while a scenario is a complete path from the root node to the leaves. The conditional structure of the tree can be conveniently represented by his nodal partition matrix, in which each row corresponds to a scenario and each column to a time stage. The tree on the right is associated with the partition matrix on the left.

The scenario generation algorithm evaluates the probability of each node conditioned on the state of the process at the previous stage. The probability of each scenario is derived as the product of the conditional probabilities of the constituent nodes.

The model described in Section 4.1 requires the specification of an extended set of random coefficients along a sequence of scenario paths: each such scenario yields a different wealth trajectory and surplus or deficit against the prescribed targets over the planning horizon.

The scenario model is built on a subset of dynamics. As in the model in Section 4.1, we label the set of scenarios by $s = 1, 2, \ldots, S$: each scenario is described by a sample path from the root node to the set of leaf nodes at the horizon. The state of the scenario process at time t in scenario s is given by $\omega_t(s)$.

Particularly for long-term financial planning problems a distinction between nominal and real values needs to be made. Consider Eq. (10) above for the individual income: the salary is typically represented as a predictable process known up to a certain horizon. Similarly the individual pension provision flows. These amounts are known over short time periods but their purchasing power over long horizons is affected by the inflation rates observed over time. A proper model of uncertainty must thus consider not simply the financial and real variables whose evolution over time cannot be predicted, but also those quantities that even if deterministic in nominal terms and thus predictable, have an impact on the decision problem which is function of a random variable, as motivated above. In this perspective the role played by economic variables such as the inflation rate, given the long-term nature of the decision problem is well documented in the literature (see Dempster et al., 2002; Mulvey and Thorlacius, 1998; Wilkie, 1995; Cariño et al., 1994) and must be considered in the model set-up.

3.3.2. Economic and capital markets modeling

Consider again the case problem of an investor who dreams a country house and a sufficient amount of wealth upon retirement. He is married with two children and a generally expensive city life with private insurance and a defined contribution pension scheme. Under these assumptions: what are the factors, whose behavior over time will influence the gap between current and target wealth and thus the success of a strategy aimed at achieving those targets?

In first instance:

- The inflation rate: the investor expects to maintain his purchase power in real terms and be able to buy a house whose revaluation value will be bounded from below by the inflation rate. Furthermore the inflation rate will determine salary increases and the monetary policy of financial authorities: thus the movements of the interest rate term structure.
- The yield curve movements will be determined by the market forces and the interventions of central authorities, these will determine the price fluctuations of risk-free and risky fixed-income securities.
- The individual saving ability: this is a function of the salary evolution and the profitability of the adopted financial strategy. Financial returns do specifically influence:
  - The value of the life insurance of the investor, as a result of the adopted defined contribution scheme.
  - The individual financial wealth dynamics.
- Financial returns for the assumed asset and liability classes depend on:
  - Market expectations on equity earnings and dividends.
  - Risky and risk-free returns of fixed income securities.
  - Dynamics of the US corporate curve.
- The dynamics of the real estate market in general and with respect to the area of investment: this is a function of the inflation rate but also of the yield curve, that will determine the convenience to borrow for the individual investors.

These are the elements of the vector process ω whose evolution over time must be captured in the stochastic programming formulation and will determine the scenarios in the stochastic program.

Each scenario represents a possible evolution of those variables from the current time up to the final horizon along trajectories that must be explicitly modeled and associated with the problem control variables in a scenario structure such as the one described in Section 3.3.1. The decision process can be revised over time as the uncertainty unveils: the first stage decisions, however, are taken here and now and cannot be revisited by the investor. The model of uncertainty, for this reason, is typically expected to be accurate and well calibrated in the short run but still able to include consistently the evolution over the planning horizon of quantities typically correlated with each other. The distinction between an economic model and a capital market model is functional to the definition of long-term equilibrium conditions in the capital market.

Several approaches have been proposed in the literature over the years (Dupačová, Consigli and Wallace, 2000) to perform this delicate task. More recently two systems worth mention in the area of market capital and economic modeling for long-term financial planning problems are Dempster's et al. (2002) capital market model, developed as part of the Pioneer Global fund management system and Mulvey's Cap:Link model originally developed in 1996 (Mulvey, 1996) and then enriched and implemented in several application areas in the last few years. These are adopted as reference models in several stochastic programming applications and represent the building blocks of the scenario generator.

The interaction between economic and market variables can be captured within hierarchical models such as the one depicted in Figure 8. Some of the dependencies discussed above are here explicitly considered.

As shown in Figure 8, the economic policy is typically treated as an exogenous variable and similarly the impact of external imbalances on the exchange rates is not directly modeled in order to avoid unnecessary model extensions. The dependence of the financial variables relevant for the ALM problem—asset returns and liabilities, real estate

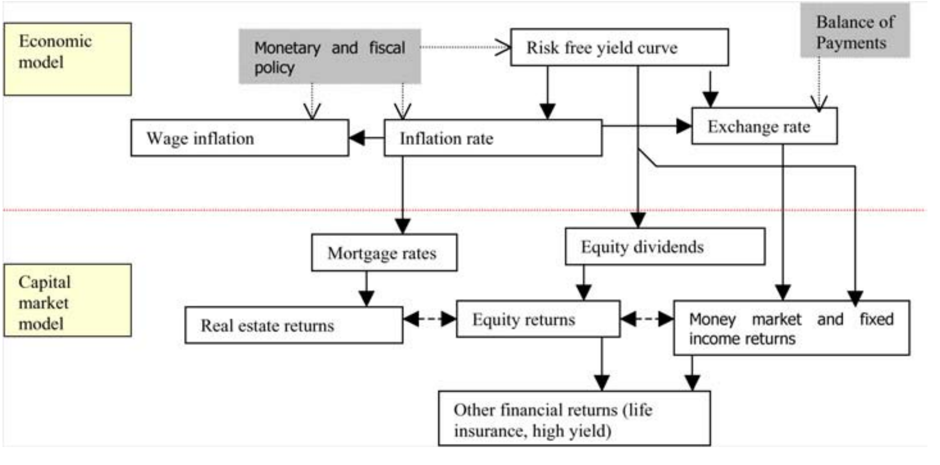


Fig. 8. Cascade structure underlying the generation of long-term financial scenarios.

prices—on economic variables is instead included within a hierarchical structure, and do imply an explicit stochastic model for a limited number of state variables—short- and long-term interest rate, inflation and exchange rates, risk premiums—and the derivation of other quantities—equity and fixed income returns as well as mortgage interest rates—as function of the above processes.

Chen et al. (1997) distinguish between the definition of the scenarios for the core random variables of the model—or core uncertainty—and the generation of the actual coefficients of the stochastic programming problem. The latter are typically derived as deterministic functions of the former, a transformation that however have implications on the properties of the associated probability space.

The implementation of the scenario generator requires the definition of the model instance, the estimation of an extended set of econometric relationships and the generation of a scenario tree whose state and time structure needs to be consistent with the general instance of the optimization problem.

The relationships in the above figure imply the estimation of an extended set of coefficients and have heavy statistical implications in terms of both parameter estimation and model validation (see Dempster et al., 2002; Ziemba, 2003).

The arrows in the model structure imply a model choice and essentially different model instances (see Mulvey and Thorlacius, 1998). Each relationship furthermore leads to a different dynamics, depending on the adopted model of evolution: first order and second order autoregressive equations (Wilkie, 1995) are often used in long-term financial planning problems to ensure the system stability. A model with stochastic volatility is described in Section 6 that allows the switching between volatility regimes.

A global return stochastic model with nonlinear dynamics and correlated variables is proposed by Dempster et al. (2002).

A target or goal such as buying a house requires a model of evolution for the real estate market with a number of specific issues. Similarly the inclusion of a financial asset such as an insurance contract, do generally require a model of evolution for the salary on which the contribution rate is defined and a return model for the pension fund itself, funded over time by the individual.

The depth of the model also implies a choice and the solution of a trade-off between the parsimony of the representation and the estimation of the model equations (see Section 4).

The outlined model of uncertainty is rather general and this type of framework is increasingly applied to general ALM problems or even for forecasting and simulation purposes.

The literature in the area of financial and economic models, with a cascade structure from core economic variables, to a certain extent falling in the class of the objectives of economic and monetary policy, towards financial market variables, also heavily influenced by market forces, is rich and growing (Vasicek, 1977; Olsen et al., 1992; Sachs and Larrain, 1993; Brennan and Schwartz, 1998; Ermolieva, MacKellar and Westlund, 1999; Vitale, 1999; Zhou, 1999; Abaffy et al., 2000; Hodrick and Vassilou, 2002).

The presence of long-term dependence from macroeconomic variables, furthermore, has general implications on the analytical structure of the asset and liability return model. Few desirable properties recently considered in capital market modeling are the return mean-reverse behavior, the presence of autoregressive components and possible volatility regimes switching within correlated markets. The inclusion of a mean-reversion element is consistent with the assumption of a long-term stationary equilibrium level, to which the return processes will tend over time. This equilibrium level is typically a function of the expected inflation and the market risk premium.

Auto-regression implies a time series which fluctuate around a certain equilibrium level and at each step the process *reacts* from a previous deviation with 1 or 2 or 3 time lags leading to AR(1-2-3) type of models. The possibility of time-varying variance–covariance matrices in Gaussian models, leads, for instance, to the estimation of heteroskedastic correlated processes of ARCH-GARCH(p, q) type.

The analytic properties of the economic and capital market model influence the techniques adopted to estimate the model parameters, an input to the scenario tree generator. Several methods have been proposed and actually implemented to achieve robust and consistent estimators for hierarchical economic and financial systems such as the one discussed above. Dempster et al. (2002) report results obtained for their global return model, relying on classical ordinary least square (OLS) estimation and seemingly unrelated regressions (SURE). Brennan and Schwartz (1998) use nonlinear SURE estimation to estimate the coefficients of their extended equity-bond model. Mulvey, Rosenbaum and Shetty (1996) discuss the steps required to perform different methods, namely generalized and simulated method of moments (GMM and SMM) estimation, integrated parameter estimation (IPE) for scenario models and adopted in the set up of the Cap:Link system (Mulvey and Thorlacius, 1998). Gouriéroux, Monfort and Renault (1993) apply indirect inference to estimate a multivariate financial model. In those cases in which the likelihood function can be specified, maximum likelihood (ML) and Quasi ML (QML) estimation is preferable as done in Section 4 and in a recent example of a three-asset model (MacLean, Consigli and Ziemba, 2004).

Whatever the method, the input of the parameter estimation technique (see also as general reference Campbell, Lo and MacKinlay, 1997) is a possibly extended set of historical data—with monthly or quarterly data—and the output is a vector of coefficients to be implemented in the scenario generator to derive a set of likely economic and financial scenarios for the optimization problem.

The dependence of the problem solution on the generated scenarios is explicit within stochastic programming developments. An issue arises when from alternative scenario generation techniques, with given outputs, a scenario tree needs to be derived with a structure consistent with the stages of the optimization problems. Indeed it is common to observe ALM applications supported by a model of uncertainty resulting in either recombining multinomial structures or simulated parallel trajectories generated via Monte Carlo techniques (Chen et al., 1997). This leads to a sampling issue, whose implications are briefly clarified here next.

Eqs. (9) to (13) highlight the driving forces underlying the evolution of wealth over time. Financial markets do operate continuously in time in different world areas and depending on the financial instruments included in a financial position, individual wealth does also evolve continuously and can take any real value. The problem discretization induced by the stochastic programming framework does inevitably lead to two types of error that need to be controlled.

The time discretization through the stages has, in particular, called for the inclusion of a possibly unequal stage partition over the planning horizon. Different problems can lead to a different stage partition, with typically stage lengths increasing as we move further in the future.

The state space discretization leads to rather saddle issues. The underlying continuous probability distribution can lead to significantly diverging wealth values. Several methods have been proposed to handle this issue (Dupačová, Consigli and Wallace, 2000).

Dempster (1988) has originally addressed the issue of optimal tree approximation relying on the expected value of perfect information as relevant importance sampling criterion and Chen et al. (1997) following up from subsequent relevant results have outlined a recursive algorithm for tree expansion and stochastic problem solution conceived to reach a sufficient description of the problem uncertainty. The key idea here is that, particularly in presence of long planning horizons and rich decision spaces, not all scenarios need to be relevant for the solution of the optimal problem. Emphasis is then put on the identification of a number of scenarios sufficient for a reasonable problem solution.

In general the discrete wealth scenarios introduced in the model are assigned equal probability of occurrence not to alter the frequency distribution directly generated by the adopted stochastic model. It is worth pointing out that the sampling procedure does alter the probability space originally adopted to represent the model of uncertainty: this is particularly relevant when the original stochastic model is constructed so to satisfy valuation criteria such as the free of arbitrage market condition. In binomial or multinomial settings: this condition does not need to be satisfied after the sampling has taken place. In general portfolio optimization problem solutions are expected to reflect individual risk preferences, as pointed out above and do not require the arbitrage-free condition to hold.

Pflug (2004) has introduced a scenario probability-weighting scheme consistent with a discrete approximation that minimizes the Wasserstein or other probability distance measures from an underlying continuous distribution.

3.4. Optimal strategies and policy evaluation

The output of a scenario generator is represented by possible future evolutions of financial scenarios expected to influence the individual wealth and thus the course of actions the investor might take. A discipline on the optimal strategy can be introduced by limiting the range of possible decisions through the introduction of policy rules (Mulvey et al., 2007). Alternative rules can then be evaluated within a Monte Carlo simulation framework: in our context, for instance, we may wish to consider the effectiveness of a life cycle strategy or rather a target-oriented fixed mix strategy (Dempster et al., 2002).

The goal of policy optimization is here to simulate individual wealth dynamics within an optimization framework. The output of the optimizer is a set of preferable policy rules given the generated set of scenarios.

Consider for instance the stochastic program (3) and the wealth model described by Eqs. (9)–(13): the model of uncertainty determines the probability distribution of the individual wealth as well as the target wealth the individual would need to achieve the selected targets given the current wealth distribution and the expected consumption flows. Rather than solving directly program (3), we may wish to introduce a rule under which the individual decisions are determined by the target wealth dynamics. A fixed mix strategy, for instance, would lead to a set of portfolio rebalancing decisions needed to keep a constant ratio between individual and target wealth across the different scenarios.

More generally a policy rule can be regarded as an additional constraint—typically in parametric form—on the dynamic decision process. In general as the rule changes so will the associated individual wealth distributions and by simulation, alternative policies can in this way be evaluated.

A further example is provided in a financial context by the identification through financial engineering (see Section 2.5) of a replicating portfolio when the terminal payoff is considered as a function of the individual targets under arbitrage free conditions. The cost of hedging against target deviations through portfolio rebalancing can in this case be estimated as customary with option contracts.

The introduction of a decision rule may cause the mathematical program to become non-convex and thus calls for the adoption of dedicated solution methods (Mulvey, Gould and Morgan, 2000; Dempster et al., 2002).

Monte Carlo simulation is currently heavily used in order to validate the outputs of scenario generators: assumptions on the stochastic nature of the financial processes are accepted when through extended back-testing generated scenarios are proven consistent with actually observed market scenarios.

Similarly the evaluation of policy rules and the back-testing of suggested policies on the wealth dynamics can provide both an important input to an SP-based ALM system and a validation tool on its own. As shown in Section 4.3, the generation of output information accessible to the decision maker represents a key and yet underestimated aspect for the widespread adoption of SP-based ALM decision models. The introduction of advanced simulation tools in integrated decision support systems is also for this reason becoming essential.

Policy optimization, furthermore, requires an effective interaction between the wealth manager and the investor: rules are to be agreed with the investor and the suggested optimal strategy becomes in this way easily understandable by the decision maker, an aspect not to be underestimated.

4. The individual investor stochastic programming model

We now consider a representative personal financial planning (pfp) problem and analyze the associated development steps in full generality. Alternative modeling choices resulting from practical application in this area are considered. We also present a generic model instance for individual AL management that can be regarded as a basic benchmark model for dedicated developments in this area. The analysis is developed introducing as a starting point a case-problem description that will also inspire the model outlined in Section 4.1 and Appendix A, and the described stochastic return model.

We consider an individual with a certain current income and a forecast of future inflows and outflows mainly related to foreseeable commitments due to his family life (son school, subscriptions, trips, etc.): these can be regarded as sources and uses of funds over a long planning horizon. He also has a number of objectives or *targets*, such as to buy a house on the seaside or buy a boat, he strongly wishes to achieve over a finite time horizon together with a sufficient wealth at the time of retirement.

We now consider the following case study, on which grounds the benchmark model will then be specified.

Case-study

A one salary family with two children and an expected time to retirement of around 25 years.

The targets are:

- Buy a country house in Tuscany with few acres of olive trees and grapes for personal wine production, currently worth 300 000 euros, and
- Retire with a lump sum payment of around 200 000 current euros.

At the time in which the individual goes to the financial intermediary the value of the country house is known.

The initial conditions of this financial planning problem are given by

- The monthly net income,
- An international portfolio in stocks and bonds,
- An estimate of life monthly expenses (house, food, car, trips, etc.),
- A current son and daughter monthly estimated costs (school and other expenses),
- Insurance monthly costs (life and health).

Resulting in a monthly income surplus of around 1000 euros.

Ceteris paribus, nothing changing, without any planned financial scheme and without life insurance, the individual can plan a total saving amount over the 25 years of $500 * 12 * 25 = 150\,000$ euros to be added to the portfolio value in order to estimate the projected difference between the individual wealth and the target wealth.

The financial intermediary must solve this problem and suggest a sustainable policy taking into account the randomness of the individual salary, the horizon target values and the possibility to invest and borrow in the future at the prevailing unknown prices and interest rates.

The definition of the asset and liability universe does precede the estimation of the individual risk aversion and is generally driven by the intermediary. In full generality we consider a Euro-based investor and the following asset and liability classes carrying both market and credit risks:

- USD and Euro denominated equities,
- Euro-denominated government bonds,
- USD risky fixed income securities,
- Money market securities,
- Life insurance minimum guarantee instruments.

No short selling is allowed and borrowing is bounded by the individual creditworthiness. The borrowing rate is given at any point in time by the current government bond yield-to-maturity plus a fixed spread of 4%.

We can refer to Figure 3 for the methodological steps linking the problem formulation to its solution. The required dynamics of the individual AL problem and the interaction between random events and decisions can be easily implemented for the described decision problem in a stochastic programming framework.

In general the following decision paradigm needs to be somehow embedded in the problem formulation and accounted for in the mathematical specification of the problem (see Figure 9).

This sequence will occur along any scenario, whatever the individual objective function.

In our representative case problem we consider a one-salary family with two parents, two children and essentially two targets: buy a house in the Italian country side and retire with a sufficient amount in his pension fund. The children are expected to stay at home until they complete their university studies.

Both targets depend on random processes that will affect the individual all along the planning horizon. The house price at any future date does depend on a real estate random

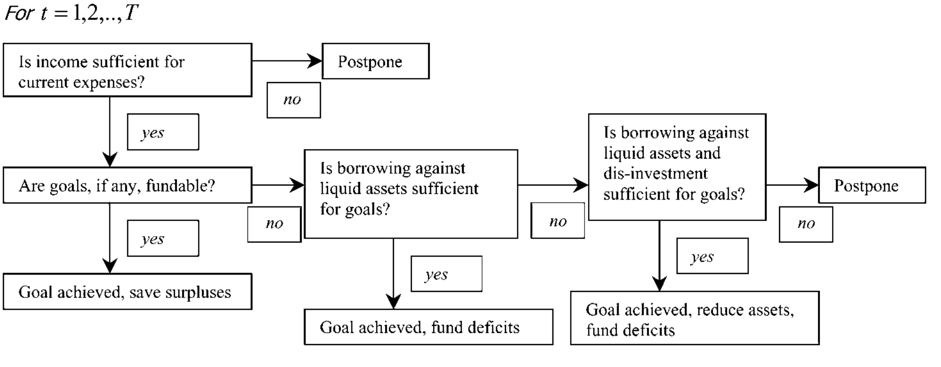


Fig. 9. Canonical decision sequence in individual ALM problems.

appreciation rate in the area of interest. The pension fund provision and the associated retrievable value will depend on the pension year-by-year installments and revaluation, which will in general depend on financial markets performance.

Targets are thus scenario dependent as the wealth process that is expected to fund those goals. The solution of the optimization problem will suggest, at the given conditions, the strategy, if any, that will maximize, under the given constraints, some function of the stated targets. In the sequel we clarify the mathematical tools and related computational implementations that at the current state-of-the-art can handle effectively this problem.

Observe that the target pension value can be included in the objective at the end of the working life: this is typically the horizon of the individual ALM problem. Also observe that depending on the individual aversion to risk, the optimal problem can be defined in such a way so to reach the targets for sure, at any cost, or on average, without incurring in excessive risk taking and/or borrowing: the definition of an appropriate risk-return trade off in a dynamic setting is central to the problem definition.

4.1. A benchmark model for individual AL management

We discuss in this section the general features of an individual ALM model incorporating the elements considered in the previous sections. The generic case problem introduced above does provide the decision framework relevant here. The following Section 5 and Appendix B consider the implementation and solution steps of the stochastic programming problem in general and with respect to a specific industrial development. The mathematical instantiation of the model in algebraic form is described in detail in Appendix A. Here we do analyse its specific features as benchmark model for individual ALM.

We consider a risk-averse individual who seeks the maximization of a nonlinear objective function that depends on the assumed intermediate and final goals, and whose decision strategy is constrained by a set of general and specific constraints to be satisfied (with probability one) along a discrete number of scenarios and time stages. The problem falls under the class of DSP problems (3) described above.

The model building blocks are (see Appendix A): the index sets, the decision and the model variables, the scenario-dependent coefficients and an objective function with a set of linear constraints.

Related to the case study in Section 4:

- The variables are indexed with respect to
  - Time stages $t = 0, 1, \dots, T$, with the horizon coinciding with the expected retirement age,
  - Investment classes $i = 1, \dots, I$ to be considered in the allocation strategy,
  - Scenarios $s = 1, \dots, S$, sample paths of the associated random process from the root node to the horizon.

  In presence of several liability classes we would also have an index for the liabilities. Typically this is not necessary for individuals that have only limited borrowing options and in our case problem the only one is a mortgage for the real estate investment.

- The decision variables are thus related to investment and borrowing decisions, excess or deficit in the personal cash account and a provision coefficient that determines the amount the individual wishes to allocate in the pension fund, a constant proportion of his current income. At each stage the individual, given his family current consumption rate, must decide:
  - The investment in the financial portfolio, that, possibly together with a mortgage, will finance the country house, and
  - How much to allocate in the pension fund for retirement purposes.

  The variables considered in the model are: $x_{i,t}(s)$, $x^+_{i,t}(s)$, $x^-_{i,t}(s)$ for the amount of asset i held, bought or sold, respectively, at time t under scenario s. $y_t(s)$, $y^+_t(s)$, $y^-_t(s)$ for the debt held, increase and reduction at time t under scenario s, the cash surplus or deficit $z^+_t(s)$, $z^-_t(s)$ with $z_t(s) = z^+_t(s) - z^-_t(s)$ for the balance again at time t under scenario s and finally the provision coefficient ψ for the pension plan.
- The following model variables provide a minimal set of equations needed for an appropriate problem instantiation. The family consumption is typically treated as the key control variable in general stochastic control problems (Merton, 1969). Here instead this is an input to the problem, exogenously determined as a given constant share of the current income. This assumption can be easily relaxed. The target wealth is here associated with the described individual goals: buy the country house around the 10-year horizon and retire at the end of the working life with a sufficient income. The latter coincides with the stochastic program horizon.

  The following notation is used with t current time and s for the scenario: $W_t(s)$ is the individual wealth, $C_t(s)$ is the consumption, $I_t(s)$ the individual income, $F_t(s)$ the pension fund value, $\widetilde{W}_t(s)$ the target wealth and $\widetilde{F}_t(s)$ the target pension fund value at the horizon.
- The above set of decision and model variables has an associated extended set of financial and economic coefficients needed for the correct representation of the problem uncertainty. A deterministic, though dynamic, problem would be generated assuming a single possible realization of the given set of prices and returns over the planning horizon. In the stochastic case we allow for several discrete coefficient realizations as described in Section 3.3. As above for each coefficient we indicate by t the current time and s the current scenario. Consistently with the problem dynamics and the required non-anticipativity of the decision sequence here time t refers to the end of the period and anticipates the recourse decision.

  We indicate with $r_{i,t}(s)$ the price return of asset i, $\zeta^{\pm}_t(s)$ the interest rates on excesses ad deficits on the cash account, $b_t(s)$ the 10-year borrowing rate, $\xi_{i,t}(s)$ the exchange rate on asset i, $\rho_{i,t}(s)$ the cash return of asset i, $\nu_{i,t}(s)$ the value of asset i, $\pi_t(s)$ the price inflation, $g_t(s)$ the output growth, $\varphi_t(s)$ the pension fund return, $h_t(s)$ the real estate return and $\delta_{t,T}(s)$ the discount factor at time t for the horizon T. The model includes a set of deterministic coefficients, namely the transaction cost coefficient τ_i for buying or selling asset i and the maximum wealth fractions β_i for investments in asset i and available for borrowing γ.

The above decision variables do determine the wealth evolution described in Figure 6 following the dynamics of the set of economic and financial coefficients described in Section 3.3.2. The coefficient vector process determines the asset and liability nodal values along the scenario tree. The 10-year borrowing interest rate is considered as a benchmark for the mortgage rate. The pension fund returns and the real estate appreciation rate are treated as endogenous to the scenario generation. This is a desirable model choice, while several individual ALM tools—see Section 6—do impose predetermined and scenario independent target values that may very often cause the problem infeasibility or require unrealistic borrowing capacity. The pension fund return is independent from the individual investment universe. In general, it is desirable that the two returns, the portfolio return for the individual and the pension fund return, are not strictly correlated in order to introduce a form of internal indirect hedging in case of adverse scenarios.

The generic individual ALM problem is defined as a nonlinear dynamic stochastic program (NLDSP).

In Section 3.2.1 we have described a set of possible candidates for the accurate representation of the individual stochastic programming asset–liability management problem. The adoption of a quadratic programming approach, relying on the introduction of a sequence of period utility functions, is very much in the spirit of an individual AL decision problem in which the investor is primarily concerned with the risk of not achieving his goals. Eq. (6) in Section 3.2.2 is consistent with this approach and can appropriately be considered in this setting. For more generality, we do assume here an objective function explicitly including the investor risk attitude, expressed by a concave power utility function—see Eq. (5) in Section 3.2.1.

The objective is the maximization of the expected utility of terminal wealth minus the target wealth cumulative absolute deviations before the horizon:

$$\max_{x \in X} E\left[\frac{1}{\lambda}(W_T)^{\lambda} - \sum_{t<T} \gamma_t \left|W_t - \widetilde{W}_t\right|\right]. \tag{14}$$

In (14) x is a decision plan in the space X and E is the expectation operator. The objective is defined by the power utility of the terminal wealth $\frac{1}{\lambda}(W_T)^{\lambda}$ with λ as risk-aversion parameter and by an additive absolute deviation distance: $\sum_{t<T} \gamma_t |W_t - \widetilde{W}_t|$. The former is a classical utility function—see Section 3.2.1—already adopted in the past (see Ziemba, 2003; Brennan and Schwartz, 1998) and its implementation requires the estimation of the parameter λ (see Section 4.3). The latter considers the deviations before the horizon of the current from the target wealth: these are weighted by the sequence $\gamma_t \geqslant 0, t = 1, \ldots, T-1$. The individual is assumed to maximize the difference between a reward function, the terminal wealth expected utility, and a cost function, the deviations from a "sufficient" wealth trajectory; sufficient with respect to the adopted goals. The sequence γ_t can be calibrated on the specific targets, with increasing values from 0 before target period. The investor in our case study seeks a sufficient amount for

retirement at the horizon: as shown below this target is embedded in the model formulation.

The specification of the objective function requires the definition of the wealth and target wealth processes W_t, $\widetilde{W}_t$, respectively.

For the former: the current wealth is defined along each scenario by the previous stage endowment plus the cash account surplus, the individual income and the portfolio and pension fund revaluation, minus consumption and borrowing. For $t = 1, \dots, T$ and $s = 1, \dots, S$

$$\begin{aligned} W_t(s) = W_{t-1}(s) + z_t(s) + I_t(s) + \big[X_t(s) - X_{t-1}(s)\big] \\ + \big[F_t(s) - F_{t-1}(s)\big] - C_t(s) - y_t(s), \end{aligned} \tag{15}$$

where the initial conditions are given by

$$W_0(s) = X_0(s) + I_0 + z_0, \qquad X_0(s) = \sum_{i \in I} x_{i,0}(s).$$

So given the initial financial portfolio, the individual income and position on the cash account, at every stage, as a consequence of the costs incurred and the returns on the financial portfolio and the pension fund, the liquid wealth may increase or decrease.

The first stage H&N decision occurs at the root node and determines the portfolio value at the end of the first period. No borrowing is allowed at the root node:

$$\begin{aligned} &x_{i,0}(s) = \vec{x}_i + x^+_{i,0}(s) - x^-_{i,0}(s), \\ &y_0(s) = 0. \end{aligned} \tag{16}$$

The individual income, the portfolio return and the pension fund returns together with the consumption process are modeled endogenously and complete the wealth specification.

For $t = 1, \dots, T$ and $s = 1, \dots, S$:

$$I_t(s) = I_{t-1}(s)\big(1 + \pi_t(s) + g_t(s)\big), \qquad I_0(s) = \hat{I}, \tag{17}$$

$$\psi_t(s) = \psi I_t(s), \tag{18}$$

$$F_t(s) = F_{t-1}(s)\big(1 + \varphi_t(s)\big) + \psi_t(s), \quad F_0 = \widehat{F}. \tag{19}$$

The individual income is function of the current inflation rate and labor productivity and the contributions to the pension fund are expressed as a function of the individual income: the contribution rate is a decision variable and the fraction to be allocated for the desired fund revaluation is in this case an output of the problem solution. In a defined contribution scheme the fund value will increase in Eq. (19) due to new contributions and as a result of the fund market performance. Consumptions are determined as a fixed percentage of the individual income (see Appendix A).

The portfolio value depends on the adopted portfolio strategy. For $t = 1, 2, \dots, T$, $i = 1, 2, \dots, I$, $s = 1, 2, \dots, S$, we have

$$X_t(s) = \sum_{i \in I} x_{i,t}(s),$$
$$x_{i,t}(s) = x_{i,t-1}(s)\bigl(1 + r_{i,t}(s)\bigr) + x_{i,t}^{+}(s) - x_{i,t}^{-}(s). \tag{20}$$

Whereas borrowing decisions will determine the individual outstanding debt as:

$$y_t(s) = y_{t-1}(s) + y_t^{+}(s) - y_t^{-}(s). \tag{21}$$

Together with a set of bounds and non-negativity constraints (see the full model instance in Appendix A) these are the required dynamics for the wealth equation.

The target wealth in the objective function (14) is a function of the country house market value at the predicted time of investment and the desired final value of the pension provision. We define this value at every stage through a pair of equations. The house value, given the current market value of the real estate investment is just a function of the market evolution forward in time. The latter can be defined as the terminal pension fund value discounted at an appropriate rate.

For $t = 1, \dots, T$ and $s = 1, \dots, S$:

$$\widetilde{W}_t(s) = H_t(s) + \widetilde{F}_{t,T}(s), \tag{22}$$
$$H_t(s) = H_{t-1}(s)\bigl(1 + h_t(s)\bigr), \quad H_0 = \widehat{H}, \tag{23}$$
$$\widetilde{F}_{t,T}(s) = \widetilde{F}_T \delta_{t,T}(s). \tag{24}$$

Additional bounds and non-negativity conditions can be required to specify the target wealth feasibility region, but in general the model key elements are all included in the set of Eqs. (15)–(24). The model instance is completed by the cash balance equation (see Appendix A) determined by the stage cash flows due to dividend and interest payments, contributions, investment and disinvestments, current consumptions and so on. As customary the inclusion of a cash account relaxes the liquidity constraints and induces the problem feasibility.

To summarize the following relevant features characterize the ALM model, a benchmark reference for dedicated developments (see Appendices A and B for ORS implementation framework):

- The individual seeks the maximization of the power utility of his terminal wealth at the horizon while minimizing along the stages of the problem the absolute deviations from the scenario dependent target values. Decisions are allowed up to $T - 1$, right before the horizon: the likelihood of reaching a sufficient amount of pension provisions at retirement is thus determined by the strategy followed until the penultimate period.
- The problem is specified as a convex nonlinear stochastic programming problem with linear constraints. It is a large-scale problem as the number of scenarios and stages do increase to yield a robust solution.
- Pension provisions are defined as a fraction of the scenario-dependent disposable income. The latter is here modeled as a function of the inflation rate and a coefficient of labor productivity: the two variables must be included in the scenario generator.

The provision coefficient is an output of the problem solution and can be calibrated in order to allow the required fund revaluation path. Or just given exogenously in order to check the target feasibility.

- The country house value is also endogenous to the decision problem and the revaluation coefficient of the house current price is also an output of the scenario generator.
- The period risk aversion coefficients in the objective function—Eq. (14)—γ_t can thus be calibrated in order to minimize the deviations from target values at specific points in time.
- The probability of reaching one or both targets is defined by the number of scenarios in which the terminal wealth is greater or equal to the target wealth specified by Eq. (24).
- At each stage the current wealth is defined by the previous-stage wealth, plus the financial portfolio and the pension fund value increments, plus the individual income and minus consumption and borrowing net of potential cash surpluses.
- Investment and borrowing decisions are bounded by the individual available wealth in the previous stage.
- The problem dynamics are specified by the stages $t = 1, \ldots, T$ that do not need to be equally distributed: the associated scenario structure determines both the number of economic and financial scenarios and their conditional evolution over the planning horizon and the number of nodal problems to be solved to reach the problem solution.

This requires, as specified in Figure 2, the generation of a mathematical program standard with scenario dependent coefficients and its solution with available solution algorithms. The exact derivation of risk aversion parameter and the model coefficients from the assumed economic and capital market model represent two preliminary steps to consider.

4.2. Risk aversion estimation

The estimation of a risk aversion parameter for the specification of the individual utility function determines the degree of risk the individual is willing to face in order to achieve the targets he has in mind. Utility theory as already discussed in Section 3.2.2 provides all the ingredients needed to address this issue. What is worth recalling here is that at this point in time almost all the dedicated individual ALM platforms do include a module for the estimation of the individual risk attitude.

The specification of the objective function in the individual ALM benchmark model requires the estimation of a set of risk aversion parameters. Figure 5 shows a set of concave utility functions with different Arrow–Pratt risk aversion measures with respect to the terminal wealth (see the discussion in Ziemba, 2003). The risk aversion coefficient is denoted in Eq. (14) by λ, while the set $\{\gamma_t,\ t = 1, \ldots, T - 1\}$ define a user input to be calibrated in order to force target achievements around desirable times.

Modern decision support systems include a module for the estimation of the risk aversion coefficient typically in the form of an interview in which the financial planner collects relevant information on the individual economic profile and decision making process and then the ALM system must convert such information into a risk aversion estimate. Here below we describe one of the many approaches that can be implemented to this purpose while in Appendix B we show how this type of information is collected within a commercially available tool. It's been pointed out that the simplification here introduced is in general not desirable: either by introducing a dependence of individual preferences on wealth scenarios or by incorporating period utilities in the dynamic stochastic programming problems we wish to overcome the restrictions associated with an assumption of long-term stability of investors risk preferences.

The risk parameter λ in (14) is typically estimated by interview using an iterative procedure (see also Ziemba, 2003) based on the definition of the certainty equivalent (CEV) of a random event, starting with the definition of the CEV of two equally likely extremes and then filling the center and finally fitting the discrete points with the polynomial utility function above.

Consider the utility of a payoff W in the set $[0, \overline{W}]$, with $u(0) = 0$, $u(\overline{W}) = 1$ (in case rescale to have maximum utility equal to 1). At the first iteration, the investor is asked which sure payoff W^1 he would consider equivalent to the gamble $1/2 * 0 + 1/2 * \overline{W}$. The latter is the expected monetary value (EMV) of the gamble. The answer W^1 is instead the certain monetary equivalent (CME) of the gamble. By construction, thus, the utility of this amount is equivalent to the expected utility of the gamble: $u(W^1) = Eu := 1/2u(0) + 1/2u(\overline{W})$.

Three possibilities arise, associated with different risk-attitudes:

(i) Risk averse investor: $u(\text{EMV}) > u(\text{CME})$,
(ii) Risk neutral investor: $u(\text{EMV}) = u(\text{CME})$,
(iii) Risk loving investor: $u(\text{EMV}) < u(\text{CME})$.

At the second iteration the investor is asked which payoffs:

- W^2 would leave him indifferent with respect to the gamble $1/2 * 0 + 1/2 * W^1$, yielding the CME for which $u(W^2) = 1/2u(0) + 1/2u(W^1)$ and which
- W^3 he would exchange with the gamble $1/2 * W^1 + 1/2 * \overline{W}$, yielding the CME for which $u(W^3) = 1/2u(W^1) + 1/2u(\overline{W})$.

Again, as before, to the inequality between the utility of the expected monetary value and the utility of the certain monetary equivalent will correspond different risk attitudes (risk averse, risk neutral and risk seeking investors).

The procedure in three iterations generates 5 estimates of the subjective utility function that together with the origin and the terminal wealth can be used to fit a polynomial utility such as the one in (14):

$$\min_{\lambda} \sum_{j=1,\dots,8} \left[\lambda^{-1} W^{\lambda} - u(W^j)\right]^2. \tag{25}$$

The solution of (25) generates the estimate of the risk-aversion parameter λ to be considered in the decision problem. In presence of high risk aversion the allocation process should over time concentrate on asset classes carrying moderate financial risk, such as bonds in liquid market. Similarly risk-seeking investors will privilege probability distributions with high dispersion around the mean. Kallberg and Ziemba (1983) provide an early but essential reference on the role played by risk aversion in alternative utility functions. The optimal strategy will be a function of the individual risk aversion and the financial scenarios upon which the contingent plans are defined.

The estimation of the problem coefficients—Eqs. (14)–(24)—allows the exact specification of the ALM problem for the single individual and the determination of the scenario-based strategies needed to achieve the targets. The phase of scenario generation thus leads to the definition of the sequence of coefficient realizations required by the optimization module to generate the optimal (stochastic) personal financial plans.

4.3. *Scenario generation*

The inputs of the scenario generator are represented by two elements:

- The set of economic and financial coefficients for the adopted dynamic capital market model, and
- A partition matrix needed to label the discrete evolution of the coefficient process and specify exactly the conditional structure of the scenario tree consistent with the stages of the decision problem.

The output is described by a sequence of return vectors, exactly matching the conditional nature of the decision problem formulated as a dynamic stochastic programming problem in scenario tree form.

Consider for instance the scenario representation described in Figure 7 in Section 3.3 above.

Investment decisions are associated with different asset classes and for each class we distinguish between a price return $r_{it}(s)$, resulting from the change in price of the asset, and a cash return $\rho_{it}(s)$, resulting from dividends paid by equity instruments and interest flows paid by fixed income instruments. When either an investment or dis-investment takes place, the asset value $\nu_{it}(s)$ along the scenario must be known in order to compute the associated cash flow. Furthermore, depending on the asset denominated currency, an associated exchange rate $\xi_{it}(s)$ must be known as well.

For the liability side: the individual is allowed to borrow at the current ten year borrowing rate and must put aside a portion of its current income for insurance provisioning. These quantities require the knowledge of the interest rate evolution $b_t(s)$ and the individual income, a function of the inflation rate $\pi_t(s)$ and work productivity $g_t(s)$. The provisioning will increase the value of the pension fund.

The pension fund evolution over time will also follow a return process described by $\varphi_t(s)$.

Targets valuation finally requires the knowledge of the rate of appreciation $h_t(s)$ of real estate investments over a certain horizon and an assumption on the discount factor $\delta_{t,T}(s)$ for the target pension fund value at retirement.

Some of these factors do depend on each other as discussed in Section 3.3.2 and this dependence yields the model structure shown in Figure 8.

Here next to limit the analytic burden and focus only on specific driving economic factors, we briefly describe the set of stochastic equations for the interest rates, the exchange rate and the inflation process in a two-currency system with five investment classes. The following assumptions have been adopted for these variables:

- The US and Euro short interest rates are modeled as mean-reverting AR(1) diffusion processes with random volatility. The yield curve can be derived from the evolution of these short rates;
- The US and Euro inflation rates as independent constant volatility models, functions of: the domestic short interest rate increments and the deviation of real interest rates from their long-term equilibrium values;
- The US–Euro exchange rate is modeled as a random volatility diffusion model which is function of interest rate and inflation rate differentials;
- Finally, the investment classes are modeled as mean-reverting autoregressive correlated processes with time-varying variance–covariance (Multivariate GARCH(1, 1)).

Scenarios for the investment portfolio and borrowing scenarios, denominated in the local currency, the euro, are ultimately derived from the above processes converted in the local currency. We can assume for the sake of simplicity that the financial and real returns required by the model in Section 4.1 can be all derived as functions of the described set of state variables.

We briefly outline the analytic structure of a scenario generator developed in support to the described financial planning problem, that follows the above structure.

System (26) describes the random processes for the short interest rates in the Euro and the US currency areas. The two processes are correlated with random volatility coefficients as shown in the second set of equations.

$$\begin{cases} \Delta b_t^{\text{Eur}} = \alpha^{b1}\left(\bar{b}^{\text{Eur}} - b_t^{\text{Eur}}\right)\Delta t + \sigma_t^{b1}\Delta W_t^{b1}, & b_0^{\text{Eur}} = \bar{b}^{\text{Eur}}, \\ \Delta b_t^{\text{US}} = \alpha^{b2}\left(\bar{b}^{\text{US}} - b_t^{\text{US}}\right)\Delta t + \sigma_t^{b2}\Delta W_t^{b2}, & b_0^{\text{US}} = \bar{b}^{\text{US}}, \end{cases} \tag{26}$$

where:

$$\begin{aligned} (\sigma_t^{b1})^2 &= \alpha^{\sigma 1}\left(\sigma_{t-1}^{b1}\right)^2 + \beta^{\sigma 1}\sigma_{t-1}^{b1}e_{t-1}^{b1}, \quad \sigma_0^{b1} = \bar{\sigma}^{b1}, \\ (\sigma_t^{b2})^2 &= \alpha^{\sigma 2}\left(\sigma_{t-1}^{b2}\right)^2 + \beta^{\sigma 2}\sigma_{t-1}^{b2}e_{t-1}^{b2} \quad \sigma_0^{b2} = \bar{\sigma}^{b2}. \end{aligned} \tag{27}$$

In (26) b_t^{Eur} and b_t^{US} define the interest rate in the Euro and USD zone, respectively, at time t and $\bar{b}^{\text{Eur}}$ and $\bar{b}^{\text{US}}$ are long-term equilibrium levels. α^{b1}, α^{b2} are the mean-reverting parameters for the Euro and the US rate, respectively. Finally σ_t^{b1}, σ_t^{b2} define the time-dependent volatility at time t as order 1 autoregressive equations with dependence expressed by the $\alpha^{\sigma 1}$, $\alpha^{\sigma 2}$, and $\beta^{\sigma 1}$, $\beta^{\sigma 2}$ parameters with $e_t^* \propto N(0, 1)$. The random terms $\Delta W_t^* \propto N(0, \Delta t)$ in (26) are Wiener processes—*white noises*—associated with each variable. We do allow ΔW_t^{b1} and ΔW_t^{b2} to be correlated.

The interest rates enter the model for the long-term inflation in each currency zone.

For the Euro zone, given the equilibrium condition $\pi_t^{\text{Eur}} = \bar{\pi}^{\text{Eur}}$ we have:

$$\Delta\pi_t^{\text{Eur}} = \alpha^{\pi 1}\Delta b_t^{\text{Eur}} + \alpha^{\pi 1}\left(\left(\bar{b}^{\text{Eur}} - \bar{\pi}^{\text{Eur}}\right) - \left(b_t^{\text{Eur}} - \pi_t^{\text{Eur}}\right)\right)\Delta t + \sigma^{\pi 1}\Delta W_t^{\pi 1}, \tag{28}$$

And for the USD zone:

$$\Delta\pi_t^{\mathrm{US}} = \alpha^{\pi 2}\Delta b_t^{\mathrm{US}} + \alpha^{\pi 2}\big(\big(\bar{b}^{\mathrm{US}} - \bar{\pi}^{\mathrm{US}}\big) - \big(b_t^{\mathrm{US}} - \pi_t^{\mathrm{US}}\big)\big)\Delta t + \sigma^{\pi 2}\Delta W_t^{\pi 2}. \quad (29)$$

Eqs. (26)–(29) allow the derivation of the exchange rate process for the dynamic of the Euro versus USD.

The exchange rate does not enter the set of equations for the Benchmark model but rather allows the conversion of USD-denominated returns in the domestic Euro currency.

The exchange rate is expressed as a function of the short rate and interest rate differentials between the two currency areas. Denoting with $\Delta\xi_t$ the increment of the exchange rate between t and $t+1$, and with $\alpha^{1,\xi}$, $\alpha^{2,\xi}$, σ^{ξ}, respectively, the drift and volatility coefficients. ΔW_t^{ξ} are Wiener increments inducing the random behavior of the variable

$$\Delta\xi_t = \big[\alpha^{1,\xi}\big(b_t^{\mathrm{Eur}} - b_t^{\mathrm{US}}\big) + \alpha^{2,\xi}\big(\pi_t^{\mathrm{Eur}} - \pi_t^{\mathrm{US}}\big)\big]\Delta t + \sigma^{\xi}\Delta W_t^{\xi}, \quad \xi_0 = \bar{\xi}. \quad (30)$$

The set of Eqs. (26)–(30) form the first part of the scenario model corresponding to the economic variables. The equation for the Euro short rate on the other hand determines the dynamics of the mortgage rates and fixed income instruments.

The model for the asset classes is constructed as a multivariate GARCH(1, 1) correlated system along the lines described by Eqs. (31). We have five equations for the market dynamics in the form a system of s.d.e's and other five equations for the variance–covariance update.

Let r_t^j, $j = 1, 2, 3, 4, 5$, denote the random returns of the five benchmarks. We have, for $t = 1, 2, \ldots, T$ and $r_0 = \bar{r}$ with equilibrium values $\hat{r}$, in matrix form

$$\begin{cases} \Delta r_t = \alpha^r(\hat{r} - r_t)\Delta t + \sqrt{\Sigma_t}\Delta W_t = \alpha^r(\hat{r} - r_t)\Delta t + \sqrt{\Sigma_t}e_t\sqrt{\Delta t} \\ \qquad = \alpha^r(\hat{r} - r_t)\Delta t + u_t\sqrt{\Delta t}, \\ \Sigma_t = \Sigma + \phi^1\Sigma_{t-1} + \vartheta^1 u_{t-1}^2. \end{cases} \quad (31)$$

In (31) we have, in the usual notation, three vectors r_t, $\hat{r}$, ΔW_t of dimension (5, 1). Σ_t is a (5, 5) variance covariance matrix, while the coefficient matrices ϕ^1, ϑ^1 define the GARCH(1, 1) set of equations and $e_t \propto N(0, 1)$ is a standard normal noise. Observe the variable substitution $u_t = \sqrt{\Sigma_t}e_t$, which implies that $\{\Sigma_t \mid \Xi^{t-1}\} = \Sigma + \Sigma_{t-1}(\phi^1 + \vartheta^1 e_{t-1})$, where Ξ^{t-1} defines the information set available at $t-1$.

This model requires the introduction of long-term average values for the associated asset classes: this is an input that will clearly influence the scenario dynamics. It is in generally included relying on independent studies and/or subjective views that are to be incorporated in the model.

The stochastic model instantiation requires the estimation of the parameters. The estimation method is largely based on maximum likelihood maximization via the BHHH algorithm (Berndt et al., 1974). For a description of the estimation method and the related statistical steps that we do not address here, we refer to Engle and Kroner (1995) and Bollerslev, Engle and Nelson (1994). Several alternative techniques have been mentioned in Section 3.3.2 and may also have been adopted in this context.

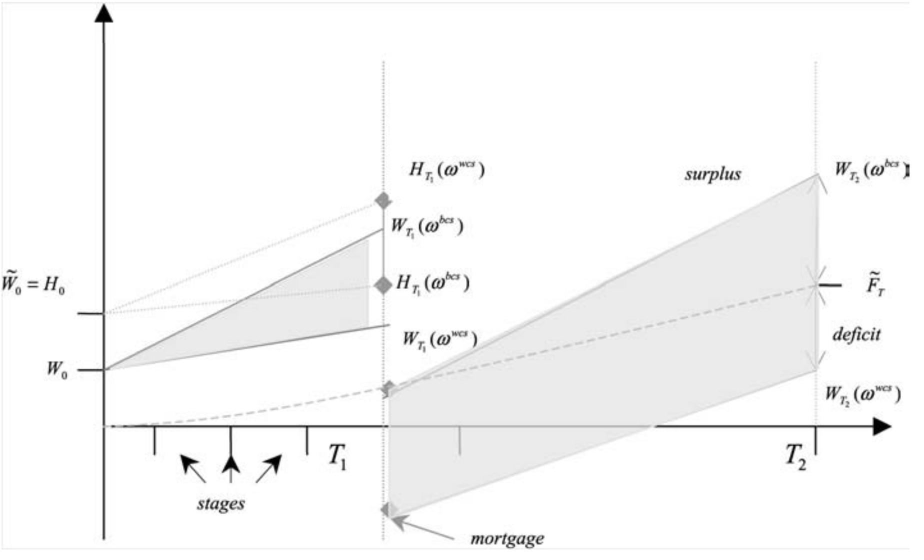


Fig. 10. Random wealth and targets in individual financial planning.

The estimation of the equations of the economic and financial market model provides an input to the scenario generator—see Section 5.2—and allows the derivation of the likely evolution of the target values over the assumed planning horizon.

The interdependence between planning horizon, disposable wealth and target wealth can be clarified in graphical form.

In Figure 10 we describe the possible wealth evolution excluding any rebalancing decision, based on the current wealth composition.

At the 10 year horizon—$T_1 = 10$—the country house value is expected to be within the $[H_{T_1}(\omega^{bcs}), H_{T_1}(\omega^{wcs})]$ range and given the current portfolio structure the individual wealth may or may not be sufficient to fund the real estate investment.

We show in Figure 10 a possible negative gap between current wealth, net of family current consumptions (see the model in Section 4.1), and 10 year house value in either cases of very positive $W_{T_1}(\omega^{bcs})$—best case scenario—or negative $W_{T_1}(\omega^{wcs})$—worst case scenario—wealth evolution.

The maximum financial gap occurs in presence of negatively correlated financial and real estate markets when the house value increases up to $H_{T_1}(\omega^{wcs})$ and the wealth evolution increases very moderately from W_0 to $W_{T_1}(\omega^{wcs})$. Following the scheme in Figure 9, several alternatives are in this case available. As shown by Eq. (15), the value of the intermediate target does not enter the wealth definition, once the investment has occurred. The wealth will accordingly fall, depending on the realized scenario and possibly become negative.

We assume that under the negative scenario the investment will be financed by a mortgage at the future prevailing interest rate. The borrowing will then reduce the wealth

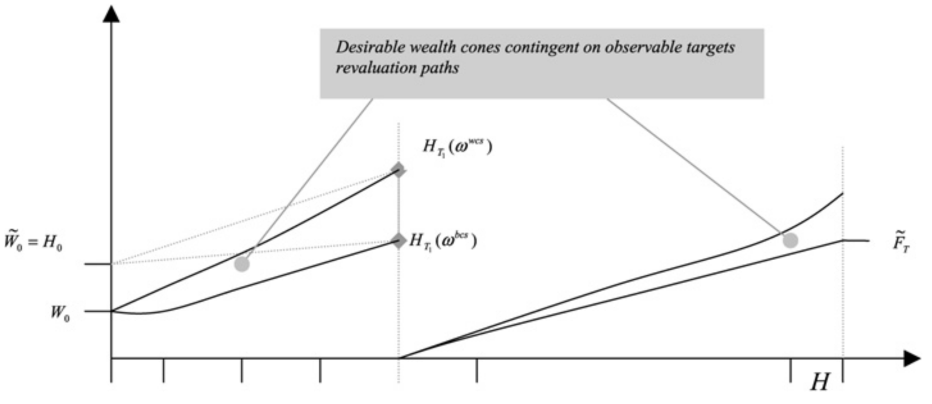


Fig. 11. Wealth trajectories resulting form an optimal contingent plan.

performance until the planning horizon. Alternatively under a positive scenario no borrowing will be needed at time T_1: indeed the individual wealth remains positive and the subsequent market evolution may lead to a financial surplus at the horizon.

The solution of the introduced DSP problem, with the given objective function, will however induce wealth trajectories consistent with the evolution of targets' values. The sequence of recourse decisions—see Figure 1—under the established constraints will at each stage adjust the wealth trajectory and maximize the likelihood of achieving the different targets.

The stochastic optimization model in Section 4.1 is formulated to minimize the difference between current and target wealth—evaluated at each stage—and maximize the individual expected utility at the horizon.

The required mortgage may not be necessary in this case, as shown in Figure 11, and the deficit at the horizon on the pension fund terminal value is expected to disappear.

We assume in Figure 11 that at targets horizon, contingent on the realized revaluation scenario, one of the possible wealth scenario will occur. It is also assumed that the wealth—net of consumption and other life expenses—at the intermediate target and at the horizon will be matching exactly the value of the target that will be funded through disinvestments making borrowing not needed. In reality, in view of the decision plan non-anticipativity constraint, this event will hardly occur, inducing a limited mortgage necessary.

The range of variation of the financial scenarios and the required investment policy, will severely depend on the adopted model of uncertainty as it is clear at this point. The definition of an appropriate financial model for long-term financial planning and wealth management applications has attracted increasing interest by financial intermediaries and fund managers in the last decade or so, inducing to a certain extent an almost independent stream of developments (Wilkie, 1995; Dempster et al., 2002; Pflug, 2004). Used in many cases for financial forecasting and strategic decision-making without requiring an explicit solution of a stochastic optimization problem.

As such, the model of uncertainty is in general carefully considered in real applications and its validation requires extended out-of-sample back testing (Consigli, 2004; Villaverde, 2003; Consiglio, Cocco and Zenios, 2004). On this point, in particular, we present here below to conclude this section an instructive exchange of information, specifically on the economic and financial model described here above, between the officer of a (large) financial (sponsoring) Institution (FinInst) and the software house' (SoftH) coordinator of a dedicated individual ALM development. The received feedback gives some interesting indications on the requirements put forward by the financial institutions for the scenario tree generation, at the time in Europe the Euro became the legal currency.

Scenario Generation in practice

FinInst

I believe that the model should be simplified.
As the world has changed (e.g., due to the introduction of the euro) one can in any case have little confidence that such historical parameter values will continue to be valid.

SoftH

Unreliable statistical estimates for expected returns and volatilities arise from the results of the estimation and the back-testing procedure. Constant coefficient models have been considered as a possible choice. As far as expected returns are concerned, the general model is mean reverting. It is not exactly a time-varying expected return model, in the classical sense of long-term predictability. It does only adjust to a constant long-term value that can be chosen by the user.
We do allow in any case the user to check the validity of the estimates by back-testing the simulated paths against the realized history for any period and at any time.

FinInst

GARCH is typically used to model short-term perturbations to the long-term volatility. It is probably of less value where the investment time horizon is long. (In these circumstances one might reasonably expect the central limit theorem to hold.)

SoftH

Volatility time varying models are not necessarily used for modeling short-term perturbations only. Indeed the most successful long-term scenario generation models available not only in the academia but also in the industry do include time-varying volatility parameters. It's plenty of references: Towers Perrin implemented the Cap:Link system developed by Mulvey (University of Princeton), Prof. Wilkie's actuarial model for pension funds which is now marketed by Bacon and Woodrow (1998), the TY model also includes random volatility for long-term inflation series (Yakoubov, Teeger and Duval, 1999). The scenario generator developed in Cambridge by Consigli and Dempster (1998a) as well.

FinInst

The model is being used to provide a strategic benchmark. It should be left to the investment manager to decide whether particular assets are offering above or below average expected returns at any given time.

SoftH

The generator allows for the inclusion of subjective views.

FinInst

The more complex the model the harder it becomes to understand the implications of the parameter values for the long-term dynamics of the model, and the harder it becomes to decide whether the parameters are appropriate.

SoftH

The development includes a back-testing module that can be applied to both scenario analysis and optimal strategies validation. In the first case simulated market dynamics are back-tested out-of-sample in order to check if historical scenarios were appropriately captured.

FinInst

Unlike the expected return on equities, the current shape of the yield curve is known. This contains some information on the expected future return on cash. The model's parameters and initial values should be chosen so that the initial value of cash equals its actual current value, and its implied long-term average value is broadly consistent with long bond yields.

SoftH

Long-term scenario generators such as the one implemented in the ALM development are typically constructed with two types of initial conditions: current values or average values over a recent period. The first is implemented in the project. The capital market model does not rely on any sort of implied information such as the one that can be extracted from the current yield curve (based on a type of rational expectations hypothesis). We do not believe that expected future rates can be extracted from forward information. This would be another approach.

FinInst

As the expected return on cash will vary over time, we suggest that the other assets be modeled as cash plus a constant risk premium plus lognormal noise. This preserves a simple structure for the non-cash assets while excluding the possibility of the expected return on cash ever exceeding that on risky assets.

SoftH

This is an interesting and possibly desirable but different modeling approach. We are estimating and generating scenarios from a set of simultaneous equations. To a certain extent, however, the suggested approach seems to lead to a single state variable model: the short-term interest rate and all the rest built on it. This is not a new idea and has been implemented in some long-term models for bond portfolios (one important reference is Zenios, 1993). They never lead to very good results we believe. Better to concentrate on long-term correlations. In general risk premiums are time-dependent: we leave the estimates and the back-testing to speak.

FinInst

The long-term expected asset risk premiums should be imposed by us and not fitted directly from history (otherwise the asset class that performed best over the fitting period will tend to dominate). Asset volatilities could probably be fitted against historical data, but I suggest that a longer period than 12 years should be used. At a minimum it would be good to take the data back to the October 1987 crash. Probably the DM should be used as a proxy for the euro pre-1999 (though this is a matter of debate).

SoftH

This is already allowed: the long-term expected returns can be input by the user. He can use whatever model or approach he wants to force the long-term dynamics of the system. We believe that an approach based on risk premiums as pointed out above is not very appropriate.

FinInst

The future behavior of Eurozone inflation is a matter of great uncertainty and history is of limited relevance. We are tempted to suggest that it simply be modeled as cash plus a (negative) constant plus noise. This assumes that Eurozone monetary policy typically responds quickly and one-to-one to changes in inflation. The implied long-term average rate of inflation should be broadly consistent with the gap between euro conventional and inflation-linked bond yields.

SoftH

The model implemented for the inflation is explicit. We will check what comes out from the proposed approach: again we do believe however that market implied information (yields differentials) is not very helpful in long-term financial planning, due to the remarkable changes of market expectations over time (it is plenty of examples in this respect).

FinInst
The model should include the facility to link future cash flows to inflation and perhaps to wage inflation. This would allow for example the inclusion of future cash inflows equal to a fixed percentage of wages. In addition it would be good to incorporate the impact of pure uncertainty on the value of the cash flows. This could be modeled as an annual perturbation (uncorrelated with the other sources of uncertainty in the model) to the expected future value of the cash flows. This has the effect of making future cash flows progressively more uncertain the further into the future they lie.

SoftH
This again has to do with the structure of the model: wages and net cash flows are an input given by the investor and not an output. We are already extending the model to incorporate wage inflation uncertainty.

FinInst
In any set of parameters provided to users it would be good also to provide one or more alternative sets with e.g. the expected return on equities 1–2% p.a. lower than our central long-term estimate, and asset volatilities half as high again. This will help users to gain a sense of the sensitivity of the results to the underlying assumptions, and emphasizes that our estimates are only estimates.

SoftH
We will try to enrich the user control and understanding of the proposed solutions: it is in the philosophy of the project to allow sensitivity analysis and make clear which inputs are relevant and how they do affect the problem solution. We do not consider however that to leave the full set of parameters open to users input would be a good idea.

FinInst
Finally, how long does the model take to run the simulations and generate optimal portfolios on a standard PC?

SoftH
Around 5 minutes with 10 years and 5,000 scenarios.

In Section 5.2 we follow up this discussion by considering the required modules for the implementation of the scenario generator within the decision support system required for the solution of the ALM problem.

Different economic and financial scenarios over an extended time period—in the above case study we consider a 25-year horizon—will result in possible problem infeasibility: the persistence of prolonged periods of recession have typically dramatic consequences on financial markets performances and individual incomes. Scenario generators should account for such events and the resulting optimal strategies will be contingent on those scenarios as well (Ziemba, 2003).

5. Problem generation and solution

Instantiations of the type of model described in the previous section can lead to very large nonlinear stochastic optimization problems involving several thousands of scenarios and millions of variables and constraints (see Consigli and Dempster, 1998a, 1998b). The steps from the problem definition to its solution are common to dynamic stochastic programming applications in general (see Ziemba, 2003; also Wallace and Ziemba, 2005) and are:

- The description of the decision problem in mathematical form.
- The definition of the asset and liability universe and the associated return models.

- The identification of a comprehensive dependence structure for the financial and real estate returns resulting into a strategic economic and capital market model.
- The statistical estimation of the model coefficients.
- The generation of a scenario data process consistent with the stochastic program conditional time and scenario structure.
- The generation of the associated mathematical program—deterministic and then stochastic—in standard MPS and SMPS format (Birge et al., 1986; Gassmann, 2005).
- Its solution either by decomposition or handling the full deterministic equivalent problem.
- The solution validation and the analysis of the output.

When dealing with the solution of a nonlinear large-scale problem, furthermore, only recently efficient algorithms have become available (see Section 5.2). We have also underlined the importance of policy optimization as an intermediate step before the solution of the stochastic programming problem.

Several decisions support systems have been prototyped in this area and are growing in number (see Wallace and Ziemba, 2005). They typically imply an initial set of developments carried out in the Academia and then at one point they become efficient enough to turn into commercial systems and be marketed. The Stochastics™ system of Cambridge System Associates (Dempster, Scott and Thompson, 2002, 2005) incorporates a set of modules developed over the years to deal with the problem definition, the generation of a stochastic programming problems—relying on the most updated version of Stochgen—and its solution as linear or nonlinear convex SP problem. The system has been recently enriched with a number of useful graphical interfaces becoming relatively user friendly. Similarly SPInE—stochastic programming integrated environment (Messina and Mitra, 1997; Valente, Mitra and Poojari, 2005)—of OptiRisk Systems (CARISMA of Brunel University directed by G. Mitra) combines an interactive modeling system based on the AMPL language with a module for scenario generation and the set of FortMP solvers for mixed integer linear and quadratic programming problems. Lattice Financial Ltd (see Section 6) has been probably the first case of years of scientific developments (particularly by J.M. Mulvey, Princeton University) turning into a software for personal financial planning integrating the above mentioned methodological steps in an efficient way.

ORS and Prometeia Calcolo in Europe did also independently start to increase in partnership with the academic world (Consiglio, Cocco and Zenios, 2004) their effort to develop stochastic decision tools.

Large software houses such as ILOG-CPLEX, IBM with SP-OSL and consulting centres as ORTEC International in the US have significantly increase their supply of numerical procedures and solution algorithm particularly efficient for the generation and solution of stochastic programming problems.

Mathematical programming modeling languages such as AMPL (Fourer, Gay and Kernighan, 2002) and OPL (ILOG 2000) have been originally developed to handle deterministic, though dynamic, optimization problems but can be used recursively to generate stochastic programs according to the scenario structure and the set of conditional probabilities at hands.

The key elements of the stochastic programming framework are the problem mathematical description, the scenario generator, the solver and the output generator. Several schemes have been proposed in the recent past (Dempster et al., 2002; Pirbhai, Mitra and Kyriakis, 2003; Pflug, 2004 for scenario generation) to fulfill these tasks—Input–Output flows—and we present their technical implications in the following sections.

5.1. An SP implementation of individual AL management

The building blocks of a state-of-the-art decision system for AL management in general and specifically for individual ALM. The ALM model can be regarded as an object at the center a computer-based system with a set of inputs—the coefficients of the model, the targets, the scenario vector data process and a set of scenario probabilities—and a set of outputs—the optimal strategies, the targets success rate, post-optimality and sensitivity analysis and so on.

When dealing with a financial service such as personal financial planning or private banking, provided by an intermediary with a distributed geographical network a web based application provides an efficient mean to handle the complexity of the core system, built according to the previous discussion—see Figure 2—and manage the required input/output information from/to the user.

We can briefly outline a possible way to handle the different modules.

The scenario model and the estimation of its coefficients do depend typically on an extended Data Base of financial and economic series collected by Data Providers worldwide. The collection of the relevant data for the running application is typically performed with a DB connector.

Similarly the output must reach individuals anywhere in the intermediary network of branches. This requires an integration layer and a communication (Com) interface from the core system running locally or at the Intermediary Headquarter to the network branches.

One possible implementation of such an integrated system is described in Figure 12.

The core building blocks of the system do include:

- A model generator (AMPL for instance or Magic in Figure 12),
- A scenario generator,
- A set of optimization subroutines for statistical estimation and problem solution (LAMPS in Figure 12, a proprietary solver of ORS), plus
- A strategy generator. This is essential to convert a complicated stochastic programming output with several hundreds of optimal decisions along a scenario tree into something immediately useful for the financial advisor. See also the discussion in Appendix B.

These modules are generally handled through a Central Model Manager and several back-testing facilities do complete the system to ease the validation of the statistical model and the optimal strategies.

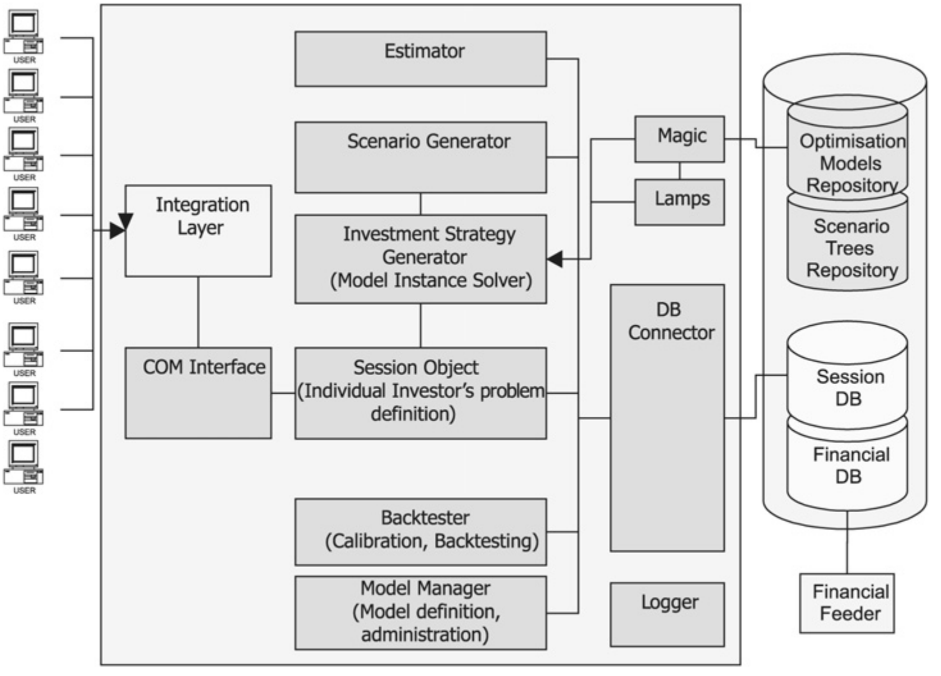


Fig. 12. ORS integrated DSP system for ALM applications.

The central column in Figure 12 shows the elements that are typically included in any prototype ALM system together with the Input Data Warehouse. The output network can vary depending on the application and the decided marketing policy (web-based, Internet, or others, see Consiglio, Cocco and Zenios, 2004).

Speed of communication is critical for distributed systems as well as for systems running locally to support demos and consulting commitments.

On the right the Optimization module is required both for the solution of the stochastic programming problem and as mentioned in the phase of coefficient estimation for the economic and capital market model. The two problems are typically different and based on different solution algorithms as discussed previously. The solution methods for the dynamic stochastic programming problem are considered in Section 5.3. The estimation of the coefficients of the scenario model relies on different methods depending on the adopted model of uncertainty. In some cases (see Barro and Canestrelli, 2004) also for stochastic programming applications, the financial scenarios following a common practice in the risk management community have been generated avoiding any analytical setting and, thus, estimation process and relying on historical scenarios.

5.2. Scenario generator: modules, input and output

We have described in Section 4.2 a characterizing part of the analytical framework underlying the generation of a relevant set of economic and financial scenarios to be considered for the solution of the individual ALM problem. We consider next a possible approach to scenario generation, built on the described model outline. The optimal decision plan is contingent on the introduced scenario paths for the extended set of coefficients introduced in the model instance Section 4.1.

Unlike the stochastic model of uncertainty, that even if rather general, does address the issue of modeling the random process underlying the individual ALM problem, the discussion that follows is rather general and can be applied to a range of financial models not necessarily associated with ALM problems. Several other approaches can be adopted and have been recently put forward in the stochastic programming community (Dupačová, Consigli and Wallace, 2000; Dempster, Scott and Thompson, 2002; Pflug, 2004; Ziemba, 2003).

The phase of scenario generation relies on the following inputs:

- The problem's planning horizon and time periods and its conditional structure, reflected on the associated nodal partition matrix,
- The set of equations and coefficient estimates of the economic and capital market model capturing the problem core uncertainty,
- The set of functions for the remaining part of the ALM problem coefficients that can be computed from the core random variables.

And generates as output the set of scenario-dependent coefficients—discrete realizations in time and space—relevant for the generation of a stochastic instance of the individual ALM problem. Each scenario is then assigned a probability of occurrence.

Broadly speaking the following steps are modular parts of the scenario generator:

- The parameter estimator for the econometric model,
- A simulator of the model state variables,
- A module to convert possibly continuous sample paths in a discrete scenario structure consistent with the optimization problem conditional structure,
- A module to assign scenario probabilities and generate as output a three-dimensional array to be interfaced with the model generator.

The scenario module for financial markets simulation can allow a great deal of flexibility to the user, with the possibility to include subjective scenarios, as well as introducing unrealistic hypotheses on correlations. Figure 13 shows a possible loop across the mentioned modules with the required set of interfaces.

The parameter estimator does typically include a set of procedures largely relying on global optimizers for maximum likelihood (ML) estimation of non-convex probability densities.

Estimation procedures alternative to ML, such as the mentioned generalized or simulated method of moments, do also rely heavily on the sequential solution of optimization techniques for the financial model estimation.

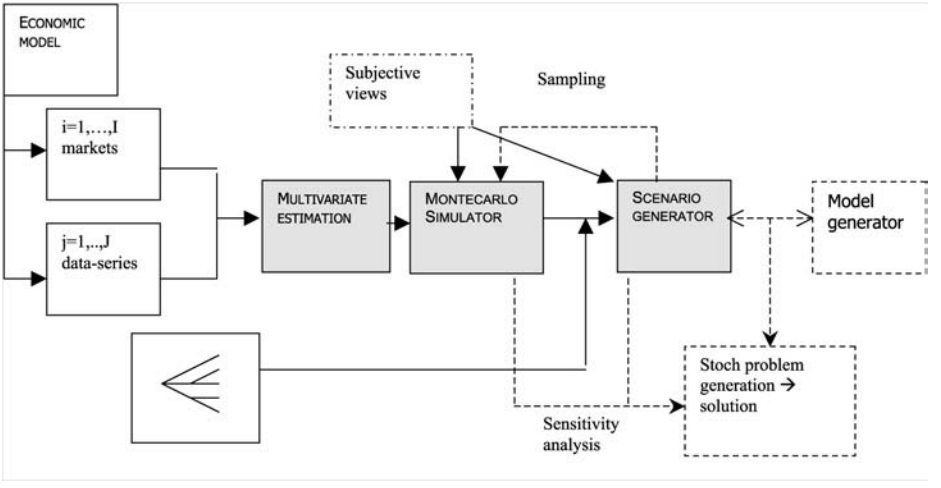


Fig. 13. Scenario generation.

The definition of a representative set of discrete scenarios out of the stochastic system evolution is widely discussed in the literature and several approaches can at this stage be considered (Dupačová, Consigli and Wallace, 2000). Pflug (2004) and Römisch (2004) do consider the issue of generating a minimal set of scenarios whose probability distribution minimizes in some sense the distance from a theoretical continuous distribution. While the construction of the set of interdependencies between economic and financial variables is primarily motivated by theoretical reasons and the attempt to capture relevant stylized evidences of financial markets behaviors, the scenario tree generation is instead essentially concerned with the identification of a limited number of scenarios that while maintaining the statistical properties of the underlying stochastic model, do maintain computational tractability and provide a limited but relevant description of the future uncertainty. Those scenarios that can have a severe impact on the problem financial feasibility do deserve in this sense specific care (Ziemba, 2003).

The core scenario module in Figure 13 can be analyzed in more detail considering the algorithm actually adopted for the definition of the stochastic equations and the array structure that can accommodate the generator output. This is considered in Figure 14.

Here above the I/O info of the scenario loop refers to the generation of a total number of S conditional scenarios over T periods for M coefficients required for the decision problem implementation. A distinction is made above between the random process ω, which captures the problem uncertainty and includes the financial and economic processes in Figure 8 and the coefficient process χ of model coefficients that is determined as a function of the former and is tailored to a specific model instance such as the one in Section 4.1. We have mentioned above the theoretical implications of the distinction on the statistical properties of the problem.

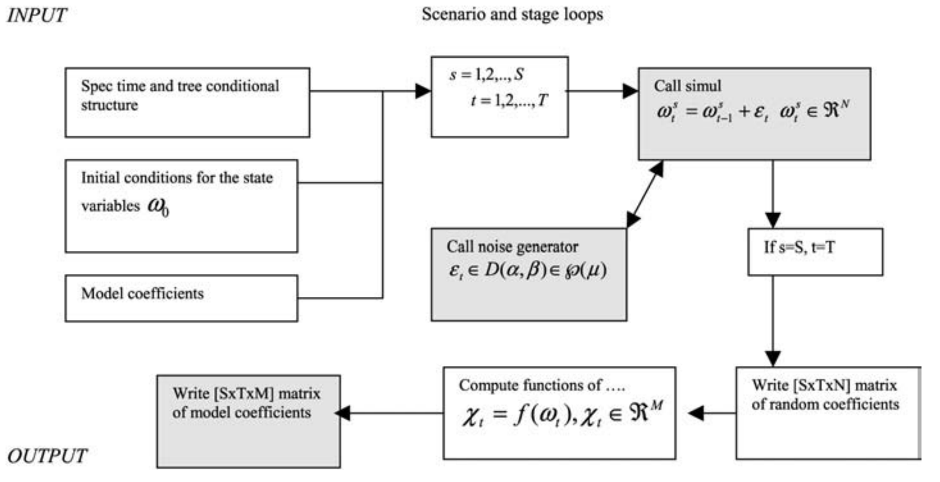


Fig. 14. Scenario generation modules.

As indicated in Figure 14, different distributions can be adopted for the residuals of the problem and this is consistent with the introduction of return distributions including, for instance, extreme scenarios against which the ALM problem solution needs to be stress-tested.

To a certain extent the dimension and statistical properties of both ω and χ processes do depend on the desired depth of the financial model and decision problem complexity: the inclusion for instance of scenario-dependent targets forces the extension of the capital market model to include the equations for the real estate. The assignment of a set of equal probabilities to the generated coefficient scenarios $\chi(s)$ is intended not to alter the probability frequency distribution introduced in the previous loop for the data process ω. In general the two set of probabilities do determine the resulting frequency distribution for χ, relevant for the stochastic program solution (Pflug, 2004). It is specifically at this stage that the inclusion of distributional constraints will affect the solution of the ALM problem and to which the distributional properties of the stochastic program must be referred to.

The introduction of the set of coefficients and probabilities in the stochastic programming problem does follow the definition of the deterministic multistage problem for the first scenario and its extension to the required number of scenarios. This extension is to be consistent with the required non-anticipativity condition discussed in Sections 2 and 4 and has heavy computational implications mentioned in the following Section 5.3. See also Dempster, Scott and Thompson (2002) for an efficient scheme on stochastic program generation.

Figure 15 displays a representative scenario tree consistent with an underlying stochastic optimization stage structure (generated by Dempster, CFR Cambridge, Dempster, Scott and Thompson, 2002). The dimension of the tree and its branching

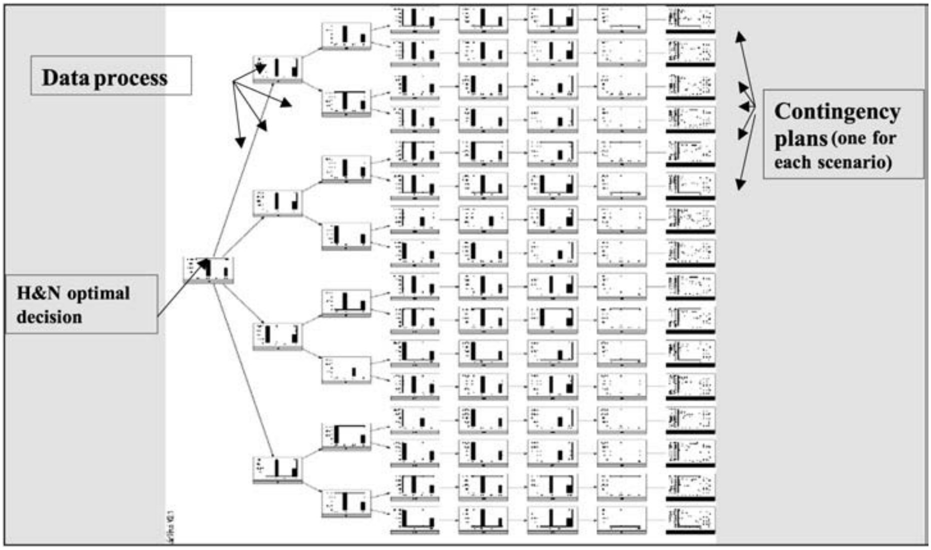


Fig. 15. Example of scenario tree for stochastic programming problems.

structure for real problems will in general be far more extended and rich. Every set of nodal problems along the scenario tree will result in an optimal contingency plan as reported on the right of Figure 15.

The H&N decision at the root node is the only decision feasible with respect to any future scenario and optimal for the associated optimization problem. The identification of the optimal decision sequence is preceded by the solution of the nonlinear optimization problem according to the methods described in the following section.

5.3. Solution methods for linearly constrained nonlinear convex programs

This section is motivated by the relevance we attach to the nonlinear representation of the individual risk preferences within ALM problems of the type considered in this chapter. Furthermore the ability to solve general nonlinear constrained optimization problems within large-scale stochastic programming framework has been achieved only recently and motivates the following remarks. We follow up from Section 3.2, where several problem formulations where associated with possible solution approaches.

A variety of solution methods have been used for several years to handle large-scale DSP problems (Consigli and Dempster, 1998a, 1998b) directly or by decomposition. For linear and quadratic problems the simplex method or interior point methods as well as nested Benders decomposition provide efficient solution algorithms, while for general nonlinear problems both nested Benders and sequential quadratic programming methods can be adopted (Dempster et al., 2002; Villaverde, 2003). Direct methods based on simplex or interior point solvers can be readily applied to so-called deterministic equivalent problems that thus have to be generated in the first place (Consigli and Dempster, 1998a, 1998b).

The inclusion of a general nonlinear convex objective function, other than the quadratic, as typically the case in ALM developments for individuals, implies either a piece wise linear approximation to handle the objective function curvature or the adoption of optimization methods relying on some form of decomposition.

Barro and Canestrelli (2005) present a solution approach based on a double decomposition scheme, with respect to time and scenarios, that combines the progressive hedging algorithm (Rockafellar and Wets, 1991) and a discrete version of Pontryagin maximum principle (Sethi and Thompson, 2000). The method appears able to handle arbitrary concave utility functions and can be naturally adopted to benchmark equivalent stochastic control problems (Purcal, 2003). An extended set of computational evidences is presented for large-scale nonlinear optimization methods by Benson, Shanno and Vanderbei (2002): they benchmark LOQO (Vanderbei and Shanno, 1999) algorithm with KNITRO (Byrd, Hribar and Nocedal, 1999) and SNOPT (Gill, Murray and Saunders, 2002), three popular solution methods for large nonlinear problems. The problem dimensions they consider, however, are not yet comparable with large stochastic programming problems with thousands of scenarios. Andersen's MOSEK (Andersen and Andersen, 2000; Hilli et al., 2004) implements a geometric homogeneous method (Andersen and Ye, 1998) able to solve very rapidly large scale conic quadratic and convex nonlinear problems such as those arising for individual financial planning.

The objective function in Eq. (14) can be regarded as a special case of a two-objective function in which the decision maker maximizes a given reward function—here represented by the utility of terminal wealth—and minimizes an associated risk measure—here represented by deviations between target and available wealth. The solution of the problem is consistent with the maximization of the difference between a risk and a reward functions in this dynamic setting (Consigli, 2004). The same remark holds for problem (4) in Section 3. In presence of Markovian decision problems resulting in block diagonal technology matrices (see Consigli and Dempster, 1998a), as typically the case in ALM applications, convex nonlinear solvers can be conveniently used in decomposition schemes such as nested Benders decomposition (Gassmann, 1990; Consigli and Dempster, 1998a, 1998a) in which the set of nested optimization problem is solved iteratively through a sequence of two-stage stochastic dynamic programming recursions with optimality conditions passed forward and feasibility conditions backward along the tree. This type of approach is implemented for instance by the Personal Financial Planner described in Appendix B where a quadratic solver is used in presence of decomposable quadratic programs while nonlinear convex problems are approximated by piecewise linear approximation.

More recently (see Ziemba, 2003), due to the increasing turbulence of financial markets, constraints on extreme losses have been required also within long-term decision systems. The inclusion of extreme scenarios on dynamic portfolio models has an impact on both the problem feasibility region and the likelihood to reach a target, particularly if an extreme event is generated in the neighborhood of a target. Stress-testing and post-optimality analysis are in this case essential and have been considered under alternative solution algorithms (see Dupačová, Bertocchi and Moriggia, 1998; Jobst and Zenios, 2002; Billio and Casarin, 2003; MacLean, Consigli and Ziemba, 2004).

The adoption of tail risk measures such as the Conditional Value-at-Risk (Rockafellar and Uryasev, 2000; Consigli, 2004) in individual AL management tools is very recent and has become possible thanks to the linear representation of this risk measure which unlike the Value-at-Risk is coherent and convex (Consigli, 2004). The inclusion of the CVaR constraint is still consistent with the above mentioned stage decomposition scheme in presence of Markovian problems and convex nonlinear objective functions.

6. Conclusions and directions

In this chapter we have analyzed in detail the key features and theoretical implications of ALM developments dedicated to individual investors. The state-of-the-art in this area has been clarified with respect to available industrial developments and directions suggested by scientific progresses in the fields of financial analysis, operations research and decision theory. The generality of this type of dynamic decision problem in the field of stochastic programming developments for AL management has also been emphasized.

We have presented in detail in Section 4 a benchmark ALM model for individual investors explicitly considering the need to model endogenously the individual targets. The mathematical model is at the centre of a set of building blocks integrated in a decision framework that links horizontally the data input associated with a DB system and a financial data provider to the output analysis and financial strategies generated to track the desired targets over time.

More generally asset–liability management systems for individuals have been shown to present a degree of complexity essentially due to challenges posed by:

- The modeling of individual decision problems constrained by real life commitments, long horizon and pending uncertainty on the future;
- The required generation of medium and long-term economic and financial scenarios to assess the likelihood of specific targets given current and foreseeable individual resources;
- The desirable inclusion of time dependent risk aversion coefficients and period utilities in the objective function of the stochastic optimization problem, consistently with the changing nature of human preferences;
- The need that the optimal decision process generated by the system is shared and fully understood by the individual: policy optimization as well as simulation analysis can in this respect help and enrich the output info of the ALM system;
- The need for computationally strong developments able to cope with market requirements and supporting in some cases several types of financial applications relying on a number of I/O interfaces;

- The intersection of developments from the areas of financial engineering, statistical analysis, optimizations theory, operations research to be integrated in a computationally efficient way.

The trend towards the production of increasingly efficient systems with a high degree of modularity able to accommodate the continuously evolving requirements of the wealth management industry is expected to continue in the future and we expect the software industry to increase its integration with the financial sector and its reliance on the stochastic programming community to enhance its development capabilities.

Acknowledgements

This chapter has benefited from several forms of co-operation. I would like to thank Pierluigi Riva of ORS for his help on the description of the ORS personal financial planner and the interest he showed on the work. A special thank goes to Monica Defend and her colleagues at Pioneer Investments for their help on the collection of European and Italian data on the mutual fund industry and again for their general interest on the development. A particular thank to William Ziemba for its accurate revision and editing of a first draft of the chapter, to John Mulvey and Stavros Zenios for their important suggestions on a first draft of the chapter and to both Editors, Stavros Zenios and William Ziemba, for their invitation to write the chapter.

Appendix A. INDIV ALM mathematical outline

This appendix presents the mathematical specification of the individual ALM benchmark model introduced in Section 4.1.

We use the following sets specification.

Sets:

$t = 0, 1, \ldots, T$ time points. The horizon T is the expected retirement age,

$i = 1, \ldots, I$ investment classes for the allocation strategy,

$s = 1, \ldots, S$ scenarios, sample paths from root node to horizon.

The decision variables are related to investment and borrowing decisions, excess or deficit in the personal cash account and a provision coefficient. At each stage essentially the individual must decide, given his family current consumption rate, his financial portfolio that together with the mortgage will finance the country house, and how much to allocate in the pension fund for retirement purposes.

Decision variables:

$x_{i,t}(s)$ amount held of asset i at time t in scenario s,

$x^+_{i,t}(s)$ amount bought of asset i at time t in scenario s,

$x^-_{i,t}(s)$ amount sold of asset i at time t in scenario s,

$y_t(s)$ debt at time t in scenario s,
$y_t^+(s)$ borrowing at time t in scenario s,
$y_t^-(s)$ debt reduction at time t in scenario s,
$z_t^+(s)$ cash surplus at time t in scenario s,
$z_t^-(s)$ cash deficit at time t in scenario s,
$z_t(s)$ cash balance at time t in scenario s, $z_t(s) = z_t^+(s) - z_t^-(s)$,
ψ provision coefficient.

The model variables provide a minimal set of equations needed for an appropriate problem instantiation. The target wealth is associated with the individual goals introduced in the chapter case study: buy the country house around the 10 year horizon and retire at the end of the working life with a sufficient income. The latter coincides with the stochastic program horizon.

Model variables:
$W_t(s)$ personal wealth at time t in scenario s,
$C_t(s)$ consumption at time t in scenario s,
$I_t(s)$ individual income at time t in scenario s,
$F_t(s)$ pension fund value at time t in scenario s,
$\widetilde{W}_t(s)$ target wealth at time t in scenario s,
$\widetilde{F}_T(s)$ target pension fund value at retirement in scenario s.

The above decision variables determine the wealth evolution described in Figure 6 following the dynamics of an extended set of coefficients that depend on the economic and financial variables described in Section 3.3.2.

Coefficients:
$r_{it}(s)$ price return of asset i at time t in scenario s,
$\zeta_t^{\pm}(s)$ interest rate on cash account at time t in scenario s,
$b_t(s)$ 10 year borrowing rate at time t in scenario s,
$\xi_{it}(s)$ exchange rate for asset i at time t in scenario s,
$\rho_{it}(s)$ cash return of asset i at time t in scenario s,
$\nu_{it}(s)$ value of asset i at time t in scenario s,
τ_i transaction cost for buying–selling asset i,
β_i maximum current wealth fraction for investments in asset class i,
γ maximum current wealth fraction available for borrowing,
c consumption rate,
$\pi_t(s)$ price inflation at time t in scenario s,
$g_t(s)$ output growth at time t in scenario s,
$\varphi_t(s)$ pension fund return at time t in scenario s,
$h_t(s)$ real estate return at time t in scenario s,
$\delta_{t,T}(s)$ discount factor at time t for planning horizon T in scenario s.

The generic individual ALM problem is defined as a nonlinear dynamic stochastic program (NLDSP). Two types of objective functions can be considered for a consistent inclusion of the individual risk attitudes. We refer to Section 4.1 for a detailed description of the equations of the model.

A.1. INDIV *ALM model set-up*

Objective functions: Max expected utility of terminal wealth minus target wealth cumulative deviations before the horizon

$$\max_{x \in X} E\left[\frac{1}{\lambda}(W_T)^{\lambda} - \sum_{t<T} \gamma_t \left|W_t - \widetilde{W}_t\right|\right].$$

Subject to:

Generic ALM constraints

First stage wealth and allocation decision equation (no borrowing at root node)

$$\begin{aligned}
&x_{i,0}(s) = \vec{x}_i + x^+_{i,0}(s) - x^-_{i,0}(s),\\
&W_0(s) = X_0(s) + I_0 + z_0,\\
&X_0(s) = \sum_{i \in I} x_{i,0}(s), \quad i = 1, 2, \ldots, I,\ s = 1, 2, \ldots, S,\\
&y_0(s) = 0.
\end{aligned}$$

Asset and liability inventory balance constraints

$$\begin{aligned}
&x_{i,t}(s) = x_{i,t-1}(s)\bigl(1 + r_{i,t}(s)\bigr) + x^+_{i,t}(s) - x^-_{i,t}(s),\\
&X_t(s) = \sum_{i \in I} x_{i,t}(s), \quad t = 1, 2, \ldots, T,\ i = 1, 2, \ldots, I,\ s = 1, 2, \ldots, S,\\
&y_t(s) = y_{t-1}(s) + y^+_t(s) - y^-_t(s).
\end{aligned}$$

Individual cash balance constraint

$$\begin{aligned}
&I_t(s) - C_t(s) - \psi_t(s) + z^+_{t-1}(s)\bigl(1 + \zeta^+_t(s)\bigr) - z^-_{t-1}(s)\bigl(1 + \zeta^-_t(s)\bigr)\\
&\quad + z^+_t(s) - z^-_t(s) + y^+_t(s) - y_{t-1}(s)b_t(s) - y^-_t(s)\\
&\quad + \sum_i \xi_{i,t}(s)\bigl[x_{i,t-1}(s)\rho_{i,t}(s) + \upsilon_{i,t}(s)(1 - \tau_i)x^-_{i,t}(s)\\
&\quad - \upsilon_{i,t}(s)(1 + \tau_i)x^+_{i,t}(s)\bigr] = 0, \quad t = 1, \ldots, T,\ s = 1, \ldots, S.
\end{aligned}$$

No horizon decisions

$$\begin{aligned}
&x^+_{i,T}(s) = 0, \qquad x^-_{i,T}(s) = 0, \quad i = 1, \ldots, I,\\
&y^+_T(s) = 0, \qquad y^-_T(s) = 0, \quad s = 1, \ldots, S.
\end{aligned}$$

Bounds and non-negativity

$$0 \leqslant x_{i,t}(s)v_{i,t}(s) \leqslant \beta_i W_{t-1}(s),$$
$$0 \leqslant y_t(s) \leqslant \gamma W_{t-1}(s), \quad t = 1, \ldots, T,\ i = 1, \ldots, I,\ s = 1, \ldots, S,$$
$$0 \leqslant z_t^{+/-}(s).$$

Specific individual ALM constraints

Individual income dynamics and consumption

$$I_t(s) = I_{t-1}(s)\bigl(1 + \pi_t(s) + g_t(s)\bigr), \quad I_0(s) = \hat{I},\ t = 1, \ldots, T,\ s = 1, \ldots, S,$$
$$C_t(s) = cI_t(s), \quad t = 1, \ldots, T,\ s = 1, \ldots, S.$$

Pension fund contributions and pension fund dynamics

$$\psi_t(s) = \psi I_t(s), \quad t = 1, \ldots, T,\ s = 1, \ldots, S,$$
$$F_t(s) = F_{t-1}(s)\bigl(1 + \varphi_t(s)\bigr) + \psi_t(s), \quad F_0 = 0,\ t = 1, \ldots, T,\ s = 1, \ldots, S.$$

Target wealth dynamics

$$H_t(s) = H_{t-1}(s)\bigl(1 + h_t(s)\bigr), \quad H_0 = \widehat{H},\ t = 1, \ldots, T,\ s = 1, \ldots, S,$$
$$\widetilde{F}_{t,T}(s) = \widetilde{F}_T \delta_{t,T}(s), \quad t = 1, \ldots, T,\ s = 1, \ldots, S,$$
$$\widetilde{W}_t(s) = H_t(s) + \widetilde{F}_{t,T}(s), \quad t = 1, \ldots, T,\ s = 1, \ldots, S.$$

Wealth dynamics

$$\begin{aligned} W_t(s) = {} & W_{t-1}(s) + z_t(s) + I_t(s) - C_t(s) - y_t(s) \\ & + \bigl[X_t(s) - X_{t-1}(s)\bigr] + \bigl[F_t(s) - F_{t-1}(s)\bigr], \\ & t = 1, \ldots, T,\ s = 1, \ldots, S. \end{aligned}$$

Non-negativity

$$C_t(s) \geqslant 0, \quad \psi_t(s) \geqslant 0, \quad I_t(s) \geqslant 0, \quad t = 1, \ldots, T,\ s = 1, \ldots, S.$$

Appendix B. Operational Research Systems *Personal Financial Planner*™

As mentioned in Section 2, an individual ALM system is able to support a variety of applications from personal financial planning to private banking and strategic asset allocation problems. In this appendix we present some of the features of a decision support system explicitly designed for a large Italian bank to support the class of applications discussed in Section 2 and that are growing in number and attracting as mentioned significant efforts by several software houses worldwide.

The tool can be analyzed with respect to different phases in the definition of the individual planning problem:

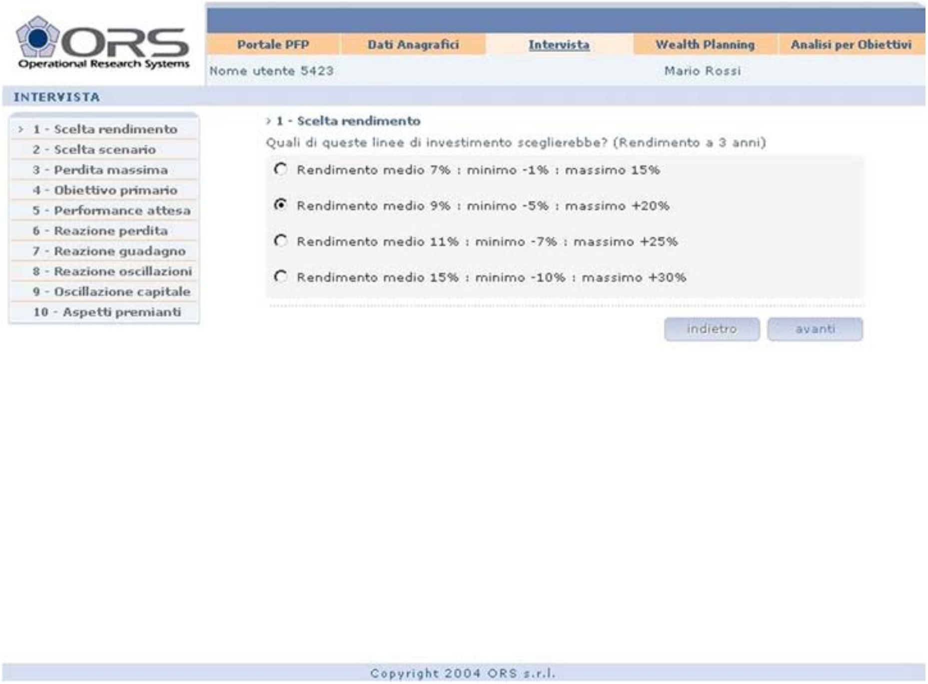


Fig. 16. The investor interview interface.

- First comes the interview of the investor to assess her-his degree of risk aversion—Sections 3.2.2 and 4.3.
- Then the determination of the required degree of assistance and the targets of the investor—Sections 3.2.1 and 4.
- The identification of the asset universe on which the financial strategy must rely upon—Section 4.
- The definition of an optimal model portfolio for the specified targets given the generated scenarios—solution of the problem outlined in Section 4.1.
- The comparison between the results of the model versus the real portfolio.

For each such phase a set of graphical interfaces is available and featured here below.

The interview aims at determining the degree of risk aversion relying on an extended set of information aimed at determining the individual psychology with respect to financial decision making in general and more specifically her–his preference order on sequence of lotteries, as described in Section 4.3.

Individual attitude to risk is analyzed considering specifically:

- The definition of a target return for the financial portfolio—determined taking also into account the tolerance with respect to a range of variation, its average and volatility over a certain horizon,

Fig. 17. Degree of individual assistance.

- Foreseeable reactions to losses or gains over time,
- An expected rate of increase of the family income and individual salary,
- A desirable targets time distribution.

Other elements.

A possible interface supporting the interview phase is shown in Figure 16 where Mr. Rossi is a potential client of the financial intermediary.

The output of the interview can be qualitative, resulting in a clustering procedure where the output is transformed into a number and the degree of risk aversion determined accordingly: this is assessed as 0.3 for Mr. Rossi (see Figure 17).

The individual risk preference is also qualified by the current financial condition of the family and the presented set of targets. These are wishes that the individual intends to pursue and their ambition depends on time and the initial endowment.

The initial endowment and set of targets are specified in Figure 17 where the degree of assistance is also specified with respect to financial and insurance management—very much in the philosophy of private banking applications, knowledge of real-estate advisory needs and current and desired future investments.

The three targets presented by Mr Rossi to the financial advisor include:

- A boat—the *sciallino*—needed to replace the boat currently owned,

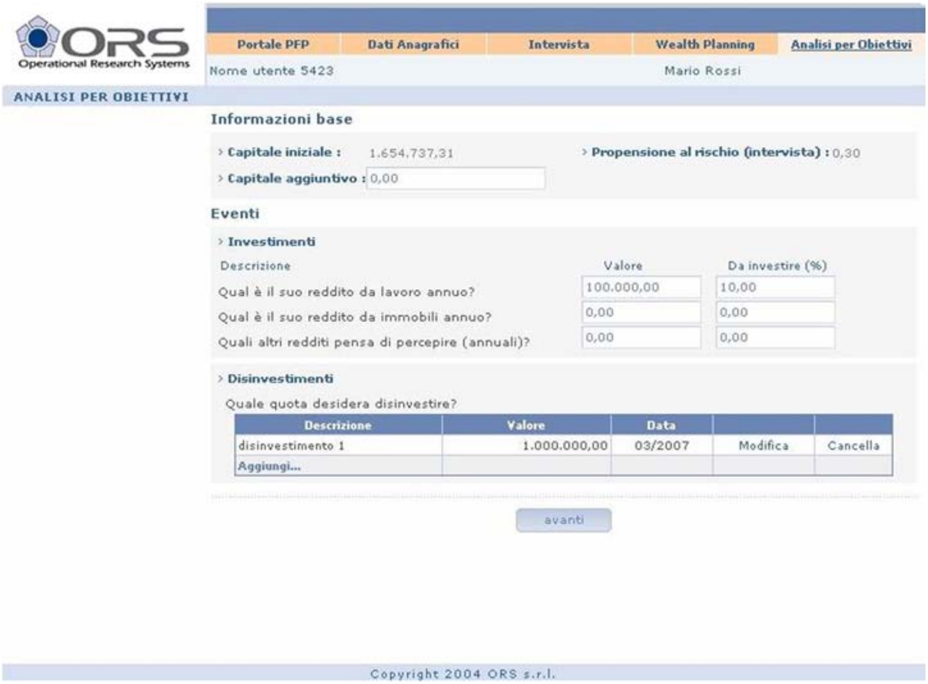


Fig. 18. Mr. Rossi financial condition and targets.

- A holiday house on the Mediterranean coast,
- A chalet in the mountains for winter holidays.

With an initial endowment of roughly 1.65 mln Euros and an annual income around 100 000 Euros these are reasonable objectives to be met on the next few years. Mr. Rossi also declares to be willing to fill his family desires disinvesting an amount up to 1 mln Euros in order to avoid any borrowing or short selling, as shown in Figure 18.

These assumptions have an immediate impact on the instantiation of the Indiv_ALM model described in Section 4.1 and Appendix A. Following the scheme in Figure 9: no borrowing must be allowed among the possible decisions and the targets can be financed by the current portfolio holdings. The assets in the current portfolio must thus be included in the asset universe to capture the possible evolution of the current portfolio to be compared with a theoretically superior portfolio—in the sense of portfolio dominance in the given risk-return space.

The definition of the asset universe must thus consider the asset classes currently owned by the investor plus other asset classes consistently with the individual risk aversion and target returns. Here below, given the initial portfolio and a selected set of asset classes, an optimal portfolio is identified.

Fig. 19. Portfolio analysis.

The asset classes considered in this application include cash and:

- Government and corporate domestic bonds in the Euro currency and government international bonds,
- Equity Europe, Equity Italy, North American equity and Emerging markets plus East Asian equity markets.

The model portfolio will include specified benchmarks for each such market. It is clear that the decision problem solution does require an underlying global capital market model such as the one outlined in Section 3.3 (see also Dempster, Scott and Thompson, 2002; Ziemba, 2003, etc.).

Bond market returns will depend on the evolutions of the corresponding yield curves, for which a specific model must be identified (see Vasicek, 1977; Zenios, 1995; Mulvey, 1996; Abaffy et al., 2000; Ziemba, 2003). In addition corporate bonds may require the inclusion of a credit spread model to account for possible defaults (here the literature is extensive, see Ziemba, 2003).

Furthermore equity returns will also require scenario generators with risk premiums components and dedicated models following the discussion in Section 3.3. The management of an international portfolio does also require the inclusion of a relevant set of exchange rates.

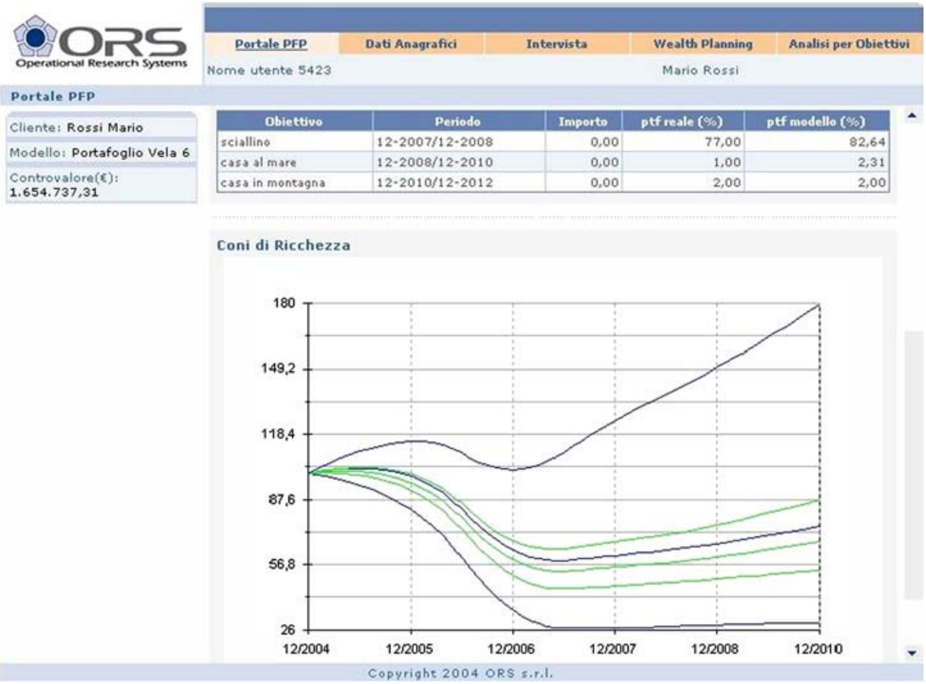


Fig. 20. Mr. Rossi stochastic wealth behavior.

The optimal model portfolio does depend on the initial portfolio and possible constraints on transaction costs and other type of constraints. The optimal portfolio value can be plotted as a function of the generated set of financial scenarios given an initial portfolio currently owned by Mr. Rossi.

Observe that the model portfolio can differ significantly from the current portfolio and it is explicitly function of the generated scenarios and the included targets. Here below in correspondence of the end of year 2006 the expected wealth decreases due to the boat target set around that period.

The above optimal portfolio is a H&N portfolio and as time evolves it may be required a subsequent run of the model and an adjustment of the optimal strategy.

The time distribution of the targets indicated in Figure 19 is handled by considering that until the first is not met the other also will not be met. So under a number of scenarios all three targets will be met, then only the first two or only the first or none, with corresponding probabilities defined by the number of scenarios over the total number of scenarios generated.

In Figure 20 the different lines do correspond to wealth scenarios on the optimal portfolio from the best to the worst scenario and the associated mean evolution over the next six years.

Following Mr. Rossi wealth cones, the required targets achievements can be studied within an easily accessible and understandable framework.

References

Abaffy, J., Bertocchi, M., Dupačova, J., Moriggia, V., 2000. On generating scenarios for bond portfolios. Bulletin of the Czech Economic Society 11, 3–27.

Andersen, E.D., Andersen, K.D., 2000. The MOSEK interior point optimizer for linear programming: An implementation of the homogeneous algorithm. In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (Eds.), High Performance Optimization. Kluwer, Academic Publishers, pp. 197–232.

Andersen, E.D., Ye, Y., 1998. A computational study of the homogeneous algorithm for large-scale convex optimization. Computational Optimization and Applications 10, 243–269.

Bank of Italy, 2004. Financial Accounts (Suppl.). In: Statistical Bulletin: Monetary and Financial Indicators, vol. 35. XIV ed., June. www.bancaditalia.it.

Barro, D., Canestrelli, E., 2004. Tracking error: A multistage portfolio model, Working paper of the Department of Applied Mathematics University Ca Foscari of Venice (IT). Corresponding author canestre@unive.it. Forthcoming in: Pflug, G. (Ed.), Volume of Proceedings of EUMOptFin—Cyprus 2004.

Barro, D., Canestrelli, E., 2005. Dynamic portfolio optimization: Time decomposition using the Maximum Principle with a scenario approach. European J. Oper. Res. 163 (1), 217–229.

Benson, H.Y. Shanno, D. Vanderbei, R.J., 2002. A comparative study of large-scale nonlinear optimization algorithms, ORFE Technical report ORFE-01-04. Princeton University, NJ.

Berger, A.J., Mulvey, J.M., 1998. The Home Account Advisor™: Asset and liability management for individual investors. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 634–665.

Berndt, E.K., Hall, B.H., Hall, R.E., Hausman, J.A., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3–4, 653–665.

Billio, M., Casarin, R., 2003. Extreme returns in a shortfall risk framework. In: Atti della giornata di studio Metodi Numerici per la Finanza. Applied Mathematics Department, University Ca' Foscari, 30 May 2003.

Birge, J.R., Louveaux, F., 1997. Introduction to Stochastic Programming. Springer-Verlag, New York.

Birge, J.R., Dempster, M.A.H., Gassmann, H.I., Gunn, E.A., King, A.J., Wallace, S.W., 1986. A standard input format for multiperiod stochastic linear programs. Committee on Algorithms Newsletter 17, 1–20.

Bollerslev, T., Engle, R., Nelson, D., 1994. ARCH models. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. IV. North-Holland, Amsterdam, pp. 2959–3038.

Borsa Italiana, 2004. BI*t* Notes, 2nd Report on Italian Shareholding, 12. www.borsaitaliana.it.

Brennan, M.J., Schwartz, E.S., 1998. The use of Treasury Bills in strategic asset allocation programs. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 205–230.

Byrd, R.H., Hribar, M.E., Nocedal, J., 1999. An interior point algorithm for large scale nonlinear programming. SIAM Journal on Optimization 9 (4), 877–900.

Campbell, J., Lo, A., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, US.

Cariño, D.R., Myers, D.H., Ziemba, W.T., 1998. Concepts, technical issues, and uses of the Russell–Yasuda Kasai financial planning model. Operations Research 46, 450–462.

Cariño, D.R., Kent, T., Myers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., Ziemba, W.T., 1994. The Russell–Yasuda Kasai model: An asset–liability model for a Japanese Insurance company using multistage stochastic programming. Interfaces 24, 29–49.

Chen, Z., Consigli, G., Dempster, M.A.H., Hicks-Pedron, N., 1997. Towards sequential sampling algorithms for dynamic portfolio management. In: Zopounidis, K. (Ed.), New Operational Tools for the Management of Financial Risk. Kluwer Academic Publisher, pp. 197–211.

Consigli, G., 2004. Estimation of tail risk and portfolio optimization with respect to extreme measures. In: Szegö, G. (Ed.), New Risk Measures for the 21st Century. John Wiley and Sons Publ., pp. 365–401.

Consigli, G., Dempster, M.A.H., 1998a. Dynamic stochastic programming for asset–liability management. Annals of Operations Research 81, 131–162.

Consigli, G., Dempster, M.A.H., 1998b. The CALM stochastic programming model for dynamic asset–liability management. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 464–500.

Consiglio, A., Cocco, F., Zenios, S.A., 2002. Scenario optimization asset and liability modelling for individual investors, Working Paper 02-07. HERMES Center of Excellence on Computational Finance & Economics, School of Economics and Management, University of Cyprus, Nicosia, CY.

Consiglio, A., Cocco, F., Zenios, S.A., 2004. www.Personal_Asset_Allocation. Interfaces 344 (4) (2004) 287–302, July–August.

Consiglio, A., Cocco, F., Zenios, S.A., 2007. The PROMETEIA Model for Managing Insurance Policies with Guarantees. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 2. (Ch. 15).

Dahl, H., Meeraus, A., Zenios, S.A., 1993. Some financial optimization models: II. Financial engineering. In: Zenios, S.A. (Ed.), Financial Optimization. Cambridge University Press, UK, pp. 37–71.

Dempster, M.A.H., 1988. On stochastic programming: II. Dynamic problems under risk. Stochastics 25, 3–59.

Dempster, M.A.H., 1993. CALM: A stochastic MIP model, Technical Report. Department of Mathematics, University of Essex, UK.

Dempster, M.A.H., Scott, G.E., Thompson, G.W.P., 2002. Stochastic modelling and optimization using STOCHASTICS™, Research Paper Centre for Financial Research. Judge Institute of Management, University of Cambridge, Cambridge, UK.

Dempster, M.A.H., Scott, G.E., Thompson, G.W.P., 2005. Stochastic modelling and optimization using STOCHASTICS™. In: Wallace, S.W., Ziemba, W.T. (Eds.), Applications of Stochastic Programming. In: MPS–SIAM Series in Optimization. Mathematical Programming Society, pp. 131–150.

Dempster, M.A.H., Germano, M., Medova, E.A., Villaverde, M., 2002. Global asset–liability management. Presented to the Institute of Actuaries (U.K.), 25 November 2002. Judge Institute of Management Studies Research Series. www.jims.cam.ac.uk.

Dupačová, J., 1995. Multistage stochastic programs: The state-of-the art and selected bibliography. Kybernetica 31, 151–174.

Dupačová, J., Bertocchi, M., Moriggia, V., 1998. Postoptimality for scenario based financial planning models with an application to bond portfolio management. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, pp. 263–285.

Dupačová, J., Consigli, G., Wallace, S.W., 2000. Scenarios for multistage stochastic programs. Annals of Operations Research 100, 25–53.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous GARCH. Econometric Theory 11, 122–150.

Ermolieva, T., MacKellar, L., Westlund, A., 1999. Robustness to stochastic shocks of alternative old-age pension arrangements: Macroeconomic stability, Interim IIASA report. www.iiasa.ac.at.

Fleming, J.P., Stein, J.L., 2004. Stochastic optimal control in international finance and debt. Journal of Banking and Finance 28, 979–996.

Fourer, R., Gay, D.M., Kernighan, B.W., 2002. AMPL: A Modeling Language for Mathematical Programming. Duxbury Press, Brooxs, Cole Pbl.

Gassmann, H.I., 1990. MSLiP: A computer code for the multi-stage stochastic programming problem. Mathematical Programming 47, 407–423.

Gassmann, H.I., 2005. The SMPS format for stochastic linear programs. In: Wallace, S.W., Ziemba, W.T. (Eds.), Applications of Stochastic Programming. In: MPS–SIAM Series in Optimization. Mathematical Programming Society, pp. 9–20.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2002. The Innovest Austrian Pension Fund financial planning model InnoALM, Working paper. University of British Columbia, CA.

Gill, P.E., Murray, W., Saunders, M.A., 2002. SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM Journal of Optimisation 12, 979–1006.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, 85–118.

Guiso, L., Haliassos, M., Japelli, T. (Eds.), 2002. MIT Press, Cambridge, MA.

Hakansson, N.H., 1974. Convergence to isoelastic utility and policy in multiperiod portfolio choice. Journal of Financial Economics 1, 201–2024.

Hilli, P., Koivu, M., Pennanen, T., Ranne, A., 2004. A Stochastic Programming Model for Asset Liability Management of a Finnish Pension Company. Kluwer, pp. 1–26. Downloadable from www.hkkk.fi.

Hodrick, R., Vassilou, M., 2002. Do we need multi-country models to explain exchange rate and interest rate and bond return dynamics? Journal of Economic Dynamics and Control 26, 1275–1299.

Høyland, K., Wallace, S.W., 2001. Generating scenario trees for multistage decision problems. Management Science 47, 295–307.

Jobst, N.J., Zenios, S.A., 2002. On the simulation of portfolios of interest rate and credit risk sensitive securities, Research paper. Hermes Center of Excellence on Computational finance & Economics, University of Cyprus, CY.

Kahneman, D., Riepe, M., 1998. Aspects of investor psychology. Journal of Portfolio Management, 52–64.

Kallberg, J.G., Ziemba, W.T., 1979. On the robustness of the Arrow–Pratt risk aversion index. Economic Letters II, 21–26.

Kallberg, J.G., Ziemba, W.T., 1983. Comparison of alternative utility functions in portfolio selection problems. Management Science 29, 1257–1276.

Karatzas, I., Lehoczky, J.P., Shreve, S.E., 1987. Optimal portfolio and consumption decisions for a 'small investor' on a finite horizon. SIAM Journal on Control and Optimization 25, 1557–1586.

King, A., Korf, L., 2001. Martingale pricing measures in incomplete markets via stochastic programming duality in the dual of L^{∞}. Stochastic Programming E-Print Series. www.speps.info.

Konno, H., Yamazaki, H., 1991. Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market. Management Science 37, 519–531.

Kusy, M.I., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34 (3), 356–376.

MacLean, L., Consigli, G., Ziemba, W.T., 2004. Modeling market returns with a speculative bubble. In: Proceedings August 2004 Conference of the American Statistical Association, pp. 1–6. lmaclean@mgmt.dal.ca or giorgio@giorgioconsigli.it.

MacLean, L.C., Ziemba, W.T., 2006. Capital Growth: Theory and Practice. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. A. Elsevier (Ch. 10).

MacLean, L., Zhao, Y., Ziemba, W.T., 2005. Wealth goals investing. In: Wallace, S.W., Ziemba, W.T. (Eds.), Applications of Stochastic Programming. In: MPS–SIAM Series in Optimization. Mathematical Programming Society, pp. 509–522.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7 (1), 77–92.

Merton, R.B., 1969. Lifetime portfolio selection under uncertainty: The continuous time case. Review of Economic Statistics 51, 247–257.

Messina, E., Mitra, G., 1997. Modelling and analysis of multistage stochastic programming problems: A software environment. European Journal of Operations Research 101, 343–359.

Mulvey, J.M., 1996. Generating scenarios for the Towers Perrin investment system. Interfaces 26, 1–13.

Mulvey, J.M., Gould, G., Morgan, C., 2000. An asset and liability management system for Towers Perrin–Tillinghast. Interfaces 30, 96–114.

Mulvey, J.M., Rosenbaum, D., Shetty, B., 1996. Parameter estimation in stochastic scenario generation systems, Technical Report SOR-96-15. ORFE Department, Princeton University, NJ.

Mulvey, J.M., Pauling, B., Britt, S., Morin, F., 2007. Dynamic Financial Analysis for Multinational Insurance Companies. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 2. (Ch. 12).

Mulvey, J.M., Thorlacius, A.E., 1998. The Towers Perrin global capital market scenario generation system. In: Mulvey, J.M., Ziemba, W.T. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 286–314.

Mulvey, J.M., Ziemba, W.T., 1998. Asset and liability management systems for long-term investors: Discussion of the issues. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK, pp. 3–38.

Ogryczak, W., Ruszczynski, A., 1999. From stochastic dominance to mean-risk models: Semi-deviations as risk measures. European Journal of Operational Research 116, 33–50.

Olsen, B.R., Dacorogna, M., Muller, U., Pictet, O., 1992. Going back to the basics—rethinking market efficiency, Discussion Paper. Olsen & Associates Research Group, RBO 1992-09-07; http://www.olsen.ch/library/.

Pennanen, T., King, A.J., 2004. Arbitrage pricing of American contingent claims in incomplete markets—a convex optimization approach. Stochastic Programming E-Print Series. www.speps.info.

Pflug, G., 2001. Optimal scenario tree generation for multiperiod financial planning. Math. Programming, Ser. B 89, 251–271.

Pflug, G., 2004. Scenario estimation and generation. In: Tutorial Tenth Stochastic Programming Conference. University of Arizona, Tucson. Downloadable from http://tucson.sie.arizona.edu/SPX/pc_tutorials.html.

Pirbhai, M., Mitra, G., Kyriakis, T., 2003. Asset Liability Management using Stochastic Programming, CARISMA Technical Reports CTR/02/03. Downloadable from http://carisma.brunel.ac.uk/workingpapers.asp.

Purcal, S., 2003. A stochastic control model for individual asset–liability management, Working Paper of the School of Actuarial Studies. University of New South Wales, Sidney, AUS.

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. The Journal of Risk 2 (3), 21–41.

Rockafellar, R.T., Wets, R.B.J., 1991. Scenario and policy aggregation in optimization under uncertainty. Mathematics of Operations Research 16, 119–147.

Römisch, W.H., 2004. Scenario modelling for multistage stochastic programs. In: Tenth International Symposium on Stochastic Programming, plenary lecture Oct. 14, 2004. University of Arizona, Tucson. http://tucson.sie.arizona.edu/SPX/.

Rudolf, M., Ziemba, W.T., 2004. Intertemporal surplus management. Journal of Economic Dynamics & Control 28, 975–990.

Sachs, D.J., Larrain, B.F., 1993. Macroeconomics in the Global Economy. Prentice-Hall, US.

Sethi, S.P., Thompson, G.L., 2000. Optimal Control Theory: Applications to Management Science and Economics. Kluwer Academic Publ.

Sharpe, W.F., 1998. Founder of Financial Engines, Inc., Financial Engines Advisors LLC, The Financial Engines Personal Online Advisor service. www.financialengines.com, last update 2003.

Thorp, E.O., 2006. The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. A. Elsevier (Ch. 9).

Valente, P., Mitra, G., 2003. The evolution of WEB based optimization: From ASP to e-service, CARISMA Technical Reports CTR/08/03. Downloadable from http://carisma.brunel.ac.uk/workingpapers.asp.

Valente, P., Mitra, G., Poojari, C.A., 2005. A Stochastic Programming Integrated Environment (SPInE). In: Wallace, S.W., Ziemba, W.T. (Eds.), Applications of Stochastic Programming. In: MPS–SIAM Series in Optimization. Mathematical Programming Society, pp. 109–130.

Vanderbei, R.J., Shanno, D.F., 1999. An interior point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications 13, 231–252.

Vasicek, O.A., 1977. An equilibrium characterisation of the term structure. Journal of Financial Economics 5, 177–188.

Villaverde, M., 2003. Global fund management using stochastic optimization. Judge Institute of Management Studies, University of Cambridge, Research Papers in Management Studies, WP 12/2003.

Vitale, P., 1999. Sterilised Central Bank intervention in the foreign exchange market. Journal of International Economics 49, 245–267.

Yakoubov, Y.H., Teeger, M.H., Duval, D.B., 1999. The TY model: A stochastic investment model for asset and liability management. Presented at the AFIR and ASTIN colloquia in Tokyo. Downloadable from http://www.sias.org.uk/papers/model.pdf.

Wallace, S.W., Ziemba, W.T. (Eds.), 2005. Applications of Stochastic Programming. MPS–SIAM Series in Optimization. Mathematical Programming Society.

Wilkie, A.D., 1995. More on a stochastic asset model for actuarial use. British Actuarial Journal 1, 777–964.

Zenios, S.A., 1993. Financial Optimization. Cambridge University Press, Cambridge, UK.

Zenios, S.A., 1995. Asset/Liability management under uncertainty for fixed-income securities. Annals of Operations Research 59, 7–98.

Zhou, G., 1999. Security factors as linear combinations of economic variables. Journal of Financial Markets 2, 403–432.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset, Liability and Wealth Management. AIMR, Charlottesville, VA.

Ziemba, W.T., 2007. The Russell Yasuda, InnoALM and Related Models. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 2. (Ch. 19).

Ziemba, W.T., Mulvey, J.M. (Eds.), 1998. Worldwide Asset and Liability Modeling. Cambridge University Press, Cambridge, UK.

This page intentionally left blank

Chapter 18

A SCENARIO APPROACH OF ALM

GUUS BOENDER

ORTEC, P.O. Box 4074, 3006 AB, Rotterdam, The Netherlands

Free University Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands
e-mail: gboender@ortec.nl

CEES DERT

ABN AMRO Asset Management, P.O. Box 283, 1000EA, Amsterdam, The Netherlands
e-mail: Cees.Dert@nl.abnamro.com

FRED HEEMSKERK and HENK HOEK

ORTEC, P.O. Box 4074, 3006 AB, Rotterdam, The Netherlands
e-mails: fheemskerk@ortec.nl; hhoek@ortec.nl

Contents

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02018-7

Abstract

ALM is very important in the strategic decision making of liability driven organizations, especially pension funds, insurance companies and banks. In this chapter we treat three components of ALM for pension plans: The ALM decision problem, the ALM-methodology, and ALM in practice.

We describe and demonstrate that the quintessence of the full scale ALM decision problem is that risk budgets of all stakeholders and all available policy instruments are taken into account to accomplish adequate pensions at acceptable cost and risk. In practice this implies that ALM leads to an optimal integral pension-, contribution-, and investment policy, recently frequently referred to as the Pension Deal. Also, since the integral approach implies that the strategic asset allocation is evaluated in terms of the costs, benefits and risks of the beneficiaries, ALM can be seen to encompass what in the investment community recently has been introduced as liability driven investing.

In the ALM-methodology to accomplish this objective a central role is played by scenario analysis, sustained by optimization methods.

The section about ALM in practice is based on three pension plans. For these plans we first describe the role of ALM in sustaining the specification of the risk profile of the stakeholders, and in the identification of the basic integral pension-, contribution-, and investment policy. Next we select three possible policy-measures to demonstrate the increase of efficiency of the basic ALM-policies, i.e., currency hedging, state dependent asset allocation, and alternative equity exposure.

Keywords

asset liability management, derivatives, liability driven investing, optimization, pension deal, pension funds, scenarios, simulation, state dependent asset allocation, strategic asset allocation

JEL classification: C15, C22, C32, C44, C61, C88, E37, E43, E47, G10, G11, G23, G31

1. Introduction

Why is ALM so extremely popular in international pension fund markets, and in newer application areas such as insurance companies, endowment funds and housing corporations?

At pension plans the main risk budgets are provided by the sponsors and the (future) beneficiaries. For these stakeholders larger risk budgets imply lower contribution rates and higher pensions, and vice versa. The stakes are high. For example, Dutch pension assets in 2005 amount to about 600 billion US$, which approximates Dutch GDP. Thus, e.g., 2% more return on the assets equals 2% of GDP per year, which in the Netherlands amounts on average to 5% of national salaries per year. Especially in the UK, the US and Switzerland the pension funds have such a large impact on the national economy as well. Therefore governance, justification, transparency, efficiency and accountability of pension management are getting crucially important. This is further enforced by the national pension regulators, and by the international pension accountants. ALM sustains decision makers in these issues by providing insight in the relevant risk-return relationships, and by identifying and communicating optimal ALM-strategies which yield each risk-provider in the pension deal a maximal benefit in return.

We treat ALM for pension plans from three approach routes: the full scale ALM decision problem, the ALM-methodology, and ALM in practice.

In Section 2 we argue that the quintessence of the ALM decision problem is that risk budgets of all stakeholders and all available policy instruments are taken into account in order to accomplish adequate pensions at acceptable cost and risk. In practice this implies that ALM should lead to an optimal integral pension-, contribution-, and investment policy, recently referred to as a fair pension deal.

In Section 3 we describe the applied ALM-models and ALM-methods. A crucial role is played by scenarios to model uncertainty, and by simulation and optimization methods to sustain the identification of efficient integral ALM-policies.

Section 4 treats ALM in practice. The results in this section are based on data of three pension plans. We first describe the scenarios which are used to simulate and optimize ALM-policies, as well as the various ALM-measures which are applied to evaluate the performance of the ALM-strategies. Given the scenarios we first concentrate on the usefulness of ALM to sustain decision makers in determining the basics of the ALM-policies, i.e., the specification of the risk profile of the stakeholders, and the identification of the basic integral pension-, contribution-, and investment policy. Next we select on three possible policy measures to demonstrate the improvement of basic ALM-policies, i.e., currency hedging, state dependent asset allocation and alternative equity exposure. The effects of other measures such as flexible pension schemes, dynamic derivative strategies, matching of interest rate risk, elimination of the “long-only” constraint, and the impact of other alternative investments such as commodities, private equity and hedge funds are the subject of other research papers. This chapter focuses on the integral approach of ALM-problem, and on the use of the available policy measures to accomplish fair and efficient pension deals.

2. Institutional setting and the definition of ALM

Employers and employees negotiate a pension scheme in which employees either earn pension rights (= Defined Benefit scheme) and/or obtain pension contributions (= Defined Contribution scheme). Increasingly, employers and employees agree on a hybrid pension schemes with both DB- and a DC-components. For more information on the development of pension schemes, see Ambachtsheer and Ezra (1998); Davis (1994); Muralidhar (2001); and Modigliani and Muralidhar (2004).

The agreed pension scheme is carried out in the pension plan under the responsibility of the board of trustees. In The Netherlands the board of trustees consists 50% of employers and 50% of employees and retirees. A board of trustees has to take into account the interests and requirements of many pension stakeholders. These concern first of all the sponsor, employees, and beneficiaries (= retired and deferred non-active members of the plan), but also the "indirect" stakeholders consisting of the pension regulators and the pension accountants. We call the employees and beneficiaries the active and non-active members of the plan.

We first consider the interests of the sponsor. On the one hand, sponsors wish to profit by low pension contributions to the plan resulting from high portfolio returns on the pension assets. On the other hand, sponsors set constraints on the extent that pension investment risk and other pension risk drivers are allowed to affect her own Profit & Loss account (P&L) and Balance Sheet (BS) under the rules of IFRS or FAS (= International Financial Reporting Standards and Financial Accounting Standards). The maximum pension risk that a sponsor is willing to bare is referred to as the risk budget provided to the pension plan. The rationale of this risk budget is that it allows the trustees ceteris paribus to create higher returns on the pension assets, thereby enabling to reward the sponsor's risk budget with lower expected contributions to the plan.

The contribution agreement materializes at which low levels of the funded ratio (ratio of the pension assets and the value of the pension liabilities) the sponsor donates additional contributions, thereby specifying her pension risk budget, and which rebates she gets in return at high levels of the funded ratio. Therefore, the contribution agreement is also referred to as the surplus/deficit agreement, and as the risk sharing agreement between the sponsor and the pension fund.

To illustrate the importance of the specification of the contribution agreement, we introduce the frequently used concept of investment-leverage (IL) of pension plans. The investment leverage is defined as the ratio of the pension assets of the pension plan, and the sum of annual salaries of the active members of the plan. In the Netherlands the average IL is about 4, ranging from 1 for very young plans to 20 for very mature plans. The IL-concept is of enormous practical importance. Consider a plan with an IL of 4, and suppose that the risk budget provided by the sponsor enables the investment managers to increase the investment risk so that the plan earns an additional expected return of 1% per year. If these additional returns would be completely allocated to the benefit of the sponsor, this would imply a reduction of the contributions of 4% of salaries per year. However, the sponsors also have to take into account the accompanying risk. Consider, for example, the case where the funded ratio has dropped to a level of 90%, which must immediately be remedied by the sponsor to a level of 100%. For the plan with an IL of 4, this would amount to a contribution of 40% of salaries. A mature pension fund, with a typical IL of 20, would have to provide a huge contribution equaling 200% of salaries. If we also take into account that deficits in the pension plan are likely to coincide with a bad profitability of the sponsor, then it is evident that the risk budget which the sponsor provides to the plan should be constrained to responsible values. Sponsors should make a responsible trade-off of low expected contributions to the pension fund, and Contribution at Risk. Extremes are the full Defined Benefit schemes where the sponsor bares all the pension risk, and Defined Contribution schemes where the sponsor donates a fixed contribution and all pension risk is carried by the members of the plan. Currently, hybrid schemes emerge with both DB- and DC-components where also the sponsor bares pension risk, but in a limited amount.

The second group of stakeholders whose interests boards of trustees have to take into account are the members of the plan, distinguished in the employees and the beneficiaries. The fund can either carry out a DC-scheme, where all members have the responsibility for their own pension investments (personal asset allocation and risk management), or a DB-scheme where members build up pension rights which are covered by collectively held pension assets. The funded ratio in DB schemes is the extent in which the value of the pension assets covers the value of the value of all the liabilities. Recently Collective DC-schemes emerge, where the sponsor carries no pension risk, and the members collectively invest. The methods in this chapter apply to DB and Collective DC (although, of course, individual DC could be viewed as a pension fund of one person, to which the methods in this chapter then would apply as well).

In many DB plans and in all Collective DC plans members bare risk by renouncing from COLA (= Cost Of Living Allowance) once the funded ratio of the plan deteriorates to too low levels. The risk baring of the employees and beneficiaries can even be taken a significant step further by allowing temporary reductions of the pensions and pension rights. The rationale of these risk sharing COLA-arrangements is that these, analogously to the pension risk budget of the sponsor, ceteris paribus allow the pension plan to carry more investment risk, thereby paying higher pensions enabled by the higher expected portfolio returns.

The COLA-agreement materializes at which low levels of the funded ratio the members abstain from full indexation, and which additional indexation they get in return at high levels of the funded ratio. Of course, if the COLA-agreement would enable the plan to create an additional expected portfolio return of 1% as the result of a higher risk profile of the investments, this could be applied to improve the pensions 1% per year. (Also recall the pension rule of thumb that 1% more return on the pension assets approximately yields 30% higher pensions.) On the other hand, missing COLA of, e.g., 2.5% per year over a period of 20 years would reduce the purchasing power of the pensions with about 40%, so that also for these stakeholders an efficient equilibrium should be accomplished between expected pensions and Pension at Risk.

Boards of trustees shape and manage these objectives of the sponsors and the members by specifying and carrying out an integral contribution-, COLA-, and investment policy, also referred to as ALM-policy (= Asset Liability Management) and as Pension Deal. ALM-projects support trustees and their stakeholders in identifying feasible and efficient ALM-policies.

In addition to negotiating and agreeing a contribution- and COLA-agreement with the sponsors and members, the trustees in specifying an ALM-policy also have to specify the maximum risk that the plan encounters a deficit, also referred to as Surplus at Risk. Clearly Surplus at Risk concerns the investment risk which cannot directly be transferred to the sponsors and the members. In many countries local laws limit Surplus at Risk. For example, in The Netherlands pension funds have to create surpluses such that the probability of deficit (i.e., funded ratio $< 100\%$) is at most 2.5% on a 1-year horizon. However, Surplus at Risk should not only be limited by laws, but also by the phenomenon that it becomes disproportional difficult to recover from increasing deficits (see the box and Table 1).

The nonlinear relationship between funded ratio and required portfolio return

Liability return is defined as the autonomous growth of the liabilities (excluding the growth due to new pension rights of the active members). Given, e.g., a 4% liability return, a funded ratio of 100% would remain at the level of 100% if a portfolio return is achieved of 4%. However, at a funded ratio of 80%, a portfolio return of 5% would have to be achieved to compensate for the growth of the liabilities, increasing to a required portfolio return of 8% in case the funded ratio would drop to a level of 50% (cf. Table 1). Hence, at increasing deficits, the trustees would have to take disproportional larger investment risk to restore the financial health of the plan. At too large deficits, this phenomenon makes complete ruin of the plan not to be averted, unless drastic additional measures are taken with respect to additional funding by the sponsors, or a reduction of the pension rights of the plan members.

Table 1 column 2 depicts the portfolio return $x1$ that is needed to match the growth of 4% of the liabilities in any year given that the fund has a deficit d. Thus $x1$ solves the equation $(1 - d) * x1 = 0.04 \Rightarrow x1 = 0.04/(1 - d)$.

Column 3 depicts the portfolio return $x2$ that in case of a deficit d is needed to eliminate the deficit in a period of 5 years. Thus, this required return not only has to match the growth of the liabilities, but also eliminates the deficit. Thus $x2$ solves the equation: $(1 - d) * (1 + x2)^5 = 1.04^5 :\Rightarrow x2 = 1.04/(1 - d)^{1/5} - 1$.

Thus, ALM first of all concerns determining the maximum allowable risk with respect to the sponsors, members and deficits, and specifying an optimal integral contribution-, indexation-, and investment policy, whose consequences satisfy the risk limits, and which provides optimal pension contributions and pension benefits in return. As mentioned, the stakes are high. In The Netherlands each 1% additional investment return which can be created as a result of the risk sharing pension deals (currently being equal to 1% of GDP) either leads to a reduction of the contributions of 4% of salaries, or cu-

Table 1
Required portfolio returns in case of deficit

Deficit (% liabilities)	Required return to match the growth of the liabilities	Required return to eliminate deficit in 5 years
0%	4%	
10%	4.4%	6.2%
20%	5%	8.7%
30%	5.7%	11.7%
40%	6.7%	15.2%
50%	8%	19.5%
60%	10%	24.9%
70%	13.3%	32.3%
80%	20%	43.5%
90%	40%	64.8%

mulatively to 30% higher pensions. Therefore ALM-projects to accomplish these deals, and to investigate investment classes and investment strategies to optimize the results of the pension deals have become crucially important. We clearly see this trend worldwide.

3. The ALM approach

In practice we approach the ALM-problem with scenario analysis, combined with simulation- and optimization techniques, see Figure 1. The objectives of this approach are twofold:

– Provide quantitative and graphical insight to the ALM-decision makers;
– Identify efficient integral ALM-strategies.

The quintessence of scenario analysis is that the external uncertainties which ALM-decision makers have to take into account, i.e., inflation, interest rates, risk premiums of equity, as well as transitions of the plan members, etc., are modeled by a set of possible plausible future developments, referred to as scenarios. Using a corporate model of the pension plan, ALM-strategies are evaluated by simulating the consequences of the ALM-strategy in each individual scenario. In practice, usually the "learn-and-react" approach is followed. That is, ALM-strategies are evaluated, analyzed, and in an iterative procedure improved, until an efficient and fair ALM-strategy is obtained, recently also referred to as the Pension Deal, which efficiently balances downside risk and upside potential of all stakeholders. In some cases, especially in ALM-projects with a large research component, this learning process is sustained by applying optimization techniques. We use the same approach, and in a large extent the same models, in ALM for insurance companies, housing corporations and banks.

Before we proceed in describing our scenario analysis approach in more detail we observe beforehand that presently this approach is still in a large extent rooted in the

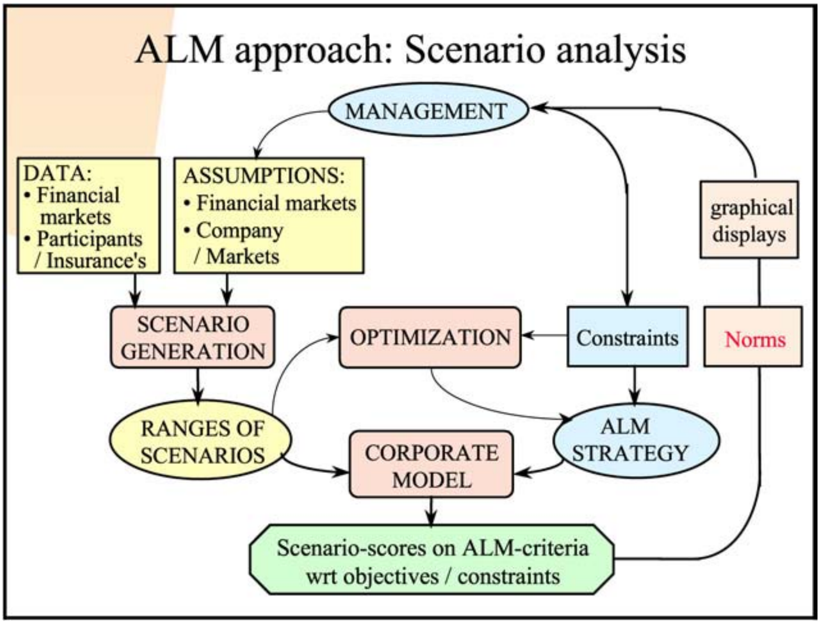


Fig. 1. Scenario approach of ALM.

approach which has been published in the classic ALM-papers by Kingsland (1982) and Winkelvoss (1982). It is interesting and relevant to repeat the following statement in Kingsland (1982):

> "The dynamic behavior of a pension plan is clearly dominated by rules and methodology which are discontinuous and nonlinear functions of its financial condition. The task of developing a closed-form solution to evaluate the potential state of a pension plan following a series of stochastic investment and inflation experiences would be extremely difficult, if not impossible. To date, the only approach that has proven feasible is the application of Monte Carlo Simulation, wherein an investment and inflation scenario is generated by random draws based on the expected probability distribution of year to year investment and inflation behavior. In order to develop an accurate assessment of the range of potential uncertainties, it is necessary to repeat this simulation process by generating dozens or hundreds possible scenarios, consistent with statistical expectations."

Since 1982 science has enormously moved their borders, see the ALM-papers in this two-volume Handbook. This especially concerns approaches for scenario generation, and optimization algorithms. But, as will appear in the remainder of this section, in order to analyze the full scale ALM-problem, taking into account the costs, benefits and risk of all stakeholders, the core of the statement of Kingsland in our practice is still valid. As can be verified from other chapters in this volume, stochastic optimization methods increasingly find their way in ALM-applications as well.

In the sequel of this section we will elaborate on the models of our approach. In Section 3.1 we describe the methodology to generate scenarios of the future economic environment. In Section 3.2 we explain the methodology which is followed to evaluate ALM-strategies with respect to the generated scenarios. In this section we explain how, given a scenario set of the external risk drivers, we simulate what the consequences of an ALM-policy are for the various SaR-, CaR-, and PaR-risk and return measures. Section 3.3 focuses on optimization. Practical cases are reported in Section 4.

3.1. Generating scenarios

The uncertainties that have to be taken into account in ALM-analyses consist of the future economic environment on the one hand, and of the pension liabilities that result from the uncertain development of the current and future active and non-active members on the other. The uncertainties are modeled as a fan of scenarios, and not as a tree, which of course is only responsible since we restrain the use of the scenarios to simulation and non-anticipating optimization. These scenarios are generated in two steps.

The scenarios of the economic environment are generated using a Vector Auto Regressive model (VAR). Denote these scenarios of the vectors of uncertainties as $\{s_{tn} = i_{tn},\ r_{tn},\ c_{tn},\ e_{tn};\ t = 1, \ldots, T,\ n = 1, \ldots, N\}$ (cf. Figure 2), where

i_{tn} = local price- and wage inflation in $\langle t, n\rangle$ (thus, i_{tn} is the value of inflation in node t of scenario n),

r_{tn} = interest rate structures in the distinguished countries in $\langle t, n\rangle$,

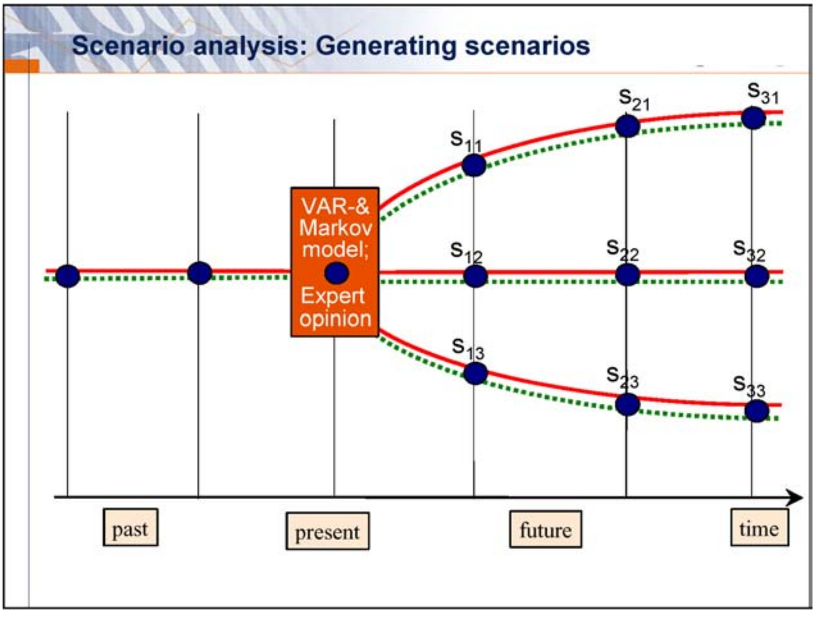


Fig. 2. Scenario generation of the future economic environment, and of the development of the actuarial quantities of the members.

c_{tn} = currency rates between of the distinguished countries in $\langle t, n\rangle$,
e_{tn} = risk premiums of the distinguished asset classes in $\langle t, n\rangle$.

In the next step the relevant actuarial quantities are developed using a Push Pull Markov probability model to determine the status (i.e., active/non-active, age, salary group, ...) of the current and future members in each node $\langle t, n\rangle$ of the scenarios, whereas the pension scheme of the plan is used to determine the corresponding actuarial quantities of the members in $\langle t, n\rangle$. That is, each s_{tn} is extended to $s_{tn} = i_{tn}, r_{tn}, c_{tn}, e_{tn}, a_{tn}, p_{tn}, l_{tn}$ where

a_{tn} = actuarial cost in $\langle t, n\rangle$,
p_{tn} = pension payments in $\langle t, n\rangle$ (excluding COLA),
l_{tn} = value of the liabilities in $\langle t, n\rangle$.

Thus, as a result of the scenario generation process that is further described in this section we know the value of each relevant economic quantity on the one hand, and the resulting actuarial quantities of the plan on the other. In our ALM-projects we typically choose to work with scenario sets of 2500 scenarios with a horizon of 25 years ($T = 25$, $N = 2500$).

The methodology to generate scenarios for the future economic environment is depicted in Figure 3. That is, a time-series model is used to extrapolate the properties of historic time series probabilistically to the future in the form of many scenarios. As illustrated in the picture, in practice some properties of the scenarios are frequently

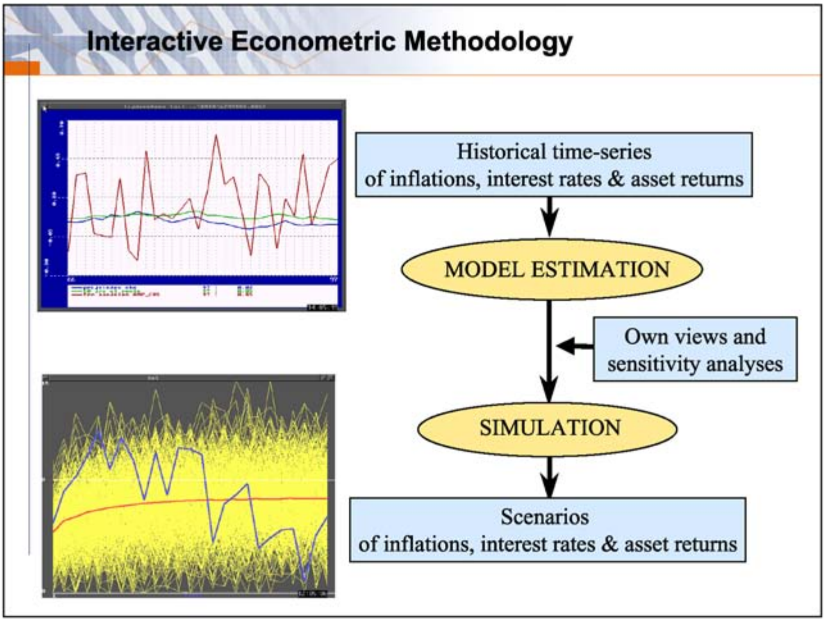


Fig. 3. Probabilistic extrapolation of time series in the form of scenario-sets.

changed by the decision-makers. This typically concerns changes of the expected development of inflation, interest rates and equity risk premiums based on expert opinion of investment advisory committees, and changes of higher moments of the scenarios in the context of sensitivity analyses.

To construct year-frequency scenarios of the future development of the economic time series we apply (log-)Normal Vector AutoRegressive VAR models. In VAR-models, the values of the economic quantities in any year follow a multidimensional (log-)normal probability distribution whose expected values are linear combinations of the realizations of the economic quantities in the previous years:

$$y_{t\cdot} \sim N\big(\mu + \Omega * [y_{t-1} - \mu], \Theta\big).$$

The model assumes stationarity, such that it may be necessary to transform the raw historic data, or that it may be necessary to include dummy variables for periods, such as the oil crisis, which violate stationarity.

The estimation of the model proceeds in two steps. First the sample estimators are determined of the variance- and covariance matrices, denoted as V and W (to preserve stationarity, the denominator of these estimators is the number of sample points, and not the number of sample points minus the number series). In the second step, applying the Yule Walker estimation method Ω and Θ are, respectively,

$$\Omega = V * W^{-1}; \quad \text{and} \quad \Theta = V - W * V^{-1} * W^T.$$

An important characteristic of the VAR-model, which is crucial for the quality of the ALM-analyses which are sustained by the model, is that if the parameters are estimated using the Yule Walker method, then also with limited historic data the scenarios which are generated by the model will asymptotically display the same expected values, standard deviations, and (auto-)correlations as observed in the applied historic data set (cf. Boender and Romeijn (1991), and Steehouwer (2005)). Thus, the scenarios are "consistent with statistical expectations", which is of crucial importance from the point of view of interpretation and decision support (cf. Bunn and Salo (1993)).

The VAR-model is continuously extended and improved. In particular:

- Yield curves

  Due to the need to analyze duration strategies, and due to the growing importance to work with mark-to-market valued liabilities, the model generates yield curves $r(t)$. Of course, in ALM this implies that a yield curve has to be generated in each year t of each scenario s, in such a way that the relevant dynamics and correlations are in accordance with statistical expectations. We accomplish this by using the Nelson Siegel model which is characterized by four parameters, i.e.:

$r(t) = \beta_0 + (\beta_2 + \beta_2) * (1 - \mathrm{e}^{-t/\tau})/(-t/\tau) - \beta_2 \mathrm{e}^{-t/\tau}$	
	Nelson Siegel yield curve
$\lim_{t\to\infty} r(t) = \beta_0$	Long interest rate
$\lim_{t\downarrow 0} r(t) = \beta_0 + \beta_1$	Short interest rate
β_2 and τ	Curvature & scaling parameters

We generate scenarios of Nelson Siegel yield curves by feeding the VAR-model with historic data of the four Nelson Siegel parameters, and, analogously to the other economic quantities, generate the four parameters in each node $\langle t, n \rangle$, which characterizes the yield curve in each $\langle t, n \rangle$. The Yule Walker estimation procedure ensures that the relevant standard deviations and (auto-)correlations coincide with the historic data. Also, the extent in which the curve shifts, tilts and flexes over time is in accordance with past observations.

- Currencies

  In many ALM-projects, hedging currency risk is an important topic. This implies that the scenario generator must be able to produce consistent realizations of currencies and interest rates. We accomplish this by assuming that the covered interest rate parity applies, adding (correlated) error terms which measure the deviations from the interest rate parity in the applied historic data.

- Consistency of simulation and evaluation

  Mainly due to the new international accounting rules and the progress in determining the fair value of the liabilities, scenarios are increasingly also used for the valuation of (embedded) options. Scenarios, which are used for the latter approach, have of course to be arbitrage free. In ALM-projects these two objectives of scenarios are frequently combined. That is, ALM-policies are evaluated by simulating their consequences on a set of scenarios, and within this evaluation process options are evaluated using a scenario-approach, which of course requires these scenarios to be consistent.

- Business cycles

  VAR-models assume stationarity. However, recently theories have revived that the economic environment is not stationary, but moves in compositions of longer term and shorter term business cycles. An important part of our research effort is focused on identifying these cycles, and replicating them in the scenarios (cf. Steehouwer (2005)).

The simulation of the liabilities in each node $\langle t, n \rangle$ is accomplished in three phases. In the first phase a so-called Push Markov model is applied to generate the status of each current active and non-active plan member in each $\langle t, n \rangle$. That is, given characteristics of the members, especially gender, age, salary group and years of service, matrices of transition probabilities are used to simulate future developments of the members with respect to survival, disability, resignation and career. This part of the model is called a Push Markov model since the stochastic behaviors of the members are independent. The survival probabilities are based on public actuarial tables. The probabilities of disablement, resignation and career are based on data in the social annual reports of the sponsors.

The expected future development of the size and structure of the employee force is input of the ALM-model, determined in cooperation of the Human Resource Manager of the company. Given the results of the Push Markov model, in the second phase a so-called Pull Markov model is applied. This model successively fills vacancies by hiring new employees until the number of employees in each category in each $\langle t, n \rangle$ is as much as possible in accordance with specified numbers.

The result of the first two phases of the generation process of the liabilities is that we know the status of each current and future active and non-active member in each node of each scenario. Then the (frequently extensive) pension scheme of the plan is applied to compute all the relevant actuarial quantities in each node $\langle t, n \rangle$, which especially concern the actuarial cost, the pension payments and the value of the pension liabilities. Of special importance is the determination of the pension liabilities in each $\langle t, n \rangle$. These are determined by discounting the future payments of the members in $\langle t, n \rangle$ by the Nelson Siegel interest rate structure in $\langle t, n \rangle$.

3.2. *Simulating the consequences of ALM-policies*

In the previous section we described how the VaR-model, the Push-Pull Markov model, and the pension scheme of the plan are used to generate scenarios of the economic environment and of the corresponding development of the pension liabilities. These scenarios are used to evaluate the risk-return consequences of ALM-policies in terms of Surplus-, Contribution-, and Pension at Risk and at Gain. To illustrate how this is accomplished we for the sake of simplicity assume:

– The ALM-policy consists of a contribution policy x_c, a COLA policy x_i, and asset allocation policy x_a, summarized in the ALM-policy vector x.
– The fund in each node $\langle t, n \rangle$ is characterized by the asset allocation α_{tn}, the funded ratio ϕ_{tn} (= ratio between the value of the pension asset and the value of the pension liabilities), the contribution rate χ_{tn}, and a possible COLA deficit ι_{tn}, summarized in the plans "state of the world" vector θ_{tn}.

Thus x denotes the set of decision rules which the ALM-decision makers analyze and specify in ALM-projects. For example, a simple choice for the investment policy x_a, might be to rebalance to a fixed strategic asset allocation in each node $\langle t, n \rangle$, whatever the state of the world θ_{tn} in $\langle t, n \rangle$. Another typical example is the specification of a COLA policy x_I where COLA is not fully granted in $\langle t, n \rangle$ if the funded ratio ϕ_{tn} in $\langle t, n \rangle$ attains too low values.

Now, applying an ALM-policy x on the initial state θ_0 will, given the vector of scenarios for year 1 $\{s_{1n},\ n = 1, \ldots, N\}$, yield the status θ_{1n} of the fund in year 1 of every scenario $n = 1, \ldots, N$: $\theta_{1n} = x(\theta_0, s_{1n})$, $n = 1, \ldots, N$. Analogously, applying the ALM-policy x to any θ_{tn}, will yield the state in $\theta_{t+1.n}$ (see Figure 4): $\theta_{t+1,n} = x(\theta_{tn}, s_{t+1,n})$, $t = 1, \ldots, T$; $n = 1, \ldots, N$.

Given the state vectors θ that result from an ALM-policy x on a scenario set s, in practice we usually determine the following ALM-scores:

– Expected contribution rate:

  The ALM-quantity "expected contribution rate" is defined as the average value of the observed contribution rates in all combinations $\langle t, n \rangle$.

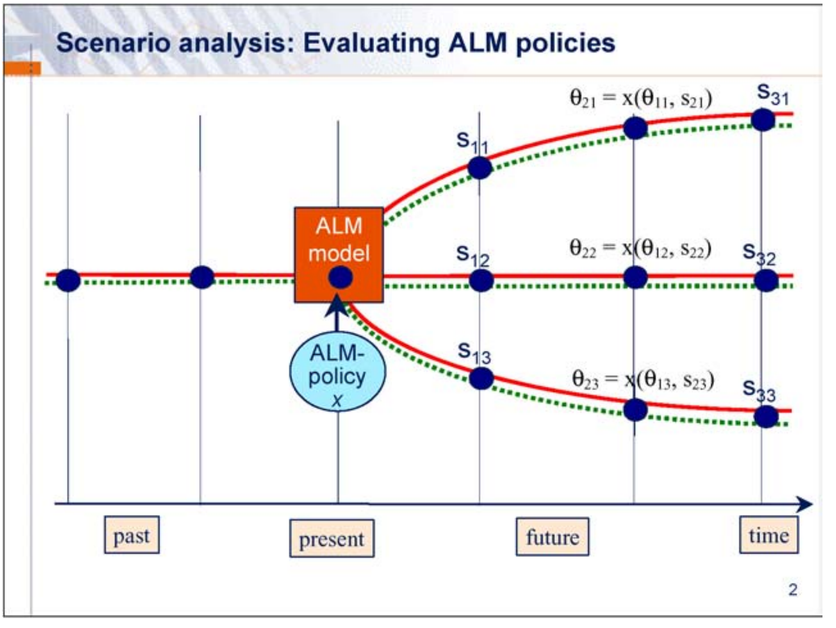


Fig. 4. Evaluation of the “state of the world” of an ALM-policy x in every node $\langle t, n\rangle$.

– Expected funded ratio:

Analogously to the previous definition the expected funded ratio is defined as the average of the observed funded ratios over all combinations $\langle t, n\rangle$.

– Downside deviation of the funded ratio:

Following the definition of downside deviation of portfolio return (cf. Sortino and van der Meer (1991)), we define the downside deviation of the funded ratio in year t as the standard deviation of the funded ratios which are smaller than 100% in year t, i.e.,

$$D_t = \left[\sum_{n=1}^{2500} (\mathrm{MIN}(\varphi_{tn} - 100, 0)^2/2500\right]^{1/2},$$

where φ_{tn} denotes the funded ratio in $\langle t, n\rangle$. The overall downside deviation of the funded ratio is defined as the average of D_t over all $t = 1, \ldots, 25$.

– Probability of underfunding:

This risk measure is defined as the percentage of $\langle t, n\rangle$ combinations where a plan is confronted with underfunding. An important alternative definition of underfunding is the percentage of scenarios in which the pension fund is ever over the horizon $t = 1, \ldots, T$ confronted with underfunding.

- 1% 1 year Surplus at Risk (SaR):
  The 1% 1-year Surplus at Risk is defined the amount of underfunding which occurs with 1% probability. For example, if the 1% 1-year SaR is equal to 10, then with a probability of 1% the funded ratio in any year will be smaller than 90%.
- Probability of incomplete COLA:
  This risk measure is defined as the percentage of $\langle t, n\rangle$ combinations where the pension rights of the non-active members will not be fully compensated for the inflation of prices.
- 5% 3-year Pension at Risk (PaR):
  This risk measure quantifies the risks of incomplete COLA over longer periods. The 5% 3-year PaR is defined as the minimal COLA deficit (in percent points) which will occur with 5% probability over a period of 3 years.
- Probability of a contribution rate larger than the basic actuarial contribution:
  This risk measure is defined as the percentage of $\langle t, n\rangle$ combinations where the contribution rate, due to required additional contributions, is larger than the basic actuarial contribution.
- 5% 3-year Contribution at Risk (CaR):
  Analogously, the 5% 3-year CaR is the minimal amount of contributions (expressed as the percentage of salaries in any year) which the sponsor has to pay with 5% probability over a period of 3 years.

3.3. Optimization

Ideally, ALM-optimization models should take into account all available policy instruments. That is, the decision variables of these models should not only concern the asset allocation, but also the contribution- and COLA policy. Secondly, ALM-optimization models should ideally take into account that a current decision can optimally be adapted in future circumstances.

Important examples of ALM-models who optimally adapt current decisions to future circumstances are the dynamic recourse optimization models in: Berkelaar (2002); Carino et al. (1994); Carino and Ziemba (1998); Dert (1995); Geyer et al. (2005); Kouwenberg (2000); Rudolf and Ziemba (2004); Sharpe and Tint (1990); Siegmann (2003); Wallace and Ziemba (2005); and Ziemba (2003), and in the many references on this research area in this two-volume Handbook.

Due to our complex definition of the ALM-problem, in particular, the measures to postpone COLA if the funded ration drops below critical values, in practice we still use the simple hybrid simulation–optimization method described in Boender (1997), Boender, Oldenkamp and Vos (1997) and Boender and Romeijn (1995). That is, the model randomly generates and evaluates tens of thousands random ALM-policies, and selects the ALM-policies which constitute the efficient frontier with respect to the applied ALM-criteria. In principle, any parameter of an ALM-policy can be a decision variable in this search process. A typical example is the search for a state dependent rule which optimally relates the strategic asset allocation (abbreviated SAA), in particular, the amount of equity, to the value of the funded ratio (cf. Section 4.6). However, such an approach is also able to identify jointly optimal investment-contribution policies, which de facto is the core of integral ALM.

Thus, the hybrid simulation–optimization can in principle search for the optimal value of any parameter of ALM-policies, and it guarantees complete consistency between optimization and simulation (actually the interface parameters of the simulation model constitute the decision variables of the random optimization model). However, such a random approach suffers from the "curse of dimensionality". In particular, if an ALM-policy is sought which determines the optimal decision in every year t of every scenario n, the resulting ALM-problem would typically contain tens of thousands of decision variables, in which case the hybrid random method is powerless. Therefore, the need to extend and improve the recourse models such that an optimal integral ALM-decision with respect to all available policy instruments can be identified in every $\langle t, n\rangle$ strongly remains.

4. ALM: Practical results

This section presents ALM-analyses for three pension plans. The characteristics of these plans are described in Section 4.1. In Section 4.2 we describe the scenarios which we apply to evaluate the consequences of ALM-policies. Section 4.3 focuses on the determination of the risk profile of the plans, and on the consequences of the maximum allowable risk for the strategic asset allocation and the basic ALM-policy. In Section 4.4 we proceed to demonstrate the economic power and efficiency of sharing pension risk between the plan sponsor and the members.

The next sections display the efficiency improvements that can be accomplished by alternative investment classes and by alternative investment strategies. In Section 4.5 we quantify the ALM-effectiveness of the policy to hedge strategic currency risks. Section 4.6 treats state-dependent asset allocation. Section 4.7 concludes by analyzing the ALM-improvements that can be accomplished by investing in an asset class that consists of an efficiently constructed portfolio of derivatives.

4.1. Experimental environment

The three pension plans which we study all carry out an average pay DB pension scheme, where the active members earn an old-age pension right of 1.75% of pensionable salary (= salary minus state pension) per year. Employees start earning pension rights at the age of 25, and retire at the age of 65, such that a complete pension amounts to 70% of average pay. Widowers pensions amount to 70% of old age pensions. In order to make the results as comparable as possible, we assume that each pension fund starts with the same funded ratio of 120%.

Other characteristics are displayed in Table 2. Plan A can be considered as an "average representative" pension plan. Plan B is a young plan, while plan C is mature. We

Table 2
Characteristics of the pension plans

	Plan A	Plan B	Plan C
No. active members	2.000	11.000	2.000
No. non-active members	6.500	6.500	20.000
Total	8.500	17.000	22.000
Liabilities active members	€175 million	€425 million	€175 million
Liabilities non-active members	€475 million	€475 million	€1450 million
Total liabilities	€650 million	€900 million	€1625 Euro
Maturity (= share of non-actives in the pension liabilities)	73%	53%	89%
Liability leverage (= salaries of the active members expressed as fraction of the total liabilities)	0.22	0.69	0.09
Asset leverage (= pension assets expressed as fraction of the salaries of the active members)	5.5	1.7	13.6
Funded ratio (discount rate liabilities 4%)	120%	120%	120%

emphasize the relevance of the liability leverages and the asset leverages of the plans. The liability leverages implicate that the plans A, B and C with 1% of salaries can improve the funded ratio with, respectively, 0.22 percentpoints (pp), 0.69 pp and 0.09 pp. Thus, from the point of view of the power of additional contributions, plan B is relatively strong, and plan C weak. On the other hand, the asset leverages implicate that 1 pp additional portfolio return for the plans A, B and C is, respectively, equal to 5.5, 1.7, and 13.6% of the salaries of the active members. Thus, the plan that profits most from high portfolio returns suffers hardest from low returns, and vice versa. These characteristics in a large extent determine which ALM-policies are optimal for each of the plans A, B and C.

In our ALM-analyses we assume that the trustees of the plans have already decided on the contribution policy and COLA policy. For reasons of comparison we assume that these core elements of the ALM-policies are equal for each of the plans:

- Contribution policy

  All the plans base their contribution policy on the actuarial contribution, which in any year is equal to the actuarial cost of the new pension rights which have been granted to the active members (excluding the cost of the indexation of the pension rights). Dependent on the value of the funded ratio in any year the plans adapt the basic contribution according to the following surplus-deficit agreement:

  – Deficit agreement

    The deficit agreements, which specifies the risk budgets provided by the sponsors and the amount of risk sharing between the plans and the sponsors, run as follows. If the funded ratio drops below 120%, the contribution is increased, taking into account two restrictions. First of all, the contribution rate, expressed as a percentage of the salaries of the active members, is only allowed to increase 2 pp from one year to the next. Furthermore, the contribution rate is limited to a maximum value of 20% of salaries. These constraints imply that the sponsors only carry part of the risk of the pension plan, with obvious consequences for the ALM-policy, in particular, for the strategic asset allocation.

– Surplus agreement

The surplus agreements, which specify the benefits which the sponsors get in return for risk budgets they have provided to the plans, run as follows. If the funded ratio exceeds the level of 140%, the contribution is decreased, taking into account two state dependent restrictions. First of all, analogously to the deficit agreement, the contribution rate is only allowed to decrease 2 pp from one year to the next. Secondly, the contribution rate is not further decreased than to the level of 0% (= contribution holiday). However, these restrictions are relaxed if the funded ratio exceeds the level of 180%. In these circumstances all excess assets are refunded to the sponsor.

– COLA policy

The three plans follow the policy that the pension rights of (only) the non-active members are compensated for the inflation of prices only if the plan is in the situation of a positive surplus (funded ratio $>$ 100%). Indexation deficits are recovered once the funded ratio has recovered to a level above the minimally required 100%. Note that the possibility to postpone COLA of the non-active members defines an important risk budget that the plans can exploit to achieve a higher return on the pension assets. The non-active members are "rewarded" for this risk budget and risk sharing by the intention to use the additional portfolio returns first of all to compensate their pension rights for inflation.

4.2. Economic scenarios[1]

In the experiments we use scenarios for the economic quantities and asset classes which are displayed in Tables 3 and 4. Table 3 displays expected values, standard deviations, autocorrelations and Information Ratios (= expected value/standard deviation) which have been observed in the period which has been used to estimate the VAR-model, as well as the values of these statistics in the 2500 scenarios which have been generated by the estimated VAR-model (horizon 25 years).[2] The reader can verify the approximate equality between the historic and future statistics. Exceptions concern the expected values, the Information Ratios, and the standard deviation of the inflation of wages and prices. This is explained as follows:

[1] With respect to the liability scenarios we assume that the number of employees is constant over time. Transition probabilities with respect to survival, career, disablement, etc. are based on national data and on data provided by the sponsoring companies.

[2] The international derivatives fund refers to the alternative asset class which is analyzed in Section 4.7.

Table 3
Historic and scenario statistics of the applied 2500 economic scenarios

	Scenarios 2001–2025					Historic data 1970–2000				
	Avg.	St. dev.	Auto corr. lag1	Auto corr. lag2	IR	Avg.	St. dev.	Auto corr. lag1	Auto corr. lag2	IR
Price inflation	2.75	1.5	0.56	0.25		4.1	3.0	0.86	0.68	–
Wage inflation	3.75	1.8	0.56	0.46		4.9	4.1	0.89	0.76	–
10 year Euro interest rate	5.75	1.3	0.70	0.48		8.1	1.6	0.75	0.53	
Euro bonds (duration 5)	5.9	6.7	−0.22	0.01						–
Equity MSCI Europe	8.75	20.7	−0.17	0.00	0.41	13.5	21.1	−0.19	−0.02	0.64
Equity MSCI USA unhedged	8.9	21.2	0.09	0.01	0.40	13.7	22.5	0.11	0.03	0.61
Equity MSCI UK unhedged	8.6	23.2	0.02	0.11	0.33	13.1	25.5	−0.04	0.11	0.51
Equity MSCI JP unhedged	11.2	34.6	−0.11	−0.12	0.29	16.5	36.3	−0.14	−0.18	0.45
Equity MSCI Emerging markets	12.9	36.1	−0.37	0.02	0.35	19.9	37.0	−0.36	−0.08	0.54
Equity MSCI world unhedged	8.2	20.6	−0.08	−0.09	0.40	12.6	20.7	−0.09	−0.12	0.61
Equity MSCI USA hedged	8.65	15.7	−0.10	−0.09		14.0[a]	16.1	−0.05	−0.21	–
Equity MSCI UK hedged	8.4	19.3	−0.06	−0.03		14.4	20.1	−0.06	0.01	–
Equity MSCI JP hedged	11.0	28.2	−0.08	−0.08		12.0	29.3	−0.06	−0.13	–
International derivatives fund	16.1	34.5	−0.05	−0.04						–

[a]The figures in this block of the table are the locally observed statistics over the period 1970–2000 in the US, UK and Japan.

- Oil crisis

  In the generated scenarios the standard deviations of the inflations are significantly lower than in the past. This is due to the requirement of the VAR-approach that the data should be stationary, which is violated by the values of inflation in the period 1971–1975. Including a dummy variable for the inflations in this period remedies stationarity, but implies of course that the standard deviations in the scenarios deviate from the past.
- Expected values

  In all ALM-projects expected values of the inflations, and the returns on asset classes are not replicated from the past, but separately decided upon, frequently in cooperation with investment advisory committees of the board of trustees.

  The risk premium of the several equity asset classes is determined as follows. Starting point is the assumption that the risk premium of European equity, defined as the difference between the expected arithmetic return on European equity and the 10 year Euro interest rate, is equal to 300 basis points (bp). This is substantially lower than we have observed in the past, cf. also the book by Dimson, Marsh and Staunton (2002), who estimate the geometric equity risk premium over the last hundred years to be 450 bp. See Mehra (2006) for further estimates. On the other hand, several economists, see, e.g., Arnott and Bernstein (2002), estimate the forward looking risk premium of equity “nowhere near the 5 percent level of the past”, and argue that the

Table 4
Correlations in the applied 2500 economic scenarios

	1	2	3	4	5	6	7	8	9	10	11	12
1. Price inflation												
2. Wage inflation	0.74											
3. Euro bonds (duration 5)	0.11	0.21										
4. Equity MSCI Europe	−0.21	−0.26	0.13									
5. Equity MSCI USA unhedged	0.02	−0.14	0.16	0.80								
6. Equity MSCI UK unhedged	−0.09	−0.19	0.20	0.79	0.69							
7. Equity MSCI JP unhedged	−0.13	−0.11	0.04	0.42	0.32	0.38						
8. Equity MSCI Emerging markets	−0.14	−0.14	0.03	0.56	0.47	0.44	0.78					
9. Equity MSCI world unhedged	−0.10	−0.20	0.17	0.87	0.86	0.76	0.60	0.67				
10. Equity MSCI USA hedged	−0.08	−0.10	0.20	0.62	0.81	0.69	0.24	0.35	0.59			
11. Equity MSCI UK hedged	−0.09	−0.12	0.21	0.74	0.63	0.92	0.43	0.58	0.69	0.58		
12. Equity MSCI JP hedged	−0.17	−0.12	−0.01	0.32	0.25	0.23	0.94	0.78	0.47	0.24	0.34	
13. International derivatives fund	−0.08	−0.15	0.13	0.65	0.65	0.57	0.45	0.50	0.75	0.44	0.52	0.35

forward looking risk premium “may well be near zero, perhaps even negative”. Also taking into account that we extrapolate historic standard deviations unchanged to the future,[3] we position ourselves between the very exuberant past, and these very pessimistic views on the future, and proceed with an arithmetic prospective risk premium of 300 basis points (bp) (cf. also Siegel (2003)). For an investment period of 25 years, an arithmetic risk premium of 300 bp for European equity (standard deviation 20.7%) implies a geometric risk premium of only 100 bp, whereas for local US-equity (standard deviation 15.7%), this implies a geometric risk premium of 200 bp.

Given the specification of the prospective risk premium of one of the equity asset classes we determine the expected risk premium of the other equity asset classes as follows. Keeping the volatility of European equity unchanged, the Information ratio of this asset class deteriorates from 0.64 in the past, to 0.41 in the generated scenarios. Given this adaptation, the risk premiums of the other equity asset classes are reduced in such an extent that the IR of these asset classes reduces in accordance with the reduction of the IR of European equity.

In addition to the “classic” asset classes, the tables also mention the characteristics of the “international derivatives fund (IDF)” of ABN-AMRO. The IDF can be seen as a complex call option with an indefinite life, of which the manager seeks to optimize the return profile by buying and selling options (cf. Dert (2002)). Investment decisions are based on the assumption that equity prices are lognormally distributed with an expected return equal to the risk-free rate +6% and a standard deviation that is equal to the

[3] The papers demonstrating low prospective risk premiums of equity never seem to address the consequences of their analysis on the volatilities. Clearly from the point of view of asset allocation low prospective risk premiums of equity would be considerably less problematic (or even not problematic) if the volatilities decline accordingly.

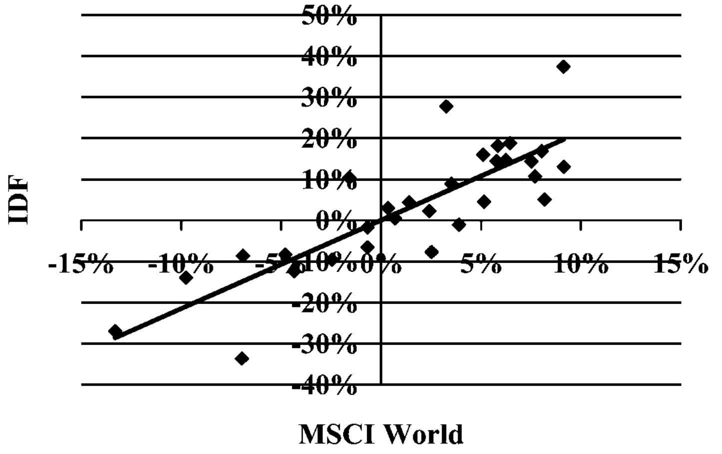


Fig. 5. Gross monthly log returns on the IDF and MSCI, and linear regression line.

implied volatility of at the money options that expire at the investment horizon. The investment policy of the IDF does not allow for stock picking or market timing.

To assess the impact of investing in the IDF, an empirical relationship between the return on the IDF and the return on the MSCI world was established. Figure 5 shows the track record that was used to estimate a relationship between the monthly returns on the IDF and the monthly returns on the MSCI world.

The ALM-model uses annual returns. Therefore, we transform the relationship between the monthly log returns to a relationship between annual log returns. This is straightforward under the assumption that neither the monthly returns on the MSCI World nor the error term in the regression are autocorrelated. We obtain the following relationship:

$$\text{Return IDF} = e^{2.15 \ln(\text{Return MSCI}) + u},$$

where:

Return MSCI = Gross annual total return on the MSCI World Index,

Return IDF = Gross annual return on the IDF,

u = Random number sampled from a normal distribution with mean 0 and standard deviation 0.319.

In IDF the coefficient of 2.15 reflects substantial leverage to equity returns whilst the functional specification does not allow for returns on the IDF worse than −100%. This is realistic because investors in the IDF only buy a participation in a mutual fund and cannot loose more than their initial investment. This combination of leveraged equity exposure and limited liability can only be attained by the use of options.

Using this relationship we generate returns on the IDF via a 2-step procedure. First, using the approach described in Section 3, we generate scenarios for all economic quantities except the return on the IDF. Then, for each point in time in each scenario the return on the IDF is computed by substituting the associated return on equity and a random number, sampled from the error distribution u, in the equation above. This procedure has two important advantages over using theoretical option pricing models. Firstly, the IDF returns are generated in a way that is fully consistent with the scenario generation of the quantities. And secondly, the returns are based on a historic relationship between real life option returns and the MSCI world which reflects all practical issues that affect returns on option portfolios. The crucial assumption that the decision makers have to make, however, is that this historic relationship is sufficiently representative for the future. But then again, it is straightforward to use the ALM-model to assess the sensitivity of the outcomes of the study to perturbations in this relationship.

4.3. Determination risk profile and strategic asset allocation

Assuming that the pension plans have already decided on the COLA- and contribution part of the ALM-policy, the first ALM-issue that we have to address is the specification of the strategic asset allocation (SAA). Evidently, more investment risk will on the one hand yield higher expected portfolio returns, at the benefit of the contribution rates, the indexations and/or strengthening the buffer, whereas the underfunding-, contribution-, and indexation risk will deteriorate on the other. In deciding on the SAA, and thereby completing the integral ALM-policy, we assume:

(i) The trustees of the boards put an explicit risk constraint on the funded ratio;

(ii) In the basic ALM-policy the SAA will consist of Euro-bonds, and of the equity classes Euro, US, UK, Japan and the emerging markets, where the returns of latter four are unhedged for currency risk. The total amount in equity will be allocated to Euro 25%, US 50%, UK 10%, JP 10% and emerging markets 5%.

Then, the SAA is the mix, structured according to (ii), which satisfies the risk constraint (i).

Thus, it is of crucial importance that the trustees properly determine the maximum allowable downside risk of the funded ratio. In practice, this is supported by the information which is displayed in Figure 6, which for plan A shows how the funded ratios develop in the 2500 scenarios if the plan would, respectively, invest 30, 43, 50 or 60% in equity. In numerous practical ALM-projects the trustees of the plans, thoroughly analyzing and discussing (if not debating) these graphical presentations of the scenario developments of their intended ALM-policies, frequently decide to limit the risk profile of the funded ratio to a downside deviation of at most 2%. For plan A this would imply an SAA with 43% internationally diversified unhedged equity.

Observe from Figure 6 that each extension of the SAA with 10 pp equity would imply that the lower envelop of the scenario developments of the funded ratios would deteriorate approximately 10 pp. Thus, taking into account that the portfolio returns on the pension assets which are required to cover the autonomous growth of the liabilities increases disproportionally with decreasing funded ratios,[4] it cannot be considered un-

[4] Let λ be the liability return, and let f be the funded ratio. Then, in order to meet the liability growth the pension assets have to render a portfolio return of λ/f. For example, if f drops from 100 to 90, 75 and 50%,

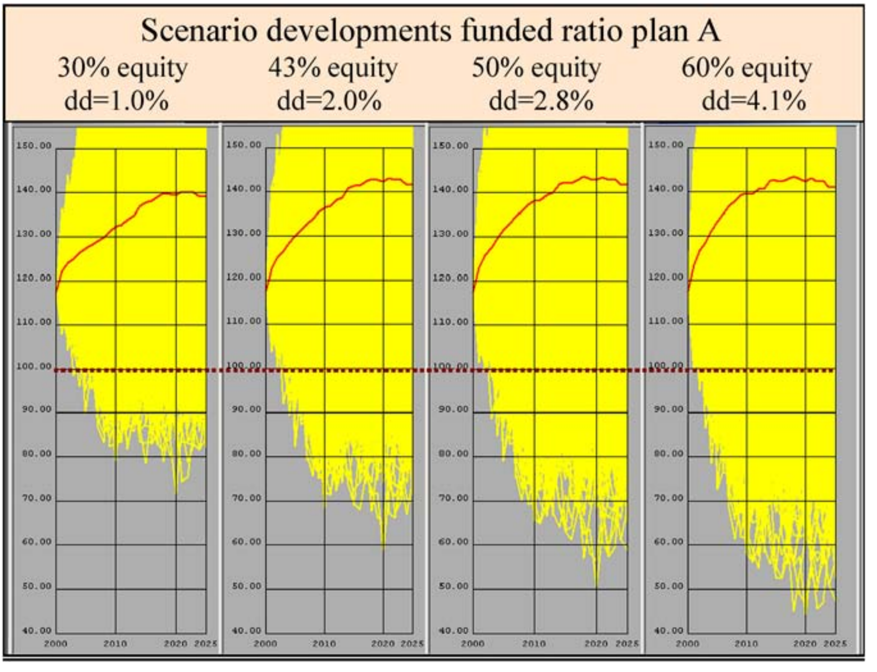


Fig. 6. Scenario developments and scores of the downside deviation (dd) of the funded ratio of plan A with 30, 43, 50 and 60% equity in the strategic asset mix.

wise that the trustees do not authorize a higher risk-profile than a downside deviation of at most 2%.

Applying the risk norm to limit the downside deviation of the funded ratio to 2% also to the plans B and C, these plans can carry a SAA with, respectively, 40 and 33% equity. In combination with the COLA- and contribution policy, which we assume to remain unchanged, the SAA's complete the basic ALM-policies of the plans. The ALM-scores of these policies are depicted in Table 5. We observe that, apparently counter-intuitive, plan B, whose liability leverage is about three times larger than for plan A, and therefore about three times stronger than plan A from the point of view of the impact of additional contributions on the funded ratio, nevertheless contains less equity in the SAA than plan A. This is because the high liability leverage, which makes the plan stronger than plan A at low levels of the funded ratio, also accounts for three times larger deterioration's of the funded ratio if the plan reduces the contributions at high levels of the funded ratio. As a result the expected funded ratio of plan B (127%) is

then the required portfolio returns increase, respectively, 11, 33 and 100%. These required returns only cover the autonomous growth of the liabilities, such that in order to improve the funded ratio, even higher returns have to be achieved.

Table 5
ALM-results for plans A, B and C, given downside deviation of the funded ratio $\leqslant$2% (horizon 25 years, 2500 scenarios)

		Max. perc. equity	Exp. contribution rate	Exp. funded ratio	Prob. funded ratio < 100%	1% SaR 1 yr	Prob. incomplete COLA	5% PaR 3 yr	Prob. contribution > actuarial contribution	5% CaR 3 yr
Plan A	Basic policy	43	5.4	137	4.2	10.7	5.9	3.3	30	59
	Max contribution 15%	37	6.3	134	4.6	10.6	6.2	3.7	33	45
	No risk sharing with non-active members	28	8.2	133	4.2	11.1	0.0	0.0	31	59
	Hedging currency risk	73	1.1	143	3.9	10.7	5.2	1.5	25	57
	State dependent equity exposure (avg)	56	1.4	133	6.4	10.2	9.3	5.1	41	60
	20% alternative equity exposure	58	1.0	143	3.8	10.8	5.0	1.8	26	56
Plan B	Basic policy	40	7.9	127	5.5	10.8	7.2	4.2	42	55
	Max contribution 15%	36	8.0	125	5.9	10.9	7.8	4.7	45	45
	No risk sharing with non-active members	29	8.7	124	6.0	11.1	0.0	0.0	46	54
	Hedging currency risk	69	6.3	132	5.1	10.9	6.5	3.6	35	50
	State dependent equity exposure (avg)	50	6.4	127	6.7	10.3	8.8	4.5	49	57
	20% alternative equity exposure	53	6.3	132	5.3	11.4	6.7	3.9	36	51
Plan C	Basic policy	33	3.7	134	5.0	10.7	6.8	4.2	37	60
	Max contribution 15%	28	5.3	132	5.7	11.3	7.7	5.2	41	45
	No risk sharing with non-active members	17	10.4	127	4.7	11.0	0.0	0.0	43	60
	Hedging currency risk	56	−3.9	142	4.2	10.5	5.4	2.4	30	60
	State dependent equity exposure (avg)	42	−4.4	128	7.5	10.1	10.8	6.3	54	60
	20% alternative equity exposure	46	−4.2	141	4.3	11.0	5.7	3.0	30	60

substantially lower than the expected funded ratio of plan A (137%), which of course reduces the maximum feasible risk exposure of the SAA of plan B. The net result appears that plan B, notwithstanding her contribution power, can invest slightly less in equity than plan A. This phenomenon emphasizes the complexity of integral (COLA-, contribution-, investment-) ALM-policies, as well as the quintessential role of buffers in strategic asset allocation of ALM.

4.4. Risk sharing and integral ALM

In previous sections we described the relevancy of integral ALM, where also the risk budgets provided by the sponsors and the non-active plan members are exploited in order to be able to carry sufficient investment risk, and earn sufficient investment returns to make it feasible to provide adequate pension payments at acceptable pension cost.

In this section this statement is quantified. That is, we analyze the ALM-consequences if the sponsor reduces the risk budget by allowing maximal contribution rates of 15% of salaries, rather than 20%. Secondly, we analyze the ALM-impact if the non-active members would withdraw their risk budget by forcing the plan to fully compensate their pension rights for inflation, whatever the value of the funded ratio. Table 5 shows the importance of the reduction of these risk budgets. That is, if the sponsor is only willing to pay 15% of salaries if the funded ratio deteriorates, then the plans A, B and C would have to reduce the percentage of equity in the SAA's from 43, 40 and 33% to, respectively, 37, 36 and 28% with obvious consequences for the contribution rates.

The risk budget provided by the non-active members appears to be even more important. If this risk budget would be withdrawn, then the percentage of equity in the SAA's would have to be reduced from 43, 40 and 33% to, respectively, 28, 29, and 17%. Taking in mind that an equity risk premium of 3.5% implies that 10% less equity in the SAA of a portfolio of 10 billion reduces investment income with 35 million p.a., the necessary reduction in equity as a result of the withdrawn risk budget can be considered enormous. This is verified by the implications on the expected contribution rates in Table 5: For plan A, the expected contribution rate would increase from 5.4% of salaries to 8.2%. For the younger plan B with the smaller liability leverage, the contribution rate is 8.7% rather than 7.9%. The mature plan C with the relatively large group of non-active members, and the large liability leverage, profits most from the risk budget of the non-active members. For this plan the withdrawal of this risk sharing leads to an expected contribution rate of 10.4% rather than 3.7%.

Finally, the risk budget provided by the non-active members may appear asymmetric to the benefit of the sponsor. Due to this risk budget the expected contribution rates decrease significantly, while the risk providers are confronted with pension risk with apparently no "upside" in return. The (historic) background of this phenomenon is that due to this risk budget the sponsors are able to provide high DB pension rights at reasonable internationally competitive cost, while the non-active plan members are only confronted with reasonably low pension risk.

4.5. *Hedging strategic currency risk*

An investment outside the home country is a composition of local asset classes and currency exposure. Pension plans increasingly decide to work with hedged strategic benchmarks. That is, they decide to measure their "strategic value added" with respect to benchmarks in which currency risk is eliminated.[5] In this section we give the arguments and counter arguments for this policy, and analyze the potential ALM-efficiency gains for the plans A, B and C.

(a) Risk-return trade-off

The most important argument to hedge strategic currency risk is the assumption that in spite of short time volatility, currencies will not yield an expected long term return. Observe that if pension plans would really in every respect be long term investors, this argument to hedge strategic currency risk would lose a lot of her power: Pension plans could just surf the currency waves without incurring hedging cost. However, also for long term pension plans the short term is extremely relevant. Drops in the funded ratio might trigger additional contributions and postponement of COLA, while also International Accounting Standards (IAS) more-and-more transfer profits and losses of the pension plans to the P&L's of the sponsors. Thus also pension plans have to take into account the short term implications of their long term strategies.

In the scenarios hedged returns are constructed from the local returns by determining expected values according to the interest rate parity (i.e., forward currency returns make up for the differences in the short interest rates between home-countries), and adding correlated error terms which are based on the deviations from the interest rate parity in the applied historic data set (1970–present). The characteristics of the hedged returns, taking into account 20 basis points per annum for the cost of hedging, are displayed in Tables 3 and 4. Clearly the hedged returns provide a much more efficient risk-return trade-off than the unhedged counterparts.

(b) Mean reversion

In addition to the risk-return trade-off, strategic currency hedging also finds its roots in the concept of mean reversion. It can be verified from Table 3 that local equity in all countries displays mean reversion properties. See, for example, Euro equity with a historic mean reversions of -0.19 (lag 1 year), and local US-equity with historic mean reversion of -0.05 and -0.21 (lag 1 year, and lag 2 years). Thus, locally equity returns below the mean with more than 50% probability are followed by returns above the mean, which, in particular, for long term investing pension plans provides a crucially important strong mitigation of risk.

However, we can also verify from Table 3 that unhedged returns, composed of local returns and currency returns, lose the property of mean reversion, especially for US-equity, where the majority of the Dutch international investments are allocated.

[5] A hedged strategic benchmark is not in contradiction with currency bets on the tactical investment level.

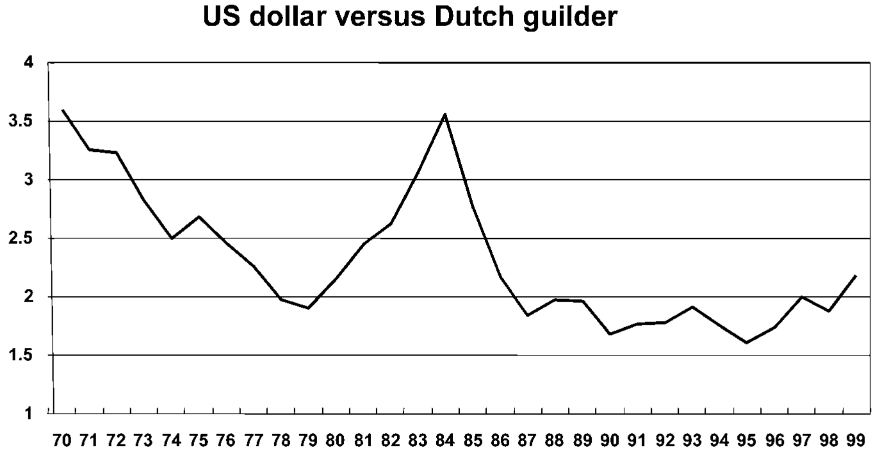


Fig. 7.

This is due to the high mean aversion (= positive autocorrelation) of +0.34 of US-dollar-return with respect to the Euro in the historic dataset (cf. also Figure 7), which obviously ruins the mean reversion of local US-equity from the point of view of the unhedged dollar-exposed European investor in US-equity. Clearly, if history would repeat itself, this would provide an additional strong argument to hedge strategic currency risk, especially for long term European investments in the US.

Table 5 displays the ALM-results if the plans would carry out their basic ALM-policy, and hedge the strategic currency risk of their equity investments in the US, UK and Japan. The ALM-results are impressive. Satisfying the risk constraint on the funded ratio plan A can increase the investments in equity form 43 to 73% of the portfolio, with a resulting reduction of the expected contribution rate from 5.4% of salaries to 1.1%. Also, the scores on all other ALM-criteria improve, in particular, the 5% 3-year Pension at Risk for the non-active members, which almost halves from 2.7 to 1.5%. Decomposition of these efficiency gains learns that about 2/3 can be attributed to the risk-return trade-off of hedged returns (a), and 1/3 to the retaining of mean reversion.

Next we address whether there are also counter-arguments for fully hedged benchmarks:

– Forward premium puzzle

A delicate issue is the impact of the forward premium puzzle on the decision to hedge strategic currency risk. That is, that the dollar-return and the forward dollar rate (which is used for hedging the currency risk) in practice are negatively correlated. In the historic period which we apply for scenario generation this correlation is −0.38. In the scenarios the puzzle presents itself in the curious phenomenon that the correlation between the dollar and local US-equity is +0.16, whereas the correlation between the forward dollar rate and local US-equity is -0.31. Thus recalling

$$\sigma^2(\text{unhedged}) = \sigma^2(\text{local equity}) + \sigma^2(\text{dollar}) + 2 * \rho(\text{local equity, dollar}) * \sigma(\text{local equity}) * \sigma(\text{dollar})$$

$$\sigma^2(\text{hedged}) = \sigma^2(\text{local equity}) + \sigma^2(\text{forward rater}) + 2 * \rho(\text{local equity, forward rate}) * \sigma(\text{local equity}) * \sigma(\text{forward rate})$$

we must conclude that the puzzle works out very negative for the volatility of unhedged returns (positive correlation between local US-equity and the dollar), and positively for hedged returns (negative correlation between local US-equity and the forward rate).

– Peer group risk

Also pension plans are sensitive for peer group risk. That is, if a plan realizes a lower portfolio return than the universe (whatever the benchmarks and the liability structure of the plan), this will lead to bad publicity in the press. Thus, in taking the hedge decision also the probability of realizing a lower performance than unhedged benchmarks has to be taken into account. Relevant probabilities are (of course) that a hedged portfolio with 50% probability will generate a lower return than an unhedged portfolio, and that a hedged portfolio with 10% probability generates a 3% lower return as an unhedged portfolio. Taking into account that 3% could be seen as a reasonable estimate for the equity risk premium, these figures again emphasize that the currency exposure strongly matters.

– Long term implementation

Finally, if boards of pension plans abstain from hedging strategic currency risk, the most overriding argument is implementation risk. Currency hedging is implemented using forward contracts. That is, if the currency depreciates, the plan is regularly compensated for the loss. But, if the currency appreciates, the plan is regularly “billed”. Figure 7 indicates that currencies tend to move in long waves. Thus, a plan that hedges currency risk may be confronted with a sequence of dozens of months in which the plan repeatedly has to pay to compensate for the appreciating currency. Boards fear the risk that during such a long sequence they might bail out of the hedging policy, thereby incurring a considerable loss, rather than waiting for the recovery which is implicit in the basic assumption that currencies yield no long term return.

4.6. State dependent asset allocation

In this section we address the issue of state dependent asset allocation. Intuitively, the more surplus over the liabilities, the more market risk a pension plan could carry in the investment portfolio. A logical consequence of this proposition (not appreciated by many investment managers faithfully acting on the principle of mean reversion) is that the smaller the surplus, the less a pension plan should expose herself to market risk, and

Table 6
State dependent percentages of equity in the strategic asset allocation: Evaluation criteria: Expected contribution rate & downside deviation funded ratio $< 2\%$

Plan A	Plan B	Plan C
Funded ratio: $\leqslant 105:20\%$	$\leqslant 110:20\%$	$\leqslant 115:10\%$
Funded ratio: $120:55\%$	$120:55\%$	$120:55\%$
Funded ratio: $130:65\%$	$130:65\%$	$130:65\%$
Funded ratio: $\geqslant 190:95\%$	$\geqslant 190:95\%$	$\geqslant 190:95\%$

thus reduce equity in the SAA. An important argument underlying this proposition is that the portfolio return that is required to cover the autonomous growth of the liabilities increases exponentially if the funded ratio deteriorates. As a result it would be rational to carry out a policy which aims to prevent underfunding by timely reducing equity in the SAA if the funded ratio deteriorates.

We investigated this proposition by the simple random search algorithm described in Section 3. The (approximate) optimal policies that have been identified by the search for the plans A, B and C are depicted in Table 6.[6] The ALM-consequences of these policies are contained in Table 5.

The structure of the optimal rules is that if the funded ratio drops to 120%, which is also the level at which the contribution rate is increased, the amount of equity is significantly reduced. Clearly this element of the optimal rules guarantees that deficits are efficiently aimed to prevent. If the funded ratio increases above the level of 120%, the optimal amount of equity increases proportionally: up to a funded ratio of 130% with 10% of equity per 10% points more funded ratio, and at funded ratios above 130% with 5% of equity per 10 pp more funded ratio.

Secondly, the ALM-scores in Table 5 indicate that also the ALM-scores of state dependent asset allocation are impressive. In relation to the basic ALM-policy the expected contribution rate for the plans A, B and C reduces from 5.4, 7.9 and 3.7% to 1.4, 6.4 and -4.4% at the same underfunding risk. Hence, given these criteria, state dependent asset allocation is extremely efficient. However, due to the state dependent investment policy the COLA- and contribution risks significantly increase. Hence, in practice pensioners might argue that such an investment policy should be accompanied by appropriate measures to reduce COLA risk.

[6] If the algorithm would randomly generate state dependent rule of arbitrary form, then the optimal rules would be V-shaped. That is, given the applied ALM-criteria, at low funded ratios it is optimal to increase the amount of equity in the SAA if the funded ratio further deteriorates. This result is also sustained by the more sophisticated recourse optimization models (see, e.g., Berkelaar (2002), Dert (1995), and Kouwenberg (2000)). Apparently, if the funded ratio is critically low, the optimal escape is to increase the expected portfolio return, and accept the accompanying higher risk. However, since the regulating authority forbids this conduct, we restricted the random optimizer to generate only rules where the market risk in the SAA increases monotonically with increasing funded ratios.

4.7. Alternative equity exposure

The use of derivatives in ALM is still a relatively uninvestigated research area, see, e.g., Boender, Oldenkamp and Vos (1997). In this final section we investigate the ALM consequences of the alternative equity exposure via the IDF derivatives product described in Section 4.2. We carried out the investigation by replacing 20% of the equity exposure of the asset mixes by the IDF product (thus, in an asset allocation with 50% equity, 10% points of equity is replaced by IDF). The results appear in Table 5. The reader can verify that the ALM efficiency gains are very significant. Why should using derivatives produce these improvements? The improvements cannot be explained from the Information Ratio of the derivatives product. The key is asymmetry. Using derivatives ALM-ers can create upside potential, while efficiently protecting downside risk. Therefore, with respect to the strategic use of derivatives in ALM, we expect that we have only seen the very beginning.

5. Conclusion

This paper emphasizes the integral approach of ALM-problems. That is, given a set of assumptions about the relevant uncertainties, for each of the stakeholders a maximum risk limit is specified. This implies that risk limits are specified with respect to pension and indexation/COLA at risk, contribution at risk, and surplus at risk. Given these risk limits an integral indexation/COLA-, contribution-, and investment policy is specified which optimizes return. The ALM-policy which is the result of the integral approach is also referred to as the pension deal.

The empirical Section 4 shows the relevance of this approach. For plan A which can carry 43% of equity and 57% bonds in the base case, the maximum amount of equity decreases to 37 and 28%, respectively, if the sponsor reduces the maximum contribution form 20% of salaries to 15%, and if the non-active members abstain from state-dependent COLA.

The paper also demonstrates the relevance of the structure of the liabilities. In the base case, where plan A can carry 43% of equity in the strategic asset allocation, plans B and C that are, respectively, less and more mature than plan B, the amount of equity in the strategic asset allocation is ceteris paribus equal to 40 and 33%. Recently the relevance of the liabilities to investing is also emphasized in the investment community (cf., e.g., inflection point III in Bernstein (2003), referred to as liability driven investing (LDI)).

The paper also demonstrates the effectiveness of alternative asset categories and alternative asset strategies. We selected to study hedging currency risk, state dependent asset allocation and a high upside/low downside derivatives product. Given the same amount of surplus at risk, for plan A these alternatives imply that the plan can risk neutrally change the strategic asset allocation to 15 and 30% more equity. The effects of other measures such as flexible pension schemes, dynamic derivative strategies, matching of interest rate risk, elimination of the "long-only" constraint, and the impact of other alternative investments such as commodities, private equity and hedge funds are not described here.

Finally, our approach is based on simulation. We look forward to integrating more stochastic optimization in our models (see, e.g., Wallace and Ziemba (2005), and the results of others in this volume), such that ALM projects in the pension-, insurance-, and banking industry both applying new products and new technologies can further improve the efficiency of their strategies.

References

Ambachtsheer, K.P., Ezra, D.D., 1998. Pension Fund Excellence. John Wiley.

Arnott, R.D., Bernstein, P.L., 2002. What risk premium is "normal". Financial Analysts Journal, 64–85, March/April.

Berkelaar, A., 2002. Strategic asset allocation and asset pricing, PhD. Erasmus University Rotterdam.

Bernstein, P.L., 2003. Points of inflection: Investment management tomorrow. Financial Analysts Journal, 18–23, July/August.

Boender, C.G.E., 1997. A hybrid simulation/optimization scenario model for asset–liability management. European Journal of Operations Research 99, 126–135.

Boender, C.G.E., Oldenkamp, B., Vos, M., 1997. Solvency insurance with optioned portfolios: An empirical investigation, in: AFIR Proceedings.

Boender, C.G.E., Romeijn, H.E., 1991. The multi-dimensional Markov chain with prespecified asymptotic means and (auto-)covariances. Communications in Statistics 20, 1.

Boender, C.G.E., Romeijn, H.E., 1995. Stochastic methods. In: Horst, R., Pardalos, P.M. (Eds.), Handbook of Global Optimization. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 829–870.

Bunn, D.W., Salo, A.A., 1993. Forecasting with scenarios. European Journal of Operations Research, 68.

Carino, D.R., Kent, T., Meyers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., Ziemba, W.T., 1994. The Russell Yasuda–Kasai model: An asset–liability model for a Japanese Insurance Company using stochastic programming. Interfaces 24 (1), 29–49.

Carino, D.R., Ziemba, W.T., 1998. Formulation of the Russell Yasuda–Kasai financial planning model. Operations Research 46 (4), 433–449.

Davis, E.P., 1994. Pension Funds: Retirement Income Security and Capital Markets: An International Perspective. Clarendon Press, Oxford.

Dert, C.L., 1995. A multistage stochastic programming approach to asset–liability management, PhD. Erasmus University Rotterdam.

Dert, C.L., 2002. Strategic Use of Options in Pension Investments.

Dimson, E., Marsh, P.R., Staunton, M., 2002. Triumph of The Optimist. Princeton University Press; plus yearly updates from: The Global Investment Returns Yearbook, 2006, ABN-AMRO joint with Rolf Elgeti.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2005. The Innovest Austrian Pension Fund Planning Model InnoALM, Mineo. University of Vienna.

Kingsland, L., 1982. Combining financial and actuarial risk: Simulation analysis. Projecting the financial condition of a pension plan using simulation analysis. Journal of Finance 37 (2), 577–584.

Kouwenberg, R.R.P., 2000. Dynamic asset liability management, PhD. Erasmus University Rotterdam.

Mehra, R., 2006. The Equity Risk Premium. Handbook of Finance, North-Holland.

Modigliani, F., Muralidhar, A.S., 2004. Rethinking Pension Reform. Cambridge University Press.

Muralidhar, A.S., 2001. Innovations in Pension Fund Management. Stanford University Press.

Rudolf, M., Ziemba, W.T., 2004. Intertemporal asset–liability management. Journal of Economic Dynamics and Control 28 (4), 975–990.

Sharpe, W.F., Tint, L.G., 1990. Liabilities: A new approach. The Journal of Portfolio Management 17, 5–10.

Siegel, J., 2003. Stocks for the Long Run, third ed. McGraw-Hill, New York.

Siegmann, A.H., 2003. Optimal financial decision making under loss averse preferences, PhD thesis. Free University Amsterdam.

Sortino, F.A., van der Meer, R., 1991. Downside risk. The Journal of Portfolio Management, 27–31.

Steehouwer, H.S., 2005. Macroeconomics and reality, PhD. Free University Amsterdam.

Wallace, S.W., Ziemba, W.T., 2005. Applications of Stochastic Programming. SIAM–Mathematical Programming Society.

Winkelvoss, H.E., 1982. PLASM: Pension Liability and Asset Simulation Model. Journal of Finance 37 (2), 585–594.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset, Liability, and Wealth Management. AIMR.

Chapter 19

THE RUSSELL-YASUDA KASAI, INNOALM AND RELATED MODELS FOR PENSIONS, INSURANCE COMPANIES AND HIGH NET WORTH INDIVIDUALS[1]

WILLIAM T. ZIEMBA

Sauder School of Business, University of British Columbia, Vancouver, BC V6T 1Z2, Canada

Contents

[1] This research was supported by NSERC. This chapter relies on my previous published papers and books: Kallberg, White and Ziemba (1982); Kusy and Ziemba (1986); Cariño, Ziemba et al. (1994); Cariño and Ziemba (1994); Cariño, Myers and Ziemba (1998); Ziemba and Mulvey (1998); Ziemba (2003); Geyer et al. (2002) and Geyer and Ziemba (2007).

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02019-9

Abstract

All institutions and individuals and have asset/liability management problems continuously. Assets must be invested to achieve sufficient returns to pay liabilities and achieve goals subject to uncertainties, policy and legal constraints, taxes, etc. This chapter discusses asset and liability management in the presence of portfolio complexities, such as transaction costs, liquidity, taxes, investor preferences (including downside risk control, policy and other constraints), uncertain returns, and the timing of returns and commitments. The issues involved have common characteristics in the management of investment portfolios for large financial institutions (such as pension funds, insurance companies, and hedge funds) and for individual life-cycle planning. The approach recommended is discrete-time, multiperiod stochastic programming. For most practical purposes, such models provide a superior alternative to other approaches, such as mean-variance, simulation, control theory, and continuous-time finance. Stochastic programming leads to models that take into account investor preferences in a simple, understandable way. The paper discusses scenarios in several ways and how economic models can be used for their generation. This involves a few case examples such as crash models for equity markets and hedge fund disasters that occurred because of overbetting and not being diversified in all scenarios. The models presented are the Russell-Yasuda Kasai and the InnoALM both of which had a major impact on the practice of ALM.

Keywords

pensions, insurance, scenarios, stochastic optimization

JEL classification: C61, C88, G11

1. Introduction

The use of scenario-based, stochastic programming optimization models in discrete time provides an approach to asset/liability modeling over time. The models provide a tool to think about, organize, and do calculations concerning how one should choose asset mixes over time to achieve goals and cover liabilities. Risk and return are balanced to achieve period-by-period goals and pay liabilities. The models force diversification and the consideration of extreme scenarios to protect investors from the effects of tail outcomes and also do well in normal times. They will not let individuals or institutions get into situations in which extreme, but plausible, scenarios would lead to truly disastrous consequences, such as losing half or more of one's assets. Because these models force consideration of all relevant scenarios, the common practice of assuming that low-probability scenarios will not occur is avoided. Hence, the disasters that frequently follow from this error are avoided.

In my discrete-time, multiperiod stochastic programming models, one typically maximizes a concave, risk-averse utility function composed of the discounted expected wealth in the final period less a risk measure composed of a risk-aversion index times the sum of convex penalties for target violations relating to investor goals of various types in various periods. The convexity means that the larger the target violation, the larger the penalty cost. Hence, risk is measured as the non-attainment of investor goals, and this risk is traded off against expected returns. This approach is similar to a mean-variance preference structure, except it is based on final wealth and the risks are downside risks that are measured across periods and investor goals. Discrete scenarios that represent the possible returns and other random parameter outcomes in various periods are generated from econometric and other models, such as those related to market dangers with increasing risk and from expert modeling. Mean-return estimation and inclusion of extreme events are important for model success. The scenario approach has a number of advantages:

- Normality or lognormality, which is used in other approaches but is not an accurate representation of actual asset prices, especially for losses, need not be assumed.
- Tail events can be easily included; studies show that downside probabilities estimated from actual option prices are 10 to 100 to 1,000 or more times fatter than lognormal.
- Scenario-dependent correlations between assets can be modeled and used in the decision-making process so that "normal" and "crisis" economic times (with higher and differing signed correlations) can be considered separately.
- The exact scenario that will occur and the probabilities and values of all the scenarios do not need to be accurately determined to provide model performance that is superior to that of other models and strategies, such as mean-variance or portfolio insurance.
- The best decisions are determined in light of relevant constraints, uncertainties, and preferences of the decision maker.
- Most of the natural, practical aspects of asset/liability applications can be modeled well in the multiperiod stochastic programming approach. Methods to solve such models are now highly developed and can be implemented on high-performance

personal computers. Model output is easy to understand and interpret if good graphical interfaces are used that are user friendly and easily understood by such non-optimization experts as pension fund trustees. The models can be tested via simulation and statistical methods and considerable independent evidence demonstrates their superiority to other standard approaches and strategies.
- The approach protects investors from large market losses by considering the effects of extreme scenarios while accounting for other key aspects of the problem. The world is very dangerous with many extreme scenarios coming from economics, terrorism, etc. For example, the use of derivatives worldwide has grown into a US$ 200 trillion industry. Although much of this trade is for hedging and reduces risk, the sheer volume yields new risks.
- Determining whether investment positions are truly diversified and of the right size across time is crucial to protect against extreme scenarios and ensure that the results will be good in normal times and avoid disasters.

The chapter discusses the basic two models in some detail. The rest of the chapter discusses various aspects of scenario analysis, hedge fund disasters and their prevention, the importance of means and other relevant portfolio analysis.

2. How to make good multiperiod models

Over the years since my 1974 course at UBC on stochastic programming where I started working on SP-ALM models especially with my students Martin Kusy and Jerry Kallberg, I have developed my own philosophy of a good way to make multiperiod SP models. This is highly related to my work on stock market anomalies which yield better scenarios and hedge fund type investment strategies. To actually do well in financial markets you must first believe that you can beat them. So, let me first introduce the various stock market camps.

2.1. The various efficient/inefficient market camps: Can you beat the stock market?

I divide market participants into five groups. There are other ways to do this categorization but my way is useful for our purpose of isolating and studying great investors and those who want to build SP models. This naturally evolves from the academic study of the efficiency of financial markets.

The Five Groups are:

1. Efficient markets (E).
2. Risk premium (RP).
3. Genius (G).
4. Hog wash (H).
5. Markets are beatable (A).

The first group are those who believe in efficient markets (E). They believe that current prices are fair and correct except possibly for transactions costs. These transaction costs, which include commissions, bid-ask spread, and price pressures, can be very large. A BARRA study made by Andy Rudd some years ago showed that these cost for small cap stocks averaged 4.6% one-way for a $50,000 institutional investor sale. The leader of this school which had dominated academic journals, jobs, fame, etc. in the 1960s to the 1980s was Gene Fama of the University of Chicago.

This group also provided many useful concepts such as the capital asset pricing model of Sharpe, Lintner and Mossin which provided a theoretical justification for index funds which are the efficient market camp's favored investment mode. They still beat about 75% of active managers. Index funds have grown and grown. Dimension Fund Advisors formed by Fama's students manages over $25 billion and others such as Barclay's in San Francisco manage over $100 billion. This is done with low fees in an efficient manner. The indices for these passive funds have grown to include small cap, foreign investments and a variety of exchange traded funds as well as the traditional market index, the S&P 500.

Over time the hard efficient market line has softened into a Risk Premium (RP) camp. They feel that markets are basically efficient but one can realize extra return by bearing additional risk. They strongly argue that if returns are above average, then the *risk* must be there *somewhere*; you simply cannot get higher returns without bearing additional risk they argue. For example, beating the market index S&P 500 is possible but not risk adjusted by the CAPM. They measure risk by Beta, which must be greater than one to receive higher than market returns. That is, the portfolio risk is higher than the market risk. But they allow other risk factors such as small cap and low book to price. But they do not believe in full blown 20 plus factor models. Fama and his disciples moved here in the 1990s. This camp now dominates the top US academic journals and the jobs in academic finance departments at the best schools in the US and Europe.

The third camp is called Genius (G). These are superior investors who are brilliant or geniuses but you cannot determine in advance who they are. MIT economics professor Paul Samuelson has championed this argument. Samuelson feels that these investors do exist but it is useless to try to find them as in the search for them you will find 19 duds for every star. This view is very close to the Merton–Samuelson criticism of the Kelly criterion: that is, even with an advantage and many plays, it is possible to lose a lot of your wealth (see MacLean and Ziemba, 2006). The evidence though is that you can determine superior investors ex ante and to some extent they have persistent superior performance. In his Quantum funds, George Soros did this with superior picking of futures to bet on. This is the traders are "made not born" philosophy. This camp will isolate members of other camps such as in (A) or (H).

The fourth camp is as strict in its views as camps (E) and (RP). They feel that efficient markets which originated in and is perpetuated by the academic world is hogwash (H). In fact, the leading proponent of this view and one it is hard to argue with as he is right at the top of the worlds richest person's list is Warren Buffett, who says he wants to give university chairs in efficient markets to further improve his own very successful trading. An early member of this group, the great economist John Maynard Keynes was an academic. We see also that although they may never heard of the Kelly criterion, this camp does seem to use it implicitly with large bets on favorable investments; see MacLean and Ziemba (2006), and Thorp (2006).

This group also feels that by evaluating companies and buying them when their value is more than their price, you can easily beat the market by taking a long term view. They find these stocks and hold them forever. They find a few such stocks that they understand well and get involved in managing them. They forget about diversification because they try to only buy winners. They also buy whole companies and let the original managers run them and bet on insurance when the odds are greatly in their favor. They well understand tail risk which they only take at huge advantages to themselves when the bet is small relative to their wealth levels.

The last group are those who think that markets are beatable (A) through behavioral biases, and security market anomalies using computerized superior betting techniques. They construct risk arbitrage situations with positive expectation. They research the strategy well and follow it for long periods of time repeating the advantage many times. They feel that factor models are useful most but not all of the time and show that beta is not one of the most important variables to predict stock prices. They use very focused, disciplined, well researched strategies with superior execution and risk control. Many of them use Kelly or fractional Kelly strategies. All of them extensively use computers. They focus on not losing, and they rarely have blowouts. Members of (A) include Ed Thorp, Bill Benter, Nikolai Battoo, Blair Hull, Harry McPike, James Simons, Jeffrey Yass, David Swensen and me. Blowouts occur more in hedge funds that do not focus on not losing and true diversification and over-bet; when a bad scenario hits them, they get wiped out, such as LTCM and Niederhoffer, see below for details. My idea of using scenario dependent correlation matrices (Geyer and Ziemba, 2007), discussed in the InnoALM section below, is important here. These A investors believe in scenarios and think they can devise them better than just relying on the past data. They use various models to do this.

2.2. How do investors and consultants do in all these cases?

All can be *multimillionaires* but the *centimillionaires* are in (G), (H) and (A) like the seven listed before me in (A) and Buffett. These people make more money for their clients than themselves but the amount they make for themselves is a huge amount: of course, these people eat their own cooking, that is, they are clients themselves with a large amount of their money in the funds they manage. (An exception is someone who founded an (RP) or (E) company kept most of the shares and made an enormous amount of fees for themselves irrespective of the investment performance given to the clients because the sheer volume of assets under management is so large.) But I was fortunate to work/consult with six of these centimillionaires and was also the main consultant to the Frank Russell Research Department for nine years which is perhaps the leading RP implementor. (A) people earn money by winning and taking a percent of the profits, Thorp returned 15.8% net with $200 million under management; fees about

$8 million/year. (E) and (RP) people earn money from fees by collecting assets through superior marketing and *sticky* investment decisions.

3. Scenarios

In this section we deal with understanding scenarios and extreme events and discuss in detail two disasters that could have been avoided with proper SP-ALM modeling.

3.1. Discrete scenarios and fat tails

What investors want is true diversification and protection from extreme scenarios when it is needed and to be able to plan in advance what they need to do in such situations.

The basic theory of modern finance uses normality as the key assumption in static or discrete-time models and lognormality in continuous-time models.

Sums of random variables converge to normal distributions by the central limit theorem, and sums of normal distributions are normal. Similarly, products of lognormal random variables are lognormal. Hence, these two assumptions are versions of a similar view of the world. These assumptions allow for a clean, elegant theory that is useful as a benchmark and to derive qualitative results. Real asset prices, however, have much fatter tails, especially during short intervals.

Table 1 provides insight into how market participants who buy and sell put and call options view the probability distribution of returns for the S&P 500 index which these yields scenarios on asset prices. The column of standard deviations corresponds to returns below the mean, and the standardized lognormal distribution values reflect this

Table 1
Cumulative probabilities of S&P 500 returns computed from daily bid–ask prices of puts and calls

Standard deviation	Standard lognormal distribution	Implied probability distributions		
		4/86–10/87	11/87–12/88	1/91–12/91
10	0.0000000000000000000000000078		0.0000083	0.0000018
9	0.00000000000000000011		0.000021	0.0000056
8	0.00000000000000063		0.000049	0.000017
7	0.0000000000013		0.00011	0.000049
6	0.000000001	0.00000016	0.00026	0.00015
5	0.00000029	0.000014	0.00076	0.00055
4	0.000032	0.00025	0.0029	0.0029
3	0.0013	0.0014	0.011	0.015
2	0.023	0.026	0.045	0.051
1	0.16	0.19	0.16	0.15
0	0.50	0.52	0.47	0.45

Source: Jackwerth and Rubinstein (1996).

distribution. The next three columns represent implied probability distributions from bid–ask prices of traded put and call options on the S&P 500 during one period before and two periods after the October 1987 stock market crash. These probabilities are not the true chances of these events occurring (which is unknown) but, rather, reflect the probabilities that the options market prices indicate for these events. Before the crash, the 1–3 standard deviation left tail event (i.e., losses) probabilities from the options are about the same as from the lognormal model underlying the Black–Scholes options' fair prices. The implied probabilities are 10 to 100 or more times fatter for 4–6 standard deviation moves. This finding is consistent with the behavioral economics finding that small probabilities are overestimated. See Kahneman and Tversky (1979) and the earlier papers in Hausch, Lo and Ziemba (1994) and Hausch and Ziemba (2007) on the favorite-longshot bias in racetrack betting. After the 1987 crash, the effect is even stronger because investors were willing to pay much more for out-of-the-money puts for portfolio insurance and other types of protection. The last two columns show tail probabilities 10 to 100 times fatter than the precrash levels.

An interesting question is: What is the return, on average, to investors who buy puts and calls at different strike prices on the S&P 500? Tompkins, Ziemba and Hodges (2003) have studied this question. Griffin (1947), McGlothlin (1956), Ali (1979), Snyder (1978) and others have documented a favorite long-shot bias in racetrack betting. These papers are all reprinted in Hausch, Lo and Ziemba (1994). The data show that bets on high-probability, low-payoff gambles have high expected value and those on low-probability, high-payoff gambles have low expected value. For example, a 1–10 horse, having more than a 90% chance of winning, has an expected value of about US$ 1.03 (for every US$ 1 bet), whereas a 100–1 horse has an expected value of about 13.7 cents for each dollar invested. Hence, for this bet, the fair odds are about 700–1. The favorite long-shot bias exists in other sports-betting markets. (See Hausch, Lo and Ziemba (1994) and Hausch and Ziemba (2007) for surveys of results, updates and references.)

Ziemba and Hausch (1986) studied the expected return per dollar bet versus the odds levels for more than 300,000 horse races. The North American public underbets favorites and overbets long shots. This bias has appeared in many years and in all sizes of racetrack betting pools. While the horseracing favorite long-shot bias is quite stable and pervasive, exceptions exist in Asian racetrack markets (see Busche and Hall (1988) and Busche (1994), reprinted in Hausch, Lo and Ziemba (1994)). The expected return varies with the odds level (see Figure 1, the top lines on the left); breakage, the rounding down of payoffs to round numbers like 6.87 to 6.80 is included here. There is a positive expected return for bets on extreme favorites, but for all other bets, the expected return is negative. The favorite long-shot bias is monotone across odds, and the drop in expected value is very large for the lower probability horses. The effect of differing track take—transaction costs—is seen in the California versus New York graphs. These are expected returns versus odds or probabilities for parimutuel pooled bets, and for fixed-odds bookie wagers, the results are similar because the bookies create odds to clear the market and equilibrate bettor demand knowing bettor biases. Figgis (1974) and Lord Rothschild (1978), and Ziemba and Hausch (1987) provide British data.

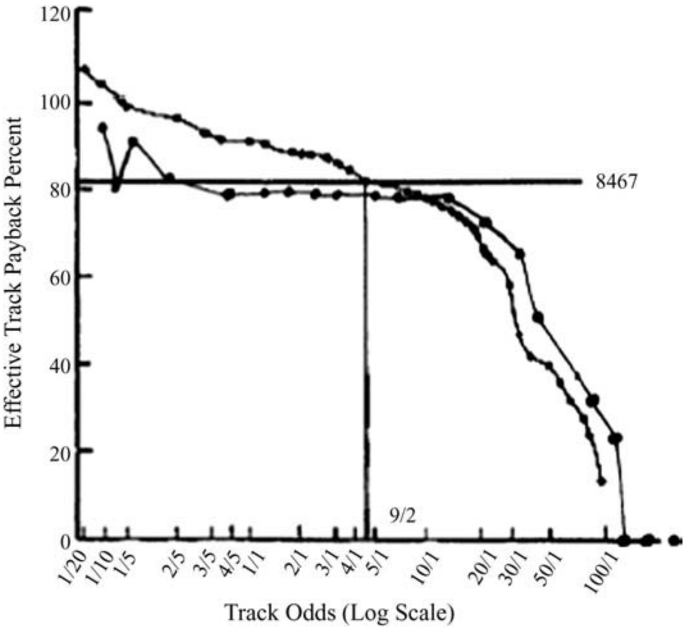


Fig. 1. Effective track payback less breakage for various odds levels in California and New York for 300,000^{+} races over various years and tracks (the top lines) and for essentially all North American races (300,000^{+} (lower line). Source: Ziemba and Hausch (1986) and Ziemba (2004).

Thaler and Ziemba (1988) suggested some reasons for this bias, such as bettors might overestimate the chances that long-shot bets will win, or as in Kahneman and Tversky (1979), bettors might overweight small probabilities of winning when the potential payout is large (in calculating their utility). Bettors may also derive utility simply from the hope associated with holding a ticket on a long shot; not only is it more fun to pick a long shot to win over a favorite, it has more bragging rights. I have used this idea in consulting on the design of lotteries for the British Columbia Lottery Corporation and Singapore pools and British Lotto. Transaction costs related to the time cost of cashing tickets also play a role because bettors prefer collecting large payoffs rather than low payoffs. Also, some bettors may choose horses for irrational reasons, such as the name of the horse or its number. This behavior occurs in Hong Kong and the Kentucky Derby (see Ziemba and Hausch, 1987) and tends to flatten the favorite long-shot bias curve. The 2004 curve corresponding to the current favorite-longshot bias at US racetracks, the bottom line on the left in Figure 1, is somewhat flatter because of rebate and Internet long-shot betting (Ziemba, 2004). Even for the most favored horses there are expected losses. Then the returns are flat with expected returns of about 80% until the 20–1 area. Then the expected returns drops dramatically and are similar to those in the 1986 graph. See also the papers in Hausch and Ziemba (2007).

Table 2
Expected return per $1 bet vs. odds levels: Three-month options on S&P 500 futures, March 1985–September 2002

Call options					Put options				
Odds (%)	t Obs.	Average payoff	Std. dev. of payoff	t-test vs. $1	Odds (%)	t Obs.	Average payoff	Std. dev. of payoff	t-test vs. $1
0.95–1.00	47	1.0010	0.3204	0.02	0.95–1.00	37	0.8998	0.4493	−1.35*
0.90–0.95	60	1.0561	0.4605	0.96	0.90–0.95	44	0.8662	0.5872	−1.50*
0.85–0.90	66	1.1231	0.5704	1.76**	0.85–0.90	50	0.8426	0.7265	−1.53*
0.80–0.85	67	1.1407	0.6990	1.66**	0.80–0.85	54	0.7937	0.8120	−1.86**
0.75–0.80	63	1.0938	0.5953	1.25	0.75–0.80	53	0.8137	0.8950	−1.51*
0.70–0.75	64	1.1366	0.7732	1.41*	0.70–0.75	51	0.7879	0.9979	−1.51*
0.65–0.70	62	1.461	0.8648	1.33*	0.65–0.70	53	0.7702	0.9648	−1.73**
0.60–0.65	59	1.1311	0.9972	1.01	0.60–0.65	54	0.6215	1.0258	−2.70***
0.55–0.60	58	1.1727	1.1154	1.18	0.55–0.60	50	0.8225	1.2458	−1.01
0.50–0.55	54	0.9890	1.0410	−0.08	0.50–0.55	56	0.5807	1.1377	−2.76***
0.45–0.50	56	1.1365	1.3925	0.73	0.45–0.50	51	0.7344	1.4487	−1.21*
0.40–0.45	58	1.2063	1.6012	0.98	0.40–0.45	56	0.6785	1.5367	−1.57*
0.35–0.40	51	0.9770	1.7015	−0.10	0.35–0.40	56	0.4744	1.2383	−3.19***
0.30–0.35	54	0.9559	1.6041	−0.20	0.30–0.35	62	0.6357	1.6791	−1.76**
0.25–0.30	59	1.2923	2.7539	0.81	0.25–0.30	64	0.6316	1.8231	−1.62*
0.20–0.25	53	1.1261	2.5378	0.36	0.20–0.35	65	0.6426	1.9854	−1.45*
0.15–0.20	55	0.8651	2.0742	−0.48	0.15–0.30	64	0.6696	2.2441	−1.18
0.10–0.15	56	1.2262	3.6982	0.46	0.10–0.15	66	0.6602	2.6359	−1.05
0.05–0.10	53	1.5085	5.3370	0.69	0.05–0.10	66	0.6432	3.4256	−0.85
0.00–0.05	39	0.0123	0.1345	−44.89***	0.00–0.05	57	0.7525	5.6025	−0.33
All options	69	1.1935	2.4124	0.67	All options	69	0.6212	2.5247	−1.25

Source: Tompkins, Ziemba and Hodges (2003).
Note: When the hypothesis is rejected at a 90% level or above, the t-statistic appears in bold. t Obs = number of observations; Std. dev. = standard deviation.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Are the buyers of puts and calls on the S&P 500 futures similar in behavior to the race-track bettors? The demand for options comes from hedging and speculative investing. The primary use of put options is for hedging. Some demand also exists for speculative investing. The hedging demand for puts implies that the expected return is negative, and more so for deep out-of-the-money options. The main hedging demand for call options is for those selling them in covered-call strategies, which depresses their price. If this strategy were the sole mechanism for dealing in call options, it should result in an increase in the expected return for out-of-the-money call options, which Tompkins, Ziemba and Hodges (2003) do not observe. The expected loss from the purchase of deep out-of-the-money call options more likely results from speculative activity similar to that for the favorite long-shot bias.

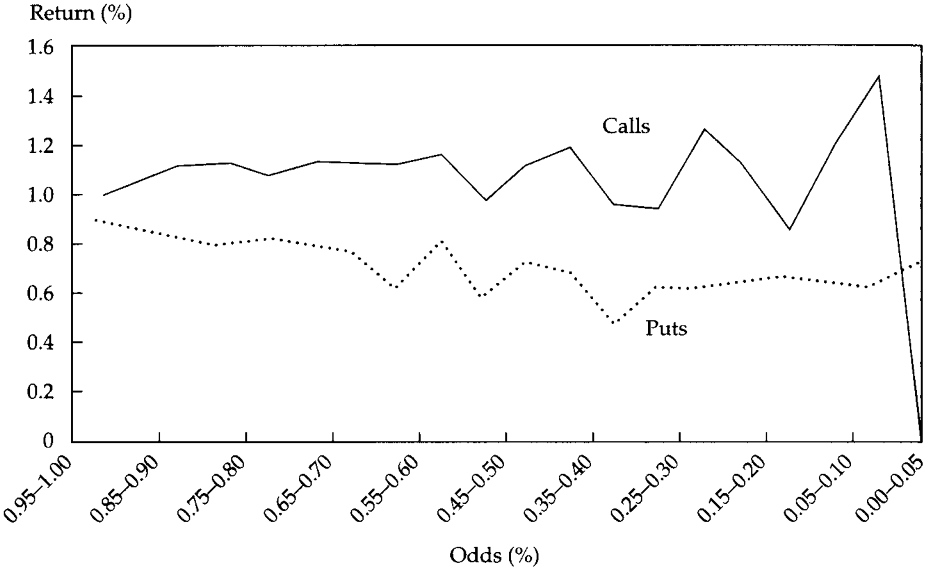


Fig. 2. Expected return per $1 bet vs. odds levels: Three-month calls and puts on S&P 500 futures, March 1985–September 2002. Source: Tompkins, Ziemba and Hodges (2003).

Tompkins, Ziemba and Hodges (2003) collected the settlement prices of one and three month put and call options on the S&P 500 index futures from the Chicago Mercantile Exchange, from March 1985 to September 2002. Table 2 and Figure 2 show their results for three-month options; the results for one month options are similar. A comparison of the racing and S&P 500 results shows that the probabilities equal the reciprocal of the odds $+1$.

In Table 2, the first column is the odds of finishing in the money, as measured by $N(d_2)$ or $N(-d_2)$ from the Black and Scholes (1973) formula. The next column is the number of options that fall into various 5% bands. The average payoff for a US$ 1 investment in that option band appears next, followed by the standard deviation of the option payoffs in that band. The final column is a one-sided t-test of the hypothesis that the mean return is equal to the initial investment of US$ 1.

For the three-month call options on S&P 500 futures, there is a favorite long-shot bias, as in horseracing. The in-the-money call options return more than the US$ 1 investment, on average. The at-the-money and out-of-the-month calls return about the US$ 1 investment. At the lowest level of 0–5%, a t-statistic could not be estimated because no call option in this range paid off during March 1985 to September 2002. This result confirms the hypothesis of Figlewiski (1989) that investors see out-of-the-money call options in the same way they see lottery tickets, and investors overpay for deep out-of-the-money call options on S&P 500 futures.

Table 3
Earthquake loss ratios per year

Year	Loss ratio	Year	Loss ratio	Year	Loss ratio
1971	17.4	1979	2.2	1987	22.8
1972	0.0	1980	9.2	1988	11.5
1973	0.6	1981	0.9	1989	129.0
1974	3.4	1982	0.0	1990	47.0
1975	0.0	1983	2.9	1991	17.2
1976	0.0	1984	5.0	1992	3.2
1977	0.7	1985	1.3	1993	3.2
1978	1.5	1986	9.3	1994	–

Source: Embrechts, Resnick and Samovodmitsky (1998).

No put options (on average) pay more than the US$ 1 initial investment. The average payoff decreases as the probabilities decrease, which is analogous to the horseracing favorite long-shot bias. This result is consistent with the contentions of Jackwerth and Rubinstein (1996) and Dumas, Fleming and Whaley (1998) that investors view put options as insurance policies and are willing to accept an expected loss to protect their equity holdings.

3.2. Extreme scenario examples

The damage measured by an index from earthquakes in California from 1971 to 1993 is shown in Table 3 and Figure 3. In these data, some years have zero damage, some have 5, and so on. The highest loss ratio is 129. The question is: How much earthquake damage occurred in California in the next year? Can we forecast the 1994 value? When I present this question in lectures, most people answer 10 or less. The 1989 peak of 129 and the 47 in 1990 are not considered in most observers' calculations. They look like outliers because the main probability mass is from 1972–86, when the maximum damage was less than 10.

The answer for 1994 is 2272.7. And the peak shown in Panel A of Figure 3 that was so high became very small in the next year. Hence, as shown in Panel B of Figure 3, extreme events can occur that are beyond the range of all previous events. There may have been earthquakes in California 400 years ago that were bigger than Northridge's (greater Los Angeles) in 1994, but few people and buildings were around then and the earthquakes could not destroy much. In Figure 3, the years 1989 and 1990 appear as similar to the 1972–86 years, and all the years 1971–93 appear to have similar values. What we have is an outcome way beyond the range of all past data. Many insurance companies in the United States declare bankruptcy on a regular basis as their business is essentially put selling which on average wins but can be disastrous if one overbets and is hit by a bad scenario.

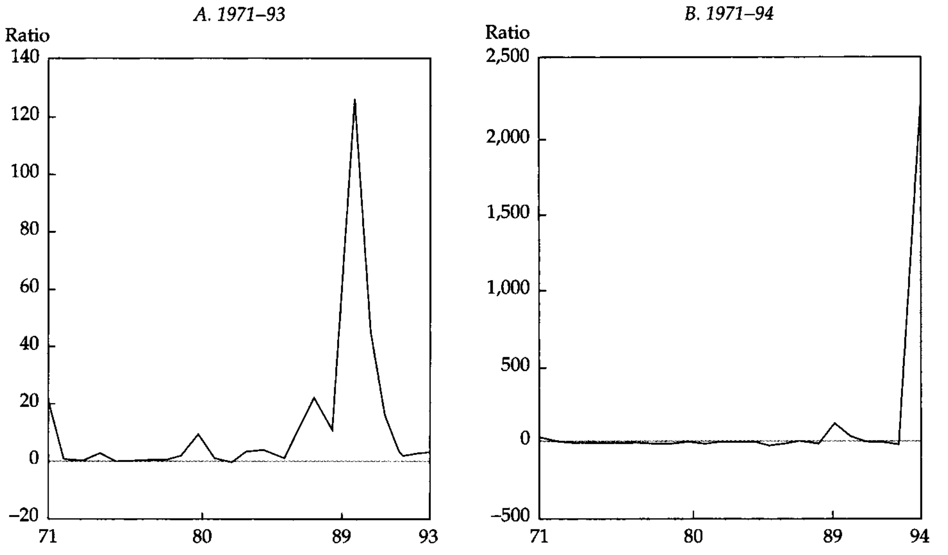


Fig. 3. Earthquake losses per year in California. Source: Ziemba (2003).

The number of such extreme events is increasing. Below is a list of events that occurred in 1998 that were beyond the range of the previous data. These rare, beyond-previous-data events are not so rare. Highly levered speculative investing has occurred for hundreds of years. But recently, more and more complex derivative instruments have become available, which is one of the causes of the growth of these rare events. The New York risk management consultants at Capital Markets Risk list "first-time" market events—events that conventional (not stochastic programming) risk-control models cannot foresee because the events have occurrences way beyond the range of previous historical occurrences. Normally, four or five such events occur each year; 77 occurred in 1998! Here are 17 of them:

- 18 May: Indonesia's rupiah collapses to 17,000 to the US dollar.
- 17 Aug: Russia defaults on ruble-denominated debt; ruble collapses by two-thirds.
- 31 Aug: The Dow plunges 512.61 points or 6.37% (on −1 day, the strongest trading day of the month).
- Jul/Sep: US banks suffer worst derivatives losses ever—US$ 445 million.
- 24 Sep: Hedge fund Long-Term Capital Management is bailed out with US$ 3.6 billion.
- 27 Sep: Japan Leasing Association files for bankruptcy with US$ 17.9 billion in liabilities, the biggest financial failure since World War II.
- 5 Oct: 30-year US Treasury yields hits record 4.74% low.
- 7 Oct: The US dollar plunges 7.8% against the yen, the largest one-day loss in 12 years.
- 8 Oct: China's yuan soars to an all-time high of 8.2777 to the US dollar.

- 9 Oct: Japan's Nikkei Index sinks to 11,542, the lowest since 1984.
- 13 Oct: London's FTSE 100 soars a record 214.2 points.
- 2 Nov: The US savings rate sinks to 0.2%.
- 5 Nov: Some leading Western banks cut yen deposit rates to negative values.
- 11 Nov: Shares of the globe.com skyrocket more than tenfold in the first day of trading.
- 30 Nov: US mortgage rates fall to 6.64%, the lowest since 1967.
- 3 Dec: Eleven European countries cut interest rates simultaneously.
- 10 Dec: World oil prices slide below US$ 10 a barrel, the lowest since 1986.

3.3. The long term capital management failure—what happens when you ignore possible scenarios

Of the many hedge fund failures, LTCM stands out as a particularly public one. The firm started with the talents of the core bond traders from John Merriwether's group at Salomon Brothers, who were successful for a number of years. When Warren Buffett came on board at Salomon, the culture of this group clashed with Buffett's apparently more conservative style, although in truth, Buffett's record is Kelly like and not all that different from Merriwether's group (see Thorp (2006) and Ziemba (2005)). A new group was formed with an all-star cast of top academics, including two future Economics Nobel Laureates and many top professors and students, many of whom were linked to the Massachusetts Institute of Technology. In addition, top government officials were involved. The team was dubbed as being "too smart to lose", and several billion was raised, despite the lack of a real track record; fees were high (25% of profits plus a 3% management fee), and the entry investment ($100 million minimum plus a lockup period) was also high. The idea, according to Myron Scholes, was to be a big vacuum cleaner, sucking up nickels all over the world.

There were many types of trades, but the essence of the bond risk arbitrage was to buy underpriced bonds in various locales and sell overpriced bonds in other locales and then wait for the prices to revert to their theoretical efficient market prices and unwind the position. These trades were similar to the Nikkei put warrant risk-arbitrage trade that Thorp and I did (see Shaw, Thorp and Ziemba, 1995). However, LTCM used much more leverage. I call such bond trades "buy Italy and sell Florence" trades The interest rate implied by the bond prices is higher in Italy than in Florence (Figure 4), but the theory is that Florence, a smaller city, would have more risk. Hence, the trade should have an advantage and be unwound once the prices reverted to their true risk-priced values.

LTCM analysts made many such trades, most of which were much more complex, all around the world. They also had many other complex and innovative trades. Their belief that markets were efficient and would snap back quickly when temporarily out of whack and the continuous lognormal assumptions of option pricing hedging led them to take large positions that, according to their theory, were close to riskless. The plan worked. Net returns for the part of 1994 that the fund operated were 19.9%. The fund also had superb results in 1995 and 1996, with net returns of 42.8 and 40.8%, respectively. In-

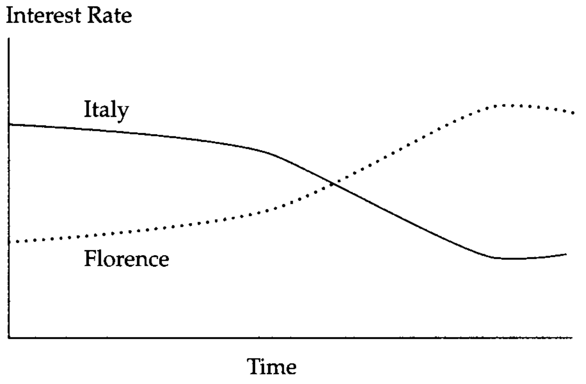


Fig. 4. "Buy Italy, sell Florence" trades. Source: Ziemba (2003).

deed for the principals whose money grew fee-less, the net returns were 63 and 57%, respectively, with taxes deferred. By 1997, however, it became harder to find profitable trades and the gains fell to 17.1%.

Although 1997 was a good record for most, it was not satisfactory to LTCM's principals. The action was to return US$ 2.7 billion of the US$ 6.7 billion investor money and to put in an additional US$ 100 million of personal loans to the principals from banks. Banks and most others were keen to loan to or invest with this group, and investors were not happy to leave the fund. The difficulties in 1998 were exacerbated by the 17 August Russian ruble devaluation and bond default. Russian bonds denominated in rubles trading for, say, 60 rubles fell rapidly to 3 rubles, whereas Russian bonds denominated in German marks or dollars fell only a few percent because they were not subject to default. So, long 60/short 95, say, became long 3/short 92.

Such losses occur from time to time in various markets, and hedge funds that overbet are vulnerable. LTCM had US$ 1.25 trillion in positions (that is, nearly 1% of the current (September 2006) value of the world derivatives and even a higher percentage in 1998) and US$ 125 billion in borrowed money, but although the trades were all over the world and hence seemed to be diversified, they were not. As a result, a scenario-dependent correlation situation occurred, such as that modeled in the Innovest pension application below. The underlying variable that frequently rears its ugly head in disasters—investor confidence—played a role. As shown in Figure 5, from August to October 1998, the difference in high yield bond rates minus US Treasury rates increased from roughly 4 to 6%. For example, emerging market debt was trading for 3.3% above US T-bonds in October 1997, then 6% in July 1998, and then an astounding 17% in September 1998.

LTCM was unable to weather the storm of this enormous crisis of confidence and lost about 95% of its US$ 4 billion, including most of the principals' and employees' considerable accumulated fees. The US$ 100 million loan put some of them into bankruptcy, although others came out better financially. It did not help that they unwound liquid positions first rather than across all liquidity levels, as the Nobel Laureates rec-

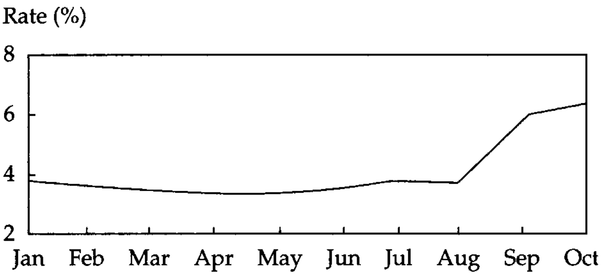


Fig. 5. Difference in yields between high-yield bonds and US Treasuries, 1998. Source: Based on data from Salomon Smith Barney.

ommended. Nor did it help that many other copycat firms had similar positions; that LTCM had created enemies by being so good and so brash; that the lack of monitoring of margin by brokers eager for business allowed the positions to grow to way overbet dangerous levels; and that the $2.7 billion was gone and they could not draw on it when it was most needed. MacLean and Ziemba (2006) on the Kelly criterion argue that investors should never bet more than the log optimal amount and betting more (as LTCM did) is stochastically dominated because of lower growth rates *and* higher risk. Smart people bounce back and learn from their mistakes, as has this group of traders with new hedge funds and other ventures. The lessons for the purposes of this chapter are:

- Do not overbet, it is too dangerous.
- VAR type systems are inadequate to measure true risk. A typical VAR implies that the trades will have a probability of less than 5% of losing $9 million or more. See Jorion (2000, 2006) on Var and Dunbar's (2000) discussion of the VAR calculations used by LTCM. LTCM analysts did a very careful analysis but the problem was that the risk control method of VAR which is used in regulations does not really protect hedge funds that are so highly levered because you are not penalized enough for large losses. Indeed if you lose $10 million it is penalized the same as losing $100 million if the VAR number is $9 million of losses. LTCM was not subject to VAR regulation but still used it. What you really need are convex penalties so that penalties are more than proportional to losses.
- Be aware of and consider extreme scenarios.
- Allow for extra illiquidity and contract defaults. LTCM also suffered because of the copycat firms which put on similar positions and unwound them at the same time in August/September 1998.
- Really diversify (to quote Soros, "we risked 10% of our funds in Russia and lost it, $2 billion, but we were still up 21% in 1998").
- Historical correlations work when you do not need them and fail when you need them in a crisis when they approach 1.0. Real correlations are scenario dependent.

Good information on the demise of LTCM and the subsequent US$ 3.5 billion bailout by major brokerage firms that was organized by the Fed are in Perold (1999), Jorion

(2000) and Edwards (1999). Eventually, the positions converged and the bailout team was able to emerge with a profit on its investment.

The Russian ruble devaluation of some two-thirds was no surprise to me. In 1992, my family and I were in St. Petersburg as guests of Professor Zari Rachev, an expert in stable and heavy-tail distributions (see Rachev, 2003). As we arrived, I gave him a $100 bill and he gave me six inches of 25 ruble notes. Our dinner out cost two inches for the four of us; and drinks were extra in hard currency. So, I am in the Soros camp: Make bets in Russia if you have an edge but do not risk too much of your wealth.

Where was the money lost? The score card according to Dunbar (2000) was a loss of $4.6 billion. Emerging market trades such as those similar to my buy Italy, sell Florence lost 430 million. Directional, macro trades lost 371 million. Equity pairs trading lost 306 million. Short long term equity options, long short term equity lost 1.314 billion. Fixed income arbitrage lost 1.628 billion.

The bad scenario of investor confidence that led to much higher interest rates for lower quality debt and much higher implied equity volatility had a serious effect on all the trades. The long-short equity options trades, largely in the CAC40 and Dax equity indices, were based on a historical volatility of about 15% versus implieds of about 22%. Unfortunately, in the bad scenario, the implieds reached 30% and then 40%. With smaller positions, the fund could have waited it out but with such huge levered positions, it could not. Equity implieds can reach 70% or higher as Japan's Nikkei did in 1990/1991 and stay there for many months.

Figlewski (1994), Ziemba (2003) and Ziemba and Ziemba (2007) discuss the general situation of how to lose money with derivatives. These include improper hedging, counter party (credit risk) default, overbetting through speculation, forced liquidation at unfavorable prices, and misunderstanding the risk exposure. Most of these amount to not diversifying in all scenarios and not over betting. If one truly diversifies and does not over bet, such disasters can usually be avoided. Let's now look at another overbetting, not diversified disaster.

3.4. The imported crash of October 27 and 28, 1997

A currency crisis developed in various Asian countries in mid 1997. It started in Thailand and moved all across the region. The problem was lack of foreign reserves that occurred because spending and expectations that led to borrowing were too high and Japan, the main driver of these economies, was facing a consumer slowdown so its imports dropped. Also loans were denominated in what was then considered a weak currency, the US dollar. So that these countries were effectively long yen and short dollars. A large increase in the US currency in yen terms exacerbated the crisis. The countries devalued their currencies, interest rates rose and stock prices fell. A well-known hedge fund failure in 1997 was Victor Niederhoffer's fund which had an excellent previous record with only modest drawdowns. A large long bet on cheap Thai stocks that became cheaper and cheaper turned $120 million into $70 million. Buying on dips added to losses. Then the fund created a large short position in out-of-the-money S&P 500 index futures puts. A typical position was November 830's trading for about $4–6 at various times around August–September 1997.

The crisis devastated the small economies of Malaysia, Singapore, Indonesia, etc. Finally it spread to Hong Kong. There, the currency was pegged to the US dollar at around 7.8. The peg was useful for Hong Kong's trade and was to be defended at all costs. The weapon used was higher interest rates which almost always lead to a stock market crash but with a lag. See the discussion below. The US S&P 500 was not in the danger zone in October 1997, by my models and I presume those of others, and the trade with Hong Kong and Asia was substantial but only a small part of the US trade. US investors thought that this Asian currency crisis was a small problem because it did not affect Japan very much. In fact, Japan caused a lot of it. I argue whenever there is a boom or bust in Asia, look to Japan and you will likely find much of the cause there: see Ziemba and Schwartz (1992).

The week of October 20–25 was a difficult one with the Hang Seng dropping sharply. The S&P was also shaky so the November 830 puts which were 60 cents on Monday, Tuesday and Wednesday but rose to 1.20 Thursday and 2.40 on Friday. The Hang Seng dropped over 20% in a short period including a 10% drop on Friday, October 25. The S&P was at 976 way above 830 as of Friday's close. A further 5% drop on Monday, October 27 in Hong Kong led to a panic in the S&P 500 futures later on Monday in the US. The fall was 7% from 976 to 906 which was still considerably above 830. On Tuesday morning there was a further fall of 3% to 876 still keeping the 830 puts out of the money. The full fall in the S&P 500 was then 10%.

But the volatility exploded and the 830's were in the $16 area. Refco called in Niederhoffer's puts mid morning. Refco took a loss of about $20 million. So, Niederhoffer's $70 million fund was bankrupt and actually in the red since the large position in these puts and other instruments turned the 70 million into minus 20 million. The S&P 500 bottomed out around the 876 area and moved violently in a narrow range then settled and then moved up by the end of the week right back to the 976 area. So it really was a tempest in a teapot. The November 830 puts expired worthless. Investors who were short equity November 830 puts were required to put up so much margin that forced them to have small positions and they weathered the storm and their $4–6, while temporarily behind at $16 did eventually go to zero. So did the futures puts, but futures shorters are not required to post as much margin so if they did not have adequate margin because they had too many positions. They could have easily been forced to cover at a large loss. Futures margins, at least for equity index products, do not fully capture their real risk inherent in these positions. None of the academic studies I know about on risk measures deals with this issue properly. When in doubt, always bet less. Niederhoffer is back in business but has had additional great gains and great losses; see Ziemba ans Ziemba (2007) for details.

One of my Vancouver neighbors, I learned later, lost $16 million in one account and $4 million in another account. The difference being the time given to cash out and cover the short puts. I was in this market also and won in the equity market and lost in futures. It did teach me how much margin you actually need in futures which I use now in such trading for myself and in private accounts which has been very profitable with a few wrinkles to protect oneself. One of the unhedged strategies won 78 of 79 times from December 1985 to March 2007. A hedged strategy had a 45% geometric mean with 74 of 79 winners with six ruled too risky by an option price danger control measure out of the 85 possible plays in those 21+ years and a seven symmetric downside Sharpe ratio; see Ziemba and Ziemba (2007).

The lessons for hedge funds are much as with LTCM. Do not overbet, do diversify, watch out for extreme scenarios. Even the measure to keep one out of potentially large falls (the 5 of 84 above) did not work in October 1997. That was an imported fear-induced crash not really based on US economics. My experience is that most crashes occur when interest rates relative to price earnings ratios are too high as discussed in the section here on the bond–stock model. A mini crash caused by some extraneous event can occur at any time. So to protect oneself, derivative positions must *never* be too large or hedged or with proper stop loss provisions. Ziemba and Ziemba (2007) discuss yet another hedge fund blowup, the 2006 Amarath natural gas disaster. Kouwenberg (2003) discusses the value added of hedge funds in investment portfolios.

4. Procedures for scenario generation

> One of the things capital markets do is consider possible worlds. The level and direction of prices reflect the markets' assessment of the probabilities of possible worlds becoming actual .... There are advocates for many of these views. Investors consider the risks and rewards and allocate their money accordingly.

So wrote Bill Miller, the famed manager of the Legg Mason Value Trust mutual fund (who beat the S&P 500 during each of the 15 years 1990–2005). He described scenarios in this way in reference to the book by Philosophy Professor David Lewis on the plurality of worlds, which argues that all possible worlds exist in the same sense that this world does.

There are many methods to estimate scenarios, I discuss the main ideas of some of them. Scenarios are a means to describe and approximate possible future economic environments. In my modeling applications, scenarios are represented as discrete probabilities of specific events. Together, all the scenarios represent the possible evolution of the future world. The basic idea is to have a set of T period scenarios of the form $S^{\mathrm{T}} = (S_1, S_2, \ldots, S_T)$, where $s_t \in S_t$ are the possible outcomes of all random problem elements and where s_t occurs in period t with probability $p_t(s_t)$.

A typical scenario tree is shown in Figure 6. S_1 has three possible outcomes, S_2 has three, and S_3 has two. There are 18 separate economic futures usually with different chances/probabilities of occurrence. Each can occur, and together, they approximate the possible future evolution of the economic environment relevant to the problem at hand. For asset/liability modeling, the most important parts of the distribution are the means and the left tail. The mean drives the returns, and the left tail, the losses. I cannot include all possible scenarios but rather focus on a discrete set that well approximates the

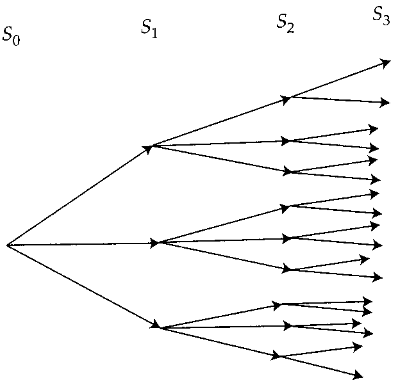


Fig. 6. Typical scenario tree.

possible important events that could happen. Because I have S^{T} total scenarios, I can include those I want. Once a scenario is included, the problem must react to what would be the consequences of that scenario. This important and flexible feature of stochastic programming modeling is not usually available in other approaches. The inclusion of such extreme scenarios means that the model must react to the possibility of that scenario occurring. This inclusion is one of the ways a stochastic programming overall model would have helped mitigate the 1998 losses and collapse of LTCM. The model would not have let them hold such large positions (see the discussion below). I frequently aggregate scenarios to pick the best N out of the S^{T} so that the modeling effort is manageable.

The generation of good scenarios that well represent the future evolution of the key parameters is crucial to the success of the modeling effort. Scenario generation, sampling, and aggregation is a complex subject, and I will discuss it by describing key elements and then provide various developed and implemented models.

Scenarios should consider the following, among other things:

- mean reversion of asset prices; see Figure 12 below;
- volatility clumping, in which a period of high volatility is followed by another period of high volatility;
- volatility increases when prices fall and decreases when they rise;
- trending of currency, interest rates, and bond prices;
- ways to estimate mean returns;
- ways to estimate fat tails; and
- ways to eliminate arbitrage opportunities or minimize their effects.

The true distribution P is approximated by a finite number of points $(w^1, \ldots, w^S)$ with positive probability p^s for each scenario s. The sum of all scenario probabilities is 1.

Scenarios come from diverse sources and are used in many applications. They can come from a known discrete probability distribution or as the approximation of a continuous or other probability distribution that is estimated from past data, economic forecasting models, or comparison with similar past events. The latter is especially useful for situations that have never occurred, for example, if there is a potential crisis in Brazil whose effects must be estimated and no data or models are available but similar crises have occurred in Russia, where the effects have been well estimated. These data are useful for scenario estimation, especially for disastrous scenarios.

Economic variables and actuarial predictions drive the liability side, whereas economic variables and sentiment drive financial markets and security prices. Hence, estimating scenarios for liabilities may be easier than for assets because there often are mortality tables, actuarial risks, legal requirements, such as pension or social security rules, well-established policies, and so on. (See Embrechts, 2000.) Such scenarios may come from simulation models embedded into the optimization models that attempt to model the complex interaction between the economy, financial markets, and liability values. Examples include Goldstein and Markowitz (1982); Kingsland (1982); Winklevoss (1982); and Boender (1997).

Abaffy et al. (2000) and Dupačovà, Consigli and Wallace (2000) survey scenario estimation and aggregation methods that represent a larger number of scenarios by a smaller number. See also Jamshidian and Zhu (1977); Birge and Louveaux (1997); Jarrow and van Deventer (1998); Edirisinghe (1999); Kouwenberg and Zenios (2007); and Wallace and Ziemba (2005). There are also technical articles by Hochreiter and Pflug (2003); Dupačovà, Gröwe-Kuska and Römisch (2003); Pflug (2001); Casey and Sen (2005); Pennanen and Koivu (2002); Heitsch and Römisch (2003); Edirisinghe and Ziemba (1992, 1994a, 1994b); Frauendorfer (1996); Birge and Wets (1986). We can use the following classification:

1. There can be full knowledge of the exact probability distribution. This knowledge usually comes from a theoretical model, but it is possible to use historical data or an expert's experience.
2. There can be a known parametric family based on a theoretical model whose parameters are estimated from available and possibly forecasted data. For example, much literature exists for scenario generations using the Vasicek (1977), the Heath–Jarrow–Morton (1992), and the Black–Derman–Toy (1990) and other interest rate models for interest rate, fixed-income, and bond portfolio management. For example, the prices of T-bonds can be computed on a lattice subject to the initial yield curve. Then, the prices of other relevant interest-rate-dependent securities can be estimated. Also, stochastic differential equation modeling can be used to generate scenarios for asset returns and liability commitments by using a cascade of models that feed one into another. See, for example, Jamshidian and Zhu (1977); Chan et al. (1992); and the Towers Perrin scenario-generation system based on Mulvey (1996) and Mulvey and Thorlacius (1998). Methods used to evaluate value at risk can also be used to create scenarios because they estimate probability distributions: see Duffie and Pan (1997) who also discuss the complex issue of "how many scenarios is enough". Examples include Jorion (2006) and Jamshidian and Zhu (1977), who estimate market

and currency risk. See also Jobst and Zenios (2003) and Duffie and Singleton (2003) who estimate market and credit risk.

3. Scenarios can be formed by sample-moment information that aggregates large numbers of scenarios into a smaller, easier set or generates scenarios from assumed probability distributions. This idea was used in the five period Russell-Yasuda Kasai insurance model in 1989 which is discussed below; see also Cariño, Myers and Ziemba (1998). Smith (1993); Keefer and Bodily (1983); Keefer (1994) have suggested such methods for static problems. Høyland and Wallace (HW) (2001) have expanded and refined the idea for multiperiod problems and they generate scenarios that match some moments of the true distribution. Typically it is the first four moments. Extension of the HW work are in Høyland, Kaut and Wallace (2003) and Kouwenberg (1999). The procedure is easy to use and provides adequate scenarios. It does have its faults because the two very different looking distributions can have the same first four moments. See Hochreiter and Pflug (2003) for an example using ideas from Heyde (1963) who showed that even infinite moment matching will not replicate all distributions. Such mathematical mismatches aside, the HW approach is to find a set of discrete scenarios that best fits the distributions first four moments minimizing least squares. The scenario tree consists of realizations and their probabilities of all random parameters in all time periods.
4. The simplest idea is to use past data that reflect comparable circumstances and assign them equal probabilities. This idea can be implemented by using the raw data or through procedures, such as vector autoregressive modeling or bootstrapping, that sample from the past data. Using subsets of past data is a viable scenario generation method. Grauer and Hakansson (1998) (see also Mulvey and Vladimirou (1992)) have used the idea in many asset-only studies. Their idea is to assume that the past will repeat in the sense that events that occurred with various probabilities will reoccur with the same probabilities but not necessarily in the same order. Hence, some of the dependence of returns is not preserved with this approach. Another objection to the approach is that it probably will miss fat tail events and, of course, those that have never occurred yet. Still Grauer and Hakansson have some very good results with this simple approach using log and negative power utility functions to determine the optimal asset weights.
5. When no reliable data exist, one can use an expert's forecasts (examples include Markowitz and Perold (1981) and Shapiro (1988) or governmental regulations. Abaffy et al. (2000) pointed out that to test the surplus adequacy of an insurer, New York State Regulation 126 suggests seven interest rate scenarios to simulate the performance of the surplus. Liability commitments are frequently easier to estimate than assets because of demographic data, regulations, and so on.
6. The Hochreiter and Pflug (HP) (2003) approach generates scenario trees using a multidimensional facility location problem that minimizes a Wasserstein distance measure from the true distribution. Kolmogorov–Smirnov distance approximations is an alternative approach but it does not take care of tails and higher moments. The HP method combines a good approximation of the moments and the tails. An

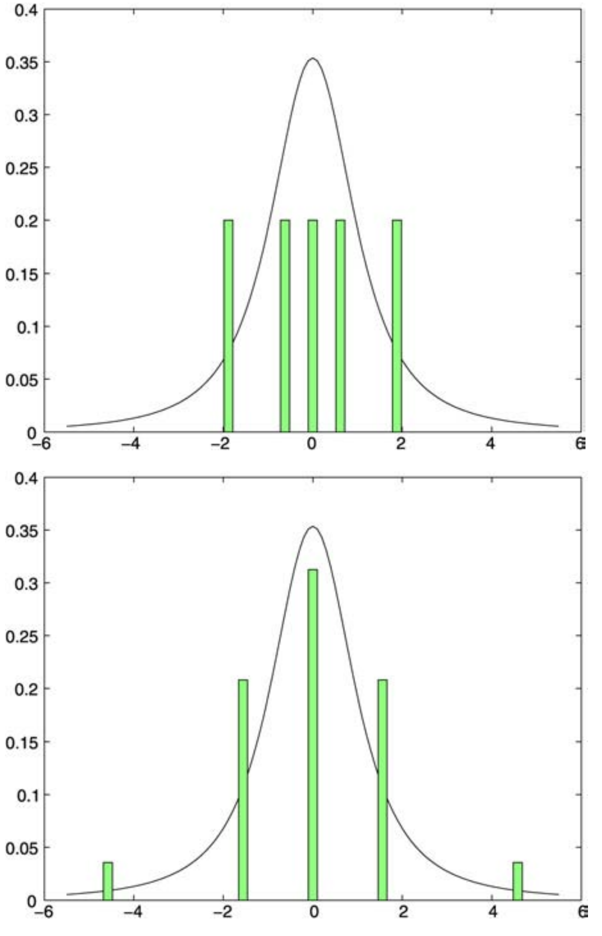


Fig. 7. Approximation of a t-distribution with a Kolmogorov–Smirnov distance (top) and the Wasserstein distance measure (bottom).

example is the t-distribution with two degrees of freedom with density $(2+x^2)^{-3/2}$. This density is approximated in Figure 7 with the KS distance on the left and the Wasserstein distance on the right. The latter gives better approximations, the location of the minimum is closer to the true value and the mean is closer as well. The KS distance is invariant with respect to monotone transformations of the x-axis and thus does not approximate the tails well like the Wasserstein distance.

The optimal approximation of a continuous distribution G by a distribution with mass points $z_1, \ldots, z_m$ with probabilities $p_1, \ldots, p_m$ using the KS distance is

$$z_i = G^{-1}\left(\frac{2i-1}{2m}\right), \qquad p_i = \frac{1}{m}.$$

Table 4
Summary of inputs for scenario analysis

Historical data
Macroeconomic factors, such as yield curves, credit spreads, dividend yields (and their growth and earnings forecasts), and currency-exchange values
Expert judgment
Sentiment and extraneous factors
Mean and tail factors
Asset class forecasts, such as cash, equities in various sectors and countries, and interest and bond yields
Scenarios

The Wasserstein distance is the solution of the non-convex program

$$\min\left\{\int \min|u - c_j|\,\mathrm{d}G(u)\colon\ c_1, \ldots, c_m \in R\right\}.$$

The locations $c_1, \ldots, c_m$ then yield the probabilities

$$p_i = \int_{\{u\colon\ |u-c_i|=\min|u-c_j|\}} \mathrm{d}G(u).$$

Figure 7 shows a typical application of the HP approach. Observe how few scenarios approximate the distribution.

The list depicted in Table 4 is typical. Whatever method is used to generate the scenarios, in relying on the meshing of decision-maker subjective estimates, expert judgment, and empirical estimation, it is crucial to validate the estimated distributions and to make sure that the decision maker has not defined the range too narrowly. Perhaps reflecting on the distribution by asking what would make the value be outside the range and then assessing the probability would help expand the range and make the probability assessment more realistic.

4.1. The bond–stock return danger model

One aspect of scenario generation is to avoid disasters and the bond–stock model help. During 1988–89 I was a visiting professor of finance at the University of Tsukuba in the Japan and consultant to the Yamaichi Research Institute. It was an extraordinary experience and got me on the right track to learn markets better. Much of what I learned appears in Ziemba and Schwartz (1991) and Stone and Ziemba (1993). My wife, Sandra Schwartz, and I also wrote *Power Japan* (1992), which discusses the Japanese economy.

In the study group, I came up with a simple stock market crash model in 1988 with only a single variable—the difference between stock and bond rates of return. The idea was that stocks and bonds compete for investment dollars, and when interest rates are

Table 5
S&P 500, P/E, Government bond yield and the yield premium over stocks, January 1984–August 1988

Month	S&P 500	P/E	30-year govt. bond (*a*)	1/(P/E) (*b*)	(*a*) − (*b*)
1986					
Jan	208.19	14.63	9.32	6.84%	2.48
Feb	219.37	15.67	8.28	6.38	1.90
Mar	232.33	16.50	7.59	6.06	1.53
Apr	237.98	16.27	7.58	6.15	1.43
May	238.46	17.03	7.76	5.87	1.89
Jun	245.30	17.32	7.27	5.77	1.50
Jul	240.18	16.31	7.42	6.13	1.29
Aug	245.00	17.47	7.26	5.72	1.54
Sep	238.27	15.98	7.64	6.26	1.38
Oct	237.36	16.85	7.61	5.93	1.68
Nov	245.09	16.99	7.40	5.89	1.51
Dec	248.60	16.72	7.33	5.98	1.35
1987					
Jan	264.51	15.42	7.47	6.49	0.98
Feb	280.93	15.98	7.46	6.26	1.20
Mar	292.47	16.41	7.65	6.09	1.56
Apr	289.32	16.22	9.56	6.17	3.39
May	289.12	16.32	8.63	6.13	2.50
Jun	301.38	17.10	8.40	5.85	2.55
Jul	310.09	17.92	8.89	5.58	3.31
Aug	329.36	18.55	9.17	5.39	3.78
Sep	318.66	18.10	9.66	5.52	4.14
Oct	280.16	14.16	9.03	7.06	1.97
Nov	245.01	13.78	8.90	7.26	1.64
Dec	240.96	13.55	9.10	7.38	1.72
1988					
Jan	250.48	12.81	8.40	7.81	0.59
Feb	258.10	13.02	8.33	7.68	0.65
Mar	265.74	13.42	8.74	7.45	1.29
Apr	262.61	13.24	9.10	7.55	1.55
May	256.20	12.92	9.24	7.74	1.50
Jun	270.68	13.65	8.85	7.33	1.52
Jul	269.44	13.59	9.18	7.36	1.82
Aug	263.73	13.30	9.30	7.52	1.78

low, stocks are favored, and when interest rates are high, bonds are favored. The main thing that I wished to focus on was that when the measure—the difference between these two rates, the long bond yield minus the earnings yield (the reciprocal of the P/E ratio)—was very large, then there was a high chance of a stock market crash. A crash is a 10% fall in the index within one year from the current index value. The model explains

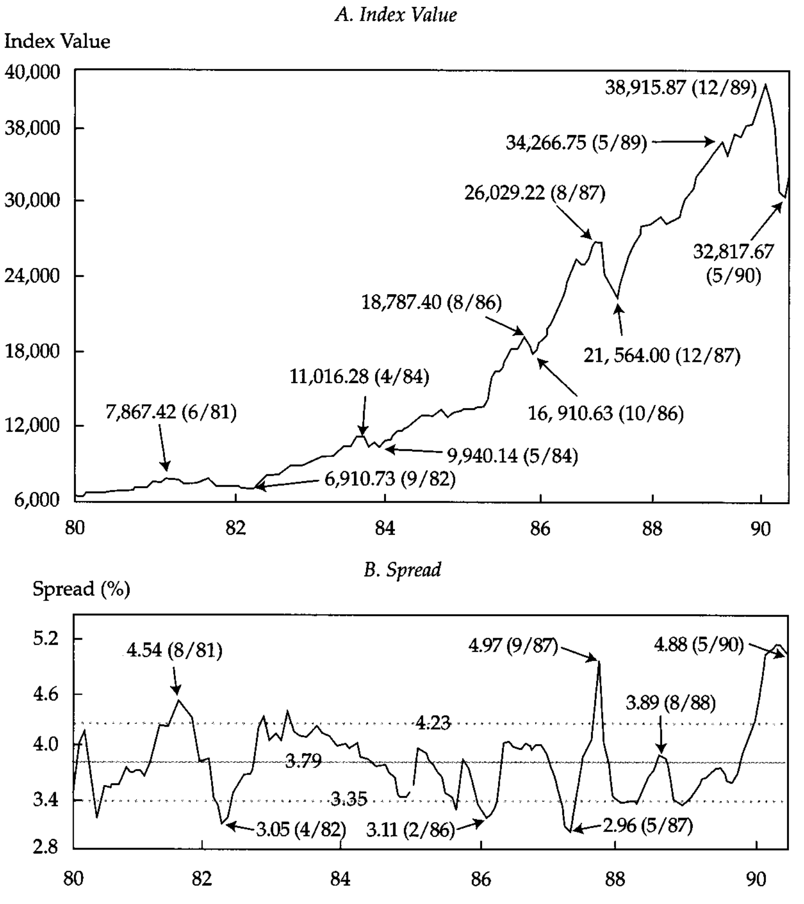


Fig. 8. Bond–stock yield differential model for the Nikkei stock average, 1980–90. Note: Data through 29 May 1990. Shaded lines in Panel B denote upper limit, mean, and lower limit. Source: Ziemba (2003).

the October 1987 crash. That application is how this idea came to me. Table 5 shows this relationship for the US market. The bold numbers in this table indicate the existence of extreme danger in the stock market because 30-year government bond yields were much higher than the stock market yields, as measured by the reciprocal of the last year's reported P/E. These high interest rates invariably lead to a stock market crash. Here, the danger indicator moved across a statistical 95% confidence line in April 1987. The market ignored this signal and moved higher but did eventually crash in October 1987. Most investors ignored a similar signal in the S&P 500 in April 1999, and then a crash began in August 2000 and a weak stock market ensued in 2001–2002, which is discussed below. For a study of this measure from 1970 to 2005 in five major markets, see Berge and Ziemba (2006) and Berge, Consigli and Ziemba (2007).

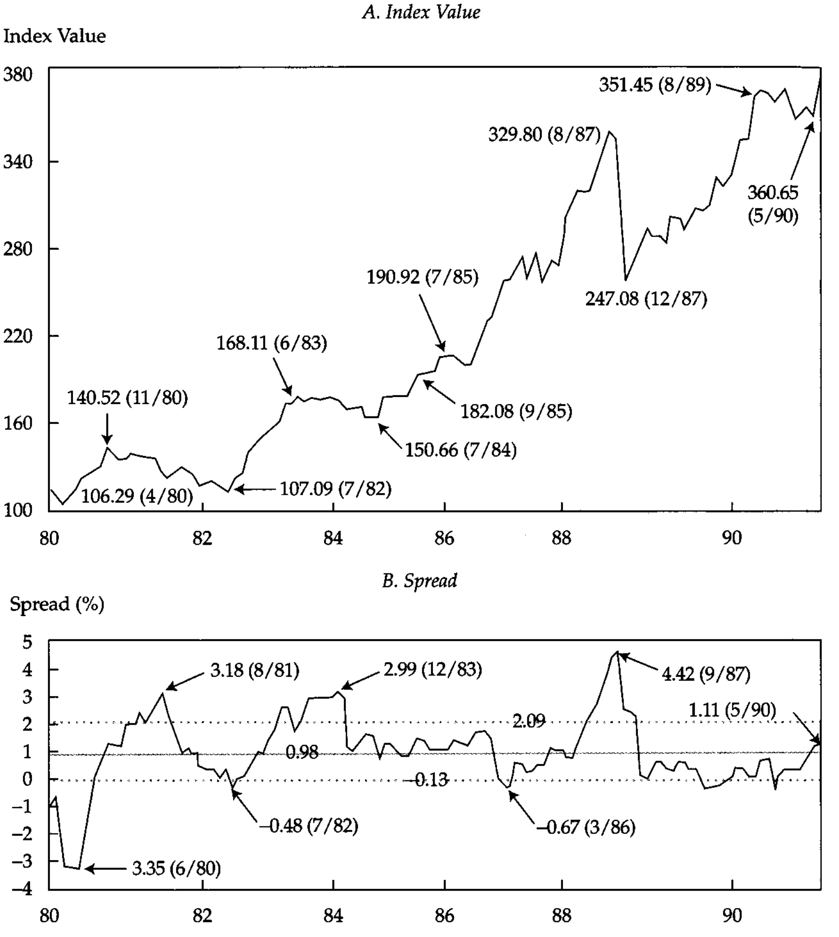


Fig. 9. Bond–stock yield differential model for the S&P 500, 1980–1990. Note: Data through 29 May 1990. Shaded lines in Panel B denote upper limit, mean, and lower limit.

In 1988–1989, I asked one of my colleagues in the study group, Sugheri Iishi, to help me check this measure for Japan, as shown in Figure 8 from 1980 to mid-1990. Twenty 10+ percent crashes occurred during the 1949–89 period. We found that whenever this measure was in the danger zone (that is, outside a 95% confidence band), within one year, a crash of 10% or more from the current level would occur. Not all crashes had the measure in the danger zone, but whenever it was, there was a crash with no misses (12 of 12). Eight of the 10+ percent declines occurred for other reasons.

So, the measure was successful at predicting future crashes—but there was no precise way to know when they would occur and how deep they would be. Long-run mean reversion, however, suggests that the longer the bull run is and the more overpriced the

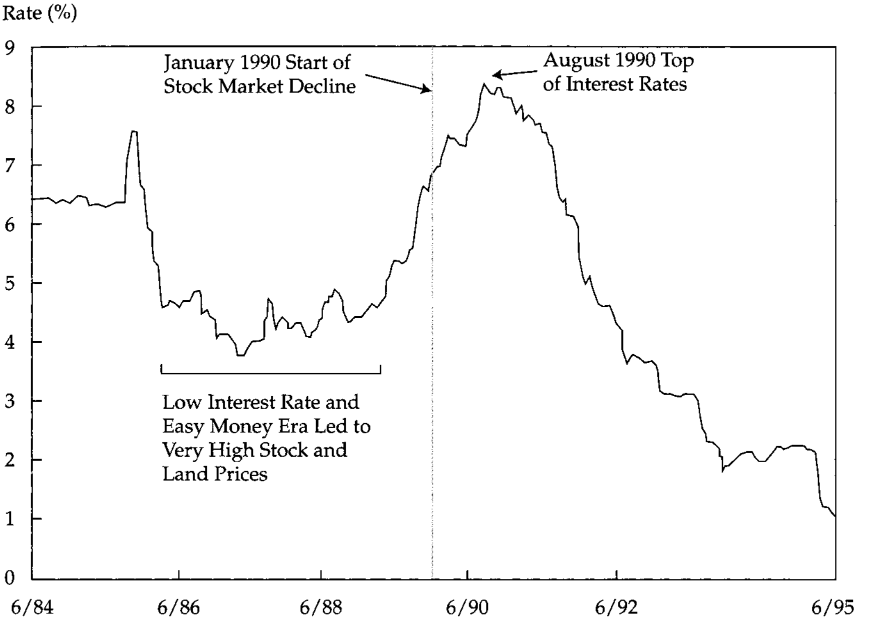


Fig. 10. Short-term interest rates in Japan, June 1984–June 1995.

measure is, the longer and deeper the decline will probably be. Then, one can use the measure as part of an econometric system to estimate future scenarios.

Each time the spread exceeded the 4.23% cutoff (which was higher than 95% confidence), a crash would occur. The measure was way in the danger zone in late 1989, and the decline (the 21st crash) began on the first trading day of 1990, with the Nikkei Stock Average (NSA) peaking at 38,916 (see Figure 8). Unfortunately, Yamaichi's top management did not listen to Iishi when I sent him up to explain our results in Japanese; there was much greater danger in the market than they thought in 1989. By 1995, Yamaichi Securities had declared bankruptcy and had ceased to exist.

The model also indicates that the valuation was still high as of 29 May 1990. Not much later, the 22nd crash began. Interestingly, at the bottom of the 22nd crash on 1 October 1990, the Nikkei Stock Average was at 20,222, which was almost exactly the mean. Meanwhile, the same calculation on 29 May 1990 for the S&P 500, as shown in Figure 9, shows that the US market was cheap—that is, below the mean—following the September 1987 peak of 4.42%. The 29 May 1990 value of 1.11% was, however, slightly above the mean level and the highest since the late fall of 1987.

Japan had weak stock and land markets for more than ten years since the beginning of 1990. This situation was been caused by many factors, political as well as economic. But the rising interest rates for eight full months leading up to August 1990 (see Figure 10) were one of them. This extreme tightening of an overlevered economy was too much. Cheap and easily available money, which caused the big run-up in asset prices in the 1980s, turned into expensive and unavailable money in the 1990s.

4.2. The 2000–2003 crash in the S&P 500

The S&P 500 was 470.42 at the end of January 1995. It was about 750 in late 1996, at the time of Alan Greenspan's famous speech on irrational exuberance in the US stock market. The closing value peaked at 1527.46 (1552.87 intraday) on 24 March 2000, fell to 1356.56 on 4 April, and then came close to this peak, reaching 1520.77 (1530.09 intraday) on 1 September, the Friday before Labor Day. The bond–stock crash model was in the danger zone virtually all of 1999, and it got deeper in danger as the year progressed and the S&P 500 rose from 1229.23 at the end of December 1998 to 1469.25 at the end of December 1999. The P/E ratio was flat, increasing only from 32.34 to 33.29, while long-bond yields rose from 5.47 to 6.69. The S&P 500 fell to 1085 on 17 September 2000.

Table 6 and Figure 11 detail this progression from January 1995 to December 1999. The spread reached 3, which was well in the 95% confidence danger band in April 1999 and rose to 3.69 in December 1999. The stage was set for a crash, which did occur, as shown in Panel A of Figure 12. Long-term mean reversion indicates that the 1996–2000 S&P 500 values were too high relative to 1991–95, and a linear interpolation of the latter period gives a value close to that in May 2003. The model for Japan was hard to interpret because of high P/E ratios, but interest rates were close to zero. One had a close to 0–0 situation, so the model did not seem to apply to Japan in 1999. The model was not in the danger zone with return differences close to zero (see Figure 13).

There was a dramatic fall in the S&P 500 from its peak of 1527.46 in March 2000 to its September 2000 low of 1085. Further declines occurred in 2001 and 2002. The lowest close was 768.63 on 10 October 2002. This decline was similar to previous crashes. There were other signals.

History shows that a period of shrinking breadth is usually followed by a sharp decline in stock values of the small group of leaders. Then the broader market takes a more modest tumble. (Bagnell, (1999, Business 1))

The Toronto Stock Exchange (TSE 300) rose during 1999 and 2000 and fell in 2001 and 2002 similar to the US markets. During 1999, the TSE 300 gained 31%, but the gain was only 3% without three very high P/E, large-cap stocks and one of them was mostly shares of one of the other two. The largest gainer in market value, Nortel Networks peaked at US$ 120 and was about $1.70 at the end of 2002 and about $2.50 per share in early May 2007 (adjusted for its 1–10 reverse split and a price about $25).

The concentration of stock market gains into very few stocks, with momentum and size being the key variables predicting performance, was increasing before 1997 in Europe and North America. Table 7 for 1998 shows that the largest-cap stocks had the highest return in North American and Europe but small-cap stocks outperformed in

Table 6
Bond–stock yield differential model for the S&P 500, 1995–1999

Year/month	S&P 500	P/E (a)	30-yr. govt. bond (b)	Return on stocks ($c = 1/a$)	Crash signal ($b - c$)
1995					
Jan	470.42	17.10	8.02%	5.85%	2.17
Feb	487.39	17.75	7.81	5.63	2.18
Mar	500.71	16.42	7.68	6.09	1.59
Apr	514.71	16.73	7.48	5.98	1.50
May	533.40	16.39	7.29	6.10	1.19
Jun	544.75	16.68	6.66	6.00	0.66
Jul	562.06	17.23	6.90	5.80	1.10
Aug	561.88	16.20	7.00	6.17	0.83
Sep	584.41	16.88	6.74	5.92	0.82
Oct	581.50	16.92	6.55	5.91	0.64
Nov	605.37	17.29	6.36	5.78	0.58
Dec	615.93	17.47	6.25	5.72	0.53
1996					
Jan	636.02	18.09	6.18%	5.53%	0.65
Feb	640.43	18.86	6.46	5.30	1.16
Mar	645.50	19.09	6.82	5.24	1.58
Apr	654.17	19.15	7.07	5.22	1.85
May	669.12	19.62	7.21	5.10	2.11
Jun	670.63	19.52	7.30	5.12	2.18
Jul	639.96	18.80	7.23	5.32	1.91
Aug	651.99	19.08	7.17	5.24	1.93
Sep	687.31	19.65	7.26	5.09	2.17
Oct	705.27	20.08	6.95	4.98	1.97
Nov	757.02	20.92	6.79	4.78	2.01
Dec	740.74	20.86	6.73	4.79	1.94
1997					
Jan	786.16	21.46	6.95%	4.66%	2.29
Feb	790.82	20.51	6.85	4.88	1.97
Mar	757.12	20.45	7.11	4.89	2.22
Apr	801.34	20.69	7.23	4.83	2.40
May	848.28	21.25	7.08	4.71	2.37
Jun	885.14	22.09	6.93	4.53	2.40
Jul	954.29	23.67	6.78	4.22	2.56
Aug	899.47	22.53	6.71	4.44	2.27
Sep	947.28	23.29	6.70	4.29	2.41
Oct	914.62	22.67	6.46	4.41	2.05
Nov	955.40	23.45	6.27	4.26	2.01
Dec	970.43	23.88	6.15	4.19	1.96
1998					
Jan	980.28	24.05	6.01%	4.16%	1.85
Feb	1,049.34	25.09	6.00	3.99	2.01
Mar	1,101.75	27.71	6.11	3.61	2.50

(*continued on next page*)

Table 6
(*continued*)

Year/month	S&P 500	P/E (a)	30-yr. govt. bond (b)	Return on stocks ($c = 1/a$)	Crash signal ($b - c$)
1998					
Apr	1,111.75	27.56	6.03	3.63	2.40
May	1,090.82	27.62	6.10	3.62	2.48
Jun	1,133.84	28.65	5.89	3.49	2.40
Jul	1,120.67	28.46	5.83	3.51	2.32
Aug	97.28	27.42	5.74	3.65	2.09
Sep	1,017.01	26.10	5.47	3.83	1.64
Oct	1,098.67	27.41	5.42	3.65	1.77
Nov	1,163.63	31.15	5.54	3.21	2.33
Dec	1,229.23	32.34	5.47	3.09	2.38
1999					
Jan	1,279.64	32.64	5.49%	3.06%	2.43
Feb	1,238.33	32.91	5.66	3.04	2.62
Mar	1,286.37	34.11	5.87	2.93	2.94
Apr	1,335.18	35.62	5.82	2.79	3.03
May	1,301.84	34.60	6.08	2.89	3.19
Jun	1,372.71	35.77	6.36	2.80	3.56
Jul	1,328.72	35.58	6.34	2.81	3.53
Aug	1,320.41	36.00	6.35	2.78	3.57
Sep	1,282.70	30.92	6.50	3.23	3.27
Oct	1,362.92	31.61	6.66	3.16	3.50
Nov	1,388.91	32.24	6.48	3.10	3.38
Dec	1,469.25	33.29	6.69	3.00	3.69

Asia and Japan. The situation was similar from 1995–1999, with 1998 and 1999 being the most exaggerated.

The influential book *Irrational Exuberance* by Robert Shiller (2000), a behavioral finance economist at Yale University, hit the market in April 2000. It was a monumental success in market timing, with an especially bearish view that is consistent with Figure 11 and Table 7. Shiller's data, as shown in Figure 14, document very high P/Es in relation to earnings in 2000, with most of the rise in the 1995–2000 period, similar to Table 6 for the S&P 500. He demonstrated that in 1996–2000, the stock market was overpriced relative to historical norms measured by PE ratios and dividend yields.

Shiller had been predicting a crash since 1996, as reported in Campbell and Shiller (1998). He remained defiantly bearish, as in Campbell and Shiller (2001), which is an update of the 1998 paper. His case has been helped by three largely unpredictable bad scenarios; the 9/11 attacks on the United States, the June/July 2002 crises of accounting confidence in the United States, and the 2003+ US/British war with Iraq. One could argue that the second bad scenario was a direct consequence of the 1996–1999 bubble period, as did the *Economist* (2002) and interviews of Warren Buffett and Peter Drucker

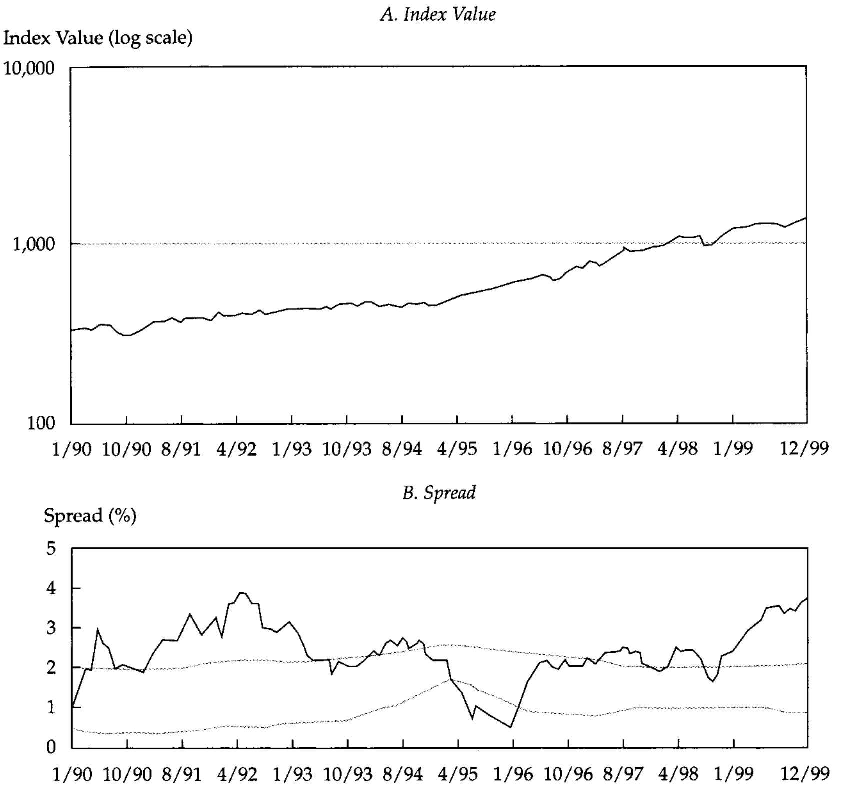


Fig. 11. Bond–stock yield differential model for the S&P 500, January 1990–December 1999. Note: Shaded lines in Panel B denote upper and lower limit. Source: Ziemba (2003).

in 2002. The latter two sages saw this scenario coming a long time ago. Also, the third bad scenario was a consequence of the first.

This chapter is about the stochastic programming approach to asset, liability, and wealth management, so the main points most relevant for my analysis are:

- We should be able to use such measures as the bond–stock return difference and such research as Shiller's figures to create better scenarios. I argue that the mean is by far the most important aspect of return distributions. Figure 15 for the S&P 500 in 2000–2002 serves as a reminder of this finding.
- The extent of such danger measures also suggests that the entire distribution from which scenarios are drawn should be shifted left toward lower and more volatile returns. We know that volatility increases as markets decline. Koivu, Pennanen and Ziemba (2005) show one way to create such *better* scenarios.
- The evidence is high that stocks outperform bonds, T-bills, and most other financial assets in the long run (see Siegel, 2002; Dimson, Marsh and Staunton, 2002, 2006,

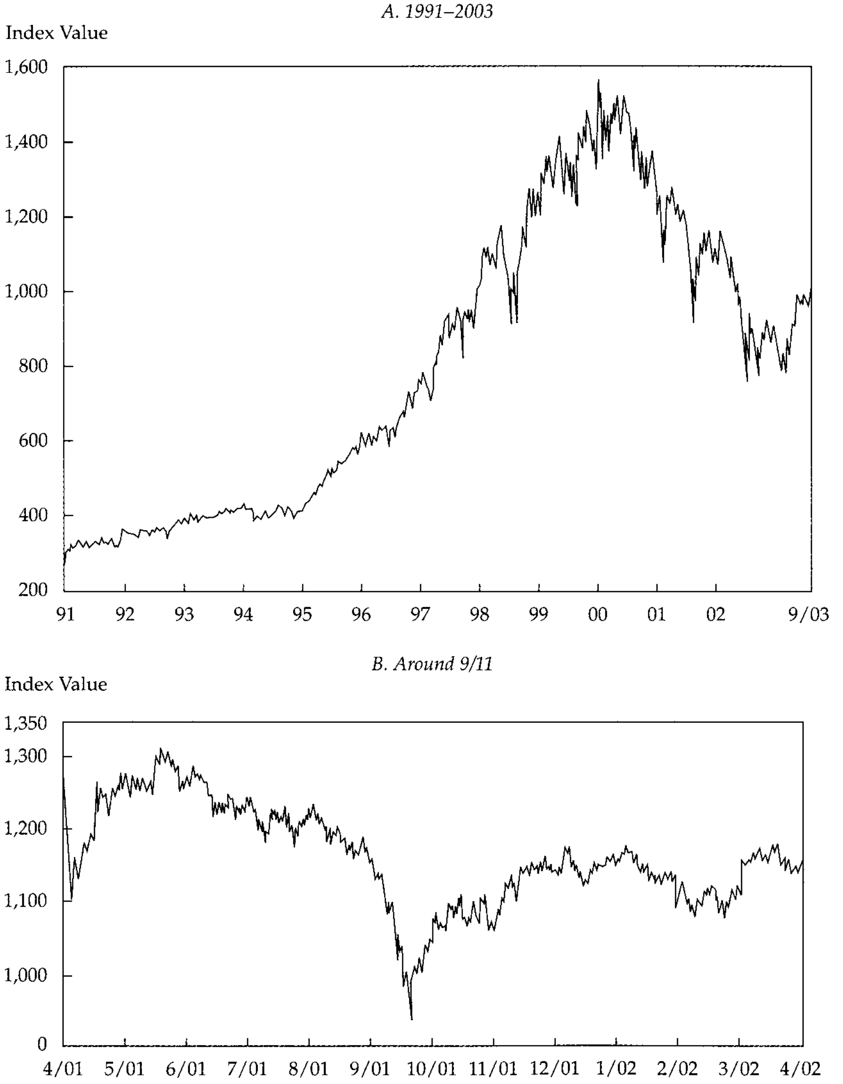


Fig. 12. The S&P 500 mean reversion and around September 11, 2001.

2007; and Table 8 and Figures 16 and 17). Stocks generally outperform in times of inflation and bonds outperform in times of deflation (see, e.g., Smith, 1924). Why do stocks generally outperform bonds? As has been said, "A major reason is that businesses retain earnings, with these going onto create more earnings and dividends too." From the review of Smith by J.M. Keynes in 1925, quoted in Buffett (2001). In times of growth, firms borrow at fixed cost with the expectation of earning positive

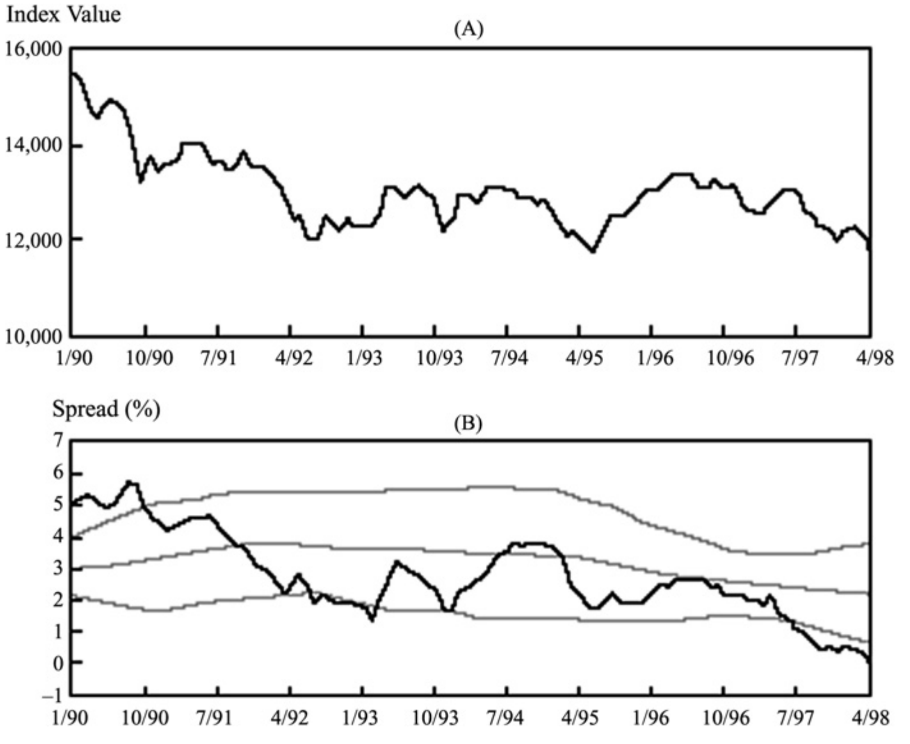


Fig. 13. Nikkei stock average and bond–stock yield differential model for 10-Year Government Bonds, January 1990–April 1998. (A) Index value. (B) Spread. Note: Shaded lines in Panel B denote upper limit, mean, and lower limit. Bond–stock yield differential mean and standard deviation calculated for October 1994–August 1998.

economic profit, so in the long term, equities, as a reflection of this positive income creation, should grow at the rate of productivity.

- Occasionally, stocks underperform alternative asset classes for long periods. Figure 17 shows this phenomenon for the DJIA from 1885 to 2001 in 2001 dollars, and Figure 15 shows the 2000–02 period for the S&P 500 and US Government bonds. When bonds outperform stocks, as in this latter period, they are usually negatively correlated with stocks as well; see Figure 18, which has rolling monthly correlations. Between 1982 and 1999, the return of equities over bonds was more than 10% a year in European Union countries. The question is whether we are moving back to a period where the two asset classes move against each other or whether the phenomenon will prove to be temporary. Moreover, the historical evidence since 1802 for the United States and 1700 for the United Kingdom indicates that the longer the period, the more likely this dominance will occur. Siegel (2002) showed that in all 20-year periods from 1926 to 2001, US equities total returns including dividends outperformed

Table 7
MSCI indexes grouped into quintiles by 31 December 1997 P/E, 31 December 1998

Quintile	World			North America			Europe			Latin America			Asia ex Japan			Japan		
		Total return			Total return			Total return			Total return			Total return			Total return	
			Market			Market			Market			Market			Market			Market
	P/E	Equity	Cap	P/E	Equity	Cap	P/E	Equity	Cap	P/E	Equity	Cap	P/E	Equity	Cap	P/E	Equity	Cap
1 Highest	57	13%	48%	48	20%	63%	55	25%	53%	31	−38%	−31%	36	−6%	7%	134	8%	−5%
2	25	13	45	26	16	43	24	24	25	19	−32	−21	18	10	10	39	16	16
3	18	9	30	20	7	24	19	16	32	14	−38	−28	13	15	11	29	15	12
4	14	−1	17	17	1	30	15	−0.4	35	9	−34	−37	8	−2	13	22	28	24
5 Lowest	8	3	17	13	−1	11	10	−3	13	5	−27	−25	5	19	35	14	38	32

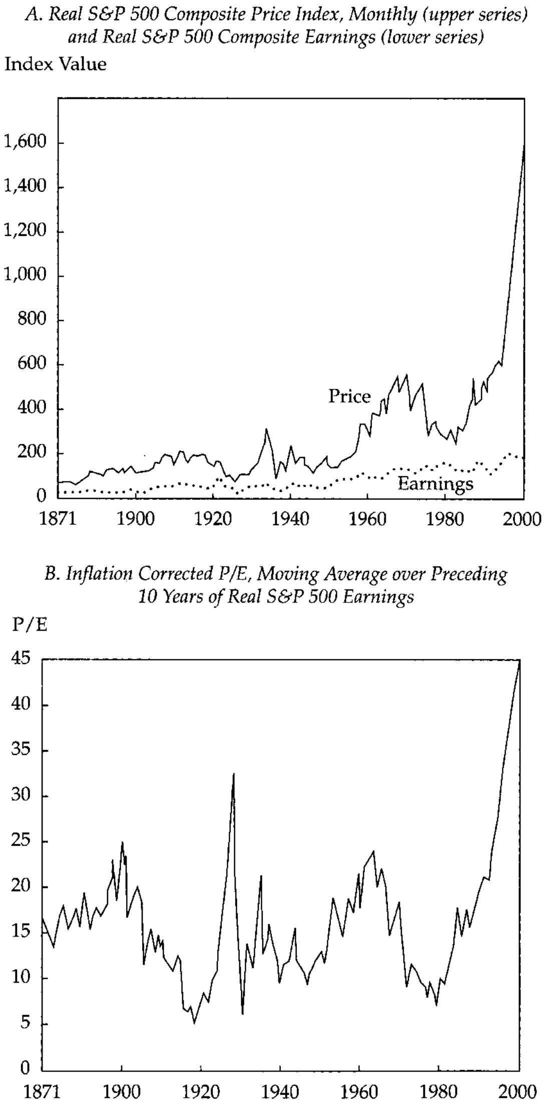


Fig. 14. Stock prices, earnings, and PE, January 1871–January 2000. Source: Shiller (2000).

bonds, and for 30-year horizons, it is optimal (with a mean-variance model) to be more than 100% in stocks and short bonds based on the past. Siegel used various risk tolerance measures, such as ultraconservative and risk taking. These measures are easy to devise using the Kallberg–Ziemba (1983) results by just assigning Arrow–Pratt risk-aversion values, as I have done in the second column of Table 9. Values

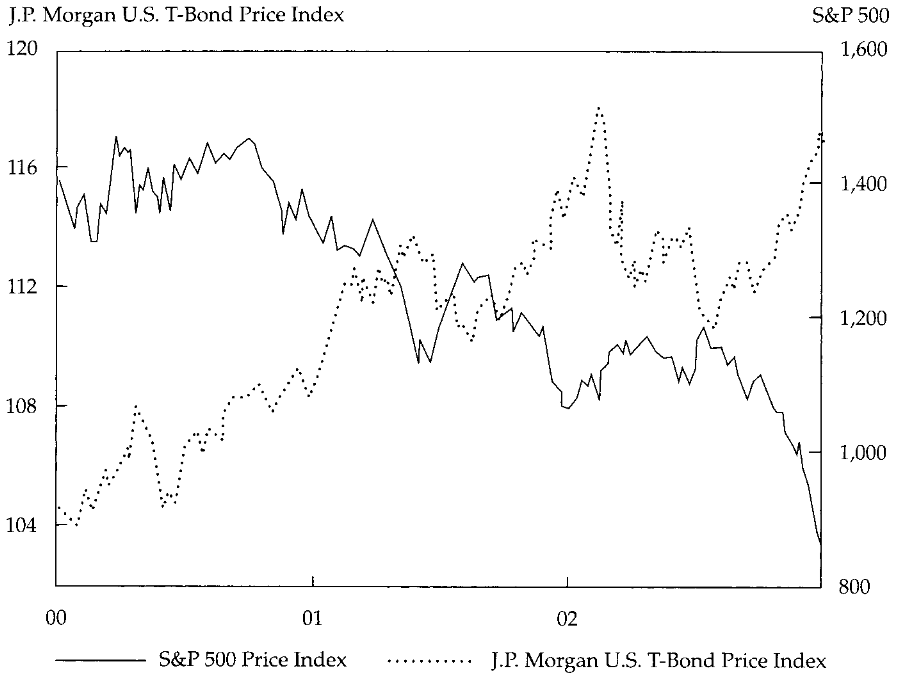


Fig. 15. S&P 500 and US Government Bonds, 2000–2002. Source: Schroder Investment Management Ltd.

Table 8
Equities' superior returns versus other asset classes, December 1925–December 1998

Asset class	Multiple
Inflation	9 times
T-bills	15 times
T-bonds	44 times
Corporate bonds	61 times
Large-cap stocks	2,351 times
Small-cap stocks	5,117 times

Source: Ibbotson Associates (1999), Swenson (2000).

over 100% mean more than 100% stocks or a levered long position, which would be short bonds or cash.

- For stochastic programming asset/liability models, we need scenarios over long periods. So, a deep major issue is how long the trouble might last.
- Stochastic programming models handle extreme event scenarios in a natural way. There is little chance of anyone predicting such events as the 9/11 attacks. However,

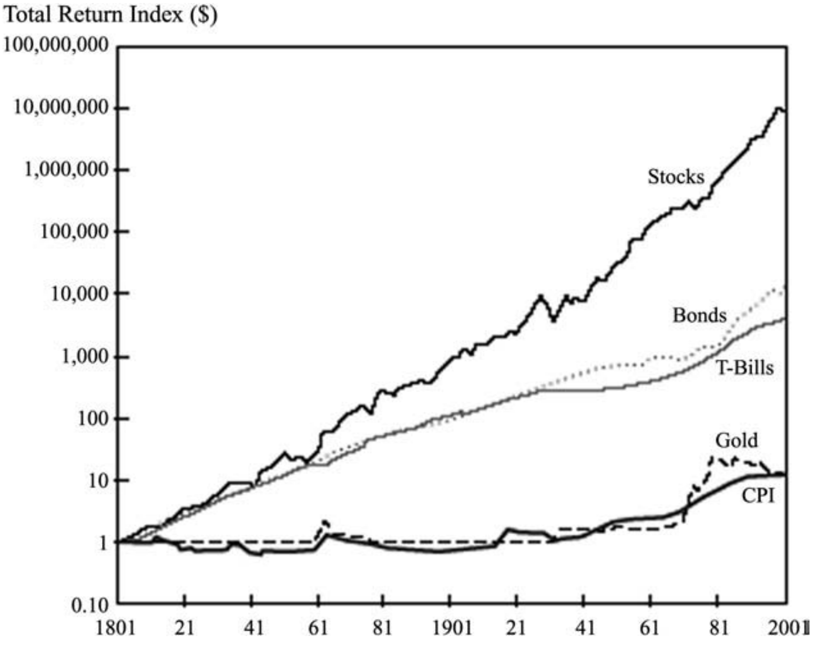


Fig. 16. Total nominal return indexes, 1801–2001. Note: CPI = Consumer Price Index. Source: Siegel (2002).

Table 9
Portfolio allocation: Percentage of portfolio recommended in stocks based on all historical data

Risk tolerance	Holding period				
	R_A	1 Year	5 Years	10 Years	30 Years
Ultraconservative	10	8.1%	23.3%	39.5%	71.4%
Conservative	6	25.0	40.6	60.1	89.7
Moderate	4	50.0	63.1	87.2	114.9
Risk taking	2	75.0	79.8	108.3	135.5

Note: R_A = risk-aversion index. Source: Siegel (2002).

scenarios that represent the effect that such events could occur in terms of their impact on market returns can be included. If such events have never occurred before, scenarios can be devised from similar events in other markets and their possible outcomes.

- Successful models will generate scenarios in many ways.

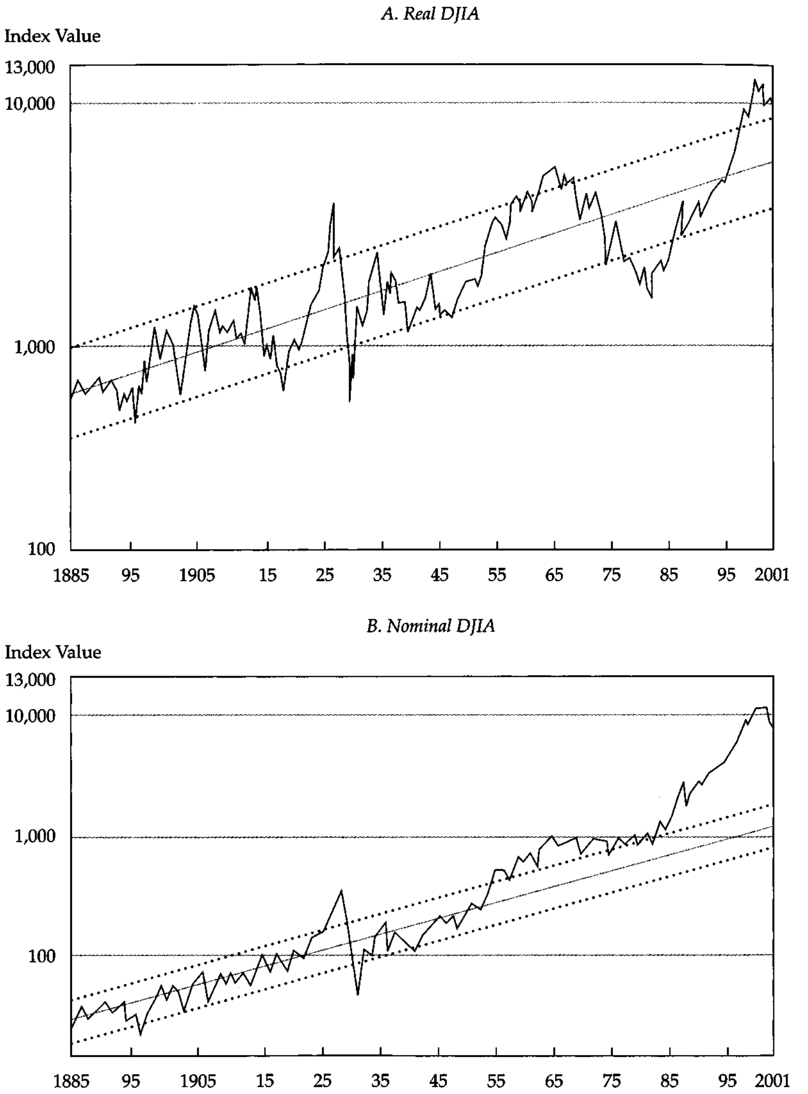


Fig. 17. The real and nominal DJIA. Source: Siegel (2002).

The next sections discuss some of the ways scenarios can be generated. Figures 15, 19 and 20 are useful in this regard. A clear pattern emerges; the position of stocks as over or under fair value has signaled market swings.

Figure 19 shows the late 2002 values for the crash indicator with the Fed (US Federal Reserve) model. The model uses 10-year bond yields and computes the ratio of the bond and stock yields in terms of a percentage over- or undervalued. This measure is

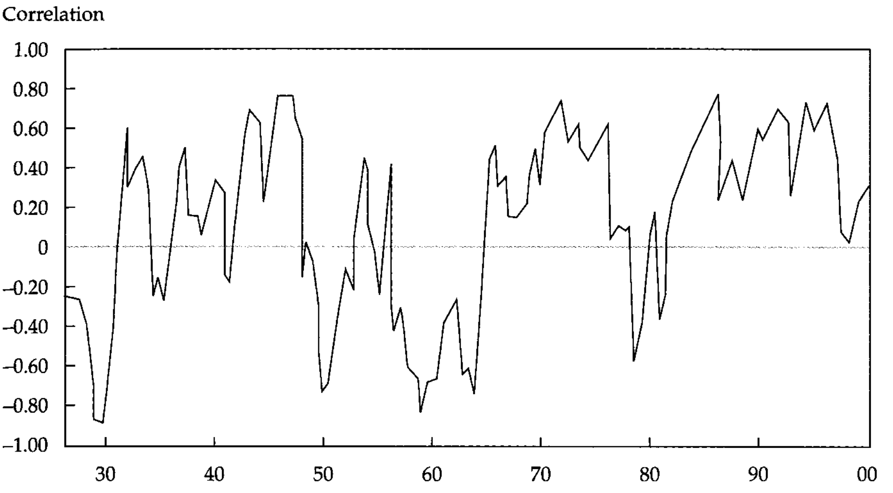


Fig. 18. The correlation between US equity and government bond returns, 1930–2000. Source: Schroder Investment Management Ltd.

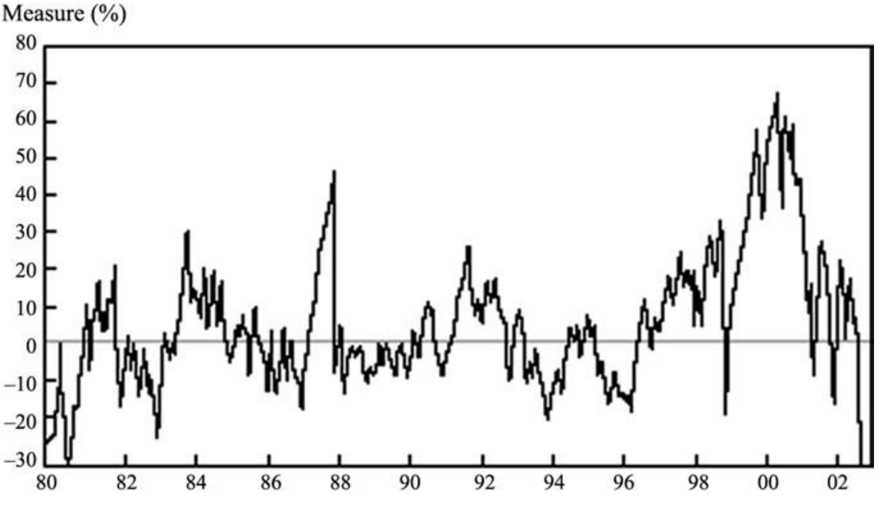


Fig. 19. Race to the bottom: percentage of S&P 500 over or under fair value, 1980–2002. Source: Ned Davis Research.

close to the difference crash model. Figure 19 indicates that very undervalued markets since 1980 have historically had high returns. When the measure is above 15%, mean S&P returns average a loss of 6.7%; when from 5–15%, the return is 4.9%; and when below −5%, the return is 31.7%. In late 2002/early 2003, the market was at one of its steepest discounts to fair value. Figure 20 provides my calculations, which mirror those

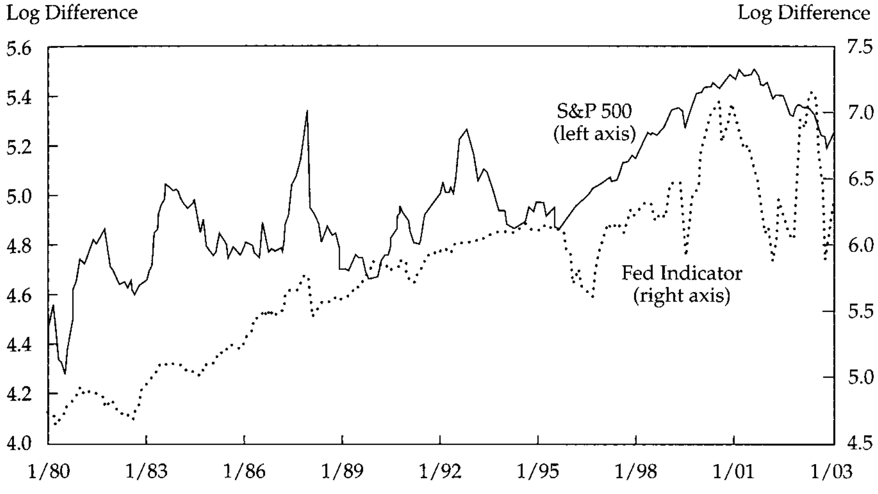


Fig. 20. The fed model, 1980–May 2003. Source: Koivu, Pennanen and Ziemba (2005).

in Figure 19. The length and depth of the 2000–03 decline is seen in the jagged parts of Figures 12, 15, and 20. One sees the initial danger zone for the measure in 1999, but then the market returned to the danger zone in 2001 and 2002 because stock prices fell and earnings fell even more. Consensus future earnings forecasts were invariably far too optimistic during this period. The S&P 500 fell from 1,460.25 at the end of December 1999 to 885.76 on 31 October 2002, down 37%.

Another valuation measure, thanks to Warren Buffett, is the market value of all publicly traded stocks relative to GNP. Buffett suggests that if the measure is 70 or 80%, it is a buy, and if it is over 200%, as in 1999, it is a sell. The measure was 133% in late 2001 and was lower in early 2003, since the stock market fell more than the GNP. This measure, like that in Figure 20, signaled a recovery in stock prices which occurred from 2003–2006.

Nonetheless, 41% of the stocks in the S&P 500 did not fall or actually rose during this period and an additional 19% declined by 10% or less annualized. These were small-cap stocks with market values of $10 billion or less. The fall in the S&P 500 was mainly in three areas—information technology, telecommunications, and large-cap stocks, with large increases prior to the crash. Information technology stocks in the S&P 500 fell 64% and telecommunications stocks fell 60% from 1 January to 31 October 2002. The largest-cap stocks (with market caps of US$ 50 billion plus) lost 37%. But most other stocks either lost only a little or actually gained. Materials fell 10% but consumer discretionary gained 4.5%, consumer staples gained 21%, energy gained 12%, financial services gained 19%, health care gained 29%, industrials gained 7%, and utilities gained 2%. Equally weighted, the S&P 500 index lost only 3%. These values include dividends. The stocks that gained were the very small-cap stocks with market caps below US$ 10 billion. Some 138 companies with market caps between US$ 5–10 billion gained 4%, on average, and 157 companies with market caps below US$ 5 billion gained 23%, on average. This is a reminder again of the dangers of markets where very few stocks are going up and dominating the market averages.

5. My philosophy

The philosophy I propose for assessing scenarios is as follows. Markets are understandable at most (95+ percent) of the time, but real asset prices have fat tails because extreme events occur much more than lognormal or normal distributions indicate.

According to Keim and Ziemba (2000), much of asset returns are *not* predictable. Hence, we must have ways to combine conventional models, options pricing, and so on, that are accurate most of the time with the irrational unexplainable aspects that occur once in a while. Whether the extreme events are predictable or not is not the key issue—what is *crucial* is that we consider that they can happen in various levels with various chances.

> Even apart from the instability due to speculation, there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than mathematical expectations, whether moral or hedonistic or economic. Most, probably, of our decision to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of animal spirits, a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities.
>
> (Keynes, 1938)

> Human behavior is a main factor in how markets act. Indeed, sometimes markets act quickly, violently with little warning. Ultimately, history tells us that there will be a correction of some significant dimension. I have no doubt that human nature being what it is, that it is going to happen again and again.
>
> (Greenspan, 1998)

As a working background hypothesis we assume that markets are basically efficient most of the time so we have to deal with that plus the anomaly and behavioral finance departures from efficiency.

An integrative asset/liability management strategy needs alternative modeling tools, such as stochastic programming, to capture the effects of costly lower-tail outcomes, as asked in Stulz (1996), Lo (1999, 2001), and Merton (2000a, 2000b).

5.1. Dynamic and liability aspects

Figure 21 shows the time flow of assets arriving and liability commitments leaving for institutions, such as insurance companies, pension funds, banks, and individuals. These problems are complex.

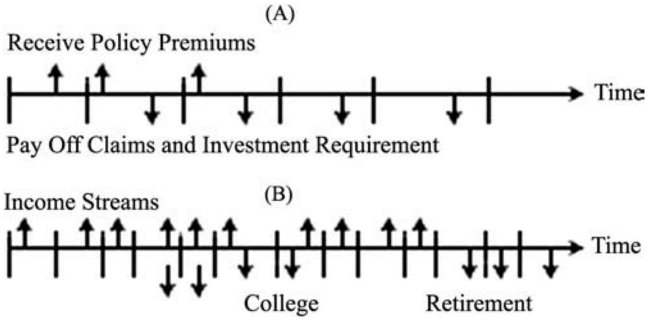


Fig. 21. Time flow of assets. (A) Institutions. (B) Individuals.

We have the following risk ladder from the introduction to Ziemba and Mulvey (1998) with various levels of details, aggregation, and model decisions.

- *Rung 5*. Total integrated risk management.
- *Rung 4*. Dynamic asset and liability management.
- *Rung 3*. Dynamic asset-only.
- *Rung 2*. Static assets-only portfolios.
- *Rung 1*. Pricing single securities.

This chapter concerns mostly Rung 3 and, especially, Rung 4. Rung 5 represents overall company wide models that involve all aspects of a business; see Rosen and Zenios (2006).

The stochastic programming approach, which considers the following aspects, is ideally suited to analyze such problems. I argue that it has the following elements.

- Multiple time periods; possible use of end effects—steady state after decision horizon adds one more decision period to the model; the trade-off is to either have an end-effects period or a larger model with one less period. The Russell-Yasuda Kasai model used end effects while the InnoALM model does not; see the discussion below.
- The scenarios should be consistent with economic and financial theory for asset returns, interest rates, and bond prices with anomaly and behavioral finance scenarios included.
- Discrete scenarios for random elements—returns, liabilities, currencies.
- Scenario-dependent correlation matrices so that correlations change for extreme scenarios. This was first implemented in the InnoALM model. Since bond prices rise during stock market crashes, these assets are negatively correlated then but they are positively correlated otherwise. So, capturing this effect with one correlation matrix will not work. You simply need multiple correlation matrices for scenarios sets. The LTCM and October 27–28, 1997 crash examples discussed above remind one of the importance of this.
- Use of various forecasting models that can handle fat tails.
- Institutional, legal, and policy constraints.
- Model derivatives and illiquid assets.

- Model the transaction costs.
- Use expressions of risk in terms understandable to decision makers item.
- I suggest simple, easy-to-understand concave, piecewise linear, risk-averse utility functions that maximize long-run expected profits, net of expected discounted penalty costs for shortfalls. You pay more and more penalty for shortfalls as they increase (highly preferable to VAR).
- Model as constraints or penalty costs for target violations in the objective function. This allows for multiple goals through targets within periods and across periods. But one must be able to specify the relative importance of these targets within periods and across periods.
- You maintain adequate reserves and cash levels and regularity requirements.
- Realistic multiperiod problems can now be solved on modern workstations and personal computers by using large-scale linear programming and stochastic programming algorithms; see the InnoALM model below for one application using the IBM code and Wallace and Ziemba (2005) for other publicly available code systems.
- The model makes you diversify the key for keeping out of trouble.

  Some possible approaches to model situations with such events are as follows:
- Simulation models have too much output to understand but are very useful as check. SP models via stress tests and what if calculations and in comparison with other models.
- Mean-variance models are static in nature and useful for such applications, but they are not very useful with liquidity or other constraints or for multiperiod problems or with liabilities, etc.
- Expected log models yield very risky short term strategies that do not diversify well; fractional Kelly with downside constraints are excellent for risky investment betting (see MacLean and Ziemba, 2006 and Thorp, 2006).
- Stochastic control models, although theoretically interesting, give hair-trigger and bang-bang policies in which one is 100% stocks and then zero percent stocks one second later. See, for example, the Brennan–Schwartz (1998) model in Ziemba and Mulvey (1998) and the Rudolf–Ziemba (2004) model. The question is how to constrain the asset weight changes to be practical? Possibly, the best work with this approach has been done by Campbell and Viceira (2002), who use the approach successfully to analyze long-term asset-only allocation decisions in which the power of the technique dominates the limitations of the model (which they acknowledge). Among other conclusions, they show that:
  (a) The riskless asset for a longterm investor is not T-bills, rather it is an inflation indexed bond since it delivers a predictable stream of real income.
  (b) A safe labor income stream is equivalent to a large position in the risk-free asset, allowing the investor to hold much more in risky assets, then this is reversed to some extent. Fixed commitments are negative income.
  (c) Risky investments are extremely attractive to young households because they have large relatively safe human wealth relative to their financial wealth.

(d) Business owners should have less equity exposure since their income stream is correlated with the stock market.
(e) Wealthy investors should be more risk averse since more of their consumption stream depends on their financial success.
(f) Since stocks are mean reverting, that is, they have lower risk over longer time horizons, investors can time the market over longer horizons; since the equity risk premium is time varying, the optimal strategic allocation mix changes over time. These are useful rules of thumb for many investors derived from a theoretically sound framework. My goal in SP-ALM models is to tailor the asset allocation mix for the particular institution or investor given their consumption and other goals, taxes, preferences, uncertainties, transactions costs, liquidity, etc.
(g) Stochastic programming models with decision rules have policy prescriptions, such as fixed mix or buy and hold; the decision rules are intuitively appealing but are suboptimal and usually lead to nonconvex difficult optimization modeling.

Stochastic programming models provide a good approach to asset liability management as we see below.

5.2. The importance of getting the mean right

In any static or multiperiod decision model, SP or otherwise, the estimates of the mean are 60% or more of the success and risk of the model. The left tail is next most important as that is where the losses are. I start with a theoretical result that the mean dominates if the two distributions cross only once; this is illustrated in Figure 22.

Theorem 1. (Hanoch and Levy, 1969) *If $X \sim F(\,)$ and $Y \sim G(\,)$ have CDF's that cross only once, but are otherwise arbitrary, then F dominates G for all concave u. The mean of F must be at least as large as the mean of G to have dominance. Variance and other moments are unimportant. Only the means count.*

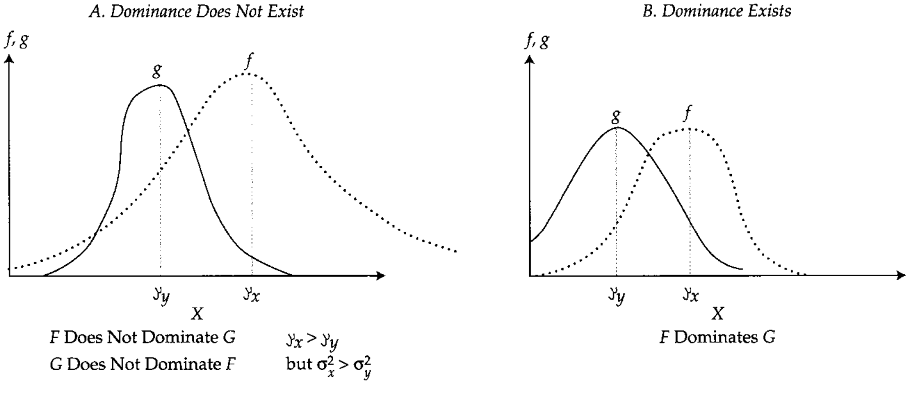


Fig. 22. The main theorem of mean-variance analysis.

With normal distributions X and Y will cross only once iff the variance of X does not exceed that of Y. That is the basic equivalence of Mean-Variance analysis and Expected Utility Analysis via second-order (concave, non-decreasing) stochastic dominance.

5.3. Errors in means, variances and covariances: empirical

Replace the true mean μ_i by the observed mean $\mu_i(1 + kZ_i)$ where Z_i is distributed $N(0, 1)$ with scale factor $k = 0.05$ to 0.20, being the size of the error. Similarly, replace the true variances and covariances by the observed variances $s_i^2(1 + kZ_i)$ and covariances $s_{ij}(1 + kZ_i)$. We use monthly data from 1980–89 on ten DJIA securities which include Alcoa, Boeing, Coke, Dupont and Sears. See Chopra and Ziemba (1993) which updates and extends Kallberg and Ziemba (1981, 1984).

The certainty equivalent, CE, of a portfolio with utility function u equals u^{-1} (expected utility of a risky portfolio). This comes from the equation:

$$u(\mathrm{CE}) = \mathrm{E}_\xi u(\xi' x) \quad \Rightarrow \quad \mathrm{CE} = u^{-1}\big[\mathrm{E}_\xi u(\xi' x)\big].$$

Assuming exponential utility and normal distributions, yields exact formulas to calculate all quantities in the

$$\text{certainty equivalent loss (CEL)} = \left(\frac{\mathrm{CE}_{\mathrm{opt}} - \mathrm{CE}_{\mathrm{appox}}}{\mathrm{CE}_{\mathrm{opt}}}\right)100.$$

Observe that the mean-variance problem is mean-($\frac{1}{2}$ risk aversion) variance.

Table 10 shows that the errors in means are about 20 times the errors in covariances in terms of CEL value and the variances are twice as important as the covariances. So

Table 10
Mean percentage cash equivalent loss due to errors in inputs

Average ratio of CEL for errors in means, variance and covariances			
t risk tolerance	Errors in means vs. covariances	Errors in means vs. variances	Errors in variances vs. covariances
75	5.38	3.22	1.67
50	22.50	10.98	2.05
75	56.84	21.42	2.68
	↓	↓	↓
	20	10	2
The error depends on the risk tolerance but roughly			
	Error mean	Error var	Error covar
	20	2	1

Source: Chopra and Ziemba (1993).

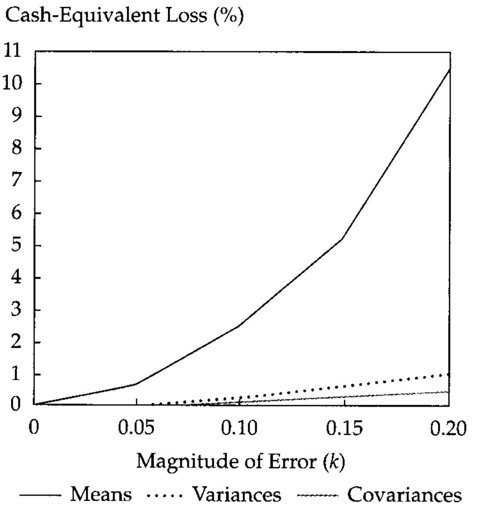


Fig. 23. The effect of errors in means, variances and covariances on optimal portfolios. Source: Chopra and Ziemba (1993).

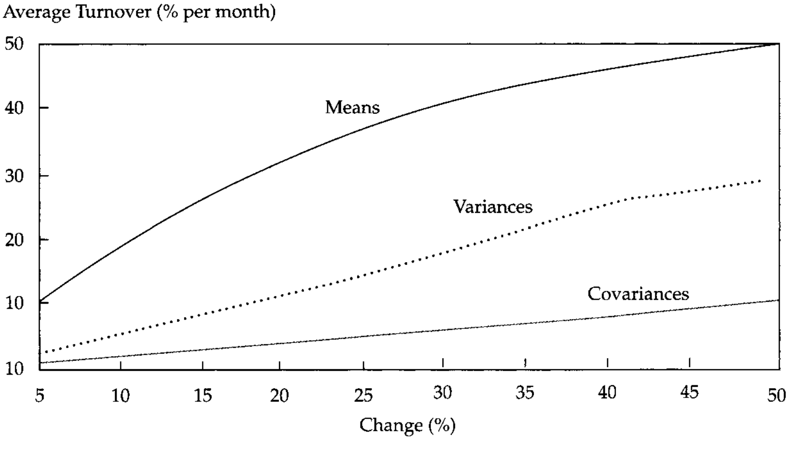


Fig. 24. Average turnover: percentage of portfolio sold (or bought) relative to preceding allocation. Source: Chopra (1993).

roughly, there is a 20:2:1 ratio in the importance of these errors. Also, this is risk aversion dependent with $t = (R_A/2)100$ being the risk tolerance. So for high risk tolerance, that is low risk aversion, the errors in the means are even greater. Hence for utility functions like the log of Kelly with essentially zero risk aversion, the errors in the mean can be 100 times as important as the errors in the other parameters. So, Kelly bettors should never overbet. See Table 10.

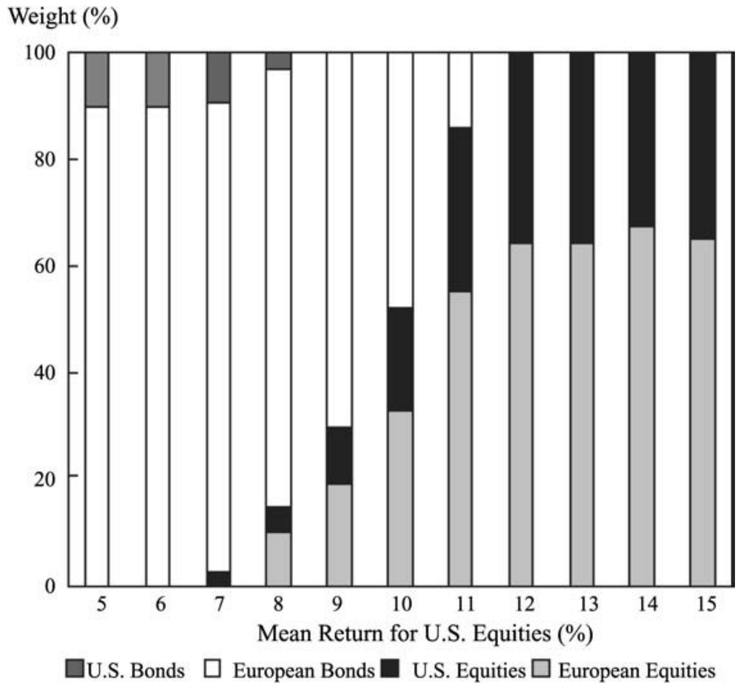


Fig. 25. Optimal asset weights at stage 1 for varying levels of US equity means in a multiperiod stochastic programming pension fund model for Siemens Austria. Source: Geyer and Ziemba (2007).

Conclusion: spend your money getting good mean estimates and use historical variances and covariances.

Chopra (1993) shows that the same relationship holds regarding turnover but it is less dramatic than for the cash equivalents, see Figure 24.

These results apply to essentially *all* models. You must get the means right to win! Figure 25 shows this effect in the five period ten year InnoALM modeldiscussed below.

If the mean return for US stocks is assumed to equal the long run mean of 12% as estimated by Dimson et al. (2002, 2006, 2007), the model yields an optimal weight for equities of 100%. A mean return for US stocks of 9% implies less than 30% optimal weight for equities. The effect is much less in later periods; see Figure 23.

5.4. Fixed-mix strategies

Fixed-mix strategies, where the asset allocation weights are constant and at each decision point the assets are rebalanced to the initial weights, are common and yield good results. An attractive feature is their volatility pumping because they rebalance by selling assets at high prices and buying them at low ones. Fixed-mix strategies compare are usually superior to buy-and-hold strategies. Figure 26 shows the 1982–94 return-risk

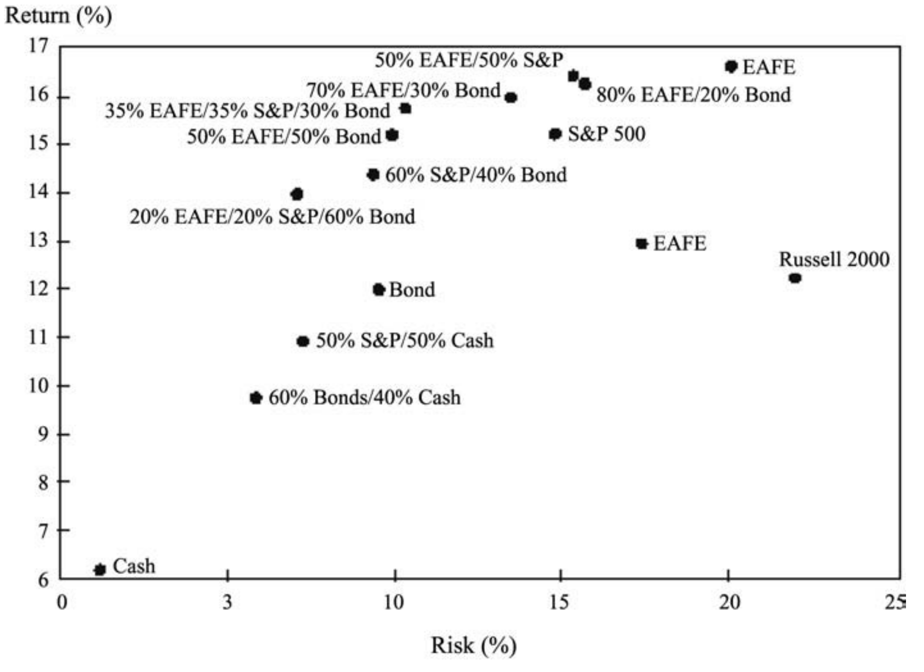


Fig. 26. Historical performance of some asset categories, 1 January 1982–1931 December 1994. Source: Ziemba and Mulvey (1998).

performance of a number of asset categories, with various asset weights of the Europe, Australia, and the Far East (EAFE) Index, S&P 500, bonds, the Russell 2000 Small-Cap Index, and cash.

The theoretical properties of fixed mix strategies are discussed by Dempster, Evstigneev and Schenk-Hoppé (2003) and Merton (1990), who show their advantages in stationary markets, where the return distributions are the same each year. The long run growth of wealth is exponential with probability one. The stationary assumption is fine for long run behavior but for short time horizons even up to 10 to 30 years using scenarios to represent the future will generally give better results.

Hensel, Ezra and Ilkiow (1991) showed the value of strategic asset allocation. They evaluated the results of seven representative Frank Russell US clients whose assets were managed by recommended superior professional managers (who are supposed to beat their benchmarks with lower risk). The study covered the 16 quarters from January 1985 to December 1988. A fixed-mix *naive* benchmark was US equity (50%), non-US equity (5%), US fixed income (30%), real estate (5%), and cash (10%). Table 11 shows the mean quarterly returns and the volatility explained. Most of the volatility (94.35% of the total) is explained by the naive policy allocation, similar to the 93.6% in Brinson, Hood and Beebower (1986) and in Brinson, Singer and Beebower (1991). T-bill re-

Table 11
Average return and return variation explained (quarterly by the seven clients)

Decision level	Average contribution	Additional variation explained (volatility)
Minimum risk (T-bills)	1.62%	2.66%
Naive allocation (fixed mix)	2.13	94.35
Specific policy allocation	0.49	0.50
Market timing	(0.10)	0.14
Security selection	(0.23)	0.40
Interaction and activity	(0.005)	1.95
Total	3.86%	100.00%
T-bills and fixed mix	3.75	

Source: Hensel, Ezra and Ilkiw (1991).

turns (1.6%) and the fixed-mix strategy (2.13%) explain most of the mean returns. The managers returned 3.86% versus 3.75% for T-bills plus fixed mix, so they added value. This added value was from their strategic asset allocation into stocks, bonds, and cash. The managers were unable to market time or to pick securities better than the fixed-mix strategy.

Further evidence that strategic asset allocation accounts for most of the time-series variation in portfolio returns and that market timing and asset selection are far less important has been provided by Blake, Lehmann and Timmermann (1999). They used a nine-year (1986–94) monthly data set on 306 UK pension funds with eight asset classes. They found a slow mean reversion in the funds' portfolio weights toward a common, time-varying strategic asset allocation. The UK pension industry had very few management companies. At the time of this study, the top four companies controlled 80% of the market. In contrast, in the US, the largest company in 1992 had a 3.7% share, according to Lakonishok, Schleifer and Vishny (1994). During the 1980s, the UK pensions were about 50% overfunded but in 2006 these pensions are largely underfunded; such as the UK university professors' pension plan which is only about 82% in 2006 funded with a shortfall of £6 billion and 90% funded in April 2007; see Board and Sutcliffe (2007). Fees are related to performance usually relative to a benchmark or peer group. BLT concluded that:

- UK pension fund managers had a weak incentive to add value and face constraints on how they try to do it. Although strategic asset allocation may be set by the trustees, these strategies are flexible, have wide tolerance for short-run deviations, and can be renegotiated.
- Fund managers know that relative rather than absolute performance determines their long-term survival in the industry.
- Fund managers earn fees related to the value of assets under management, not to their relative performance against a benchmark or their peers, with no specific penalty for underperforming nor reward for outperforming.

- The concentration in the industry leads to portfolios being dominated by a small number of similar *house positions* for asset allocation to reduce the risk of relative underperformance.

The asset classes from WM Company data were UK equities, international equities, UK bonds, international bonds, cash, UK property, and international property. UK portfolios are heavily equity weighted. The 1994 weights for these eight asset classes for the 306 pension funds were 53.6, 22.5, 5.3, 2.8, 3.6, 4.2, 7.6, and 0.4%, respectively. In contrast, US pension funds had 44.8, 8.3, 34.2, 2.0, 0.0, 7.5, 3.2, and 0.0%, respectively.

Most of the 306 funds had very similar returns year by year. The semi-interquartile range was 11.47 to 12.59%, and the 5th and 95th percentiles were less than 3% apart.

Leaving out international property, the returns on different asset classes were not exceptional. The eight classes averaged value weighted 12.97, 11.23, 10.76, 10.03, 8.12, 9.01, 9.52, and −8.13 (for the international property) and overall 11.73% a year. Bonds and cash had returns similar to the equities in this period. BLT found, similar to the previous US studies cited above, that for UK equities, a high percentage (91.13%) of the variance in differential returns across funds occurred because of strategic asset allocation. For the other asset classes, this variance is lower—60.31% (international equities), 39.82% (UK bonds), 16.10% (international bonds), 40.06% (UK index bonds), 15.18% (cash), 76.31% (UK property), and 50.91% (international property). For these other asset classes, variations in net cash flow differentials and covariance relationships explain the rest of the variation.

5.5. Stochastic programming is superior to fixed mix

Despite their good results, fixed-mix and buy-and-hold strategies do not use new information from return outcomes in their asset allocation. Indeed, their asset allocation is fixed independent of all new information. Hence, a stochastic programming model which reacts to such information should be superior. Here, we use an example of Cariño and Turner (1998) to show this. This section also illustrates the use of scenario dependent strategies within a multiperiod stochastic programming framework.

The SP model has three periods of one, two and two years. The investor starts at Year 0 and ends at Year 5, with the goal of maximizing expected final wealth, net of risk. Risk is measured as one-sided downside risk based on non-achievement of a target wealth goal at Year 5. The target is 4% return a year, or 21.7% at Year 5. Figure 27 shows the shortfall cost function. The penalty for not achieving the target is steeper and steeper as the non-achievement increases. For example, at 100% or more of the target, there is no penalty; at 95–100%, it is a steeper, more expensive penalty; and at 90–95%, it is steeper still. This shape preserves the concave risk averse behavior of the objective function using a convex risk-penalty function. The piecewise linear function means that the stochastic programming model remains linear.

The objective function is

$$\text{Maximize } E\left(\text{final wealth} - \frac{R_A}{2}(\text{accumulated penalized shortfalls})\right)$$

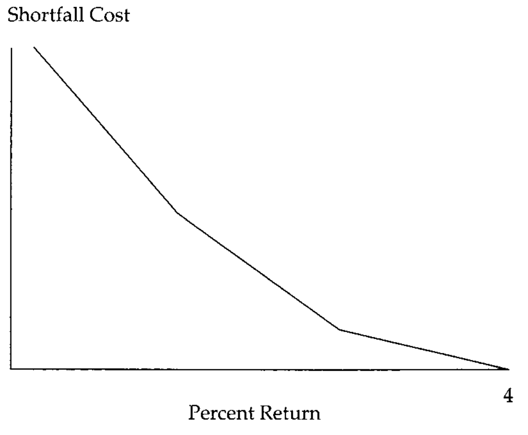


Fig. 27. A convex piecewise linear shortfall cost function: target 4% a year.

Table 12
Means, variances, and covariances of six asset classes

Asset class	Expected return	Std. dev.	US large-cap equity	US small-cap equity	Non-US equity unhedged	Emerging markets unhedged	US bonds	US cash
US large-cap equity	11.0%	17.0%	1.0					
US small-cap equity	11.0	25.0	0.8	1.0				
Non-US equity unhedged	11.0	21.0	0.5	0.3	1.0			
Emerging markets unhedged	11.0	25.0	0.3	0.3	0.3	1.0		
US bonds	7.0	7.0	0.4	0.3	0.2	0.0	1.0	
US cash	5.7	1.0	0.0	0.0	0.0	0.0	0.3	1.0

Source: Cariño and Turner (1998).

where $R_A = -u''(w)/u'(w)$ is the Arrow–Pratt absolute risk-aversion index and balances risk and return.

The six asset classes have the means, variances, and covariances shown in Table 12. Scenarios represent possible future outcomes. In this situation, that is all the possible paths of returns that can occur over the three periods. The goal is to make 4% each period, so cash that returns 5.7% will always achieve this goal. Bonds return 7.0%, on average. They return at least 4%, much of the time but often have returns below 4%. Equities return 11% and beat the 4% hurdle most of the time but also fail to achieve 4% some of the time. Assuming that the returns are independent and identically distributed with lognormal distributions, by sampling $4 \times 3 \times 2$, there are 24 scenarios for cash, bonds, and equities, as shown in Figure 28. The heavy line is the 4% threshold, or 121.7% at Year 5. To simplify, the scenarios are visualized over two periods. The scenario tree has nine nodes with six distinct scenarios. Three outcomes are possible in

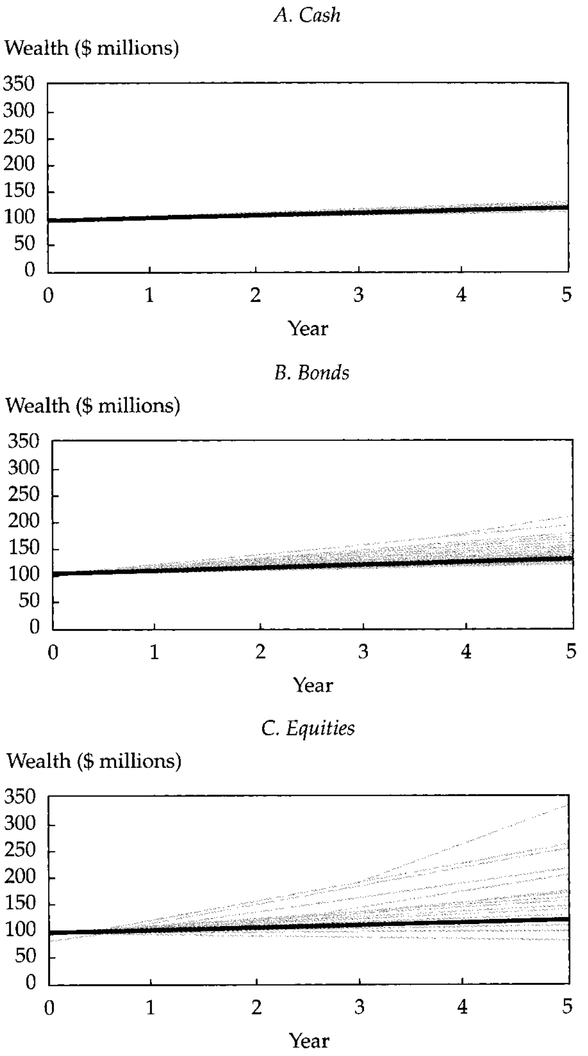


Fig. 28. Scenarios. Source: Cariño and Turner (1998).

Period 1 and two are possible in Period 2, for six in total, as shown in Figure 29 and in Table 13. For example, with one-third probability, US equity large caps will return 0.90754, US equity small cap, 0.534592, and so on; in Period 2, these two return either 0.119713 or 0.461739 and −0.130465 or 0.392537, with equal probability. The strategy layout is

Period 1 1 Year	Period 2 2 Years	Period 3 2 Years

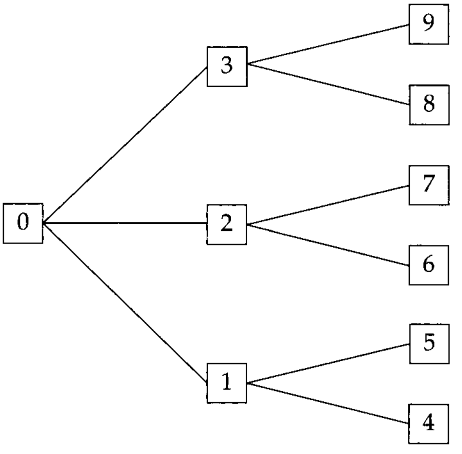


Fig. 29. Scenarios in the three periods. Source: Cariño and Turner (1998).

Table 13
Example scenario outcomes listed by node

Node	Conditional probability	US large-cap equity	US small-cap equity	Non-US equity	Emerging markets	US bonds	US cash
1	0.3333	0.032955	0.341701	0.041221	0.279216	0.027300	−0.014084
2	0.3333	−0.091184	0.049939	0.109955	0.082171	−0.128904	0.024156
3	0.3333	0.090754	0.534592	0.120825	0.204917	0.162770	0.132663
4	0.5000	0.035930	0.056592	−0.000627	−0.304342	0.061070	0.000830
5	0.5000	0.119713	−0.130465	0.193180	0.519016	0.069383	0.028540
6	0.5000	0.461739	0.392537	0.116938	0.360205	0.089025	0.050224
7	0.5000	0.245134	0.122433	0.568656	0.180286	0.110467	0.092815
8	0.5000	−0.090452	−0.292077	−0.292757	0.001132	0.129944	0.121655
9	0.5000	0.041096	0.054468	0.118764	−0.048986	0.065222	0.088793

Source: Cariño and Turner (1998).

We then compare the following two strategies:

1. the dynamic stochastic programming strategy, which is the full optimization of the multiperiod model; and
2. the fixed-mix strategy, in which the portfolios from the mean-variance frontier have allocations rebalanced back to that mix at each stage—buy when low and sell when high. This strategy resembles covered calls, which is the opposite of portfolio insurance.

There is a continuum of fixed mix strategies all dominated by the stochastic programming strategies as shown in Figure 30. We focus on two such strategies: Fixed-Mix

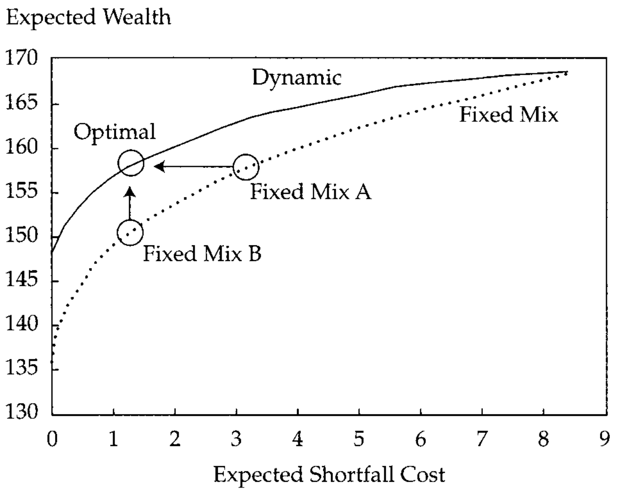


Fig. 30. Optimal stochastic strategy vs. fixed-mix strategy. Source: Cariño and Turner (1998).

Strategy A (64/36 percent stock/bond mix) and Fixed-Mix Strategy B (46/54 percent stock/bond mix).

How does the dynamic strategy achieve lower shortfall cost or higher expected wealth? For any given fixed-mix strategy, there is a dynamic strategy that has either the same expected wealth and lower shortfall cost or the same shortfall cost and higher expected wealth. The portfolio allocation for the optimal strategy starts out in a similar way to the Fixed-Mix Strategy A, with the same expected return. The split between equities and fixed-income assets is about 60/40 percent. The Fixed-Mix Strategy B, having the same risk, has a much lower allocation to equities (see Panel A of Figure 31).

The allocations at the end of Year 1 provide insight into the dynamic strategy (see Figure 31). The allocations depend on the outcome of the first year's returns. The optimal allocations are dependent upon to the outcomes. The allocations shift to lower-return, less volatile assets as the excess over the wealth target is reduced. When the return is high, the strategy moves to a high-return riskier asset.

The model suggests that by taking advantage of the opportunities to adapt the asset mix given the current wealth level, the chances of exceeding the target hurdle are increased.

This example is "in sample". Fleten, Høyland and Wallace (2002) compared two versions of a portfolio model for the Norwegian life insurance company Gjensidige NOR, namely, multistage stochastic linear programming and fixed-mix. They found that the multiperiod stochastic programming model dominated the fixed-mix approach. However, the degree of dominance was much smaller out of sample than in sample (see Figure 32).

Why does the SP advantage decrease so much? The answer seems to be that out of sample, the random input data are structurally different from those in sample, so the

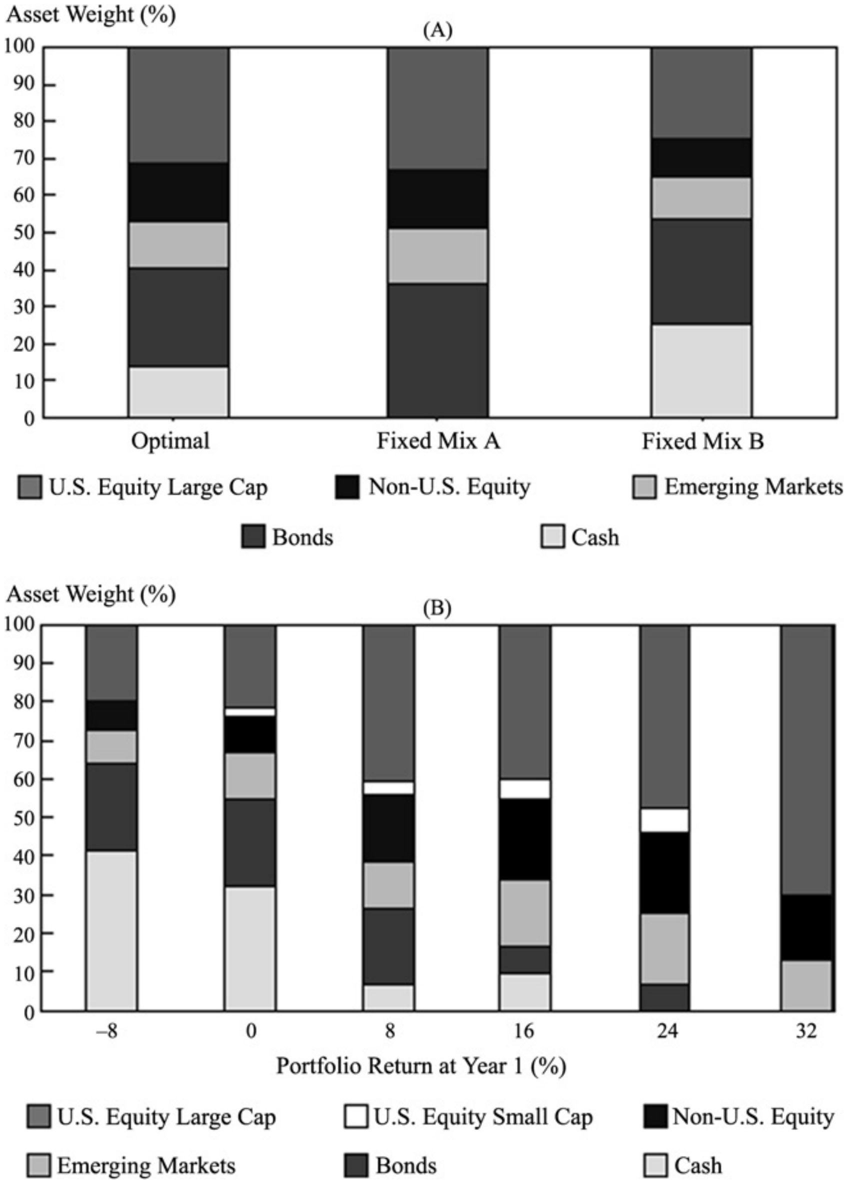


Fig. 31. Initial and period 1 portfolios. (A) Initial portfolios of the three strategies. (B) Contingent allocations at Year 1. Source: Cariño and Turner (1998).

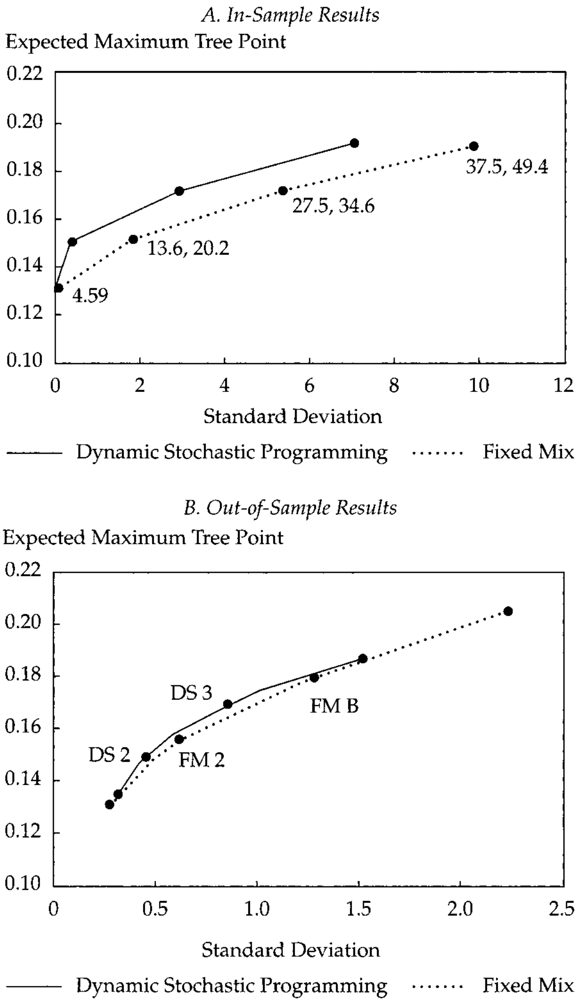


Fig. 32. Advantage of stochastic programming over fixed-mix model. Source: Fleten, Høyland and Wallace (2002).

stochastic programming model loses its advantage in optimally adapting to the information available in the scenario tree. Also, the performance of the fixed-mix approach improves because the asset mix is updated at each stage through its volatility pumping procedure.

6. The Russell-Yasuda Kasai model

The Russell-Yasuda Kasai model was the first large-scale multiperiod stochastic programming model implemented for a major financial institution (see Henriques, 1991).

I designed the model while working a consultant to the Frank Russell Company in Tacoma, Washington, from 1989–1991. Under the direction of Research Head Andy Turner, the team of David Cariño, Taka Eguchi, David Myers, Celine Stacy, and Mike Sylvanus at Russell in Tacoma, Washington, implemented the model for the Yasuda Fire and Marine Insurance Company in Tokyo. Professors Roger Wets and Chanaka Edirishinghe served as consultants in Tacoma, and Professor Katsughe Sawaki was a consultant to Yasuda Kasai in Japan. Kats, a member of my 1974 University of British Columbia class in stochastic programming in which we started to work on asset/liability models (ALMs), was then a professor at Nanzan University in Nagoya and acted independently of our Tacoma group. Kouji Watanabe headed the group in Tokyo, which included Y. Tayama, Y. Yazawa, Y. Ohtani, T. Amaki, I. Harada, M. Harima, T. Morozumi, and N. Ueda.

In 1990–1991, computations were a major focus of concern. I knew how to formulate the model, an outgrowth of the Kusy and Ziemba (1986) model for the Vancouver Savings and Credit Union and the Kallberg, White and Ziemba (1982) model. David Cariño, who had a PhD from Stanford's Engineering Economic Systems Department, did much of the formulation of the details. Originally, we had a 10-period model with 2048 scenarios. The model was far too big to solve at that time and became an intellectual challenge for the stochastic programming community. Drs. Robert Entriken, D. Jensen, R. Clark, and Alan King of IBM Research worked on its solution but never quite cracked it. We quickly realized that 10 periods made the model far too difficult to solve and also too cumbersome to collect the data and interpret the results, and the 2048 scenarios were, at that time, a large number to deal with. About two years later, Professor Hercules Vladimirou, working with Alan King at IBM Research and using parallel processing on several workstations, was able to effectively solve the original model. The state of the art in 2007 is such that huge models can now be solved; see, e.g., Gonzio and Kouwenberg (2001), and the model systems in Wallace and Ziemba (2005).

The Russell-Yasuda Kasai model was designed to satisfy the following need, as articulated by Kunihiko Sasamoto, director and deputy president of Yasuda Kasai:

> The liability structure of the property and casualty insurance business has become very complex, and the insurance industry has various restrictions in terms of asset management. We concluded that existing models, such as Markowitz mean-variance, would not function well and that we needed to develop a new asset/liability management model.
> The Russell-Yasuda Kasai model is now at the core of all asset/liability work for the firm. We can define our risks in concrete terms, rather than through an abstract, in business terms, measure like standard deviation. The model has provided an important side benefit by pushing the technology and efficiency of other models in Yasuda forward to complement it. The model has assisted Yasuda in determining when and how human judgment is best used in the asset/liability process.
>
> Cariño et al. (1994, p. 49)

The model was a huge success and was of great interest to both the academic and institutional investment asset/liability communities. Our research and development team

Table 14
Russell business engineering models

Model	Type of application	Year delivered	Number of scenarios	Computer hardware
Russell-Yasuda (Tokyo)	Property and casualty insurance	1991	256	IBM RISC 6000
Mitsubishi Trust (Tokyo)	Pension consulting	1994	2,000	IBM RISC 6000 with Parallel Processors
Swiss Bank Corp. (Basle)	Pension consulting	1996	8,000	IBM UNIX2
Daido Life Insurance Company (Tokyo)	Life insurance	1997	25,600	IBM PC
Bank Fideuram (Milan)	Assets only (personal)	1997	10,000	IBM UNIX2 and PC
Consulting clients	Assets only (institutional)	1998	Various	IBM UNIX2 and PC

in Tacoma and Tokyo won second prize in the Franz Edelmann Practice of Management Science competition in Chicago in May 1993. Our work was summarized in the January 1994 issue of *Interfaces*, in which all six finalists had their papers published (see Cariño et al., 1994). The full model is described in Cariño and Ziemba (1998) and Cariño, Myers and Ziemba (1998). The main points are summarized here.

The success of the Russell-Yasuda Kasai model led to the formation of a new business unit called Russell Business Engineering. That group built three other large custom models for insurance and pension funds in Japan and Switzerland, an ALM planning system for individuals in Italy, and an assets-only system for use in consulting for large pension fund clients of Russell (see Table 14 for general details of these models). There are no publicly available papers on these models but there are slides from talks in Cariño et al. (1997), Fan et al. (1998), Murray (1998), and Murray and Fan (1998). This development at Russell helped spawn the field of stochastic programming in finance related to asset/liability modeling. In May 1995, I was asked to organize a two-week seminar at the Issac Newton Institute, Cambridge University, as part of a six-month program on financial mathematics. The week on ALMs led to the book by Ziemba and Mulvey (1998), and the week on security market anomalies led to the book by Keim and Ziemba (2000). Figure 33 with model origins, early models, and modern models from the introduction to Ziemba and Mulvey shows how much the field grew in the 1990s. It has grown even more since, and this *Handbook* is an assessment of the current state of the literature and applications. Before discussing the Russell-Yasuda Kasai model, I review a few aspects of the insurance business.

6.1. Elements of the insurance business

Insurance businesses have two basic elements—the collection of premiums for bearing the risks of other people and organizations and the investment of those premiums. They are classic asset/liability enterprises. The investment side parallels that of other financial

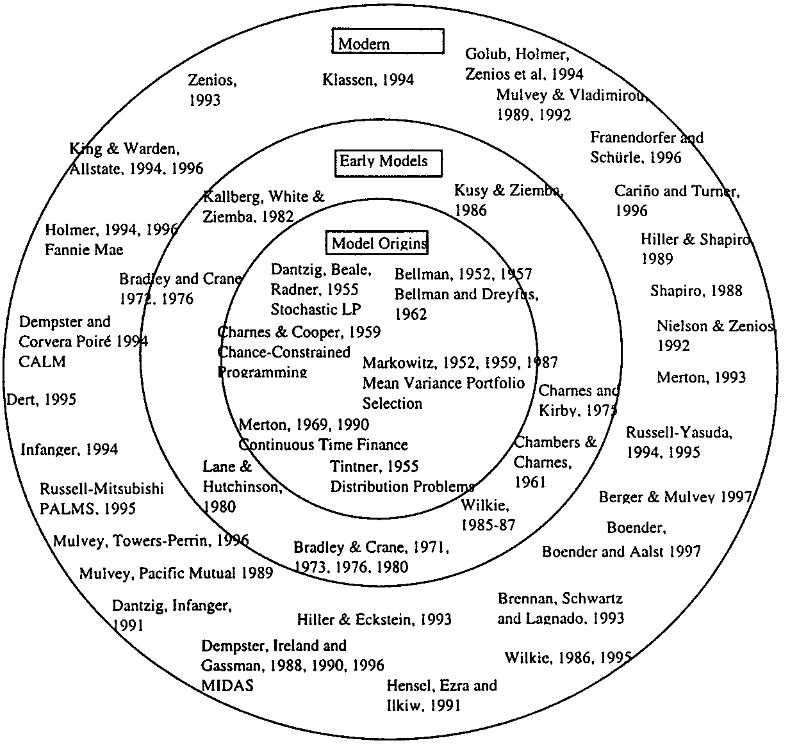


Fig. 33. Stochastic programming asset–liability models over time. Source: Ziemba and Mulvey (1998).

institutions. They must invest their premiums and previous investments to provide satisfactory returns over time and provide resources for insurance claims. They must have enough capital to weather storms from bad scenario outcomes from both the investment business and insurance claims.

In some insurance businesses, the claims have little risk once aggregated, if properly diversified, as in life insurance. Since insurance is put selling, this is profitable on average, since puts are almost always overvalued, but very dangerous if one over bets and is not diversified in all scenarios; see Section 3.4. But in most insurance businesses, the claim risk is substantial.

The insurance side generates premiums that, on average, cover future claims. These claims have distributions of losses, as with typical liabilities. The danger is in the tails. Insurance companies are frequently insuring against rare events for which the loss is many times the premium paid. They frequently diversify across such events by passing on some of this tail risk to re-insurers.

Insurance companies have been in business for hundreds of years and have developed sophisticated mathematical analyses of claims. Still, disasters can easily occur and many insurance companies go bankrupt. The danger is that the insurance policies written may not be really as diversified as the calculations and experience might indicate. Then, if, in effect, overbetting and a disastrous scenario occurs, serious trouble may ensue.

6.2. The Yasuda Kasai problem

The Yasuda Fire and Marine Insurance Company, called "Yasuda Kasai" meaning *fire*, is based in Tokyo. It began operations in 1888 and is the second largest Japanese property and casualty insurer and seventh largest in the world by revenue. Its main business was voluntary automobile (43.0%), personal accident (14.4%), compulsory automobile (13.7%), fire and allied (14.4%), and other (14.5%). The firm had assets of ¥3.47 trillion (US$ 26.2 billion) at the end of fiscal 1991 (31 March 1992). In 1988, Yasuda Kasai and Russell signed an agreement to deliver a dynamic stochastic asset-allocation model by 1 April 1991. For some months Russell researchers tried to formulate such a model without success. Then they asked me to formulate a stochastic programming model and work on that began in September 1989. The goal was to implement a model of Yasuda Kasai's financial planning process to improve their investment and liability payment decisions and their overall risk management.

The business goals were to:

1. maximize long-run expected wealth;
2. pay enough on the insurance policies to be competitive in current yield;
3. maintain adequate current and future reserves and cash levels; and
4. meet regulatory requirements, especially with the increasing number of savings-oriented policies being sold that were generating new types of liabilities.

The model needed more-realistic definitions of operational risks and business constraints than the return variance of previous mean-variance models used at Yasuda Kasai. The implemented model determines an optimal multiperiod investment strategy that enables decision makers to define risks in tangible operational terms, such as cash shortfalls. The risk measure used is convex and penalizes target violations more and more as the violations of various kinds and in various periods increase. The concave, risk averse objective is to maximize the discounted expected wealth at the horizon, net of expected discounted penalty costs incurred during the five periods of the model.

This objective is similar to that of a mean-variance model, except it is for five periods and counts only downside risk through target violations. Targets are used for multiple objectives but one must weight relative values of target violations across targets and time. This approach is preferable to value at risk (VAR) or conditional value at risk (CVAR) and its variants for ALM applications because, for most people and organizations, the non-attainment of goals is more and more damaging as the non-attainment increases. The loss is not linear in the non-attainment (as in CVAR) and VAR does not consider the size of the non-attainment at all. References on VAR and CVAR as risk measures are Artzner et al. (1999); Duffie and Pan (1997); and Jorion (2006). Krokhma, Uryasev and Zrazhevsky (2005) apply these measures to hedge fund performance. My risk measure is coherent in their sense, and its theoretical properties from an axiomatic

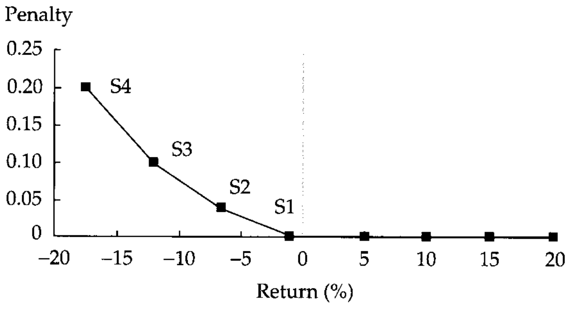


Fig. 34. Convex piecewise linear risk measure. Note: S = slope.

point of view are discussed in Rockefeller and Ziemba (2000). Figure 34 shows this measure. The piecewise linear function is used to maintain the model as a large stochastic linear program.

The model formulates and meets the complex set of regulations imposed by Japanese insurance laws and practices. The most important of the intermediate-horizon commitments is the need to produce income sufficiently high to pay the required annual interest in the savings-type insurance policies without sacrificing the goal of maximizing long-run expected wealth. During the first two years of use, fiscal 1991 and 1992, the investment strategy recommended by the model yielded a superior income return of 42 bps (US$ 79 million) over what a mean-variance model would have produced. Simulation tests also demonstrate the superiority of the stochastic programming scenario-based model over a mean-variance approach. In addition to the revenue gains, considerable organizational and informational benefits are evident.

6.3. *Formulation of the Russell-Yasuda Kasai model*

The basic Russell-Yasuda Kasai model has the following elements; see Cariño and Ziemba (1998) for additional details. This simplified formulation does not include additional types of shortfalls, indirect investments (tokkin funds and foreign subsidiaries), regulatory restrictions, multiple accounts, loan assets, the effects of taxes and end effects.

The stages are $t = 0, \ldots, T$.

Decision variables all at end t:

$$
\begin{aligned}
V_t &= \text{total fund market value at } t, \\
x_{nt} &= \text{market value in asset } n \text{ at } t, \\
w_{t+1} &= \text{income shortfall at } t+1, \ \text{and} \\
v_{t+1} &= \text{income surplus at } t+1.
\end{aligned}
$$

Random variables are:

$$RP_{n,t+1} = \text{price return of asset } n \text{ from } t \text{ to } t+1 \text{ and}$$
$$RI_{n,t+1} = \text{income return of asset } n \text{ from } t \text{ to } t+1.$$

Random variables appearing in the right-hand side are:

$$E_{t+1} = \text{deposit inflow from } t \text{ to } t+1,$$
$$P_{t+1} = \text{principal payout from } t \text{ to } t+1,$$
$$I_{t+1} = \text{interest payout from } t \text{ to } t+1,$$
$$g_{t+1} = \text{rate at which interest is credited to policies from } t \text{ to } t+1, \text{ and}$$
$$L_t = \text{liability valuation at } t.$$

The piecewise linear convex shortfall risk measure is $c_t(\cdot)$.

The objective of the model is to allocate discounted fund value among asset classes to maximize the expected wealth at the end of the planning horizon T less expected penalized shortfalls accumulated throughout the planning horizon.

$$\text{Maximize E}\left[V_T - \sum_{t=1}^{\text{T}} c_i(w_t)\right]$$

subject to budget constraints

$$\sum_n X_{nt} - V_t = 0$$

asset accumulation relations

$$V_{t+1} - \sum (1 + RP_{nt+1} + RI_{nt+1}) X_{nt} = F_{t+1} - P_{t+1} - I_{t+1}$$

income shortfall constraints

$$\sum_n RI_{nt+1} X_{nt} + w_{t+1} - v_{t+1} = g_{t+1} L_t$$

and non-negativity constraints

$$X_{nt} \geqslant 0, \quad v_{t+1} \geqslant 0, \quad w_{t+1} \geqslant 0$$

for $t = 0, 1, 2, \ldots, T-1$. Liability balances and cash flows are computed to satisfy the liability accumulation relations

$$L_{t+1} = (1 + g_{t+1}) L_t + F_{t+1} - P_{t+1} - I_{t+1}, \quad t = 0, \ldots, T-1.$$

The model has the base linear program which has a single scenario in the multiperiod deterministic formulation. Then the random variable scenarios are generated with a tree specification. Then, combining the base linear program and the tree, gives the model as a very large extensive form linear program.

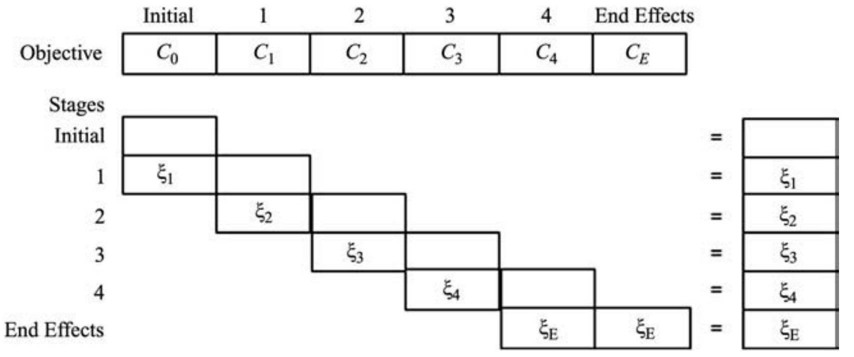


Fig. 35. Multistage stochastic linear programming structure of the Russell-Yasuda Kasai model. Note. C = cost coefficient. Source: Cariño and Ziemba (1998).

A stochastic linear program like Figure 35 plus a scenario tree like Figure 37 together yield a large linear program in extensive form.

The dimensions of the implemented problem were

Par	BrPar	Scan	Assets	Var	Alloc			GLP		
					Rows	Cols	Coeff	Rows	Cols	Coeff
INI	1	1	7	59	22	60	257	22	60	257
Q01	8	8	7	59	48	35	573	384	680	4584
Y01	4	32	7	59	54	96	706	1728	3072	22392
Y02	4	128	7	29	52	66	557	6658	8448	71296
Y05	2	256	7	29	52	66	407	13312	16896	104192
YFF	1	256	5	21	35	58	450	8960	14848	115200
Total					263	431	2950	31062	44004	318121

The model equations are shown in block form in Figure 35, where a ξ_t means there is uncertainty in that block. The model has 256 scenarios over four periods plus a fifth end-effects period. The model is flexible regarding the time horizon and length of decision periods, which are multiples of quarters. A typical application has initialization plus Period 1 to the end of the Quarter 1; Period 2, the remainder of Fiscal Year 1; Period 3, the entire Fiscal Year 2; Period 4, Fiscal Years 3, 4, and 5; and Period 5, the End-Effects Years 6 on to forever; see Figure 35. Figure 36 shows the decision-making process.

Yasuda Kasai faced the following situation:

- An increasing number of savings-oriented policies were being sold that had new types of liabilities.
- The Japanese Ministry of Finance insurance laws created restrictions which led to complex constraints.

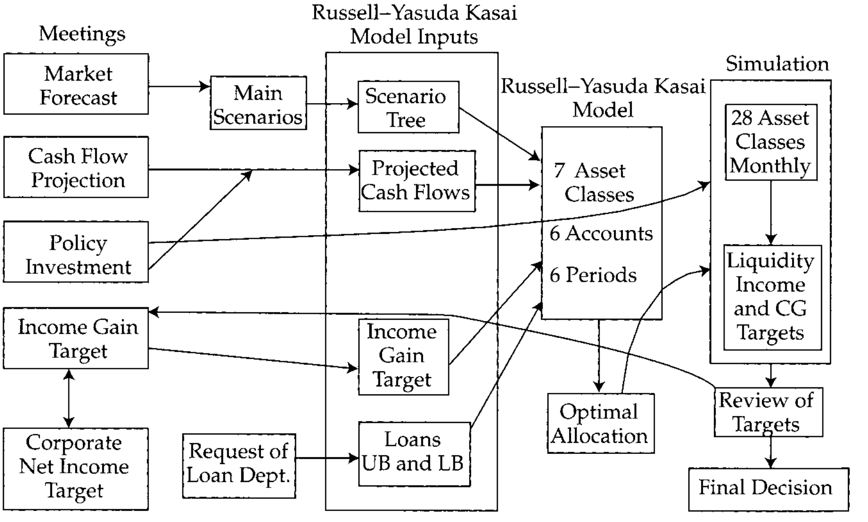


Fig. 36. Yasuda Kasai's asset/liability decision-making process. Note: UB = upper bound; LB = lower bound; CG = company growth. Source: Cariño et al. (1994).

Table 15
Yasuda Kasai's balance sheet

Assets	Liabilities
Cash	General account
Loans	Savings account
Fixed income	Special savings 1
Equities	Special savings 2
Foreign fixed income	Special savings 3
Foreign equities	Special savings 4
Other	Net worth

Source: Cariño and Ziemba (1998).

- The firm's goals included both current yield and long-run total return, so risks and objectives were multidimensional.

The insurance policies were complex, being part actual insurance and part an investment with a fixed guaranteed amount plus a bonus, dependent on general business conditions in the industry. The insurance contracts were of varying length; maturing, being renewed or starting in various time periods, and subject to random returns on assets managed, insurance claims paid, and bonus payments made. Table 15 shows the insurance company's balance sheet with various special savings accounts.

Table 16
Example of regulations on foreign investments

	Account 1				Account 2			
	Direct	Tokkin	Foreign	Subsidiaries	Direct	Tokkin	Foreign	Subsidiaries
Cash	x				x			
Fixed income	x				x			
Equities		x	x	x		x	x	x
Foreign fixed income	x	x	x		x	x	x	
Foreign equities	x	x	x		x	x	x	

Note: x means the investment is allowed, and a blank means it is not. Source: Cariño and Ziemba (1998).

Table 17
Asset classes for the Russell-Yasuda Kasai model

Asset	Associated index
Cash bonds	Nomura Bond Performance Index
Convertible bonds	Nikko Research Convertible Bond Index
Domestic equities	Tokyo Stock Price Index (TOPIX)
Hedged foreign bonds	Salomon Brothers World Bond Index (or hedged equivalent)
Hedged foreign equities	Morgan Stanley World Equity Index (or hedged equivalent)
Unhedged foreign bonds	Salomon Brothers World Bond Index.
Unhedged foreign equities	Morgan Stanley World Equity Index
Loans	Average lending rates (trust/long-term credit or long-term prime rates)
Money trusts, etc.	Call rates (overnight with collateral)

Note: Life insurance company general accounts are an asset class but have no associated index. Source: Cariño and Ziemba (1998).

There were many regulations on assets, including restrictions on equity, loans, real estate, foreign investment by account, foreign subsidiaries, and tokkin funds (pooled accounts). Table 16 depicts an example of a regulation on investing in foreign investments. How should one manage investment assets in this environment? The goal was to implement a model of Yasuda Kasai's financial planning process to improve the investment decisions and overall risk management. Yasuda Fire and Marine staff would operate the software model independently. The model was to run on a single IBM RS/6000 workstation with a computing time of three hours or less. All significant input was parameterized. Risks were defined in operational terms as the difference between actual and targeted income, cash flow shortages, or capital losses in cash money terms, not less interpretable concepts like variance. The asset classes are shown in Table 17.

Dividing the universe of available investments into a manageable number of asset classes involves a trade-off between detail and complexity. A large number of asset classes increases detail at the cost of increasing size. So the model allows the number

Table 18
Typical asset class list

Indentifier	Description	Identifier	Description
CJ	Cash	BF	Bonds (foreign)
LL	Loans (floating rate)	EU	Equity (US)
LF	Loans (fixed rate)	EN	Equity (non-US foreign)
LO	Loans	EF	Equity (foreign)
BJ	Bonds (domestic)	FO	Foreign assets
EJ	Equity (domestic)	RE	Real estate
BU	Bonds (US and Canada)	OT	Other
BA	Bonds (UK and Australia)	GE	Generic assets
BE	Bonds (European continent)		

Source: Cariño and Ziemba (1998).

Table 19
Investment types

Identifier	Description
D	Direct
T	Tokkin
C	Capital to foreign subsidiaries
L	Loans to foreign subsidiaries

Source: Cariño and Ziemba (1998).

and definition of asset classes to be specified by the user. There can be different asset classes in different periods. For example, asset classes in earlier periods can be collapsed into aggregate classes in later periods.

Table 18 presents a typical asset class list. Investment in asset classes may be done either directly or indirectly through *tokkin* funds (pooled asset vehicles), capital ownership of foreign subsidiaries, or loans to foreign subsidiaries. The regulatory rules that apply to indirect investments yield the investment types shown in Table 19.

The model chooses an asset allocation. The variables represent the market values chosen for each class. Even though the number of allocation variables could potentially equal the product of the number of asset classes, the number of investment types, the number of accounts, and some asset class investment-type account combinations are disallowed either by regulations or by company policy. An input table controls the available allocation classes of the model, and the user can specify fewer available allocation classes in later periods. A typical number of allocation variables by stage is shown in Table 20.

The primary recommendations from the model are

- a market value allocation for each asset class in each period,

Table 20
Allocation variables by stage

t	Stage name	Account					Total
		S	1	2	G	E	
0	INI	15	13	15	15	1	59
1	Q01	15	13	15	15	1	59
2	Y01	15	13	15	15	1	59
3	Y02	7	7	7	7	1	29
4	Y05	7	7	7	7	1	29
5	YFF	5	5	5	5	1	21

Note: S = a saving product; G = general; and E = exogenous.

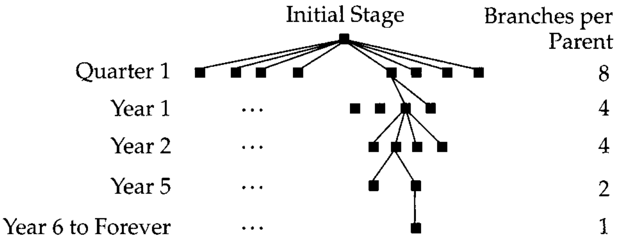


Fig. 37. Scenario tree. Note: Total number of scenarios was 256. Each node represents a joint outcome of all the asset class returns. Source: Cariño and Ziemba (1998).

- a book value allocation for each asset class in each time period,
- amount of asset purchases and sales in each time period,
- expected wealth at model horizon points, and
- shortfalls versus model targets in each time period.

The model had the 256 scenarios shown in Figure 37. The end-effects period is Year 6 to forever and is a steady-state period in insurance policy growth. All scenario trees have one scenario in the last stage.

The scenarios were generated using three models:

(1) independent across periods;

(2) a vector autoregressive factor model that uses interest rates measured by the long-term government bond yields and the first section of the Tokyo Stock Exchange (Tokyo Stock Price Index [TOPIX] returns), and currency measured by the yen versus US dollar return; and

(3) a general model in which the user can specify the scenarios as desired.

The model has the following elements:

Matrix generator. Creates the base linear program (LP) plus a set of random coefficient specifications for each period.

Table 21
Expected allocations for initialization period: INI

	Total (100 million)	Percentage
Cash	2,053	9%
Loans (floating rate)	5,598	26
Loans (fixed rate)	5,674	26
Bonds	2,898	13
Equity	1,436	7
Foreign bonds	3,277	15
Foreign equity	875	4
Total	21,800	100%

Note: Total book value 1 = 22,510 (¥100 million). Total book value 2 = 34,875 (¥100 million). Source: Cariño and Ziemba (1998).

Table 22
Expected allocations in the end-effects period (¥100 million)

	General	Savings	Special Savings 1	Special Savings 2	Exogenous	Total	Percentage
Cash	0	44	0	36	0	80	0.1%
Bonds	5,945	17	14,846	1,311	0	22,119	40.1
Equity	0	0	4	0	18,588	18,592	33.7
Foreign bonds	2,837	1,094	0	0	0	3,931	7.1
Foreign equity	0	4,650	6,022	562	0	11,234	20.4
Total	8,782	5,804	20,072	1,908	18,588	55,154	

Note: Total book value 1 = 28,566. Total book value 2 = 50,547. Source: Cariño and Ziemba (1998).

Scenario generator. Builds the decision-tree structure and generates the random returns for each asset class.

Liability generator. Projects the random liabilities for each decision node.

Coefficient generator. Combines the random coefficient specifications with the random variables to generate random coefficients.

Solver. Generates the optimal solution.

A major part of the information from the model consists of reports with tables and figures of model output. Actual year by year asset-allocation results from the model are confidential. But we have the following expected allocations in the initialization (Table 21) and the end-effects periods (Table 22). These results are averages across all 256 scenarios in ¥100 million units and percentages by account.

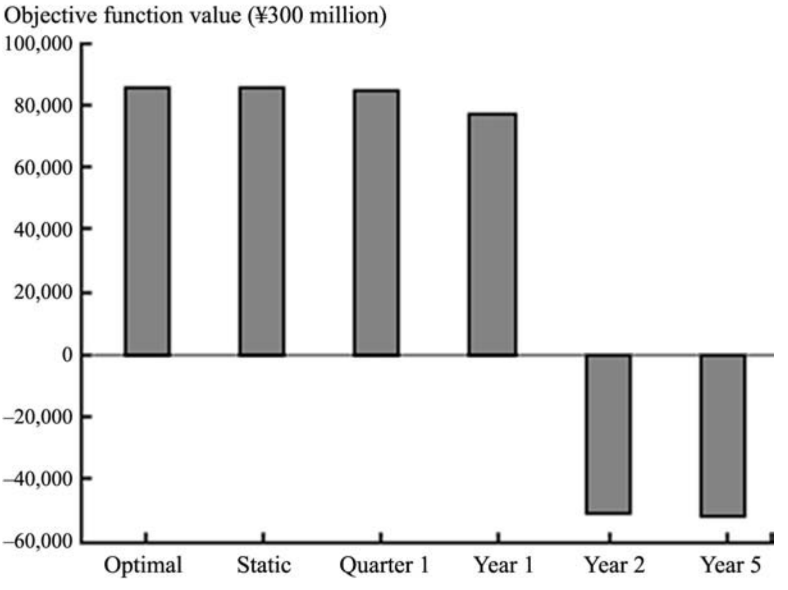


Fig. 38. The effect of mean-variance approximations on the optimal stochastic programming solution. Source: Cariño and Ziemba (1998).

6.4. How good is the model?

When one builds and implements a model, the question is: How good is it? A first test is: Is it better than the model currently being used (which was a mean-variance model)? This comparison follows Kusy–Ziemba's (1986) comparison of their multi-period stochastic linear programming model (solved via two-period technology), with the dynamic decision-tree approach of Bradley and Crane (1972, 1973). The comparison set up plausible situations and had each model compute optimal solutions. Then, by simulation, it compared how good the solutions were using statistical tests. They found that the stochastic linear programming model was greatly superior to the decision-tree dynamic programming approach. The flexibility of stochastic programming was a major reason for this superiority.

For the Russell-Yasuda Kasai application, after one year, the mean-variance model gives extremely poor solutions. This finding is shown by the effect of the number of periods that are fixed with the mean-variance allocations versus those having the stochastic programming solution. For example, Figure 38 shows the optimal multiperiod stochastic programming solution. This is compared to the suboptimal policies of using the mean variance approximation for more and more periods. Extensive use of the mean variance approximation leads to poor performance relative to the optimal solution.

A second test is: How did the model allocations do in actual use with real money in fiscal 1991 and 1992 (1 April–31 March)? *From Russia with Love* fans will remember the phrase, "we use live targets as well"; out-of-sample tests, especially with real money, are the ultimate test of a model's success. The results are as follows:

Fiscal 1991. Positive comparison versus previous mean-variance constant mix strategy: +15 bps (US$ 25 million) in income yield.

Fiscal 1991–1992. Combined +42 bps better (US$ 79 million) in income yield.

In fiscal 1991, both the actual income return and total return were more than 7%.

During the same period, the Nikkei Stock Average lost 1.84%. Total return in 1991–1992 combined was +5 bps better (US$ 9 million). Also, the risk management was greatly improved.

Summary:

- The 1991 Russell-Yasuda Kasai model was then the largest application of stochastic programming in financial services.
- There was a significant ongoing contribution to Yasuda Kasai's financial performance —US$ 79 million and US$ 9 million in income and total return, respectively, over fiscal 1991–1992—and the model has been in use since then.
- The basic structure was portable to other applications because of its flexible model generation.
- The model had a substantial potential impact on the performance of financial services companies.
- The top 200 insurers worldwide had in excess of US$ 10 trillion in assets.
- The industry moved toward more complex products and liabilities and risk-based capital requirements.

7. Pension models: Aging of the World's populations

In January 2003, the pension fund consultants at Watson Wyatt Worldwide estimated that global pension funds had balance sheets of minus US$ 2.5 trillion (assets minus liabilities). Assets totaled US$ 10.7 trillion, the same as in 1997. Liabilities are growing because people are living longer, and assets are falling because of the equity declines in 2000–2002.

Pensions around the world and those who manage and guarantee them will need to deal with the serious problem that the world's populations are aging rapidly. Also, an enormous group of retirees will be needing to cash out of pension and other portfolios at the same time. In Europe, the percentage of people in the 65 and older group will roughly double from 1990–2030, from 20 to 40% (see Table 23). By 2030, two workers will have to support each pensioner, compared with four workers in 1990. Better living conditions, more-effective medical systems, declining fertility rates, and low immigration have all contributed to this aging phenomenon.

The European, Japanese, British, Irish, American, and Canadian pension fund situations are quite different. In Europe, the bulk of pension payments are paid by the state. These are called "Pillar 1" and amount to about 88% of total pension costs. Without changes, the pension payouts in much of the European Union (EU) will grow from

Table 23
Elderly dependence ratio projections in Europe and OECD projections of pension cost as a percentage of GDP, 1990–2030

Country	Dependency ratio			Pension cost as percentage of GDP					
	1990	2010	2030	1995	2000	2010	2020	2030	2040
Austria	22.4	27.7	44.0	NA	NA	NA	NA	NA	NA
Belgium	22.4	25.6	41.1	10.4%	9.7%	8.7%	10.7%	13.9%	15.0%
Denmark	22.7	24.9	37.7	6.8	6.4	7.6	9.3	10.9	11.6
Finland	19.7	24.3	41.1	10.1	9.5	10.7	15.2	17.8	18.0
France	20.8	24.6	39.1	10.6	9.8	9.7	11.6	13.5	14.3
Germany	21.7	30.3	49.2	11.1	11.5	11.8	12.3	16.5	18.4
Ireland	18.4	18.0	25.3	3.6	2.9	2.6	2.7	2.8	2.9
Italy	21.6	31.2	48.3	13.3	12.6	13.2	15.3	20.3	21.4
Netherlands	19.1	24.2	45.1	6.0	5.7	6.1	8.4	11.2	12.1
Portugal	19.5	22.0	33.5	7.1	6.9	8.1	9.6	13.0	15.2
Spain	19.8	25.9	41.0	10.0	9.8	10.0	11.3	14.1	16.8
Sweden	27.6	29.1	39.4	11.8	11.1	12.4	13.9	15.0	14.9
United Kingdom	24.0	25.8	38.7	4.5	4.5	5.2	5.1	5.5	5.0
European Union average	21.4	25.9	40.3						

NA = not available. Note: Ratio is for those 65 and over percentage of population. OECD = Organization for Economic Cooperation and Development. Source: Based on data from Bos et al. (1994) and Roseveare et al. (1996).

about 10% of GDP in 1997 to more than 15% in 2030. Contribution rates will have to be raised significantly to enable the public social security system to remain solvent. Also, effective private pension plans will need to play a more major role, given the demand for health care and other social services in addition to pensions. For some countries, however, such as the UK and Ireland, where pension schemes linked to employment (second pillar) and individual pensions (third pillar) are more prevalent, the pension costs will remain stable over the projection period (see Table 23).

How much in assets do these countries have in reserve? Table 24 shows that, except for the United Kingdom, the Netherlands, and to a lesser extent Ireland and Sweden, pension fund assets as a percentage of GDP are very low. Burless (2002), using data from 1927 to 2001, shows that in five major industrial countries (France, Germany, Japan, the UK and the US), individual savings accounts are not sufficient for a safe retirement. Hence, retirees need additional resources.

One way to deal with the problem of large pension payments is to have good risk-adjusted returns, and the models in this *Handbook* are an attempt to provide such results by taking into account various problem elements and uncertainties. The countries with the more aggressive pension fund managers—Ireland, the United Kingdom, and the United States—not surprisingly, have had higher returns than other European countries (see Table 25).

Table 24
Pension fund assets: Total, as a percentage of GDP, and allocation, 1997

Countries	Assets (US$ billions)	GDP (US$ billions)	Percentage of GDP	Allocation of assets				
				Equity	Fixed income	Real estate	Cash	Other
Austria	20.9	181.8	11.5%	4.1%	82.4%	1.8%	1.6%	10.0%
Belgium	10.3	213.8	4.8	47.3	41.3	5.2	5.6	0.6
Denmark	29.3	143.7	20.4	23.2	58.6	5.3	1.8	11.1
Finland	8.9	103.6	8.6	13.6	55.0	13.0	18.2	0.0
France	84.4	1,229.1	6.9	12.6	43.1	7.9	6.5	29.9
Germany	270.7	1,865.4	14.5	9.0	75.0	13.0	3.0	0.0
Greece	4.6	105.0	4.4	7.0	62.9	8.3	21.8	0.0
Ireland	34.5	64.1	53.8	58.6	27.1	6.0	8.0	0.4
Italy	21.6	1,010.7	2.1	4.8	76.4	16.7	2.0	0.0
Luxembourg	0.03	13.7	0.2	23.7	59.0	0.0	6.4	11.0
Netherlands	361.7	320.0	113.0	36.6	51.3	5.2	1.5	5.2
Portugal	9.4	86.0	10.9	28.1	55.8	4.6	8.8	2.7
Spain	18.7	470.4	4.0	11.3	60.0	3.7	11.5	13.5
Sweden	96.2	202.4	47.5	40.3	53.5	5.4	0.8	0.1
United Kingdom	891.2	1,127.3	79.1	72.9	15.1	5.0	7.0	0.0
Total								
European Union	1,862.4	7,137.0	26.1	53.6	32.8	5.8	5.2	2.7
United States				52	36	4	8	na
Japan				29	63	3	5	na

na = not applicable. Source: Based on data from the 1996 report of the European Federation for Retirement Provision.

Table 25
Average real annualized pension fund net returns, 1984–93

Country	Return
Restrictive	
Belgium	8.8%
Denmark	6.3
Germany	7.1
Netherlands	7.7
Spain	7.0
Average	7.4%
Aggressive	
Ireland	10.3%
United Kingdom	10.2
United States	9.7
Average	10.1%

Source: European Commission (1997).

These superior results are related to the asset-allocation weights; the advanced countries have more equities and most of the EU countries listed in Table 24 have the bulk of their capital invested in bonds. In the models developed here, these weights are scenario and time dependent.

7.1. Why do European pension fund managers invest so much in bonds?

European pension plans have a strong preference for bond holdings, as shown in Table 24. More “mature” Pillar 2 countries, such as the United Kingdom and Ireland, which have managed portfolios for foreign as well as domestic investors for a long time, have a higher equity exposure, which may better reflect the long-term aspect of pension obligations. In countries such as Austria, Germany, Italy, Spain, and France, equity markets that were not developed until recently have pension plans that are invested more heavily in local government bonds.

In addition, such asset structures also reflect the attitude toward equities in various countries. With the introduction of the euro in 1999, a first important step toward a more integrated capital market, especially for equities, was made. In Austria, pension funds are now starting to increase their equity positions, but it will take some time to reach a structure similar to that in well-established US, UK, and Irish pension industries. Strict regulations, availability of investment products, fear of foreign investment, a short-term outlook, and tradition led to this policy in the past. The regulations, especially the perception of them, are still not flexible enough to allow pension managers to diversify their portfolios across asset classes, currencies, and worldwide markets.

Some changes, however, are on the horizon from the European Commission. The new proposals would allow European pensions more freedom to invest in equities and in foreign assets and currencies. The limit for worldwide equities would rise to 70% versus the current average of about 35% in EU countries.

The high percentage of bond allocations in European pension funds, shown in Table 24, has had a substantial effect on actual performance. Table 25 shows annualized real pension fund returns in a subset of five EU countries versus those in the US, the UK, and Ireland. Not surprisingly, the more advanced, more aggressive investment styles in the United States, the UK, and Ireland led to returns that were about 3% higher a year which over a ten year period amounts to a huge under performance in Austria and similarly managed pension funds in other countries in Europe.

Using measures such as the Sharpe ratio or the capital asset pricing model, studies, such as Dimson, Marsh and Staunton (2002, 2006, 2007), Keim and Ziemba (2000), and Jorion and Goetzmann (1999), have indicated that over long periods, equity returns outperform bond returns in risk-adjusted terms. Moreover, the historical evidence since 1802 for the United States and 1700 for the UK indicates that the longer the period, the more likely this dominance by equities. Recall the calculations in Table 9 from Siegel (2002); for all 20-year periods from 1926–2001, US equities outperformed bonds, and for 30-year horizons, based on the past, it has been optimal (with a mean-variance model) to be more than 100% in stocks and short bonds.

Hensel and Ziemba (2000) showed how the slow but steady outperformance of assets can lead to dramatically higher total wealth levels over long periods. For example, for the United States, during the 1942–97 period, a strategy that was 100% in US small-cap stocks with Democratic administrations and 100% in large-cap stocks in Republican administrations had, in 1997, 24.5 times as much wealth as the typical 60/40 percent stock/bond mix used in most US pension funds. How much to invest in cash, stocks, and bonds over time is a deep and complex issue. For a theoretical analysis in which the uncertainty of mean reversion is part of the model, see Barberis (2000). One thing is clear: Equities have had an enormous advantage over cash and bonds during most periods in most countries, so the optimal blend is to have much more equity than 4.1% of Austria.

The case for equities, however, is not as clear-cut as Figure 16 might indicate. Figure 17 shows how bumpy the gains have been for the DJIA in real terms. Hence, investors who need funds for liability commitments, such as pensioners, may well have much poorer results. Indeed, despite the near linearity of the growth of equity values in Figure 16, there were three long subperiods of essentially zero nominal equity growth, not counting dividends, in the 20th century: 1900–1920, 1929–1954, and 1964–1981.

For example, the DJIA was 66.08 on 31 December 1899, and 71.95 on 31 December 1920, a rise of only 0.4% a year. By September 1929, the DJIA was 381, a 430% increase in less than nine years. But then it fell to 177 in 1946, half its 1929 level. Then, on 31 December 1964, the DJIA was 874.12, and it was essentially the same at 875.00 on 31 December 1981, 17 years later. But 17 years later on 31 December 1998, it was more than 10 times higher at 9181.43. Interest rates were crucial because long-term US government bonds were yielding 4.20% on 31 December 1964, 13.65% on 31 December 1981, and 5.09% on 31 December 1998.

So, the conclusion, to paraphrase Warren Buffett (2001), is that equity prices have risen dramatically since 1900 in the United States, but during three long periods of 20, 25, and 17 years, stocks had essentially zero gains, or even losses, in nominal terms.

In the US, notable examples of institutions close to pension funds that have had very high risk-adjusted returns from a variety of private placement hedge fund and other investments without high equity exposures are the endowments of Harvard and Yale universities and the Ford Foundation (see Swensen, 2000; Clifford, Kroner and Siegel, 2001; Ziemba, 2005; and Ziemba and Ziemba, 2007). The higher equity proportions or other ways to increase real returns would have resulted in better-funded pension plans, in higher pension payments, or lower contribution rates for companies. Of course, this outperformance is predicated on a continuing high equity risk premium and is volatility dependent. Between 1982 and 1999, the return of equities over bonds was more than 10% a year in EU countries. These high equity returns of the distant past and the bull market of the late 1990s, however, led to valuations of P/E and other measures that were at historically high levels in Europe, the United States, and elsewhere. Studies by Siegel (2002), Campbell and Shiller (1998, 2001), Shiller (2000), Koivu, Pennanen and Ziemba (2005), Ziemba (2003), Berge and Ziemba (2006), and Ziemba and Ziemba (2007) have suggested that this outperformance is unsustainable, and the weak equity

Table 26
National investment restrictions on pension plans

Country	Investment restrictions
Germany	Maximum: 30% equities/5% foreign bonds
Austria	Maximum: 40% equities/45% foreign securities; minimum: 40% Eurobonds
France	Minimum: 50% Eurobonds
Portugal	Maximum: 35% equities
Sweden	Maximum: 25% equities
United Kingdom	Prudent man rule
United States	Prudent man rule

Source: European Commission (1997).

returns are consistent with this view. The long-run results indicate equity outperformance, however, and in the future, this historical result may well be continued and thus needs to be reflected in the scenarios.

Pension fund managers that have been mostly invested in bonds face a dilemma. Should they move more into assets that have historically had higher mean returns along with higher variance or stick with what has worked satisfactorily, if not spectacularly, in the past? Of course, what other pension funds do is a factor in evaluating fund performance, especially in the use of specified benchmark performance evaluation levels. The specification of the type of benchmark (a linear combination of assets) around which the fund is to be evaluated greatly influences pension investment behavior. InnoALM, the Innovest Austrian pension fund financial planning model described below, was designed to help pension fund managers prudently make these choices by taking essentially all aspects of the problem into account.

For example, Austrian pension fund managers have had considerably more flexibility in their asset allocation decisions than the investment rules in shown in Table 26 might indicate. For example, if an investment vehicle was more than 50% invested in bonds, then that vehicle was considered to be a bond fund. So, investment in 45% equities and 55% in bond funds (whose average bond and stock weightings are 60/40), yields an average equity is 67%, which is similar to that of the higher-performing UK managers. Moreover, currency hedged assets are considered to be euro denominated. Hence, the minimum of 40% in Eurobonds is effectively a 40% limit on worldwide bonds, but because of the above rules on the weighting of assets, this limit is not really binding. In addition, the 5% rule on option premiums means that managers have had effectively full freedom for worldwide asset allocations. Such use of the rules, however, was not typical by actual pension fund managers. In some scenarios, such allocations away from the asset allocation typical in other Austrian pension funds could have led to disaster. So, without being armed with a model such as InnoALM that would calculate the possible consequences of asset-weight decisions, it was safest for managers to go with the crowd.

The European Commission (1999) stressed the importance of a relaxation of restrictive quantitative rules on pension fund investing. The diversification of investments is more important than rules on different investments. It recommends the use of modern asset and liability management techniques to achieve this diversification goal.

The following section describes a model for the effective operation of private pension funds in Austria. These funds usually work on a funded basis in which the pension benefits depend on an employment contract or the pursuit of a particular profession. Schemes are administered by private institutions, and benefits are not guaranteed by the state. These occupational pension schemes are widespread throughout Europe. Normally, contributions to such systems are made by the employer and on an optional basis for additional benefits by employees. The contribution level may depend on the wage level or the position within a company. Defined-contribution plans (DCPs) have fixed contributions, and the payout depends on the capital accumulation of the plan. Defined-benefit plans (DBPs) have payouts guaranteed by the company, and the contribution is variable, depending on the capital accumulation over time.

An important difference between these two methods is the risk-bearer position. In DBPs, the employer guarantees the pension payment, which is usually tied to some wage at or near retirement. Hence, the company has to inject money into the pension plan if asset returns do not cover pension liabilities. However, the company gains, or, equivalently, reduces future contributions if the asset returns of the plan are higher than required to fund the liabilities. For DCPs, which have become more popular, the employees and pensioners bear the risk of low asset returns. Their pensions are not fixed and depend on the asset returns. High returns will increase pensions and vice versa. No direct financial risk for the employer is incurred, although with poor returns, the employer could suffer from a negative image, if, for example, the following headline appeared: "Pensions for the Siemens Pensioners Have to be Reduced by 3 Percent in the Next Year." The Siemens pension plan for Austria is a DCP, but InnoALM was designed to handle either pension system.

The steep fall in worldwide equity prices in 2000–2003, especially in 2002, caused a major crisis for insurance companies and pension funds and many others that have considerable long equity positions. Especially in trouble are DBPs that guarantee fixed payments until the death of the pensioners and their spouses. In England, this issue is already causing many discussions of rule changes, such as changing the age at which someone can receive retirement benefits to 70 instead of the current 65, and asset-allocation decisions, such as shifting into bonds from equities at very depressed prices, thereby locking in large losses. The UK company Boots Group has moved completely into bonds, a move that had good success in the short term but appears to be suboptimal in the long term. The University of Toronto pension fund lost $484 million in the year ending April 2003 using similar suboptimal strategies. At least Boots did the bond allocation before the equity fall and bond rise and matched liabilities to these assets.

The case for bond-only pension funds was made by Bodie (2001). The theoretical idea is to eliminate liabilities with a series of bonds of varying durations held to maturity. Then, the liabilities are taken care of and the interest rate risk is eliminated. TIPS (US Treasury inflation-indexed securities) that return 3.4% plus inflation over a 30-year cycle can be used to mitigate the effects of inflation. My view in this chapter is to consider all assets in relation to the future scenarios in one's asset-allocation mix, including TIPS. A policy that uses only one asset class must be suboptimal. Nonetheless, the idea is of interest to those who cannot predict scenarios, such as strict efficient market proponents. Oxford, Cambridge, Stanford, Princeton, and other universities have successfully used a similar idea to buy land and housing near the university to subsidize students and faculty. But Oxford and Cambridge had poor endowment returns in 2000–2003 because of their high equity allocations.

7.2. The Innovest Austrian Pension Fund financial planning model

Siemens Oesterreich, part of the global Siemens Corporation, is the largest privately owned industrial company in Austria. Its businesses, with revenues of €2.4 billion in 1999, include information and communication networks, information and communication products, business services, energy and traveling technology, and medical equipment. Its pension fund, established in 1998, is the largest corporate pension plan in Austria and is a DCP. More than 15,000 employees and 5,000 pensioners are members of the pension plan, which had €510 million in assets under management as of December 1999.

Innovest Finanzdienstleistungs, founded in 1998, is the investment manager for Siemens Oesterreich, the Siemens pension plan, and other institutional investors in Austria. With €2.2 billion in assets under management, Innovest focuses on asset management for institutional money and pension funds. Of 17 plans analyzed in the 1999–2000 period, it was rated as the best plan in Austria. The motivation to build InnoALM, as described in Geyer and Ziemba (2007) and their earlier working paper (Geyer et al., 2002), was based on the desire to have superior performance and good decision aids to help achieve this ranking.

Various uncertain aspects, possible future economic scenarios, stocks, bonds, and other investments, transaction costs, liquidity, currency aspects, liability commitments over time, Austrian pension fund law, and company policy suggested that a good way to approach this asset–liability problem was via a multiperiod stochastic linear programming model. These models evolved from Kusy and Ziemba (1986); Cariño et al. (1994); Cariño and Ziemba (1998); Cariño, Myers and Ziemba (1998); and Ziemba and Mulvey (1998). This model has innovative features, such as state-dependent correlation matrices, fat-tailed asset-return distributions, simple computational schemes, and output.

InnoALM was produced in six months in 2000, with Geyer and Ziemba serving as consultants and with assistance from Herold and Kontriner who were Innovest employees. InnoALM demonstrates that a small team of researchers with a limited budget can quickly produce a valuable modeling system that can easily be operated by nonstochastic programming specialists on a single personal computer (PC). The IBM OSL stochastic programming software provides a good solver. The solver was interfaced

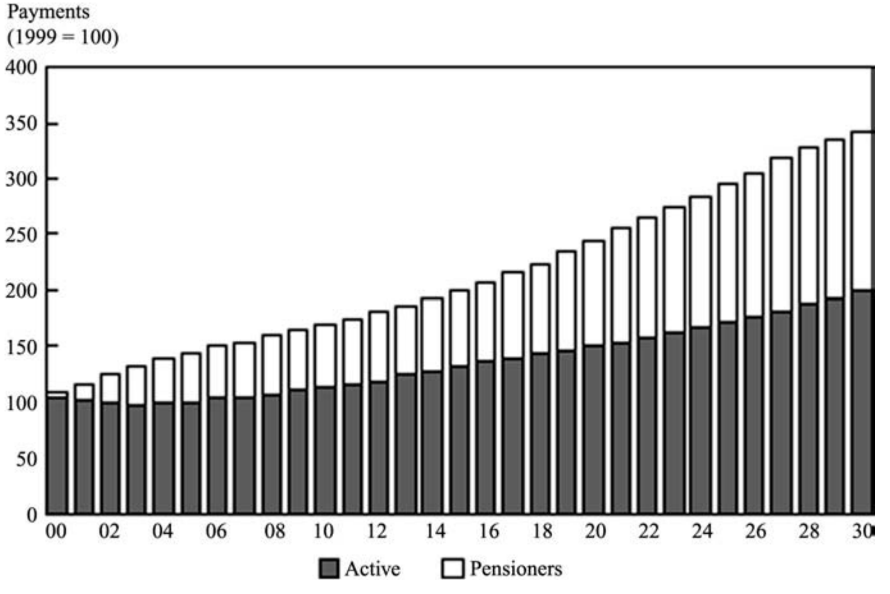


Fig. 39. Index of expected payments for active and retired employees, 2000–2030. Source: Geyer and Ziemba (2007).

with user-friendly input and output capabilities. Calculation times on the PC are such that different modeling situations can be easily developed and the implications of policy, scenario, and other changes can be seen quickly. The graphical output provides pension fund management with essential information to aid in the making of informed investment decisions and understanding the probable outcomes and risk involved with these actions. The model can be used to explore possible European, Austrian, and Innovest policy alternatives.

The liability side of the Siemens pension plan consists of employees for whom Siemens is contributing DCP payments and retired employees who receive pension payments. Contributions are based on a fixed fraction of salaries, which varies across employees. Active employees are assumed to be in steady state; thus, employees are replaced by a new employee with the same qualification and sex so that there is a constant number of similar employees. Newly employed staff start with less salary than retired staff, which implies that total contributions grow less rapidly than individual salaries. Figure 39 shows the index of expected total payments for active and retired employees until 2030.

The set of retired employees is modeled using Austrian mortality and marital tables. Widows receive 60% of the pension payments. Retired employees receive pension payments after reaching age 65 for men and 60 for women. Payments to retired employees are based on the individually accumulated contribution and the fund performance during active employment. The annual pension payments are based on a discount rate of

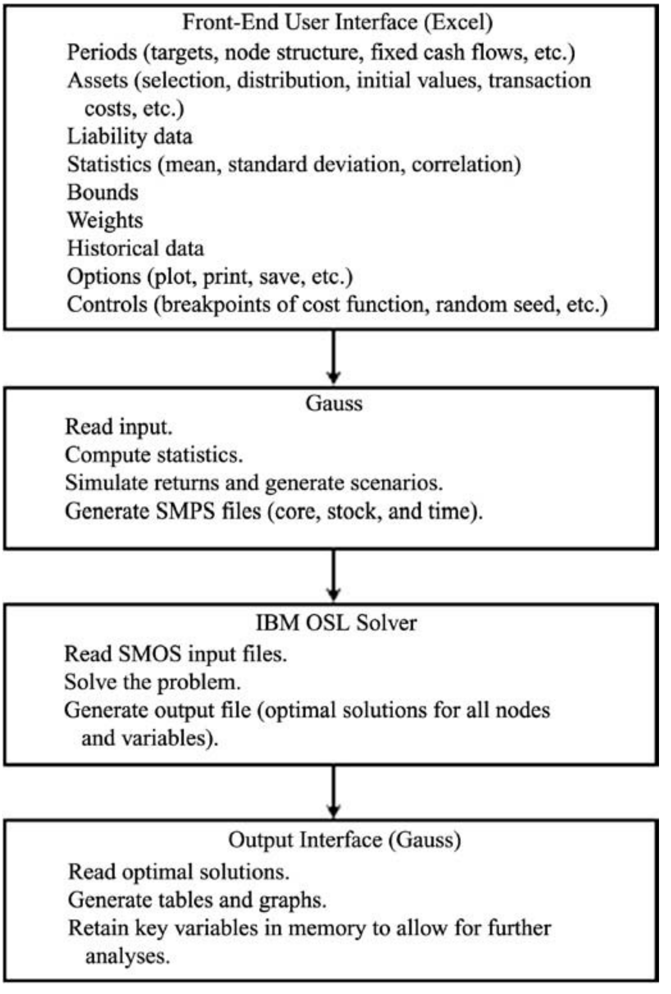


Fig. 40. Elements of InnoALM. Source: Geyer and Ziemba (2007).

6% and the remaining life expectancy at the time of retirement. These annuities grow by 1.5% annually to compensate for inflation. Hence, the wealth of the pension fund must grow by 7.5% a year to match liability commitments. Another output of the computations is the expected annual net cash flow of plan contributions minus payments. Because the number of pensioners is rising faster than plan contributions, these cash flows are negative and the plan is declining in size.

The model determines the optimal purchases and sales for each of N assets in each of T planning periods. Typical asset classes used at Innovest are US, Pacific, European, and emerging market equities and US, UK, Japanese, and European bonds. The objective is to maximize the concave risk-averse utility function "expected terminal wealth" less convex penalty costs subject to various linear constraints. The effect of such constraints is evaluated in the examples that follow, including Austria's limits of 40% maximum in equities, 45% maximum in foreign securities, and 40% minimum in Eurobonds. The convex risk measure is approximated by a piecewise linear function, so the model is a multiperiod stochastic linear program. Typical targets that the model tries to achieve (and is penalized for if it does not) are for growth of 7.5% a year in wealth (the fund's assets) and for portfolio performance returns to exceed benchmarks. Excess wealth is placed into surplus reserves, and a portion of the excess is paid out in succeeding years.

The elements of InnoALM are described in Figure 40. The interface to read in data and problem elements uses Excel. Statistical calculations use the program Gauss, and these data are fed into the IBM OSL solver, which generates the optimal solution to the stochastic program. The output, some of which is shown in the next section, uses Gauss to generate various tables and graphs and retains key variables in memory to allow for future modeling calculations.

7.3. *Formulating InnoALM as a multistage stochastic linear programming model (based on Geyer and Ziemba, 2007)*

The non-negative decision variables are wealth (after transactions costs) W_{it}, and purchases P_{it} and sales S_{it} for each asset ($i = 1, \dots, N$). Purchases and sales are in periods $t = 0, \dots, T-1$. Purchases and sales are scenario dependent except for $t = 0$.

Wealth accumulates over time for a T period model according to

$$\begin{aligned}
W_{i0} &= W_0^{\text{init}} + P_{i0} - S_{i0}, && t = 0,\\
\widetilde{W}_{i1} &= \widetilde{R}_{i1} W_{i0} + \widetilde{P}_{i1} - \widetilde{S}_{i1}, && t = 1,\\
\widetilde{W}_{it} &= \widetilde{R}_{it} \widetilde{W}_{i,t-1} + \widetilde{P}_{it} - \widetilde{S}_{it}, && t = 2, \dots, T-1,\\
\widetilde{W}_{iT} &= \widetilde{R}_{it} \widetilde{W}_{i,T-1}, && t = T.
\end{aligned}$$

W_i^{init} is the prespecified initial value of asset i. There is no uncertainty in the initialization period $t = 0$. Tildas denote scenario-dependent random parameters or decision variables. Returns are associated with time intervals. $\widetilde{R}_{it}$ ($t = 1, \dots, T$) are the (random) gross returns for asset i between $t-1$ and t. The scenario generation and statistical properties of returns are discussed below.

The budget constraints are

$$\sum_{i=1}^{N} P_{i0}(1 + tcp_i) = \sum_{i=1}^{N} S_{i0}(1 + tcs_i) + C_0, \quad t = 0,$$

and

$$\sum_{i=1}^{N} \widetilde{P}_{it}(1 + tcp_i) = \sum_{i=1}^{N} \widetilde{S}_{it}(1 + tcs_i) + C_t, \quad t = 1, \dots, T-1,$$

where tcp_i and tcs_i denote asset-specific linear transaction-costs for purchases and C_t sales, and is the fixed (non-random) net cashflow (inflow if positive).

Portfolio weights can be constrained over linear combinations (subsets) of assets or individual assets via

$$\sum_{i \in U} \widetilde{W}_{it} - \theta_U \sum_{i=1}^{N} \widetilde{W}_{it} \leqslant 0,$$

and

$$-\sum_{i \in L} \widetilde{W}_{it} + \theta_L \sum_{i=1}^{N} \widetilde{W}_{it} \leqslant 0, \quad t = 1, \dots, T-1,$$

where θ_U is the maximum percentage and θ_L is the minimum percentage of the subsets U and L of assets $i = 1, \dots, N$ included in the restrictions. The θ_U's, θ_L's, U's and L's may be time-dependent.

Risk is measured as a weighted discounted convex function of target violation shortfalls of various types in various periods. In a typical application, the deterministic wealth $\overline{W}_t$ target is assumed to grow by 7.5% in each year. The wealth targets are modeled via

$$\sum_{i=1}^{N} \big(\widetilde{W}_{it} - \widetilde{P}_{it} + \widetilde{S}_{it}\big) + \widetilde{M}_t^W \geqslant \widetilde{W}_t, \quad t = 1, \dots, T,$$

where $\widetilde{M}_t^W$ are non-negative wealth-target shortfall variables. The shortfall is penalized using a piecewise linear convex risk measure using the variables and constraints

$$\widetilde{M}_t^W = \sum_{j=1}^{m} \widetilde{M}_{jt}^W, \qquad t = 1, \dots, T,$$

$$\widetilde{M}_{jt}^W \leqslant b_j - b_{j-1}, \quad t = 1, \dots, T; \;\; j = 1, \dots, m-1,$$

where $\widetilde{M}_{jt}^W$ is the wealth target shortfall associated with segment j of the cost-function, b_j is the jth breakpoint of the risk measure function ($b_0 = 0$), and m is the number of segments of the function. A piecewise linear approximation to the convex quadratic risk measure is used so the model remains linear. The appropriateness of the quadratic function is discussed below. Convexity guarantees that if $\widetilde{M}_{jt}^W > 0$ then $\widetilde{M}_{j-1,t}^W$ is at its maximum and if $\widetilde{M}_{jt}^W$ is not at its maximum then $\widetilde{M}_{j+1,t}^W = 0$.

Stochastic benchmark goals can also be set by the user and are similarly penalized for underachievement. The benchmark $\widetilde{B}_t$ target is scenario dependent. It is based on stochastic asset returns and fixed asset weights α_i defining the benchmark portfolio

$$\widetilde{B}_t = W_0 \sum_{j=1}^{t} \sum_{j=1}^{N} \alpha_i \widetilde{R}_{ij}.$$

The corresponding shortfall constraints are

$$\sum_{j=1}^{N}\left(\widetilde{W}_{it} - \widetilde{P}_{it} + \widetilde{S}_{it}\right) + \widetilde{M}_t^B \geqslant \widetilde{B}_t, \quad t = 1, \ldots, T,$$

where $\widetilde{M}_t^B$ is the benchmark-target shortfall. These shortfalls are also penalized with a piecewise linear convex risk measure.

If total wealth exceeds the target, a fraction $\gamma = 10\%$ of the exceeding amount is allocated to a reserve account and invested in the same way as other available funds. However, the wealth targets at future stages are adjusted. Additional non-negative decision variables $\widetilde{D}_t$ are introduced and the wealth target constraints become

$$\sum_{i=1}^{N}\left(\widetilde{W}_{it} - \widetilde{P}_{it} + \widetilde{S}_{it}\right) - \widetilde{D}_t + \widetilde{M}_t^W = \overline{W}_t + \sum_{j=1}^{t-1} \gamma \widetilde{D}_{t-j}, \quad t = 1, \ldots, T-1,$$

where $\widetilde{D}_1 = 0$.

Since pension payments are based on wealth levels, increasing these levels increases pension payments. The reserves provide security for the pension plan's increase of pension payments at each future stage.

The pension plan's objective function is to maximize the expected discounted value of terminal wealth in period T net of the expected discounted penalty costs over the horizon from the convex risk measures $c_k(\cdot)$ for the wealth- and benchmark-targets, respectively,

$$\operatorname{Max} \mathrm{E}\left[d_T \sum_{i=1}^{N} \widetilde{W}_{iT} - \lambda \sum_{t=1}^{T} d_t w_t \left(\sum_{k \in \{W,B\}} v_k c_k(\widetilde{M}_t^k)\right)\right].$$

Expectation is over T period scenarios S_T. The discount factors d_t are related to the interest rate r by $d_t = (1+r)^{-t}$. Usually r is taken to be the three or six month Treasury-bill rate. The v_k are weights for the wealth- and benchmark-shortfalls and the w_t are weights for the weighted sum of shortfalls at each stage normalized via

$$\sum_{k \in \{W,B\}} v_k = 1 \quad \text{and} \quad \sum_{i=1}^{T} w_t = T.$$

Such concave objective functions with convex risk measures date to Kusy and Ziemba (1986), were used in the Russell-Yasuda model (Cariño and Ziemba, 1998), and are justified in an axiomatic sense in Rockafellar and Ziemba (2000). Nontechnical decision makers find the increasing penalty for target violations a good approach and easy to understand.

In the InnoALM model the penalty function $c_k(M^k)$ is a quadratic function. Kallberg and Ziemba (1983) show for normally distributed asset returns that varying the average Arrow–Pratt absolute risk aversion index R_A traces out the whole spectrum of risk attitudes of all concave utility functions. The most aggressive behavior is log utility which has $R_A = 1/\text{wealth}$ which is essentially zero. Typical 60–40 stock-bond pension funds have $R_A = 4$. The Kallberg and Ziemba (1983) results indicate that for computational purposes the quadratic utility function $u(w) = w - R_A/2w^2$ will suffice and is easier to use in the optimization. The error in this approximation is close to zero and well below the accuracy of the data.

The parameter λ in the objective corresponds to $R_A/2$ which in the quadratic utility function is the weight assigned to risk measured in terms of variance. The objective function of the InnoALM model only penalizes wealth and benchmark target *shortfalls*. If the target growth is roughly equal to the average return of the portfolio shortfalls measure only negative deviations form the mean, whereas variance is based on positive and negative deviations. This implies that shortfalls only account for about half of the variance. Therefore, to obtain results in agreement with a quadratic utility function we use $\lambda = R_A$ rather than $R_A/2$, in the objective function. To obtain a solution to the allocation problem for general levels of total initial wealth w_0 we use the rescaled parameter $\lambda = R_A/w_0$ in the objective function.

Using a quadratic function, the penalty function $c_k(M^k)$ is

$$c_k(M^k) = \sum_{j=1}^{m} \widetilde{M}_{jt}^k (b_{j-1} + b_j), \quad \widetilde{M}_{jt}^k \leqslant b_j - b_{j-1}, \quad \text{with } b_0 = 0.$$

Uncertainty is modeled using multiperiod discrete probability scenarios using statistical properties of the assets' returns. A scenario tree is defined by the number of stages and the number of arcs leaving a particular node. Figure 41 shows a tree with a 2–2–3 node structure for a three-period problem with four stages and introduces some definitions and terminology. The tree always starts with a single node which corresponds to the present state ($t = 0$). Decisions are made at each node of the tree and depend on the current state which reflects previous decisions and uncertain future paths. A single scenario s_t is a trajectory that corresponds to a unique path leading from the single node at stage 1 ($t = 0$) to a single node at t. Two scenarios s'_t and s''_t are identical until $t - 1$ (i.e., $s'_{t-1} = s''_{t-1}$) and differ in subsequent periods $t, \ldots, T$. The scenario assigns specific values to all uncertain parameters along the trajectory, i.e., asset returns and benchmark targets for all periods. Given all T period scenarios S_T and their respective probabilities one has a complete description of the uncertainty of the model.

Allocations are based on optimizing the stochastic linear program with IBM's optimization solutions library using the stochastic extension library (OSLE version 3). IBM has ceased all sales of this product in 2004. While existing installations of OSLE may still be used, new implementations require alternative software such as the open source project COIN-OR (see http://www.coin-or.org). The library uses the Stochastic Mathematical Programming System (SMPS) input format for multistage stochastic programs (see King et al., 2005). The *core*-file contains information about the decisions variables, constraints, right-hand-sides and bounds. It contains all fixed coefficients and dummy entries for random elements. The *stock*-file reflects the node structure of the scenario tree and contains all random elements, i.e., asset and benchmark returns, and probabilities. Non-anticipatory constraints are imposed to guarantee that a decision made at

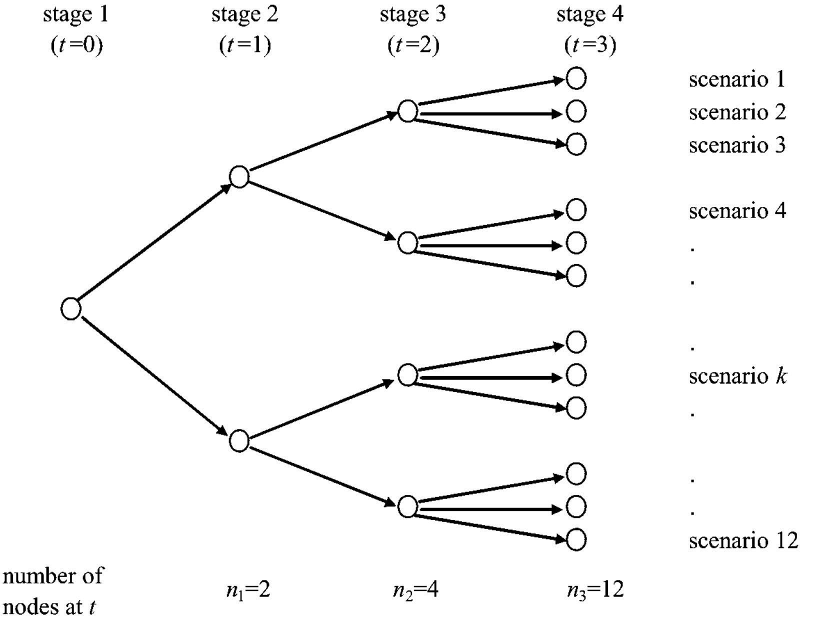


Fig. 41. Scenario tree with a 2–2–3 node structure (12 scenarios).

a specific node is identical for all scenarios leaving that node so the future cannot be anticipated. This is implemented by specifying an appropriate scenario structure in the *stock* input file. The *time*-file assigns decision variables and constraints to stages. The required statements in the input files are automatically generated by the InnoALM system.

7.4. Some typical applications

To illustrate the model's use, Geyer and Ziemba (2007) presented results for a problem with four asset classes (European stocks, US stocks, European bonds, and US bonds) with five periods (six stages). The periods are twice 1 year, twice 2 years, and 4 years (10 years in total). They assumed discrete compounding, which implies that the mean return for asset i (μ_i) used in simulations is $\mu_i = \exp(\bar{y}_i) - 1$ where $\bar{y}_i$ is the mean, based on log returns. Using a 100–5–5–2–2 node structure, they generated 10,000 scenarios. Initial wealth equals 100 units, and the wealth target is assumed to grow at an annual rate of 7.5%. To make the results more general, they do not consider a benchmark target or cash in- and outflows in this sample application. They used a risk-aversion index of $R_A = 4$, and the discount factor equals 5%, which corresponds roughly with a simple static mean-variance model to a standard 60/40 stock/bond pension fund mix (see Kallberg and Ziemba, 1983).

Table 27
Statistical properties of asset returns

Returns	European stocks		US stocks		European bonds	US bonds
	1/70–9/00	1/86–9/00	1/70–9/00	1/86–9/00	1/86–9/00	1/86–9/00
Monthly						
Mean (%)[a]	10.60	13.30	10.70	14.80	6.50	7.20
Standard deviation (%)[a]	16.10	17.40	19.0	20.200	3.70	11.3
Skewness	−0.90	−1.43	−0.72	−1.04	−0.50	0.52
Kurtosis	7.05	8.43	5.79	7.09	3.25	3.30
Jarque–Bera Test	302.60	277.30	151.90	155.6	7.70	8.50
Annual						
Mean (%)	11.10	13.30	11.0	15.20	6.50	6.90
Standard deviation (%)	17.20	16.20	20.10	18.40	4.80	12.10
Skewness	−0.53	−0.10	−0.23	−0.28	−0.20	−0.42
Kurtosis	3.23	2.28	2.56	2.45	2.25	2.26
Jarque–Bera Test	17.40	3.90	6.20	4.20	5.00	8.70

Source: Geyer and Ziemba (2007).
[a] Annualized.

Assumptions about the statistical properties of returns measured in nominal euros are based on a sample of monthly data from January 1970 for stocks and 1986 for bonds to September 2000. Summary statistics for monthly and annual log returns are shown in Table 27. The US and European equity means for the longer 1970–2000 period were much lower and slightly less volatile than those for the 1986–2000 period. The monthly stock returns were non-normal and negatively skewed. Monthly stock returns were fat tailed, whereas monthly bond returns were close to normal (the critical value of the Jarque–Bera test for $a = 0.01$ is 9.2).

For long-term planning models such as InnoALM, with its one-year review period, however, properties of monthly returns are less relevant. The bottom panel of Table 27 shows statistics for annual returns. Although average returns and volatilities remained about the same, one year of data is lost when annual returns are computed and the distributional properties changed dramatically. There was negative skewness, but no evidence existed for fat tails in annual returns, except for European stocks (1970–2000) and US bonds.

The mean returns from this sample are comparable to the 1900–2000 101-year mean returns estimated by Dimson, Marsh and Staunton (2002). Their estimate of the nominal mean equity return was 12.0% for the United States and 13.6% for Germany and the United Kingdom (the simple average of the two countries means). They estimated mean of bond returns of 5.1% for the United States and 5.4% for Germany and the United Kingdom.

Assumptions about means, standard deviations, and correlations for the applications of InnoALM appear in Table 28 and are based on the sample statistics presented in

Table 28
Mean, standard deviation, and correlation assumptions

Asset class	European stocks	US stocks	European bonds	US bonds
Normal periods (70% of the time)				
US stocks	0.755			
European bonds	0.334	0.286		
US bonds	0.514	0.780	0.333	
Standard deviation	14.6%	17.3%	3.3%	10.9%
High volatility (20% of the time)				
US stock	0.786			
European bonds	0.171	0.100		
US bonds	0.435	0.715	0.159	
Standard deviation	19.2%	21.1%	4.1%	12.4%
Extreme periods (10% of the time)				
US stocks	0.832			
European bonds	−0.075	−0.182		
US bonds	0.315	0.618	−0.104	
Standard deviation	21.7%	27.1%	4.4%	12.9%
Average period				
US stocks	0.769			
European bonds	0.261	0.202		
US bonds	0.478	0.751	0.255	
Standard deviation	16.4%	19.3%	3.6%	11.4%
All periods				
Mean	10.6%	10.7%	6.5%	7.2%

Source: Geyer and Ziemba (2007).

Table 29. Projecting future rates of returns from past data is difficult. The equity means from the 1970–2000 period are used because the 1986–2000 period had an exceptionally high performance of stocks that is not assumed to prevail in the long run.

The correlation matrices in Table 28 for the three different regimes are based on the regression approach of Solnik, Boucrelle and Le Fur (1996). Moving average estimates of correlations among all assets are functions of standard deviations of US equity returns. The estimated regression equations are then used to predict the correlations in the three regimes shown in Table 28. Results for the estimated regression equations appear in Table 29. Three regimes are considered, and the assumption is that 10% of the time, equity markets are extremely volatile; 20% of the time, markets are characterized by high volatility; and 70% of the time, markets are normal. The 35% quantile of US equity return volatility defines "normal" periods. "Highly volatile" periods are based on the 80% volatility quantile, and "extreme" periods, on the 95% quartile. The associated correlations reflect the return relationships that typically prevailed during those market conditions. The correlations in Table 28 show a distinct pattern across the three regimes.

Table 29
Regression equations relating asset correlations and US stock return volatility

Correlation	Constant	Slope with respect to US stock volatility	t-Statistic of slope	R^2
European stocks–US stocks	0.62	2.7	6.5	0.23
European stocks–European bonds	1.05	−14.4	−16.9	0.67
European stocks–US bonds	0.86	−7.0	−9.7	0.40
US stocks–European bonds	1.11	−16.5	−25.2	0.82
US stocks–US bonds	1.07	−5.7	−11.2	0.48
European bonds-US bonds	1.10	−15.4	−12.8	0.54

Source: Geyer and Ziemba (2007).

Correlations among stocks tend to increase as stock return volatility rises, whereas the correlations between stocks and bonds tend to decrease. European bonds may serve as a hedge for equities during extremely volatile periods because bond and stock returns, which are usually positively correlated, are then negatively correlated. The latter is a major reason using scenario-dependent correlation matrices is a major advance over the sensitivity of one correlation matrix.

Optimal portfolios were calculated for seven cases—with and without mixing of correlations and with normal, t-, and historical distributions. The “mixing” cases NM, TM, HM use mixing correlations. Case NM assumes normal distributions for all assets. Case HM uses the historical distributions of each asset. Case TM assumes t-distributions with five degrees of freedom for stock returns, whereas bond returns are assumed to have normal distributions. The “average” cases NA, HA, and TA use the same distribution assumptions, with no mixing of correlation matrices. Instead, the correlations and standard deviations used in these cases correspond to an “average” period in which 10, 20, and 70% weights are used to compute the averages of correlations and standard deviations in the three different regimes. Comparisons of the average (A) cases and mixing (M) cases are mainly intended to investigate the effect of mixing correlations. TMC maintains all assumptions of case TM but uses Austria’s constraints on asset weights (see Table 26). Eurobonds must be at least 40% and equity at most 40%, and these constraints are binding.

7.5. *Some test results*

Table 30 shows the optimal initial asset weights at Stage 1 for the various cases. Table 31 shows results for the final stage (expected weights, expected terminal wealth, expected reserves, and shortfall probabilities). These tables exhibit a distinct pattern: The mixing-correlation cases initially assign a much lower weight to European bonds than the average-period cases. Single-period, mean-variance optimization, and average-period cases (NA, HA, and TA) suggest an approximate 45/55% stock/bond mix. The

Table 30
Optimal initial asset weights at stage 1 by case

Case	European stocks	US stocks	European bonds	US bonds
Single-period, mean-variance optimal weights (average periods)	34.8%	9.6%	55.6%	0.0%
NA: No mixing (average periods) normal distributions	27.2	10.5	62.3	0.0
HA: No mixing (average periods) historical distributions	40.0	4.1	55.9	0.0
TA: No mixing (average periods) t-distributions for stocks	44.2	1.1	54.7	0.0
NM: Mixing correlations normal distributions	47.0	27.6	25.4	0.0
HM: Mixing correlations historical distributions	37.9	25.2	36.8	0.0
TM: Mixing correlations t-distributions for stocks	53.4	11.1	35.5	0.0
TMC: Mixing correlations historical distributions; constraints on asset weights	35.1	4.9	60.0	0.0

Source: Geyer and Ziemba (2007).

Table 31
Final stage results

Case	European stocks	US stocks	European bonds	US bonds	Expected terminal wealth	Expected reserves at Stage 6	Probability of target shortfall
NA	34.3%	49.6%	11.7%	4.4%	328.9	202.8	11.2%
HA	33.5	48.1	13.6	4.8	328.9	205.2	13.7
TA	35.5	50.2	11.4	2.9	327.9	202.2	10.9
NM	38.0	49.7	8.3	4.0	349.8	240.1	9.3
HM	39.3	46.9	10.1	3.7	349.1	335.2	10.0
TM	38.1	51.5	7.4	2.9	342.8	226.6	8.3
TMC	20.4	20.8	46.3	12.4	253.1	86.9	16.1

Source: Geyer and Ziemba (2007).

mixing-correlation cases (NM, HM, and TM) imply a 65/35% stock/bond mix. Investing in US bonds is not optimal at Stage 1 in any of the cases, an apparent result of the relatively high volatility of US bonds.

Table 31 shows that the distinction between A and M cases becomes less pronounced over time. European equities, however, still have a consistently higher weight in the mixing cases than in the no-mixing cases. This higher weight is mainly at the expense of Eurobonds. In general, the proportion of equities at the final stage is much higher than in the first stage. This result may be explained by the fact that the expected portfolio wealth at later stages is far above the target wealth level (206.1 at Stage 6), and the higher risk associated with stocks is less important. The constraints in case TMC lead to lower expected portfolio wealth throughout the horizon and to a higher shortfall probability than in any other case. The calculations show that initial wealth would have to be 35% higher to compensate for the loss in terminal expected wealth stemming from those constraints. In all cases, the optimal weight of equities is much higher than the historical 4.1% in Austria.

The terminal wealth and the shortfall probabilities at the final stage shown in Table 31 make the difference between mixing and no-mixing cases even clearer. The mixing-correlation cases yield higher levels of terminal wealth and lower shortfall probabilities.

If the level of portfolio wealth exceeds the target, the surplus, $\widetilde{D}_j$, is allocated to a reserve account. The reserves in t are computed from $\sum_{j=1}^{t} \widetilde{D}_j$ and are shown in Table 10 for the final stage. These values are in monetary units given an initial wealth level of 100. They can be compared with the wealth target 206.1 at Stage 6. Expected reserves exceed the target level at the final stage by up to 16%. Depending on the scenario, the reserves can be as high as 1,800. Their standard deviation (across scenarios) ranges from 5 at the first stage to 200 at the final stage. The constraints in case TMC lead to a much lower level of reserves compared with the other cases, which implies, in fact, less security against future increases of pension payments.

Optimal allocations, expected wealth, and shortfall probabilities are mainly affected by considering mixing correlations but the type of distribution chosen has a smaller impact. This distinction is primarily the result of the higher proportion allocated to equities, if different market conditions are taken into account by mixing correlations.

The results of any asset-allocation strategy depend crucially on the mean returns. Geyer and Ziemba investigated the effect by parameterizing the forecasted future means of equity returns. Assume that an econometric model forecasts that the future mean return for US equities is some value between 5–15%. The mean of European equities is adjusted accordingly so that the ratio of equity means and the mean bond returns shown in Table 10 are maintained. Geyer and Ziemba retain all other assumptions of case NM (normal distribution and mixing correlations). Figure 25 summarizes the effects of these mean changes in terms of the optimal initial weights. As expected, the results are sensitive to the choice of the mean return; see Chopra and Ziemba (1993) and Kallberg and Ziemba (1981, 1983). If the mean return for US stocks is assumed to equal the long-run mean of 12%, as estimated by Dimson, Marsh and Staunton (2002, 2006), the model yields an optimal weight for equities of 100%. A mean return for US stocks of 9%, however, implies an optimal weight of less than 30% for equities.

7.6. Model tests

Because scenario-dependent correlations have a significant impact on asset-allocation decisions, it is worthwhile to further investigate their nature and implications from the perspective of testing the model. Positive effects on the pension fund performance induced by the stochastic, multiperiod planning approach will be realized only if the

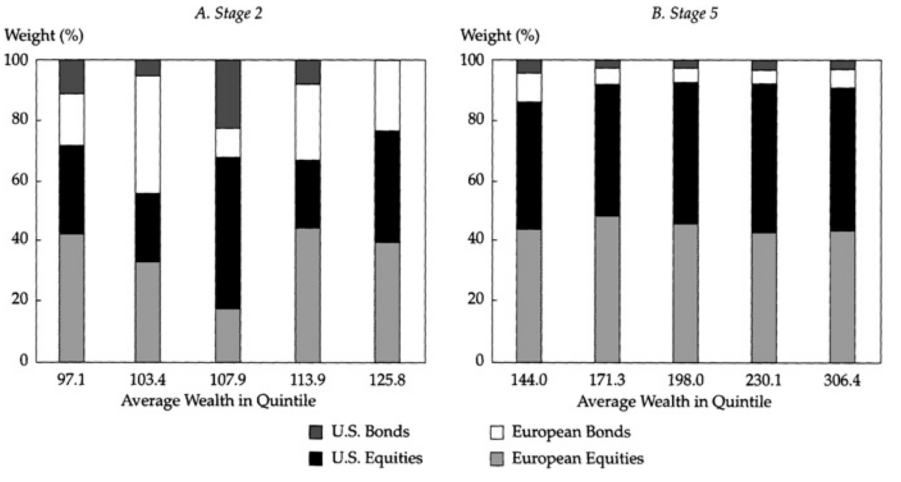


Fig. 42. Optimal weights conditional on quintiles of portfolio wealth at Stages 2 and 5. Source: Geyer and Ziemba (2007).

portfolio is dynamically rebalanced, as implied by the optimal scenario tree. Geyer and Ziemba tested the performance of the model considering this aspect. As a starting point, they broke down the rebalancing decisions at later stages into groups of achieved wealth levels. This process reveals the "decision rule" implied by the model, depending on the current state. Consider case TM. They formed quintiles of wealth at Stage 2, computed the average optimal weights assigned to each quintile, and did the same using quintiles of wealth at Stage 5.

Panels A and B of Figure 42 depict the distribution of weights for each of the five average levels of wealth at the two stages. Although the average allocation at Stage 5 is essentially independent of the wealth level achieved (the target wealth at Stage 5 is 154.3), the distribution at Stage 2 depends on the wealth level in a specific way. If average attained wealth is 103.4, which is slightly below the target, the model chooses a cautious strategy. Bonds have the highest weight in this case (almost 50%). In this situation, the model implies that the risk of even stronger underachievement of the target is to be minimized. The model relies on the low, but more certain, expected return of bonds to move back to the target level. If attained wealth is far below the target (97.1), the model implies more than 70% equities and a high share (10.9%) of relatively risky US bonds. With such strong underachievement, a cautious strategy has no room to attain the target level again. If the average attained wealth equals 107.9, which is close to the target wealth of 107.5, the highest proportion would be invested in US assets, with 49.6% in equities and 22.8% in bonds. The US assets are more risky than the corresponding European assets, which is acceptable because the portfolio wealth is close to the target and risk does not play a big role. For wealth levels above the target, most of the portfolio are switched to European assets, which are safer than US assets.

Table 32
Results of asset-allocation strategies using the decision rule implied by the optimal scenario tree

Case	Complete sample 1/92–1/02		Out of sample 10/00–1/02	
	Mean	Standard deviation	Mean	Standard deviation
NA	11.6%	16.1%	−17.1%	18.6%
NM	13.1	15.5	−9.6	16.9
HA	12.6	16.5	−15.7	21.1
HM	11.8	16.5	−15.8	19.3
TA	10.0	16.0	−14.6	18.9
TM	14.9	15.9	−10.8	17.6
TMC	12.4	8.5	0.6	9.9

Source: Geyer and Ziemba (2007).

This “decision” may be interpreted as an attempt to preserve the high levels of attained wealth.

The decision rules implied by the optimal solution can be used to perform a test of the model using the following rebalancing strategy. Consider the 10-year period from January 1992 to January 2002. In the first month of this period, Geyer and Ziemba assumed that wealth is allocated according to the optimal solution for Stage 1, given in Table 30. In each of the subsequent months, they rebalance the portfolio as follows. First, they identify the current volatility regime (extreme, highly volatile, or normal) based on the observed US stock return volatility. Then, they search the scenario tree to find a node that corresponds to the current volatility regime and has the same or a similar level of wealth. The optimal weights from that node determine the rebalancing decision. For the no-mixing cases NA, TA, and HA, the information about the current volatility regime cannot be used to identify optimal weights. In those cases, they use the weights from a node with a level of wealth as close as possible to the current level of wealth.

Table 32 presents summary statistics for the complete-sample and out-of-sample periods. The mixing-correlation solutions, assuming normal and t-distributions (cases NM and TM), provided a higher average return with lower standard deviation than the corresponding no-mixing cases (NA and TA). The advantage may be substantial, as indicated by the 14.9% average return of TM compared with 10.0% for TA. The t-statistic for this difference was 1.7 and was significant at the 5% level (one-sided test). Using the historical-distribution and mixing-correlation case (HM) yielded a lower average return than the no-mixing case (HA). In the constrained case (TMC), the average return for the complete sample was in the same range as for the unconstrained cases. This result stems primarily from the relatively high weights assigned to US bonds; US bonds performed

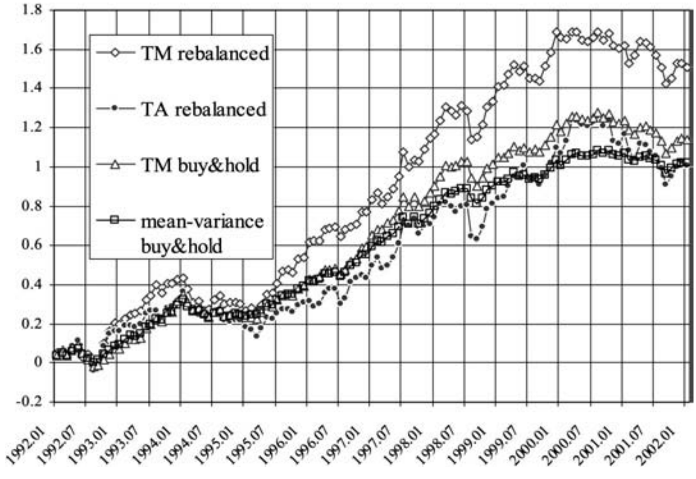


Fig. 43. Cumulative monthly returns for different strategies, 1992–2002. Source: Geyer and Ziemba (2007).

well during the test period, whereas stocks performed poorly. The standard deviation of returns was much lower because the constraints imply a lower degree of rebalancing.

To emphasize the difference between the cases TM and TA, Figure 43 compares the cumulated monthly returns obtained from the rebalancing strategy for the two cases with a buy-and-hold strategy that assumes that the portfolio weights on January 1992 were fixed at the optimal TM weights throughout the test period. In comparison to the buy-and-hold strategy or the performance using TA results, for which rebalancing does not account for different correlation and volatility regimes, rebalancing on the basis of the optimal TM scenario tree provided a substantial gain.

Such in- and out-of-sample comparisons depend on the asset returns and test period. To isolate the potential benefits from considering scenario-dependent correlations, Geyer and Ziemba performed the following controlled simulation experiment. Consider 1,000 10-year periods in which simulated annual returns of the four assets are assumed to have the statistical properties summarized in Table 27. One of the 10 years is assumed to be an "extreme" year, two years correspond to "highly volatile" markets, and seven years are "normal" years. They compare the average annual return of two strategies: (a) a buy-and-hold strategy using the optimal TM weights from Table 30 throughout the 10-year period, and (b) a rebalancing strategy that uses the implied decision rules of the optimal scenario tree as explained in the in- and out-of-sample tests above. For simplicity, they assume that the current volatility regime is known in each period. The average annual returns for 1,000 repetitions of the two strategies are 9.8% (rebalancing) and 9.2% (buy and hold). The t-statistic for the mean difference is 5.4, indicating the highly significant advantage of the rebalancing strategy, which exploits the information about state-dependent correlations.

For comparison, Geyer and Ziemba use the optimal weights from the constrained case TMC and repeat the same experiment. They obtain the same average mean of 8.1% for both strategies. This result indicates that the constraints imply insufficient rebalancing capacity. Therefore, knowledge about the volatility regime cannot be sufficiently exploited to achieve superior performance relative to a buy-and-hold strategy. This result also shows that the relatively good performance of the TMC rebalancing strategy in the sample 1992–2002 period was positively biased by the favorable conditions during that time.

The model, once developed in 2000, proved to be very useful for Innovest. In 2006, Konrad Kontriner (Member of the Board) and Wolfgang Herold (Senior Risk Strategist) of Innovest stated that:

> The InnoALM model has been in use by Innovest, an Austrian Siemens subsidiary, since its first draft versions in 2000. Meanwhile it has become the only consistently implemented and fully integrated proprietary tool for assessing pension allocation issues within Siemens AG worldwide. Apart from this, consulting projects for various European corporations and pensions funds outside of Siemens have been performed on the basis of the concepts of InnoALM.
>
> The key elements that make InnoALM superior to other consulting models are the flexibility to adopt individual constraints and target functions in combination with the broad and deep array of results, which allows to investigate individual, path dependent behavior of assets and liabilities as well as scenario based and Monte-Carlo like risk assessment of both sides.
>
> In light of recent changes in Austrian pension regulation the latter even gained additional importance, as the rather rigid asset based limits were relaxed for institutions that could prove sufficient risk management expertise for both assets and liabilities of the plan. Thus, the implementation of a scenario based asset allocation model will lead to more flexible allocation restraints that will allow for more risk tolerance and will ultimately result in better long term investment performance.
>
> Furthermore, some results of the model have been used by the Austrian regulatory authorities to assess the potential risk stemming from less constrained pension plans.

8. Conclusions

Some key points to remember about the stochastic programming approach to asset, liability, and wealth management are:

- *Point 1*. Means are by far the most important part of the distribution of returns, especially the direction. Thus, you must estimate future means well or you can quickly travel in the wrong direction, which usually leads to losses or underperformance, or complete disaster if one is over levered.
- *Point 2*. Mean-variance models are useful as a basic guideline when you are in an assets-only situation. Professionals adjust means using mean-reversion, James–Stein, or truncated estimators and constrain output weights. Do not change asset positions unless the advantage of the change is significant. Do not use mean-variance analysis with liabilities and other major market imperfections, except as a first test analysis.

- *Point 3*. Trouble arises when you overbet and a bad scenario occurs. Thus, do not overbet when there is any possibility of a bad scenario occurring, unless the bet is protected by some type of hedge or stop loss.
- *Point 4*. Trouble is exacerbated when the expected diversification does not hold in the scenario that occurs. Thus, you must use scenario-dependent correlation matrices because simulations around historical correlation matrices are inadequate for extreme scenarios.
- *Point 5*. When a large decline in the stock market occurs, the positive correlation between stocks and bonds fails and they become negatively correlated. Thus, when the mean of the stock market is negative, bonds are more attractive, as is cash.
- *Point 6*. Stochastic programming scenario-based models are useful when you want to look at aggregate overall decisions, with liabilities, liquidity, taxes, policy, legal, and other constraints, and have targets and goals you want to achieve. It thus pays to make a complex stochastic programming model when a lot is at stake and the essential problem has many complications.
- *Point 7*. Other approaches, such as continuous-time finance, decision-rule-based stochastic programming, control theory, and so on, are useful for problem insights and theoretical results. But in actual use, they may lead to disaster unless modified. Black–Scholes option pricing theory says you can hedge perfectly with lognormal assets, which can lead to overbetting. Fat tails and jumps arise frequently and can occur without warning. The S&P 500 opened limit down 60 points or 6% when trading resumed after 9/11, and it fell 14% that week. Thus, be careful of the assumptions, including implicit ones, of theoretical models. Use the results with caution no matter how complex and elegant the math or how smart or famous the author. Remember, you have to be very smart to lose millions and even smarter to lose billions.
- *Point 8*. Do not be concerned with getting all the scenarios exactly right when using stochastic programming models. You cannot do so, and it does not matter that much anyway. Instead, worry about having the problem periods laid out reasonably and make sure the scenarios basically cover the means, the tails, and the chance of what could happen. If the current situation has never occurred before, use one that is similar to add scenarios. For a crisis in Brazil, use Russian crisis data, for example. The results of stochastic programming will give you good advice when times are normal and keep you out of severe trouble when times are bad. Those using stochastic programming models may lose 5, 10, or 15%, but they will not lose 50, 70, or 95%, as some investors and hedge funds have. Thus, if the scenarios are more or less accurate and the problem elements are reasonably modeled, stochastic programming will give good advice. You may slightly outperform in normal markets, but you will greatly outperform in bad markets when other approaches may blow up.
- *Point 9*. Stochastic programming models for asset/liability management were very expensive in the 1980s and early 1990s but are not expensive now. Years ago, Vancouver analysts using a large linear programming model to plan lumber operations at MacMillan Blodel used to fly to San Francisco to use a large computer that would run all day to run the model once. Now, models of this complexity take only seconds

to run on inexpensive personal computers. Thus, advances in computing power and modeling expertise have made stochastic programming modeling much less expensive. Such models, which are still complex and require approximately six months to develop and test, cost a couple hundred thousand dollars. A small team can make a model for a complex organization quite quickly at fairly low cost compared with what is at stake.

- *Point 10*. Eventually, as more disasters occur and more successful stochastic programming models are built and used, they will become popular. Thus, the ultimate goal is to have them in regulations, such as value at risk. Although VAR does more good than harm, its safety is questionable in many applications. Conditional value at risk is an improvement, but for most people and organizations, the non-attainment of goals is more than proportional (i.e., convex) in the non-attainment.

References

Abaffy, J., Bertocchi, M., Dupačovà, J., Moriggia, V., 2000. On generating scenarios for bond portfolios. Bulletin of the Czech Economic Society, 3–27, February.

Ali, M.M., 1979. Some evidence on the efficiency of a speculative market. Econometrica 47, 637–642.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Barberis, N.M., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.

Berge, K., Consigli, G., Ziemba, W.T., 2007. The predictive ability of the bond stock earnings yield differential. Journal of Portfolio Management, in press.

Berge, K., Ziemba, W.T., 2006. The predictive ability of bond stock earnings yield differences model, 1970–2005, Working paper. Sauder School of Business, UBC.

Birge, J.R., Louveaux, F., 1997. Stochastic Programming. Springer-Verlag.

Birge, J.R., Wets, R.J.-B., 1986. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. Stochastic programming 84. I. Math. Programming Stud. 27, 54–102.

Black, F., Derman, E., Toy, W., 1990. A one-factor model of interest rates and its application to Tresury bond options. Financial Analysts Journal 1, 33–39.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Blake, D., Lehmann, B.N., Timmermann, A., 1999. Asset allocation dynamics and pension fund performance. Journal of Business 72, 429–461.

Board, J., Sutcliffe, C., 2007. Joined-Up Pensions Policy: An Asset–Liability Model for Simultaneously Setting the Asset Allocation and Contribution Rate. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 2. Elsevier, pp. 1029–1067 (Ch. 21).

Bodie, Z., 2001. Retirement investing: A new approach, Working Paper. Boston University.

Boender, G.C.E., 1997. A hybrid simulation/optimisation scenario model for asset liability management. European Journal of Operational Research 99, 126–135.

Bos, E., Vu, M.T., Massiah, E., Bulatao, R., 1994. World population projections, 1994–95 edition, Technical report. The World Bank, Washington, DC.

Bradley, S.P., Crane, D.B., 1972. A dynamic model for bond portfolio management. Management Science 19, 139–151.

Bradley, S.P., Crane, D.B., 1973. Management of commercial bank government security portfolios: An optimization approach under uncertainty. Journal of Bank Research 4, 18–30.

Brennan, M.J., Schwartz, E.S., 1998. The use of Treasury bill futures in strategies asset allocation program. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modeling. Cambridge University Press, pp. 205–228.

Brinson, B., Hood, L.R., Beebower, G.L., 1986. Determinants of portfolio performance. Financial Analysts Journal 42, 39–45.

Brinson, G.P., Singer, B.D., Beebower, G.L., 1991. Determinants of portfolio performance II: An update. Financial Analysts Journal 47 (3), 40–48, May/June.

Buffett, W.B., 2001. Warren Buffett on the stock market. Fortune, December 10.

Burless, G., 2002. What do we know about the risk of individual account pensions? Evidence from industrial countries, Working paper. Brookings Institution.

Busche, K., 1994. Efficient market results in an Asian setting. In: Hausch, D.B., Lo, V., Ziemba, W.T. (Eds.), Efficiency of Racetrack Betting Markets. Academic Press, San Diego, pp. 615–616.

Busche, K., Hall, C.D., 1988. An exception to the risk preference anomaly. Journal of Business 61 (3), 337–346, July.

Campbell, J.Y., Shiller, R.J., 1998. Valuation ratios and the long-run stock market outlook. Journal of Portfolio Management 24, 11–26.

Campbell, J.Y., Shiller, R.J., 2001. Valuation ratios and the long-run stock market outlook: an update, Working paper W8221, NBER.

Campbell, J.Y., Viceira, L.M., 2002. Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. Oxford University Press.

Cariño, D., Myers, R., Ziemba, W.T., 1998. Concepts, technical issues and uses of the Russell-Yasuda Kasai financial planning model. Operations Research 46, 450–462.

Cariño, D., Turner, A.L., 1998. Multiperiod asset allocation with derivative assets. In: Ziemba, W., Mulvey, J. (Eds.), World Wide Asset and Liability Modeling. Cambridge University Press, pp. 182–204.

Cariño, D., Ziemba, W.T., 1998. Formulation of Russell-Yasuda Kasai financial planning model. Operations Research 46, 433–449.

Cariño, D.R., Kent, T., Myers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., Ziemba, W.T., 1994. The Russell-Yasuda Kasai model: An asset/liability model for a Japanese insurance company using multistage stochastic programming. Interfaces 24, 29–49.

Cariño, D., et al., 1997. MTB Pension asset/liability management model. Presentation, The Frank Russell Company.

Casey, M.S., Sen, S., 2005. The scenario generation algorithm for multistage stochastic linear programming. Mathematics of Operations Research 30 (3), 615–631.

Chan, K.C., Karolyi, G.A., Longstaff, F., Sanders, A.B., 1992. An empirical comparison of alternative models of the short term interest rate. Journal of Finance 47, 1209–1227.

Chopra, V.K., 1993. Improving optimization. Journal of Investing, 51–59.

Chopra, V.K., Ziemba, W.T., 1993. The effect of errors in mean, variance and co-variance estimates on optimal portfolio choice. Journal of Portfolio Management 19, 6–11.

Clifford, S.W., Kroner, K.F., Siegel, L.B., 2001. In pursuit of performance: the greatest return stories ever told. Investment insights. Barclays Global Investor 4 (1), 1–25.

Dempster, M.A.H., Evstigneev, I.V., Schenk-Hoppé, K.R., 2003. Exponential growth of fixed mix assets in stationary markets. Finance and Stochastics 7 (2), 263–276.

Dimson, E., Marsh, P., Staunton, M., 2002. Triumph of the Optimists. Princeton University Press.

Dimson, E., Marsh, P.R., Staunton, M., 2006. Global Investment Returns Yearbook. ABN Ambro, London School of Business School.

Dimson, E., Marsh, P.R., Staunton, M., 2007. The worldwide equity premium: A smaller puzzle. In: Mehra, R. (Ed.), Handbook of the Equity Risk Premium, Elsevier/North-Holland.

Duffie, D., Pan, J., 1997. An overview of value at risk. Journal of Derivatives 4, 7–49.

Duffie, D., Singleton, K., 2003. Credit Risk Modeling. Princeton University Press.

Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: Empirical tests. Journal of Finance 53 (6), 2059–2106.

Dunbar, N., 2000. Inventing Money: The Story of Long-Term Capital Management and the Legends Behind It. John Wiley & Sons, New York.

Dupačovà, J., Consigli, G., Wallace, S.W., 2000. Scenarios for multistage stochastic programs. Annals of Operations Research 100, 25–53.

Dupačovà, I., Gröwe-Kuska, N., Römisch, W., 2003. Scenario reduction in stochastic programming. An approach using probability metrics. Mathematical Programming, Ser. A 95, 493–511.

Edirisinghe, N.C.P., 1999. Bound-based approximations in multistage stochastic programming: nonanticipativity aggregation. Annals of Operations Research 85, 103–127.

Edirishinghe, N.C.P., Ziemba, W.T., 1992. Bounds for stochastic programmes with recourse. Operations Research 40, 660–677.

Edirishinghe, N.C.P., Ziemba, W.T., 1994a. Bounds for two-stage stochastic programs with fixed recourse. Mathematics of Operations Research 19, 292–313.

Edirisinghe, N.C.P., Ziemba, W.T., 1994b. Bounding the expectation of a saddle function, with application to stochastic programming. Mathematics of Operations Research 19, 314–340.

Edwards, F.R., 1999. Hedge funds and the collapse of long term capital management. Journal of Economic Perspectives, 189–210, Spring.

Embrechts, P., 2000. Actuarial versus financial pricing of insurance. Journal of Risk Finance 1 (4), 17–26.

Embrechts, P., Resnick, S., Samovodmitsky, G., 1998. Living on the edge. Risk 11 (1), 96–100.

European Commission, 1997. Towards a Single Market for Supplementary Pensions Commission Communication, COM.

European Commission, 1999. Towards a Single Market for Supplementary Pensions Commission Communication, COM.

Fan, Y.-A., et al., 1998. Daido asset/liability management model. Presented at VIII International Conference on Stochastic Programming, August 14.

Figgis, E.L., 1974. Rates of return from flat race betting in England in 1973. Sporting Life 11, March.

Figlewiski, S., 1989. What does an option pricing model tell us about option prices. Financial Analysts Journal 45 (5), 12–15.

Figlewiski, S., 1994. How to lose money in derivatives. Journal of Derivatives, 75–82, Winter.

Fleten, S.-E., Høyland, K., Wallace, S., 2002. The performance of stochastic dynamic and fixed mix portfolio models. European Journal of Operational Research 140 (1), 37–49.

Frauendorfer, K., 1996. Barycentric scenario trees in convex multistage stochastic programming. Approximation and computation in stochastic programming. Mathematical Programming, Ser. B 75 (2), 277–293.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2002. The Innovest Austrian pension fund financial planning model InnoALM, Working paper. UBC.

Geyer, A., Ziemba, W.T., 2007. The Innovest Austrian pension fund financial planning model InnoALM. Operations Research, in press.

Goldstein, A.B., Markowitz, B.G., 1982. Sofasim: A dynamic insurance model with investment structure, policy benefits and taxes. Journal of Finance 37, 595–604.

Gonzio, J., Kouwenberg, R., 2001. High performance computing for asset liability management. Operations Research 49, 879–891.

Grauer, R.R., Hakansson, N.H., 1998. On naive approaches to timing the market: The empirical probability assessment approach with an inflation adapter. In: Ziemba, W.T., Mulvey, J.M. (Eds.), World Wide Asset and Liability Modelling. Cambridge University Press, pp. 149–181.

Greenspan, A., 1998. Speech before the Committee on Banking and Social Services, US House of Representatives, 24 July.

Griffin, R.M., 1947. Odds adjustments by American horse-race bettors. American Journal of Psychology 62, 290–294.

Hanoch, G., Levy, H., 1969. The efficiency analysis of choices involving risk. Review of Economic Studies 36, 335–346.

Hausch, D.B., Lo, V., Ziemba, W.T. (Eds.), 1994. Efficiency of Racetrack Betting Markets. Academic Press, New York.

Hausch, D.B., Ziemba, W.T. (Eds.), 2007. Handbook of Investment: Volume Sports and Lottery Investment Markers. North-Holland, Amsterdam. In press.

Heath, D., Jarrow, R., Morton, D., 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. Econometrica 60 (1), 77–105.

Heitsch, H., Römisch, W., 2003. Scenario reduction algorithms in stochastic programming. Stochastic programming. Computational Optimization and Applications 24 (2–3), 187–206.

Henriques, D.B., 1991. A better way to back your assets. New York Times, p. 11. Sunday, March 31, Section 3 (Business).

Hensel, C.R., Ezra, D.D., Ilkiw, J.H., 1991. The importance of the asset allocation decision. Financial Analysts Journal, 65–72, July/August.

Hensel, C.R., Ziemba, W.T., 2000. How did Clinton stand up to history? US stock market returns and presidential party affiliations. In: Keim, D.B., Ziemba, W.T. (Eds.), Security Market Imperfections in World Wide Equity Markets. Cambridge University Press, pp. 203–217.

Heyde, C.C., 1963. Some remarks on the moment problem II. Quart. J. Math. Oxford Ser. (2) 14, 97–105.

Hochreiter, R., Pflug, G.C., 2003. Scenario generation for stochastic multi-stage decision processes as facility location problems, Technical report. Department of Statistics and Decision Support Systems, University of Vienna.

Høyland, K., Wallace, S.W., 2001. Generating scenario trees for multistage decision problems. Management Science 47, 295–307.

Høyland, K., Kaut, M., Wallace, S.W., 2003. A heuristic for moment-matching scenario generation. Computational Optimization and Applications 24 (2–3), 169–185.

Ibbotson Associates, 1999. Stocks, Bonds, Bills and Inflation, 1999 Year Book. Ibbotson Associates, Chicago.

Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51 (5), 1611–1631.

Jamshidian, F., Zhu, Y., 1977. Scenario simulation: theory and methodology. Finance and Stochastics 1, 43–76.

Jarrow, R.A., van Deventer, D.R., 1998. Integrating interest rate risk and credit risk in ALM. In: Kamakura Corporation (Ed.), Asset & Liability Management, A Synthesis of New Methodologies. Risk Books.

Jobst, N.J., Zenios, S.A., 2003. Tracking bond indices in an integrated market and credit risk environment. Quantitative Finance 3 (2).

Jorion, P., 2000. Risk management lessons from long-term capital management. European Financial Management 6, 277–300.

Jorion, P., 2006. Value-at-Risk: The New Benchmark for Controlling Market Risk, third ed. Irwin, Chicago.

Jorion, P., Goetzmann, W., 1999. Global stock markets in the twentieth century. Journal of Finance.

Kahneman, D., Tversky, A., 1979. Choices, values and frames. Econometrica 47, 263–291.

Kallberg, J.G., Ziemba, W.T., 1981. Remarks on optimal portfolio selection. In: Bamberg, G., Opitz, O. (Eds.), Methods of Operations Research, vol. 44. Oelgeschlager, Gunn and Hain, pp. 507–520.

Kallberg, J.G., Ziemba, W.T., 1983. Comparison of alternative utility functions in portfolio selection problems. Management Science 29 (11), 1257–1276.

Kallberg, J.G., Ziemba, W.T., 1984. Mis-specifications in portfolio selection problems. In: Bamberg, G., Spremann, K. (Eds.), Risk and Capital. Springer-Verlag, New York, pp. 74–87.

Kallberg, J.G., White, R., Ziemba, W.T., 1982. Short term financial planning under uncertainty. Management Science XXVIII, 670–682.

Keefer, D.L., 1994. Certainty equivalents for three-point discrete-distribution approximations. Management Science 40, 760–773.

Keefer, D.L., Bodily, S.E., 1983. Three-point approximations for continuous random variables. Management Science 29, 595–609.

Keim, D., Ziemba, W.T. (Eds.), 2000. Security Market Imperfections in Worldwide Equity Markets. Cambridge University Press.

Keynes, J.M., 1938. Investment Policy Report on the Chest Fund. Kings College, Cambridge, UK.

King, A.J., Wright, S.E., Parija, R., Entriken, R., 2005. The IBM stochastic programming system. In: Wallace, S.W., Ziemba, W.T. (Eds.), Applications of Stochastic Programming, MPS–SIAM Series on Optimization. Cambridge Univ. Press, pp. 21–36.

Kingsland, L., 1982. Projecting the financial condition of a pension plan using plan simulation analysis. Journal of Finance 37 (2), 577–584.

Koivu, M., Pennanen, T., Ziemba, W.T., 2005. Cointegration analysis of the FED model. Finance Research Letters 2, 248–259.

Kouwenberg, R., 1999. Scenario generation and stochastic programming models for asset liability management. European Journal of Operational Research 134, 51–64.

Kouwenberg, R., 2003. Do hedge funds add value to a passive portfolio: Correcting for non-normal returns and disappearing funds. Journal of Asset Management 3, 361–382.

Kouwenberg, R., Zenios, S.A., 2007. Stochastic programming models for asset liability management. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 1: Theory and Methodology. In: Handbooks in Finance. North-Holland, pp. 253–303.

Krokhma, P., Uryasev, S., Zrazhevsky, G., 2005. Numerical comparison of CVAR and CDAR approaches: application to hedge funds. In: Wallace, S.W., Ziemba, W.T. (Eds.), Applications of Stochastic Programming. In: SIAM–Mathematical Programming Series on Optimization, pp. 609–631.

Kusy, M.I., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34 (3), 356–376.

Lakonishok, J., Schleifer, A., Vishny, R., 1994. Contrarian investment extrapolation and risk. Journal of Finance 49.

Lo, A., 1999. The three p's of total risk management. Financial Analysts Journal 55 (1), 13–26.

Lo, A., 2001. Risk management for hedge funds: introduction and overview. Financial Analysts Journal 57 (6), 16–33.

Lord Rothschild, 1978. Royal Commission on Gambling, vols. I and II. Presented to Parliament by Command of Her Majesty, July.

MacLean, L., Ziemba, W.T., 2006. Capital growth theory and practice. In: Zenios, S.A., Ziemba, W.T. (Eds.), Theory and Methodology, Handbook of Asset and Liability Modeling, vol. 1. North-Holland, Amsterdam, pp. 429–475.

McGlothlin, W.H., 1956. Stability of choices among uncertain alternatives. American Journal of Psychology 69, 604–619.

Markowitz, H.M., Perold, A., 1981. Portfolio analysis with factors and scenarios. Journal of Finance 36 (4), 871–877, September.

Merton, R.C., 1990. Continuous-Time Finance. Blackwell Publishers, Cambridge, MA.

Merton, R.C., 2000a. Finance and the role of financial engineering in the 21st century. Nikkei Market's Japan Investor's Watch, The Nikkei Weekly Asia Web Guide.

Merton, R.C., 2000b. Future possibilities in finance theory and finance practice. In: Geman, H., Madan, D., Pliska, S.R., Vorst, T. (Eds.), Mathematical Finance: Bachelier Congress. Springer, pp. 47–74.

Mulvey, J.M., 1996. Generating scenarios for the Towers Perrin investment system. Interfaces 26, 1–13.

Mulvey, J.M., Thorlacius, A.E., 1998. The Towers Perrin global capital market scenario generation system. In: Ziemba, W.T., Mulvey, J.M. (Eds.), World Wide Asset and Liability Management. Cambridge University Press, pp. 286–312.

Mulvey, J.M., Vladimirou, H., 1992. Stohastic network programming for financial planning problems. Management Science 38 (11), 1642–1664.

Murray, S., 1998. Stochastic programming: asset liability management in practice with special constraints and challenges. Presented at the VIII International Conference on Stochastic Programming, Vancouver.

Murray, S., Fan, Y.-A., 1998. A model for individual asset/liability management in the Italian context. Presented at the VIII International Conference on Stochastic Programming, August 14.

Pennanen, T., Koivu, M., 2002. Integration quadratures in discretization of stochastic programs. Stochastic programming e-print series.

Perold, A., 1999. Long Term Capital Management, L.P.C. Harvard Business School case.

Pflug, G.Ch., 2001. Scenario tree generation for multiperiod financial optimization by optimal discretization. Mathematical programming and finance. Mathematical Programming, Ser. B 89 (2), 251–271.

Rachev, Z. (Ed.), 2003. Handbook of Heavy Tailed Distributions in Finance. Handbooks in Finance Series. North-Holland. www.elsevier.nl/homepage/sae/hf/menu.htm.

Rockafellar, T., Ziemba, W.T., 2000. Modified risk measures and acceptance sets. Mimeo, University of Washington, July.

Rosen, D., Zenios, S.A., 2006. Enterprise-wide asset and liability management: Issues, institutions, and models. In: Zenios, S.A., Ziemba, W.T. (Eds.), Theory and Methodology, Handbook of Asset and Liability Modeling, vol. 1. North-Holland, Amsterdam, pp. 1–23.

Roseveare, D., Leibfritz, W., Fore, D., Wurzel, E., 1996. Ageing populations, pension systems and government budgets: simulation for 20 OECD countries. Economics Department working paper no. 168. OECD, Paris.

Rudolf, M., Ziemba, W.T., 2004. Intertemporal surplus management. Journal of Economic Dynamics and Control 28 (4), 975–990.

Shapiro, J.F., 1988. Stochastic programming models for dedicated portfolio selection. In: Mitra, G.B. (Ed.), Mathematical Models for Decision Support. In: ASI Series, vol. F48. Springer-Verlag, Berlin, pp. 587–611.

Shaw, J., Thorp, E.O., Ziemba, W.T., 1995. Convergence to efficiency of the Nikkei put warrant market of 1989–90. Applied Mathematical Finance 2, 243–271.

Shiller, R., 2000. Irrational Exuberance. Princeton University Press.

Siegel, J.J., 2002. Stocks for the Long Run. Wiley.

Smith, E.L., 1924. Common stocks as long term investments.

Smith, J.E., 1993. Moment methods for decision analysis. Management Science 39, 340–358.

Snyder, W.W., 1978. Horse Racing: Testing the efficient markets model. Journal of Finance 33, 1109–1118.

Solnik, B., Boucrelle, C., Le Fur, Y., 1996. International market correlation and volatility. Financial Analysts Journal 52, 17–34.

Stone, D., Ziemba, W.T., 1993. Land and stock prices in Japan. Journal of Economic Perspectives, 149–165, Summer.

Stulz, R., 1996. Rethinking risk management. Journal of Applied Corporate Finance 9, 8–24.

Swensen, D.W., 2000. Pioneering Portfolio Management: An Unconventional Approach to Institutional Investments. The Free Press.

Thaler, R., Ziemba, W.T., 1988. Anomalies: Parimutuel betting markets: Racetracks and lotteries. Journal of Economic Perspectives 2, 161–174.

Thorp, E.O., 2006. The Kelly criterion in blackjack, sports betting and the stock market. In: Zenios, S.A., Ziemba, W.T. (Eds.), Theory and Methodology, Handbook of Asset and Liability Modeling, vol. 1. North-Holland, Amsterdam, pp. 385–428.

Tompkins, R.G., Ziemba, W.T., Hodges, S.H., 2003. The favorite-longshot bias in S&P 500 futures options: the return to bets and the cost of insurance, Working Paper. Sauder School of Business, UBC.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Wallace, S.W., Ziemba, W.T. (Eds.), 2005. Applications of Stochastic Programming. Mathematical Programming Series on Optimization. SIAM.

Winklevoss, H.E., 1982. PLASM: Pension liability and asset simulation model. Journal of Finance 37 (2), 585–594.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset Liability and Wealth Management. AIMR, Charlottesville, VA.

Ziemba, W.T., 2004. Behavioral finance, racetrack betting and options and futures trading. Presented to the Mathematical Finance Seminar, Stanford, January.

Ziemba, W.T., 2005. The symmetric downside risk Sharpe ratio and the evaluation of great investors and speculators. Journal of Portfolio Management, 108–122, Fall.

Ziemba, W.T., Hausch, D.B., 1986. Betting at the Racetrack. Dr Z Investments, San Luis Obispo, CA.

Ziemba, W.T., Hausch, D.B., 1987. Dr. Z's Beat the Racetrack. William Morrow.

Ziemba, T., Mulvey, J.M. (Eds.), 1998. World Wide Asset and Liability Modelling. Cambridge University Press.

Ziemba, W.T., Schwartz, S.L., 1991. Invest Japan. Probus Publishers, Chicago, IL.

Ziemba, W.T., Schwartz, S.L., 1992. Power Japan. Probus Publishers, Chicago, IL.

Ziemba, R.E.S., Ziemba, W.T., 2007. Scenarios for Risk Management and Global Investment Strategies. Wiley.

Chapter 20

DYNAMIC ASSET AND LIABILITY MANAGEMENT FOR SWISS PENSION FUNDS

GABRIEL DONDI* and FLORIAN HERZOG*

ETH Zurich, Measurement and Control Laboratory, 8092 Zurich, Switzerland
and
swissQuant Group AG, Zurich, Switzerland
e-mails: gdondi@alumni.ethz.ch; florian.herzog@alumni.ethz.ch

LORENZ M. SCHUMANN

Schumann Investment Management, Zurich, Switzerland

HANS P. GEERING

ETH Zurich, Measurement and Control Laboratory, 8092 Zurich, Switzerland

Contents

* Corresponding authors.

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02020-5

Abstract

We present an asset and liability model for Swiss pension funds. This includes an asset dynamics model and an optimisation technique to solve the problem of allocating the funds considering the liabilities maturity structure. Our liability model is based on the current and projected future cash outflows of all members, taking into account: projection of the individuals' income, probabilities of entry and exit of members, and probabilities of death and invalidity of members. For the modelling of the various probabilities, we use a life insurance mathematics approach. This results in a dynamic, stochastic description of the pension fund liabilities. The projected uncertain future cash flows are sorted by their date of payment. Payments in a certain period are summed up in liability buckets. Furthermore, we compute the obligations that arise from the current wealth of funds where future contributions are not taken into account. Similar to the liability buckets, the obligations are also summed up into obligation buckets. The buckets give a manageable description of the pension fund's liabilities (and obligations) and their term structure. The assets are modelled from the perspective of a Swiss investor. We use a dynamic factor model with heavy tailed residuals to model stock and bond market prices. We propose an optimisation technique for the asset liability management problem where the liability buckets are matched with available wealth of the pension fund. The optimisation problem is to minimise the shortfall of bucket funding while reaching a required future surplus. The solution results in an asset allocation for each liability bucket based on its time horizon. In this way we realise the life-styling hypothesis for each individual across the entire pension fund. In a case study we apply this method to the data of a Swiss pension fund with over 3500 members and over 1 billion (10^9) Swiss Francs of wealth.

Keywords

asset and liability management, life insurance model, asset and liability portfolio optimisation, factor models

JEL classification: C61, G11, G23

1. Introduction

Pension funds manage a significant amount of wealth. It is therefore highly relevant that they manage their wealth in a responsible way while always taking into account their very long time horizon. Asset and liability management for pension funds has several different issues. There is the traditional asset management and portfolio optimisation with the appropriate asset price models. There are also models describing the pension funds liabilities which depend on the type of pension fund. The combination of asset management with the liabilities of a pension fund leads to a true asset liability management.

1.1. Pension fund liability modelling

An important aspect of pension fund liability modelling is the way in which the pension fund members incomes are modelled. The income is central, since it defines the contributions into the pension fund. Very generally, Merton models a deterministic wage income in the framework of dynamic portfolio optimisation with consumption (Merton, 1992, Part II). Boulier, Huang and Thaillard (2001) then provide a pension fund model in continuous time with deterministic salaries. Further, pension fund related papers, such as Battocchio and Menoncin (2004), Cairns, Blake and Dowd (2000) or Bodie et al. (2004) provide stochastic models for wages, modelled with one or more Brownian motions for salary changes due to interest rate changes and company stock price changes, among others. Our model for salaries is based on Cairns (2003), where salaries are increased in relation to a cost-of-living index and an age related function.

Pension-fund models have been used in simulation tools which take into account the mortalities of pension fund members and what this implies on the pension fund liabilities (Kingsland (1982); Winklevoss (1982); Bacinello (1988); Chang (1999); and Ziemba (2003, Chapter 4)). Motivated by an actuarial approach based on the mathematics used for life insurances as in Gerber (1997) or the technical summary given in connection with a Swiss life table EVK (2000) we also use mortality probabilities in our model. A different approach to modelling pension funds is given by Blake (1998), where pension fund schemes are modelled in a framework of financial options. A full book on modelling pension systems (Simonovits, 2003) covers many issues from life cycle, funded, and unfunded systems to issues of demographics and the transition of pension systems.

1.2. Asset and liability management optimisation for pension funds

Contribution rates are an important factor in pension fund management. Since the main interest of (young) pension fund members is mostly to pay small contributions and generate higher returns by investing a larger proportion in riskier assets (such as stocks) in the financial markets. Optimisation is based on a trade-off between contribution rate and the pension fund liquidity, resulting in an investment strategy (O'Brien (1986);

Haberman and Sung (1994); Reichlin (2000); Taylor (2002); and Josa-Fombellida and Rincon-Zapatero (2004)). Further optimisation methods for pension fund investment strategies in continuous time are given in Cairns, Blake and Dowd (2000) and in Haberman and Vigna (2002) for defined contribution pension plans. Optimal investment strategies for special cases of pension funds, such as minimum guarantees or pension fund accumulation and decumulation, are solved with CRRA utility functions in continuous time in Deelstra, Grasselli and Koehl (2003) and in Battocchio, Menoncin and Scaillet (2003), respectively. Furthermore, Devolder, Bosch-Princep and Fabian (2003) and Charupat and Milevsky (2002) describe stochastic optimal control solutions for optimal asset allocation for life annuities. Another approach tries to capture the investment preferences of an individual investor whose investment horizon shortens with advancing age and to implement this behaviour into an investment strategy applicable to the total wealth of the pension fund. The asset allocation is a function of risk aversion and time horizon and has been described in Brennan, Schwartz and Lagnado (1997) and Campbell and Viceira (2002) for the individual and also in Cairns, Blake and Dowd (2003), called stochastic life-styling, as a strategy for pensions saving.

Common risk measures used in asset and liability management situations are Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). Blake, Cairns and Dowd (2001) estimate the VaR with regard to pension plan design and Dowd, Blake and Cairns (2003) investigate long-term VaR. Further, CVaR has been applied to a pension fund in Bogentoft, Romeijn and Uryasev (2001) and Bosch-Príncep, Devolder and Domínguez-Fabián (2002). Several other studies considered asset liability management, although not in the context of pension fund management, they are very relevant and applicable for the specific case of the pension fund. Ziemba and Mulvey (1998), Ziemba (2003), Zenios (2002) authored books covering the broad field of asset and liability management, specially for long-term financial planning. Papers on long-term planning are Mulvey, Pauling and Madey (2003), Mulvey and Shetty (2004), Kusy and Ziemba (1986) and Cariño and Ziemba (1998). Further references on asset and liability management are also given in Section 8.

1.3. Pension funds in the Swiss pension system

We describe a model for the liabilities of a Swiss pension fund. To give a broad overview on the Swiss pension system, we first describe the function of the Swiss pension funds within the total pension system.

The Swiss old-age insurance system consists of three pillars. The first pillar is intended to secure existence, the second pillar is to retain the living standard, and the third consists of individual retirement savings. The first pillar, the state pension, is financed through a national pay-as-you-go pension system. The contributions are split up between employer and employee and they are a fixed percentage of earned wages (currently 8.4%). There is no limitation on contributions (even for above-average salaries).

However, there is a minimum and a maximum pension amount which is paid. The actual pensions are adjusted yearly (e.g., for 2004, monthly pensions for individuals were minimum 1055 Swiss Francs (CHF) and maximum 2110 CHF).

The second pillar, occupational pension, is financed through an employer-specific, earnings-related, and fully funded pension fund. The pension system is designed such that the first and second pillar together approximately result in a pension of 60% of the last wages which, after taxes should make it possible to retain one's living standard after retirement. The participation is mandatory as soon as a given minimal salary is earned (currently 25,320 CHF). Above a certain wage level (75,960 CHF), there are possibilities for a more flexible investment of the additional wealth summarised by a term called "above-mandatory". Contributions towards the pension fund are split up between employer and employee and are a percentage of earned wages (between 7 and 18%, depending on pension fund).

Technically, the funds are legally independent institutions whose funds are not linked with the sponsoring company or any other institution. The sponsoring company decides on whether the pension fund is to be run as a defined benefit (DB) or defined contribution (DC) plan. For the "above-mandatory" savings, more choices can be made by the employee regarding investment strategy. There is a minimum guaranteed annual return on the accumulated wealth specified by law. This return remained unchanged at 4% from 1985 until 2002, in 2003 it was reduced to 3.25% and in 2004 to 2.25%. For the DC plan, the meaning of the minimum return is obvious. For the DB plans the guarantee on the benefits remained the same and the minimum guaranteed return remains merely an accounting issue. Pension funds may be run by autonomous institutions for one (large) company alone. Others may consist of several companies of whose employees are pooled into one single pension fund. Still others may consist of very few members and be run by an actuary. Finally, certain large insurance companies also offer their services to run pension funds. Depending on its structure, a pension fund may be run as a non-profit organisation, whereas, for instance, the insurance companies that offer to run a pension fund want to make a profit from that business.

In the pension plan, the employees are guaranteed pension benefits upon reaching a given retirement age. For receiving full benefits, certain conditions apply with regard to years of service. In the DC scheme, annual pensions (benefit payments) are calculated using a formula based on accumulated wealth at retirement and a predefined factor (currently 7.2% of final wealth is given as annual pension, as specified by law). The factor is based on calculations related to average life expectancy at retirement. In the DB plan, the employees are guaranteed a percentage of their last earned salary (typically around 60%).

In addition to the first and second pillar, the third pillar consists of privately paid, tax-privileged savings. This kind of pension is voluntary and is used as a means to supplement the mandatory pension or, for example, to enable early retirement.

We focus on the second pillar, the (occupational) pension funds.

1.4. Analysis of major issues at managing a fully funded pension fund

In a pension system where the young active members' and the retired passive members' wealth is managed as fully funded pension fund, several (and possibly opposing) interests need to be met. In addition, there are legal requirements, such as minimal return and quantitative diversification rules, as well as uncertainties due to investment strategies and to demographic trends. All parties involved in the pension fund have their own particular interests: the pensioners and the active members shortly before retirement are interested in a secure and stable pension which is best achieved by a low-risk, conservative investment strategy. The younger active members are interested in the highest possible returns in order to augment their future pensions. The sponsoring companies' interests, finally, lie in minimising the need for paying supplemental funds due to under-coverage of the pension fund.

The task of the pension fund manager is to achieve the goals of all parties involved, while observing the legal requirements, achieving a minimal guaranteed return (in order not to become under-covered), and with the additional difficulties of uncertain market returns, liquidity needs and demographic trends.

In addition to the management problem with the opposing interests, Swiss pension funds used to re-distribute gains that were above the guaranteed minimal return. As reported in the press (e.g., in the Swiss newspaper "Neue Zürcher Zeitung" (NZZ, 2004)), a study that was done for a Swiss parliamentary commission found, that it is no longer possible to back-trace both the actual surpluses nor where these were distributed in the past. This surplus could either have been used to build reserves, have been re-distributed to the members, or used for (completely) other purposes that may have not benefited the pension fund members at all. The Swiss pension funds started to get into more and more trouble after the market downturn which started in the year 2000. Although academic models clearly indicated danger in the stock market in advance (Ziemba, 2003, Chapter 2), the general investment community did not expect the downturn after many years of constantly rising markets. As a consequence the pension funds reserves and investment strategies were often not up to their actual risk potential. Together with the guaranteed return which was fixed by a federal law, this led to the under-coverage of many pension fund. According to statistics and the press (e.g., NZZ (2004)), by the end of 2002 over 50% of pension funds were under-covered, i.e., their financial assets were not sufficient to cover their promised (expected) liabilities in the future. We could also observe that assets and liabilities were considered as separate entities leading to short-term liquidity management combined with an asset investment strategy that did not match the liabilities.

In the following sections we propose a solution to the problem of the opposing interests. We do this by analysing the pension funds population by means of life insurance mathematics, since we believe that a lot of information is contained in the structure of the pension fund members which can be exploited in order to find an investment and liquidity management strategy. Therefrom we can derive an asset and liability management strategy which should visibly tackle the interests of the involved parties. The solution not only yields an investment allocation strategy which takes into account the pension fund population structure, but may also be used as a means to fairly distribute possible surpluses and to still fulfil legal constraints.

1.5. Brief comparison of the Swiss and other pension fund systems

The following very brief comparison of international pension fund systems is taken from a larger comprehensive study of the OECD insurance and private pension compendium for emerging markets (OECD, 2001). We only discuss the second pillar or what is commonly organised under occupational pension. Every country we regarded has a very distinct and characteristic pension system. The pension systems mostly grew with the countries individual requirements and historical, economical and demographic issues. It is therefore hard to compare pension systems. We provide this overview for interested readers, who may wish to compare their own (and best known) pension system with the Swiss system.

Switzerland's occupational pensions:

Funding and risk bearing: Occupational pensions are an earnings-related fully funded system. Most funds are DC funds, whereas DB schemes are represented mainly in the public sector. There is a minimum nominal rate on the pensions savings (1985–2002: 4%, 2003: 3.5%, 2004: 2.25%). Excess returns must be re-distributed within the pension fund with a quota of at least 90% in the form of interruption of contribution payment (for active members and sponsoring companies), additional benefits (for pensioners) or reserves. The **Administration** of the pension funds can be done either by non-profit foundations, cooperative societies or as institutions incorporated under public law. **Participation** is mandatory for employees. The employers mandatorily decide on the pension fund for their employees. Contributions towards the pension fund are split between employee and employer, whereby the employer must at least pay 50%. **Minimum diversification requirements** apply: Investment in debt instruments of a single entity (except government bonds, banks, and insurance companies) is limited to 10% (5% for foreign assets). Investment in equity of a single company is limited to 10% (5% for foreign assets). Self-investment in the sponsoring company is limited to 10%. Investment in derivatives is allowable for hedging purposes only. There is an overall limit in investments in foreign bonds of 30%, of 25% in foreign equities and of 20% in foreign currency bonds. There are aggregate limits for domestic and foreign equity (50%), foreign bonds and foreign currency bonds (30%), as well as real estate and Swiss and foreign equity (70%). **Supervision** is effected by the federal office.

USA occupational pensions:

Funding and risk bearing: Occupational pensions may be made available as funded DB, hybrid or DC plans without any guarantees. They are **administered** by the sponsor (assets are managed in a closed pension fund by trustees), life insurance companies or as collective investment schemes (401(k)). **Participation** is voluntary. There are general requirements for **diversification**: For all DB plans and some DC plans: 10% limit on investment in employer securities or real property; no transactions with parties in interest (i.e., a fiduciary, provider of services, participant, plan sponsor, beneficiary, or some other party with a relationship to the plan). Assets must be under the jurisdiction of US courts.

USA Individual Retirement Accounts (IRA):

Funding and risk bearing: IRAs are fully funded DB or DC schemes. IRAs invested in mutual funds or bank deposits offer no guarantees. IRAs invested in annuities offer their guarantees. They are **administered** either by a collective investment scheme provider, a bank or an insurance company (annuity). **Participation** is voluntary. There are general requirements for **diversification**: For all DB plans and some DC plans: 10% limit on investment in employer securities or real property; no transactions with parties in interest. Assets must be under the jurisdiction of US courts.

UK contracted-out schemes:

In contracted-out schemes members have their state pension (SERPS) rights included within their private scheme. **Funding and risk bearing**: They are funded DB or DC schemes. For DC plans, there is minimum mandatory annuisation (annuity must be bought at age 75). They are **administered** by trustees (via a closed pension fund) or a life insurance company. **Participation** is voluntary, but in case of an opt-out, they must sign up with an appropriate personal pension plan. There are general requirements for **diversification** and suitability. Employer related investment is limited to 5%. There are no quantitative portfolio restrictions.

UK contracted-in schemes:

In contracted-in schemes members have their state pension (SERPS) rights treated separately from their private scheme (private pensions benefits are paid in addition to SERPS). **Funding and risk bearing**: They are funded DB or DC schemes. Arrangements which may be unfunded to provide executives with extra retirement provision on earnings above the earnings cap. They are **administered** by trustees (via a closed pension fund) or a life insurance company. **Participation** is voluntary. There are general requirements for **diversification** and suitability. Employer related investment is limited to 5%. There are no quantitative portfolio restrictions.

Netherlands occupational pension:

Funding and risk bearing: The occupational pensions are funded and are mainly DB schemes. They are **administered** by a closed pension fund (foundation) or an insurance company. **Participation** for Employers is compulsory under collective bargaining arrangements and by statute in certain sectors. For employees participation is mandatory. **Diversification** is required but there are no quantitative rules.

2. Liability model for a Swiss pension fund

2.1. General assumptions

The pension fund's liabilities consist of all current and future payments towards pensioners and insured active members. Pensioners receive a pension which is an annuity based on their income at retirement age or wealth accumulated by the time of retirement. Active members pay contributions during their working years in order to accumulate wealth for retirement age. Wealth is compounded with a minimal return. The rate of minimal return is specified by law. Active members may leave the pension fund, in which case the accumulated wealth must be transferred to their new pension fund. To get a detailed description of the pension funds payment streams we base our model on every payment to every pension fund member. This means that we regard the (remaining) expected payments to pensioners as well as projected expected outflows to today's active members. In the calculation of liabilities we make several assumptions:

- The pension fund has a given population of active members, pensioners and disabled members. We know age, current salary and wealth of each of the active members and pensions of retired and disabled members.
- Individuals start accumulating wealth at age 25. Men retire at 65, women at 63.
- Individuals salaries have age- and cost-of-living related increases.
- Uncertainty factors for active members are mortality, risk of becoming disabled and early exit from the pension fund. For pensioners we regard the mortality risk.
- We regard a pension fund large enough for statistical properties to apply.
- All payments are done at the end of the year.

2.2. Pension fund liabilities and obligations

The pension fund's liabilities can be modelled in two different ways, each having its later use in the asset and liability management optimisation. We differentiate between obligations and liabilities in the following way:

Obligations summarise the pension fund's promised payments to pensioners and active members based on their current wealth. For obligations, we do not take into account any future contributions into the pension fund by active members. Pensioner's obligations are the remaining pension benefit payments until death. For the active members, the obligations consist of the wealth accumulated at the moment which is compounded with the minimal return.

Liabilities consist of the pension fund's promised payments to pensioners and active members taking into account current wealth and including outstanding future contributions. Pensioners liabilities are the remaining pension benefit payments until death. For active members, we make assumptions on their projected future wages which leads to future contributions into the pension fund. When we project the future balance of the members' accumulated wealth, we do not only compound with the guaranteed return, but future contributions are also included in the compounding process.

For pensioners there is no difference in obligations and liabilities. For obligations no assumptions are made on future wages and contribution payments. For liabilities these assumptions are used. Since contributions are not included and thus the wealth only accumulates with the (guaranteed) return, the obligations will always be smaller than the liabilities until the moment of retirement. By recalculating obligations every year (after contributions have been paid) we can update the past obligations calculation with the new obligations based on accumulated wealth plus new contributions. This way, the older the member gets, the difference between obligations and liabilities grows smaller and smaller. These characteristics of obligations and liabilities are described in the sections on modelling obligations and liabilities (Sections 4–6).

2.3. Pension fund bucket structure for asset management

We consider payments to every member individually. We thereby not only regard the amount, but also the instant at which the payment is due. This provides us with the expected payment stream over the regarded time horizon for every member in the pension fund. We further collect all payments due in a certain time period (e.g., one year) and call this the liability bucket for liabilities or obligations bucket for obligations. Liability payments due in the next year are then collected in the one-year liability bucket, payments due in two years in the two-year liability bucket, and payments due in j years are collected in the j-year liability bucket. The same can be done for obligations. This procedure results in the long-term structure of payments out of the pension fund. The bucket structure is then needed as an instrument for an advanced term-structure-oriented asset management.

3. Basic principles of the pension fund model

3.1. An actuarial perspective for pension funds

Not only do we regard guaranteed pension benefits and payment streams, but also the uncertainties that they are subjected to. Our model and the notation are influenced by the description used by actuaries in life insurance mathematics (e.g., as in Gerber (1997) and Koller (2000)). Apart from using survival probabilities and probabilities of becoming disabled, we need to know the probabilities for members leaving the pension fund before retirement, e.g., due to changing their employer, or leaving the country, summarised here by the term “labour mobility”. These figures have been derived in a study by Rufibach et al. (2001) based on data of the pension fund for the employees of all Swiss federal institutions.

All of the cases of uncertainty mentioned above have an influence on the pension fund payment streams. However, we cannot know whether or when they will occur. They can therefore be considered as risk factors of the pension fund. Since every individual member is assumed to be independent of the other, we can aggregate individual risks over the entire pension fund. The risk factors of the pension fund are highly dependent on the pension funds population and there are big differences between different population groups, young and old, or men and women, for instance. We will first go into the details of the different probabilities and then define a measure which gives us the total probability for a member to remain in the pension fund for a given number of years at a given age.

3.2. Death probability and probability of invalidity

Death probabilities are given in “life tables” consisting of the one-year death probabilities. They are built from statistical data which can be derived from a whole population or from a more specific, smaller group, such as pension fund members. When we consider a person aged α years, in year t we denote this by $\alpha(t)$. The one-year death probability for this so-called “life aged $\alpha(\cdot)$” is then $q_{\alpha(\cdot)}$. The one-year survival probability is given as $p_{\alpha(\cdot)} = 1 - q_{\alpha(\cdot)}$. Death probabilities for men and women differ mainly at high ages. We use the table (EVK, 2000) given by the Swiss pension fund for the Swiss federal employees based on its own statistics. This table is updated and published every ten years. The maximum age considered here is 105 years denoted by $\alpha_{\max}$. The survival probability over several years is

$$_jp_{\alpha(t)} = \prod_{i=0}^{j-1} p_{\alpha(t+i)}, \quad j \geqslant 1 \tag{1}$$

which is the probability of a life aged α at time t to survive the next j years. The one-year invalidity probability $^*i_{\alpha(\cdot)}$ for a life aged $\alpha(\cdot)$ is the probability of a person becoming disabled within one year at a given age. The invalidity probability is given in tables similar to the known life tables until retirement age α_R (i.e., it is also contained in (EVK, 2000)). There is a much higher invalidity probability (almost with a factor of 2) for women below 50 than there is for men at the same age, whereas the invalidity probability for men rises rapidly for ages above 50.

3.3. Exit probability and entry rate

Exit probabilities depend strongly on the sample from which the statistical data is taken. We consider the one-year exit probabilities for a given age $\alpha(\cdot)$, denoted by $to_{\alpha(\cdot)}$, as found in (Rufibach et al., 2001) for the pension fund of the Swiss federal employees. The ages are restricted to 60 years for men and 57 for women since it is assumed that labour mobility can be disregarded within five years before retirement. The entry rate is defined as the number of new entries at a given age in relation to the actual number of members at the same age. For a given age $\alpha(\cdot)$, we denote the entry rate by $ti_{\alpha(\cdot)}$. Here too, there is a big difference between men and women and different ages.

3.4. Total exit probability

Active members face the risk of not surviving the next year, $q_{\alpha(\cdot)} = 1 - p_{\alpha(\cdot)}$, risk of falling disabled, $^*i_{\alpha(\cdot)}$ and they may leave the pension fund (i.e., change employer), with probability $to_{\alpha(\cdot)}$. From one year to the next, the age-dependent probability for an active member to stay in the pension fund is

$$n_{\alpha(\cdot)} = 1 - (1 - p_{\alpha(\cdot)}) - {}^*i_{\alpha(\cdot)} - to_{\alpha(\cdot)}. \tag{2}$$

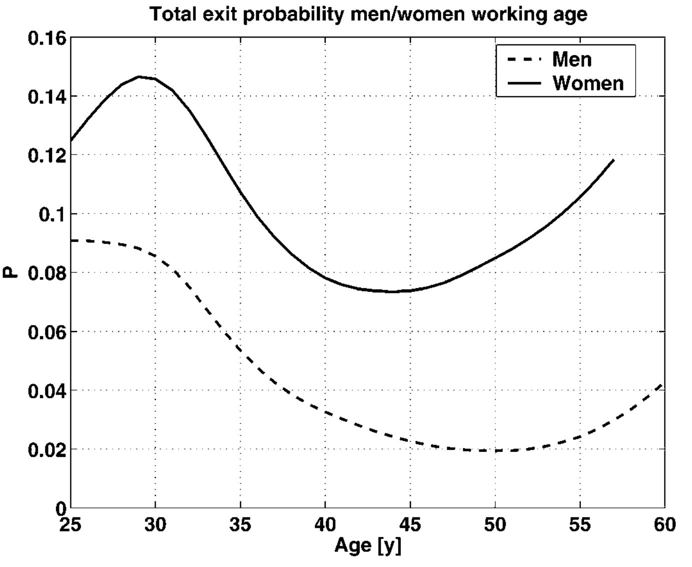


Fig. 1. Total exiting probability for men and women at working ages.

Then, $1 - n_{\alpha(\cdot)}$ is the one-year total exit probability for a member aged $\alpha(\cdot)$. This probability is shown in Figure 1 for men and women. The probability of an active member with age α at time t reaching an age α_τ in the future still being a member of the pension fund is

$$_{\alpha_\tau}n_{\alpha(\cdot)} = \prod_{i=\alpha(\cdot)}^{\alpha_\tau} \left[p_i - {}^*i_i - to_i\right]. \tag{3}$$

With Eq. (3), we can find the probabilities of staying a number of years in the pension fund for every age $\alpha(\cdot)$. Figure 2 shows the curves generated for men of four different ages (25, 35, 45, and 55 years). As an example, we can find here that the probability for a man 25 years old to stay with the pension fund until 35 is approximately 0.43. To stay until pension, the probability for a 25-year-old is at 20%. Similar curves can be calculated for women. These probabilities depend on the sample and very different figures might appear for another set of statistical data of another pension fund.

3.5. *Earnings projections*

3.5.1. *Projected wealth and salary*

To project future pensions, we need to know final expected salaries or accumulated wealth at retirement (depending on the type of pension fund). Salaries may rise due to a multiplicative age-related factor (as described in Cairns (1994)), denoted by

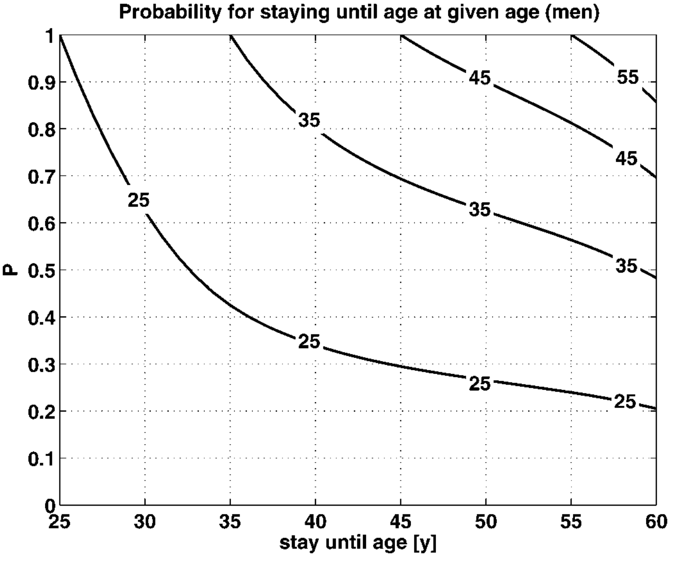


Fig. 2. Total exiting probability for men and women at working ages.

$ar(S(t), \alpha(t))$ and similarly, a multiplicative cost-of-living related factor denoted by $col(t)$, respectively. Let $\widetilde{S}(\cdot)$ be the projected salary based on this year's salary. Next year's expected salary is

$$\widetilde{S}(t+1) = S(t)ar\bigl(S(t), \alpha(t+1)\bigr)col(t+1).$$

Salaries for $j \geqslant 1$ are projected by

$$\widetilde{S}(t+j) = S(t)\prod_{i=1}^{j} ar\bigl(S(t), \alpha(t+i)\bigr)col(t+i). \tag{4}$$

Factors for salary rises are given as follows:

Age-related increases $ar(S(t), \alpha(t))$**:** Salaries rise as a function of age due to work experience or career advances and also as a function of the current salary level. On average, higher salaries have greater increases than smaller ones and salary rises are larger for younger employees. The age-related effect may also become negative as employees get older. For a more detailed description of the effect, also with a differentiation between sex, education, and management levels, see Dondi (2003). The effects of earnings over a lifetime, also summarised under "life-cycle model", are also described in Campbell and Viceira (2002), Clark et al. (2004), and Simonovits (2003). Detailed country studies of salary distributions are in Pittau and Zelli (2002) for Italy and for Brazil in Cowell, Ferreira and Litchfield (1998), respectively.

Cost-of-living-related increases $col(t)$**:** Cost-of-living-related increases reflect the effect of inflation on salaries. Here, we model the instantaneous rise in cost of living with the relative change of the consumer price index, denoted by $r_{col}(t)$, with a first-order autoregressive process (AR(1)),

$$r_{col}(t) = Ar_{col}(t-1) + b + \nu\varepsilon_r(t), \tag{5}$$

with factor loading parameter A, constant b, variance parameter ν, and with the Gaussian white noise $\varepsilon_r(t)$. As factor, we use the Swiss consumer price index (CPI). The parameters are estimated using ordinary least square, as described in Hamilton (1994).

For obligations, the projected wealth in j years $\widetilde{W}_{ob}(t+j)$ is obtained just by compounding the actual wealth with the guaranteed return

$$\widetilde{W}_{ob}(t+j) = W(t)(1+r_G)^j. \tag{6}$$

To calculate the liabilities projected wealth in j years $\widetilde{W}_{li}(t+j)$, the active member's wealth is accumulated by adding up regular payments and investing the already saved wealth at the guaranteed interest r_G. Contribution payments consist of an (age-dependent) fraction $\beta(\alpha(t))$ of the salary $S(t)$. The projected wealth is

$$\widetilde{W}_{li}(t+j) = W(t)(1+r_G)^j + \sum_{i=0}^{j-1} S(t+i)\beta\big(\alpha(t+i)\big)(1+r_G)^{j-i-1}, \tag{7}$$

which is the current wealth compounded with r_G, the guaranteed return, plus the sum of all projected contribution payments over the next j years. The latter are also compounded with r_G for every year after they have been paid into the fund. Define the projected wealth accumulated at retirement age $\alpha(t+j) = \alpha_R$ as $\widetilde{W}_{ob}^{(\alpha_R)}(t+j)$ for obligations and $\widetilde{W}_{li}^{(\alpha_R)}(t+j)$ for liabilities. The final salary at retirement age is $\widetilde{S}^{(\alpha_R)}(t+j)$.

3.6. Pension fund plans

For the two different kinds of pension plans (DB and DC), the expected annual benefit payments at retirement are calculated as follows:

DC plan: In the DC plan, the annual pension is calculated as a fraction λ of the wealth accumulated at retirement age $\alpha(t) = \alpha_R$, $\widetilde{W}_{ob}^{(\alpha_R)}(\cdot)$ or $\widetilde{W}_{li}^{(\alpha_R)}(\cdot)$ for obligations and liabilities, respectively. The projected pension for an active member is then

$$\begin{aligned} &\text{Obligations:} \quad \tilde{b}_A(t) = \lambda \widetilde{W}_{ob}^{(\alpha_R)}(t+j), \quad \text{when } \alpha(t+j) \geqslant \alpha_R,\\ &\text{Liabilities:} \quad \tilde{b}_A(t) = \lambda \widetilde{W}_{li}^{(\alpha_R)}(t+j), \quad \text{when } \alpha(t+j) \geqslant \alpha_R. \end{aligned} \tag{8}$$

DB plan: In the DB plan, the annual pension is given as a fraction κ of the salary earned in the last year before retirement

$$\begin{aligned} &\text{Obligations:} \quad \tilde{b}_A(t) = \kappa \widetilde{S}^{(\alpha_R)}(t+j), \quad \text{when } \alpha(t+j) \geqslant \alpha_R,\\ &\text{Liabilities:} \quad \tilde{b}_A(t) = \kappa \widetilde{S}^{(\alpha_R)}(t+j), \quad \text{when } \alpha(t+j) \geqslant \alpha_R. \end{aligned} \tag{9}$$

3.6.1. Member specific wealth and salaries

In the previous sections we have given a model for wages and wealth accumulation for every individual member. When regarding the whole pension fund, we need to be able to differentiate every single member's wealth and salary and their projections. We do this by assigning every member an identification value, which is typically done by a unique member number using a variable θ. We can specify the age of member θ in year t by $\alpha(\theta, t)$. For instance, the current wealth of θ in year t is then $W(\theta, t)$.

4. The pension funds' current obligations

Obligations to pensioners consist of pensions payments. For active members the wealth $W(\theta, t)$ at time t that the pension fund has accumulated for member θ may be further accumulated until retirement or else have to be paid when the member leaves the fund early.

4.1. Current obligations to pensioners

For pensioners, we do not necessarily keep the information of wealth and wages at retirement. We simply know the annual benefit payments $b_P(\cdot)$ that are paid until the pensioner dies. They may be increased to account for rises in cost of living. Mortality is considered the only uncertainty for pensioners. Then the expected payment stream of the obligations to a pensioner θ for next year is $\mathrm{E}[ob_P(\theta, t)] = b_P(\theta, t)\, p_{\alpha(\theta,t)}$. The expected obligations to a pensioner θ in j years are then

$$\mathrm{E}\big[ob_P(\theta, t + j)\big] = {}_j p_{\alpha(\theta,t)} b_P(\theta, t + j), \tag{10}$$

the pension payment $b_P(\theta, t + j)$ multiplied with the one-year survival probability after having survived the next j years, ${}_j p_{\alpha(\theta,t)}$. Figure 3 shows the expected payment stream of obligations to pensioner θ with age 65 and pension $b_P(\theta, t + j)$.

4.2. Current obligations to active members arising before retirement

For active members, when we look a certain amount of time ahead, we need to distinguish their status at that future time. The expected payments are different whether

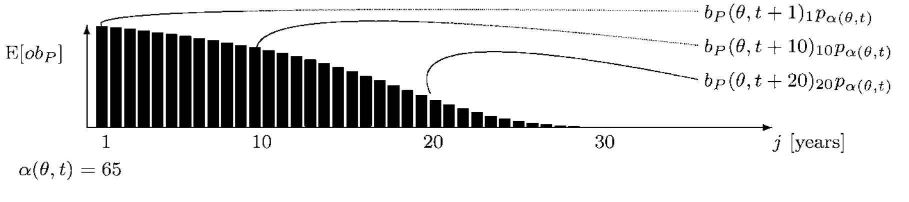


Fig. 3. Expected obligations payment stream to pensioner with age 65.

the member has reached retirement age in the mean-time or is still in the working age. When members have retired, expected obligations are based on their pension payments. While they are still in the working age, expected obligations are based on payments due to early leaving of the pension fund.

An active member θ at time t, which is still active at time $t+1$ (i.e., younger than retirement age) may still leave the pension fund early. The obligation which the pension fund must expect in this case, ob_A is

$$\mathrm{E}\big[ob_A(\theta, t+1)\big] = \widetilde{W}_{ob}(\theta, t+1) to_{\alpha(\theta,t+1)}, \quad \text{when } \alpha(\theta, t+1) < \alpha_R. \tag{11}$$

The expected payment for next year $\mathrm{E}[ob_A(\theta, t+1)]$ is the current accumulated wealth for member θ, $W(\theta, t)$, compounded with the guaranteed return r_G for one year, multiplied with the exiting probability depending on θ's age next year, $to_{\alpha(\theta,t+1)}$. In j years, assuming the member to remain active, the expected obligation arising with today's accumulated wealth is

$$\mathrm{E}\big[ob_A(\theta, t+j)\big] = {}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)} \widetilde{W}_{ob}(\theta, t+j) to_{\alpha(\theta,t+j)}, \\ \alpha(\theta, t+j) < \alpha_R. \tag{12}$$

This is the current wealth, compounded with the guaranteed return over j years, $\widetilde{W}_{ob}(\theta, t+j)$, multiplied with the probability of the member remaining active until $j-1$ (the year before), ${}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)}$, multiplied with the exiting probability at the age in j years, $to_{\alpha(\theta,t+j)}$.

4.3. Current obligations to active members arising after retirement

When we compound the actual wealth with the guaranteed return r_G for the number of years remaining until pension ($\alpha_R - \alpha(\theta, t)$), we obtain the wealth accumulated for retirement obligations as $W(\theta, t)(1+r_G)^{\alpha_R - \alpha(\theta,t)}$. The expected pension for a DC plan arising from the expected wealth is

$$\tilde{b}_A(\theta, t) = \lambda W(\theta, t)(1 + r_G)^{\alpha_R - \alpha(\theta,t)}, \tag{13}$$

as given in Eq. (8). For the DB plan, $\tilde{b}_A(\theta, t)$ is calculated using to Eq. (9) depending on expected wages at retirement age. The obligation to the active member who has retired in the year considered is ob_{AP}. The expected obligation is

$$\mathrm{E}\big[ob_{AP}(\theta, t+j)\big] = {}_{\alpha_R}n_{\alpha(\theta,t)} \tilde{b}_A(\theta, t+j)\, {}_j p_{\alpha_R}, \quad \alpha(\theta, t+j) \geqslant \alpha_R. \tag{14}$$

This is the projected pension $\tilde{b}_A(\theta, t+j)$ multiplied with the one-year survival probability in j years (where the age in j years is above retirement) and with the probability of the member staying in the fund until pension ${}_{\alpha_R}n_{\alpha(\theta,t)}$.

Figure 4 shows the different expected payment streams for the obligations arising before and after pension. The structure of the payment stream before pension has an distinctive shape. The peak at the beginning is explained by the high exit probabilities in the young years between 25 and 35 as described in Section 3.3. The valley at ages

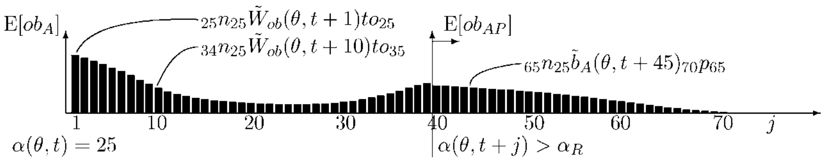


Fig. 4. Expected payment stream of obligations to an active member aged 25 while being an active member and after retirement.

40 to 55 ($j = 15$–30) is due to low exiting probability in combination with small invalidity probabilities. In the years just before pension, the effects of higher probability of falling invalid (Section 3.2) in combination with the very long compounding time (35 to 40 years) causes the expected obligations to rise again (Eq. (12)). The wealth which remains for pension (in the case of the member who stays in the pension fund until retirement) is taken for the calculation of the pension and is finally distributed over the expected lifetime as pensioner, as described by Eq. (14).

5. The pension funds' projected liabilities

Pensioners' liabilities consist of the payments of the remaining pensions. For active members, we distinguish between wealth $W(\theta, t)$ at time t that the pension fund has accumulated for member θ which might be further accumulated until retirement and what might be paid out before that date. Here, we also include future projected contributions into the pension fund and the interest earned on the contributions for every year.

5.1. Projected liabilities to pensioners

The annual benefit payment to pensioner θ is $b_P(\theta, t)$. The liability arising from this year's pension payment to pensioner θ is $li_P(\theta, t)$. Annual payments are paid until the pensioner dies. They may be increased to account for rises in the cost of living after retirement. Mortality is the only uncertainty for pensioners.

Taking into account mortality, the expected liability arising from the payment of θ's pension for next year is $\mathrm{E}[li(\theta, t)] = b_P(\theta, t)p_{\alpha(\theta,t)}$. The expected liabilities to θ in j years are

$$\mathrm{E}\big[li_P(\theta, t+j)\big] = {}_j p_{\alpha(\theta,t)} b_P(\theta, t+j). \tag{15}$$

This is the survival probability for a year in j years ${}_j p_{\alpha(\theta,t)}$, multiplied with the pension benefit $b_P(\theta, t+j)$.

5.1.1. Projected future liabilities to active members before retirement

As seen in the case of obligations, the pension fund must take into account two possible types of future cash outflows for liabilities for its active members. Members either stay with the pension fund until retirement or will leave before that date. In retirement, they will receive a regular pension depending on their pension scheme. In case of a member leaving before retirement, the pension fund must transfer the full amount of accumulated wealth to the member's new retirement plan. A considerable amount of funding is transferred in that manner and it cannot be disregarded since it is of importance with regard to short term liquidity planning. The liability arising for active member θ which is due to labour mobility is $li_A(\theta, t)$.

For member θ, with current accumulated wealth $W(\theta, t)$ and the probability of leaving the pension fund $to_{\alpha(\theta,t)}$, next year's expected payment (due to labour mobility) is

$$\mathrm{E}\big[li_A(\theta, t+1)\big] = to_{\alpha(\theta,t+1)} \widetilde{W}_{li}(\theta, t+1). \tag{16}$$

In j years' time, we expect the total payment of

$$\mathrm{E}\big[li_A(\theta, t+j)\big] = {}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)} \widetilde{W}_{li}(\theta, t+j) to_{\alpha(\theta,t+j)}, \quad \alpha(\theta, t+j) < \alpha_R, \tag{17}$$

where ${}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)}$ is the probability of the member to still be an active member in $j-1$ years. Further, $\widetilde{W}_{li}(\theta, t+j)$ is the expected accumulated wealth of member θ at time $t+j$ according to Eq. (7) and $to_{\alpha(\theta,t+j)}$ is the exit probability, given the age of member θ in year $t+j$.

5.2. *Projected liabilities to active members after retirement*

The liabilities arising after retirement to member θ who is active at time t but retired in year $t+j$ are $li_{AP}(\theta, t+j)$. The expected liabilities in j years arising from the pension benefits are

$$\mathrm{E}\big[li_{AP}(\theta, t+j)\big] = {}_{\alpha_R}n_{\alpha(\theta,t)} \tilde{b}_A(\theta, t+j)\, {}_j p_{\alpha_R}, \quad \alpha(\theta, t+j) \geqslant \alpha_R. \tag{18}$$

This is the probability for reaching retirement age as a member of the pension fund given today's age, ${}_{\alpha_R}n_{\alpha(\theta,t)}$, multiplied with the projected benefit payment for the active member $\tilde{b}_A(\theta, t+j)$, and multiplied with the probability of surviving j years after retirement ${}_j p_{\alpha_R}$. The projected pension benefit for the DC plan is given as $\tilde{b}_A(\theta, t+j) = \lambda \widetilde{W}_{li}^{(\alpha_R)}(\theta, t+j)$ and for the DB plan with final projected wages $\tilde{b}_A(\theta, t+j) = \kappa \widetilde{S}^{(\alpha_R)}(\theta, t+j)$.

6. Construction of the bucket structure

6.1. *Bucket structure of expected undiscounted liabilities*

The pension and labour mobility liability structure shows the pension fund liabilities, taking into account when the payment is due. Next year's payments are calculated as the sum of all payments to current pensioners, plus the payments to the members that reach retirement age next year, plus all payments to active members leaving the pension fund next year due to labour mobility. This results in

$$\mathrm{E}\big[LB_1(t)\big] = \sum_{\theta=1}^{\Theta(t)} \big(\Gamma_{P1}(t)1_P(\theta,t) + \Gamma_{AP1}(t)1_{AP}(\theta,t+1) + \Gamma_{A1}(t)1_A(\theta,t+1)\big), \tag{19}$$

with

$$\Gamma_{P1}(t) = b_P(\theta,t+1)\,p_{\alpha(\theta,t)},$$
$$\Gamma_{AP1}(t) = {}_{\alpha_R}n_{\alpha(\theta,t)}\tilde{b}_A(\theta,t+1)\,p_{\alpha_R},$$
$$\Gamma_{A1}(t) = to_{\alpha(\theta,t+1)}\widetilde{W}_{li}(\theta,t+1),$$

where $\Theta(t)$ is the set of pension fund members in year t, and the indicator function $1_P(\theta,t) = 1$ if member θ is a pensioner in year t, and 0 otherwise. Furthermore the indicator function $1_{AP}(\theta,t+1) = 1$ for member θ who is an active member in year t and retires in year $t+1$; zero otherwise, and $1_A(\theta,t+1) = 1$ for member θ who is an active member in year $t+1$, zero otherwise. The factors $\Gamma_{P1}(t)$, $\Gamma_{AP1}(t)$ and $\Gamma_{A1}(t)$ are the contributions of the members towards the buckets. The factor $\tilde{b}_A(\cdot)$ is the projected pension for the active members at the time they retire. For a DC fund, it is given as $\tilde{b}_A(\theta,t+1) = \lambda\widetilde{W}_{Li}(\theta,t+1)$, and for a DB fund $\tilde{b}_A(\theta,t+1) = \kappa\widetilde{S}(\theta,t+1)$, as defined in Eqs. (8) and (9).

The expected non-discounted payments due in j years are

$$\mathrm{E}\big[LB_j(t)\big] = \sum_{\theta=1}^{\Theta(t)} \big(\Gamma_{Pj}(t)1_P(\theta,t) + \Gamma_{APj}(t)1_{AP}(\theta,t+j) + \Gamma_{Pj}(t)1_A(\theta,t+j)\big), \tag{20}$$

where

$$\Gamma_{Pj}(t) = b_P(\theta,t+j)\,{}_jp_{\alpha(\theta,t)},$$
$$\Gamma_{APj}(t) = {}_{\alpha_R}n_{\alpha(\theta,t)}\tilde{b}_A(\theta,t+j)\,{}_jp_{\alpha_R},$$
$$\Gamma_{Aj}(t) = {}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)}\widetilde{W}_{li}(\theta,t+j)\,to_{\alpha(\theta,t+j)}.$$

This liability bucket structure shows the distribution of future cash flows and indicates what the undiscounted actual values of payments will be in future years.

6.2. *Bucket structure of non-discounted obligations*

Summarising from Sections 4.2 and 4.3 we can define the obligations arising from current accumulated wealth

$$\mathrm{E}\big[OB_j(t)\big] = \sum_{\theta=1}^{\Theta(t)} \big(\Lambda_{Pj}(t)1_P(\theta,t) + \Lambda_{Aj}(t)1_A(\theta,t+j) + \Lambda_{APj}(t)1_{AP}(\theta,t+j)\big), \tag{21}$$

Table 1
Pension fund members with their pension fund related data as explained in the text (annual salaries and wealth in thousands of CHF)

θ	$\alpha(\theta,t)$	$S(\theta,t)$	$\widetilde{S}^{(\alpha_R)}(\theta,t+j)$ at pension	$W(\theta,t)$	$\widetilde{W}_{li}^{(\alpha_R)}(\theta,t+j)$ at pension	DB plan $\kappa=0.65$	${}_{\alpha_R}n_{\alpha(\theta,t)}$	$p_{\alpha(\theta,t)}$
A	25	69	131	0	1600	$\tilde{b}_A = 85$	0.17	0.9993
B	45	113	126	238	1080	$\tilde{b}_A = 82$	0.56	0.9980
C	65	–	–	695	–	$b_P = 48$	–	0.9852
D	75	–	–	356	–	$b_P = 33$	–	0.9623

where

$$\Lambda_{Pj}(t) = {}_jp_{\alpha(\theta,t)}b_P(\theta,t+j),$$
$$\Lambda_{APj}(t) = {}_{\alpha_R}n_{\alpha(\theta,t)}\tilde{b}_A(\theta,t+j){}_jp_{\alpha_R},$$
$$\Lambda_{Aj}(t) = {}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)}\widetilde{W}_{ob}(\theta,t+j)to_{\alpha(\theta,t+j)}.$$

As in Section 6.1, the indicator functions 1_P, 1_A, and 1_{AP} delimitate pensioners, active members, and active members who became pensioners, respectively.

6.3. *Example: Buckets for pension fund liabilities and obligations*

We consider four male members of the pension fund, $\Theta = \{A, B, C, D\}$, and $\theta = A$, B, C or D. In Table 1, we find their ages $\alpha(\theta,t)$, momentary salaries $S(\theta,t)$, projected salary at retirement $\widetilde{S}(\theta,t+j)$, their momentary accumulated wealth $W(\theta,t)$, projected wealth at retirement $\widetilde{W}(\theta,t+j)$, the probability of staying with the pension fund until retirement ${}_{\alpha_R}n_{\alpha(\theta,t)}$, and the expected survival probability for the next year $p_{\alpha(\theta,t)}$.

Figure 5 shows the expected payment streams for every member θ as the bucket structure. For the total pension liabilities (left column), the pension fund sums up all of the payments for every year as shown in the bottom graph. The sum of all the obligations is shown in the right column.

6.4. *Bucket structure uncertainties*

In previous sections, we derived expected values for liabilities and obligations for pensioners and active members based on life tables and other probabilistic figures. We assumed that the members are independent and are not influenced by the other members, thereby, for instance, neglecting slight differences between life tables for widowers and widows and their married peers. In this section, we derive the bucket uncertainties described by the variances of the bucket values.

When we look j years ahead, we can observe two possible outcomes for pensioners. Either they will die before j years have passed, or they will still live. An active member

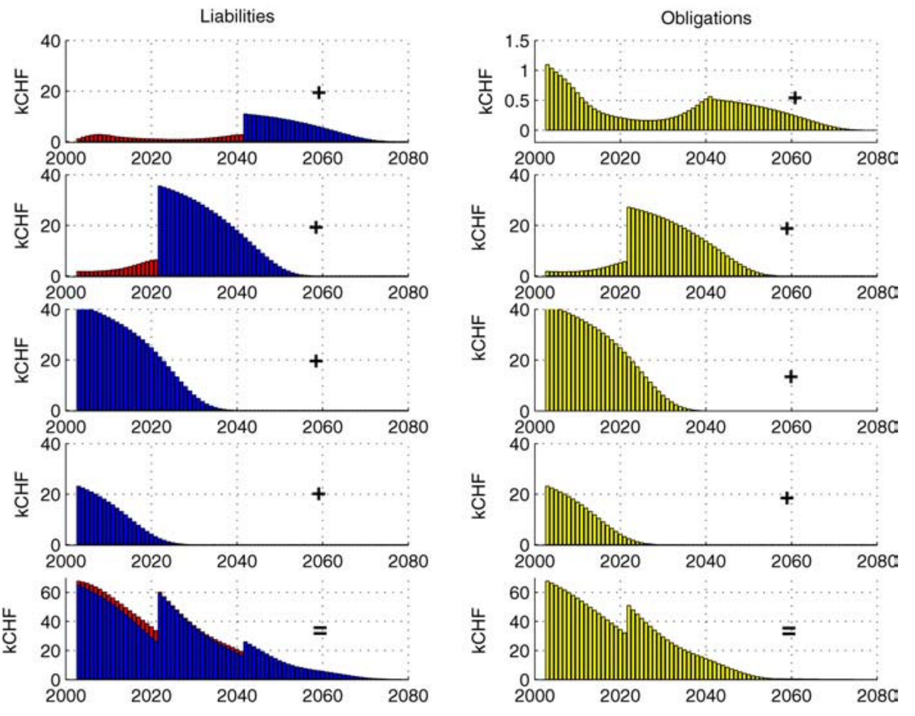


Fig. 5. Pensions liability buckets (left) and obligation buckets (right) for four members and the sum for the pension fund (bottom). Note the different scales of the axes.

who is still active in j years is either still in the pension fund or has left before. And finally, active members who are older than the retirement age in j years, they could have left the pension fund before retirement, or, if they stayed until pension, there are the possibilities of their still being alive or being dead by now. In all three cases, we can observe two possible outcomes, so we can use binomial trees to model the probabilities and the outcomes, which in our case define the payments due to the members. Such a binomial tree is shown for the general case in Figure 6 where the probability of moving on the "up" branch is p invoking a payment of Pay_u. The probability of moving on the "down" branch is given by $1 - p$, which invokes a payment of Pay_d.

In this general case, the expected value of the payment in year $t + j$ is

$$\mathrm{E}\big[\text{Pay}(t + j)\big] = p\text{Pay}_u + (1 - p)\text{Pay}_d, \tag{22}$$

with variance

$$\text{var}\big[\text{Pay}(t + j)\big] = \big(\text{Pay}_u - \mathrm{E}[\text{Pay}]\big)^2 p + \big(\text{Pay}_d - \mathrm{E}[\text{Pay}]\big)^2 (1 - p). \tag{23}$$

We can define the probabilities and the payments which occur. The different cases are shown in Table 2. We use general forms for wealth and benefits ($\widehat{W}(\cdot)$ and $\hat{b}(\cdot)$), which

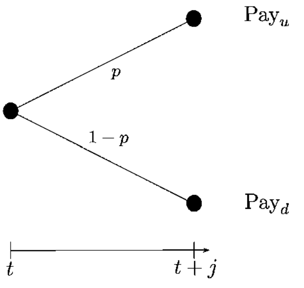


Fig. 6. Tree for payments and probable outcomes.

Table 2
General case applied to pension fund probabilities and payments. Wealth $\widehat{W}(\cdot)$ and benefits $\hat{b}(\cdot)$ must be adjusted to the specific case (BD or DC)

Case	$\alpha(t)$	$\alpha(t+j)$	p	Pay_u	Pay_d
P	$> \alpha_R$	$> \alpha_R$	${}_jp_{\alpha(\theta,t)}$	$\hat{b}(\theta, t+j)$	0
AP	$< \alpha_R$	$> \alpha_R$	${}_{\alpha_R}n_{\alpha(\theta,t)}\,{}_jp_{\alpha(\theta,t)}$	$\hat{b}(\theta, t+j)$	0
A	$< \alpha_R$	$< \alpha_R$	${}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)}to_{\alpha(\theta,t+j)}$	$\widehat{W}(\theta, t+j)$	0

can be specified further for the different cases, such as DB and DC funds, and active and passive members. All we need to know are the probabilities and the payments arising for the different cases of pensioners and active members in order to calculate uncertainties for the buckets.

"Case P" is the pensioner who survives another year in j years with probability ${}_jp_{\alpha(\theta,t)}$ and therefore gets paid a pension worth $\mathrm{Pay}_u = b_P(\theta, t+j)$. In the case of death, with probability $1 - {}_jp_{\alpha(\theta,t)}$, a payment of $\mathrm{Pay}_d = 0$ is due.

"Case AP" is the active member who is older than the retirement age in j years. This member gets paid a pension worth $\mathrm{Pay}_u = \hat{b}(\theta, t+j)$ depending on the pension scheme (DB or DC). Further payment depends on projected earnings and wealth at retirement age and the probability of actually staying with the pension fund until pension, and then surviving the j years up to the current year, ${}_{\alpha_R}n_{\alpha(\theta,t)}\,{}_jp_{\alpha(\theta,t)}$.

"Case A" is the active member, still active in j years. This member gets a payment of $\mathrm{Pay}_d = \widehat{W}(\theta, t+j)$, the accumulated wealth in j years under the probability of staying in the fund for the last $j-1$ years (${}_{\alpha(\theta,t+j-1)}n_{\alpha(\theta,t)}$) and then leaving the pension fund in the jth year (with probability $to_{\alpha(\theta,t+j)}$).

6.4.1. *Uncertainties of obligations and liabilities to pensioners*

The general benefit payment $\hat{b}(\theta, t)$ shown in Table 2 is replaced by the payment for pensioners $b_P(\theta, t)$. The general form for the expected value in Eq. (22) yields the expected value for the obligations

$$\mathrm{E}\big[ob_P(\theta, t+j)\big] = {}_j p_{\alpha(\theta,t)} b_P(\theta, t+j). \tag{24}$$

We then insert the values of Table 2 into the general variance equation (23) for the variance for the pensioners obligations

$$\begin{aligned}\mathrm{var}\big[ob_P(\theta, t+j)\big] &= \big(b_P(\theta, t+j) - b(\theta, t+j)\, {}_j p_{\alpha(\theta,t)}\big)^2\, {}_j p_{\alpha(\theta,t)} \\ &\quad + \big(0 - b_P(\theta, t+j)\, {}_j p_{\alpha(\theta,t)}\big)^2 (1 - {}_j p_{\alpha(\theta,t)}) \\ &= {}_j p_{\alpha(\theta,t)} (1 - {}_j p_{\alpha(\theta,t)}) b_P(\theta, t+j)^2. \end{aligned} \tag{25}$$

For pensioners the obligations and liabilities are equal. Therefore, we can replace the $ob_P(\cdot)$ by $li_P(\cdot)$ for the variance of the liabilities as follows

$$\mathrm{var}\big[li_P(\theta, t+j)\big] = {}_j p_{\alpha(\theta,t)} (1 - {}_j p_{\alpha(\theta,t)}) b_P(\theta, t+j)^2. \tag{26}$$

6.4.2. *Uncertainties on obligations and liabilities to active members*

By plugging the values for the active members (Cases "*A*" and "*AP*") of Table 2 into the general variance formula (23) and replacing the general $\widehat{W}(\cdot)$ with $\widetilde{W}_{ob}(\cdot)$ and $\hat{b}(\cdot)$ with $\tilde{b}_A(\cdot)$, yields the expected value for "*A*"

$$\mathrm{E}\big[ob_A(\theta, t+j)\big] = {}_{\alpha(\theta,t+j-1)} n_{\alpha(\theta,t)} \widetilde{W}_{ob}(\theta, t+j) to_{\alpha(\theta,t+j)} \tag{27}$$

and for "*AP*"

$$\mathrm{E}\big[ob_{AP}(\theta, t+j)\big] = {}_{\alpha_R} n_{\alpha(\theta,t)} \tilde{b}_A(\theta, t+j)\, {}_j p_{\alpha(\theta,t)}. \tag{28}$$

The variances are then for "*A*"

$$\begin{aligned}&\mathrm{var}\big[ob_A(\theta, t+j)\big] \\ &\quad = {}_{\alpha(\theta,t+j-1)} n_{\alpha(\theta,t)} to_{\alpha(\theta,t+j)} (1 - {}_{\alpha(\theta,t+j-1)} n_{\alpha(\theta,t)} to_{\alpha(\theta,t+j)}) \\ &\qquad \times \big(\widetilde{W}_{ob}(\theta, t+j)\big)^2 \end{aligned} \tag{29}$$

and for "*AP*"

$$\mathrm{var}\big[ob_{AP}(\theta, t+j)\big] = {}_{\alpha_R} n_{\alpha(\theta,t)}\, {}_j p_{\alpha_R} (1 - {}_{\alpha_R} n_{\alpha(\theta,t)}\, {}_j p_{\alpha_R}) \big(\tilde{b}_A(\theta, t+j)\big)^2. \tag{30}$$

For the uncertainties on the liabilities we must note that the projected wealth $\widetilde{W}_{li}(\cdot)$ and projected pensions $\tilde{b}_A(\cdot)$ themselves are stochastic due to the model used in Section 3.5.1, where there is a stochastic factor for cost-of-living increases in salaries. The variance for the liabilities is

$$\begin{aligned}&\mathrm{var}\big[li_A(\theta, t+j)\big] \\ &\quad = {}_{\alpha(\theta,t+j-1)} n_{\alpha(\theta,t)} to_{\alpha(\theta,t+j)} (1 - {}_{\alpha(\theta,t+j-1)} n_{\alpha(\theta,t)} to_{\alpha(\theta,t+j)}) \\ &\qquad \times \big(\mathrm{E}\big[(\widetilde{W}_{li})\big]^2 + \mathrm{var}[\widetilde{W}_{li}]\big), \end{aligned} \tag{31}$$

where we use the short notation for projected wealth $\widetilde{W}_{li}$ meaning $\widetilde{W}_{li}(\theta, t+j)$. Similarly, "*AP*"

$$\begin{aligned}\operatorname{var}\big[li_A P(\theta, t+j)\big] \\ = {}_{\alpha_R}n_{\alpha(\theta,t)\,j}p_{\alpha_R}(1 - {}_{\alpha_R}n_{\alpha(\theta,t)\,j}p_{\alpha_R})\big(\mathrm{E}\big[(\tilde{b}_A)\big]^2 + \operatorname{var}[\tilde{b}_A]\big),\end{aligned} \tag{32}$$

$\tilde{b}_A(\theta, t+j)$ which depends on the pension fund (DB or DC). The values for the expected pensions $\tilde{b}_A(\cdot)$ can be calculated according to Eqs. (9) and (8).

6.4.3. *Uncertainties of obligations and liabilities buckets*

For the uncertainty of the buckets we sum up all individual uncertainties since every member is assumed to be an individual and not influenced by the other members. For the obligations buckets $OB(t)$ this results in

$$\begin{aligned}\operatorname{var}\big[OB_j(t)\big] = \sum_{\theta=1}^{\Theta(t)}\Big(&\operatorname{var}\big[ob_P(\theta, t+j)\big]1_P(\theta, t) + \operatorname{var}\big[ob_A(\theta, t+j)\big]1_A(\theta, t+j) \\ &+ \operatorname{var}\big[ob_A P(\theta, t+j)\big]1_{AP}(\theta, t+j)\Big).\end{aligned} \tag{33}$$

We use the indicator functions $1_P(\theta, t)$, $1_A(\theta, t+j)$ and $1_{AP}(\theta, t+j)$ as before, where $1_P(\theta, t)$ indicates that θ is a pensioner at time t, $1_A(\theta, t+j)$ indicates that θ is an active member in t and still so in $t+j$, and $1_A P(\theta, t+j)$ indicates that θ is an active member in t and has retired at $t+j$. The variances of $ob_P(\cdot)$, $ob_A(\cdot)$ and $ob_{AP}(\cdot)$ are in Eqs. (25), (29), and (30).

For uncertainties of the liabilities buckets $LB(t)$ we have the similar equation with the same indicator functions

$$\begin{aligned}\operatorname{var}\big[LB_j(t)\big] = \sum_{\theta=1}^{\Theta(t)}\Big(&\operatorname{var}\big[li_P(\theta, t+j)\big]1_P(\theta, t) + \operatorname{var}\big[li_A(\theta, t+j)\big]1_A(\theta, t+j) \\ &+ \operatorname{var}\big[li_A P(\theta, t+j)\big]1_{AP}(\theta, t+j)\Big),\end{aligned} \tag{34}$$

where the variances of $li_P(\cdot)$, $li_A(\cdot)$ and $li_{AP}(\cdot)$ are in Eqs. (26), (31), and (32).

6.4.4. *Scenario generation for obligation and liability buckets*

The future payments of the pension fund are stochastic since the fund faces several uncertainties that arise from its liability structure. The uncertainties arise from the risks individual members face as well as from macroeconomic factors such as wage inflation. To realistically evaluate future payments to members, the pension fund should know likely scenarios as well as worst-case and best-case scenarios. In terms of the bucket structure, obligation scenarios are $OB^s_j(t)$ and the liability scenarios are $LB^s_j(t)$. Based on the expectations and variances for each individual bucket, we are able to simulate different scenarios. Since the liabilities are the summation of the individual's contribution to the aggregated risk, we use the normal distribution to generate the scenarios. Alternatively, we can simulate all the different binomial random variables and aggregate them into the buckets. When the pension fund is large enough that the central limit theorem applies, the simulation based on the normal distribution is sufficient to generate statistically significant scenarios.

6.5. Total and discounted liabilities and obligations and current coverage ratios

The bucket structure for obligations and liabilities represents the actual, non-discounted values of the benefits which are due at given times in the future. By summing up all the buckets, we can calculate the total expected liabilities and obligations. We can also use the bucket structure to calculate the present value of the liabilities and obligations by discounting the buckets with the appropriate discounting factor. The present value of the buckets can be used to specify the current coverage ratio by comparing the present value with the present value of the current wealth.

6.5.1. Total expected obligations and liabilities to pensioners

The total (non-discounted) expected remaining pension outflow to a pensioner $B_P(\theta, t)$ is

$$\mathrm{E}\big[B_P(\theta, t)\big] = \sum_{i=1}^{\alpha_{\max}-\alpha(\theta,t)+1} {}_i p_{\alpha(\theta,t)} b_P(\theta, t+i). \tag{35}$$

This is the sum of every year's payment considering mortality. Values for p_α in life tables are usually set to zero for very high ages (above a given $\alpha_{\max}$). Since obligations and liabilities for pensioners are equal, the expected remaining pension outflow for pensioners is also equal for obligations and liabilities arising from these payments.

6.5.2. Total expected obligations and liabilities to active members

For active members the total expected pension benefit outflow $\widetilde{B}_A(\theta, t)$ is

$$\begin{aligned}\mathrm{E}\big[\widetilde{B}_A(\theta, t)\big] = {}_{\alpha_R} n_{\alpha(\theta,t)} &\sum_{i=1}^{\alpha_{\max}-\alpha_R} {}_i p_{\alpha_R} \hat{b}_A(\theta, t+i-1) \\ &+ \sum_{j=1}^{\alpha_R-\alpha(\theta,t)} {}_{\alpha(t+j-1)} n_{\alpha(\theta,t)} \widehat{W}(\theta, t+j) to_{\alpha(\theta,t+j)}.\end{aligned} \tag{36}$$

Here we sum up the yearly benefit payments after retirement $\hat{b}_A$ with the uncertainty of mortality after pension ${}_i p_{\alpha_R}$, while taking into account the possibility that the active member will not reach retirement age as member of the pension fund (${}_{\alpha_R} n_{\alpha(\theta,t)}$). We also add the expected payments before retirement due to labour mobility (i.e., the probability of being in the fund until the year before the jth year $({}_{\alpha(t+j-1)}n_{\alpha(\theta,t)})$ multiplied with the projected wealth for the jth year $(\widetilde{W}_{li}(\theta, t+j))$ multiplied with the probability of leaving during the jth year $(to_{\alpha(\theta,t+j)})$. From Eq. (36) both values for obligations and liabilities can be derived by using the general forms for benefits $\hat{b}_A(\cdot)$ and $\widehat{W}(\cdot)$ for wealth, respectively. For benefit payments $\hat{b}_A(\cdot)$ is $\hat{b}_A(\cdot) = \lambda \widehat{W}^{(\alpha_R)}(\cdot)$ for the DC pension plan and $\hat{b}_A(\cdot) = \kappa \widetilde{S}^{(\alpha_R)}(\cdot)$ for the DB plan.

6.5.3. Discounted obligations and liabilities buckets

Since the obligations and liabilities buckets have a term structure they can be discounted with the appropriate discount-factor and added up, which leads to the present value of obligations or liabilities. For an obligations bucket $OB_j(t)$ the present value is

$$PVOB_j(t) = \mathrm{E}\big[OB_j(t)\big]\big(1 + r_d^{(j)}\big)^{-j}, \tag{37}$$

where $r_d^{(j)}$ is the interest rate used for discounting over j years. The present value of a liabilities bucket $LB_j(t)$ is

$$PVLB_j(t) = \mathrm{E}\big[LB_j(t)\big]\big(1 + r_d^{(j)}\big)^{-j}. \tag{38}$$

The present value of the pension fund's total obligations then is given by the sum over all discounted buckets

$$PVOB_{\Sigma}(t) = \sum_{i=1}^{\alpha_{\max}-25} \mathrm{E}\big[OB_j(t)\big]\frac{1}{(1 + r_d^{(j)})^j}, \tag{39}$$

where the number of buckets over which we sum up is given by the maximum reachable age minus the age of the youngest member (here 25 years) and $(1 + r_d^{(j)})^{-j}$ is the discounting factor for the bucket in j years. As an example, $r_d^{(j)}$ could be given by the yield curve of government bond interest rates with various maturities. Similarly, for liabilities

$$PVLB_{\Sigma}(t) = \sum_{i=1}^{\alpha_{\max}-25} \mathrm{E}\big[LB_j(t)\big]\frac{1}{(1 + r_d^{(j)})^j}, \tag{40}$$

where the only difference is that we sum up the discounted liability buckets instead of obligations buckets.

6.5.4. Current coverage ratio

An important measure for the pension funds financial health is the coverage ratio. It is the present value of assets divided by the present value of liabilities. The coverage ratio is also found in the literature by the name of funding ratio (i.e., Rudolf and Ziemba

(2004)). When the coverage ratio falls below one, there is currently not enough wealth in the pension fund to pay for the future liabilities.

The coverage ratio for the obligations is $CR_{\Sigma,ob}(t)$. Since obligations consist only of the payments due, based on compounding current wealth, we can write for the coverage ratio

$$CR_{\Sigma,ob}(t) = \frac{W(t)}{PVOB_{\Sigma}(t)}, \tag{41}$$

which is the ratio of current wealth to the present value of obligations. For liabilities, we need to take into account the present value of all future contributions into the pension fund, such that the coverage ratio for liabilities $CR_{\Sigma,li}(t)$ becomes

$$CR_{\Sigma,li}(t) = \frac{W(t) + pvc(t)}{PVLB_{\Sigma}(t)}, \tag{42}$$

where $pvc(t)$ is the present value of all future expected contributions. The two coverage ratios must be equal.

7. The information value of the bucket structure

7.1. Information value of the life-insurance model

Based on this “life-insurance” model of a pension fund, we are able to analyse the relevant questions about the liabilities. For each individual member, we know the probability of exit from the pension fund. This helps us to answer questions such as mean time of membership for any given age and sex. This knowledge is crucial for the management of a fund, since we want to match the duration of the assets and liabilities. Furthermore, this knowledge allows the pension fund to earmark the wealth of each individual. Given the wealth of individual members and their exit probabilities, the obligation buckets indicate when and how much cash-flow can be expected from the current wealth. The description of uncertainty, i.e., the variances of the buckets, allows us to model scenarios for the obligations. For instance, active members might live longer than expected, which increases our liabilities.

Moreover, the different scenarios for inflation and wage levels help the pension funds to assess future contributions and the resulting liabilities. This feature, combined with the probabilistic description of the time of membership and exit from the fund, allows the pension fund to compute the expected cash-flows due to obligations and liabilities. The cash-flow description includes all major uncertainties that an individual member faces and which trigger a cash-flow from the fund. By assuming that all active members will stay in the fund until they reach the pension age, major cash-flows due to labour mobility and invalidity are not projected and the expected time of membership is overestimated. This might lead to inadequate investment decisions, since we assume longer maturity and a smaller intermediate outflow of funds than necessary.

Since we know the probabilities of the exit times, we are able to earmark the wealth and the contributions to the liability buckets. In this way, the pension fund's wealth can be exactly matched to its liabilities. This in turn allows the fund to be managed according to its expected maturity. Furthermore, a sensible liquidity planning is another result, since provisions are taken for cash-flows due to labour mobility and invalidity. The idea of earmarking is used in Sections 8.5 and 8.6 as a central aspect of the pension fund optimisation.

7.2. *Transparency and health of the pension plan*

The earmarking of the individual wealth and of future contributions allows us to compute the coverage ratio for each bucket. Instead of working with a lump-sum coverage ratio, we can compute a term structure (maturity structure) of coverage ratios. This allows us to judge much better the financial "health" of the pension fund and indicates possible (unwanted) subsidies from one group to another. A lump-sum coverage ratio slightly below 100%, e.g., 95%, shows an under-funding of current obligations. When the under-funding is mostly concentrated on bucket coverage ratios with a long maturity, in which a higher investment into risky assets is possible, while the coverage ratios of the short maturities are above 100%, the under-funding must not be critical. In this situation, an increase in contributions is probably unnecessary. However, coverage ratios under 100% for buckets with short maturities indicate a clear financial stress situation. The bucket coverage ratio improves the assessment of the financial situation and allows for measured responses in a shortfall situation. The earmarking of current wealth and contributions allows the pension fund to detect possible cross-subsidies, where one group of members' contributions is used to finance the liabilities (obligations) for another group of pension members. The method helps detecting such situations and shows the extent of such subsidies. The pension fund may also analyse how to correct such situations and evaluate their cost.

7.3. *Utilisation of the information*

The precise modelling of expected payments helps the pension fund simulate future scenarios and analyse its current situation. We can compute the influence of different fund structures, such as the guaranteed interest rate or studying differences between a DB plan or a DC plan. The precise modelling also increases the transparency of the fund structure and the utilisation of current pension fund wealth and contributions. Moreover, different contribution schemes, which may depend, for example, on the coverage ratio or the current underlying economic situation, can be evaluated. Scenarios of future trajectories of assets and liabilities may serve as inputs to optimise the asset allocation. The bucket structure simulation together with a simulation of the asset allocation, supports the pension fund management in determining critical scenarios, such as high labour mobility while asset returns are falling. In this way, concentrations of hidden risks can be detected and addressed early. The term structure of obligations allows us to use classical asset liability techniques such as duration matching and immunisation.

7.4. Pooling of pension funds

Another important question this method helps to analyse is the pooling of different pension funds. When two or more pension funds want to pool their operations, we can judge the situation much better with the life-insurance modelling technique. We can assess the obligations and liability buckets which arise from the pooling and compare this to the individual pension fund's situation. This helps to determine for which fund such pooling is beneficial and for which it is not. For example, a pension fund with low labour mobility combined with a fund with high labour mobility and a similar distribution of members ages, might result in a more manageable pension fund, since for one fund it reduces the duration of liabilities and for the other it increases it. Additionally, the cash-flows become smoother over time and less volatile for each given period.

7.5. Cost vs. benefits of the model

This modelling approach of pension fund obligations and liabilities is costly in the sense that considerable data is needed for the analysis. However, the insights gained by this type of model justifies the additional work needed. Especially in long-term optimisation and analysis, a precise description of the liabilities ensures applicable and robust decisions.

8. Asset–liability management optimisation

8.1. Introduction

The globalisation of financial markets and the introduction of various new and complex products, such as options or other structured products, have significantly increased the volatility and risk for participants in the markets. Moreover, advances in communication technology and computers have dramatically increased the reaction speed of financial markets to world events. This has occurred within Switzerland, the country in which the fund resides, as well as across markets internationally. The long-term nature of a pension fund amplifies the financial rewards for good decisions as well as the penalties for bad decisions. Furthermore, the dynamic and uncertain nature of both the asset and the liability trajectories greatly complicates the investment problem. Therefore, the need to integrate the liability and the asset management has dramatically increased.

In recent years, a growing number of applications of integrated risk management have emerged. Insurance companies and pension funds pioneered these applications, which include the Russell–Yasuda investment system (Cariño et al., 1994), the Towers Perrin System (Mulvey, 1995), and the Siemens Austria Pension Fund (Ziemba (2003) and Geyer et al. (2004)). In each of the applications, the investment decisions are linked with liability choices, and the system's funds are maximised over time using multi-stage stochastic programming. The integrated risk management approach is therefore the best suited way of managing a pension fund. It includes the dynamics of the assets, the dynamics of the liabilities, the long-term nature of the pension fund, and the uncertainty faced by the fund.

Most financial planning systems today still rely on the classical mean-variance framework pioneered over 50 years ago. Despite its huge success, the single-period setting possesses some significant deficiencies. First, it is difficult to use in a long-term application where investors are able to rebalance their portfolio frequently. Second, for situations where investors face liabilities or goals at specific future dates, the investment decisions must be taken with regard to the dynamics and time structure involved. The multi-period approach may also provide superior performance over the single-period approach, see Dantzig and Infanger (1993). Third, the definition of risk, such as variance or semi-variance, does not transfer any information regarding the chances of matching the obligations or goals. Furthermore, variance or semi-variance are not coherent risk measures in the sense of Artzner et al. (1999). Fourth, the mean variance framework is extremely sensitive to the model inputs, i.e., mean values and covariances. Fifth, the mean variance framework cannot easily handle issues such as taxes and transaction costs.

Nevertheless, economic growth theory recommends that a multi-period investor should maximise the expected logarithmic wealth at each time period, as suggested by Luenberger (1998). However, it has been shown in Rudolf and Ziemba (2004) that a logarithmic utility may lead to a too high risk tolerance. In addition, the theory depends on various assumptions such as no transaction cost, i.i.d. asset returns, and neither liabilities nor in- or outflows to be time-dependent. When these assumptions are violated, as they are in the case of pension funds, a multi-period setting is the appropriate framework to handle such a problem.

For all these stated reasons, it is necessary to use a dynamic multi-period optimisation rather than the classical single-period framework.

8.2. Multi-period asset model

Many different formulations of multi-period investment problems can be found in the literature, see Ziemba and Mulvey (1998), Kall and Wallace (1994), Kusy and Ziemba (1986), or Louveaux and Birge (1997). We adopt the basic model formulation presented in Mulvey and Simsek (2002) and Mulvey and Shetty (2004), with various modifications. The asset and liability management horizon consists of τ time steps represented by $T = \{t, t+1, t+2, \dots, t+\tau\}$, where t is the current time and $t+\tau$ is the planning horizon. At every time step, the pension fund is able to make a decision regarding its investments and faces the inflow of funds due to contributions and outflows due to obligations.

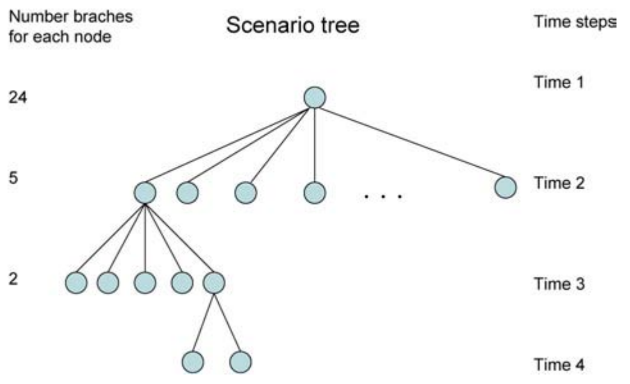


Fig. 7. Graphical description of a scenario tree.

The investment classes are defined as the set $I = \{1, 2, \ldots, N\}$. This set should reflect all important asset classes, such as stocks (large, small, international, emerging markets), bonds, cash, and real estate. The asset classes chosen for the optimisation should reflect important market segments and should be available as investable security, such as index funds or future contracts. Examples are the Swiss Performance Index, which tracks the largest 100 Swiss stocks or the S&P 500 in the US.

The uncertainties faced by the pension fund, either from the investments or from the liabilities, are modelled by the so-called scenario approach. By utilising a sufficiently large number of scenarios, we are able to represent all of the random effects the pension fund faces. The scenarios are defined as the set S that represents a reasonable description of the future uncertainties. A scenario $s \in S$ describes a unique path through consecutive nodes of the scenario tree as depicted in Figure 7. Many mathematical techniques exist to generate scenarios. Most authors use various econometric methods to describe future asset returns. Examples in the case of asset liability management applications include Dert (1998), Boender (1997), Boender, van der Aalst and Heemskerk (1998), Koivu, Pennanen and Ranne (2005), and Wilkie (1995). An important aspect of the scenario modelling is its forward-looking and dynamic character. Instead of analysing historical returns, such as mean and covariances, we can build forward-looking models which evaluate the pension fund's situation. The idea is to construct sets of scenarios that represent the pension fund's unique situations, such as possible assets, cost, liabilities, and legal requirements.

Let $x_i^s(t+k)$ be the amount of wealth invested in instrument i at the beginning of the time step $t+k$ under scenario s. The units used are the pension fund's home currency (Swiss Francs). Foreign assets, either hedged or unhedged, are also denoted in the fund's home currency. At time $t+k$ the total wealth of the fund is

$$\Omega^s(t+k) = \sum_{i=1}^{N} x_i^s(t+k), \quad \forall s \in S, \tag{43}$$

where $\Omega^s(t+k)$ denotes the total wealth under scenario s. Given the returns of each investment class, the asset values at the end of the time period are

$$x_i^s(t+k)\bigl(1+\tilde{r}_i^s(t+k)\bigr) = y_i^s(t+k), \quad \forall s \in S,\ \forall i \in I, \tag{44}$$

where $\tilde{r}_i^s(t+k)$ is the return of investment class i at time $t+k$ under scenario s. The returns are obtained from the scenario generation system. Therefore, $y_i^s(t+k)$ is the ith asset value at the end of the time period $t+k$ under scenario s. For stocks, assume that dividends are part of the return and that they are automatically reinvested. The sales or purchases of assets occur at the beginning of the time period, where $d_i^s(t+k) \geqslant 0$ denotes amount of asset i sold at time $t+k$ under scenario s, and $p_i^s(t+k) \geqslant 0$ denotes the purchase of asset i at time $t+k$ under scenario s. The asset balance equation for each asset is

$$\begin{aligned} x_i^s(t+k) &= y_i^s(t+k-1) + p_i^s(t+k)(1-\delta_i) - d_i^s(t+k), \\ &\forall s \in S,\ \forall i \in I \setminus \{1\}, \end{aligned} \tag{45}$$

where δ_i is the proportional transaction cost of asset i. We make the assumption that the transaction costs are not a function of time, but depend only on the investment class involved. This assumption can be easily relaxed to include a time dependency. We treat the cash component of our investments as a special asset. The balance equation for cash is

$$\begin{aligned} x_1^s(t+k) = {}& y_1^s(t+k-1) + \sum_{i=2}^{N} d_i^s(t+k)(1-\delta_i) \\ & - \sum_{i=2}^{N} p_i^s(t+k) + c^s(t+k) - o^s(t+k), \quad \forall s \in S, \end{aligned} \tag{46}$$

where $x_1^s(t+k)$ is the cash account at time $t+k$ under scenario s, $c^s(t+k)$ and $o^s(t+k)$ are the contributions received by the fund and the outflows due to payments by the fund at time $t+k$ under scenario s, respectively. The cash account equals the interest rate earned from the cash account's value of the last period, plus all money earned from sales of assets, minus all money used for the purchase of assets, plus the contributions minus the payments due to obligations of the fund. Furthermore, we restrict all assets, including the cash account, to be non-negative, i.e., $x_i^s(t+k) \geqslant 0$. This means that we do not allow for any borrowing. This restriction may be dropped in other situations.

All variables in Eqs. (43)–(46) are dependent on the actual scenario s. These equations could be decomposed into subproblems for each scenario where we anticipate which scenario will evolve. To model reality, we must, however, impose non-anticipativity constraints. All scenarios which inherit the same past up to a certain time period must evoke the same decisions in that time period, otherwise the non-anticipativity requirement would be violated. So, $x_i^s(t+k) = x_i^{s'}(t+k)$ when s and s' have same past until time $t+k$.

Another way of imposing the non-anticipativity requirement is to use control policies which do not depend on the future path of the assets and liabilities. Such a policy is the fixed-mix strategy, where we require that a fixed proportion of the wealth is invested in a certain asset (Ziemba (2003)). Alternatively, we could also allow that the mix is time-varying, but still independent of the scenarios. This results in a dynamic-mix (or time-varying mix) strategy. The fixed-mix strategy has a thorough theoretical underpinning in the case where the asset returns show no inter-temporal dependency. This is a well-known result from continuous-time finance and stochastic control theory as derived by Merton (1969, 1973). When asset returns exhibit inter-temporal dependencies, which is the case when asset returns are described by factor models, the constant-mix is not the optimal asset allocation strategy. For this case, Campbell, Rodriguez and Viceira (2004), and Herzog et al. (2004) show, by employing optimal stochastic control theory, that a dynamic-mix is the optimal asset allocation strategy. The fraction of wealth at time $t+k$ invested in the ith asset is

$$u_i(t+k) = \frac{x_i^s(t+k)}{\Omega^s(t+k)}, \quad s \in S. \tag{47}$$

For the fixed-mix strategy, $u_i(t+k) = u_i$, since it does not depend on time. The mix strategies reduce the number of decision variables to a large extent, but they introduce nonlinearity into the problem. To analyse Eq. (47), we use Eqs. (43)–(46) and obtain

$$\begin{aligned} x_j^s(t+k) &= y_j^s(t+k-1) + p_j^s(t+k)(1-\delta_j) - d_j^s(t+k) \\ &= u_j(t+k)\left(\sum_{i=1}^{N}\left\{y_i^s(t+k-1)\right\} + c^s(t+k) - o^s(t+k)\right. \\ &\quad \left. - \sum_{i=2}^{N}\left\{\delta_i\left(p_i^s(t+k) + d_i^s(t+k)\right)\right\}\right), \quad j \in I \setminus \{1\},\ s \in S. \end{aligned} \tag{48}$$

The asset sales and purchases under each scenario and at any time can be computed in this manner. The wealth at time $t+k$ before rebalancing minus the net transaction costs is divided among the N assets according to the mix rule $u_j(t+k)$. The various nonlinearities in Eq. (48) are visible such as the term $u_j(t+k)(\sum_{i=2}^{N}\delta_i(p_i^s(t+k) + d_i^s(t+k)))$.

When we neglect transaction costs and use the mix rules, the evolution of the wealth can be computed from

$$\begin{aligned} \Omega^s(t+k) &= \sum_{i=1}^{N}\left\{\left(1+\tilde{r}_i(t+k-1)\right)u_i(t+k-1)\right\}\Omega^s(t+k-1) \\ &\quad + c^s(t+k-1) - o^s(t+k-1), \quad s \in S. \end{aligned} \tag{49}$$

The system's N balance equations and constraints are reduced to a one-dimensional equation and we do not need to compute the sales and purchases for each scenario and time in order to know the evolution of the fund's wealth.

8.3. Optimisation objectives for the pension fund asset–liability management

We outline the general objectives of the fund and how they can be translated into objective functions and constraints. Since we deal with decisions under uncertainty, the definition of risk for the pension fund is crucial for the appropriate asset liability management optimisation. From the regulatory point of view in Switzerland, the health of a pension fund is judged by the criteria known as (current) coverage ratio. The interest rate for discounting the liabilities set by the regulators is called technical interest rate. The coverage ratio is required to be above 100% and various measures have to be taken in case it falls below 100%. These measures range from changes in the investment strategy to increases in contributions from both employer and employees up to in the worst case, the termination of the fund.

Many coherent risk measures have been found, such as maximum loss or conditional Value-at-Risk (CVaR). The risk measure used for asset liability management applications, however, must be tailored to the specific situation and the specific objectives of the fund. Standard coherent risk measures, such as CVaR, can be used in ALM situations, when applied to the fund's net wealth, e.g., the sum of the assets minus all the remaining liabilities (obligations). CVaR penalises linearly all events which are below the VaR limit for a given confidence level. The inherent VaR limit is a result of the CVaR optimisation, see Rockafellar and Uryasev (2000) and Rockafellar and Uryasev (2002). The VaR limit therefore depends on the confidence level chosen and the shape of the distribution. The VaR limit (quantile) may be a negative number, i.e., a negative net wealth result. In the situation of a pension fund, we do not only want to penalise scenarios that are smaller than a given quantile, but all scenarios where the net wealth is non-positive. The most important objective of the pension fund is to meet its obligations and achieve a positive net wealth. For these reasons, we define risk as a penalty function for the net wealth. The net wealth can be either measured in absolute terms, e.g., Swiss Francs, or in relative terms, i.e., as a percentage of the obligations. Second, we want to penalise small "non-achievement" of obligations differently from large "non-achievement". Therefore, the penalty function should have an increasing slope with increasing "non-achievement". Third, the penalty function should be a convex function. The risk of the pension plan is measured as one-sided downside risk based on non-achievement of the obligations.

The expectation is a convexity preserving operation. The penalty function is not a coherent risk measure. However, the conditional expectation of the non-positive net wealth is coherent and coincides with the penalty function where only one slope with an incline of one is used. The reason why the general penalty function is not coherent is that it is not a monetary measure of potential losses, but rather a non-monetary (artificial) number of non-achievement. The advantage of the penalty function approach is that large losses are more heavily penalised than smaller losses. For optimisation purposes, we can use the penalty function for the terminal net wealth of the fund or compute the penalty function for each future time. A practical method is to define a non-negative weighted sum of the penalty functions for all time. In this way we control the risk of non-achievement in a multi-period setting.

The second objective of the fund is to achieve a surplus from the funds under management. For this reason, we want to maximise the expected net wealth. We either formulate this goal as a terminal date objective, or again, as a weighted sum over all time periods.

8.4. Optimisation approaches for multi-period ALM models

For the mathematical model of asset and liability dynamics described in Section 8.2, many different optimisation methods exist. Here, we will briefly describe two different approaches, namely the stochastic programming approach and the dynamic stochastic control approach. Other possible optimisation techniques, such as dynamic programming, are not discussed here. For our pension fund application we briefly describe a system which sets up an account for each liability which is funded, in order to cover the associated obligation. Each account is individually managed and funded.

8.4.1. Stochastic programming approach

The stochastic programming approach finds the optimal sales variables $p_i^s(t+k)$ and the optimal purchase variables $d_i^s(t+k)$ given the current time and scenario including the non-anticipativity constraints. By inspecting Eqs. (43)–(46), we see that these equations are linear. Furthermore, when we use a linear objective function, then the optimisation problem becomes a large-scale linear programming (LP) problem. The objective function is

$$\max \sum_{s=1}^{S} \bigl\{ p_s \bigl(\Omega^s(t+\tau) - ROB^s(t+\tau) \bigr) - \gamma p_s PF^s(t+\tau) \bigr\}, \tag{50}$$

where the $ROB^s(t+\tau) = \sum_{j=\tau+1}^{J} OB_j(t)$ represents all remaining obligations at the end of the planning horizon under scenario s, $PF^s(t+\tau)$ denotes the piecewise linear penalty function, and p_s denotes the probability of scenario s. Alternatively, we can formulate the objective function based on the liability buckets. The penalty function is also applied to net wealth, i.e., $\Omega^s(t+\tau) - ROB^s(t+\tau)$. The objective (50) is a piecewise linear function of the decision variables. Many specialised algorithms exist to solve this special form LP, such as the L-shaped algorithm, see van Slyke and Wets (1969), Kall and Wallace (1994), Birge et al. (1994), and Birge and Holmes (1992). Other approaches, known not to work satisfyingly, are known as the progressive hedging algorithm, see Rockafellar and Wets (1991), Berger, Mulvey and Ruszczynski (1994), and Barro and Canestrelli (2005).

The advantage of the stochastic programming approach over the dynamic stochastic control approach is that the decisions are computed similarly to a feedback control system. Given the new realisations, i.e., when we know the current state of our system, we have a prescribed rule of actions that helps us to achieve the objectives. The decision rule observes the actual path our system has taken in the past and uses the conditional information of the current state for the future uncertainty. Furthermore, under different scenarios, the fraction of wealth invested in the different asset classes differs, which sets this apart from the fixed-mix approach.

The main disadvantage is that for systems with many future time periods, the number of scenarios $S = \prod_{j=1}^{T} n_j$, where n_j is the number of branches at each time step j, increase exponentially. Furthermore, developing a representative tree of scenarios is a challenging research field. There are several objectives to meet when building a stochastic scenario tree. The scenarios must be based on sound and realistic economic principles. Basic observations from econometrics should be included, such as volatility clustering in equity markets or mean reversion of interest rates. Projections of future scenarios should be investigated with regard to how well they fit past observations. The main feature of scenario generation is to reflect a universe of possible outcomes. This must include likely scenarios, as well as best and worst case scenarios.

8.4.2. *Dynamic stochastic control approach*

The dynamic control approach works by using control policies that are independent of the current and future scenarios. The basic decision variables are fixed-mix or dynamic-mix variables. The main advantage of the stochastic control policies is that they reduce the size of the problem drastically and that they are well understood by investment professionals. Furthermore, theoretical results from continuous-time finance support this approach, see Merton (1992). By allowing dynamic-mix strategies, i.e., the mix is not constant throughout the planning horizon, we can also introduce feedback into the system. This feedback depends on the time dependencies of the asset returns.

The main disadvantage of the stochastic control strategies is that they introduce nonlinearities into the optimisation problem regardless of the functional form of the objective function. Very often we are also faced with the situation of a non-convex nonlinear optimisation, which requires specialised global solution algorithms, see Maranas et al. (1997) and Konno and Wijayanayake (2002), or other global optimisation techniques, such as genetic algorithms.

8.4.3. *Application to the problem of a pension fund*

The pension fund investment problem is marked by its long-term nature and its term structure of payments due to pensions and other obligations. In Section 8.3 we briefly outlined the fund's ALM objectives.

To match all the goals for each obligation bucket (or liability bucket), a corresponding account is set up. The account consists of several assets which are to provide the funds needed to pay the obligation of that bucket. The accounts have different terminal dates depending on maturity of the associated bucket. Due to the actuarial treatment of the various risks, we know the distribution of possible exit times of members and, therefore, are able to earmark the contributions to the different liability buckets.

Furthermore, each individual account has its own investment policy which depends not only on the risk-aversion but also on the investment time horizon. Using stochastic control theory, Brennan, Schwartz and Lagnado (1997) and Campbell and Viceira (2002) have shown that the investment policy depends both on risk aversion and time horizon. For each account, a part of the current wealth is earmarked. The decision variables are therefore, the earmarking of funds for the accounts and the investment policy of the accounts. To optimise the entire fund and not only each individual account, the objective function is the superposition of surplus wealth of all of the accounts and the superposition of penalty functions of all of the accounts. By using the accounts, we can translate the preferences of the individual pension fund members into the goals and objectives of the pension fund. The life-style hypothesis, see Cairns, Blake and Dowd (2003), is thus realised across the pension fund. Moreover, in this way we can control the coverage ratio for each individual bucket. The coverage ratio for the entire fund is not as meaningful as the bucket coverage ratio together with the bucket term structure.

Alternatively, we could set up an asset liability management optimisation for the entire wealth of the fund. From the liability model and the bucket structure, we have a stochastic model for the time structure of contributions and payments due to liabilities. The optimisation then determines an investment policy for the entire fund. The fund's objective in terms of risk and surplus can similarly be incorporated. However, using the accounts and buckets better matches the individuals' preferences with the overall management of the pension fund. Furthermore, the earmarking of funds is an important decision variable which determines the health of the fund, i.e., coverage ratios for buckets provide much clearer information on possible shortfalls and on the time horizon in which they might occur than the lump-sum coverage ratio.

In Section 8.5, we propose an optimisation for the asset obligation management which does not take transaction cost into account and uses relative measures for the surplus and the risk involved. This solution focuses on the management of current funds (wealth) and the obligations that arises therefrom. Since we do not consider any future contributions, the pension fund's wealth must be invested to match the obligations. In Section 8.6, we work with absolute measures and include transaction costs. We also include future contributions and therefore match the current funds and contributions with the resulting liabilities. Both optimisations use the idea of buckets, accounts, and earmarking. The optimisation approach is the dynamic stochastic control approach.

8.5. Optimisation of asset obligation management with earmarking and investment policy without transaction costs

We describe an optimisation for the pension fund's asset obligation management based on the projected obligations and current wealth of the pension fund. For each obligation bucket, we set up an account to manage the funds earmarked for the specific obligation. We assume that transaction costs and market impact can be neglected.

8.5.1. Projected account value

From the life insurance model of the pension fund the future guaranteed payments for each obligation bucket are known. To meet these obligations, the pension fund assigns funds from its current wealth and determines an investment policy. All accounts are denoted by the set J. For each bucket we allocate a part of the current wealth using

$$\Omega_j(t) = v_j(t)\Omega(t), \tag{51}$$

where $\Omega(t)$ is the current wealth, $\Omega_j(t)$ is the wealth allocated to cover the obligations in the jth bucket, and $v_j(t)$ is a fraction of the current wealth. The following constraints apply

$$\sum_{j=1}^{J} v_j(t) = 1, \quad v_j(t) > 0. \tag{52}$$

For $v_j(t) = 0$ we take the jth bucket out of consideration since there are no liabilities to be covered for this bucket. The fraction of wealth assigned to the jth bucket should be invested so that the guaranteed future obligations are matched or even exceeded. The projected account value assigned to jth bucket is

$$\Omega_j^s(t+m_j) = \prod_{i=1}^{m_j} \left\{ \sum_{l=1}^{N} \big(1+\tilde{r}_l(t+i)\big) u_{jl}(t) \right\} \Omega_j(t), \tag{53}$$

where $M_j^s(t+m_j)$ is the account for the jth bucket and m_j is the number of rebalancing periods until the account matures at time $t+j$. The returns of the individual assets at time $t+i$ under scenario s are $\tilde{r}_l^s(t+i)$. Assume that there exists a finite number of scenarios $s = 1, \ldots, S$. Then $u_{jl}(t)$ is the fraction of bucket wealth invested into the lth asset. Assume that the investments into the N assets are rebalanced at every rebalancing period. To write (53) more concisely, we stack all returns plus one in the vector $r^s(t+i) = (1+\tilde{r}_1^s(t+i), \ldots, 1+\tilde{r}_N^s(t+i))^{\mathrm{T}}$ and write $u_j(t) = (u_{j1}(t), \ldots, u_{jN}(t))^{\mathrm{T}}$. Eq. (53) becomes

$$\Omega_j^s(t+m_j) = \prod_{i=1}^{m_j} \left\{ \big(r^s(t+i)\big)^{\mathrm{T}} u_j(t) \right\} \Omega_j(t). \tag{54}$$

We assume that no transaction cost or taxes apply, which is of course an approximation of reality. However, there are investable securities with minimal brokerage fees, such as futures on indices. Furthermore, we tacitly assume that our trading impact is small enough that we do not influence the asset prices.

The logarithm of both sides of (54) yields

$$\begin{aligned} ML_j^s(t+m_j) &= \ln\big(\Omega_j^s(t+m_j)\big) \\ &= \sum_{i=1}^{m_j} \left\{ \ln\Big(\big(r^s(t+i)\big)^{\mathrm{T}} u_j(t)\Big) \right\} + \ln v_j(t) + \ln \Omega(t), \end{aligned} \tag{55}$$

where $ML_j^s(t+m_j)$ is the logarithm of $\Omega_j^s(t+m_j)$. The decision variables for the jth account are current investment policy $u_j(t)$ and the earmarking of funds $v_j(t)$.

Theorem 1. *The logarithmic value of the jth account $ML_j^s(t+m_j)$ is concave with respect to the decision variables $u_j(t)$ and $v_j(t)$.*

Proof. By straight forward calculation verify that the Hessian of Eq. (55) is negative semi-definite. □

8.5.2. Optimisation problem

The aim of the fund is to meet its obligations with high certainty while attempting to generate a surplus. Therefore, we define risk as the possibility that the account value does not cover the obligations. The coverage ratio at the time of the maturity of the bucket is

$$CR_j^s(t+m_j) = \frac{\Omega_j^s(t+m_j)}{OB_j^s(t)}, \tag{56}$$

where $OB_j^s(t)$ denotes the future obligations in bucket j under scenario s and $CR_j^s(t+m_j)$ is the coverage ratio at maturity. Taking the logarithm of (56) yields

$$\begin{aligned} cr_j^s(t+m_j) &= \ln\bigl(\Omega_j^s(t+m_j)\bigr) - \ln\bigl(OB_j^s(t)\bigr) \\ &= ML_j^s(t+m_j) - \ln\bigl(OB_j^s(t)\bigr). \end{aligned} \tag{57}$$

The log-coverage ratio is larger than zero if the account's fund is larger than the earmarked obligations and smaller than zero if the fund's wealth does not cover the obligations. If the account value does not cover the obligations, the log-coverage ratio is a negative number.

The number is similar to a total percentage number of either over- or underachievement minus 100%, e.g., a coverage ratio of 120% results in a log-coverage ratio number of 18.23%. The log-coverage-ratio actually slightly underestimates coverage ratios above 100% and overestimates coverage ratios below 100%, e.g., a coverage ratio of 80% results in a log-coverage-ratio number of −22.8%. In this manner, the measure errs on the conservative side, since it underestimates surpluses and overestimates shortfalls. When we divide the log-coverage number by the investment periods of this account, we obtain a compounded return number.

We define the aggregated weighted shortfall under scenario s as the non-negative summation of all individual weighted negative log-coverage ratios for each bucket, or mathematically as

$$\begin{aligned} SF^s(t) &= \sum_{j=1}^{J} \max\big(-cr_j(t), 0\big) f_j^{SF} \\ &= \sum_{j=1}^{J} f_j^{SF} \max\Big(\ln\big(OB_j^s(t)\big) - ML_j^s(t+m_j), 0\Big). \end{aligned} \tag{58}$$

Here $f_j^{SF} \geqslant 0$ is a weighting factor for each bucket. The risk of the pension plan is measured as a one-sided downside risk based on non-achievement of the obligations. As penalty function we choose piecewise linear function

$$PF^s(t) = \begin{cases} 0 & \text{if } SF^s(t) = 0, \\ p_1 SF^s(t) & \text{if } 0 \leqslant SF^s(t) \leqslant c_1, \\ p_2(SF^s(t) - c_1) + p_1 c_1 & \text{if } SF^s(t) \geqslant c_1. \end{cases} \tag{59}$$

Here $p_1 > 0$ and $p_2 > 0$ are slopes of the penalty function and $c_1 > 0$ is the point where the penalty increases more steeply, as discussed in Section 8.3. In this way, we do not penalise the achievement or over-achievement of the obligation, but penalise the shortfall. The shortfall is divided into two sections, one for slightly missing the obligations and one for larger deviations from the obligations.

The surplus of the pension fund is computed with the help of the aggregated weighted surplus

$$SP^s(t) = \sum_{j=1}^{J} f_j^{SP} \Big(ML_j^s(t+m_j) - \ln\big(OB_j^s(t)\big)\Big), \tag{60}$$

where $f_j^{SP} \geqslant 0$ is a weighting factor of the surplus of each bucket. The weighting factors are also necessary to interpret the result since the surpluses arise from buckets with very different time periods. The objective function is the expected penalty minus the weighted expected surplus, given by

$$\min_{u_j(t), v_j(t)} \sum_{s=1}^{S} p_s \big(PF^s(t) - \gamma SP^s(t)\big), \tag{61}$$

where $\gamma > 0$ is the weighting factor for the aggregated surplus and p_s denotes the probability for each scenario. The optimisation problem is to choose the investment policy for each bucket $(u_j(t))$ and the fraction of pension fund wealth earmarked for the jth bucket $(v_j(t))$, while minimising the objective function. This also allows us to control the coverage ratio for each bucket. Furthermore, we impose linear (or convex) constraints on the decision variables. For example, we do not allow short-selling or leverage for the investment policy for each bucket. The earmarking of funds for the different buckets must stay within a given range in order to keep the current coverage ratio within predetermined levels. Furthermore, we need to observe institutional, legal, and regulatory constraints. For instance, in Switzerland a pension fund is prohibited from investing more than 30% of its funds in the Swiss stock market.

The optimisation problem to determine the investment policy and the funding of each bucket is

$$\min_{u_j(t),v_j(t)} \sum_{s=1}^{S} p_s\big(PF^s(t) - \gamma SP^s(t)\big)$$

$$\text{s.t.} \qquad PF^s(t) = \begin{cases} 0 & \text{if } SF^s(t) \leqslant 0, \\ p_1 SF^s(t) & \text{if } 0 \leqslant SF^s(t) \leqslant c_1, \ \forall s \in S, \\ p_2(SF^s(t) - c_1) + p_1 c_1 & \text{if } SF^s(t) \geqslant c_1, \end{cases}$$

$$SF^s(t) = \sum_{j=1}^{J} f_j^{SF}\big(\ln\big(OB_j^s(t)\big) - ML_j^s(t+m_j)\big) \quad \forall s \in S,$$

$$SP^s(t) = \sum_{j=1}^{J} f_j^{SP}\big(ML_j^s(t+m_j) - \ln\big(OB_j^s(t)\big)\big) \quad \forall s \in S,$$

$$ML_j^s(t+m_j) = \sum_{i=1}^{m_j}\big\{\ln\big(\big(r^s(t+i)\big)^{\mathrm{T}} u_j(t)\big)\big\} + \ln v_j(t) + \ln \Omega(t)$$

$$\forall s \in S,\ \forall j \in J,$$

$$u_j(t) \in \mathcal{U},\ v_j \in \mathcal{V} \quad \forall j \in J. \tag{62}$$

Here $\mathcal{U}$ denotes the set of all linear or convex constraints for the investment policy and $\mathcal{V}$ the set of all linear or convex constraints for the earmarking of funds.

Theorem 2. *The optimisation problem given in (62) is a convex optimisation problem with respect to the decision variables $u_j(t)$ and $v_j(t)$.*

Proof. In Theorem 1 we have shown that $ML_j^s(t+m_j)$ is a concave function with respect to the decision variables. Furthermore, the aggregated weighted shortfall (58) is convex with respect to the decision variables since $-ML_j^s(t+m_j)$ is convex and the nonnegative weighting preserve convexity, see Boyd and Vandenberghe (2004, 3.2.1). The penalty function (59) is piecewise linear which again preserves the convexity. The expectation operation over all scenarios also preserves the convexity. The expected penalty part of the objective function is thus convex. The aggregated weighted surplus (60) is a concave function with respect to the decision variables, but the multiplication with $-\gamma$ ($\gamma > 0$) changes this part to a convex function. The expectation operation preserves the convexity of the negative weighted surplus. Furthermore, we assume that $\mathcal{U}$ and $\mathcal{V}$ are both sets that describe linear or convex inequalities and equalities. □

The optimisation problem is convex and can be thus solved by a suitable convex programming software package. The main advantage is that we arrive at a unique solution and do not need specialised algorithms to search for the global solution. The problem

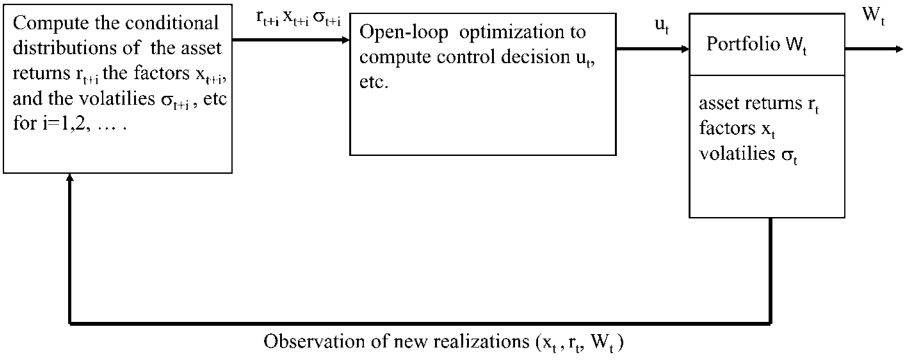


Fig. 8. Graphical description of the MPC idea.

can be solved with many decision variables, which allows the pension funds to set up a suitable number of accounts and include a wide array of investment vehicles.

8.5.3. *Control algorithm*

The asset obligation optimisation prescribes an investment rule for each bucket. However, each year the pension fund pays out the obligation of the first year bucket and collects all contributions. The bucket structure is calculated anew and the buckets for the new obligations are established. Because of the contributions, the exit and entry of members, and the effect of the guaranteed interest rate, the obligations look different from the last year's obligations, even when we take into account that the second-year obligation is now the new first-year obligation and so on.

For this reason, we have to calculate the optimisation problem (62) again. This recalculation introduces feedback into the problem, since we base our solution on the latest states of our assets and obligations. Moreover, when the rebalancing frequency is higher than the recalculation frequency of the obligations, we also solve the problem for the intermediate time periods. We then base the new optimisation also on the latest investment results and therefore, the pension fund is managed with a feedback control approach. The approach is shown in Figure 8. In control engineering this approach is known as either model predictive control, or receding horizon control, or sometimes as open-loop feedback control. This approach has thorough theoretical underpinnings, see Garcia, Prett and Morari (1989) and Bemporad et al. (2002) for deterministic applications. References to the model predictive control approach in stochastic optimisation problems in the area of finance include Herzog, Dondi and Geering (2004), Herzog, Geering and Schumann (2004), and Meindl and Primbs (2004), and, in other areas, e.g., Kouvaritakis, Cannon and Tsachouridis (2004).

By solving the problem for each time step again, we solve the feedback problem for the given state and time. The model predictive control approach always solves the optimisation problem conditional on the latest measured information. The optimisation problem corresponds to an open-loop control solution. Furthermore, at each resolving of the optimisation problem, we may also recalculate the asset return model. We can refit the asset return model to the data set that includes the new information. In this manner, we adapt our economic models to the new situation. Similarly we can also include new observations from our liability models, even though demographic statistics tend to change rather slowly. This enables us to build a truly adaptive control algorithm for the pension fund optimisation.

8.6. Asset–liability management optimisation with earmarking and investment policy and transaction costs

We describe a pension fund asset liability optimisation based on the projected liabilities and projected contributions. As described in Section 8.5, for each liability bucket we set up an account to manage the funds earmarked for the specific liability. Additionally, we assume that transaction costs and market impact may not be neglected and therefore must be included in the asset dynamics as described in Section 8.2.

8.6.1. Account dynamics

Based on the results of Section 6, we possess a stochastic description of current obligations that arise from current funds, future contributions and future (uncertain) liabilities resulting from the obligations and contributions. Since we know the probability of individual members leaving the fund, e.g., death or labour mobility, we are able to earmark their contributions accordingly to the different liability buckets. These earmarked contributions are paid into the respective account.

The control variables are the earmarking of current wealth to the individual accounts and the investment policy of the individual accounts. The asset dynamics for the accounts are

$$v_j(t)\Omega(t) = \sum_{i=1}^{N} x^s_{j,i}(t), \quad \forall s \in S, \tag{63}$$

$$\tilde{\Omega}^s_j(t+m_j) = \sum_{i=1}^{N} x^s_{j,i}(t+m_j) - LB^s_j(t), \quad \forall s \in S,\ \forall j \in J, \tag{64}$$

$$\Omega^s_j(t+k) = \sum_{i=1}^{N} x^s_{j,i}(t+k), \quad \forall s \in S,\ \forall j \in J,\ k = 0, 1, \dots, m_j, \tag{65}$$

$$\begin{aligned} &x^s_{j,i}(t+k)\big(1+\tilde{r}^s_i(t+k)\big) = y^s_{j,i}(t+k), \\ &\quad \forall s \in S,\ \forall j \in J,\ \forall i \in I, k = 0, 1, \dots, m_j, \end{aligned} \tag{66}$$

$$\begin{aligned} &x^s_{j,i}(t+k) = y^s_{j,i}(t+k-1) + p^s_{j,i}(t+k)(1-\delta_i) - d^s_{j,i}(t+k), \\ &\quad \forall s \in S,\ \forall j \in J,\ \forall i \in I \setminus \{1\}, k = 0, 1, \dots, m_j, \end{aligned} \tag{67}$$

$$x^s_{j,1}(t+k) = y^s_{j,1}(t+k-1) + \sum_{i=2}^{N} d^s_{j,i}(t+k)(1-\delta_i)$$
$$- \sum_{i=2}^{N} p^s_{j,i}(t+k) + c^s_j(t+k),$$
$$\forall s \in S,\ \forall j \in J, k = 0, 1, \ldots, m_j, \tag{68}$$
$$u_{j,i}(t) = \frac{x^s_{j,i}(t+k)}{\Omega^s_j(t+k)}, \quad s \in S,\ k = 0, 1, \ldots, m_j. \tag{69}$$

The transaction costs (or market impact) are in Eqs. (67) and (68) as linear costs, as previously shown in Eq. (45). These two equations, together with Eq. (66), describe the asset dynamics in each account j. The earmarking of current wealth is shown in (63) and the net wealth, i.e., the account's wealth at the terminal date minus the liability ($LB^s_j(t)$), is given in (64), where $\widetilde{\Omega}^s_j(t+m_j)$ denotes the account's net wealth at maturity time $t+m_j$ under scenarios s. The wealth of the account at time $t+k$ is $\Omega^s_j(t+k)$ and given in (65). The investment policy for each account is $u_{j,i}(t)$, whereas the asset values at the beginning of the time period $t+k$ are determined by Eq. (69) which is the fixed-mix decision rule. Furthermore, the contributions into the account are $c^s_j(t+k)$, which are determined from the liability bucket calculations. No outflows are included in the cash accounts, since we assume that each account at its terminal date pays the corresponding liability. Before the terminal date, no outflows occur. Unfortunately, the asset dynamics are non-linear and non-convex, since we use the dynamic stochastic control approach. Even if we use a convex objective function, the resulting optimisation problem is non-convex and requires global optimisation routines.

8.6.2. *Optimisation problem*

The aims of the fund are optimised not on the level of the individual accounts but rather on the level of the entire fund. For this reason, we use the same objectives as in Section 8.5. Since the dynamics of the accounts are given in absolute values, e.g., Swiss Francs, rather than relative (percentage) values, the objective function is also set up in absolute values. We define the aggregated weighted shortfall under scenario s as the non-negative summation of all individual weighted negative net wealth values of the accounts, or mathematically as

$$SF^s(t) = \sum_{j=1}^{J} f^{SF}_j \max\bigl(-\widetilde{\Omega}^s_j(t+m_j), 0\bigr)$$
$$= \sum_{j=1}^{J} f^{SF}_j \max\bigl(LB^s_j(t) - \Omega^s_j(t+m_j), 0\bigr). \tag{70}$$

Here $f_j^{SF} \geqslant 0$ is a weighting factor for each bucket. The risk of the pension plan is measured as a one-sided downside risk based on non-achievement of the liabilities. As penalty function use the piecewise linear function

$$PF^s(t) = \begin{cases} 0 & \text{if } SF^s(t) = 0, \\ p_1 SF^s(t) & \text{if } 0 \leqslant SF^s(t) \leqslant c_1, \\ p_2(SF^s(t) - c_1) + p_1 c_1 & \text{if } SF^s(t) \geqslant c_1. \end{cases} \tag{71}$$

Here $p_1 > 0$ and $p_2 > 0$ are slopes of the penalty function and $c_1 > 0$ is the point where the penalty increases more steeply. The shape of the penalty function can be changed arbitrarily, as long as the convexity is preserved.

The surplus of the pension fund is

$$SP^s(t) = \sum_{j=1}^{J} f_j^{SP} \big(\Omega_j^s(t + m_j) - LB_j^s(t)\big), \tag{72}$$

where $f_j^{SP} \geqslant 0$ is a weighting factor of the surplus of each bucket. The weighting factors are also necessary to interpret the result since the surpluses arise from buckets with very different time periods. For instance, one could select the weighting factors to be the discount factor in order to compare two different surpluses from two different time periods.

The objective function is the expected penalty minus the weighted expected surplus

$$\min_{u_j(t), v_j(t)} \sum_{s=1}^{S} p_s \big(PF^s(t) - \gamma SP^s(t)\big), \tag{73}$$

where $\gamma > 0$ is the weighting factor for the aggregated surplus and p_s denotes the probability for each scenario.

The entire optimisation problem is

$$\begin{aligned}
&\min_{u_j(t), v_j(t)} \sum_{s=1}^{S} p_s \big(PF^s(t) - \gamma SP^s(t)\big) \\
&\text{s.t.} \quad PF^s(t) = \begin{cases} 0 & \text{if } SF^s(t) \leqslant 0, \\ p_1 SF^s(t) & \text{if } 0 \leqslant SF^s(t) \leqslant c_1, \ \forall s \in S, \\ p_2(SF^s(t) - c_1) + p_1 c_1 & \text{if } SF^s(t) \geqslant c_1 \end{cases} \\
&\qquad SF^s(t) = \sum_{j=1}^{J} f_j^{SF} \max\big(LB_j^s(t) - \Omega_j(t + m_j), 0\big), \quad \forall s \in S, \\
&\qquad SP^s(t) = \sum_{j=1}^{J} f_j^{SP} \big(\Omega_j(t + m_j) - LB_j^s(t)\big), \quad \forall s \in S, \\
&\qquad v_j(t)\Omega(t) = \sum_{i=1}^{N} x_{j,i}^s(t), \quad \forall s \in S,
\end{aligned}$$

$$
\begin{aligned}
&\widetilde{\Omega}_j^s(t+m_j) = \sum_{i=1}^N x_{j,i}^s(t+m_j) - LB_j^s(t), \quad \forall s \in S,\ \forall j \in J,\\
&\Omega_j^s(t+k) = \sum_{i=1}^N x_{j,i}^s(t+k), \quad \forall s \in S,\ \forall j \in J,\ k = 0, 1, \dots, m_j,\\
&x_{j,i}^s(t+k)\bigl(1+\tilde{r}_i^s(t+k)\bigr) = y_{j,i}^s(t+k),\\
&\quad \forall s \in S,\ \forall j \in J,\ \forall i \in I,\ k = 0, 1, \dots, m_j,\\
&x_{j,i}^s(t+k) = y_{j,i}^s(t+k-1) + p_{j,i}^s(t+k)(1-\delta_i) - d_{j,i}^s(t+k),\\
&\quad \forall s \in S,\ \forall j \in J,\ \forall i \in I \setminus \{1\},\ k = 0, 1, \dots, m_j,\\
&x_{j,1}^s(t+k) = y_{j,1}^s(t+k-1) + \sum_{i=2}^N d_{j,i}^s(t+k)(1-\delta_i)\\
&\qquad\qquad - \sum_{i=2}^N p_{j,i}^s(t+k) + c_j^s(t+k),\\
&\quad \forall s \in S,\ \forall j \in J,\ k = 0, 1, \dots, m_j,\\
&u_{j,i}(t) = \frac{x_{j,i}^s(t+k)}{\Omega_j^s(t+k)}, \quad s \in S,\ k = 0, 1, \dots, m_j.
\end{aligned} \tag{74}
$$

The optimisation problem in Eq. (74) is a non-linear and non-convex problem, in contrast to Eq. (62) of Section 8.5, which is a convex problem. The advantage in this setup, however, is that we include transaction costs and future contributions. By also using the idea of buckets, accounts, and earmarking, we are capable of formulating an optimisation problem that takes the fund's and its members' goals and objectives into account.

We can also utilise the idea of the control algorithm given in Section 8.5.3 in order to base the investment policy and the earmarking on the newest observations of the states of our system. However, due to the inclusion of the contributions and liabilities, the need for recalculation is not as pressing as in the asset obligation optimisation case.

8.6.3. Return calculation for the individual members' contributions

The optimisation procedure based on earmarking must also ensure that the individual members are rewarded fairly with respect to the risk that their contributions take. Since we know the exit probabilities of the pension fund's members, we are able to determine the stake (share) of each individual's wealth in each of the accounts. For instance, the wealth of active members, who are close to retirement age is mostly invested in the short-term accounts, since their obligations are part of the short-term buckets. Thus, the returns achieved by the short-term accounts are the realised returns for their wealth. For active members who are still far away from the retirement age, the contributions and wealth are mostly earmarked to long-term accounts and therefore their returns are the realised returns of the long-term accounts. In this way the life-style hypothesis is realised for individual members and they are fairly rewarded with regard to their real risks.

9. Case study

In the case study we use real data to show the bucket structure of an existing pension fund. We use a factor model to describe asset price dynamics for a money market account, Swiss and international bonds, and Swiss and international stocks into which the pension fund can invest its wealth. The investments should achieve the returns necessary to cover the obligations (or liabilities) that grow with the minimal guaranteed interest rate. With the knowledge of the bucket structure and with the asset return model, we use the optimisation method in Section 8.5, which results in the investment strategy best suited to the term structure of the pension fund's expected outflows.

9.1. Description of the pension fund

Data is taken from a Swiss pension fund that is large enough to cover its own invalidity and longevity risks, called an autonomous pension fund. Among other data, the pension fund knows every member's date of birth, sex, and civil status. For active members the specific data is the momentary accumulated wealth, momentary salary, average salary, if applicable, grade of invalidity, and employment rate. For pensioners there is specific data on current available wealth and annual pension. Data is taken on 31 December 2002. The pension fund has a population of 2503 active members and 1042 pensioners (794 pensioners and 248 widows and widowers receiving pensions). Current accumulated wealth of all active members and current (remaining) wealth of the pensioners is 1123 million CHF.

Figure 9 shows the bucket structure of the expected obligations obtained from the pension fund's population. We can see the payments due in the next eighty years, when according to the life table the last of today's 25-year-old members have died. First, the light bars show the expected outflow to active members. This includes the expected outflows due to retirement and labour mobility. For example, in the first bucket ($j = 1$), for the year 2003 the pension fund expects to pay 5.5 million CHF to active members that were 64 in 2002 and have retired by now and for active members of all ages that leave the pension fund early. The dark bars indicate the expected payments to all pensioners. Since the maximum expected age is $\alpha_{\max} = 105$ (i.e., survival probability $p_{\alpha_{\max}} = 0$) years, the furthest payment to a pensioner can be in 40 years, when today's 65-year-old pensioners have reached the maximum expected age of 105 years. When there are no more of today's pensioners left, the only remaining population are today's 25-year-old members that are then retired. The slope of the expected payment stream at $j > 40$ is due to the death rate of the last retirees of the fund.

The resulting 80 buckets from Figure 9 could now be used for the optimisation described in Section 8.5. We can aggregate the 80 buckets to 9 buckets as shown in Table 3,

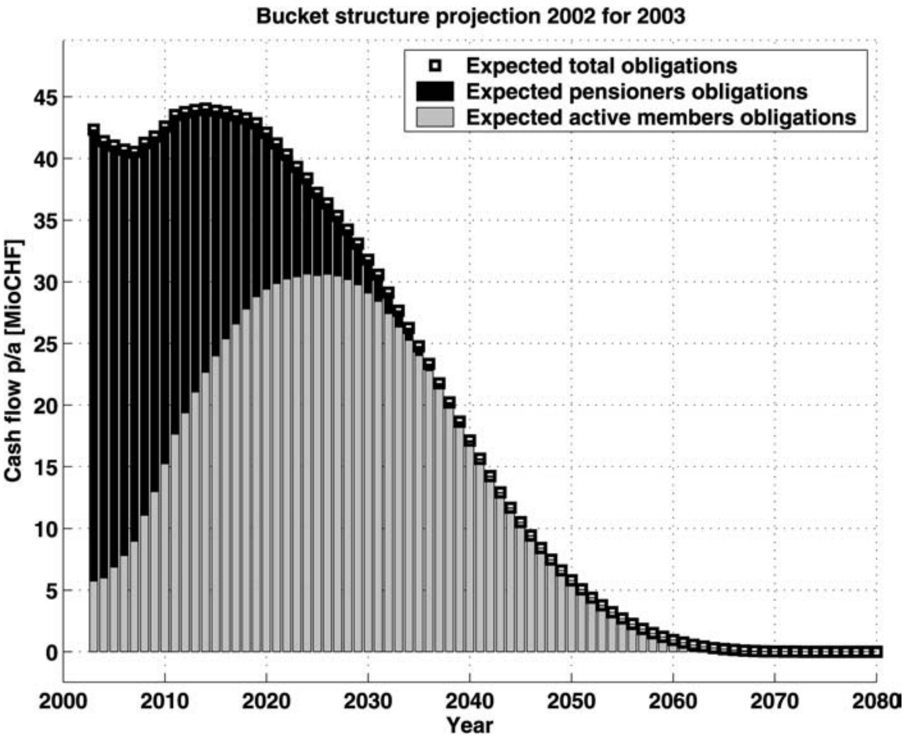


Fig. 9. Bucket structure projection 2002 for real pension fund.

Table 3
Pension fund buckets aggregated from 80 one-year buckets to 9 buckets to make the problem easier to handle

	Aggregation	j	Maturity (value weighted)
4	Single-year buckets	$j = 1, 2, 3, 4$	1, 2, 3, 4 years
4	Five-year buckets	$j = $ 5–9, 10–14, 15–19, 20–24	7, 12, 17, 22 years
1	All remaining buckets	$j = $ 25–end	32 years

the first four of which each contain one of the first four years of the original bucket structure. The next four buckets each aggregate five years into one bucket and, finally, the ninth bucket just adds up the remaining buckets. We also give these buckets a value weighted maturity, such that the term structure of the buckets is kept upright. This aggregation of buckets has been done arbitrarily and with the “common sense” argument that the finer aggregation should be in the close range and the very broad aggregation towards the end. Due to the “tail” of the bucket structure the aggregation of the very long-term buckets tends to be small in comparison to the short-term buckets.

9.2. Asset return scenario model

The optimisation procedure is based on the generation of scenarios for assets and liabilities. The liability model is described in Sections 3 to 6, whereas we have not described how we generate the asset returns.

9.2.1. Asset return model

The returns of n risky assets (or asset classes) in which we are able to invest are

$$r_t = \mu_t + \varepsilon_t^r, \tag{75}$$

where $r_t \in \mathbb{R}^n$ is the vector of asset returns, ε_t^r is a white noise process with $\mathrm{E}[\varepsilon_t^r] = 0$, and $\mathrm{E}[\varepsilon_t^r \varepsilon_t^{r\mathrm{T}}] = H$, with μ_t being the expected return and H the covariance matrix. We assume that the conditional expectation is time-varying and stochastic and that a money market account with interest rate r_t^B is available. The interest rate of the money-market account is

$$r_t^B = F_B z_{t-1} + h_B, \tag{76}$$

where the money-market account interest rate is an affine function of the factor levels z_t. For example, if one of the factors is the short-term interest rate, F_B simply selects this factor level.

9.2.2. Conditional expected returns

The expected returns of the n risky assets μ_t are

$$\mu_t = G z_{t-1} + g, \tag{77}$$

where $G \in \mathbb{R}^{n\times m}$ is the factor loading matrix, $g \in \mathbb{R}^n$ is a constant, and $z_t \in \mathbb{R}^m$ is the vector of factor levels. The factor process z_t allows us to model variables of either a macroeconomic or an industry-specific nature which affect the mean returns of the assets and vary over time. We assume that the factors are driven by a linear stochastic processes

$$z_t = A z_{t-1} + b + \varepsilon_t^z, \tag{78}$$

with $A \in \mathbb{R}^{m\times m}$, $b \in \mathbb{R}^m$, and initial condition z_0. The white-noise process can be written as $\varepsilon_t^z = \nu \xi_t^z$, with $\nu \in \mathbb{R}^{m\times m}$ and where the standard residuals are characterised by $\mathrm{E}[\xi_t^z] = 0$ and $\mathrm{E}[\xi_t^z \xi_t^{z\mathrm{T}}] = I$. The factor process z_t allows us to model variables of either macroeconomic, industry-specific, or company-specific nature that affect the mean return of the risk-bearing assets. Examples of these factors are

Macroeconomic factors	Industry-specific	Company-specific
GDP growth	Sector growth	Dividends
Long-term interest rate	Industry rate of returns	Earnings
Inflation	Industry leverage	Cash-flow

By selecting external variables with some predictive capacity, we can model the time-varying expected return of the risky asset price dynamics. For instance, many authors of empirical studies have found evidence that macroeconomic and financial factors, such as long-term interest rates or the dividend-price ratio, are suitable return predictors for US stock returns, see Glosten, Jaganathan and Runkle (1993), Campbell and Schiller (1988), and Campbell and Schiller (1991). To describe asset returns for asset liability management studies other authors used also factor models, such as Boender, van der Aalst and Heemskerk (1998) and Koivu, Pennanen and Ranne (2005). Correctly specifying input parameters such as drift values or equilibrium values matter most for successful predictive asset liability optimisations, see Hendry and Dornik (1997) or Ziemba (2003).

9.2.3. Asset return model for the Swiss pension funds

We select five major asset classes in which the pension fund can invest its wealth. Since we present a study for the strategic asset allocation (rather than the tactical asset allocation), we work with broad asset classes rather than specific investment opportunities such as individual stocks or bonds. The pension fund may invest domestically in the Swiss stock market, the Swiss bond market, and in a money-market account that pays Swiss short-term interest rates. Outside Switzerland, the fund invests in international bonds, which are in our case the Euro-zone bond market, and in international stock markets, which are restricted here to be the European stock markets (excluding Switzerland) and the US stock market. Bonds and stock markets are considered risky assets, whereas the money-market account is considered to be a risk-free investment.

We use representative total return indices for the four risky assets. For the Swiss stock market we use the DATASTREAM (DS) Swiss total market index and for the Swiss bond market we choose the DS Swiss all lives government bond index. In the case of the Euro zone bond market, we use the DS Euro-zone all lives government bond index for which data is available starting in 1999. Before 1999, we approximate the Euro-zone index by the equivalent German government bond index, since German interest rates were used as the benchmark for EU investments before the adoption of the Euro. We only use this simplified model of the early Euro-zone bond markets for model fitting only and not for a historical simulation. For the international stock markets, we choose the DS US and DS Europe total market index. For all indices we use the total return index, which includes the reinvestment of dividends in the case of stock markets and the gains or losses of the price variation in the case of bond markets, respectively. Except as noted above, all

Table 4
Factors and factor loading with statistics

Risky asset	Factors	Factor loading
Swiss stock market	log(Swiss P/E ratio) − log(10y SGB)	−0.046* (−2.04)
	Swiss DDY	0.054** (3.13)
	3m SLIBOR	−0.018** (−2.98)
	36m US Avg	0.014** (2.92)
	constant	−0.0613* (−2.56)
Swiss bond market	Swiss redemption yield	0.0037** (4.15)
	10y SGB – 3m SLIBOR	0.003** (3.73)
	constant	−0.013** (−3.13)
Euro zone bond index	Euro P/E ratio – 10y EGB	−0.01** (−2.66)
	constant	0.012** (4.39)
European stock index	log(Euro P/E ratio) – log(10y EGB)	0.032* (2.43)
	log(Euro DDY)	0.052** (2.59)
	36m US Avg	0.012* (2.07)
	constant	−0.042* (−2.16)
US stock market index	log(US P/E ratio) – log(10y USGB)	0.08** (4.03)
	36m US Avg	0.013** (2.98)
	constant	0.011* (2.25)

Swiss P/E ratio: Swiss stock market price-earnings ratio. 10y SGB: 10-year Swiss government bond interest rate (constant maturity). 3m SLIBOR: Swiss 3-month LIBOR rate. 36m US Avg: 36-month moving average US total stock market index total return. constant: constant g_i. Swiss red. yield: Swiss bond market index redemption yield. Euro P/E ratio: European stock market index P/E ratio. 10y EGB: Euro-zone government bonds 10-year interest rate (constant maturity). Euro DDY: European stock market index dividend yield. US P/E ratio: US total stock market index P/E ratio. 10y USGB: 10-year US treasury bond interest rate (constant maturity).

* denotes significance at 5% level.

** indicates significance at the 1% level. t-statistics in parentheses.

data sets start in January 1986 and end in September 2004, with a frequency of one month.

The four expected returns of the four risky indices are modelled by a factor model as explained in Section 9.2. The factors are selected such that the factor loading matrix G and the constant g are statistically significant. To achieve this, a set of predetermined factors were selected and the model is fitted to this data. The fitting algorithm is maximum pseudo-likelihood (Hamilton, 1994). A wide range of factors are preselected as recommended by Oberuc (2004, Chapters 3 and 4). Then each factor that is not statistically significant at significance level 5% is eliminated and the data is fitted again. This procedure is iterated until we find for each index a small number of significant factors.

The factors selected for the risky assets and the factor loadings are shown in Table 4. Based on this factor model of the risky assets, we simulate 5000 scenarios for

Table 5
Long-term asset model properties

Asset class	Return	Volatility	Note
Money market	2.5%	n.a.	
Swiss bonds	4.2%	3.2%	
International bonds	5.3%	5.3%	not hedged in CHF
Swiss stocks	9.9%	16.0%	
International stocks	10.9%	16.5%	hedged in CHF
International stocks	12.8%	21.0%	not hedged in CHF

32 years with quarterly frequency into the future. The simulation uses the empirical standard residuals from the data fitting as standard residuals to drive the simulation. The simulation is thus a bootstrap method that randomly selects an empirical standard residual from the factors and asset returns. In this way, we do not need to make any explicit distributional assumptions for the standard residuals, but keep any dependence that the residuals exhibit. In the bootstrap method, we select an observation of the residuals where simultaneously the empirical residuals of the risky assets and the factors are used. Since we generate 5000 scenarios and from the data fitting we have 225 observations from the fitting, we use the empirical residuals sufficiently to get a statistically significant set of scenarios. For the application of bootstrap methods, see Davison and Hinkley (1999).

Furthermore, we assume that the pension fund invests into a balanced index of European and US stocks. Therefore, we calculate an international stock market index which consists of 40% of the European stock market index and of 60% of the US stock market index. The international stock index is constantly rebalanced. The international stocks are hedged against currency movements where the pension fund uses currency futures for hedging. For international bonds, we assume that the currency movements are not hedged, since the volatility between the Euro and the Swiss Franc is moderate. The long-term steady-state properties of the five asset classes are presented in Table 5. We include the cost of hedging in the generation of the asset scenarios. The cost of hedging is the spread between the Swiss short-term interest rates and the short-term US and Euro interest rates, respectively.

In Figure 10 the path of 50 scenarios of the Swiss stock market are shown. Figure 11 shows the path of 50 scenarios of the European bond market index for 2004 to 2014. The index values are normalised such that both indices start with value 100.

9.3. Asset allocation optimisation for the pension fund

We use the optimisation methods from Section 8.5. The obligations are summarised into nine buckets with different maturities. To set up the optimisation problem, we need to define the parameters f_j^{SF}, f_j^{SP}, γ, and the shape of the penalty function. We use a penalty function that penalises two regions of shortfalls, the first where the shortfall

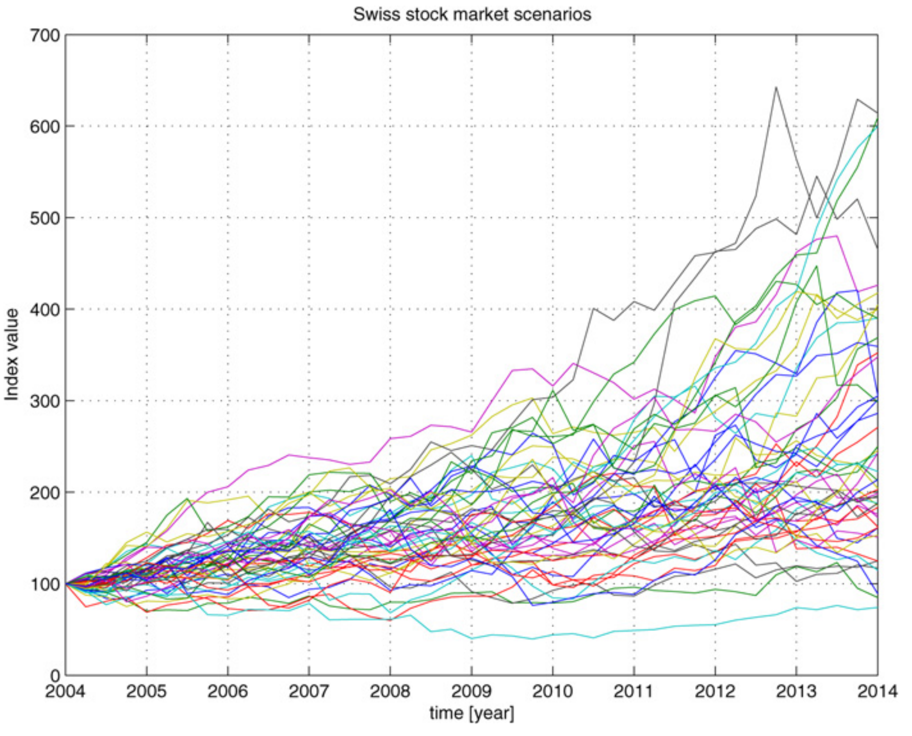


Fig. 10. Trajectories of 50 scenarios for the Swiss stock market index.

is 10% or smaller, i.e., the assets' value only cover 90% of the obligations or more, and the second where the shortfall is larger than 10%. The shortfall is penalised in the first region with the factor $p_1 = 1$ and in the second region with factor $p_2 = 4$. Large deviations of the assets' values from the obligations are penalised four times more than small deviations.

The weighting factors for the shortfall and the surplus are listed in Table 6. The weighting factors for the surplus are selected such that the first four buckets do not contribute to the surplus of the fund (in the sense of the objective function). The factors from buckets five to nine are chosen to be the discount factors for the maturity date of the respective bucket. In this way, the contributions to the surplus of buckets with different terminal dates are correctly summed up. The weighting factors for the shortfall are chosen such that shortfall projections that occur at short-term buckets are seen as much more severe than shortfall projections in the long run. The philosophy behind this selection of the factor weights is: "short buckets safety first, long buckets make profits". The optimisation thus chooses an investment policy and earmarking strategy that ensures a high probability of covering the short-term obligations on the one hand and long-term profits from the long-tern accounts on the other.

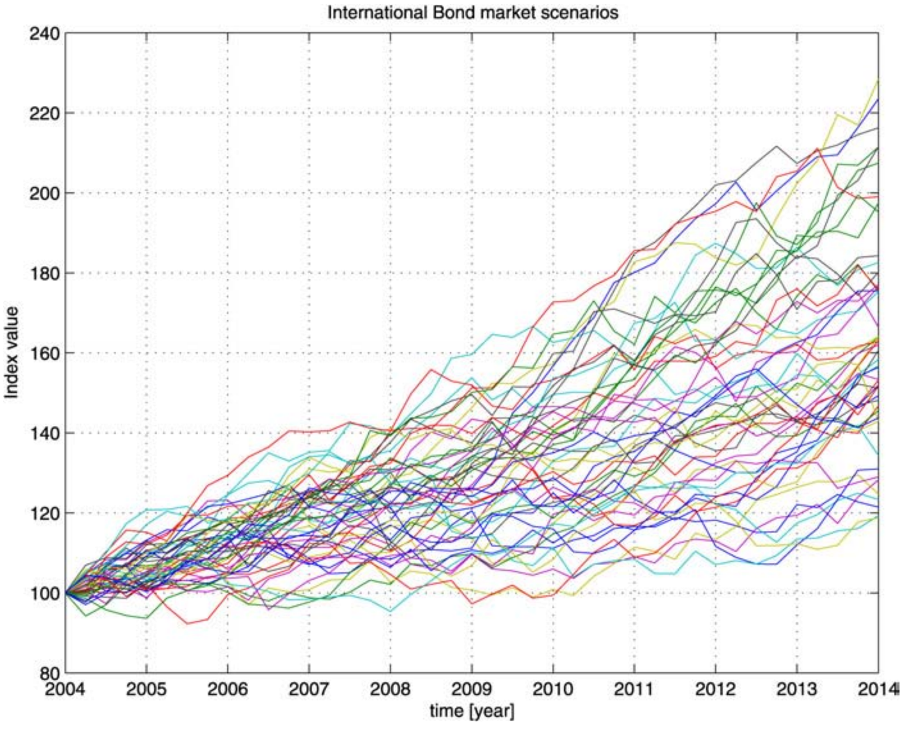


Fig. 11. Trajectories of 50 scenarios for the Euro-bond market index.

Table 6
Optimisation weighting parameters

Bucket	1	2	3	4	5	6	7	8	9	
f_j^{SF}	6	6	6	6	3	3	3	1	1	shortfall aggregation
f_j^{SP}	0	0	0	0	0.75	0.62	0.51	0.42	0.29	surplus aggregation

9.3.1. Conservative asset allocation

So far we have not discussed how we are to choose the parameter γ, the weighting factor for aggregated surplus in Eq. (61). With the help of γ we are able to decide whether we put the emphasis on covering the obligations or on generating a surplus from the funds under management. Here we look at the results of the optimisation when we set $\gamma = 0.05$. This means that we are predominantly concerned with meeting our obligations. The conservative strategy ($\gamma = 0.05$) results in the bucket earmarking and asset allocation as summarised in Table 7.

Table 7
Conservative strategy, $\gamma = 0.05$. Earmarking and current coverage ratios for aggregated buckets

j-Bucket	earm. (v_j)	CCR	M. Mk.	Sw. Bd.	Int. Bd.	Sw. Sk.	Int. Sk.
$j = 1$	7.71%	103.8%	91.3%	8.7%	0%	0%	0%
$j = 2$	7.3%	108.3%	49.8%	29.5%	17.8%	2.9%	0%
$j = 3$	6.96%	112.5%	10.2%	60.4%	22.7%	6.8%	0%
$j = 4$	6.51%	114.6%	6.4%	63.6%	21.6%	6.4%	0%
$j = 5$–9	27.04%	117.2%	0%	21.3%	48.2%	26.1%	4.4%
$j = 10$–14	17.9%	110.6%	0%	0%	40.0%	27.0%	33.0%
$j = 15$–19	13.12%	124.6%	0%	0%	5.7%	33.7%	60.6%
$j = 20$–24	7.37%	115.9%	0%	0%	0%	33.2%	66.8%
$j = 25$–end	6.09%	85.8%	0%	0%	0%	23.0%	77.0%

M. Mk.: Swiss money market account. Sw. Bd.: Swiss Bonds. Int. Bd.: international bonds. Sw. Sk.: Swiss stock market. Int. Sk.: international stock market. CCR: current coverage ratio. earm.: earmarking of funds.

The conservative strategy ensures high liquidity for the first four buckets. With very high probability, the obligations are covered, since the investment policy is conservative, i.e., with most investments being in the money market and the Swiss Bond market. The earmarking results in very high coverage ratios. For instance, the first bucket possesses a current coverage ratio of 103.8% and is almost exclusively invested in the money market. Since the obligation increases with the guaranteed interest rate of 4%, the first bucket has enough funds to cover the obligation even when the realised interest rate is almost zero. The next three buckets also have current coverage ratios that are very close to the amount needed to cover the obligations at maturity. The investment policy shifts from the money market account to predominantly Swiss and International bonds.

The buckets with medium-term maturity, i.e., buckets five and six, have a far more diversified investment policy, with investments in most of the available asset classes. The earmarking of funds ensures high coverage ratios for these two buckets. The medium- to long-term accounts are almost exclusively invested in the stock market, where the share of international stocks increases with increasing maturity. The buckets seven and eight also have high coverage ratios, however, bucket nine possesses a coverage ratio smaller than 100%. Since bucket nine matures in 32 years, the shortfall is not regarded as very critical and is a result of the high earmarking of funds for short-term obligations.

Table 8 lists the statistics for the buckets. Based on the asset and obligation model, we can compute the mean shortfall, the expected coverage ratio at maturity, and the probability of a shortfall. The mean shortfall is the expectation of all scenarios that do not cover the obligation value.

The statistics show that with a very high probability the short-term obligations (buckets 1–4) are covered. Furthermore, in the case that funds do not cover the obligations, the average shortfall is rather small. The expected shortfall for the long-term obligations can be substantial, but this case occurs also with a low probability. The expected coverage ratios are above one for the short-term buckets and high for the long-term buckets.

Table 8
Statistics for the conservative strategy

Bucket number	1	2	3	4	5	6	7	8	9
Mean shortfall (%)	0.08	0.8	1.15	1.31	4.75	7.57	10.72	4.4	17.7
Exp. coverage ratio	1.01	1.05	1.11	1.13	1.32	1.87	3.49	4.5	6.19
Probability shortfall (%)	0.4	0.3	0.28	0.28	1.2	0.5	0.26	0.08	0.2

Table 9
Aggressive strategy $\gamma = 10$. Earmarking and current coverage ratios for aggregated buckets

j-Bucket	earm. (v_j)	CCR	M. Mk.	Sw. Bd.	Int. Bd.	Sw. Sk.	Int. Sk.
$j = 1$	7.61%	102.3%	61.5%	31.4%	7.1%	0%	0%
$j = 2$	6.84%	101.5%	0%	70.6%	29.5%	0%	0%
$j = 3$	6.36%	102.8%	0%	70.3%	29.7%	0%	0%
$j = 4$	5.85%	103%	0%	63.0%	33.6%	3.4%	0%
$j = 5$–9	22.73%	98.5%	0%	0%	66.2%	33.8%	0%
$j = 10$–14	17.62%	108.8%	0%	0%	20.1%	61.0%	18.9%
$j = 15$–19	13.92%	132.1%	0%	0%	0%	46.8%	53.2%
$j = 20$–24	11.37%	178.7%	0%	0%	0%	31.9%	68.1%
$j = 25$–end	7.70%	108.5%	0%	0%	0%	15.4%	84.6%

M. Mk.: Swiss money market account. Sw. Bd.: Swiss bonds. Int. Bd.: international bonds. Sw. Sk.: Swiss stock market. Int. Sk.: international stock market. CCR: current coverage ratio. earm.: earmarking of funds.

This is expected, since the investment policy for the long-term bucket uses mostly assets with high expected returns. The obligations, however, grow at 4% and thus the expected coverage ratios at maturity are very high.

9.3.2. *Aggressive asset allocation*

In contrast to the conservative strategy, we now choose $\gamma = 10$ in order to force the optimisation to concentrate on the generation of a surplus. The result of the optimisation for the aggressive study is shown in Table 9, where the earmarking and the investment policy are given. In the aggressive case, the earmarking towards short-term buckets is lower than in the conservative case, which in turn results in current coverage ratios slightly above 100%. Accordingly, the investment policy for the short-term buckets must generate higher returns and thus, more investments in Swiss and international stocks.

The intermediate buckets (buckets 5–6) have increasing coverage ratios in which the investment policy replaces international bonds by Swiss and international stocks. The long-term buckets (buckets 7–9) are solely invested in the stock markets. With increasing maturity the share of international stocks is increased. Especially buckets 7 and 8 have very high coverage ratios. Bucket 9 has a sufficient coverage ratio. The three long-

Table 10
Statistics for the aggressive strategy

Bucket number	1	2	3	4	5	6	7	8	9
Mean shortfall (%)	0.8	2.5	2.8	3.0	8.3	15.5	15.1	1.0	5.1
Exp. coverage ratio	1.01	1.01	1.01	1.02	1.14	1.87	3.7	7.04	8.1
Probability shortfall (%)	47.1	41.0	38.9	35.4	18.7	4.36	0.3	0.0	0.0

term buckets mostly contribute to the aim of achieving a surplus from the funds under management. Table 10 shows the statistics for the different buckets. Noteworthy are the high probabilities for a possible shortfall for the first four buckets.

However, the mean shortfall is small, which indicates that frequently shortfalls occur, but with small deviations from the obligations. When we use the control algorithm, the obligations are paid out regardless of whether or not the account completely covers the obligations by using funds earmarked for other buckets. When we resolve the problem, the funds are smaller than anticipated and the new investment policy and earmarking take this into account. The risks of the aggressive strategy are highlighted by this fact. The expected coverage ratios at maturity are larger than in the other strategy. The aggressive strategy aims for profits in the long run and therefore uses the wealth to invest more in the long-term accounts. For this reason, the current coverage ratios for the medium- and long-term buckets are quite high. The investment policy uses mostly stock market investments in order to increase the expected returns. Especially in the case of the last two buckets the mean shortfall and the shortfall probability decrease significantly. The expected coverage ratios are very high for the long-term buckets. This is expected since the investment policy for the long-term bucket uses mostly assets with high expected returns.

9.3.3. Comparison of the conservative and aggressive strategies

The two strategies are very similar in their investment policy for the long-term accounts, especially for buckets 8 and 9. The long-term nature of the investment problem, i.e., 25 years and more, and the resulting long-term risk-return trade-off leads to very similar investment decisions.

However, as Figure 12 shows, the earmarking of wealth for these two strategies are very different. In the aggressive case the long-term buckets have more funds available than in the conservative case. The conservative case uses the funds to cover the short- to medium-term obligations, whereas in the aggressive strategy the funds are used to generate long-term profits. The earmarking of the pension fund's wealth (assets) is an essential feature of the risk management. The investment policy depends both on the investment horizon and on the initial funds available to each account. For example, in the conservative case, bucket three is mostly invested in Swiss bonds and the money-market account, with some small investments in the Swiss stock market.

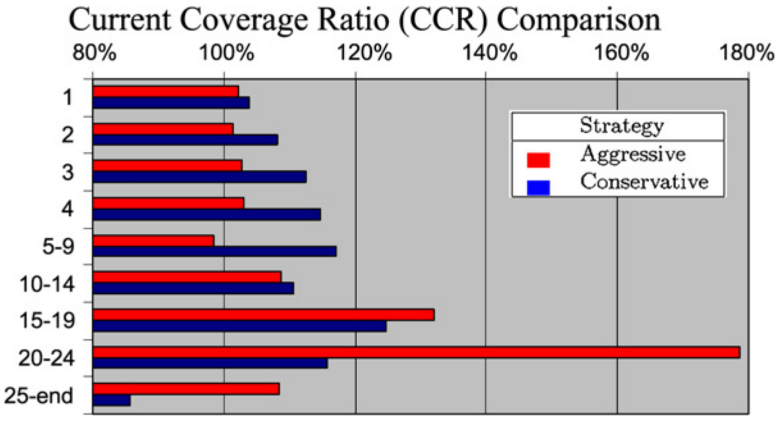


Fig. 12. Comparison of current coverage ratios.

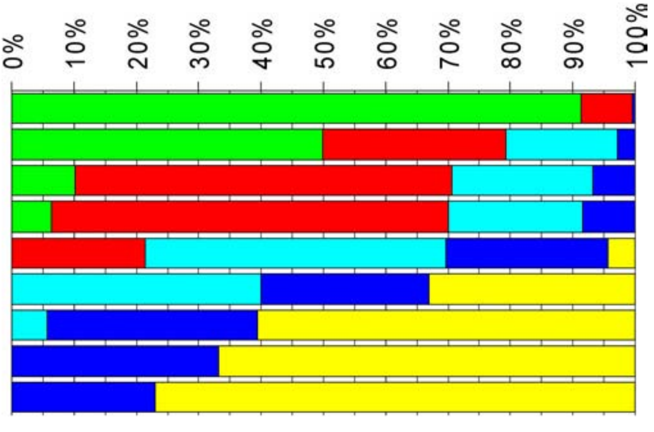


Fig. 13. Conservative asset allocation for nine buckets.

In the aggressive case, bucket three has only bond investments. The probability of a shortfall, however, is much smaller in the conservative case than in the aggressive case since the earmarking is much larger, i.e., CCR 112.5% vs. 102.8%. The asset allocation for both cases, as shown graphically in Figures 13 and 14, is similar for the long-term accounts. However, earmarking and the resulting current coverage ratios are totally different, since earmarking determines mainly the probability of a possible shortfall.

The aggregation of the different investment classes over all buckets result in the percentages given in Tables 11 and 12. Both cases have the same investments in international bonds and stocks. The main differences are the investment in the Swiss stock

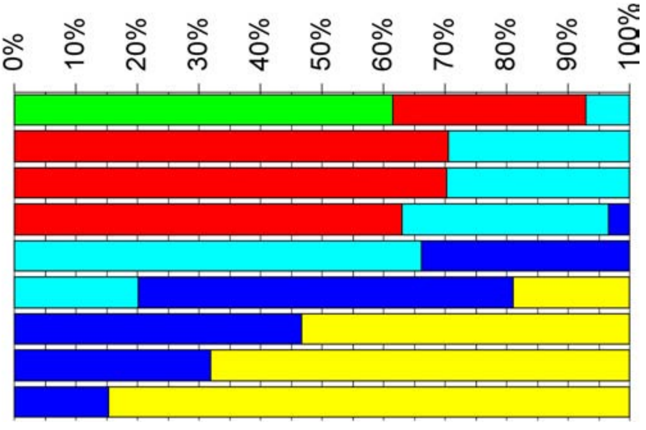


Fig. 14. Aggressive asset allocation for nine buckets.

Table 11
Conservative strategy $\gamma = 0.05$. Aggregation in investment classes

	Asset class	Σ
	Money market	11.8%
	Swiss bonds	16.9%
	Int'l bonds	25.2%
	Swiss stocks	21.4%
	Int'l stocks	24.6%

Table 12
Aggressive strategy $\gamma = 10$. Aggregation in possible investment classes

	Asset class	Σ
	Money market	4.6%
	Swiss bonds	15.4%
	Int'l bonds	25.0%
	Swiss stocks	30.0%
	Int'l stocks	25.0%

market and the money market. The conservative case places more funds in the money-market account and invests less in the Swiss stock market. The aggressive strategy uses the money market only for the first bucket and for no other bucket. Both strategies result in the same international investments, because the maximum exposure to international assets for a Swiss pension fund is limited to 50%. Furthermore, the international bonds are used to generate the returns for the medium-term buckets, while the international stocks are used for the long-term buckets. The investment policy in the case of medium- and long-term buckets is very similar and this results in the same aggregated investments in international assets. The limit for Swiss stocks is 30%, which limits the exposure to Swiss stocks in the case of the aggressive strategy.

9.4. Control algorithm for pension fund asset liability management

Given the pension fund bucket structure for the future projected cash-flows and given the model for asset returns, we can now use the optimisation method in Section 8.5 to find the optimal asset allocation and earmarking. If we sequentially update our estimations, update the available data, and calculate a new asset allocation based on the new knowledge, we obtain a control algorithm for the pension fund asset liability management. The first step is to establish the bucket structure and determine the current net wealth which is to be invested. In the next step we estimate parameters for the factor model, given the new observations of the asset returns. We then simulate the asset scenarios. Based on the knowledge of the bucket structure and asset return scenarios, we solve the optimisation problem and invest accordingly. Depending on the rebalancing period, the parameter estimation, solving of the optimisation problem, and rebalancing of the portfolio can be repeated frequently, i.e., monthly, or less frequently, such as quarterly or semi-annually. After one year the returns on the portfolio are realised and the first bucket needs to be paid to pensioners and active early leaving members. The algorithm begins anew by establishing the new bucket structure and by determining current net wealth, and it finds the "open-loop" investment strategy by solving the optimisation problem.

The control algorithm results in a regular updating and adaptation of the investment strategy to new situations on both financial markets and member structure of the pension fund. This resembles a closed loop control algorithm. The control algorithm is similar to a model predictive control approach. Where the open loop optimisation problem is solved for the long-term problem however the solution is not implemented over the full time horizon. The problem is newly solved after the first time step (or first few time steps) has (have) passed and new data is available. The result is a closed-loop like strategy.

10. Conclusion

The pension fund receives contributions during the working age of the members, invests the contributions and thereby accumulates wealth which is then paid as pensions after retirement of the members. We use data of the members of a pension fund in order to model the payment streams and construct a structure of future cash-flows out of the pension fund. The cash flows are based on the pension payments which are "promised" until the death of the pensioners, comparable to an annuity. Also considered are payments to active members which may leave the pension fund prematurely, i.e., prior to retirement and eligibility for a pension. Since the premature exit of active members and death of pensioners are given by probabilities known from the life insurance industry we calculate the expected payments and when they become due. This results in a structure of payment streams due in the future. By bundling the payments into yearly sums we create the bucket structure for pension funds. The bucket structure is used to define an optimisation method that assigns the available wealth into the different buckets. It also specifies the necessary investment strategy in order to reach the goal of the pension fund, which is to be able to pay pensions when they are due. The method is applied to the data of a real pension fund in a case study. Two different investment strategies, one conservative and one aggressive, are calculated. The conservative strategy clearly prefers liquidity and safer investment strategies to longer-term investment into more risky investments than the aggressive strategy does.

References

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Bacinello, A.R., 1988. A stochastic simulation procedure for pension funds. Insurance: Mathematics and Economics 7 (3), 153–161.

Barro, D., Canestrelli, E., 2005. Dynamic portfolio optimization: Time decomposition using the maximum principle with a scenario approach. European Journal of Operations Research 163 (1), 217–229.

Battocchio, P., Menoncin, F., 2004. Optimal pension management in a stochastic framework. Insurance: Mathematics and Economics 34 (1), 79–95.

Battocchio, P., Menoncin, F., Scaillet, O., 2003. Optimal asset allocation for pension funds under mortality risk during the accumulation and decumulation phases, Working Paper No. 26. NCCR Finrisk, Zürich, Switzerland.

Bemporad, A., Morari, M., Dua, V., Pistikoulos, E.N., 2002. The explicit linear quadratic regulator for constrained systems. Automatica 28, 3–20.

Berger, A., Mulvey, J.M., Ruszczynski, A., 1994. An extension of the DQA algorithm to convex stochastic programs. SIAM Journal of Optimization 4, 735–753.

Birge, J.R., Holmes, D., 1992. Efficient solution of two-stage stochastic linear programs using interior point methods. Computational Optimization and Applications 1, 245–276.

Birge, J.R., Donohue, C., Holmes, D., Svintsitski, O., 1994. A parallel implementation of the nested decomposition algorithm for multistage stochastic linear programs. Mathematical Programming 75, 327–352.

Blake, D., 1998. Pension schemes as options on pension fund assets: Implications for pension fund management. Insurance: Mathematics and Economics 23 (3), 263–286.

Blake, D., Cairns, A.J.G., Dowd, K., 2001. Pensionmetrics: Stochastic pension plan design and value-at-risk during the accumulation phase. Insurance: Mathematics and Economics 29 (2), 187–215.

Bodie, Z., Detemple, J.B., Otruba, S., Walter, S., 2004. Optimal consumption portfolio choices and retirement planning. Journal of Economic Dynamics and Control 28 (6), 1115–1148.

Boender, G.C., 1997. A hybrid simulation/optimisation scenario model for asset/liability management. European Journal of Operational Research 99 (1), 126–137.

Boender, C., van der Aalst, P., Heemskerk, F., 1998. Modelling and management of asset and liability of pension plans in the Netherlands. In: Ziemba, W., Mulvey, J. (Eds.), Worldwide Asset and Liability Management. Cambridge University Press, pp. 561–580.

Bogentoft, E., Romeijn, H.E., Uryasev, S., 2001. Asset/liability management for pension funds using CVaR constraints. Journal of Risk Finance 3 (1), 57–71.

Bosch-Príncep, M., Devolder, P., Domínguez-Fabián, I., 2002. Risk analysis in asset–liability management for pension funds. Belgian Actuarial Bulletin 2 (1).

Boulier, J.-F., Huang, S., Thaillard, G., 2001. Optimal management under stochastic interest rates: the case of a protected defined contribution pension fund. Insurance: Mathematics and Economics 28 (2), 173–189.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Brennan, M.J., Schwartz, E.S., Lagnado, R., 1997. Strategic asset allocation. Journal of Economic Dynamics and Control 21 (8–9), 1377–1403.

Cairns, A.J.G., 1994. Continuous-time pension-fund modelling, Working paper. Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh.

Cairns, A.J.G., 2003. Pension-fund mathematics, Discussion Paper Pension Institute PI-0315. Birkbeck College, University of London, UK.

Cairns, A.J.G., Blake, D., Dowd, K., 2000. Optimal dynamic asset allocation for defined-contribution pension plans, Working paper. Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh.

Cairns, A.J.G., Blake, D., Dowd, K., 2003. Stochastic lifestyling: Optimal dynamic asset allocation for defined contribution pension plans, Working paper. Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh.

Campbell, J.Y., Rodriguez, J., Viceira, L.M., 2004. Strategic asset allocation in a continuous-time VAR model. Journal of Economic Dynamics and Control 28 (11), 2195–2214.

Campbell, J.Y., Schiller, R., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1 (3), 195–228.

Campbell, J.Y., Schiller, R., 1991. Spreads and interest rates: a bird's eye view. Review of Economic Studies 58 (3), 495–514.

Campbell, J.Y., Viceira, L.M., 2002. Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. Oxford University Press.

Cariño, D.R., Ziemba, W.T., 1998. Formulation of the Russell–Yasuda Kasai financial planning model. Operations Research 46 (4), 433–449.

Cariño, D., Kent, T., Myers, D., Stacy, C., Sylvanus, M., Turner, A., Watanabe, K., Ziemba, W., 1994. The Russell–Yasuda Kasai model: An asset liability model for a Japanese Insurance Company using multistage stochastic programming. Interfaces 24, 29–49.

Chang, S.-C., 1999. Optimal pension funding through dynamic simulations: The case of Taiwan public employees retirement system. Insurance: Mathematics and Economics 24 (3), 187–199.

Charupat, N., Milevsky, M.A., 2002. Optimal asset allocation in life annuities: a note. Insurance: Mathematics and Economics 30 (2), 199–209.

Clark, R.L., Burkhauser, R.V., Moon, M., Quinn, J.F., Smeeding, T.M., 2004. The Economics of an Aging Society. Blackwell Publishing, Oxford.

Cowell, F.A., Ferreira, F.H., Litchfield, J.A., 1998. Income distribution in Brazil 1981–1990: Parametric and non-parametric approaches. Journal of Income Distribution 8 (1), 63–76.

Dantzig, G., Infanger, G., 1993. Multi-stage stochastic linear programs for portfolio optimization. Annals of Operations Research 45, 59–76.

Davison, A.C., Hinkley, D.V., 1999. Bootstrap Methods and their Application. Cambridge University Press.

Deelstra, G., Grasselli, M., Koehl, P.-F., 2003. Optimal investment strategies in the presence of a minimum guarantee. Insurance: Mathematics and Economics 33 (1), 189–207.

Dert, C., 1998. A Dynamic Model for asset liability management for defined pension funds. In: Ziemba, W., Mulvey, J. (Eds.), Worldwide Asset and Liability Management. Cambridge University Press, pp. 501–536.

Devolder, P., Bosch-Princep, M., Fabian, I.D., 2003. Stochastic optimal control of annuity contracts. Insurance: Mathematics and Control 33 (2), 227–238.

Dondi, G., 2003. Working Paper LSE2000: Data interpretation and statistical analysis. ETH Zürich, Switzerland.

Dowd, K., Blake, D., Cairns, A., 2003. Long-term value at risk. Centre for Risk & Insurance Discussion Paper Series—2003. I, University of Nottingham.

EVK, 2000. Technische Grundlagen der Eidgenössischen Versicherungskasse EVK2000. Eidgenössische Versicherungskasse / PUBLICA, Bern.

Garcia, C., Prett, D., Morari, M., 1989. Model predictive control: theory and practice. Automatica 25 (3), 335–348.

Gerber, H.U., 1997. Life Insurance Mathematics, third ed. Springer, New York, NY.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2004. The innovest Austrian Pension Fund financial planning model InnoALM. Conference on Asset and Liability Management: from Institutions to Households, Nicosia, May 2001.

Glosten, L.R., Jaganathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of nominal excess returns on stocks. Journal of Finance 48, 1779–1802.

Haberman, S., Sung, J.-H., 1994. Dynamic approaches to pension funding. Insurance: Mathematics and Economics 15 (2–3), 151–162.

Haberman, S., Vigna, E., 2002. Optimal investment strategies and risk measures in defined contribution pension schemes. Insurance: Mathematics and Economics 31 (1), 35–69.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press.

Hendry, D., Dornik, J., 1997. The implications for economic modelling of forecast failures. Scottish Journal of Political Economy 44, 437–461.

Herzog, F., Dondi, G., Geering, H.P., 2004. Strategic asset allocation with factor models for returns and GARCH models for volatilities. In: Hamza, M.H. (Ed.), Proceedings of the 4th IASTED International Conference on Modelling, Simulation, and Optimization, IASTED, Calgary, Canada, pp. 59–65.

Herzog, F., Geering, H.P., Schumann, L., 2004. Strategic portfolio management with coherent risk measures for dynamic asset models. In: Hamza, M.H. (Ed.), Proceedings of the 2nd IASTED International Conference on Financial Engineering and Applications—FEA 2004, IASTED, Calgary, Canada, pp. 114–123.

Herzog, F., Dondi, G., Geering, H.P., Schumann, L.M., 2004. Continuous-time multivariate strategic asset allocation. In: Proceedings of the 11th Annual Meeting of the German Finance Association, Tübingen, Germany.

Josa-Fombellida, R., Rincon-Zapatero, J.P., 2004. Optimal risk management in defined benefit stochastic pension plans. Insurance: Mathematics and Economics 34 (3), 489–503.

Kall, P., Wallace, S.W., 1994. Stochastic Programming. John Wiley, New York, NY.

Kingsland, L., 1982. Projecting the financial condition of a pension plan using simulation analysis. Journal of Finance 37 (2), 577–584.

Koivu, M., Pennanen, T., Ranne, A., 2005. Modelling assets and liabilities of a Finnish Pension Company: VEqC approach. Scandinavian Actuarial Journal 1, 46–76.

Koller, M., 2000. Stochastische Modelle in der Lebensversicherung. Springer, Berlin.

Konno, H., Wijayanayake, A., 2002. Portfolio optimization under D.C. transaction costs and minimal transaction unit constraints. Journal of Global Optimization 22, 137–154.

Kouvaritakis, B., Cannon, M., Tsachouridis, V., 2004. Recent developments in stochastic MPC and sustainable development. Annual Reviews in Control 28, 23–35.

Kusy, M., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34 (3), 356–376.

Louveaux, F., Birge, J.R., 1997. Introduction to Stochastic Programming. Springer, New York, NY.

Luenberger, D.G., 1998. Investment Science. Oxford University Press.

Maranas, C., Androulakis, I., Floudas, C., Berger, A., Mulvey, J.M., 1997. Solving long-term financial planning problems via global optimization. Journal of Economic Dynamics and Control 21 (8–9), 1405–1425.

Meindl, P.J., Primbs, J.A., 2004. Dynamic hedging with transaction costs using receding horizon control. In: Hamza, M. (Ed.), Proceedings of the 2nd IASTED International Conference on Financial Engineering and Applications—FEA 2004, Cambridge, IASTED, Calgary, Canada, pp. 142–147.

Merton, R.C., 1969. Lifetime portfolio selection under uncertainty: The continuous case. Review of Economics and Statistics 51 (3), 247–257.

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.
Merton, R.C., 1992. Continuous-Time Finance, second ed. Blackwell Publishers Inc., Oxford UK.
Mulvey, J.M., 1995. Generating scenarios for the towers Perrin investment system. Interfaces 26, 1–15.
Mulvey, J.M., Pauling, W.R., Madey, R.E., 2003. Advantages of multiperiod portfolio models. The Journal of Portfolio Management 2003, 35–45, Winter.
Mulvey, J.M., Shetty, B., 2004. Financial planning via multi-stage stochastic optimization. Computers and Operations Research 31 (1), 1–20.
Mulvey, J.M., Simsek, K.D., 2002. Rebalancing strategies for long-term investors. In: Kontoghiorghes, E. (Ed.), Computational Methods in Decision-Making, Economics, and Finance. Kluwer Academic Publishers, pp. 15–33 (Ch. 2).
NZZ, 2004. Neue Zürcher Zeitung: NZZ Fokus: Berufliche Vorsorge. NZZ AG, CH-8021 Zürich.
Oberuc, R.E., 2004. Dynamic Portfolio Theory and Management. McGraw-Hill, New York.
O'Brien, T., 1986. A stochastic-dynamic approach to pension funding. Insurance: Mathematics and Economics 5 (2), 141–146.
OECD, 2001. Insurance and Private Pensions Compendium for Emerging Markets, Book 2. Part 2:1.a. OECD.
Pittau, M.G., Zelli, R., 2002. Income distribution in Italy: A nonparametric analysis. Statistical Methods and Appliances 10 (1–3).
Reichlin, A., 2000. Asset liability management for pension funds. University of Zürich.
Rockafellar, R., Uryasev, S.P., 2000. Optimization of conditional value-at-risk. Journal of Risk 2 (3), 21–41.
Rockafellar, R., Uryasev, S.P., 2002. Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26 (7), 1443–1471.
Rockafellar, R.T., Wets, R.-B., 1991. Scenario and policy aggregation in optimization under uncertainty. Mathematics of Operations Research 16, 650–669.
Rudolf, M., Ziemba, W.T., 2004. Intertemporal surplus management. Journal of Economic Dynamics and Control 28, 975–990.
Rufibach, K., Bertschy, M., Schüttel, M., Vock, M., Wasserfallen, T., 2001. Eintrittsraten und Austrittswahrscheinlichkeiten EVK2000. Mitteilungen der Schweiz. Aktuarvereinigung 49 (1).
Simonovits, A., 2003. Modelling Pension Systems. Palgrave Macmillan, New York, NY.
Taylor, G., 2002. Stochastic control of funding systems. Insurance: Mathematics and Economics 30 (3), 323–350.
van Slyke, R., Wets, R.-B., 1969. L-shaped linear programs with applications to optimal control and stochastic programming. SIAM Journal of Applied Mathematics 17, 833–846.
Wilkie, A., 1995. More on a stochastic asset model for actuarial use. British Actuarial Journal 39, 341–403.
Winklevoss, H.E., 1982. Plasm: Pension liability and asset simulation model. Journal of Finance 27 (2), 585–594.
Zenios, S.A., 2002. Financial Optimization. Cambridge University Press.
Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset Liability and Wealth Management. AIMR Publications, Charlottesville, VA.
Ziemba, W.T., Mulvey, J.M., 1998. Worldwide Asset and Liability Modelling. Cambridge University Press.

Chapter 21

JOINED-UP PENSIONS POLICY IN THE UK: AN ASSET–LIABILITY MODEL FOR SIMULTANEOUSLY DETERMINING THE ASSET ALLOCATION AND CONTRIBUTION RATE

JOHN BOARD and CHARLES SUTCLIFFE

The ICMA Centre, University of Reading, P.O. Box 242, Reading RG6 6BA, UK
e-mails: j.board@reading.ac.uk; c.m.s.sutcliffe@reading.ac.uk

Contents

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02021-7

Abstract

The trustees of funded defined benefit pension schemes must make two vital and interrelated decisions—setting the asset allocation and the contribution rate. While these decisions are usually taken separately, it is argued that they are intimately related and should be taken jointly. The objective of funded pension schemes is taken to be the minimization of both the mean and the variance of the contribution rate, where the asset allocation decision is designed to achieve this objective. This is done by splitting the problem into two main steps. First, the Markowitz mean-variance model is generalized to include three types of pension scheme liabilities (actives, deferreds and pensioners), and this model is used to generate the efficient set of asset allocations. Second, for each point on the risk-return efficient set of the asset–liability portfolio model, the mathematical model of Haberman (1992) is used to compute the corresponding mean and variance of the contribution rate and funding ratio. Since the Haberman model assumes that the discount rate for computing the present value of liabilities equals the investment return, it is generalized to avoid this restriction. This generalization removes the trade-off between contribution rate risk and funding ratio risk for a fixed spread period. Pension schemes need to choose a spread period, and it is shown how this can be set to minimize the variance of the contribution rate. Finally, using the result that the funding ratio follows an inverted gamma distribution, shortfall risk and expected tail loss are computed for funding below the minimum funding requirement, and funding above the taxation limit. This model is then applied to one of the largest UK pension schemes—the Universities Superannuation Scheme.

Keywords

pension scheme, portfolio theory, asset–liability modeling, contribution rate risk, solvency risk

JEL classification: G11, G23

In a funded defined benefit scheme the employer and employees both make contributions to a fund which is invested to provide the pension, and any other benefits due under the scheme. The benefits received under such schemes are defined in advance, usually as a proportion of the employee's final salary. Many UK companies have recently chosen to close their defined benefit pension schemes. In the 5.5 years up to February 2003, 63% of UK final salary schemes were closed to new entrants, while an additional 9% of schemes were also closed to future accruals (Association of Consulting Actuaries, 2003). The reasons given for closure include the introduction of Financial Reporting Standard 17, the substantial deficits on final salary schemes (caused by the fall in interest rates, the major stock market decline after the peak in December 1999, the extended contribution holidays and contribution reductions for employers, increases in benefits, the conversion of discretionary benefits into non-discretionary benefits, the use of pension schemes to finance early retirement on very favorable terms, and the tax limit on scheme surpluses); the effective move from limited price indexation to fully indexed pensions, with the fall in annual increases in RPI to below 5% since July 1991; the regulatory burden of administering these schemes; the increased cost due to rising life expectancy; the increased size of pension liabilities, relative to the size of the employer; the increase in stock market volatility; the risks that such schemes impose on employers (e.g., the risk that the fund will be insufficient to pay the pensions, the credit rating of the employer may be reduced because of the possibility of pension shortfalls); the abolition of tax relief on dividends from UK companies in 1997; the changes in actuarial technique leading to more volatile surpluses; the risk that new legislation or decisions by the law courts will increase the liabilities; the lower priority given to retaining staff; the opportunity to establish defined contribution schemes with a lower cost to the employer, and the much greater portability of defined contribution schemes.

This paper develops an approach to the simultaneous analysis of two critical and interrelated decisions which must be made by any fund's trustees: the fund's asset allocation and its contribution rate. The model developed in this paper is applicable to a wide range of pension schemes, and is illustrated with reference to a particular very large pension scheme—the Universities Superannuation Scheme.

Some previous authors have used multi-period stochastic programming (MPSP) to analyse the investment and contribution rate decisions of defined benefit pension schemes (Bogentoft, Romeijn and Uryasev (2001); Dert (1998); Drijver, Klein Haneveld and Van Der Vlerk (2002, 2003); Gondzio and Kouwenberg (2001); Hilli et al. (2005); Kouwenberg (1997, 2001); and Mulvey, Simsek and Pauling (2003)).[1] While

[1] See Ziemba (2003) for an introduction to the application of multi-period stochastic programming to asset–liability management. Geyer et al. (2005) developed a multi-period stochastic programming model for both defined benefit and defined contribution schemes, which was applied to the Siemens AG Österreich defined contribution scheme. Other papers in related areas include Cariño et al. (1994), Cariño and Ziemba (1998), Cariño, Myers and Ziemba (1998) and Ziemba (2006). Fabozzi, Focardi and Jonas (2005) surveyed 28 pension schemes in the UK, USA, Netherlands and Switzerland with average assets of $16 billion. One third of these schemes used MPSP.

MPSP permits the relaxation of many of the assumptions required by other methods, it requires extensive model building, has large data requirements and, until recently, has been difficult to solve. Dynamic stochastic control theory was applied to a small Swiss pension scheme by Dondi et al. (2006), while Rudolf and Ziemba (2004) applied stochastic control theory to a hypothetical US example. However, stochastic control theory requires the solution of nonlinear problems, assumes that the portfolio is constantly revised, can generate very large short and long positions, and may require big changes in asset proportions from period to period, see Ziemba (2003). A third approach that has been applied to analyse the investment and contribution rate decisions of defined benefit pension schemes is stochastic simulation (Boender (1997); Boender, Van Aalst and Heemskerk (1998); Boender and Vos (2000); Boender, Dert and Hoek (2006); Haberman et al. (2003); Kingsland (1982); Mulvey et al. (2005); Mulvey, Gould and Morgan (2000); Mulvey and Thorlacius (1998); and Wright (1998)). Although simulation models are flexible, they do not generate optimal decisions and require considerable effort to formulate. However, they are useful to check the validity of more complex models. Fourth, Frankfurter and Hill (1981) developed a multi-period linear programming asset–liability model which minimizes the present value of the contributions. However, it approximates the nonlinearity introduced by risk, does not generate a risk-return frontier and treats the liabilities as certain. Finally, Tepper (1974) used stochastic dynamic programming to minimize the present value of the contributions, but did not include either asset or liability risk.

This paper proposes a different methodology based on mean-variance portfolio theory, which is well understood, has modest data requirements and is both general and simple to apply. This makes the methods used in this paper straightforward to operationalize, while still jointly optimizing the asset allocation and contribution rate decisions. An enhanced portfolio model, which includes the scheme's liabilities (subdivided into active members, deferred pensioners and pensioners) as well as its chosen assets, is solved to generate efficient asset allocations.

For efficient portfolios, a generalization of the mathematical model of Haberman (1992) is used to compute the implied mean and variance of the contribution rate and funding ratio (i.e., the ratio of the fund value to its actuarial liabilities). The extension of the Haberman model is critical as we are able to relax a major inflexibility of the original model to allow the discount rate used in the actuarial calculations to differ from the expected investment return, and thus to model contribution rates in a way which conforms to finance theory. This generalization also removes any trade-off between contribution rate risk and funding ratio risk (for a fixed spread period), as one is simply a linear function of the other. It also removes the need to recompute the actuarial valuation of the liabilities as the asset allocation changes. We then further enhance the model to allow an investigation of the choice of the spread period used to adjust the contribution rate. This mathematical model is also used to estimate the distribution of the funding ratio, and to investigate the regulatory and solvency risk implied by the asset allocation and contribution rate decisions.

Section 1 discusses why the asset allocation and contribution rate decisions must be taken jointly. Section 2 presents the portfolio model of the asset–liability problem, while Section 3 shows how the means and variances of efficient asset–liability portfolios can be transformed into the means and variances of the contribution rate and funding ratio. Section 4 sets out the assumptions of the Haberman model. Section 5 investigates the issue of choosing the spread period, and Section 6 considers regulatory and solvency risk. Section 7 briefly describes the pension scheme studied—the Universities Superannuation Scheme; while Section 8 contains the data. Sections 9 and 10 contain the results for the portfolio model and the transformation of these results into contribution rates and funding ratios. Section 11 has the results for the optimal spread period, Section 12 considers the effect of triennial valuations, and Section 13 deals with regulatory and solvency risk. Section 14 concludes.

1. Linkage between the asset allocation and the contribution rate

The initial point for the analysis is the calculation of the actuarial liability of the fund. This calculation is divided into three parts, the value of the liability in respect of active members (those currently contributing to the fund before retirement) and those in respect of non-contributing members (deferred pensioners and pensioners). Eq. (1) sets out a very simple calculation of the actuarial liability for active members using the projected unit method.

The projected unit method "is now the natural method to use. . . . We see no strong reason to use any other method than the projected unit method for funding large schemes expected to have a continuing flow of new entrants". A survey found "that the majority of actuaries are now using the projected unit method", Thornton and Wilson (1992). FRS 17 requires the use of the projected unit method, while it is prescribed by Financial Accounting Statement 87 (Employers' Accounting for Pensions) issued in 1985 by the US Financial Accounting Standards Board. (However, the valuation method used for company accounts under FRS 17 could differ from that used in setting the contribution rate.) For the projected unit method, "the actuarial liability for active members either as at the valuation date or as at the end of the control period is calculated taking into account all types of decrement. In such calculations pensionable pay is projected from the relevant date up to the assumed date of retirement, date of leaving service or date of death as appropriate." Faculty and Institute of Actuaries (2003)

This paper uses a simple actuarial model. However, a very wide range of alternative actuarial models could be used without changing the main conclusions. In a fully specified model, additional terms would be included to allow for withdrawals, transfers in and out, deferment, death in service, early retirement, ill-health retirement, the option for a lump sum payment on retirement, etc. The formulae are based on Actuarial Education Company (2002).

$$AL_A = N_A \times \left(\frac{P \times S}{A}\right) \times \left\{\frac{(1+e)}{(1+h)}\right\}^{R-G} \times \left\{\left[1 - \left(\frac{(1+h)}{(1+p)}\right)^{-W}\right] \bigg/ \left(\frac{(1+h)}{(1+p)} - 1\right)\right\}, \quad (1)$$

where

AL_A is the actuarial liability for the active members of the scheme,
P is the average member's past years of service as at the valuation date,
S is the average member's annual salary at the valuation date,
A is the accrual rate,
e is the forecast nominal rate of salary growth per annum between the valuation date and retirement,
h is the nominal discount rate between now and retirement, and is assumed equal to the expected investment return on the assets for this period,
R is the average member's forecast retirement age,
G is the average age of the member at the valuation date,
W is the life expectancy of members at retirement,
p is the rate of growth of the price level, and
N_A is the current number of active members of the scheme.

The final term in Eq. (1) is the capital sum required at time R to purchase an index-linked annuity of £1 per year.

A simple model for the computation of the actuarial liability for pensioners is

$$AL_P = N_P \times PEN \times \left\{\left[1 - \left(\frac{1+h}{1+p}\right)^{-q}\right] \bigg/ \left(\frac{1+h}{1+p} - 1\right)\right\}, \quad (2)$$

where

AL_P is the actuarial liability for pensioners,
N_P is the current number of pensioners,
PEN is the average current pension

and the final term is the capital sum required now to purchase an index-linked annuity of £1 per year for the life expectancy, q, of pensioners. Adjustments to this simple model are required for dependents' pensions, death lump sum, etc.

A similar expression for the liability of deferred pensioners is

$$AL_D = N_D \times \left(\frac{P_D \times S_D}{A}\right) \times \left\{\frac{(1+p)}{(1+h)}\right\}^{R-G} \times \left\{\left[1 - \left(\frac{1+h}{1+p}\right)^{-W}\right] \bigg/ \left(\frac{1+h}{1+p} - 1\right)\right\}, \quad (3)$$

AL_D is the actuarial liability for the deferred pensioners of the scheme,
N_D is the current number of deferred pensioners of the scheme,
S_D is the average deferred pensioners' leaving salary, compounded forwards to the valuation date at the inflation rate (p), and
P_D is the average deferred pensioner's past years of service as at the valuation date.

The total actuarial liability (AL_T) is

$$AL_T = AL_A + AL_P + AL_D \tag{4}$$

which is the sum of the actuarial liabilities for every active member, pensioner and deferred pensioner. The precise form of the actuarial computations in Eqs. (1)–(3) is irrelevant for the model developed below for setting the asset allocation and the contribution rate.

The trustees must invest the funds to ensure that the scheme is able to meet its liabilities. To do this they make the asset allocation decision, which involves setting the proportions of the fund invested in different classes of asset. Classes of asset might include domestic equities, foreign equities, domestic gilts, domestic index linked gilts, foreign bonds, property, cash, private equity, commodities, etc. Because it is generally accepted that asset classes with higher expected returns also have higher risks (Dimson, Marsh and Staunton, 2002; Cornell, 1999; Constantinides, 2002; and Siegel, 2002), the asset allocation has an important effect on both the risk and return of the fund. While the selection of specific stocks, bonds or properties may also be important in determining the investment performance of the fund, it is not usually possible for the trustees to become involved in this level of detail, and so the asset selection decision is usually delegated to fund managers. This delegation can be further justified by the evidence that the main determinant of investment performance for UK and US pension funds is asset allocation, rather than asset selection (Blake, Lehmann and Timmermann, 1999; Brinson, Hood and Beebower, 1986; Brinson, Singer and Beebower, 1991; and Ibbotson and Kaplan, 2000).

The trustees must also determine the employer's and employees' contribution rates. The employees' contribution rate (the percentage of their salary that each employee must pay to the pension scheme), is usually constant. In contrast, the employer's contribution rate is set (or proposed) by the pension scheme's trustees, and is periodically reappraised. The modified (or recommended) contribution rate is equal to the standard (or normal) contribution rate plus or minus a contribution rate adjustment to correct for any difference between the actual and target funding level of the scheme. The contribution rate adjustment can be computed in a variety of ways. In the UK the commonly used methods are the spread (or percentage of pay) method, the mortgage method and the straightline method. The US and Canada use the amortization of losses method. Because the employees' rate is usually constant, any change to the overall contribution rate made by the trustees will result in a change to the employer's rate. Obviously, increasing the contribution rate has a direct effect on the fund's value, while increasing employment costs to the employer.

Using the projected unit method, the standard contribution rate is defined by the Actuarial Education Company (2002) as "the present value of all benefits that will accrue in the year following the valuation date (by reference to service in that year and projected final earnings) divided by the present value of all members' earnings in that year”. The standard contribution rate (SCR) is

$$SCR = \frac{AL_A}{P \times N_A \times S \times a_{\overline{1}|}} + AE, \tag{5}$$

where

$a_{\overline{1}|}$ is an annuity to give the present value of earnings by the member over the next year, and

AE is the administrative expenses of the scheme, expressed as a proportion of the current salaries of the active members.

The asset allocation and contribution rate decisions are interrelated as both affect the level and volatility of the contribution rate and the value of the fund. Throughout this paper assets are valued using current market prices. Actuaries can use other methods of valuation (e.g., the dividend discount model) which tend to smooth out variations in the value of the fund and contribution rate. But the actuarial profession is adopting market values, and smoothing the value of the fund and the contribution rate by ignoring changes in market value is diminishing in importance. If the scheme chooses an equity tilt in its asset allocation, in the expectation that this will increase returns on the fund, the average contribution rate may be reduced. However, an equity tilt will increase the volatility of the fund’s returns. The degree of over or under funding of the scheme will also tend to be volatile, and this will increase the volatility of the contribution rate. The extent to which the volatility of an equity tilt feeds through to the contribution rate depends on the way in which the contribution rate is adjusted. For these reasons, the asset allocation and contribution rate strategies need to be considered jointly. Haberman et al. (2003) have also argued that the funding and investment strategies of a pension scheme should be considered jointly. In essence, the trustees choose the level and variance of the contribution rate which they prefer; and this then determines the asset allocation.

2. A multi-period portfolio model of the asset–liability problem

When portfolio models are applied to assets, the conventional objective is to maximize the return for a given level of risk; or minimize the risk for a given level of return. Previous models of pension schemes have used a range of objectives reflecting risk and return. The risk measures used include the minimization of the variance of the contribution rate, the variance of fund value, the variance of the funding ratio, and solvency risk (defined in a variety of ways). While the aim for an asset portfolio is to maximize its returns, the objective of a pension scheme is to minimize its cost. Therefore the “return” measures used by previous studies include the minimization of the expected contribution rate, the minimization of the present value of total future contributions, and the maximization of expected utility. This study minimizes the contribution rate and its variance.

Pension schemes have liabilities that may fall due up to sixty years (the life expectancy of a young academic) in the future, and so face a multi-period portfolio problem. Although no general solution to the multi-period portfolio problem exists, it can be solved if some additional assumptions are made. A number of authors including Hakansson (1970, 1971), Mossin (1968) and Campbell and Viceira (2002, pp. 33–35) have noted that, if portfolio returns are expected to be stationary over time (that is, returns are independently and identically distributed, or i.i.d.) and have a normal distribution, the investor's attitude to risk is wealth independent, and all dividends are immediately reinvested; then the problem is stationary, and the one-period solution is also the multi-period solution. If some aspect of the problem changes, the model can easily be re-solved. Since the contribution rate is usually fixed for three years, the asset allocation decisions in the second and third years are constrained to generate portfolios with risks and returns that are similar to those of the initial portfolio chosen in the first year of each triplet.

The strong assumption of normal i.i.d. returns is widely accepted and generally works reasonably well (it is, for example, made in the derivation of the Black–Scholes option pricing model). While asset returns can be approximated by a normal distribution, this is less clear for liabilities. If the maturity of the scheme is changing over time, the correlation between the scheme's liabilities and the various asset classes will also change. However, since the liabilities will be disaggregated into active members, pensioners and deferred pensioners; a change in scheme maturity need not change the correlations used in the model. The assumption of wealth independence fits with the evidence that the risk premium has not trended up or down over the last century as society has become much richer, and with the fact that pension schemes are organizations with an infinite life that do not themselves consume goods and services. Black (1995) refers to pension schemes as "conduits". Finally, the immediate re-investment of all dividends is common practice. Therefore, while the assumptions underlying myopia and the use of a one-period model are simplifications, they appear to offer a reasonable approximation to reality.

Pension schemes have liabilities to present and future pensioners, and the purpose of the pension fund is to meet these liabilities. To allow for liability risks, the portfolio model used to determine the asset allocation is modified by the inclusion of scheme liabilities (Sharpe and Tint, 1990; Sharpe, 1990; Ezra, 1991).[2] Instead of viewing the pension scheme as a separate entity, it can be treated as an integral part of the employer. In which case the portfolio problem includes not only the assets and liabilities of the pension scheme, but also the assets and liabilities of the employer (Bagehot, 1972). Chun, Chiochetti and Shilling (2000) and Craft (2001, 2005) have applied the Sharpe–Tint model to US corporate pension funds. The returns on shares in some employers may be highly correlated with those of a particular industrial sector. In which case the portfolio allocation decision of the fund should make allowance for this situation. However, the very large public sector pension scheme studied below (USS) has no such problems, and so this feature is not incorporated into the model.

[2] Waring (2004) deflated by the liabilities, rather than the assets; and expressed the risk and return of the asset–liability portfolio in terms of the alphas and betas of the capital asset pricing model. Nijman and Swinkels (2003) applied the Sharpe–Tint model, but with the simplification that the importance of the liabilities (k) equals the initial funding ratio.

Since the value of the liabilities is assumed to be unaffected by the asset allocation of the fund, the portfolio problem can be stated in terms of the mean and variance of returns on the fund; but with the addition of a term for the covariances of returns on each asset class with the liabilities. There is no explicit consideration of matching the duration of the assets and liabilities. However, if assets with a range of durations are included, the portfolio model implicitly takes duration matching into account.

The model of Sharpe and Tint (1990) is extended by disaggregating pension fund liabilities into three components (active members, deferred pensioners and pensioners), where each of these components has different correlations with the various asset classes. Pensions in payment can take the form of a fully index-linked annuity when pension increases are linked to the retail price index (RPI), while deferred pensions are usually based on final salary, indexed to the retirement date for subsequent increases in the RPI. Index-linked gilts are likely to represent a good match for such liabilities. Full price indexation and an absence of deflation is assumed so that there are no limited price indexation complications.

The size of the pension that will be received by active members depends on their final salaries; and other asset classes are likely to provide a better match for this salary risk than UK government bonds (gilts). The model assumes, as does Haberman's (1992) model, which is discussed below, that the growth rates of total benefits and total contributions are non-stochastic. If the growth rates of total benefits and contributions are stochastic, the variances of the portfolios produced by the Sharpe and Tint model must be expanded to include the correlations between wages and the assets, and between benefits and the assets, see Yang (2003).

The expanded portfolio model, including different types of liability, is

$$\text{Minimize} \quad V_{al} = \sum_{i=1}^{N+B} \sum_{j=1}^{N+B} x_i x_j V_{ij} \tag{6a}$$

$$\text{subject to:} \quad E_a = \sum_{i=1}^{N} x_i E_i, \tag{6b}$$

$$\sum_{i=1}^{N} x_i = 1, \tag{6c}$$

$$x_i \geqslant 0, \quad i = 1 \dots N, \tag{6d}$$

$$x_i = w_i, \quad i = N+1 \dots N+B, \tag{6e}$$

where

V_{al} is the variance of the asset–liability portfolio,

i and j represent asset or liability classes,

N is the number of assets and B the number of liabilities,

V_{ij} are covariances of returns between asset or liability classes i and j,

E_i and E_a are the expected arithmetic returns on asset or liability class i and the chosen asset portfolio, respectively,

w_i are the initial portfolio proportions of the B types of scheme liability, which are assumed fixed. Thus, for three types of liability, $w_1 = -L_0^1/A_0$, $w_2 = -L_0^2/A_0$, and $w_3 = -L_0^3/A_0$, where L_0^1 represents the current liability to active members, L_0^2 is the current liability to deferred pensioners, L_0^3 represents the current liability of pensions in payment, and A_0 is the current value of the fund's assets, and

x_i are the investment proportions in each of the $N + B$ asset or liability classes.

An efficient frontier can be constructed by repeatedly solving this quadratic programming problem for a range of required expected returns on the portfolio of assets held, E_a. Short selling is excluded by (6d) because pension schemes choose not to engage in this activity. The exclusion of short selling (and of borrowing money) has important implications for the optimal asset allocation (Sutcliffe, 2005). Because the liability proportions are fixed, the returns on the liabilities and the covariances between returns on different liabilities play no part in determining the asset proportions of the efficient frontier. The returns on the liabilities are the proportionate changes in value of the liabilities during the period. Liability returns may be due to changes in accrued years, the number of members and pensioners, the level of salaries and the RPI, variations from the actuary's demographic assumptions, and, most importantly, changes in the discount rate.

A continuous time model for the asset allocation decision of defined benefit pension schemes, based on the Sharpe and Tint (1990) model and Merton (1992), was derived by Rudolf and Ziemba (2004). Their model has four-fund separation, with investors determining their optimal weights across these four funds. The objective is to maximize the intertemporal scheme surplus, and Rudolf and Ziemba show that the proportion of the scheme's assets invested in securities providing a hedge for its liabilities should be equal to a constant (which is a linear function of the asset and liability covariances), divided by the funding ratio. Therefore, the proportion of the fund invested in assets hedging the liabilities is independent of preferences; and becomes lower as the funding ratio rises.

3. Transformation of the portfolio returns to contribution rates and funding ratios

MacBeth, Emanuel and Heatter (1994) report than trustees find it much easier to make judgments about contribution rates and funding ratios than about return distributions. Since the asset allocation decision should be taken simultaneously with the contribution rate decision, it is helpful to respecify the objective from a mean-variance analysis of returns to using the mean and variance of the contribution rate and funding ratio as the criteria. Haberman (1997b) observes that there is a difference between the variance of the present value of all future contributions, and the long-run variance of contribution rates. The usual choice, which is followed in this paper, is the long-run variance of contribution rates.

Beginning with the work of Dufresne (1986, 1988, 1989, 1990a, 1990b), mathematical expressions have been derived for the first two moments of the contribution rate and the funding ratio. These models provide formulae for the mean and variance of the total value of contributions and the total value of the fund. However, if AL_T and Q (the total value of annual salaries currently paid to active members) are fixed, it is more convenient to work with the mean and variance of the contribution rate and the funding ratio. A series of papers have developed and elaborated this approach: Bédard (1999); Booth et al. (1999); Cairns (1995, 1996a, 2000); Cairns and Parker (1997); Chang and Chen (2002); Gerrard and Haberman (1996); Haberman (1990b, 1992, 1993a, 1993b, 1994a, 1994b, 1995, 1997a, 1997b, 1998); Haberman, Butt and Megaloudi (2000); Haberman and Dufresne (1991); Haberman and Owadally (2001); Haberman and Wong (1997); Mandl and Mazurová (1996); Owadally and Haberman (1999, 2000); and Zimbidis and Haberman (1993).

Some studies have used stochastic control theory to investigate the effects of allowing the asset proportions in the risky and riskless assets to be altered over time according to some assumed rule (Boulier, Trussant and Florens, 1995; Boulier, Michel and Wisnia, 1996; Cairns, 1996b, 1997; Bédard and Dufresne, 2001; Josa-Fombellida and Rincón-Zapatero, 2001; and Rudolf and Ziemba, 2004). This usually involves modeling a hypothetical pension scheme with two classes of asset, one risky and one risk-free, and a riskless liability; to derive expressions for the mean and variance of both the value of contributions and the value of the fund for combinations of the following aspects of the problem:

- Spread period—the use of the spread method or the amortization of losses method when setting the number of years over which the over or under-funding of the scheme is to be eliminated.
- Returns distribution—the assumption of i.i.d. returns, first- or second-order autoregressive returns, or first-order or second-order moving average returns on the investments of the scheme.
- Funding method—the use of individual or aggregate funding methods by the actuary when valuing the liabilities of the scheme and setting the standard contribution rate.
- Lagged adjustments—the presence of a lag of zero, one or more years in revising the contribution rate.
- Valuation timing—annual or triennial actuarial valuations of the scheme.

Many different funding methods have been developed to compute the contribution rate and funding ratio, among them are the attained age, entry age, projected unit and current unit methods. The choice of funding method affects the level and stability of the contribution rate. For example, the entry age method produces a stable contribution rate over the life of each member, and if the distribution of entry ages and sexes remains equal to those assumed, the contribution rate for the scheme is constant over time. Similarly, if the forecast return on investments exceeds the forecast rate of salary growth, then the contribution rate generated by the projected unit method is a positive linear function of the member's age. If the age, sex and salary distribution of members remains constant, this method also produces a stable contribution rate for the scheme.

The model which is closest to the circumstances of many large UK pension schemes is that of Haberman (1992). Among this model's assumptions are that the scheme uses the spread method for adjusting the contribution rate. The spread and the amortization of losses methods have been compared by Cairns (1995, 1996a), Haberman (1998), Haberman and Owadally (2001), and Owadally and Haberman (1999, 2000). The minimum variance of the contribution rate that can be achieved using the spread method is below that achievable using the amortization of losses method. In addition, for a given variance of the funding ratio, the corresponding variance of the contribution rate is lower for the spread method. Therefore, the spread method is preferable on the grounds of giving lower variances for both the contribution rate and the funding ratio.

It is also assumed that the valuation, or discount, rate is certain and equal to the expected rate of return on investments. The use of the return on the assets as the discount rate is permitted by SSAP 24 (ASB, 1988), and has been in widespread use by actuaries for many years. Recently other discount rates have been suggested—long-term bond yields, bond yields plus a risk premium, and returns on a portfolio that replicates the liabilities, Faculty and Institute of Actuaries (2003) and Exley, Mehta and Smith (1997). FRS 17 (ASB, 2000) proposes that the return on the matching portfolio be proxied by the return on AA grade corporate bonds. If the return on the assets is higher than these alternatives, its use as the discount rate reduces the actuarial liability and the expected contribution rate. Therefore the contribution rates given by the Haberman (1992) model are usually lower than those produced by a model using the return on a liability matching portfolio as the discount rate.

Additional assumptions are that returns are i.i.d., an individual funding method (e.g., the projected unit method) is in use, actuarial valuations are annual, there is a lag of one year in adjusting the contribution rate after each actuarial valuation, there are no benefit improvements (other than full price indexation), the target funding ratio is 100%, the demographic assumptions of the actuary are realized, scheme membership is stationary in size and structure and the rate of salary growth is constant and certain. The model shows that, in these circumstances, the actuarial liability and the standard contribution rate are constant over time, and the average funding ratio is 100%.

Although the Haberman (1992) model assumes that the size of the scheme is constant, scheme growth need not affect the standard contribution rate computed using the projected unit method if the age, sex and salary distribution of members remains unchanged. To allow for growth in salaries and benefits, the Haberman (1992) model uses the deflated investment return (v_a)

$$1 + v_a = \frac{1 + E_a}{1 + e}, \tag{7}$$

where the rate of salary growth between now and retirement is assumed to increase at the same rate as benefits. To ensure stationarity, Haberman also assumes that $e = p$, and that there is no promotional scale. These restrictions are not imposed on the actuarial models in Eqs. (1) to (5).

If expected returns are to be deflated by earnings growth, it follows that the variance of the asset–liability portfolio should also be deflated, and so σ_{al}^2 is

$$\sigma_{al}^2 = \frac{V_{al}}{(1 + e)^2}. \tag{8}$$

4. Relaxing the assumptions of the Haberman (1992) model

The most restrictive assumption of the Haberman (1992) model is that the discount rate equals the rate of return on investments. This assumption, which is also widely used in other actuarial models of pension schemes, has the strange consequence that, by investing in a high-risk high-return portfolio of assets, the liabilities of the scheme get smaller. To avoid making this undesirable assumption, we generalize the Haberman model to allow the discount rate to differ from the investment return. Full details of the generalization are available from the authors, but it follows the similar generalization of the Dufresne (1988) model presented by Cairns (1995, 1996a). As well as improving the economic realism of the model, this generalization greatly simplifies its empirical application. This is because the actuarial liability is unaffected by the asset allocation decision, obviating the need to re-compute the actuarial liability for every asset allocation with a different rate of return. When the investment rate of return is used to estimate h (as in Haberman, 1992) the actuarial liabilities change as the asset allocation changes, and the actuarial formulae for the computation of the liabilities are necessary to operationalize this model. However, if the rate of return on a portfolio that matches the liabilities is used to estimate h (as in the generalized Haberman model), the return on the matching portfolio is invariant with respect to changes in the asset allocation, and the values of the actuarial liabilities are constant. In which case, the actuarial valuation of the liabilities is simply a fixed input number to the model. The generalized Haberman model also drops the requirement that the funding ratio be 100%.

The models of Haberman et al. assume that the liabilities are riskless. This implies there is no discount rate risk; and that the actuarial demographic assumptions such as longevity, withdrawals and early retirement are satisfied. The discount rate for liabilities in this paper is the riskless rate. However, when computing investment risk, the asset variance is replaced by the variance of the asset–liability portfolio. Since the portfolio model allows the liabilities to be risky, both discount rate risk and actuarial demographic risks are indirectly incorporated into the model.

The Haberman (1992) model assumes that actuarial valuations are annual, while most schemes have triennial valuations. As a result, the model tends to understate the true variance of the contribution rate and funding ratio. Triennial valuation could have been allowed for using the model of Haberman (1993b), but at the expense of assuming that the contribution rate was adjusted instantly on the date of the actuarial valuation. Cairns (1996a) concludes that a one year lag in adjusting the contribution rate has a much bigger effect on the variance of the contribution rate than does allowance for triennial valuations. We investigate the size of this effect by comparing the mean and variance of the contribution rate and funding ratio using the model of Dufresne (1988), which assumes annual valuations and instant revision of the contribution rate, with those obtained using the model of Haberman (1993b), which assumes instant revision of the contribution rate, but triennial valuations. Details of these models appear in Appendix A.

The Haberman model does not incorporate benefit improvements which may be granted when the funding ratio becomes strongly favorable, nor does it include any defaults which may occur when the funding ratio becomes very unfavorable. This is because the inclusion of such effects would considerably complicate the model.

While the Haberman model could be applied to determine the asset allocation of the pension schemes of companies, there are tax arbitrage arguments for such schemes simply selecting the asset allocation which minimizes the risk of the asset–liability portfolio (Ralfe, 2001; Ralfe, Speed and Palin, 2003; Sutcliffe, 2005). If these arguments are accepted, the Haberman model only applies to pension schemes whose employer does not pay tax. Among examples of such schemes in the UK are those run by local authorities, the British Broadcasting Corporation, the Universities Superannuation Scheme, the Church Commissioners, the Financial Services Authority, the Civil Aviation Authority, London Transport, British Coal, the Post Office and the Merchant Navy.

The following equations give the first two moments of contributions and the value of the fund for a given investment return and variance of asset–liability returns under the projected unit method using both the Haberman (1992) model and its generalized version, denoted respectively by the subscripts H and G. The expected value of the fund (F) is

$$E[F]_H = AL_T, \tag{9H}$$

$$E[F]_G = gAL_T, \tag{9G}$$

$$g = \frac{(1+v_a)(k+kd-d)}{(1+d)(k+v_a k - v_a)}, \tag{10}$$

$$1+d = \frac{1+d'}{1+e}, \quad \text{where } d' \text{ is the discount rate for liabilities,} \tag{11}$$

$$k = \frac{1}{\sum_{z=0}^{M-1}(1+d)^{-z}}, \tag{12}$$

i.e., k is the reciprocal of a compound interest rate annuity with a life of M years calculated at the rate d.

The expected modified level of contributions (C) is equal to the standard level of contributions (SC, where $SC = SCR \times Q$, Q is the total value of annual salaries currently paid to active members) plus an additional term in the case of the generalized model

$$E[C]_H = SC, \tag{13H}$$

$$E[C]_G = SC + kAL_A(1-g). \tag{13G}$$

When $g > 1$, the term $kAL_A(1-g)$ becomes negative, and if $|kAL_A(1-g)| > SC$, $E[C]$ becomes negative, and expected contributions are negative. A negative contribution rate is only possible if permitted by the scheme rules and sanctioned by the trustees; and a contribution holiday, i.e., $E[C] = 0$, is much more likely. The use of a contribution holiday, rather than negative contributions, means that the funding level of the scheme will tend to grow over time, and this conflicts with the assumption of the Haberman model that the scheme is in long run equilibrium. Therefore, the generalized Haberman model excludes situations where $E[C]_G < 0$.

The corresponding variances of F and C are

$$\mathrm{Var}[F]_H = AL_T^2 b, \tag{14H}$$

$$\mathrm{Var}[F]_G = AL_T^2 bg^2, \tag{14G}$$

$$\mathrm{Var}[C]_H = AL_A^2 k^2 b, \tag{15H}$$

$$\mathrm{Var}[C]_G = AL_A^2 k^2 bg^2, \quad \text{where} \tag{15G}$$

$$b = \frac{\sigma^2(1+uk)}{u^2[1+uk-(\sigma^2+u^2)(1-uk+k^2+uk^3)]}, \tag{16}$$

$$u = (1+v_a), \quad \text{and } \sigma^2 \text{ is the variance of } v_a.$$

These equations give the first two moments of the total levels of the value of the fund and annual contributions to the scheme. The equivalent numbers for the funding ratio (*FR*) and the contribution rate (*CR*) for the Haberman (1992) and the generalized Haberman models are

$$E[FR]_H = 1, \tag{17H}$$

$$E[FR]_G = E[F]_G/AL_T = g, \tag{17G}$$

$$E[CR]_H = SC/Q, \tag{18H}$$

$$E[CR]_G = E[C]_G/Q, \tag{18G}$$

$$\mathrm{Var}[FR]_H = b, \tag{19H}$$

$$\mathrm{Var}[FR]_G = bg^2, \tag{19G}$$

$$\mathrm{Var}[CR]_H = \mathrm{Var}[C]_H/Q^2, \tag{20H}$$ [3]

$$\mathrm{Var}[CR]_G = \mathrm{Var}[C]_G/Q^2. \tag{20G}$$

It can be seen from Eqs. (15), (19) and (20) that, for both the Haberman (1992) and the generalized Haberman models, the variance of the contribution rate is equal to the variance of the funding ratio multiplied by $AL_A^2k^2/Q^2$. For the Haberman (1992) model, AL_A varies as the investment return varies, and so the relationship between Var[*CR*] and Var[*FR*] is nonlinear. However, for the generalized Haberman model, AL_A is computed using a fixed discount rate, resulting in a constant proportional relationship between Var[*CR*] and Var[*FR*], i.e., $AL_A^2k^2/Q^2$. Therefore, for the generalized Haberman model

[3] When making this adjustment for the Haberman (1993b) model Q is increased to $3Q$ because this model deals with three year periods.

with a fixed spread period (but not the (Haberman, 1992), model) there is no trade-off between contribution rate risk and funding ratio risk.

In these equations, the values of v_a and σ^2 relate to a particular efficient portfolio generated by the quadratic programming model, AL_T, AL_A, Q and SC are the result of actuarial calculations, d is the deflated discount rate, and M is a policy variable.

5. The choice of the spread period

Although UK accounting rules require M, the number of years used in the spread method to be equal to the average future working life of the membership, actuaries are free to choose M. A number of researchers have examined the choice of M for computing the contribution rate adjustment: Bédard (1999); Booth et al. (1999); Cairns (1995, 1996a); Cairns and Parker (1997); Chang and Chen (2002); Dufresne (1986, 1988, 1989, 1990b); Haberman (1990a, 1993b, 1994a, 1994b, 1995, 1997a, 1997b, 1998); Haberman, Butt and Megaloudi (2000); Haberman and Dufresne (1991); Haberman and Wong (1997); and Owadally and Haberman (1999, 2000). Their models reveal that, as the spread period is lengthened, the variance of the contribution rate first decreases, but then, after a critical value of M, denoted M^*, begins to increase. In contrast, because they use the return on investments as the discount rate, the variance of the funding ratio increases monotonically with M.

For Haberman (1992) the optimal spread period M_H^* is:

$$M_H^* = -\log\bigl(1 - v_a/\bigl[(1+v_a)k_H\bigr]\bigr)/\log(1+v_a) \quad \text{for } v_a > 0, \tag{21H}$$

$$\text{where } k_H = \bigl[-(2-y) + \bigl(y(5y-4)\bigr)^{0.5}\bigr]/2u(1+y). \tag{22H}$$

Similarly, the optimal spread period for the generalized Haberman model, M_G^*, is:

$$M_G^* = \frac{-\log(1 - v_a/[(1+v_a)k_G])}{\log(1+v_a)} \quad \text{for } v_a > 0, \tag{21G}$$

where k_G is one of the solutions to the quintic equation

$$\begin{aligned}
&k_G^5(v_a y - dy)u^2 + k_G^4\bigl(2y\{v_a - d\}/u + y + v_a + d + v_a y + dy + dv_a + dv_a y \\
&\quad + 1\bigr)u^2 + k_G^3\bigl(2 + y\{v_a - d\}/u - 2uv_a + 2v_a + 2d - y - 2udv_a - v_a y \\
&\quad + 2dv_a - dv_a y + 2udv_a y - dy - 2uv_a y\bigr)u \\
&\quad + k_G^2\bigl(1 - 4udv_a + uv_a y + dyu - v_a y + dv_a - dy - 4uv_a - y + v_a + d \\
&\quad - dv_a y + u^2 dv_a + 2dv_a yu + u^2 dv_a y\bigr) \\
&\quad + 2k_G(y - d + ud + dy - dyu/2 - 1)v_a + dv_a(1-y) = 0,
\end{aligned} \tag{22G}$$

where $y = (1+v_a)^2 + \sigma^2$.

For regulatory and solvency reasons, the scheme may also be concerned about the variance of the funding ratio, which is a positive function of M, so that the more slowly any over or under-funding is eliminated, the higher will be the variance of the funding ratio.

Eqs. (21) and (22) reveal that M^* decreases as higher risk and higher return asset portfolios are chosen, and so a one-size-fits-all policy of determining the spread period separately from the fund's investment decision is inappropriate. The spread period used in adjusting the contribution rate is endogenous and should be selected in the light of the risk and return on the chosen portfolio. In the Haberman (1992) model, M is the only available contribution rate policy variable. However, in reality contribution rate policy can be more complex than always eliminating any over or under funding. While there is a linear relationship between contribution rate risk and funding ratio risk for the generalized Haberman model with a fixed spread period, this ceases to be the case when the spread period is endogenous.

6. Regulatory and solvency risk

Although the discussion in this section is couched in terms of UK legislation, the arguments are general and will apply in most regulatory environments. UK legislation places upper and lower limits on the funding ratio of pension schemes. Under the Pensions Act 1995,[4] a scheme which is less than 90% funded on the minimum funding requirement, MFR, basis, must be returned to 90% funding within three years, and to 100% funding within ten years. Under the Finance Act 1986, Schedule 13, Part 2, schemes which are more that 105% funded, on the prescribed valuation basis,[5] must be reduced to below 105% over the next five years.[6] The likelihood of breaching these requirements must, therefore, be considered when making the asset allocation and contribution rate decisions.

Value at risk (VaR) is a popular risk measure for financial institutions. It gives the estimated maximum loss that can occur over a stated time horizon for a chosen probability, λ; and so the probability that the actual loss exceeds the specified VaR is $(1-\lambda)$, which is called the shortfall probability (*SP*). However, although VaR is useful, it suffers from some serious theoretical and applied difficulties. For example, it does not satisfy the coherency axioms of Artzner et al. (1999). Some of these difficulties are overcome by using the expected tail loss, *ETL*, to quantify the effects of breaching the solvency and regulatory constraints. The *ETL* is also known as the conditional value at risk (CVaR), the mean shortfall, the mean excess loss or tail VaR. The *ETL* has been applied to pension schemes by Bogentoft, Romeijn and Uryasev (2001); while there is also a literature

[4] The rules are specified in the Occupational Pension Schemes (Minimum Funding Requirement and Actuarial Valuations) Regulations 1996; as amended by the Occupational Pension Schemes (Minimum Funding Requirement and Miscellaneous Amendments) Regulations 2002.

[5] The valuation basis is specified in the Pension Scheme Surpluses (Valuation) Regulations 1987, Statutory Instrument no. 412.

[6] This upper limit will shortly be abolished, as will the MFR.

on the use of shortfall risk in the context of pension schemes (see Leibowitz, Bader and Kogelman, 1996 and Haberman et al., 2003).[7]

The *ETL* computes the expected size of any breach of the funding requirements which exceeds the specified VaR. In the present context, two VaR values, representing the regulatory restrictions on the maximum and minimum funding ratio, are of interest. Breaches of the upper regulatory constraint can be analyzed in the same way as breaches of the lower constraint, except that breaches are greater, rather than smaller than the specified VaR. While the VaR is defined as a loss, it is convenient in the present circumstances to treat the VaR as a specified funding ratio, rather than a deviation from the specified upper or lower bound. Such *ETL*s have been termed conditional tail expectations. As well as the *ETL*s, the probability of a particular funding ratio breaching each of the regulatory limits (i.e., $(1-\lambda)$, the shortfall probability, *SP*) is also computed.

The computation of the *ETL*s and *SP*s requires a knowledge of the probability distribution of the funding ratio, and this probability distribution has been studied by Dufresne (1990b), Cairns (1995, 1996b, 1997, 2000), and Cairns and Parker (1997). Cairns (1995) concludes that, in discrete time, the inverted gamma distribution (also known as the reciprocal or inverse gamma distribution, and the Pearson Type *V* distribution), provides a very good approximation to the distribution of the fund value in a wide variety of cases, and so this distribution is used to compute cumulative probabilities for the funding ratio. Using the results in Evans, Hastings and Peacock (1993) and Johnson et al. (1994), it can be shown that the reciprocal of the funding ratio has a two parameter gamma distribution with parameters α (shape) and β (scale), where

$$E[FR] = 1/\beta(\alpha-1), \tag{23}$$

$$\mathrm{Var}(FR) = 1/\beta^2(\alpha-1)^2(\alpha-2). \tag{24}$$

From this, the two parameters of the gamma (and inverted gamma) distribution can be obtained from the mean and variance of the funding ratio of the generalized Haberman model:

$$\alpha = \left\{\left(E[FR]\right)^2/\mathrm{Var}(FR)\right\} + 2, \tag{25}$$

$$\beta = \mathrm{Var}(FR)/\left\{E[FR]\left[\left(E[FR]\right)^2 + \mathrm{Var}(FR)\right]\right\}. \tag{26}$$

The probability density function (*PDF*) of the inverted gamma distribution of t is equal to $1/t^2$ times the *PDF* of the gamma distribution of $1/t$. The probability density function for the inverted two parameter gamma distribution for the variable t is:

$$p(t) = t^{-(\alpha+1)}\mathrm{e}^{-1/t\beta}/\beta^\alpha\Gamma(\alpha), \quad \text{where } \alpha > 0 \text{ and } \beta > 0, \tag{27}$$

[7] For stochastic programming models, Kusy and Ziemba (1986), Cariño et al. (1994), Cariño and Ziemba (1998), and Cariño, Myers and Ziemba (1998) proposed the use of a convex tail risk measure which leads to a concave utility function that can be modeled using a piecewise linear approximation. This convexity means that shortfalls from target levels are penalized more and more as the shortfall increases.

where the term $\Gamma(\alpha)$ is the gamma function which acts as an adjustment factor to ensure that the probabilities sum to one. The *SP* and the *ETL* for over-funding are computed by integrating the probability density function from zero to the chosen value of $1/t$; while the corresponding figure for an under-funding involves integration from the chosen value of $1/t$ to infinity. The cumulative density function (*CDF*) of the inverted gamma distribution of t (i.e., inverted $CDF(t, \alpha, \beta)$) is equal to one minus the *CDF* for the gamma distribution of $1/t$ (i.e., $1 - CDF(1/t, \alpha, \beta)$).

The *ETL*s can be approximated to any degree of accuracy by computing the average of the VaR values throughout the tail (i.e., for all losses greater than the specified VaR). Let λ represent the chosen probability for the specified VaR, and divide the tail beyond this VaR into ϕ parts. Let $\lambda_i = \lambda + i(1-\lambda)/\phi$ where $i = 0 \ldots (\phi - 1)$, and then, for each value of λ_i, the cumulative inverted gamma distribution is used to find the corresponding value of the funding ratio. The *ETL* is then the equally weighted arithmetic average of these funding ratios. Thus:

$$ETL = \frac{1}{\phi} \sum_{i=0}^{\phi-1} \mathrm{VaR}(\lambda_i) \tag{28}$$

where $\mathrm{VaR}(\lambda_i)$ is the VaR corresponding to a probability of λ_i. The accuracy of this method is reported to be reasonably good for values of $\phi > 50$ (Dowd, 2002).

7. Description of the Universities Superannuation Scheme

The asset–liability model described above was applied to the Universities Superannuation Scheme (USS). USS was created in 1974 as the main pension scheme for academic and senior administrative staff in UK universities and other higher education and research institutions (Logan, 1985). From 10 December 1999, the rules of USS were changed to allow any employee, non-academic or academic, at any higher education institution (or associated establishment) in the UK to become a member. By 2002 there were over 300 institutional members (i.e., employers) participating in USS, which was the third largest pension fund in the UK, with assets of £20 billion, and 180,000 members, pensioners and deferred pensioners. USS is a defined benefit scheme with an 80^{ths} accrual rate and is managed by the trustee company, USS Ltd. The employers' contribution rate in 2002 was 14% of salary, while the employee contribution rate was 6.35%. The principal benefits are an index-linked pension and a tax-free lump sum on retirement, ill-health retirement with an index-linked pension and tax-free lump sum, and index-linked pensions for spouses and dependents on the death of the member or pensioner. USS is an immature scheme with a net cash inflow of £550 million in 2002. While the maturity of the scheme will probably increase in the future, it is expected to have a positive cash flow for many years. The USS actuary uses the projected unit method, which is an individual funding method, and triennial valuations.

8. Data

The numerical results presented below relate to the application of the asset–liability model described in this paper to USS, using data for its 2002 actuarial valuation (Universities Superannuation Scheme, 2003). This actuarial valuation allows the calculation of the values of the initial liability proportions as $w_1 = -0.5523121$, $w_2 = -0.0642698$, and $w_3 = -0.3788936$. The valuation of the liabilities was approximately equivalent to a buyout valuation. The trustees of the fund allocate its assets between five principal asset classes: UK equities, overseas equities, property, fixed interest and UK index-linked gilts. In addition, three types of liability are recognized—active members, deferred pensioners and pensioners. Table 1 shows the modest data requirements of this model; with only 25 correlation forecasts (of which 15 involve pension liabilities), 8 forecasts of the expected returns, and 8 forecasts of the standard deviations of returns. The forecasts used in generating the subsequent illustrative results are based on annual data for the period 1981–2002—the IPD index of total returns on all UK property, the MSCI World ex UK total returns index in £, the FTSE All Share index including dividends, yields on long term UK Government bonds supplied by Datastream, and UK index linked gilt yields for a constant maturity. Since the assets are proxied by indices, to the extent that the fund engages in active management, as opposed to tracking the index, the risk is understated and returns and correlations altered.

The liability data was constructed using the top point on the lecturer scale together with the simple actuarial models in Eqs. (1)–(3). For each liability estimate, the discount rate used was the current long term UK Government bond yield. Since salaries are subject to public sector pay policy, it is possible that salary increases are related to the macroeconomic situation in a way that differs from the private sector. The numerical results from this empirical analysis were then adjusted using estimates supplied by

Table 1
Correlation matrix, expected returns and standard deviations

	UK equities	Overseas equities	Property	Fixed interest	UK index-linked gilts	Active members	Deferreds	Pensioners
UK equities	–	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}
Overseas equities	–	–	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}	a_{28}
Property	–	–	–	a_{34}	a_{35}	a_{36}	a_{37}	a_{38}
Fixed interest	–	–	–	–	a_{45}	a_{46}	a_{47}	a_{48}
UK index-linked gilts	–	–	–	–	–	a_{56}	a_{57}	a_{58}
Active members	–	–	–	–	–	–	a_{67}*	a_{68}*
Deferreds	–	–	–	–	–	–	–	a_{78}*
Expected return	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
Standard deviation	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8

* The covariances between returns on different liabilities play no part in determining the asset proportions of the efficient frontier.

Schroders and Watson Wyatt. Craft (2005) used annual data on the projected pension benefit obligation of 647 US firms for 1988–2002 to construct a series of aggregate liability returns. These returns included changes in liabilities due to additional benefits, etc., adjusted for changes in the number of employees. The correlations of these liabilities with domestic equity, bonds and property; as well as the returns and standard deviation of liabilities, were used in adjusting the forecasts in Table 1.

Portfolio theory treats the means, variances and co-variances as free of estimation error. However, in reality, these parameters are subject to estimation risk. Therefore, the computed risks and returns of the portfolios are imprecise. Since portfolio theory seeks out those portfolios with low risk and high return, the presence of estimation risk tends to result in the choice of portfolios where the return is overstated and the risk is understated, see Michaud (1989). The size of this estimation risk depends on the accuracy of the forecasting procedures adopted. Board and Sutcliffe (1994) provide an example of the performance of different forecasting procedures in constructing efficient portfolios. It is generally accepted that for large pension schemes, actuarial forecasts of the demographic factors are reasonably accurate, and the key forecasts concern the investment and discount rates. Chopra and Ziemba (1993) demonstrate that for investors with high risk aversion the accuracy of mean asset returns is about three times more important than the forecasts of the variances; while the variance forecasts are about twice as important as the covariance forecasts. For investors with low risk aversion they can be in the ratios 60:3:1. Kallberg and Ziemba (1984) found that forecasts of mean returns are about ten times as important as forecasts of the covariance matrix. Therefore the greatest effort should be focused on obtaining accurate expected mean returns. In the present paper, estimation risk is not explicitly considered, and allowance for this should be made in interpreting the results presented below. Decision makers can get some idea of the robustness of the results of this model to estimation risk by conducting a sensitivity analysis and re-solving the problem using alternative estimates of asset and liability returns.

USS does not hedge the currency risk of foreign securities, and so the returns and correlations are expressed in terms of the sterling-equivalent returns. The returns are gross, since pension schemes are exempt from paying taxation. Prior to July 1997, UK pension schemes received a tax refund equal to the value of the advance corporation tax (ACT) paid by the company on the dividends declared. Therefore, before this date, the net dividend income of pension schemes exceeded their gross dividend income by the amount of this tax refund. This is no longer the case and pension schemes simply receive gross dividends. USS pensions are fully index linked, so that in the absence of deflation, no adjustment is required to allow for limited price indexation.

9. Solving the asset–liability portfolio model

The data in Table 1, together with the liability proportions (w_1, w_2 and w_3) for the pension scheme were used to solve the extended portfolio model set out in Section 2.

Table 2
Efficient portfolios—returns, standard deviations and asset proportions

	Assets (%)		Asset–liability (%)		Effectiveness (%)	Efficient asset proportions (%)				
	Exp ret.	Std dev.	Exp ret.	Std dev.		UK equities	Overseas equities	Property	Fixed interest	Index-linked gilts
1	2.20	2.500	−1.03	2.454	15	0.0	0.0	0.0	0.0	100.0
2	2.88	2.335	−0.35	2.112	37	0.0	0.0	2.1	20.7	77.2
3	3.56	2.452	0.33	1.916	48	0.0	0.2	9.4	31.9	58.5
4	4.24	2.676	1.01	1.823	53	0.0	3.2	15.3	39.0	42.6
5	4.92	2.940	1.69	1.828	53	4.4	3.2	20.2	43.8	28.3
6	5.60	3.234	2.37	1.921	48	9.1	3.2	25.1	48.6	14.0
7	6.28	3.550	3.05	2.090	38	13.8	3.2	30.0	53.0	0.0
8	6.96	4.057	3.73	2.452	15	27.8	5.3	31.8	35.1	0.0
9	7.64	4.679	4.41	3.036	−30	41.7	7.5	33.6	17.2	0.0
10	8.32	5.367	5.09	3.742	−97	56.4	9.6	34.0	0.0	0.0
11	9.00	6.599	5.77	5.360	−305	88.9	11.1	0.0	0.0	0.0
USS	8.12	5.462	4.89	4.070	−133	53.1	21.0	11.4	11.8	2.7

The results of this asset–liability pension model are sets of five asset proportions and three fixed, liability proportions. This allows the calculation of two sets of portfolio risks and returns: for the asset component of the portfolio only (representing the gross portfolio performance), or for the asset–liability portfolio (representing the net portfolio performance). Results were also obtained for the portfolio model using only assets, but these are not reported.

Table 2 shows the risk and return of both the asset and the asset–liability portfolios, as well as the investment proportions themselves, for a range of points on the efficient frontier. The asset–liability portfolios use the USS funding ratio of 101%. For low risk portfolios, the size of the optimal holding by USS of index linked gilts represents a substantial proportion of the market. This can be overcome by using the index linked swaps market, or by placing constraints on the maximum holding of index linked gilts. The last line of Table 2 shows the actual USS asset allocation on 31 January 2002. This portfolio is plotted in Figure 1 as USS_a (USS assets only) and USS_{al} (USS assets and liabilities). The efficient frontiers for these two alternative sets of portfolios are plotted in Figure 1. While the asset–liability quadratic programming model must generate convex efficient sets, the assets-only frontier may not necessarily be convex. Table 2 and Figure 1 show that the asset–liability numbers have lower risk and lower return than the corresponding asset-only numbers, highlighting the fact that the model allows the pension scheme to hedge the liability risk.

The USS Statement of Investment Principles (SIP) supports the objective of “funding the scheme’s benefits at the lowest cost over the long term, having regard to the minimum funding requirement of the Pensions Act 1995 and having regard to the attitude of the Committee of Vice-Chancellors and Principals and of the Management Committee

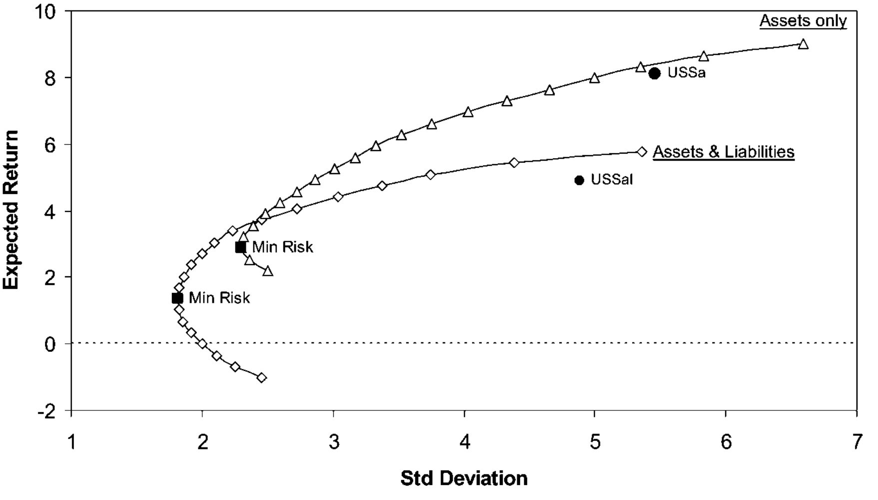


Fig. 1. Expected returns frontiers.

towards the risk of higher contributions at some time in the future", Universities Superannuation Scheme (2002). Given these objectives, the relevant values for decision taking by the pension scheme are those reflecting its net position, rather than the unhedged, or gross, asset returns and risks. The scheme is taking positions on the spread between investment returns and the rate of increase in retail prices and salaries. Thus, while small changes in the difference between these two returns may be important for the pension scheme, big changes in both may not. The asset–liability efficient frontier can be viewed as the outcome of a risk-minimizing generalized hedge, subject to the constraints of a given rate of return and a hedge ratio of $(w_1 + w_2)$ or -0.9918747.

Table 2 also shows the Ederington (1979) measure of hedging effectiveness, which gives the reduction in the variance of the asset–liability portfolio, relative to the variance of the fund's liabilities. This shows that the best hedge occurs for portfolios 4 and 5, which offer a 50% reduction in risk. It also shows that for portfolios 9 to 11 (and USS), the risk of the hedged portfolio exceeds that of the liabilities alone, so that the fund faces additional asset risk in addition to the basic risk of its liabilities.

The expected return for portfolios 1 and 2 for the assets and liabilities together is negative; a fact obscured if only the asset returns are considered. The results also demonstrate the interaction between assets and liabilities in the model, as the asset–liability frontier is not simply a linear transformation of the assets-only frontier (e.g., each point on the asset frontier shifted the same distance to the south west). For example, the risk-minimizing portfolio for the asset–liability model is portfolio 4, while for the assets-only portfolio it is portfolio 2. The expected return on the assets in the risk-minimizing portfolio 4 is 4.24%, which compares with a discount rate of 5.5% used in computing the actuarial liabilities. Thus, the asset–liability results reveal that portfolios 1, 2 and 3 are dominated, while the assets-only numbers incorrectly suggest that only portfolio 1 is dominated. Similarly, the slopes of the efficient frontiers differ, and the correct risk-return trade-off facing the pension scheme for any specified rate of return (or risk) is that provided by the asset–liability results, not the assets-only results. This confirms the view that any pension scheme should adopt an asset–liability based analysis of the asset allocation, rather than attempting to consider the asset allocation separately from the liabilities.

10. Transformation of the portfolio returns to contribution rates and funding ratios

As noted in Section 3, pension scheme trustees usually prefer to judge the results from portfolio models in terms of the implied level and dispersion of the contribution rate and funding ratio, rather than the risk and return of the asset or asset–liability allocation. For example, the declared objectives of USS can be summarized as simultaneously minimizing the average contribution rate and the variance of the contribution rate. The academic literature has considered three main objectives: (a) minimize the expected contribution rate, (b) minimize the variance of the contribution rate, and (c) minimize the variance of the funding ratio. USS takes the view that the first two objectives are the most important criteria for a well-funded scheme where insolvency is very unlikely. For the generalized Haberman model with a fixed spread period, the variance of the contribution rate is a linear function of the variance of the funding ratio, and so the choice between these two measures of risk is of little consequence. However, for regulatory reasons, the distribution of the funding ratio will also be considered, to ensure that the asset allocation decision is unlikely to lead to any regulatory problems.

In 2002 USS had a funding ratio of 101% and set the employers' contribution rate for the next three years at 14%, the same rate as in the preceding six years. This suggests that USS was in a fairly stable position, which is consistent with the Haberman (1992) model's requirement that the pension scheme is in long run equilibrium.

Eqs. (17)–(20) were used to calculate the mean and variance of the contribution rate and funding ratio of each efficient portfolio for both the Haberman (1992) and generalized Haberman models. For the generalized Haberman model the only additional parameters are M and d, while for the Haberman (1992) model many more parameters are required to re-compute the three actuarial liabilities. These parameters were based on Universities Superannuation Scheme (2003). The USS actuary used one rate for computing the standard contribution rate (i.e., a nominal yield of 6%) and a different rate for computing the funding ratio of the scheme, and hence the contribution rate adjustment (i.e., a current yield of 5%). The values of $M = 12$ and $d = 5.5\%$ were used.

The results for the Haberman (1992) model and the actual USS portfolio on 31 January 2002 are in Table 3. The fund's objective is assumed to be to minimize both the contribution rate and the standard deviation of the contribution rate, so that for the Haberman (1992) model portfolios to the north east of the minimum contribution rate

Table 3
The first two moments of the contribution rate and funding ratio—Haberman (1992) model with $M = 12$

	$E(r_a)$ (%)	σ_{AL} (%)	$E(CR)$ (%)	$SD(CR)$ (%)	$E(FR)$ (%)	$SD(FR)$ (%)
1	2.20	2.45	39.88	2.75	100.00	6.02
2	2.88	2.11	33.80	2.11	100.00	5.24
3	3.56	1.92	28.75	1.71	100.00	4.81
4	4.24	1.82	24.54	1.45	100.00	4.64
5	4.92	1.83	21.01	1.31	100.00	4.71
6	5.60	1.92	18.05	1.23	100.00	5.02
7	6.28	2.09	15.56	1.21	100.00	5.53
8	6.96	2.45	13.46	1.29	100.00	6.59
9	7.64	3.04	11.68	1.45	100.00	8.28
10	8.32	3.74	10.17	1.62	100.00	10.37
11	9.00	5.36	8.87	2.13	100.00	15.15
USS	8.12	4.07	10.59	1.82	100.00	11.24

risk portfolio (i.e., portfolios 1 to 6 in Table 2) are mean-variance dominated and need not be considered further. Thus, the transformation from the mean and variance of portfolio returns to the mean and variance of the contribution rate results in the exclusion of portfolios 4 to 6 from further consideration. This is in addition to portfolios 1 to 3, which were found to be dominated using asset–liability returns, see Table 5. The funding ratio for the Haberman (1992) model (*FR*) is constrained to be 100%, and the USS portfolios have been converted to the same basis. If the funding ratio is computed using a fixed discount rate of 5.5% (as in the generalized Haberman model), it ceases to be 100%.

Table 4 and Figure 2 show that, for the generalized Haberman model, the lowest contribution rate risk occurs for portfolios 2 and 3, while portfolio 3 has the lowest funding ratio risk. So, in this case, only portfolio 1 is dominated. Table 5 indicates that, while portfolios 2 and 3 are dominated when using asset–liability returns, they are not dominated when the generalized Haberman model is used. The generalized Haberman model allows *FR* to depart from 100%, as shown in Table 4. (The USS plots in Figures 2 and 3 use the funding ratio of 101%.) For portfolios 1–5, the funding ratio is below 100% because, although the contribution rate is high, the expected return on assets is low.

Figure 3 shows the trade-off between the contribution rate and the standard deviation of the funding ratio. Figure 3 reveals that, for the generalized Haberman model with a fixed spread period, there is a positive linear relationship between the standard deviations of the contribution rate and the funding ratio. This is in sharp contrast to the convex relationship for the Haberman (1992) model, which is also shown in Figure 3. The efficient set for the Haberman (1992) model in Figure 3 is the curve AB (portfolios 7 to 11).

Table 4
The first two moments of the contribution rate and funding ratio—generalized Haberman model with $M = 12$

	$E(r_a)$ (%)	σ_{AL} (%)	$E(CR)$ (%)	$SD(CR)$ (%)	$E(FR)$ (%)	$SD(FR)$ (%)
1	2.20	2.45	25.95	0.99	70.09	3.96
2	2.88	2.11	24.76	0.93	74.82	3.72
3	3.56	1.92	23.43	0.93	80.16	3.70
4	4.24	1.82	21.91	0.97	86.23	3.88
5	4.92	1.83	20.16	1.08	93.19	4.33
6	5.60	1.92	18.14	1.28	101.27	5.09
7	6.28	2.09	15.77	1.57	110.74	6.27
8	6.96	2.45	12.95	2.11	122.01	8.43
9	7.64	3.04	9.53	3.04	135.63	12.12
10	8.32	3.74	5.32	4.42	152.45	17.66
11	9.00	5.36	0.00	7.69	173.71	30.69
USS	8.12	4.07	6.68	4.57	147.01	18.26

Table 5
Mean-variance dominance

Model	Dominated portfolios
Assets only portfolios	1
Asset–liability portfolios	1, 2, 3
Contribution rates for Haberman (1992) with $M = 12$	1, 2, 3, 4, 5, 6
Contribution rates for generalized Haberman with $M = 12$	1

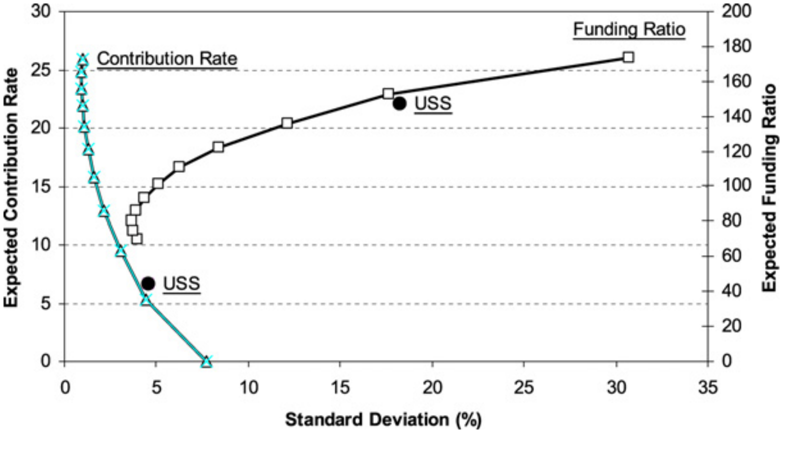


Fig. 2. Efficiency frontiers for CR and FR risk—generalized Haberman, $M = 12$.

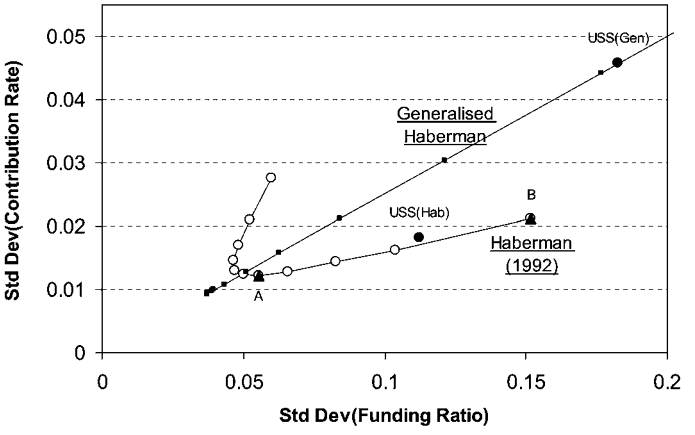


Fig. 3. $SD(CR)$ and $SD(FR)$—Haberman (1992) and generalized Haberman, $M = 12$.

11. Choice of the spread period

In Section 10, the spread period, M, was set to the value used by USS (12 years). However, it is possible that a different choice of M would improve the risk-return performance of the scheme. Section 5 described the estimation of M^*, the spread period which minimizes the contribution rate. M^* cannot be computed when the rate of salary growth exceeds the expected rate of return on the assets (because v_a is then negative), which is the case for the low return portfolios 1, 2 and 3. As Table 6 shows, for the Haberman (1992) model these are dominated portfolios, and so this is not a significant issue. However, for the generalized Haberman model Table 7 shows that this condition rules out portfolios 1 to 6, of which only the first is dominated.

Tables 6 and 7 show the optimal spread period for each portfolio (including the USS portfolio on 31 January 2002) for the Haberman (1992) and generalized Haberman models. For the Haberman (1992) model the fund's chosen spread period of 12 years is always less than the optimal spread period. As discussed in Section 5, for the Haberman (1992) model using a spread period shorter than M_H^* reduces $SD(FR)$, but increases $SD(CR)$. Therefore, although $SD(FR)$ can be reduced by lowering the spread period, this comes at the expense of an increase in $SD(CR)$. Since the two principal objectives used in this paper are to minimize both the mean and variance of the contribution rate, the possibility of minimizing $SD(FR)$ is not pursued. For the generalized Haberman model, both $SD(CR)$ and $SD(FR)$ rose as M_G^* fell.

The transformation of the portfolio results into the mean and variance of the contribution rate and funding ratio in the previous section was repeated using the relevant value of M^* (rounded to the nearest integer) from Tables 6 and 7. As can be seen by comparing Tables 3 and 6 for the Haberman (1992) model, the use of M_H^* in place of the fund's standard spread period of 12 years, increases $SD(FR)$ in every case, sometimes by substantial amounts. However, as expected, there were reductions in $SD(CR)$.

Table 6
The first two moments of the contribution rate and funding ratio—Haberman (1992) model with $M = M_H^*$

	$E(r_a)$ (%)	σ_{AL} (%)	$M = M_H^*$ (years)	$E(CR)$ (%)	$SD(CR)$ (%)	$E(FR)$ (%)	$SD(FR)$ (%)
4	4.24	1.82	129	24.54	0.66	100.00	17.10
5	4.92	1.83	60	21.01	0.84	100.00	11.70
6	5.60	1.92	39	18.05	0.94	100.00	9.91
7	6.28	2.09	29	15.56	1.02	100.00	9.31
8	6.96	2.45	24	13.46	1.16	100.00	10.01
9	7.64	3.04	20	11.68	1.36	100.00	11.34
10	8.32	3.74	17	10.17	1.57	100.00	12.90
11	9.00	5.36	15	8.87	2.10	100.00	17.49
USS	8.12	4.07	18	10.59	1.75	100.00	14.50

Note: M_H^* cannot be computed for portfolios 1–3.

Table 7
The first two moments of the contribution rate and funding ratio—generalized Haberman model with $M = M_G^*$

	$E(r_a)$ (%)	σ_{AL} (%)	$M = M_G^*$ (years)	$E(CR)$ (%)	$SD(CR)$ (%)	$E(FR)$ (%)	$SD(FR)$ (%)
7	6.28	2.09	19	15.18	1.50	119.55	8.93
8	6.96	2.45	14	12.51	2.11	127.26	9.70
9	7.64	3.04	11	9.96	3.00	131.35	11.07
10	8.32	3.74	9	7.53	4.11	133.56	12.63
11	9.00	5.36	8	4.81	6.47	137.56	17.82
USS	8.12	4.07	10	7.98	4.37	135.47	14.79

Note: M_G^* cannot be computed for portfolios 1–6.

A comparison of Tables 4 and 7 reveals that, for the generalized Haberman model, moving to M_G^* leads to a reduction in $SD(CR)$, with the size of the reduction depending on the size of the change in M. There is no reduction in $SD(CR)$ for portfolio 8 due to the rounding of the spread period and the insensitivity of $SD(CR)$ to M. For portfolios 7 and 8 $M_G^* > 12$, and the values of $SD(FR)$ rise; while for portfolios 9 to 11 and USS, $M_G^* < 12$ and the values of $SD(FR)$ fall. This shows that, when the spread period is not fixed, the trade-off between $SD(CR)$ and $SD(FR)$ reappears.

The relationship between $SD(CR)$ and M is illustrated in Figure 4 for portfolios 7 and 11 using the generalized Haberman model. This shows that for portfolio 7, increasing M from 12 to 19 years reduces $SD(CR)$ from 1.57% to 1.50%; while for portfolio 11,

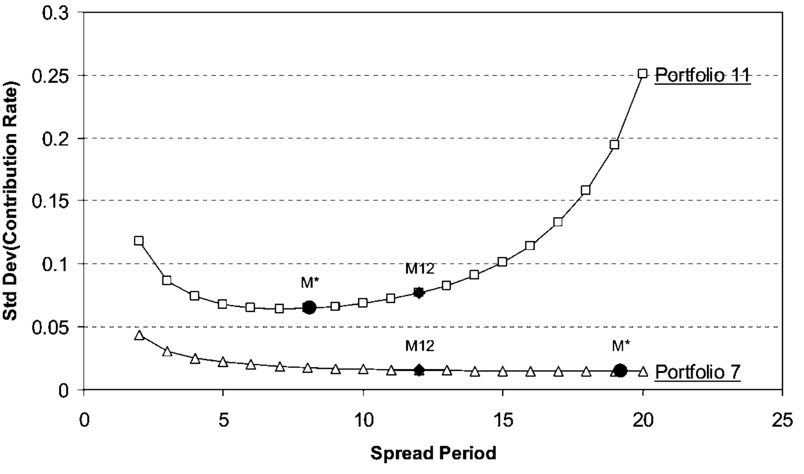


Fig. 4. Optimal spread periods—generalized Haberman.

reducing M from 12 to 8 years reduces this risk from 7.69% to 6.47%. Figure 4 also reveals that, over the relevant range, $SD(CR)$ is not very sensitive to M.

12. Allowance for triennial valuations

Both the Haberman (1992) and the generalized Haberman models assume annual actuarial valuations. However, most schemes have triennial actuarial valuations. The effects of triennial valuations on the variances of the contribution rate and the funding ratio were investigated by comparing the Dufresne (1988) and Haberman (1993b) models, which are summarized in Appendix A. Triennial valuations may increase or decrease the variance of the contribution rate, see Haberman (1993b) and Cairns (1996a). The Dufresne (1988) model is similar to the Haberman (1992) model, except that there is no lag in adjusting the contribution rate, and differs from the Haberman (1993b) only in that the latter assumes triennial actuarial valuations. Therefore a comparison of the results from these two models will reveal the increase in the variances caused by the introduction of triennial valuations. Using the USS data described above, but with M set equal to the value of M^* for the model concerned, it was found that the understatement of the variances of *CR* and *FR* due to the absence of a one year lag was only about 2.6% (or 2.7% when $M = 12$ years). This suggests that a move to triennial valuations would make little difference to the variances of *CR* and *FR*, or to the other results in this chapter.

Even if there had been a material effect on the variances by switching to triennial valuations, a model which assumes annual valuations may still be preferable. This is because many pension schemes, including USS, make an annual actuarial check on the funding ratio, after which the contribution rate could be adjusted. This is effectively an annual review of the contribution rate, even though a full actuarial valuation is performed only every three years.

13. Regulatory and solvency risk

In order to investigate solvency risk, a knowledge of the probability distribution of the funding ratio is required. The parameters of the distribution of the funding ratio were computed using the equations in Section 6 for the generalized Haberman model with $M = 12$ and $M = M_G^*$. The inverted gamma distribution was then used to compute the values of the shortfall probability (*SP*) and the *ETL* for the upper and lower regulatory restrictions. The values of *SP* and *ETL* were for one year. To obtain results for say a 3-year period, assuming that proportionate changes in the funding ratio are independent over time, the mean and variance of the *FR* used in Eqs. (23) and (24) must be multiplied by 3. If proportionate changes in the *FR* are correlated, a more complicated adjustment is required (Jorion, 2000).

As a simple approximation to the lower solvency bound that will apply after the abolition of the *MFR*, the specified VaR for the lower tail was set at 70%, which roughly approximates to 100% *MFR* funding. While the *MFR* funding ratio is due to be replaced by scheme-specific funding requirements (Secretary of State for Work and Pensions, 2002), it was decided to use the current *MFR* as the lower bound. Because each of the regulatory restrictions requires the use of a valuation basis that differs from that used to determine the contribution rate, the funding ratios of 90% and 105% were adjusted to be comparable with those used elsewhere in this paper. At the 2002 actuarial valuation the *MFR* funding ratio for USS was 144%, as against a funding ratio of 101% using the assumptions of the USS actuary. This implies a lower solvency bound of $101/144 = 70\%$. The upper bound, imposed to prevent over-funding, was set at $100/70 = 142.86\%$.

The results are in Tables 8 and 9. The *ETL*s in Tables 8 and 9 were computed by setting m, the number of samples, to 100. When $M = 12$, SP_1 drops to zero for portfolios 4–11, while for portfolio 1 it is over 50%. This reflects the negative expected asset returns for portfolios 1 and 2, and the larger expected returns for higher numbered portfolios. Conversely, for portfolios 1–7, $SP_2 = 0$, but as the expected asset return and risk rise, SP_2 rises to over 85% for $M = 12$. Even though portfolio 1 has a 50% chance of breaching the *MFR* when $M = 12$, the expected value of the funding ratio (*ETL*) is still 67%, i.e. a 3% breach. Portfolio 11 has an 85% probability of breaching the upper bound when $M = 12$, and the expected funding ratio is 180%, i.e., a breach of 37%. Thus, the average breach of the upper bound tends to be much bigger than the average breach of the lower bound, reflecting the strong positive skewness of the distribution of the funding ratio.

When $M = M_G^*$, Table 9 shows that the values of SP_1 and ETL_1 are little changed, while for SP_2 and ETL_2 the values become more even across the range, with falls for

Table 8
Solvency and regulatory risk: Shortfall probabilities and *ETL*s: Generalized Haberman model with $M = 12$

	$E(r_a)$ (%)	σ_{AL} (%)	$E(FR)$ (%)	$SD(FR)$ (%)	alpha	beta	70%		142.8%	
							SP_1 (%)	ETL_1 (%)	SP_2 (%)	ETL_2 (%)
1	2.20	2.45	70.09	3.96	315	4.54	50.57	67.03	0.00	–
2	2.88	2.11	74.82	3.72	407	3.29	9.26	68.49	0.00	–
3	3.56	1.92	80.16	3.70	472	2.65	0.14	69.16	0.00	–
4	4.24	1.82	86.23	3.88	496	2.34	0.00	69.45	0.00	–
5	4.92	1.83	93.19	4.33	466	2.31	0.00	69.58	0.00	–
6	5.60	1.92	101.27	5.09	397	2.49	0.00	–	0.00	144.01
7	6.28	2.09	110.74	6.27	314	2.89	0.00	–	0.00	144.67
8	6.96	2.45	122.01	8.43	212	3.89	0.00	–	1.14	146.42
9	7.64	3.04	135.63	12.12	127	5.84	0.00	–	26.18	150.99
10	8.32	3.74	152.45	17.66	77	8.68	0.00	69.26	68.98	160.67
11	9.00	5.36	173.71	30.69	34	17.43	0.00	68.70	85.33	180.10
USS	8.12	4.07	147.01	18.26	67	10.34	0.00	69.11	56.05	159.17

Table 9
Solvency and regulatory risk: Shortfall probabilities and *ETL*s: Generalized Haberman model with $M = M_G^*$

	$E(r_a)$ (%)	σ_{AL} (%)	$E(FR)$ (%)	$SD(FR)$ (%)	alpha	beta	70%		142.8%	
							SP_1 (%)	ETL_1 (%)	SP_2 (%)	ETL_2 (%)
7	6.28	2.09	119.55	8.93	181	4.64	0.00	–	0.88	146.64
8	6.96	2.45	127.26	9.70	174	4.54	0.00	–	6.13	147.79
9	7.64	3.04	131.35	11.07	143	5.37	0.00	–	14.86	149.40
10	8.32	3.74	133.56	12.63	114	6.63	0.00	69.36	22.03	151.05
11	9.00	5.36	137.56	17.82	62	12.00	0.00	68.91	35.25	156.35
USS	8.12	4.07	135.47	14.79	86	8.69	0.00	69.19	28.83	153.26

Note: M_G^* cannot be computed for portfolios 1–6.

high risk and high return portfolios, and rises for lower risk and lower return portfolios. Note that for M_G^*, the values of SP_2 and ETL_2 for portfolios 7 and 8 are larger than when $M = 12$; while for portfolios 9–11, the reverse is the case. This is because $M_G^* > 12$ for portfolios 7 and 8, and any surplus is removed more slowly, permitting high funding ratios to be attained; while for portfolios 9–11, $M_G^* < 12$, and any surplus is eliminated more quickly.

14. Conclusions

This paper had modeled a number of aspects of pension trustees' decisions concerning the asset allocation and contribution rate. A simple extension of the portfolio model permits the inclusion of different types of liability in the computation of the efficient asset proportions for various levels of risk and return. Pension schemes generally state that their asset allocation decision has taken account of their liabilities. However, they have usually not explicitly incorporated the liabilities into the asset allocation decision. The asset allocations derived from this extended model are different from those derived from an assets-only formulation, and offer significant hedging of the schemes' liabilities. Because of this, it is important that the results of the portfolio analysis are presented to decision makers in terms of the risks and returns on the combination of the assets and liabilities of the scheme, rather than just the assets alone.

An analysis of the asset allocation decision in terms of its effects on the risks and returns of the assets and liabilities of the scheme does not explicitly consider the effects on the contribution rate and the funding ratio. This paper shows how the results of a portfolio model can be re-expressed in terms of the first two moments of the contribution rate and funding ratio. By presenting decision makers with the implications of alternative asset allocations for the mean and variance of the contribution rate and funding ratio, the problem is translated into the variables which ultimately concern the trustees. This transformation can also reveal additional or reduced mean-variance dominance between alternative asset allocations.

Computing the effects of the asset allocation decision on the contribution rate, enables the actuary to calculate the spread period which minimizes contribution rate risk for the chosen asset allocation. This paper generalizes the Haberman (1992) model by dropping the requirement that the discount rate equals the rate of return on the investments. As well as improving the economic realism of the model, this greatly simplifies its empirical application because the actuarial valuation need not be repeated for a range of different discount rates. In addition, it removes the trade-off between contribution rate risk and funding ratio risk (for a fixed spread period). The Haberman (1992) model and the generalized Haberman model were applied to data for the Universities Superannuation Scheme, and the efficient frontiers plotted for both fixed and optimized spread periods. The trustees then choose a particular combination of the contribution rate and contribution rate risk, which determines the asset allocation.

Finally, the distribution of the funding ratio was considered, in conjunction with the upper and lower statutory limits. This allows trustees to investigate the regulatory and solvency risks associated with a particular asset allocation/contribution rate choice.

The application of the model proposed in this paper requires a change in the way pensions schemes operate. Currently, the contribution rate is usually recommended by the scheme actuary on the basis of an assumed asset allocation; while the actual asset allocation is set separately on the basis of investment advice. The proposed model requires a joint decision by a single person or group, probably at the time the contribution rate is set. The results for USS suggest that this can lead to superior decisions.

Acknowledgements

The authors wish to thank Shirin Hashemi (University of Southampton) for help in deriving equation (22G), Stephen Court (AUT) for supplying the data on academic salaries, Mike Orszag (Watson Wyatt LLP) and Craig Gillespie (Watson Wyatt LLP), Bill Ziemba (University of British Columbia), Peter Casson (University of Southampton) and participants in the Inquire UK, Inquire Europe and Q Group meeting in Versailles, and seminars at the ICMA Centre, Reading, and Watson Wyatt, Reigate for comments on an earlier draft. Although one of the authors is a director of USS Ltd, this paper represents independent research conducted using publicly available information. The views expressed in this paper are those of the authors, and may not be shared by USS.

Appendix A

A.1. Annual actuarial valuations and instant adjustment of the contribution rate (Dufresne, 1988)

$$E[F]_D = AL_T, \tag{A.1}$$

$$E[C]_D = SC, \tag{A.2}$$

$$\mathrm{Var}(F)_D = \sigma^2 AL_T^2/u^2\big[1-(\sigma^2+u^2)(1-k_D)^2\big], \tag{A.3}$$

$$\mathrm{Var}(C)_D = k_D^2\sigma^2 AL_A^2/u^2\big[1-(\sigma^2+u^2)(1-k_D)^2\big], \tag{A.4}$$

$$M_D^* = -\log\big(1-v_a/\big[(1+v_a)k_D\big]\big)/\log(1+v_a), \tag{A.5}$$

$$\text{where } v_a > 0, \quad k_D = 1-1/y, \quad y = (1+v_a)^2+\sigma^2.$$

A.2. Triennial actuarial valuations and instant adjustment of the contribution rate (Haberman, 1993b)

$$E[F]_3 = AL_T, \tag{A.6}$$

$$E[C]_3 = SC, \tag{A.7}$$

$$\mathrm{Var}(F)_3 = \big[\big\{y_3-(1+v_a)^6\big\}AL_T^2\big]/\big[(1+v_a)^6\big\{1-y_3(1-k_3)^2\big\}\big], \tag{A.8}$$

$$\mathrm{Var}(C)_3 = \big[k_3^2\big\{y_3-(1+v_a)^6\big\}AL_A^2\big]/\big[(1+v_a)^6\big\{1-y_3(1-k_3)^2\big\}\big], \tag{A.9}$$

$$M_3^* = -3\log\big(1-s/\big\{(1+s)\big[1-1/y_3\big]\big\}\big)/\log(1+s) \tag{A.10}$$

$$\text{where } v_a > 0, \quad k_3 = \big(1-\big\{1/(1+v_a)\big\}^3\big)/\big(1-\big\{1/(1+v_a)\big\}^M\big),$$

$$y_3 = (1+2v_a+v_a^2+\sigma^2)^3 \quad \text{and} \quad s = (1+v_a)^3-1.$$

References

Accounting Standards Board, 1988. Statement of Standard Accounting Practice 24—Accounting for Pension Costs. ASB Publications, London, May.
Accounting Standards Board, 2000. Financial Reporting Standard 17—Retirement Benefits. ASB Publications, London, November.
Actuarial Education Company, 2002. ActEd Study Materials: 2002 Examinations, Subject 304, Course Notes. Actuarial Education Company, Oxford.
Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9 (3), 203–228, July.
Association of Consulting Actuaries, 2003. Occupational Pensions 2003—Pensions Reform: Too Little, Too Late? ACA, March.
Bagehot, W., 1972. Risk and reward in corporate pension funds. Financial Analysts Journal 28 (1), 80–84, January–February.
Bédard, D., 1999. Stochastic Pension funding: Proportional control and bilinear processes. ASTIN Bulletin 29 (2), 271–293, November.
Bédard, D., Dufresne, D., 2001. Pension funding with moving average rates of return. Scandinavian Actuarial Journal 2001 (1), 1–17, March.
Black, F., 1995. The plan sponsor's goal. Financial Analysts Journal 51 (4), 6–7, July–August.
Blake, D., Lehmann, B.N., Timmermann, A., 1999. Asset allocation dynamics and pension fund performance. Journal of Business 72 (4), 429–461, October.
Board, J.L.G., Sutcliffe, C.M.S., 1994. Estimation methods in portfolio selection and the effectiveness of short sales restrictions: UK evidence. Management Science 40 (4), 516–534, April.
Boender, C.G.E., 1997. A hybrid simulation–optimization scenario model for asset–liability management. European Journal of Operations Research 99 (1), 126–135, May.
Boender, C.G.E., Dert, C.L., Hoek, H., 2006. ALM for pension funds. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management. Elsevier Science.
Boender, C.G.E., Van Aalst, P.C., Heemskerk, F., 1998. Modelling and management of assets and liabilities of pension plans in the Netherlands. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modelling. Cambridge University Press, pp. 561–580.
Boender, C.G.E., Vos, M., 2000. Risk return budgeting at pension plans. The Institutional Investor, pp. 80–88, May.
Bogentoft, E., Romeijn, H.E., Uryasev, S., 2001. Asset–liability management for pension funds using CVaR constraints. Journal of Risk Finance 3 (1), 57–71, Fall.
Booth, P., Chadburn, R., Cooper, D., Haberman, S., James, D., 1999. Modern Actuarial Theory and Practice. Chapman and Hall/CRC, Boca Raton, FL, Chapter 25.
Boulier, J.F., Michel, S., Wisnia, V., 1996. Optimizing investment and contribution policies of a defined benefit pension fund. In: Albrecht, P. (Ed.), Proceedings of the 6th AFIR International Colloquium, Nuremberg, 1996, pp. 593–607.
Boulier, J.F., Trussant, E., Florens, D., 1995. A dynamic model for pension funds management. In: Janssen, J. (Ed.), Proceedings of the 5th AFIR International Colloquium, Brussels, 1995, pp. 361–384.
Brinson, G.P., Hood, L.R., Beebower, G.L., 1986. Determinants of portfolio performance. Financial Analysts Journal 42 (4), 39–44, July–August.
Brinson, G.P., Singer, B.D., Beebower, G.L., 1991. Determinants of portfolio performance II: An update. Financial Analysts Journal 47 (3), 40–48, May–June.
Cairns, A.J.G., 1995. Pension funding in a stochastic environment: The role of objectives in selecting an asset allocation strategy. In: Janssen, J. (Ed.), Proceedings of the 5th AFIR International Colloquium, Brussels 1995, pp. 429–453.
Cairns, A.J.G., 1996a. An introduction to stochastic pension fund management, Working Paper 9607. Pensions Institute.
Cairns, A.J.G., 1996b. Continuous time pension fund modelling. In: P. Albrecht (Ed.), Proceedings of the 6th AFIR International Colloquium, Nuremberg, 1996, pp. 609–624.

Cairns, A.J.G., 1997. A comparison of optimal and dynamic control strategies for continuous time pension plan models. In: Proceedings of the 7th AFIR International Colloquium, Cairns, 1997, pp. 309–326.

Cairns, A.J.G., 2000. Some notes on the dynamics and optimal control of stochastic pension fund models in continuous time. ASTIN Bulletin 30 (1), 19–55, May.

Cairns, A.J.G., Parker, G., 1997. Stochastic pension fund modelling. Insurance: Mathematics and Economics 21 (1), 43–79, October.

Campbell, J.Y., Viceira, L.M., 2002. Strategic Asset Allocation: Portfolio Choice for Long Term Investors. Oxford University Press.

Cariño, D.R., Ziemba, W.T., 1998. Formulation of the Russell–Yasuda Kasai financial planning model. Operations Research 46 (4), 433–449, July–August.

Cariño, D.R., Myers, D.H., Ziemba, W.T., 1998. Concepts, technical issues and uses of the Russell–Yasuda Kasai financial planning model. Operations Research 46 (4), 450–462, July–August.

Cariño, D.R., Kent, T., Myers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., Ziemba, W.T., 1994. The Russell–Yasuda Kasai model: An asset–liability model for a Japanese Insurance Company using multistage stochastic programming. Interfaces 24 (1), 29–49, January–February.

Chang, S.C., Chen, C.C., 2002. Allocating unfunded liability in pension valuation under uncertainty. Insurance: Mathematics and Economics 30 (3), 371–387, June.

Chopra, V.K., Ziemba, W.T., 1993. The effect of errors in means, variances and covariances on optimal portfolio choice. Journal of Portfolio Management 19 (2), 6–11, Winter.

Constantinides, G.M., 2002. Rational asset prices. Journal of Finance 57 (4), 1567–1591, August.

Cornell, B., 1999. The Equity Risk Premium: The Long-Run Future of the Stock Market. John Wiley and Sons.

Craft, T.M., 2001. The role of private and public real estate in pension plan portfolio allocation choices. Journal of Real Estate Portfolio Management 7 (1), 17–23, January–March.

Craft, T.M., 2005. How funding ratios affect pension plan allocations. Journal of Real Estate Portfolio Management 11 (1), 29–35, January–April.

Chun, G.H., Chiochetti, B.A., Shilling, J.D., 2000. Pension plan real estate investment in an asset–liability framework. Real Estate Economics 28 (3), 467–491, Fall.

Dert, C.L., 1998. A dynamic model for asset liability management for defined benefit pension funds. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modelling. Cambridge University Press, pp. 501–536.

Dimson, E., Marsh, P., Staunton, M., 2002. The Triumph of the Optimists: 101 Years of Global Investment Returns. Princeton University Press.

Dondi, G., Herzog, F., Schuman, L., Geering, H.P., 2006. Dynamic asset and liability management for Swiss pension funds. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management. Elsevier Science.

Dowd, K., 2002. An Introduction to Market Risk Measurement. John Wiley.

Drijver, S.J., Klein Haneveld, W.K., Van Der Vlerk, M.H., 2002. ALM model for pension funds: Numerical results for a prototype model, Research Report. University of Grongingen.

Drijver, S.J., Klein Haneveld, W.K., Van Der Vlerk, M.H., 2003. Asset liability management modelling using multi-stage mixed integer stochastic programming. In: Scherer, B. (Ed.), Asset and Liability Management Tools: A Handbook for Best Practice. Risk Books, pp. 309–324.

Dufresne, D., 1986. Pension funding and random rates of return. In: Goovaerts, M., De Vylder, F., Haezendonck, J. (Eds.), Insurance and Risk Theory. D. Reidel Publishing, pp. 277–291.

Dufresne, D., 1988. Moments of Pension contributions and fund levels when rates of return are random. Journal of the Institute of Actuaries 115 (3), 535–544, September.

Dufresne, D., 1989. Stability of pension systems when rates of return are random. Insurance: Mathematics and Economics 8 (1), 71–76, March.

Dufresne, D., 1990a. Fluctuations of pension contributions and fund level. Actuarial Research Clearing House, pp. 111–120.

Dufresne, D., 1990b. The Distribution of a perpetuity with applications to risk theory and pension funding. Scandinavian Actuarial Journal (1–2), 39–79.

Ederington, L.H., 1979. The hedging performance of the new futures markets. Journal of Finance 34 (1), 157–170, March.

Evans, M., Hastings, N., Peacock, B., 1993. Statistical Distributions, second ed. John Wiley and Sons, Chapter 18—Gamma distributions.

Exley, C.J., Mehta, S.J.B., Smith, A.D., 1997. The financial theory of defined benefit pension schemes. British Actuarial Journal 3, 835–966, Part 4.

Ezra, D.D., 1991. Asset allocation by surplus optimization. Financial Analysts Journal 47 (1), 51–57, January–February.

Fabozzi, F.J., Focardi, S.M., Jonas, C.L., 2005. Market experience with modelling for defined benefit pension funds: Evidence from four countries. Journal of Pension Economics and Finance 4 (3), 313–327, November.

Faculty and Institute of Actuaries, 2003. Pensions and other Benefits, Subject 304 Core Reading, GN26: Pension Fund Terminology. Faculty and Institute of Actuaries, Oxford.

Frankfurter, G.M., Hill, J.M., 1981. A normative approach to pension fund management. Journal of Financial and Quantitative Analysis 16 (4), 533–558, November.

Gerrard, R., Haberman, S., 1996. Stability of pension systems when gains–losses are amortized and rates of return are autoregressive. Insurance: Mathematics and Economics 18 (1), 59–71, May.

Geyer, A., Herold, W., Kontriner, K., Ziemba, W.T., 2005. The Innovest Austrian Pension Fund financial planning model InnoALM, Working paper. University of British Columbia, February.

Gondzio, J., Kouwenberg, R., 2001. High performance computing for asset–liability management. Operations Research 49 (6), 879–891, November–December.

Haberman, S., 1990a. Stochastic approach to pension funding methods. In: Proceedings of the 1st AFIR International Colloquium, Paris, 1990, pp. 93–112.

Haberman S., 1990b. Variability of pension contributions and fund levels with random and autoregressive rates of return. Actuarial Research Clearing House, pp. 141–171.

Haberman, S., 1992. Pension funding with time delays: A stochastic approach. Insurance: Mathematics and Economics 11 (3), 179–189, October.

Haberman, S., 1993a. Pension funding with time delays and autoregressive rates of investment return. Insurance: Mathematics and Economics 13 (1), 45–56, September.

Haberman, S., 1993b. Pension funding: The effect of changing the frequency of valuations. Insurance: Mathematics and Economics 13 (3), 263–270, December.

Haberman, S., 1994a. Autoregressive rates of return and the variability of pension contributions and fund levels for a defined benefit pension scheme. Insurance: Mathematics and Economics 14 (3), 219–240, July.

Haberman, S., 1994b. Defined benefit pension funding models and stochastic investment returns. Paper presented to a joint meeting of the Staple Inn Actuarial Society and the Royal Statistical Society, March, 50 p.

Haberman, S., 1995. Pension funding with time delays and the optimal spread period. ASTIN Bulletin 25 (2), 177–187, November.

Haberman, S., 1997a. Stochastic investment returns and contribution rate risk in a defined benefit pension scheme. Insurance: Mathematics and Economics 19 (2), 127–139, April.

Haberman, S., 1997b. Risk in a defined benefit pension scheme. Singapore International Insurance and Actuarial Journal 1, 93–103.

Haberman, S., 1998. Stochastic modelling of pension scheme dynamics, Actuarial Research Report, No. 106, February, 35 p.

Haberman, S., Butt, Z., Megaloudi, C., 2000. Contribution and solvency risk in a defined benefit pension scheme. Insurance: Mathematics and Economics 27 (2), 237–259, October.

Haberman, S., Dufresne, D., 1991. Variability of pension contributions and fund levels with random rates of return. In: Cummins, J.D., Derrig, R.A. (Eds.), Managing the Insolvency Risk of Insurance Companies. Kluwer Academic Press, pp. 133–145.

Haberman, S., Owadally, I., 2001. Modelling defined benefit pension schemes: funding and asset valuation. Paper presented to the International Actuarial Association International Pensions Seminar, Brighton, 36 p.

Haberman, S., Wong, L.Y.P., 1997. Moving average rates of return and the variability of pension contributions and fund levels for a defined benefit pension scheme. Insurance: Mathematics and Economics 20 (2), 115–135, September.

Haberman, S., Day, C., Fogarty, D., Khorasanee, Z., McWhirter, M., Nash, N., Ngwira, B., Wright, I.D., Yakoubov, Y., 2003. A stochastic approach to risk management and decision making in defined benefit pension schemes, Paper presented to the Institute of Actuaries, 27th January, 95 p.

Hakansson, N., 1970. Optimal investment and consumption strategies under risk for a class of utility functions. Econometrica 38 (5), 587–607, September.

Hakansson, N., 1971. On optimal myopic portfolio policies with and without serial correlation of yields. Journal of Business 44 (3), 324–334, July.

Hilli, P., Koivu, M., Pennanen, T., Ranne, A., 2005. A stochastic programming model for asset liability management of a Finnish pension company. Annals of Operations Research.

Ibbotson, R.G., Kaplan, P.D., 2000. Does asset allocation policy explain 40, 90 or 100 percent of performance? Financial Analysts Journal 56 (1), 26–33, January–February.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, vol. 1, second ed. John Wiley and Sons, New York, Chapter 17—Gamma distributions.

Josa-Fombellida, R., Rincón-Zapatero, J.P., 2001. Minimization of risks in pension funding by means of contributions and portfolio selection. Insurance: Mathematics and Economics 29 (1), 35–45, August.

Jorion, P., 2000. Value at Risk: The New Benchmark for Managing Financial Risk, second ed. McGraw-Hill.

Kallberg, J.G., Ziemba, W.T., 1984. Mis-specification in portfolio selection problems. In: Bämberg, G., Spremann, K. (Eds.), Risk and Capital. Springer-Verlag, pp. 74–87.

Kingsland, L., 1982. Projecting the financial condition of a pension plan using simulation analysis. Journal of Finance 37 (2), 577–584, May.

Kouwenberg, R., 1997. Asset liability management for pension funds: Elements of Dert's model. In: Zopounidis, C. (Ed.), New Operational Approaches for Financial Modelling. Physica-Verlag, pp. 37–48.

Kouwenberg, R., 2001. Scenario generation and stochastic programming models for asset liability management. European Journal of Operations Research 134 (2), 279–292, October.

Kusy, M.L., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34 (3), 356–376, May–June.

Leibowitz, M.L., Bader, L.N., Kogelman, S., 1996. Return Targets and Shortfall Risks: Studies in Strategic Asset Allocation. Irwin Professional Publishing.

Logan, D., 1985. The Birth of a Pension Scheme: A History of the Universities Superannuation Scheme. Liverpool University Press.

MacBeth, J.D., Emanuel, D.C., Heatter, C.E., 1994. An investment strategy for defined benefit plans. Financial Analysts Journal 50 (3), 34–41, May–June.

Mandl, P., Mazurová, L., 1996. Harmonic analysis of pension funding methods. Insurance: Mathematics and Economics 17 (3), 203–214.

Merton, R.C., 1992. Continuous-Time Finance, Revised ed. Blackwell Publishing.

Michaud, R.O., 1989. The Markowitz optimization enigma: Is 'optimized' optimal? Financial Analysts Journal 45 (1), 31–42, January–February.

Mossin, J., 1968. Optimal multiperiod portfolio policies. Journal of Business 41 (2), 215–229, April.

Mulvey, J.M., Gould, G., Morgan, C., 2000. An asset and liability management system for towers Perrin–Tillinghast. Interfaces 30 (1), 96–114, January–February.

Mulvey, J.M., Simsek, K.D., Pauling, B., 2003. A stochastic network approach for integrating pension and corporate financial planning. In: Nagurney, A. (Ed.), Innovations in Financial and Economic Networks. Edward Elgar, pp. 67–83.

Mulvey, J.M., Thorlacius, A.E., 1998. The towers Perrin global capital market scenario generation system. In: Ziemba, W.T., Mulvey, J.M. (Eds.), Worldwide Asset and Liability Modelling. Cambridge University Press, pp. 286–312.

Mulvey, J.M., Fabozzi, F.J., Pauling, W.R., Simsek, K.D., Zhang, Z., 2005. Modernizing the defined benefit pension system. Journal of Portfolio Management 31 (2), 73–82, Winter.

Nijman, T., Swinkels, L., 2003. Strategies and tactical allocation to commodities for retirement savings schemes, Discussion paper no. 20, Tilburg University, February.

Owadally, M.I., Haberman, S., 1999. Pension fund dynamics and gains/losses due to random rates of investment return. North American Actuarial Journal 3 (3), 105–117.

Owadally, M.I., Haberman, S., 2000. Efficient amortization of actuarial gains and losses in pension plans. Actuarial Research Clearing House, pp. 275–317.

Ralfe, J., 2001. Why bonds are right for pension funds. Risk 14 (11), 54–55, November.

Ralfe, J., Speed, C., Palin, J., 2003. Pensions and capital structure: Why hold equities in the pension fund? Society of Actuaries Symposium on the Great Controversy: Current Pension Actuarial Practice in Light of Financial Economics, Vancouver, June.

Rudolf, M., Ziemba, W.T., 2004. Intertemporal surplus management. Journal of Economic Dynamics and Control 28 (5), 975–990, February.

Secretary of State for Work and Pensions, 2002. Simplicity, security and choice: Working and saving for retirement, Green Paper, Cm 5677, December.

Sharpe, W.F., 1990. Asset allocation. In: Maginn, J.L., Tuttle, D.L. (Eds.), Managing Investment Portfolios: A Dynamic Process. second ed. Warren, Gorham and Lamont, pp. 7.1–7.71, Chapter 7.

Sharpe, W.F., Tint, L.G., 1990. Liabilities—a new approach. Journal of Portfolio Management 16 (2), 5–10, Winter.

Siegel, J.J., 2002. Stocks for the Long Run, third ed. McGraw-Hill.

Sutcliffe, C.M.S., 2005. The cult of the equity for pension funds: Should it get the boot? Journal of Pension Economics and Finance 4 (1), 57–85, March.

Tepper, I., 1974. Optimal financial strategies for trusteed pension plans. Journal of Financial and Quantitative Analysis 9 (3), 357–376, June.

Thornton, P.N., Wilson, A.F., 1992. A realistic approach to pension funding. Journal of the Institute of Actuaries 119, 229–277.

Universities Superannuation Scheme, 2002. Report and Accounts for the Year Ended 31 March 2002.

Universities Superannuation Scheme, 2003. Actuarial Valuation Report as at 31 March 2002, March 2003.

Waring, M.B., 2004. Liability-relative investing II. Journal of Portfolio Management 31 (1), 40–53, Fall.

Wright, I.D., 1998. Traditional pension fund valuation in a stochastic asset and liability environment. British Actuarial Journal 4, 865–901, Part 4.

Yang, T., 2003. Defined benefit pension plan liabilities and international asset allocation, Pension Research Council Working Paper no. 2003-5. University of Pennsylvania.

Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset, Liability, and Wealth Management, The Research Foundation of AIMR, Charlottesville, VA.

Ziemba, W.T., 2006. The Russell Yasuda, InnoALM and Related Models for Pensions, Insurance Companies and High Net Worth Individuals. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. B. Elsevier Science.

Zimbidis, A., Haberman, S., 1993. Delay, feedback and variability of pension contributions and fund levels. Insurance: Mathematics and Economics 13 (3), 271–285, December.

Chapter 22

ALM ISSUES IN SOCIAL SECURITY*

SANDRA L. SCHWARTZ

Vancouver, BC V6N 2C9, Canada
e-mail: slschwartz@mac.com

WILLIAM T. ZIEMBA

Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC V6T 1Z2, Canada

Contents

* This research was supported by NSERC. Thanks go to Rachel Ziemba for comments on an earlier version of this paper.

Handbook of Asset and Liability Management, Volume 2
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)02022-9

Abstract

In this chapter we explore the issues surrounding the social security shortfall and the proposals for transforming social security into a sustainable retirement system. For the most part the description of the crisis is based on purely accounting type deterministic models that rely on demographic parameters. These models can also be used to show how simple tinkering with the variables can stave off the actual crisis. However to actually reform the system might require more creative design, of which some new approaches are investigated here. Section 1 introduces the US Social Security system and reviews what retirees can expect from social security and other assets. Section 2 reviews the crisis social security faces. Section 3 provides a review of the models and proposals for saving the system. Section 4 discusses models for redesign and reforming the system keeping in mind the original objectives of the system. Section 5 concludes. Unless otherwise noted we will be referring to US Social Security.

Keywords

social security, old age pensions, defined benefits, defined contributions

JEL classification: H55, J26

1. Background introduction to US Social Security

Social security, the first pillar of retirement income, is where demographics and public finance are most explosively entwined. In public perception it stands for a guarantee of income for retirement and the expectation that this promise will be fulfilled. The other pillars are occupational plans and personal savings. While it is difficult for a large modern nation to go bankrupt or renege on its obligations, fulfilling current expectations of publicly guaranteed income in retirement could cripple most nations. There are weaknesses or uncertainty in each of the other pillars as well. While default would be disruptive, large company pension plans or insurance companies can and have gone bankrupt leaving recipients without funds or avoiding bankruptcy by reneging on their obligations. Personal savings are hard to accumulate and manage for the average worker. Thus social security is *the* main pillar that most people rely on. Changes in social security present a moral dilemma for governments since the public expects to receive the promised benefits

Different countries face varying degrees of vulnerability in their funds for retirement and each came to the current reliance on social security by different path. The first income security insurance program was established by Otto van Bismarck in 1881. This plan paid out the insurance at 70, at which age few working males were still alive (at that time Bismarck himself was 74). The purpose was to offer protection from poverty when workers physically could no longer work and to protect widows and orphans of working males who died young from loss of income. The minimum age was lowered to 65 in 1916 and became the initial standard (model?) of the US Social Security program. See Table 1 for a history of the German public pension system.

The US program began in the mid-1930s. Ida May Fuller of Ludlow, Vermont, received the first social security check 00-000-001 on January 31, 1940. She lived to age 100 and collected $22,889 having made contributions of $24.75. Here we see the first overhang or deficit created by initial social security recipients who were already well into their working years when the program was implemented. By 2001, one in six Americans were receiving monthly benefit checks either as elderly, disabled or children whose parent had died. See Table 2 for the pension and health care spending as a percent of public spending in 2005 and projections for 2050 in various OECD countries.

Social security was originally defined as an insurance against poverty in old age. Its formal name is Old Age and Survivors Insurance. As such it has been very effective. Without social security 47.6% of the elderly would live in poverty, with Social Security payments, poverty rates are brought down to 11.9%.

At some point there was a switch in perception of the program from insurance to entitlement—compensation for a lifetimes contribution. It was no longer insurance but became replacement income. This expectation in turn led to higher benefit payments, cost of living adjustments, etc. All of which created more of a financial overhang.

The changes in the last 60 years in the US are striking: In 1910 the average retirement age for men was 75, in 1940 it was 68, by 2001 it was about 62. In 1960 men were expected to spend 50 of their 68 years of life in paid work. In 2000 they worked for only

Table 1
The German Public Pension System from Bismarck until today

1889/1891	Capital funded disability pension Old-age pension for workers 70+ Employer and employee share equally	1977	Pension splitting for divorced couples
1913	Retirement at age 65 (white-collar workers only)	1978	Minimum reserves are reduced to one month
1916	Decrease in retirement age for disability pensions from 70 to 65	1986	Benefits for child education (usually one year) Equal treatment for men and women for survivor's pensions
1921–1923	Inflationary compensation	1992	Integration of the German Democratic Republic Indexing of pensions to net wages and salaries Step-wise increase of retirement ages for unemployed, disabled, and women Introduction of actuarial adjustments for early retirement Significant reduction in years of education counting toward service life Benefits for child education are raised to three years of service life
1923	Retirement at age 65 (blue-collar)	1998	Value-added tax is increased to stabilize contributions to the GRV Introduction of the demographic factor
1929	Retirement at age 60 for elderly unemployed (white-collar workers)	1999	Introduction of demographic factor is revoked Early retirement for women and unemployed are restricted; only for the long-insured and with benefit adjustments Exceptions for disabled persons Ecological tax is increased to stabilize contributions to the GRV
1957	Conversion into PAYG system Contribution related pension benefits Objective: safeguarding the standard of living in old age Benefits: indexed to gross wages and salaries Normal retirement age 65, age 60 for unemployed (blue-collar workers), age 60 for women	2001	Transition to multipillar pension system (Riester reform) Reduction of first pillar pensions through modified gross indexation Strengthening of capital funded second and third pillars by subsidies and tax relief Redefinition of "disability" Further allowances for child education Higher value in terms of recorded years of service life Additive recording of employment possible Bonus for part-time employment Reform of survivors pensions Expansion of eligible income base Reduction of survivor's pension benefits Introduction of a child bonus Optional pension splitting for married couples

(*continued on next page*)

Table 1
(*continued*)

1968	Pure PAYG-system with minimum reserves for three months	2002	Minimum reserves are reduced to two weeks
1972	Public retirement insurance system open for all citizens (self-employed, housewives) Ex post payment of contributions becomes possible Flexible early age for insured with a long service life (63) and disabled persons (60) New minimum pension mechanism	2004	Introduction of a "sustainability" factor, effectively transforming the PAYG pillar into a quasi-NDC system

Source: Börsch-Supan and Wilke (2003).

Table 2
Ageing-related public spending pressures

	Pension		Health care		Total		Change
	2000	2050	2000	2050	2000	2050	
Belgium	6.3	13.0	6.3	10.6	12.6	23.6	11.0
Canada	4.7	6.4	6.3	10.5	11.0	16.9	5.9
France	12.1	14.3	6.9	9.4	19.0	23.7	4.7
Germany	11.8	13.8	5.7	8.8	17.5	22.6	5.1
Italy	14.2	14.4	5.5	7.6	19.7	22.0	2.3
Japan	7.9	8.5	5.8	8.2	13.7	16.7	3.0
Netherlands	5.2	8.3	7.2	12.0	12.4	20.3	7.9
Sweden	9.2	10.8	8.1	11.3	17.3	22.1	4.8
Switzerland	7.2	10.8	5.8	10.3	13.0	21.1	8.1
United Kingdom	5.0	5.6	7.9	11.0	12.9	16.6	3.7
United States	4.4	6.2	2.6	7.0	7.0	13.2	6.2

Source: OECD (2005, p. 21).

38 of their 76 years. As recently as 1965, about two-thirds of workers did not begin drawing Social Security benefits until they were 65 or older. Now, more than half retire at 62 or younger, and three-quarters receive their first benefit checks before they are 65 (Toner and Rosenbaum, 2005). The notion of an active retirement has been invented and retirement has become a 20-year stage in life. Company pensions clearly encompass three aspects: insurance (being old is like a disability), compensation (reward for a faithful career); and severance (payment to allow termination). Together these factors have led to an increase in the liabilities of the overall system. Under the social insurance program, workers earn entitlement to family benefits upon retirement, disability or

death. This breaks down into retired workers (61%), disabled workers (10%), families of retired and disabled workers (12%), and survivors of deceased workers (17%).

People have been surviving longer: in the US the life expectancy at birth is 76, but the additional expectancy at 65 is 17 years—and growing, giving a revised life expectancy of about 82. Along with lower birth rates, this is contributing to the aging of the population. OECD (2005) simulations indicate how age distribution changes will increase pressure on public spending. Pension and health care costs are only partially offset by changes in child care and family benefits but will be accompanied by pressure for expenditures on labor force upgrading and life-long learning.

Social security began on a purely pay as you go (PAYG) basis: current receipts paying for current liabilities. This was fine when there were only a few retirees in the system. However, the dependency rates—the number of workers per retiree, has been falling. Originally there were 42 workers for each beneficiary. By 1960 this was 5 to 1, by 2000 3.4 to 1, by 2004 3.25 to 1. By 2050 every worker would be supporting half a retiree! This is a 50% dependency ratio. If trends continue, by the same year, in Japan and Italy the dependency ratio is expected to be 70%. Clearly this is not sustainable. In aggregate financial terms social security payments are estimated to be 4.2% of US GDP in 2010 and 6.2% by 2040. Spending on Medicare and Medicaid combined would rise from 4.6% of GDP in 2010 to 10.1% in 2040.

The number of retirees is predicted to increase from about 40 million in 2000 to more than 76 million by 2030. The oldest baby boomers turn 65 in 2011, and by one estimate a husband and wife who retire that year are likely to collect $700,000 in benefits before they die.

1.1. Some facts about US Social Security—what retirees can expect to get

Social security has reduced the poverty rate of the elderly from 35% in 1959 to 9.7% in 1999. But it does not provide a luxurious retirement. In 2006, the average monthly retirement benefit after a 4% cost of living adjustment is $1002 ($1648 for a couple both receiving benefits). That is just over $12,000 a year. The maximum benefit depends on the age at which a worker chooses to retire. Full retirement age in 2006 is 65 years and 8 months, rising to 67 for those born in 1960 or later. The maximum amount for 2006 for a person retiring at full retirement age is $2,053 based on earnings at the maximum taxable amount for every year after age 21. In 1961 this was $4,800 and in 2005 it was $90,000.

Forty quarters worth of points (a maximum four in a year) are needed to qualify for benefits. There is a minimum income for each credit earned: in 2006 it is $970, so to earn four credits you must have earned $3,880 in the year. Average income is based on the highest 35 years of income. If only 20 years were spent in employment the remaining 15 years are zeros, reducing the average. Based on this average, a primary insurance amount (PIA) is calculated. If one retires early there is a reduction in the benefit calculated at 0.00556 for each month retirement was taken early. Thus if one retires three years (36 months) early one receives 80% of the PIA. Also, for early retirement, benefits are reduced if the recipient continues to work and earns above a threshold. In 2006, beneficiaries below the normal retirement age lost one dollar of benefits for every two dollars of earnings above $12,480. All benefits are inflation adjusted annually measured by the Consumer Price Index (CPI). The PIA is calculated to replace 56, 42 and 28% of low, average and high earner income. Low income is defined near the minimum wage and was $11,554 in 1997 when the average was $25,676 and the high is at the maximum taxable amount.

The complication of calculating benefits raises the issue of the return on social security contributions. There are two aspects. The first is the political and regulatory one that sets benefits and contribution rates. It includes issues such as the fact that early participants received benefits with few years of contributions and the shift in contribution rates, benefits and retirement age recently mandated. This is considered an internal rate of return. The second relates to the return on the social security trust fund itself since it was created compared to other investments. We look at each in turn.

1.2. The role of social security versus other assets at retirement

Popular financial advice recommends that households should replace 65–85% of their preretirement income in retirement (Butrica, Goldwyn and Johnson, 2005). But there is little scientific basis for this estimate. An analysis using multiperiod stochastic programming is in Ziemba (2003); see also Consigli (2007). Housing and health care are two of the largest categories of expenditure. About 25% of older people have mortgages—some have refinanced them for retirement while those who do not qualify for mortgages might fund some of their retirement with reverse mortgages.

The overall role of social security in funding retirement is ambiguous. While a poll found that only 20% of Americans who are not yet retired expect social security to be their major source of income when they stop working, 39% of ***retirees*** said that it was their major source of income, On the other hand, social security appears to be becoming less important in overall funding. Leland and Wilgoren (2005) reported in the New York Times on the role of social security for the elderly. They found that the lowest third of the population rely on social security for more than 90% of their income; while the wealthiest third relies on social security for less than 50% of their income and the middle third for 50–90%. Thus the lowest third, who really depend on social security face a relatively large decline in standard of living as social security strives to replace about half or less of their preretirement income and they typically have been unable to garner other savings to fill the gap. To highlight the difference, Leland and Wilgoren report on two brothers which captures the range in benefits: one who had held a variety of low-paying jobs receives $502 a month; the other and his wife, who had steady careers at the local school district, collect a combined $2,400. Poterba et al. (2006) estimated that social security represents 50% of the discounted wealth of those with less than a high school education but less than 25% for those with college or graduate education.

Declining income replacement rates are now built into the social security law. See Table 3. From 2000 to 2030 income replacement is set to decline for low earners from

Table 3
Wage levels and replacement rates, 2003 and 2030

Earner type	Career average earnings 2002	Benefit at 65 as % of	
		2003	2030
Low	$15,600	55.6	48.9
Medium	$34,700	41.3	36.2
High	$65,500	34.8	30.0
Maximum	87,000	29.6	34.0

Source: Apfel and Graetz (2005).

55.5 to 49.1%; for medium earners 41.2 to 36.5% and for maximum earners 27.3 to 24%. In that same period Medicare premiums will amount to more of the social security benefit, going from 6 to 9.2% with further increases set bring it to 13.6% in 2050. Further erosion will occur through taxation of benefits. With all the various deductions and adjustments the replacement rate in 2050 will be only 26.9% for those retiring at 65 and 20.8% for those retiring at 62.

Where is the rest of the income coming from? In 2001, about half of all US families owned a tax-favored retirement account with median balance of $29,000. Older households had somewhat larger tax-favored savings, with a median value of $55,000 for the 59% of families age 55–64 who had such accounts. As expected, tax-favored savings are concentrated among high-income households; the top 20% of the income distribution held two-thirds the retirement savings accounts. The heavy reliance on social security among retirees up through the middle of the income distribution, the shift away from defined-benefit pensions, and increased use of 401(k) plans amplifies the importance of payout options that convert savings into guaranteed incomes during retirement (Reno, 2005). Table 4 shows the income shares to the 65+ population and demonstrates the heavy reliance of the bottom 40% on social security while the top quintile relies more on earnings, asset income and pensions.

To improve income in retirement there have been shifts toward both working part time and working longer (5% greater share) as well as receiving marginally more from employer pensions and out of social security and other assets in the period 1980 to 2000 (Social Security Administration). At the same time there has been a great shift from Defined Benefit (DB) (from 60 to 15%) to Defined Contribution (DC) (from 18 to 55%) pensions (with a shift from 22 to 30% covered by both). See, e.g., Poterba et al. (2006). This shows an increase in financial risk for the retired.

Total savings is falling though a greater share of net worth is in housing equity. Alicia Munnell (2003), who heads the retirement research center at Boston College states: “Nobody is enjoying much in terms of growth in net worth. No one has enough to support themselves in retirement for 20 years.” This is supported by the Fed’s 2004 survey which reports that just under half of all families held retirement accounts. The typical family’s savings (including retirement accounts) fell to $23,000, down $7,000

Table 4
Shares of income by source for persons aged 65+, 2002

	Social security	Pensions	Earnings	Asset income	Other
Lowest quintile	83	3	1	3	10
Second	82	7	3	5	4
Middle	64	15	7	7	4
Next to highest	46	24	14	13	3
Highest	19	19	36	24	2

Source: Apfel and Graetz (2005).

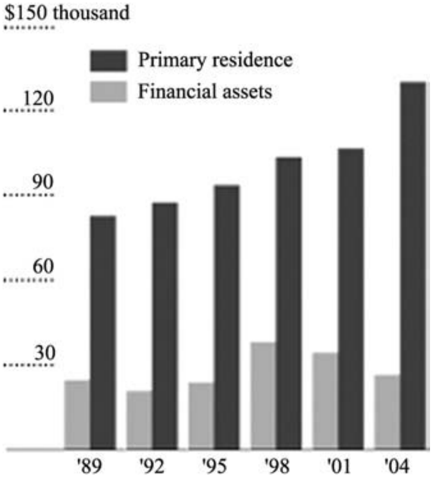


Fig. 1. Shifting assets from financial to housing among US retirees in 2004 dollars. Source. Federal Reserve Board.

from three years earlier. For households headed by a retired person, the typical savings fell from $34,400 to $26,500. Also, the 95% of Americans 55 to 64 who had any savings at all typically had $78,000, which is only about 1.5 times their median annual earnings.

Rich and Porter (2006) reported that housing wealth is increasingly funding retirement via reverse mortgages (the income of many of the elderly does not qualify them for normal mortgages). From 2001 to 2004 household net worth barely increased, to $93,100 from $91,700. While total savings dropped by 23%, the value of homes rose 22%. Just over 69% of Americans owned their own homes in 2004 (76% for retiree households). The median value of their homes jumped to $160,000 in 2004 from $131,000 three years before, a rise of 22% (for retirees value $130,000). While few plan to draw equity from their homes it is an insurance policy that more and more are calling on. Figure 1 shows the growing role of housing equity as compared with financial assets among retirees from 1989 to 2004.

1.3. Internal rate of returns—what is the return on the social security tax?

Early participants in Social Security were grandfathered in, receiving a windfall. But as the social security system was seen primarily as insurance this seemed justified. As the system matured, the huge windfall to early participants was eliminated and the real return on contributions declined. The benefit-to-contribution ratio for future generations of retirees is projected to decline further, assuming no changes are made to the system. A single man with low earnings aged sixty-five in 2029 is expected to have an average annual real rate of return from the OASI program amounting to 2.4%, slightly higher than the 2% average real return on government bonds. A couple with one worker who earned an average income would receive a substantially higher average real rate of return, 3.75%, upon turning 65 in 2029. On the other hand, a single male with high earnings would receive a much lower rate of return (0.72%). As an example for a single male worker born in 2000 with average earnings, the real annual return on his currently-scheduled contributions to Social Security will be only 0.86%. For workers who earn the maximum amount taxed (currently $80,400, indexed to wages) the real annual return is *minus* 0.72% indicating a redistribution of income toward the lower wage earners. As an insurance policy, the average-wage employee retiring in 1997 would regain social security tax contribution with interest after 6.2 years, minimum 4.4, maximum 8.1 (assuming increases match inflation).

Shoven and Slavov (2005) investigate the political or funding risk to social security from the increase in the future gap funding gap. This study is interesting because it clearly shows the impact of changes in the law on the internal rate of return (IRR) of individual cohorts. They measure the IRR for various age groups as payouts and contributions have changed over time. They consider age cohorts from 1900 to 1985, assuming that one starts work at 20, retires at the normal age of 65 and lives until 80. That is 20 years schooling, 45 years working and 15 years in retirement. With a constant population, that translates to a dependency ratio of 33% or, including youth, of 78% overall. They assume rational expectations, that is optimizing foresight. The base case results are shown in figure (for average earnings). Their results are presented in Figure 2. Observe the much higher IRR of the earliest cohorts. Other figures show that higher income individuals receive lower IRRs. What is significant is the variation in the IRR over a cohort's lifetime is unrelated to income but determined by funding risk as seen from the dips before major legislative changes. Geanakopolos, Mitchell and Zeldes (1998a, 1998b) also showed that IRRs for social security are below historical returns. While early recipients received more than they put in; future recipients will receive negative transfers. About $7.9 trillion net was transferred to those born before 1917; $1.8 trillion net to those from 1918 to 1937 creating a very high rate of return indeed! They estimate that starting in 1997, recipients need to give up about 3% of the value of the tax—this meshes with the estimate of about 3% extra to close the gap (the $10 trillion transferred to past cohorts) of promises.

There is considerable variability and the model's IRRs fluctuate for two reasons: changing law and changing expectations of inflation and wages. The IRR for the 1975

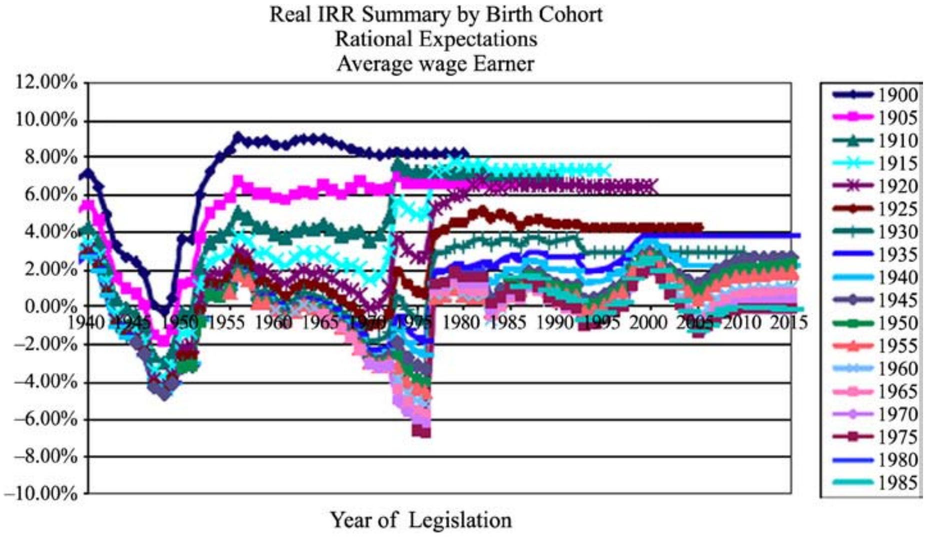


Fig. 2. Internal rates of return for social security. Source: Shoven and Slavov (2005).

cohort average earning single male was −0.95% in 1993; it rose to +2.04% by 2000 and fell to −0.39% by 2004. Workers' expectations were updated to reflect strong real wage growth in the late 1990s, resulting in higher IRRs through wage indexation. On the other hand, the impacts of the law changes are seen in 1983 and 1994. The IRRs for the average and high earners fall significantly as their benefits become subject to taxation.

Shoven and Slavov have exposed the political or funding risk of pure PAYG plans. In contrast defined contribution (DC) plans are subject to market risk. They ask which is better and conclude that a combination is best. Including both can address the mixed goals of social security and ownership in an annuity like asset instrument.

2. The crisis—is there one and what should be done about it?

The good news is that we are living longer. The bad news is that we cannot retire. The projections show the unsustainability of the current system, forcing governments to respond with various reforms or system redesigns. See Table 5 for an overview of some of the reforms mandated by the G10 countries. The individuals, think tanks and research groups who have assessed the vulnerabilities include:

- US Social Security, www.publicagenda.org/issues/frontdoor.cfm?issue_type=ss.
- Center for Retirement Research, Boston College, www.bc.edu/centers/crr/publications.shtml.
- NBER, /www.nber.org/socialsecurity/.

Table 5
Reformed pension systems in G10 countries

	Last reform	Contribution rate	Replacement rate	Public benefit type[5]	Mandatory private	Last reforms	
						Level of DB	Increase CR
Belgium	1997	16.4	41	DB	no	reduced	yes
Canada	1997	9.9	43	DB, P	no	no	yes
France	2003	16.5	53	DB	no	reduced	yes
Germany	2001	19.5	46		no	reduced	yes
Italy[1]	2004	32.7	79	NDC		abolished	no
Japan[2]	2004	18.3	50	DB, NF	no	reduced	yes
Netherlands	2004	28.1	68	DB	quasi	reduced	yes
Sweden[3]	1998	18.9	65	NDC	quasi	reduced	no
Switzerland	2003	23.8	58	DB	yes	abolished	no
UK[4]	2004	23.8	37	DB	no	no	no
US	1983	12.4	39	DB, NF	no		no

Source: OECD (2005, p. 23).

[1] for people retiring after 2010;

[2] projection to 2025;

[3] revaluation factor of 1.6%;

[4] also funds other things;

[5] DB defined benefit, NDC notational defined contribution, NF nominally funded, P prefunded.

- MRRC Michigan Retirement Research Center, www.mrrc.isr.umich.edu/.
- The Pension Research Council, University of Pennsylvania, prc.wharton.upenn.edu/prc/prc.html.
- CATO Institute on Social Security Choice, www.socialsecurity.org/.

  The proposed reforms to save social security are:

(1) to improve the PAYG balance by increasing contributions, lowering benefits and increasing the retirement age in recognition of the increasing dependency ratio, and
(2) to prefund the liabilities in an attempt to catch up with the overhang created when the system went into effect.

The Panel on the Privatization of Social Security (1998) summarized the questions facing social security: (1) PAYG vs. prefunding; (2) diversification vs. T-bonds; (3) individual vs. collective, and if collective should there be; (4) individual choice and (5) should it be centralized or decentralized. These issues are explained below. Prefunding has created the social security trust fund (SSTF) which has transformed social security from a purely PAYG system to a mixed system. This in turn has raised questions as to the role of the trust fund and how it should be managed. Once the trust fund was created many issues of property rights, privatized accounts, etc. could be raised. But at heart, social security remains a social insurance policy to prevent destitution in old age or incapacity to work.

There are two important aspects to the social security crisis: one is the financial and the other is the security. What is at risk of being lost is the implicit guarantee that old age pensioners will not live in poverty. Unlike corporate pensions and private savings for retirement, social security systems were designed to be PAYG schemes. As a PAYG scheme—equivalent to a tax on current income—this type of system insures that there are sufficient savings in real product at each year to cover the needs of pensioners. This relies on the productive capacity of the economy and the ability to tax earners—thus the importance of the dependency. Corporate pensions though prefunded, still depend on the savings propensities of current workers to release sufficient goods and services for the retired, non-working population. Enforced savings may look more benign than taxes but they need to accomplish the same end.

The first level accounting imbalance will be felt about 2017 when the current contributions will not cover the current payouts. At this point the US government will dip into the SSTF and begin to draw it down. In accounting terms a true crisis does not arrive until the SSTF is depleted about 2042 at which point higher taxes would need to be collected for the purpose of paying benefits. However, even as soon as the government dips into the SSTF, current earners will have to begin saving a higher proportion of their incomes to cover the real goods demand of the retirees without inflation and this will be felt. When the SSTF is exhausted the system's revenue will cover 74% of promised benefits. Full funding would require an increase in the total tax rate from payroll and benefit taxation from the current 12.4 to 17.5%. By 2075, the tax rate necessary to fund full promised benefits would equal 19.4% of payroll. (This is about where Germany is in 2006.)

As well as demographic issues there are a variety of intragenerational equity issues. Social security is a regressive tax with an upper limit on earned income taxed. Further compounding this inequity, poorer groups have lower life expectancy, e.g., only 65% of black men age 20 will live to age 65 so 1 in 3 will pay and not receive benefits and divorced women do not share benefits accrued by their former spouses.

Social Security is not the only such liability. Social Security retirement, disability and survivor insurance, unemployment insurance, and the Medicare health insurance for those age 65 and older together accounted in 2003 for 37% of federal government spending and more than 7% of GDP.

Estimates of the cost to close the gap vary. The range includes: borrowing as much as $11–40 trillion over the next several decades to make up for the lost revenues and pay retirees benefits earned under the old system. Eliminating the 75-year deficit without touching the benefits of those currently 55+ and the disabled would require a cut in benefits of 20% (13% if unconstrained).

Though he investigates privatization, Feldstein (1997) first discredits the idea that the current system is bankrupt and unsustainable. Bankruptcy does not apply to government and the system could be sustained with an increase in taxes to about 16% after 2030 and then to 19% from then current 12.4%. This would increase the payroll tax from 5 to 7% of GDP. While significant it is not impossible. He postulates a *typical* individual who works from 25 to 65 and retires from 65 to 85—making the mean in each block 45, 75.

He compares the contribution with the payout and calculates that the payout is 7 times the contribution—he assumes a 9% rate of return on savings and an implicit interest of 2% would have the contribution grow to 1.8 leaving a deficit of 7+ and from this he concludes that a prefunded system would require an extra 2.7% contribution per year. He also considers the transition to privatized social security, which we look at later.

The extent of the problem is about an $11 trillion present value gap. Krugman, Tanner and Marshall (2005) put this in perspective by comparison with the Bush tax cuts that created a $20 trillion loss of revenue. Reversing these cuts would more than pay for the gap. Whatever type of reform happens, this over hang of promises must be paid for.

As with much of economics, there are two well-articulated camps for this issue. On the one side are those who see bankruptcy and unsustainability in the numbers and those who believe that incremental changes will keep the system going. The difference between the two is related to issues of social insurance versus a fully funded property right in an asset; those who see a role for government and those who do not. We now further explore some of these issues.

3. Plans for saving social security

The ideas to save social security encompass two main ideas: one changes the contribution rates, benefit payout and age to retirement to achieve balance and the other attempts to improve the return on the funds held in trust by exploiting higher returns in stocks.

To rephrase the problem: in the beginning there was an extra return to early participants and the ratio of dependents to supporters was very low. This windfall is in the very nature of a PAYG system as early retirees will not have contributed to the fund for their entire working life thus creating an immediate 'deficit'. This continued in effect while coverage expanded and while the population was growing. Geanakopolos, Mitchell and Zeldes (1998a,1998b) estimated that the transfer to those born before 1917 was $15.7 trillion. Now with a more stable population and greatly improved life expectancy the bonus is dropping and the return on contributions is falling to the steady state rate first identified by Samuelson (1958) as the rate of growth of total wages—labor growth plus productivity improvements. This has led to the consideration of ways to improve the return to cover all the economy including the return to capital. For this there would need to be a direct way for the social security savings to tap into capital formation. This has led to the proposals for investment of funds in the stock market. Unfortunately there is not a direct link between stock market investment and capital formation. This gives rise to two opposing views: that investing some of the contributions will improve the accumulation in the social security accounts and the other attesting that the impact will be nil as the added investment would raise share prices and benefit current owners and later depress prices just when people would need the funds. So part of the issue is how to deal not just with the financial issues of allocating the funds but also the real economic effects of creating sufficient goods and services to meet the needs of the retired.

For the most part the models used are simple economic equilibrium models, simulations and accounting balances. Some investigate the impact on savings, age at retirement, deadweight loss. There are no real optimization models except to the extent that economic models maximize utility. Some studies use overlapping generation models: mostly they use two or three generations (work and retirement and possibly youth) but probably need cohort models of at least five groups: youth, early work, later work, retirement—then could have a better analysis of the time frame and each could represent 20 year periods.

One important aspect of the debate is the types of risks faced by retirees: longevity risk; spousal survivorship risk; inflation risk; investment risk—and specifically how they are borne. For example, a life annuity is a financial product that allows a retiree to shift longevity risk and investment risk to an insurance company but opens up credit risk.

3.1. Increase contributions, cut benefits, extend working life

Public pension plans around the world face the same demographic challenges as US social security. In many cases the demographics of comparable countries have even higher dependency ratios. Almost universally the reforms have been met with protests. Canada however has been an exception and will be discussed below.

Increasing reserve funds by increasing the contribution rate would provide security that the system will not default but is a drain on the system similar to a tax increase. It is possible that slower growth in some European countries is due to increased contribution rates and which lowers consumption and growth. Model simulations at the OECD (2005) show that with the proposed types of reforms (increased retirement age, lower payouts, increased contribution rates) only about a 1% increase in private savings would be expected.

3.2. Lowering the liability by increasing the retirement age

In the US the retirement age is set to increase from 65 to 67. However, there are obstacles to extending the age to retirement. In Japan and many European countries, one of the biggest is a pay scale that increases with age. In France and Germany 50- to 65-year-olds earn 60–70% more than 25- to 30-year-olds. In Britain where they earn about the same, employment of older people is higher. This is seen in a growing disparity of unemployment rates among 55- to 64-year olds in various European countries: in 1990 it was roughly even at about 7% while in 2003 it was only 3.3% in Britain and 9.7% in Germany (some of this may be related to the reunification of east and west).

There is a correlation between the average retirement age and the rate of employment of older workers. In Sweden with the highest average retirement age of almost 64 the labor force participation rate of 55–64 age group is almost 70% while in Poland with the lowest average the retirement age of 57 has a participation rate is about 25%. The other countries fall in between. Switzerland is an anomaly with a retirement age of 61 but a participation rate of almost 70% (equal to Sweden) (The Economist, 2006).

There are social aspects that limit the extension of the working age. In some countries like China, early retirement is institutionalized to provide a good career path for younger workers with women retiring at 50 and men at 55. On the other hand, later retirement might also mean that younger workers would end up supervising older ones.

Switzerland increases the state pension by SFr 5,000 ($3,825) per year for people who stay at work for up to five years beyond the statutory retirement age. This has been effective: more than 60% of all 55- to 64-year-olds are in work, compared with less than 30% in Italy and Belgium. In Britain a person cannot receive a pension and a salary from the same employer and in the US, pension plans often withhold benefits from a retired person who is rehired or works for more than 40 hours a month. In Japan companies rehire retired employees on one-year renewable contracts (The Economist, 2006).

A good retirement scheme should not discourage workers who want to continue to work. Diamond (2003a, 2003b) explores age related issues close to retirement. The appropriate design combines eligibility and benefit levels. He assumes a single payroll tax rate that does not vary with age and continues until eligibility or as long as work continues and assumes that once benefits start they are paid as a real annuity. One issue is in the cases where benefits vary according to age, earnings and service if they are delayed. With fixed benefits, the decision is whether to stop work and how much extra savings to make.

Estimating the deficit as equivalent to a 2% payroll tax, the Social Security Network identified the savings from each of the major proposed incremental changes to reduce the deficit. The sum of these incremental changes would close about 70% of the gap.

		Gap closed
1.	Increase the expected years of work from 35 to 38 this would include an additional 3 years of earlier lower pay work and would reduce the benefits.	12%
2.	Change the way in which benefits are taxed. For example, any benefits received beyond what was contributed would be taxed as ordinary income. Some individual retirement accounts (IRAs), already are taxed in this manner. Currently, taxes are only paid for benefits above a certain level and only on a portion.	14%
3.	Extend Social Security coverage and participation to state and local government employees who are currently excluded from the program bringing more revenues into the system.	10%
4.	Accelerate the scheduled increase in the retirement age so it reaches 67 by 2011 (now scheduled for 2025) and index it to longevity after 2025.	22%
5.	Adjust the Consumer Price Index (CPI). The Bureau of Labor Statistics has a revised definition that would reduce the CPI by 0.21% per year.	14%

Various reforms have been suggested to deal with investing the SSTF in equities, means-testing Social Security benefits, and “privatizing” Social Security.

Kotlikoff (2003) uses a “menu of pain” to resolve the funding gap. The choices range from an increase in federal income taxes of 69% and a cut in Social Security and Medicare benefits of 45%. His simulations on closed and open developed economies project a large rise in interest rates even with doubling of immigration as dependency rates increase. Kotlikoff and Burns (2004) discuss families intergenerational conflict issues.

Kotlikoff, Smetters and Walliser (2004) using a general-equilibrium overlapping-generations model with realistic patterns of fertility and lifespan show that future generations would be harmed during the demographic transition due to rising payroll taxes and declining real wages. A faster rate of technological progress would mitigate only some of the payroll tax increase and would not prevent the decline in real wages. Addressing the problem by reducing Social Security benefits as needed or by raising the eligibility age imposes major welfare losses on current or near term retirees. The model suggests the need for a 50% phased in reduction of benefits. By contrast, prefunding Social Security through consumption taxes more evenly spreads the welfare losses across generations, and it helps future generations, especially the poor, by stimulating capital formation.

3.3. The role of the social security trust fund

In the 1983 Greenspan Commission Plan moved social security from a pure PAYG to a partially funded plan by creating the social security trust fund. This is not treated as a separate entity but is only an accounting entry and the funds are used ("borrowed") under the 1970 unified federal budget. Diamond (2003a, 2003b) found that the existence of trust funds increased national saving. However, Nataraj and Shoven (2004) found that what has happened instead is that the government has spent more. This exacerbates the problem that will exist when PAYG runs out. If the federal funds budget instead of the unified budget had been balanced, the total debt of the government and the publicly held debt would have been \$3 trillion lower and future generations would have been \$3 trillion wealthier. Of course, this would have meant that there were no large surpluses in 1998–2000 and the 2004 budget deficit would actually be about \$200 billion higher than reported and, the tax cuts were not a rebate of surpluses. Therefore while the existence of the trust fund will prolong the financing of social security, as the trust fund does represent real claims on the rest of the government, from the perspective of future generations of workers, the trust funds do not represent incremental wealth and will still require additional taxation. In reality then the social security contributions are just another layer of taxation.

Oshio (2004) investigates the size of the social security trust fund the government should have and how it should manage the fund in the face of demographic shocks. He uses an overlapping-generations model with both closed and open economies. He shows that with an aging population, a trust fund could achieve the (modified) golden rule (pairs of tax and trust fund) to offset the negative income effect of a PAYG system. In a closed economy where factor-prices effects dominate, using the trust fund as a buffer for demographic shocks could lead to a widening of intergenerational inequality. He shows that with lower population growth it is difficult to justify a pure PAYG system that has no trust fund because a PAYG system tends to depress capital accumulation in an aging population. Secondly, lower population growth needs a larger trust fund to achieve the modified golden rule to offset the negative impact of a PAYG system on national savings by holding a larger trust fund. He applies this to Japan where

Table 6
How the annual contribution to a privatized account will impact the replacement earnings with an annuity

	Age (years of contribution)			
	2044 (40)	2034 (30)	2024 (20)	2014 (10)
Account balance as % of career average annual earnings at 2% rate	171	119	60	22
Total accumulation for $34,700 earner	59,337.00	41,293.00	20,820.00	7,634.00
Single life annuity as % of career average annual earnings, 2% annuity	11.6	7.8	4.2	1.6

Source: Apfel and Graetz (2005).

the government has a huge social security trust fund equivalent to about five years of benefits, compared to about three years of benefits in the United States and only one to two months of benefits in the United Kingdom, Germany, and France. Policy implications depend on the "initial state". If in the PAYG system the government has promised the current elderly (and even the old working generations) to more benefits than their own tax payments, the postponed burdens will at least partly offset the factor-price effects which future generations are expected to enjoy and reducing the trust fund may adversely affect intergenerational equity.

Just transferring a balance from the Trust Fund to a private account is not sufficient to gain in real terms. Table 6 shows the effect of saving 2% of your income at a 2% rate of return on replacement income. A medium-earner, averaging about $34,700 lifetime earnings and retiring at 65 after 20 years in the 2% account system, would have a balance of about $21,000 (60% annual earnings) and single-life annuity income at age 65 of about $1,500 a year, or about $125 a month. A scaled medium-earner who participated in the account system for 40 years would accumulate approximately $59,300, yielding a single-life annuity of about $4,000 a year, or $333 a month. So with saving 10% per year, the worker would end up with a replacement income of about 55%. This 10% is a lot in a country with an almost zero savings rate.

3.4. *Use the contributions to buy stocks instead of government bonds*

Some suggest that the trouble with social security is that the Trust Fund is not optimally invested. It earns the interest on government bonds while the market has returned on average 4% more per year (see, for example, Dimson, Marsh and Staunton, 2006). Is there a way to tap into this to solve the problem?

- Saving $1,000 yearly for a 44-year working career would accumulate $63,000 in the SSTF.
- Over the last 75 years the compounded annual real return from a balanced fund of 70 percent stocks and 30 percent bonds was 6.2%. Saving $1,000 per year for 44 years at this rate of return would result in about $211,000.

William G. Shipman in the *New York Post*, August 14, 2002

Canada instituted such a reform in 1997 without much fanfare. The government was cognizant of the desire of Canadians to preserve the Canada pension plan yet needed higher assets. In 1997 the federal and provincial governments adopted a balanced approach to reform to bring the plan to sustainability a manageable contribution rate of 9.9%. There was limited change in benefits, a moderate increase in contribution rates and the creation of an asset pool to be invested in the markets and managed at arm's length from government for the best possible rates of return. This moved the CPP from a PAYG system to partially funded system. The market investment policy is implemented by an independent organization, the Canada Pension Plan Investment Board, set up in 1998 investing by 1999. The CPPIB reflects a fundamental policy change in policy. Prior to the CPPIB, the policy was to invest surplus funds in provincial government bonds at the federal government's investment rate. This was an undiversified portfolio with an interest rate subsidy to the provinces much as the US social security trust fund. The CPPIB investment policy is similar to other large public pension plans in Canada, such as the Ontario Teachers' Pension Plan and the Ontario Municipal Employees' Retirement System. It invests in a diversified portfolio of market securities (including international) under investment rules which require the prudent management of pension plan assets in the interests of plan contributors and beneficiaries and is free to hire its own independent professional managers.

The Canadian Pension Plan (CPP) was established relatively late in 1966. The contribution rate was 1.8% of gross income. In 1997 the contribution rate was increased to 9.9% of pensionable earnings. The CPP is transitioning to have 20% prefunded by 2017 though the creation of the CPP reserve fund. Their goal is a 4.41% real (inflation adjusted) return and to date they have achieved 4.6%. The CEO, David Denison, has remarked that barely 25% of Canadians are aware that the CPP has been successfully reformed and that the CPP has a pool of assets to sustain it into the future, figuring instead that its vulnerabilities likely mirror that of the US. Instead of the US system of holding government debt, the CPP is diversified: 58.5% in publicly traded stocks, 27.7% in bonds, 4.5% in private equities, 8.7% in real return assets and 0.6% in cash and money market securities in both domestic and foreign assets. Total assets are projected to reach C$250 billion by 2016. While Canada, about 10% the size of the US, can accomplish this shift with little impact on the market, it would be hard for the US to do the same. Also many of the US plans are more complicated than just investing the funds in assets, calling for a switch to user-managed private accounts that are defined contributions not defined benefits, but we will look at a number of the variants.

The first stumbling block is the US government debt and the role the social security trust fund plays in sustaining that debt. The SSTF has been used in general revenues.

1. If the Social Security Trust Fund had not been used in general revenues, what would have happened to interest, foreign exchange, etc.?

2. What would be the impact on stock market returns if a greater percentage of social security assets are invested in the stock market? Would returns be lower than historical returns?[1] What is relative share of social security savings to market capitalization?
3. What is the distribution of risk and ability to deal with shocks?

"Using government resources to buy stocks and bonds, without other spending and tax changes, would not automatically lead to an increase in the nation's pool of investment resources," the budget office concluded. "There is no such thing as a free lunch." The AARP says that "many analyses that tout individual accounts employ an unfair comparison with Social Security. They claim rates of return on stocks assuming that there are no transition or administrative costs." When you account for the increased risk of equities standard finance theory would suggest that this would not add a lot of return and the return would be compensated by the extra risk and therefore eventually an added burden on the state to ensure in any event that the elderly poor are provided for. This is supported by Diamond and Geanakopolos (2001), who show that diversification into equities helps but only up to a limit. Diversification is likely to cause prices to rise and taxes to rise as other sources of funding for government services become necessary—so this is not an unmixed blessing. However, some redistribution is good. Orszag (1999) of the Brookings Institution argues that "when analytically accurate comparisons are undertaken, widely trumpeted gaps between rates of return for individual accounts and Social Security contributions essentially disappear."

The OECD (2005) report on *Ageing and Pension System Reform* acknowledges that the shift from public pensions to market based provision for retirement by market investments also shifts risk to the individual. The report recommends that the government ameliorate this risk by facilitating new financial instruments including issuing more long term debt and index-linked bonds. It notes that this might increase the cost of carrying the debt. As well, they observe the need to deal with longevity risks (tail risks) to which the government is already exposed but which it might need to absorb more of as the market is unlikely to do so.

Alternatively, the SSTF could be invested in socially relevant projects to expand productivity and capacities in goods and services that will be consumed by the elderly-investment in health care and health care research, public transit and other infrastructure projects.

Many studies begin by asking how to get more higher returns without losing out on benefits guaranteed to participants. Feldstein and Samwick (2001) is typical. They investigate saving more and investing it in a productive way. They impose the constraints that those who are now retired or will soon retire will receive the full PAYG benefits specified in current law, the existing payroll tax rate and base will not be increased. Future retirees would have at least the projected PAYG benefits and that there will be a permanent financing solution for Social Security that establishes solvency over the actuaries' 75-year forecasting period.

[1] Impact of placing 12.6% of wagebill (60% GDP) in the market . . . how does this compare to market cap?

Feldstein (2005a, 2005b) investigates the issue from an economic welfare point of view and calculates the deadweight loss (loss to the economy without any compensating benefit in reallocation) from additional payroll taxes to balance social security. In 2004, a 6.5% increase in marginal tax on wage income would increase the deadweight loss by $96 billion or 1% of GDP. Accounting for wages as a share of GDP and the impact of reduced wages together leads to an estimated 50% deadweight loss for a true cost of $1 in tax of $1.50. Applying this information he suggests designing programs that have less of the element of a tax and appear more like an investment. The current rules linking benefits to contributions is murky making the social security payroll tax look like a pure tax. He suggests a PAYG notional defined contribution system (more on notational defined contributions later). In fact, he suggests that the social security program needs to be redesigned to keep pace with changing economic conditions and knowledge. When social security was originally designed during the great depression, its unfunded nature was important. The reforms should evolve as the economy evolves. In order to have the best design it is important to separate social insurance from income redistribution goals. He concludes that shifting to an investment-based or mixed system depends on four issues: the transition process and cost, administrative costs, riskiness and impact on the poorest participants. Feldstein points out that the mass of investment funds, starting at about half of GDP and rising to GDP would require individual accounts.

Earlier studies had investigated these issues from a variety of approaches. Feldstein (1997) compared two transitions to privatized accounts: immediate with *recognition bonds* or gradual. The immediate route would require $7–12 trillion depending on the real the rate of return (4–2%). This is the amount of the PV of the underfunding. This could be funded by a temporary contribution increase of 2%, then, as the existing cohort dies, the rate can be decreased—even accounting for a lower rate of workers to dependents. Mitchell and Zeldes (1996) investigated a two-pillar plan: a small indexed pension, equal for the full lifetime of work and DC accounts financed by payroll taxes. They found a reduction of political risk, an increase in household portfolio choice and improved work incentives but also less redistribution (one of the original purposes of social security) and risk sharing and increased administrative costs. The return is not likely to be higher as one would need to compensate for risk and still cover underfunded promises.

Feldstein and Samwick (2001) suggest funding the transition by a temporary increase of 2%, then as the existing cohort dies the rate can be decreased—even accounting for a lower rate of workers to dependents. In calculating the impact on consumers and the economy, they used 9% for the real total annual return on stock, dividends, interest, etc. (this is rather high), and suggest that the average return on portfolio investment in pension accounts etc. is about 5.5%. He allows for increasing tax receipts and has the government contribute 3.6% to individual accounts from this increase in taxes. Their analysis suggests a lower return resulting from an increase in capital stock which would likely lower the return from 9 to 7.2% and have higher administrative costs. There is inherent risk in both the current and proposed systems, just different types.

Biggs (2002), assuming a 60–40 stock-bond allocation for young workers going to 40–60 for workers 60–65, calculates the loss of only 3% of the portfolio in the 2001's portion of the crash of 2000–2001 comprised of 21.6% loss on the stock portion almost matched by the 9.9% gain on the larger bond portion, for a total year-end loss of 3.25%. There was also a 22% fall in 2002 in the S&P 500. See Ziemba (2003) and Ziemba and Ziemba (2007) for a discussion of these crashes. One thing in favor of bond-stock mixes is that most of the time both assets rise together but when stocks fall, bonds generally rise.

Some advocate for such changes stating: "stocks have never lost value over the long term". However, this is a statement of a stock index or the market overall—as individual stocks have lost value and companies have gone bankrupt. Also there is the risk that the market can crash just when the money is needed for retirement. For example, there were three periods with nominal gains of zero: 1899–1919, 1929–1954, 1964–1981. However, with dividends, all 20-year periods since 1899 had gains (Siegel, 2002) and Figure 6. One thing that diversification into stocks would do would be to raise the wealth of current owners, those that are already wealthy and older versus younger workers.

Four countries—Chile, Australia, England, Mexico have already begun some form of privatization, however, it is not a cure-all. Chile's move to privatized social security was highly touted. The Chile plan put 10% into a privately managed fund plus 3% for administrative costs and disability insurance. In the transition, recognition bonds were issued. Now it appears to have problems: recent studies reveal high management fees, low participation rates, unexpectedly heavy dependence on an inadequate safety net, and prohibitively high costs to government. This has led to largely disappointing results, leaving many Chilean workers with no reliable retirement plan. The UK has had some problems with fees and misselling such that the UK pensions commission warns not to emulate them.

Geanakopolos, Mitchell and Zeldes (1998a) find that privatization would not pay a higher rate of return than the current system. In their view the problem stems from mixing three effects: privatization (replace unfunded DB with funded DC), diversification and prefunding. Geanakopolos, Mitchell and Zeldes (1998b) provide more detail and a simple model. After-tax returns on privatized accounts would be identical to the low returns received under the old system. For example, in their stylized economy, workers receive only 71 cents in benefits in present value terms for each dollar of contributions to the current system. In a privatized system, each dollar of contribution would have to be taxed 29 cents to make payments on the bonds replacing accrued benefit obligations, yielding the same 71 cents of benefits.

The three effects are independent and any one could be implemented without the others. In part, the cost of taxes to cover these would eliminate the higher returns from diversification. Mexico has an individual account system where all the funds are invested in government inflation indexed bonds. Singapore collects sufficient taxes to prefund the system. Chile's system is prefunded as well as privatized.

Abel (2003) investigates the effect of a baby boom on the economy using an overlapping generations model with a random birth rate and the price of capital determined

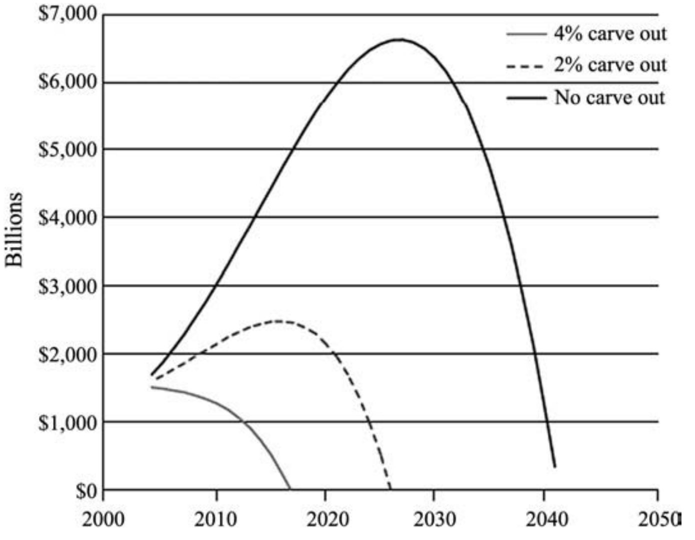


Fig. 3. SSTF with and without carveouts.

endogenously by a convex cost of adjustment. A baby boom increases national saving and investment and thus causes an increase in the price of capital. The price of capital is mean-reverting so the initial increase in the price of capital is followed by a decrease. Social Security can potentially affect national saving and investment, though in the long run, it does not affect the price of capital. This model has some restrictive assumptions including the lack of bequests and the lack of a link between rate of return and capital accumulation.

Anrig and Wasow (2005) bring up a variety of arguments against privatization. First, the current system is more than a defined benefit pension it also protects against death and disability. Privatization would make it worse not better by taking out more of the money from the trust fund, see Figure 3. This is supported by other studies including Diamond and Orszag (2004).

Privatization would limit government access to the trust fund dampening economic growth and increasing the debt, so in effect is only a shift in assets. The 2004 Economic Report of the President reported that the federal budget deficit would be higher by 1% of GDP every year for roughly two decades, reaching an increase of 1.6% of GDP in 2022. The national debt levels would be increased by an amount equal to 23.6% of GDP in 2036 for an increased debt burden for every man, woman, and child of $32,000.

The odds are against individuals investing successfully. They tend to pick investments with equal weightings therefore making the design of opportunities offered of crucial importance; see Bernartzi and Thaler (2001).

Stock market returns are variable. Burtless (1999) has studied the effect of this variability on the annuity that can be purchased on retirement. Figure 4 shows the 15-year

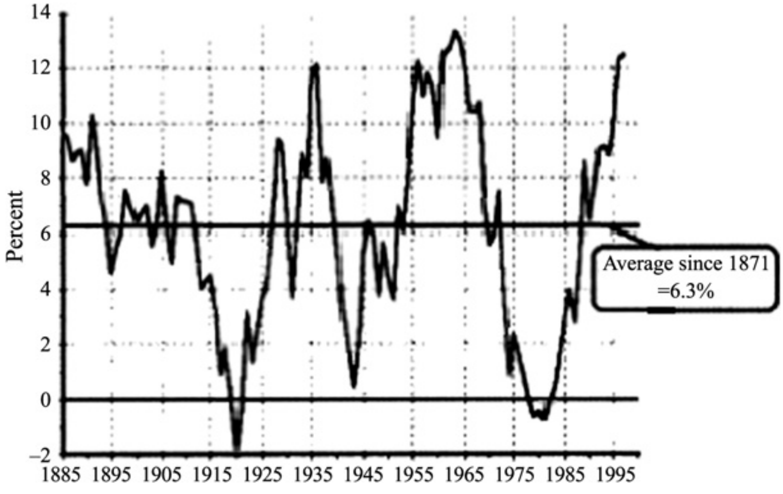


Fig. 4. Real stock market returns, 15-year training rates, 1871–1998. Source: Burtless (1999).

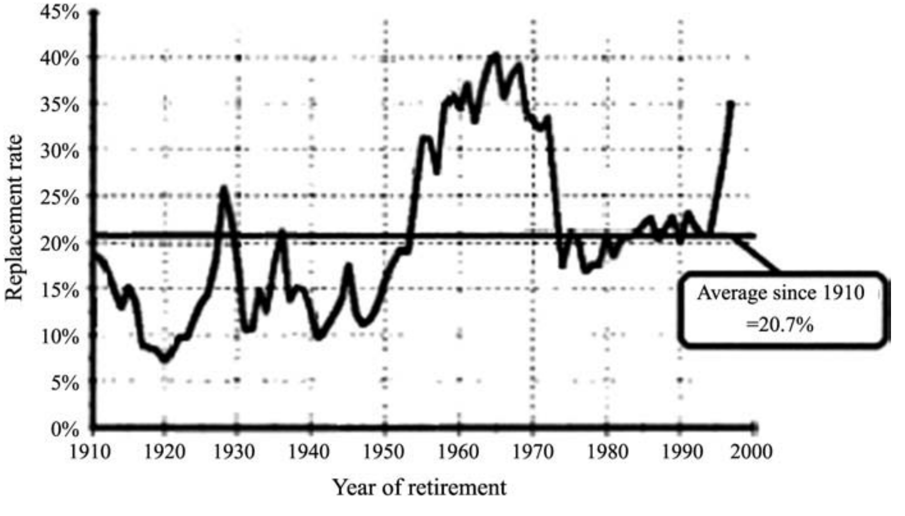


Fig. 5. The value of annuity, male single-life, as a percent of career high annual earnings at age 62.

trailing annual returns assuming $1,000 is invested in the S&P 500 and dividends are reinvested. Not only that but what you get on retirement will depend on whether you retire in a good year or a bad year, see Figure 5. He assumes that each worker puts 2% of earnings in the stock market every year (reinvesting dividends) and earned the actual historical return, year by year assuming no transaction costs and the purchase of a fair (no profit) annuity. On the basis of these assumptions the real internal rate of return averaged 6.4%. Replacement rates averaged 20.3%. The income from the annuities would be subject to erosion of purchasing power by inflation.

Where implemented, privatization has been disappointing. While greeted with enthusiasm in Chile, the plan covers only about half the workers and the investment accounts are not as large as expected due to high commissions and administrative costs. The returns have averaged only 5.1% instead of the 11% expected between 1982 and 1999 so that a regular savings account would have been had higher returns. This has meant that 41% of those eligible for pensions continue to work. Fully 28 to 33% of the contributions made by employees retiring in 2000 went toward fees. The military, which imposed mandatory private accounts, does not participate and continues to receive pensions under the old governmental system.

Investment firms would stand to gain. Goolsbee (2004) calculated the costs of privately managed individual accounts would likely reduce the retirement value of the accounts by 20% and cost nearly $1 trillion over seventy-five years. Since 1988 when UK workers have been able to open private accounts, management fees and marketing costs among financial intermediaries have taken an average of 43% of the return on investment.

Private accounts would create a new government bureaucracy to establish and track these new accounts, hiring perhaps ten thousand highly trained workers to oversee the accounts and answer questions from workers. By contrast, Social Security is a very efficient program with minimal administrative costs of less than 1% of annual revenues.

Young people would be worse off given historical rates of return. A study comparing President Bush's proposal (when Governor of Texas) to existing social security found that benefits for an average earning worker who retired in 2037 at age sixty-seven (someone aged thirty-five today) would be 20% lower than they are now, given historical rates of return over a fifty-year period.

Certain groups would be disadvantaged. Retirees would not be protected against inflation. Women stand to lose the most as they rely more on social security. Women over age sixty-five who are not part of a couple receive 51% of their incomes from Social Security versus 39% for unmarried men. African-Americans and Hispanics also would be more vulnerable under privatization.

Abel (2001) assumes that the fixed costs of participating in the stock market results in those with lower income will not participate and will instead increase their consumption thus lowing capital accumulation. His general equilibrium model indicates that privatization could reduce the aggregate capital stock substantially, by about 50% per capita. In his model, investors react to any shift of the social security system's portfolio to neutralize the effect of the change. For example, if the social security system sells a dollar of bonds and purchases a dollar of equity, private investors would buy a dollar of bonds and sell a dollar of equity. This is based on a set of assumptions about optimizing behavior and perfect markets.

Valdés-Prieto (2003, 2005) observes that social security systems periodically require adjustments due to demographic changes and economic uncertainty. Relying on discretionary legislation creates political risk for workers and beneficiaries and can raise risks

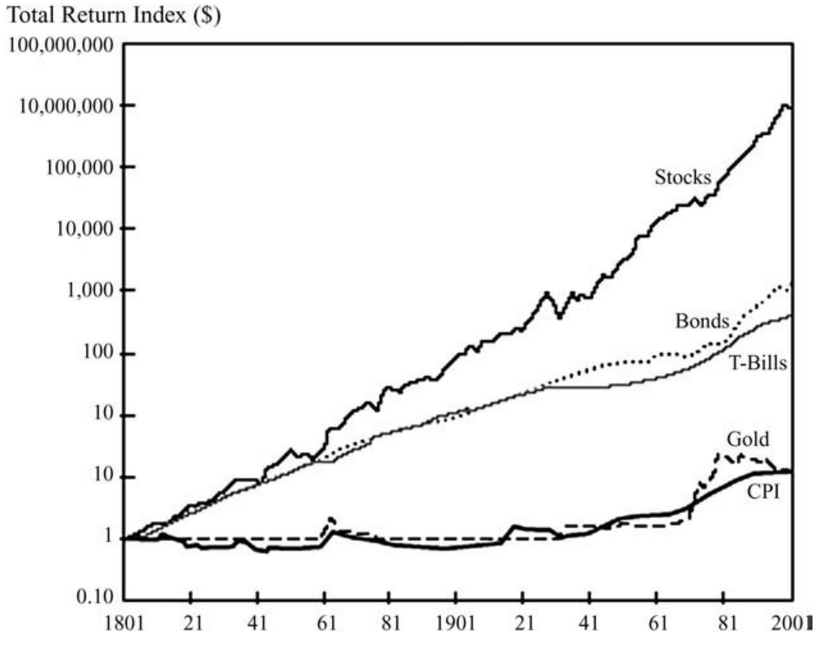


Fig. 6. Total nominal US stock market returns, 1801–2001. Note: CPI = Consumer Price Index. Source: Siegel (2002).

borne by taxpayers. An alternative approach is a rule-based adjustment, as in the case of funded mutual funds and life insurance plans that offer annuities. It is suggested that rule based adjustments can be adopted in an unfunded system as well, without incurring transition costs and without increasing the public debt. The approach would endow the Trust Fund with property rights over the revenue of a residual payroll tax paid by future workers. Revenue on the future taxes would be securitized and the resulting securities priced in financial markets. The new securities created in the process would allow beneficiaries to obtain safe real pensions protected from investment risk.

1. Determine the mix of tax increases and benefit cuts to honor current commitments;
2. In the reformed system, determine the size of average net lifetime taxes as a share of earnings. Convert this rate, currently an *implicit tax on earnings*, into an *explicit* residual payroll tax, and endow the Social Security Trust Fund with property rights over the revenue from this residual payroll tax. The remaining portion of the current payroll tax would be transformed into contributions to personal saving accounts;
3. Securitize the revenue of the residual payroll tax, creating new “Covered Wage Bill” (CWB) securities and set their prices in competitive financial markets;
4. Change the social security benefit formula to link benefits to the overall financial return achieved by the Trust Fund. Benefits for initial workers are similar to the current ones because they would initially get portfolios heavily invested in CWB securities.

Initial workers would benefit from the cut in the political risk they currently bear; and

5. Each participant can select from a limited set of balanced portfolios and choose his own asset manager to administer his personal account assets.

Proposals that a portion of the SSTF assets be invested in equities raise the possibility that the fund's assets will fall below the level needed to pay currently mandated benefit levels. In such a scenario, existing taxpayers might have to become insurers of last resort. Constantinides, Donaldson and Mehra (2005) calculate the cost of a put option that would guarantee to cover this shortfall. They calibrate a model using realistic equity premia. Their formulation accounts for the non-stationarity of security returns resulting from trust fund purchases and other phenomenon. If 20% of the fund's assets are in equities, the highest level in current discussions, the cost of the put is 1% of GDP or equivalently a temporary increase in social security taxation of at most 20%. If only 90% of the benefits are guaranteed, then the put only costs 0.03% of GDP. Since all such puts generally have negative expected value, see Tompkins, Ziemba and Hodges (2003), this cost is expensive so it is one of the alternatives to consider. Other options include investment in the best hedge funds, mutual funds, private placements, levered buyouts, commodities stocks and alternative investments with superior records and other assets projected to have superior returns; see Ziemba (2005) and Ziemba and Ziemba (2007) for studies of such superior investors.

Not wanting to destroy a successful program to 'save' it, Diamond and Orszag (2004) present a plan for saving Social Security without diverting funds into individual accounts. Social security has been successful at providing benefits to workers and their families in the form of a real annuity after the disability, retirement, or death; providing higher annual benefits relative to earnings for those with lower earnings; and providing similar replacement rates to age cohorts. Their system provides a basic level of benefits for workers and their families that cannot be decimated by stock market crashes or inflation, and that lasts for the life of the beneficiary while eliminating the need for deficit financing. They increase contributions and lower benefits. Building in an ongoing assessment of increased life expectancy starting in 2012, they increase the payroll tax to match needs eliminating the terminal year effect and reducing the deficit to 0.55%. Next they address two aspects of growing inequality. One would increase the taxable earnings phased in over time until 87% of payrolls are covered. This would lower the deficit to about 0.3%. Secondly, as life expectancy improvements are greatest for those with higher education and earnings, this further distorts the payouts, so they recommend making the benefit formula more progressive. Finally they address the legacy debt (about $11.6 trillion) that accumulated because current and past cohorts did not contribute enough to cover their benefits; see Figure 7. This initial debt accumulation was a policy decision when the program was set up but continues to haunt the program and any reforms.

They propose financing the debt in three ways: through universal coverage under Social Security (state and local government employees not currently covered so do not bear the burden, this would reduce the actuarial deficit by another 0.19%); through a

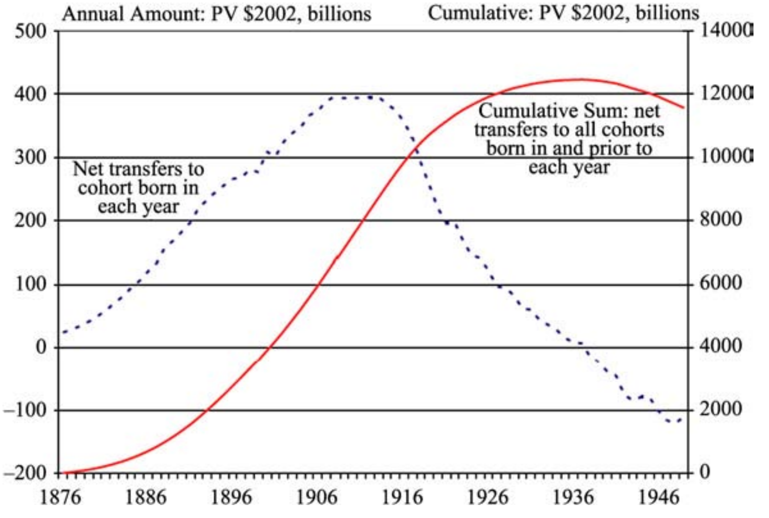


Fig. 7. The legacy debt. Source: Diamond and Orszag (2004).

legacy tax on earnings above the maximum taxable earnings base, with the tax rate beginning at 3% and gradually increasing over time; and through a universal legacy charge that would apply to workers and beneficiaries in the future. Another source could be a reformed estate tax.

The difference is that a move toward privatized accounts would take expected funds out of the SSTF while the Diamond–Orszag plan would pay off the legacy debt and put funds into the SSTF.

Diamond and Geanakopolos (2001) explore the general equilibrium impact of social security portfolio diversification into private securities, either through the trust fund or private accounts. Their analysis recognizes that there are differences in savings and investment among income groups as well as differences in the distribution of social security income and private pension and asset income. They are particularly interested in the relative impact on workers versus savers (low income earners who save little outside social security versus high income groups who save and invest) and the distributional shifts between young and old. They also explore the impact on investment and tax rates. Their conclusions depend heavily on certain assumptions about technology and it's relationship to investment. Social security is modeled as a combination of a PAYG system together with a defined contribution system simplifying the partially funded defined benefits system. Diversification is modeled by shifting the asset mix in the defined contribution system is from bonds toward equities. This would be like giving workers discretionary accounts and some of them would choose to invest part of their accounts in equities. Formally, they suppose there is a social security trust fund, which suddenly sells some of its bonds and shifts between defined benefit and defined contribution systems.

Diamond and Geanakopolos (2001) present a sequence of models to investigate different effects. Since diversification changes the level and riskiness of social security benefits this raises the expected utility of workers if they prefer some equity to an all-bond portfolio. But diversification has no effect on the utility of savers. Limited diversification increases interest rates, reduces the expected return on short-term investments (and the equity risk premium) by narrowing the expected returns between bonds and equities. Social security diversification likely changes the rate of interest, requiring higher income taxes. However, the increased income tax burden likely falls on households in different proportions than the social security taxes and benefits.

When there are infinitely-lived assets to introduce redistribution across generations they find that a rise in the riskless interest rate lowers the value of these infinitely-lived assets, hurting those savers holding these assets and benefiting young savers in the future.

They assume given fixed demands by savers for consumption when young, and for safe and risky consumption when old and that increases in government bond interest payments raise the payments on government bonds held by social security more than they raise workers' income taxes (and thus raise savers' taxes more than they raise savers' income on the government bonds they hold). Diamond and Geanakopolos also assume that the level of risky investment does not affect the relative outputs across states so that short-term risky production is on a ray in state space and that the output from both short- and long-term risky production is independently and identically distributed each period.

They find that social security diversification yields the potential for welfare improvements. Diversification away from zero exposure to equities raises the total weighted utility in the economy if household utilities are weighted so that the expected marginal utility of a dollar for sure is the same for all savers and retired workers. Social security diversification also likely causes welfare redistributions, so a Pareto improvement might require additional policies.

Those favoring social security diversification into equities argue that it favors young savers since equities have historically had higher returns than social security is projected to yield. However, they forget the unfunded liability embodied in social security commitments to the current elderly, and ignore the riskiness of stock returns. The Diamond and Geanakopolos model recognizes both of these factors, and shows that the equity premium falls with equity diversification and that the welfare of young savers depends on at least two more factors: (1) their income taxes must rise to pay the higher interest on government bonds; and (2) the assets they trade change in value. Young savers are net buyers of long-term assets, hence their welfare is lowered by an land price increases. Unless long-term capital values fall substantially after diversification, young (and future) savers are worse off with social security diversification. Meanwhile, the current elderly savers are made better off if long-term capital values rise.

Using a general equilibrium model with heterogeneous households and production, Diamond and Geanakopolos considered the implications of changing the social security trust fund portfolio policy away from 100% government debt. Social security diversification changes relative prices, and the composition and amount of real investment in turn raising total welfare and causing a redistribution among households. The assumed structure of marginal taxes along with social security diversification raises the interest rate, increasing the payoffs of the social security trust fund and helping workers, but necessitating an increase in income taxes hurting savers. There is usually a transfer of wealth between young savers and old savers, with the direction depending on the change in land values. Diversification reduces real safe investment, but increases risky investment. The effect on aggregate real investment and long-term asset prices is ambiguous. They show that the supposition that trust fund diversification would increase real investment and increase stock market value is questionable in both the risky linear case and the safe linear case. In both cases one or the other of the common sense predictions is reversed.

They make four points relative to the current policy debate:

1. the heterogeneity of the population implies that trust fund portfolio choice does have real effects on the economy;
2. it is unlikely that workers are so risk averse that a portfolio completely invested in Treasury bonds is optimal. This is reinforced by the ability of the government to spread risk over successive cohorts since social security is a defined benefit system. That is, if a defined benefit system is well-run, there is a stronger case for trust fund investment in private securities than in the models they analyzed which assumed a defined contribution system;
3. the marginal social benefit to diversification declines as the level of diversification increases (exposing workers to more risk), which puts a limit on the amount of socially desirable diversification;
4. the models substitute equity investment for bond investment, holding constant the level of funding. Many proposals for investment in equities, whether through the trust funds or through individual accounts, use equity investments as a reason to increase or decrease the financing of social security (at least in the short run) relative to what might be proposed without such investment. Such a change involves intergenerational redistribution that is not incorporated in their analysis, though their model could incorporate it. Their analysis applies to proposals that would substitute a portfolio change for cuts in future benefits.

The Social Security Network also analyzed plans for improving beyond the incremental items that closed 70% of the gap as we discussed earlier. These proposals include investing in the stock market and substantial privatization. Investing part of the Trust Fund in an unmanaged index fund holding a diversified portfolio of shares managed by an independent investment board would make the government a major stockholder in many American companies. Though the board is to be independent, there is a threat of political interference in corporate governance. Investing the Trust Fund in stock markets add risk to the system without doing anything to improve economic wellbeing. Privatization schemes generally allow individuals to choose their own level of risk-reward but does expose them to the risk of lower returns.

In creating private accounts there is a phase-in period in which contribution-benefits gap for the current retirees and those close to retirement must be funded out of higher taxes. The well-off could invest aggressively while lower- and middle-income retirees could choose more secure though lower-yield investments. Some individuals could obtain higher returns than they would under the current system. Individuals would own their own accounts, so they would face less political risk if future generations defaulted on promised benefits. National savings would likely increase if payroll taxes were raised to finance the transition to a privatized system. By shedding the enormous financial obligations of the government to future retirees, Social Security privatization, when fully implemented, would virtually eliminate the challenge of generating sufficient taxes to provide promised benefits. However, social security's success in keeping most of the elderly from poverty would be threatened. Workers who invest poorly or unwisely would not be assured of staying out of poverty. The system would become much less progressive. Low-income workers would have lower returns on their investments, and smaller retirement accounts and smaller incomes than high-income workers, because they tend to make poorer investment choices.

The shift to a privatized system would create huge transition costs. The current generation of workers would "pay twice", continuing to pay for the benefits of current retirees. Administrative costs of the Social Security program would increase, especially for individual investors, who would have to pay management fees to investment advisors which could be many times larger than the administrative costs for social security. Experience with 401(k)'s shows that individuals in the plans are typically charged retail rates rather than the lower rates charged to institutional investors (Hamilton, Kristof and Friedman, 2006). If many older Americans fell into poverty, either their families or other government programs would have to take up the slack. Many members of the baby-boom generation might find themselves supporting aging parents while trying to raise their own children.

Kotlikoff (2003) proposes to fix social security and medicare for good. The present value gap between receipts and expenditures is estimated to be $45 TR. Some extreme measures to pay this according to a Treasury Dept report are:

- raising federal income taxes, personal and corporate, by 60%;
- raising payroll taxes by 95%;
- cutting federal discretionary spending by 106%, more than eliminating it;
- cutting Social Security and Medicare benefits by 45%.

Kotlikoff proposes to replace the current social security with the Personal Security System (PSS) phasing out the existing system and topping up the trust fund to cover liabilities to existing participants. Firstly he would immediately end accrual into the retirement program, eliminating the payroll tax and replacing it with a retail sales tax of 12% going to zero to cover current liabilities. Contributions would then be made into the PSS with the government matching contributions. The government would contribute on behalf of the disabled and unemployed. The PSS balances would be invested in a single market-weighted global index fund and the real principle would be guaranteed. At retirement the PSS balances would be gradually sold to buy inflation protected annuities. Prior to retirement and the purchase of annuities, the PSS accounts could be bequeathed. Also the PSS would be shared 50–50 with spouses, eliminating one source of inequity.

European governments face a dual problem in dealing with the future of social security pensions. They, like other governments, must deal with the rising cost of pensions resulting from increasing life expectancy. This is complicated by the free movement of labor within the EU. Feldstein (2001) and Holzmann (2003) investigate this significant problem. They each develop a solution based on a notational defined contribution as the core combined with a fully funded second pillar. While Holzmann does not elaborate on the second pillar, Feldstein suggests individual investment accounts. Notational defined contribution plans are further discussed below.

Feldstein (2001)[2] predicts a need for a 50% increase in the payroll tax rate and even more for funding health costs. He simulates the probability distribution of returns relative to the social security benchmark of saving 6% of earnings during a working life and investing the savings in a 60–40 bond-stock portfolio. On retirement, the accumulated savings would purchase a variable annuity also invested in the 60–40 mix. The results are reported in Table 7 for retirees at age 67, 77, and 87. The median investment-based annuity at age 67 is more than twice the benchmark; even the savings rate is only one-third that which would be needed with the PAYG system. However there is almost a 20% chance of receiving less and an equal probability of receiving at least four times the benchmark benefit.

The risk of receiving about half the benefit is about 2% and there is about a 1% chance of receiving less than 40% of the benchmark. The risk of a relatively low benefit payout increases with individual, due to the increasing variance overtime. At age 77, there is a 1% chance that the investment based retirement benefit would be less than 56% of the benchmark benefit and a 1% chance that it would be less than 21%. At age 87, there is a 10% chance that it would be less than 40% of the benchmark and a 1% chance that it would be less than 12% of the benchmark.

Feldstein suggest adding guarantees or using a mixed-system. Using half the predicted PAYG tax, 9.5%, to finance half the benchmark would guarantee at least a 50% benefit level, putting a floor on the benefit. This could then be combined with a 3% saving and portfolio investment would cut the risk of less than half the bene-

[2] More details of this model are provided in Feldstein and Ranguelova (1998). They examined the risk aspects of a fully investment-based defined contribution Social Security plan. Individuals save a fraction of their wages in a Personal Retirement Account (PRA) invested in a 60:40 equity-debt mix and receive a similarly invested variable annuity from age 67. The value of the portfolio follows a random walk with the historical (1946–1995) mean log real return of 5.5% and standard deviation of 12.5% and administrative costs of 0.4%. For each of 10,000 simulations, they generate a mean real log return on the portfolio from a normal distribution with a mean of 0.055 and a standard deviation of 0.125 to generate a 71 year sequence of portfolio returns from 2000 to 2070.

Table 7
Probability distribution of investment-based retiree benefits as a multiple of the PAYG benchmark benefits, variable annuity purchased at age 66

Cumulative probability	Age 67	Age 77	Age 87
0.01	0.40	0.21	0.12
0.02	0.47	0.26	0.17
70.05	0.61	0.39	0.26
0.10	0.79	0.56	0.40
0.20	1.08	0.84	0.65
0.30	1.38	1.16	0.95
0.40	1.71	1.52	1.34
0.50	2.12	1.95	1.83
0.60	2.57	2.54	2.49
0.70	3.26	3.34	3.45
0.80	4.29	4.72	5.04
0.90	6.30	7.49	8.84

Source: Feldstein (2001).

fit to zero and the probability of less than 80% of the benchmark to 5%. This mixed system would eliminate the risk of very low retirement income and would provide an upside potential for less cost than the current projections. Another variant is discussed below.

Samuelson (1958) showed how, without any capital investment, individuals could earn an implicit rate of return on their social security tax payments. In steady state equilibrium, this implicit rate of return is equal to the real rate of growth of the social security tax base, i.e., the sum of the growth rates of the labor force and real wages. That is the rate of growth of total wages, which had averaged 2–2.5% but recently has been negative.

An investment-based program would eliminate the PAYG shortfall but would involve an initial reduction of consumption and a concurrent increase in the national capital stock. The rate of return in an investment-based plan is therefore the marginal product of capital which is greater than the rate of growth of aggregate wages. Poterba (1998), using revised national income and product account data, estimates that the marginal product of capital in the US non-financial corporate sector averaged 8.5% between 1959 and 1996. This relatively high rate of return on incremental capital makes it possible to finance benefits at a much lower cost in an investment-based program than in a PAYG program. It is the overall return on capital relative to the growth rate of wages that matters and not the return on equity investments relative to the return on government bonds (Feldstein and Ranguelova, 1998).

The potential long-run gains from switching to an investment-based system ignores a number of issues:

1. financing the transition, which would cost about 2% but within 20 years the cost would be less than the existing 12.4% with an assumed marginal return on capital of 9% (with 5.5% it would take 28 years);
2. the relative riskiness of defined benefit and defined contribution plans; and
3. the distributional consequences of shifting from a DB to DC plan.

While these studies show the feasibility of shifting to privatized accounts, there is no analysis of the impact of the shift of these funds to the market and the attendant increase in the government deficit. On the one hand, since the evidence is that index funds beat about 75% of the managers and it is hard to predict future currency moves, a well defined index of world wide assets might be wise. However, given the volume of assets required, it is not at all clear that they could all be invested in indices without inflating the value beyond the real economic worth. There is a paradox here that indices are only good investments on the margin. For example, if the volume in the S&P 500 index got into a level to take a substantial portion of the social funds, then its return distribution would be altered. It is clear that many places to invest are needed to absorb these funds and actually turn them into real capital investment and not just inflate asset values.

These calculations are for wealthier individuals and not for the low wage earner. Other studies suggest the need to rethink the entire social security issue and we will explore these now.

4. Rethinking and redesigning the social security system as part of a retirement package

There are several broad directions for redesigning social security systems to get both the benefits of market investment and some retirement income security. All break the link with a predetermined monetary guarantee at retirement. Some use various derivative securities to create benefits if the system were to rely more on private accounts as well as a variety of new ways to keep track of defined contributions without giving them market valuations so that the payout can be related to the economic output at the time of payout. This eliminates the possibility of not being able to meet liabilities as the liabilities are relative. This class of programs is known as non-financial defined contributions.

4.1. Feldstein's PRA with guarantees

If social security contributions are invested in the equity markets, retirement income is exposed to all the attendant market risks. Feldstein (2005c) presents a market-based approach to reducing the risk of investment based social security that could be tailored to individual risk preferences. He proposes a form of risk reduction which substitutes an investment-based personal retirement account (PRA) for the traditional pure PAYG plan to achieve both a significantly higher expected retirement income and a very high probability that the investment-based annuity would be at least as large as the PAYG benefit.

The guarantee is purchased each year based on that year's PRA savings. The basic contract would guarantee the individual a "No Lose" investment, that the amount saved in each year would be guaranteed to retain at least its real value by age 66. Such a guarantee could be provided by the firm that manages the PRA product (i.e., the mutual fund, bank, insurance company, etc.).

Feldstein suggests that the simplest way to achieve a No Lose PRA account would be to combine TIPS (Treasury Inflation Protected Securities, which have a guaranteed real return) with equities. The ratio of these would depend on the age of the saver and the rate of return on the TIPS for the relevant maturity. For example, if the saver is 21 years old and the real return on TIPS is 2%, a $1000 PRA saving would be divided between $410 in TIPS and the remaining $590 in equities. At older working ages, there are fewer years for the TIPS to accumulate and therefore a larger fraction of the initial saving must be invested in TIPS. A 40 year old would have to invest $598 out of each $1000 of new saving in TIPS.

When the individual reaches age 66, all of the annual PRA accounts would be combined to provide a single retirement fund and a conventional annuity would be purchased. The No Lose approach could be continued into this phase as well with the annuity provider offering a guarantee that the annual annuity payments would be at least as large as the individual's retirement fund could purchase with a zero real return. Other modifications could be made as desired.

To simulate the distribution of the "no lose" plan Feldstein used historical data on the S&P 500 from 1946 to 2003 and Lehman corporate bonds from 1973 to 2003. These have gross mean log real returns of 6.9 and 4.4%, respectively (with cost of 0.4% for administration). The mean individual starts work at 21 with a salary of $25,000, which rises 2% real/annum reaching $60,950 at 66 assuming benefits at 67 are 40% of the age 66 salary.

Figure 8 shows the results of replacing the 12.4% PAYG tax with 6.2% PAYG and 6.2% PRA. This would result in a median value retirement income of 147% of the traditional PAYG benefit and a 95% probability of exceeding the traditional benefit. There would be less than one chance in one hundred that the combined retirement income would be less than 96% of the traditional benefit.

Figure 8 shows that none of the initial plans clearly dominate. As no distribution dominates, he uses utility analysis and finds that the poor improvement at the bottom end is more than matched by the upside potential for the no guarantee plan is preferred. These are the first five distributions in Figure 9. Next he tailored the guarantees to individual preferences with no cost "collars", the final two distribution. This "collar" using puts and calls has significantly lowers the upside for a low improvement on the downside and a small shift in the mid range. The oval highlights where the different proposals begin to diverge (in Figures 8, 9 and 10).

Next Feldstein designed mixed plans to limit the tax increases of the current PAYG plan which some estimate will require and increase from the existing 12.4 to 18.6%. He experiments with a package that is 80% of current costs or 9.92% split between PAYG and PRA. The result is that with a reduction of the contribution (tax) and zero

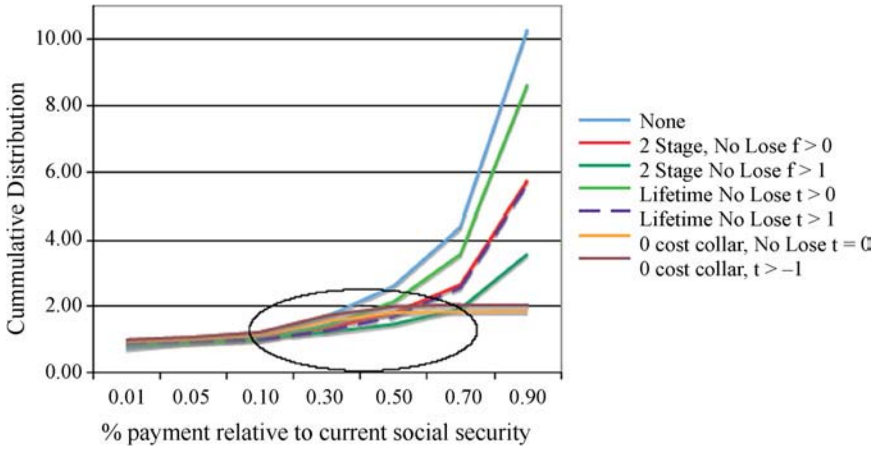


Fig. 8. Guarantee based on combination with TIPS compared with zero cost collar; combined pension at age 77, 50–50 PAYG and PRA; benchmark based on PAYG = 12.4% contribution. Source of data: Feldstein (2005c).

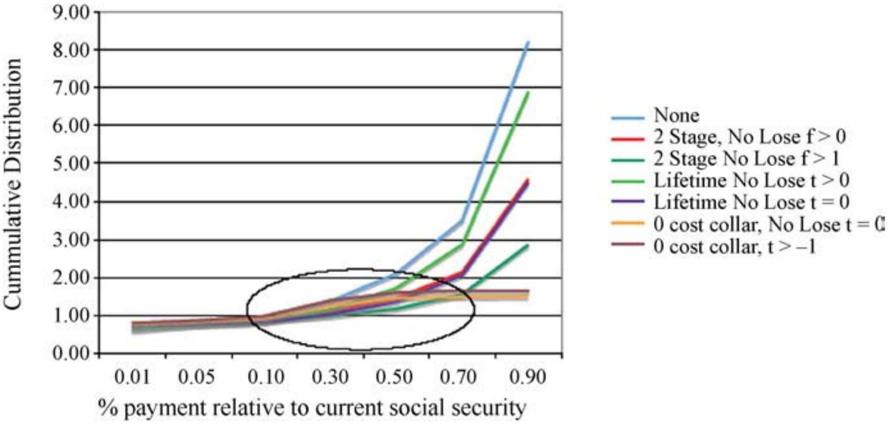


Fig. 9. Low cost mixed plans. Contribution rate is 9.92% versus 12.4% of current plan. Source of data: Feldstein (2005c).

cost collar, there is a high probability of increasing the returns above the current law, though there is more than a 10% chance that the returns will fall short. Individuals with constant relative risk aversions γ (CRRA) up to 4.0 would still prefer the no guarantee plan; is shown in Figure 9.

In the final simulations all the funds go to the PRA. Under expected utility those with $\gamma \leqslant 2.5$, prefer the pure equity investment with no guarantee; those with $3 \leqslant \gamma \leqslant 4$, prefer the zero cost collar with the guaranteed real return of at least -1%, see Figure 10.

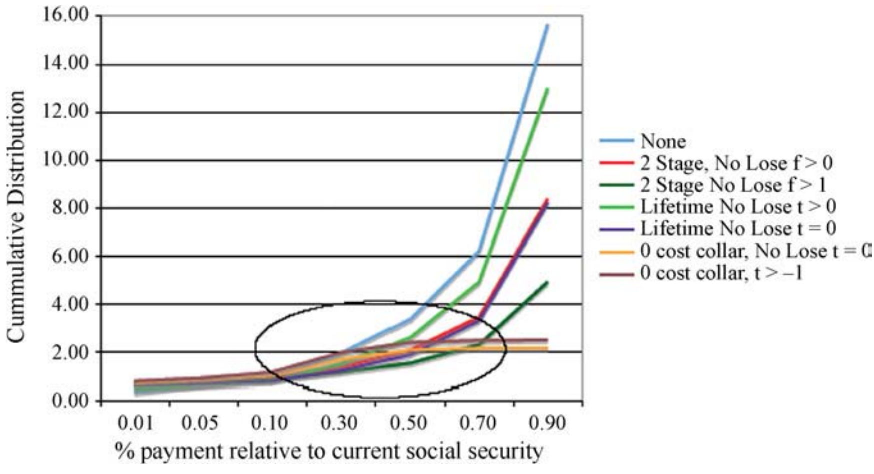


Fig. 10. Low cost (contribution rate = 9.92%) pure investment plans. Source of data: Feldstein (2005c).

Those with $4.5 \leqslant \gamma \leqslant 5$, prefer the two-stage guarantee of a real return greater than zero based on investment in TIPS.

Feldstein mentions two other utility calculations to highlight the use of mixed stock-bond plans. When the plan is mixed between PAYGO and PRA (no guarantees) individuals prefer 100% equity versus a 60:40 split. However, when the plan is pure PRA, the 100% equity plan is preferred only by individuals with low risk aversion ($\gamma \leqslant 3.0$).

Feldstein's study demonstrates how expensive the guarantees are. If the society truly wanted to guarantee some minimal level of retirement security, then a grand societal type "self-insurance" like the PAYGO system might well be the most cost effective. Also it is not at all clear that these guarantees could be purchased without even higher costs for small PRAs. In addition the increased demand for insurance would likely greatly increase the costs. In the end, those who most need the *security* would be least likely to benefit.

4.2. NDC—notational or non-financial defined contributions

One broad method to 'save' social security systems is to break the link between financial contributions and retirement payments. Notational or non-financial defined contribution (NDC) systems as opposed to financial defined contributions (FDC) that we are familiar with do this. NDCs are designed to eliminate the monetary risk from the liability of PAYG systems. Palmer (2003a) defines the major attributes of a NDC system.

Over a working life individuals accumulate notational credits that are converted to monetary payments at retirement based on the contributions working cohort not the accumulated financial contributions over the working life. Values continue to accumulate as long as the participant works and makes contributions. There is no "full-benefit" age, but instead all new contributions add to the size of the individual's annuity when claimed. There is a *minimum* pension age but not a mandatory one. When cashed in the annuity is calculated by dividing the account balance with an estimate of life expectancy at retirement.

NDCs address the fiscal, political, social, and economic needs of reform while keeping the fiscal burden of low. Börsch-Supan (2003) provides a framework for evaluating NDC systems. Holzmann and Palmer (2003a, 2003b) provides a review of recent experiences with these systems. Sweden (1994), Italy (1995), Latvia and Poland were early adapters. NDC promises to address the issues confronting pension policy in the twenty-first century.

NDC is designed to balance liabilities and assets thus providing stability. An NDC scheme is a *defined contribution*, pay-as-you-go (PAYG) pension scheme. Contributions are defined in terms of a fixed contribution rate on individual earnings that are *noted* on an individual account however they are not "funded". Diamond (2003a, 2003b) succinctly highlights the differences between an NDC and a DB plan by observing that an NDC makes adjustments on the benefit side given the contributions while a DB makes adjustment on the contribution side given the promised benefits. By definition, the payments reflect life expectancy improvements, productivity (and real wage) improvements and changing demographics. The internal rate of return reflects these factors rather than a financial market rate of return. It does this by treating all in the same cohort equally but allowing for differential treatment across cohorts. So when the total contributions fall because of a smaller working population (increased dependency ratio), the financial payout is lower. Thus, the NDC establishes a limit for the public pension commitment and thus encourages individual saving for retirement. Given the valuation of accumulated credits upon retirement, there is much scope for flexibility. If the demographics or productivity is unfavorable, the actual monetization of the NDC is lower.

An NDC provides account information to participants as well as a system-wide financial statement which can be made public. In practice the transition from the previous plan to an NDC may place some restrictions on the implementation of generic NDC.

The NDC benefit is a life annuity. It can be claimed at any time after the minimum retirement age. The annuities issued at each point in time reflect life expectancy, so in principle, an NDC is an actuarially fair pension system. It also distributes *individual* resources over the life cycle, as part of a countrywide (universal) PAYG insurance scheme. It is like to an illiquid, individual cash balance scheme, but the annuity aspect is like an *insurance* plan by redistributing the capital to the survivors. It fulfills the function of insuring against the individual risk of outliving the average participant. The NDC benefit is based on individual contributions from earnings, so it may not provide sufficient resources in old age. An NDC scheme must be supplemented with some form of low-income support, in the form of minimum income or minimum benefit guarantee. NDC accounts can also be supplemented with noncontributory rights, e.g., for childcare years. What is essential is that these be financed with revenues exogenous to the NDC scheme, i.e., general tax revenue.

How is NDC different from a funded DC plan? Most PAYG systems have some reserve fund: in Sweden, the reserve fund is about five years of expenditures, while the German system has only about a few days reserves. A crucial question is whether the accumulated balances are collateralized and which claims represent the collateral. NDC systems are "notional" in the sense that there is no collateral at all. Balances are claims against future taxpayers, and they are not backed by a financial instrument. Funded DC plans are usually understood as being collateralized against physical capital, mostly through financial instruments such as commercial bonds or stocks. We use the word "funded" only for these plans. Some authors also call those DC plans that are collateralized with government bonds "funded". We think that this is a misuse of the word "funded". Although benefit claims of such plans are marketable and yield a well-defined rate of interest, they do not represent claims on physical capital. NDC systems may permit a "natural" way to make the implicit debt of a PAYG system explicit by linking the NDC balances to government bonds, and the resulting insights about future benefits and contributions may create saving incentives on the microeconomic level. However, NDC systems and conventional DB systems share the crucial macroeconomic features of PAYG systems: NDC systems do not accumulate savings in real assets with the potential beneficial side effects on the national saving rate, capital market development, and growth. NDC systems are therefore no substitute for prefunding.

4.2.1. Case studies

4.2.1.1. Comparison of French and German point systems. Legros (2003) focuses on the differences and similarities between NDC and the French and German point systems particularly in how they respond to shocks such as demographics and productivity.

The French system is for employees in private companies and covers about 60% of workers. It is composed of both a minimum pension and a supplementary pension. It has no automatic indexing but can be adjusted each year according to the forecasts of changes in the economic and demographic environment. This however requires reliable and frequent forecasts, and a governing board independent of the retiree and worker lobbies willing to enact the changes. During retirement the two components of the pension are revalued on different indices. The basic pension is adjusted for inflation while the supplementary pension adjusts to the changes in the points value. Also during one's working life, the earned points are adjusted to the price of the points such that if the price of the points increases faster than wages, the contribution earns fewer points. As this scheme was a consolidation of numerous individual sectoral schemes, these parameters vary by sector.

The German scheme has adopted an automatic method to correct the plan by reducing the pension yield in relation to a desired contribution rate and explicitly accounting for changes in the dependency ratio and life expectancy. This makes it likely that the German scheme will move into surplus. It is an example of a DB system that mimics an NDC system. There are three main elements: a point system of credits, actuarial adjustments, and a benefit indexation formula. Recent pension reforms have changed these elements in the direction of NDC-type pensions. The *point system* has been a feature of the German pension system since its conversion to a PAYG system in 1957. *Actuarial adjustments* were legislated by the 1992 reform; they have been phased in since 1997, with the bulk of adjustments in force by 2007. As a result, effective retirement ages are expected to increase by about two years within ten years. In 2001, the reform took the first step from a purely pay-as-you-go to a capital-funded pension system. It established upper limits to the contribution rate, discontinued the benefits indexation formula, and substantially lowered pension levels. At the same time, state-subsidized supplementary private pensions were introduced to fill the pension gap. However, these reforms were not enough to stabilize public pension funding. In the late fall of 2002, the government established a reform commission to achieve "sustainability in financing the social insurance systems". This commission decided that the contribution rate must be adhered to so to balance accounts they changed the *benefit indexation formula*. This in effect turned the German DB system into an NDC-type pension system.

The German public pension system computes benefits according to a formula that has an individual component determined by each person's earnings history and retirement age, which stays fixed for the entire retirement period, and an aggregate component which adjusts benefits over time equally for all pensioners. A typical worker who works for 40 years and who earns the average labor income in each of these 40 years receives 40 earnings points. If this worker retires at age 65, no actuarial adjustments take place ($AA = 1$). In the second half of 2002, the typical worker received a pension of 1,034.40 euros per month. A worker who has worked for 20 years at average earnings, or a worker who has worked for 40 years at 50% of average earnings, will receive half these pension benefits, while workers who earn twice the average labor income for 40 years will receive twice as much as the 40-year average earner. Age 65 is the "pivotal age" below which benefits will be reduced by 3.6% (maximum 10.8%) for each year of early retirement and above which for each year of postponement, benefits increase by 6%. Even at this rate it is considered that there are still negative to retire early.

Adding a sustainability factor links annual increases in pensions to productivity growth and the growth of the contribution base. The weighting factor α spreads the financial burden among contributors and pensioners. If α equals zero, the current benefit indexation formula would remain unchanged and the financial burden generated by a higher proportion of pensioners in the population would mainly be shouldered by the labor force. The condition α equals one is a contribution-oriented pension expenditure policy. The German Reform Commission set α to $1/4$ to target the contribution rates anchored in the Riester reform (contribution rate below 20% until 2020, and under 22% until 2030). Benefits are based on accounts rather than entitlements and making the transition easier.

4.2.1.2. The Italian and Swedish reforms. Pension reform has been at the center of the effort to consolidate Italian public finances. First steps were taken in 1992 when 25% of prospective public sector pension liabilities was abruptly canceled. Then in 1995 NDC was introduced for new entrants and on a pro rata basis for those with less than 18 years work. Franco and Sartor (2003) provide and discuss empirical evidence on the results of this reform. The 1995 reform was implemented without a thorough analysis and an open public debate. This led to resistance and later reform has progressed in an incremental way, generating uncertainty and perhaps limiting the benefits of the NDC approach. They conclude that the future of NDC in Italy is uncertain. They suggest the need to: (1) accelerate the spread of notional accounts; (2) avoid increases in the ratio of pension expenditure to GDP; and (3) strengthen the self-equilibrating mechanisms.

Gronchi and Nisticò (2003) use a theoretical framework to highlight the implications of the different implementations of NDC in Italy and Sweden. In both there is a real growth norm factored into the life annuities that has not been implemented in other plans. In Italy's case, there is a real growth factor of 1.5% per annum over the life of the annuity and in Sweden it is 1.6%. The main conceptual defect of the "Italian-style" NDC scheme is that the indexation of pensions should reflect the difference between the sustainable rate of return (with GDP growth as a proxy) and the value of 1.5% used in calculating annuities, which is the case in Sweden. The absence of this correction jeopardizes the objectives of sustainability and fairness. Other shortcomings and inconsistencies of the Italian reform are discussed. The authors conclude that, although blueprints have been available since 1995, successive governments have not taken the necessary step to form a committee of experts with a mandate to "bridge the gap between the theoretical foundations of the NDC scheme and the countless details that implementation would inevitably bring".

4.2.1.3. Lessons. An NDC system does not change the basic PAYG mechanism in which the children pay for the pensions of their parents, and it does not create savings unless it generates a benefit cut, which in turn precipitates savings. However, if correctly designed, an NDC system will automatically respond to changes in the demographic and macroeconomic environment because benefits are indexed to longevity (due to the annuitization mechanism), fertility, and employment (through the notional rate of interest, if indexed to the contribution bill). Moreover, an NDC system has potentially important microeconomic effects. It will create a sense of *actuarial fairness* (because annual benefits are in line with lifetime contributions) and *actuarial neutrality* (because the system creates automatic adjustments to retirement age). It exposes redistribution because any noncontributory credits appear clearly marked on the account statements. An NDC system changes the *rhetoric of pension systems*. It makes people think in terms of accounts rather than entitlements and thus may make the transition to partial funding psychologically easier. Moreover, by exposing the dwindling balance of first pillar pensions, it may actually create incentives to save in the second and third pillar. An NDC system makes workers and administrators think in terms of "pension wealth", which may ease portability both within a country and between countries. It enables interpersonal transfers (for example, between husband and spouse) and eases replacement of survivor pensions by independent pension claims.

An NDC system takes some issues out of the political agenda. It minimizes the role of the "normal retirement age" and permits a more flexible choice between working longer and getting a lower replacement rate. The new format helps pension reform because it provides a framework to introduce actuarial adjustments (since they come "automatically"), a framework to diffuse the explosiveness of changes in the retirement age (since a flexible choice of retirement age minimizes opposition), and a framework to change intergenerational redistribution.

In sum: NDC plans deal well with *longevity* and handle slow *changes in fertility*. However sudden changes, like the baby boom/baby bust transition, still require prefunding as the older generation will not be able to pass the entire burden onto the next. Design flaws of current DB systems (such as labor supply disincentives) are relatively manageable in NDC systems. NDC are still PAYG and do not change the basic macroeconomic characteristics such as growth, savings, or improvements in capital markets. To change the growth path of an economy, these systems should be supported by an improvement of second and third pillar pensions.

As with any reform, the real key is in the implementation and this must be handled well politically—there needs to be effective communication with the stakeholders and a high level of trained administrators. The strength of NDC is also a weakness if through a large economic shock it must respond dramatically with lowered payouts. The benefits can only accrue after the tax overhang of current DB systems are dealt with. Finally, NDC must be seen as part of a multi-pillar pension system. Some attention should be paid to how the entire system functions including occupational groups that are hard to handle: farmers or persons involved in the informal or semiformal agricultural setting in developing economies. Educating people to respond to the concept that benefits depend only on one's own contributions, with life expectancy determining the size of the life benefit, overall, that for any given contribution rate and retirement age the benefit level will be lower—perhaps markedly lower—than under the old NDB scheme. Individuals can adjust by choosing a later retirement age and increasing savings. It is important that individual choices do not undermine the system by creating more people who fall below the poverty level in old age.

4.2.2. Conversion to NDC (Palmer, 2003b)

In the 1990s, Italy, the Kyrgyz Republic, Poland, and Sweden began to gradually transition from mandatory PAYG-DB to NDC schemes. In 1996, Latvia was the first to make a total conversion for all workers. Others countries are considering implementing NDC. Using the experiences of the original five NDC countries, Palmer (2003b) discusses issues involved in the transition to NDC plans. Table 8 highlights some the transition models for initial capital during the introduction to NDC plans for various countries and Table 9 reviews other issues in the transition. It is important to treat previous contributions fairly. One problem is that of any overhang from prior promises including disability insurance (which needs to be dealt with by taxes in any event). This is one of the problems of the existing PAYG system that should be dealt with by a tax transfer so the new system gets off to a good start.

Table 8
Transition models for initial capital and introduction of NDC

Country	Transition model
Italy 1995	No initial capital. NDC for those with fewer than 18 years under old system as of 1995. Pro rata benefit for those with fewer than 18 years (combine old and new) New entrants covered only by NDC.
Kyrgyz Republic 1996	No initial capital. Transition benefit to full NDC, based on service years (max 30 prior to 1996) and best yearly earnings in 1991–1995 + NDC (based on earnings from 1996) + universal flat rate (12% of average wage). Earnings 1996 and after are covered for the *entire population*.
Latvia 1997	Initial capital based on service years with various adjustments depending on year of retirement based on average wage of covered population. They are still determining cutoff year for the initial capital. Initial capital is calculated for *all workers* claiming a benefit in 1996 or after. NDC capital is calculated for women born 1940 and later and men born 1936 and later (given the pension age in 1996 of 56 for women and 60 for men).
Poland 1999	Initial capital = present capital value of acquired rights from the old system calculated as of December 31, 1998, the day before implementation of the new scheme on January 1, 1999. New rights are based on a contribution rate of 12.22%. NDC applies to all employees and nonagricultural self-employed workers born 1949 and after.
Sweden	Initial capital based on individual earnings histories from 1960, for everyone born 1938 and after. Contribution rate of 18.5% through 1994; 16.5% for 1995–1997; and 16% beginning 1998. Average covered earnings 18.5% for earnings through 1995. Persons born 1938: 4/20 of benefit with new and 16/20 with old rules, etc. Persons born 1954 + later, full NDC.

Source: Adapted from Palmer (2003b).

4.2.2.1. Coordinated Europe-wide system. An issue confronting further integration of European labor and financial markets is pension schemes based on the individual nations. Holzmann (2003, p. 225) suggests a scheme with an "(NDC) system at its core and coordinated supplementary funded pensions and social pensions at its wings". This would balance harmonization and the preferences of individual countries for contributions, benefits and eligibility. A big challenge would be the transition to such a system. We saw that transitions are difficult for an individual country but for this harmonization to work NDC and quasi NDC countries would need to be harmonized with unfunded and semifunded traditional systems all with differing contributions, benefits, eligibility and debt overhang from the past. As well, the various social assistance programs would have to be clearly separated from the pension program. Perhaps before this is possible, more countries would need to transition to an NDC-type scheme and sort out past problems. it is interesting to consider this future challenge.

Table 9
Other issues in the transition to NDC

Country	Occurrence of "unfinanced" special rights	Rate of return on notional capital	Rate of return on annuities	Reserve fund
Italy	The contribution rate for employees is 32%, but 33% is used to calculate account values; for the self-employed, the contribution rate is 15%, but 20% is used to calculate account values.	GDP index	GDP index	No
Kyrgyz Republic	Special rights from the old regime were reduced by about 70% in legislation from 1997.	75% of wage growth	Indexation by ad hoc political decision	No
Latvia	Accounts values for persons with special rights under the old regime are adjusted according to specific formulas, and all special eights are phased out under the new NDC law. Beginning in 2003, "special rights increments" are financed from general state budget revenues.	Covered wage sum	Inflation adjustment 1996–2002, and inflation adjustment plus 25% of the real wage sum from 2002, and 50% real wage sum from 2011	Not at time of implementation, but reserves are projected to turn positive from ∼2005
Poland	Farmers are excluded from the NDC scheme, but are covered by a separate scheme, approximately 95% subsidized by general tax revenues. Bridging pensions were created for persons with special rights, and are financed with state budget revenues.	75% of the covered wage sum	Inflation adjustment plus at least 20% of real wage growth, the latter by ad hoc political decision	A contribution rate was earmarked to build up reserves, but general budget deficits have thus far prohibited this in practice
Sweden	No special rights in either the old or new NDC scheme.	Covered contributions per participant + automatic balancing	Covered contributions per participant + automatic balancing	Large reserve fund from the outset

Source: Palmer (2003b).

4.2.2.2. Review of advantages and disadvantages (from Holzmann and Palmer, 2003a, 2003b).

Advantages

- Adapts automatically to the changed balance of contributors to pensioners (*baby boom/baby bust problem*) discretionary intervention.

- Adapts automatically to changed life expectancies (*longevity problem*) when the notional pension wealth is converted into a lifelong pension.
- Avoids *arbitrariness* of benefit indexation rules, adjustment factors, etc.
- Adds *transparency* to the PAYG by identifying individual contributions and the benefit claims.
- Strengthens the principle that pensions are based on *lifelong earnings*.
- Separates *transfers* so they are easily identified, e.g., tax-financed credits for higher and vocational education and similar credits for educating children.
- Produces a suitable framework for *independent pensions of spouses*.
- Creates a *homogeneous paradigm* for the other forms of retirement savings.
- Permits a *flexibility* for employees in choosing their retirement age; makes the inflexible and politically problematic fixation of a "normal" retirement age superfluous; and exposes the trade-off between accumulated contributions and retirement age in an internally consistent fashion.
- Permits easy *portability* of pension rights between jobs, occupations, and sectors.

Disadvantages

- If the contribution rate is fixed, the replacement rate becomes uncertain as it depends on the future earnings and output.
- The system does not change the tension between declines in business cycle-related earnings and long-term spending commitments. So a genuine liquidity reserve is necessary.
- If the annuity is frozen at the beginning of retirement, a stabilizing feedback mechanism is missing if there is an unexpected rise in life expectancy.
- Discretionary decisions take place at the choice of life table, computation rules (such as the averaging) for the internal rate of return, the determination of a minimum retirement age, and so on.
- The system does not change the fact that only prefunding can change which generation pays for a given pension benefit. If one wants to have the workers of generation X at least partially pay for their own pension, rather than their children in generation $X + 1$, some extent of prefunding is necessary. An NDC system is no replacement for such partial funding. It only optimizes of the PAYG system.

4.3. The PAAW (personal annuitized average wage security), a variant of the NDC

Geanakopolos and Zeldes (2005) argue that what separates Democrats and Republicans is a different focus, one on the insurance aspects and one on the individual security aspects. Democrats want to retain the insurance aspects of transferring income to lower earners while at the same time sharing the risks across generations. Republicans want to guarantee individual ownership of the assets and be able to value the assets to improve individual retirement planning. The system designed by Geanakopolos and Zeldes is based on personal accounts (progressive personal accounts or PPA) and a derivative security based on the average wage (personal annuitized average wage security of PAAW). The PPA can be defined in a way to redistribute lifetime retirement credits through a variable matching of contributions by the government—the higher the income, the lower the matching. This is done on an annual basis and depends on total lifetime contributions. A PAAW is a derivative that securitizes the contingent liability of the government. Each PAAW pays one inflation corrected dollar for every year of life after a fixed date t (retirement age) multiplied by the economy wide average wage at t. Individuals would get PAAWs in exchange for social security contributions and these would be held in the PPAs.

The PPA ensures intra-generational equity? The PAAW creates risk sharing across generations. If young workers are doing well and receiving high wages, the old will get higher payoffs from their PAAWs, and conversely. Such a system maintains the core of the current system, but would increase transparency and enhance property rights as the accounts would be personal, and lower the political risk of legislation removing benefits.

Initially the PPAs could be required to hold all their wealth in PAAWs, without any opportunity to trade them in financial markets. If done this way, it is similar to the NDC. If the PAAW is tradeable, then a value true not notational value will be attached and individuals would be better able to plan for retirement assets outside social security knowing what they would get. To prevent people from selling off all their PAAWs, they propose limiting to 10% which they claim would be enough to get a fair value without destroying the risk sharing aspects of the plan.

4.3.1. PAAWs vs. notional accounts

Typically the money in notional accounts is legislated to grow at the rate of the growth of wages. By the year of retirement the money in the account is proportional to the wages of the next generation's workers. PAAWs are real securities and can be traded so their market prices would convey useful information.

They go on to discuss incorporating risk aversion and other aspects.

A number of papers have proposed the creation of related new financial securities. Geanakopolos and Zeldes suggest that this follows a number of papers which have proposed the creation of related financial securities; for example, Shiller (1993) proposes GDP-linked securities, Blake and Burrows (2001) propose longevity or survivor bonds, and Bohn (2002) and Goetzmann (2005) propose aggregate wage-related securities.

5. Conclusions

The social security debate is an important arena in which market economics and social welfare vie. We must reflect back to the original insurance purpose of social security provision—that when people were too old, infirm or disabled to work, there would be a safety net to protect them, funded by a general tax on earnings. This evolved into an expectation of an active retirement made possible by benefits earned from contributions to a social account. Looking at the data it is clear that for the most part it still is the poorer quintile of elderly that rely solely on social security while other segments use social security as a supplement to other sources of income when they are no longer working. This aspect of the social security dilemma must be emphasized in any attempt to make the system sustainable.

Of course retirement income is complicated by the growing costs of health care in the final years of life and in all cases the overall increase in the dependency ratio—not narrowly seen as the social security dependency but overall as financial resources are turned into real demand for currently produced goods and services. So, while social security crisis might be solved by manipulating contributions, benefits and retirement age, the broader issues must be resolved through a new understanding of the intergenerational social contract. Furthermore, in the end, there is a fallacy of composition. Goods and services must be produced sufficient to meet the needs of all participants in society—the employed as well as the retired.

References

Abel, A.B., 2001. The effects of investing social security funds in the stock market when fixed costs prevent some households from holding stocks. American Economic Review 91, 128–148.

Abel, A.B., 2003. The effects of a baby boom on stock prices and capital accumulation in the presence of social security. Econometrica 71, 551–578.

Anrig Jr., G., Wasow, B., 2005. Twelve Reasons why Privatizing Social Security is a Bad Idea. The Century Foundation, New York and Washington, DC.

Apfel, K.S., Graetz, M.J., 2005. Uncharted Waters: Paying Benefits from Individual Accounts in Federal Retirement Policy. National Academy of Social Insurance.

Bernartzi, S., Thaler, R.H., 2001. Naive diversification in defined contribution savings plans. American Economic Review 91 (1), 79–98.

Biggs, 2002. How recent stock market declines affect the social security reform debate, CATO briefing paper, September 10.

Blake, D., Burrows, W., 2001. Survivor bonds: Helping to hedge mortality risk. Journal of Risk and Insurance 68 (2), 339–348, June.

Bohn, H., 2002. Retirement savings in an aging society: A case for innovative government debt management. In: Auerbach, A., Herrmann, H. (Eds.), Ageing, Financial Markets and Monetary Policy. Springer.

Börsch-Supan, A.H., 2003. What are NDC systems? What do they bring to reform strategies? In: Holzmann, R., Palmer, E. (Eds.), 2003a.

Börsch-Supan, A.H., Wilke, C.B., 2003. The German Public Pension System: How it will become an NDC system look-alike. In: Holzmann, R., Palmer, E. (Eds.), 2003a.

Burtless, G., 1999. Risk and returns of stock market investments held in individual retirement accounts. Testimony before Task Force on Social Security Reform, House Budget Committee, May 11. Available online at http://www.brookings.edu/views/testimony/burtless/19990511.htm.

Butrica, B., Goldwyn, J.H., Johnson, R.W., 2005. Understanding expenditure patterns in retirement CRR WP 2005-03.

Consigli, G., 2007. ALM for Individual Investors. In: Zenios, S.A., Ziemba, W.T. (Eds.), Handbook of Asset and Liability Management, vol. 2. Elsevier, pp. 751–827 (Ch. 17).

Constantinides, G.M., Donaldson, J.B., Mehra, R., 2005. Junior must pay: pricing the implicit put in privatizing Social Security. Annals of Finance 1, 1–34.

Diamond, P.A., 2003a. Social security rules that vary with age, Working paper. MIT, December 2.

Diamond, P.A., 2003b. Conceptualization of non-financial defined contribution systems, Working paper. MIT.

Diamond, P.A., Geanakopolos, J., 2001. Social security investment in equities. American Economic Review 93 (4), 1047–1074.
Diamond, P.A., Orszag, P.R., 2004. A summary of: Saving social security: A balanced approach, Working paper. MIT.
Dimson, E., Marsh, P.R., Staunton, M., 2006. Global Investment Returns Yearbook, ABN Ambro. London Business School.
Feldstein, M., 1997. Transition to a fully funded pension system: five economic issues, NBER 6149.
Feldstein, M., 2001. The future of social security pensions in Europe, NBER 8487.
Feldstein, M., 2005a. Structural reform of social security, NBER 11098, February.
Feldstein, M., 2005b. Rethinking social insurance, NBER 11250, March.
Feldstein, M., 2005c. Reducing the risk of investment-based social security, NBER 11084, January.
Feldstein, M., Ranguelova, E., 1998. Individual risk and intergenerational risk sharing in an investment-based social security system, NBER 6839, December.
Feldstein, M., Samwick, A., 2001. Potential paths of social security reform, NBER 8592, November.
Franco, D., Sartor, N., 2003. NDCs in Italy: Unsatisfactory present, uncertain future. In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Geanakopolos, J., Zeldes, S.P., 2005. Reforming social security with progressive personal accounts, NBER 05-07.
Geanakopolos, J., Mitchell, O.S., Zeldes, S.P., 1998a. Would a privatized social security system really pay a higher rate of return? NBER WP No. 6713, August.
Geanakopolos, J., Mitchell, O.S., Zeldes, S.P., 1998b. Social security money's worth, NBER WP No. 6722, September.
Goetzmann, W.N., 2005. More social security, not less, Yale ICF Working Paper No. 05-05.
Goolsbee, A., 2004. The fees of private accounts and the impact of social security privatization on investment managers, Working paper. University of Chicago Graduate School of Business, September.
Gronchi, S., Nisticò, S., 2003. Implementing the NDC theoretical model: A comparison of Italy and Sweden. In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Hamilton, W., Kristof, K.M., Friedman, J., 2006. Fees eat away at employees' 401(k) nest eggs, LA Times, April 23.
Holzmann, R., 2003. Toward a coordinated pension system in Europe: Rationale and potential structure.
Holzmann, R., Palmer, E. (Eds.), 2003a. Pension reform: Issues and prospects for non-financial defined contribution (NDC) schemes. In: Proceedings of the NDC Conference in Sandhamn, Sweden, September 28–30.
Holzmann, R., Palmer, E., 2003b. The status of the NDC discussion: Introduction and overview. In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Kotlikoff, L.J., 2003. Fixing social security and medicare for good, WP Boston College, Paper presented at the University of Maryland.
Kotlikoff, L.J., Burns, S., 2004. The Coming Generational Storm. MIT Press.
Kotlikoff, L.J., Smetters, K., Walliser, J., 2004. Diminishing America's demographic dilemma through prefunding social security, NBER 04-075.
Krugman, P., Tanner, M., Marshall, J.M., 2005. Social security: Is it really a crisis? Talkingpointsmemo.com, New York Society for Ethical Culture, New York City.
Legros, F., 2003. NDCs: A comparison of the French and German point systems. In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Leland, J., Wilgoren, J., 2005. Living with social security: Small dreams and safety nets, New York Times, June 19.
Mitchell, O.S., Zeldes, S.P., 1996. Social Security privatization: a structure for analysis, NBER 5512.
Munnell, A.H., 2003. The declining role of social security. Just the facts on retirement issues, 6. Center for Retirement Research at Boston College.
Nataraj, S., Shoven, J.B., 2004. Has the unified budget undermined the federal government trust funds? NBER.

OECD, 2005. Ageing and pension reform systems. Financial Market Trends (Suppl.) 1.
Orszag, P., 1999. Individual accounts and social security: Does social security really provide a lower rate of return? Center on Budget and Policy Priorities, p. 1, March 9.
Oshio, T., 2004. Social security and trust fund management, NBER 10444, April.
Palmer, E., 2003a. What is NDC? In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Palmer, E., 2003b. Conversion to NDCs—issues and models. In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Panel on the Privatization of Social Security, 1998. Evaluation issues in privatizing social security, National Academy of Social Insurance.
Poterba, J.M., 1998. The rate of return to corporate capital and factor shares: New estimates using revised national income accounts and capital stock. Carnegie–Rochester Conference Series on Public Policy 48, 211–246.
Poterba, J., Rauh, J., Venti, S., Wise, D., 2006. Defined contribution plans, defined benefit plans, and the accumulation of retirement wealth, NBER working paper, September.
Reno, V., 2005. Payouts in individual accounts pose new questions, TIAA-CREF Policy Brief.
Rich, M., Porter, E., 2006. Increasingly, the home is paying for retirement, New York Times, February 24.
Samuelson, P.A., 1958. An exact consumption loan model of interest with or without the social contrivance of money. Journal of Political Economy 66, 467–482.
Shiller, R.J., 1993. Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks. Oxford University Press.
Shoven, J.B., Slavov, S.N., 2005. Political risk versus market risk in social security, NBER N05-02, July.
Siegel, J., 2002. Stocks for the Long Run. Wiley.
The Economist, 2006. Turning boomers into boomerangs, February 16.
Tompkins, R.G., Ziemba, W.T., Hodges, S.H., 2003. The favorite-longshot bias in S&P 500 futures options: the return to bets and the cost of insurance, Working paper. Sauder School of Business, UBC.
Toner, R., Rosenbaum, D.E., 2005. In overhaul of social security, age is the elephant in the room. New York Times, June 12.
Valdés-Prieto, S., 2003. A market method to endow NDC systems with automatic financial stability. In: Holzmann, R., Palmer, E. (Eds.), 2003a.
Valdés-Prieto, S., 2005. Market-based social security as a better means of risk-sharing, Salvador PRC WP 2005-16.
Ziemba, W.T., 2003. The Stochastic Programming Approach to Asset–Liability and Wealth Management. AIMR, Charlottesville, VA (text and appendix).
Ziemba, W.T., 2005. The symmetric downside risk Sharpe ratio and the evaluation of great investors and speculators. Journal of Portfolio Management, 108–122, Fall.
Ziemba, R.E.S., Ziemba, W.T., 2007. Scenarios for Risk Management and Global Investment Strategies. Wiley.

This page intentionally left blank

AUTHOR INDEX

n indicates citation in footnote.

SUBJECT INDEX

This page intentionally left blank