The Dynamics and Control of Pension Funding

Muhammad Iqbal Owadally

Thesis submitted for the degree of
Doctor of Philosophy

Department of Actuarial Science & Statistics
School of Mathematics
The City University
London, England

November 1998

Contents

List of Tables

List of Figures

Acknowledgements

I would like to acknowledge the help of my supervisor, Prof. Steven Haberman, Dean of the School of Mathematics at The City University, London, England.

I am grateful to the Editor of the *North American Actuarial Journal* for granting me permission to reproduce in Chapter 3 (§§3.2, 3.5) and Appendix B parts of a paper due for publication in 1999.

I am also grateful to the following for their various contributions: Dr. Daniel Dufresne, University of Melbourne, Australia; Dr. Russell Gerrard, The City University, London, England; Mrs. Sally Grover, Librarian, Institute of Actuaries, Oxford, England; Dr. Thomas O'Brien, Bowling Green State University, Ohio; Prof. Arnold F. Shapiro, Pennsylvania State University, Pennsylvania; Mr. Richard Q. Wendt, Towers Perrin, Philadelphia, Pennsylvania; Dr. Howard E. Winklevoss, University of Pennsylvania, Pennsylvania, and Winklevoss Consultants, Connecticut.

My research was funded by way of a scholarship from the Department of Actuarial Science & Statistics at The City University, London, England.

Declaration

Abstract

This thesis is concerned with the funding of retirement benefits in a defined-benefit final-salary pension plan. A simplified model is set up in order to investigate the evolution of the pension funding system in a random economic and demographic environment. The long-term objectives of benefit security, contribution stability and flexibility are highlighted and are shown to relate to the motivation for advance funding of benefits. Actuarial valuations are construed as control processes to achieve these objectives through the determination of a suitable funding policy. One aspect of the funding policy is the choice of a suitable contribution as economic and demographic experience unfolds and deviates from actuarial valuation assumptions. Efficient actuarial methods of liquidating such deviations are considered when general autoregressive rates of return are projected, when new entrants into the plan vary randomly, and when other stochastic perturbations, such as discretionary contributions, are included. Two particular methods of determining contributions are compared and one is found to be more efficient at achieving the long-term objectives of security and stability. Stochastic optimal contribution and asset allocation decisions over a finite term, under certain rigorous assumptions and in a two-asset model, are derived using the dynamic programming principle. The optimal contribution control resembles the proportional spreading of surpluses and deficits while the optimal asset allocation is found to be a portfolio that dynamically hedges against the risk of inadequate benefit provision and unstable contributions. The methodology of actuarial valuations is examined qualitatively and the concept of a hedging or matching portfolio is found to be central to the valuation of pension plans, whether a market-oriented or a cash flow-oriented method is adopted. Various pension fund asset valuation methods are contrasted. The mathematical symmetry between asset gain/loss amortization and asset valuation is emphasised and an efficient way of determining actuarial asset values is investigated. Finally, the concept of economic prudence in actuarial valuations is explored in terms of the margins allowed in the choice of the discount rate (net of salary inflation) employed to value liabilities. The reasons for such prudence and its implications on funding are considered. It is well-known that conservatism leads to surpluses. Excessive and volatile surpluses as well as variable sponsor contributions must be avoided while retaining prudent funding objectives. A few simple methods (including an original method) to achieve this are studied. Some suggestions for further work are also discussed.

Chapter 1

Introduction

1.1 Aim and Outline

The aim of this thesis is to investigate the dynamics and actuarial control of defined benefit pension funds in a variable economic and demographic environment. Various actuarial methods are used in practice to achieve certain fundamental objectives of retirement benefit provision (Chapter 2). The determination of suitable contribution and asset allocation decisions (Chapter 3), pension fund asset and liability valuation methods (Chapter 4), and the use of prudence in the valuation basis (Chapter 5) are the methods of actuarial control that are investigated in this thesis. For this purpose, a pension plan model is set up under some very simplified assumptions. This admits a mathematical analysis and an objective evaluation of these actuarial methods. In each chapter, simple measures of the performance of pension funds towards meeting certain objectives are used and efficient or optimal methods are then derived.

1.2 Method

As in other disciplines, various methods need to be used to investigate pension funding problems. Different methods are required to satisfy different purposes. For many problems professional experience may generate satisfactory intuition. Sometimes, narrowly-defined problems of a practical and commercial nature dictate the use of numerical methods. In order to test actuarial methods and design pension fund systems, a more rigorous, mathematical method may be useful. This inevitably requires that a number of idealising assumptions be made. It follows that reality is simplified and possibly distorted. The consequent loss of realism and applicability must be balanced against certain advantages which justify the mathematical approach taken here:

Tractability: In order to obtain mathematical results that can be interpreted, simple assumptions are necessary at the outset.

Generality: A pension funding model that is very realistic, such as one based on simulations, is likely to be specific to particular situations, jurisdictions, economic circumstances, professional practice and custom. Results from a simpler model can be generalised and are usually less ephemeral; they may always be tested in more realistic models for particular practical purposes.

Optimality: Qualitative models based on experience or professional judgement may be difficult to ameliorate and 'tune': ideas that appear sensible *prima facie* turn out sometimes to be wrong. Complex 'real-world' models may not admit performance optimisation, except by trial and error.

Robustness: A model should allow us to test for robustness to different practical scenarios, even though it may not yield optimal solutions or complete accuracy.

Objectivity: Realistic models will not yield results that are definitive and any conclusion drawn is liable to subjective interpretation: we can never simulate all possibilities and are limited by preconceptions and computational expense. A mathematical model may be constructed with well-defined assumptions and may give results that are clearly interpretable.

Parsimony: A useful model is usually simple to understand and apply and is parsimonious.

Heuristics: A mathematical model can strengthen professional judgement and understanding of real pension funding issues by giving greater insight. It can also provide a suitable basis for teaching and learning about the fundamentals of pension funding.

1.3 Model

The simple mathematical model used in this thesis is based on the discrete-time stochastic model described and investigated by Dufresne (1986). This is itself an extension of the seminal model of Trowbridge (1952, 1963) designed to study the mathematics of actuarial pension funding methods in a simplified demographic and economic context that is static and deterministic. Bowers *et al.* (1976, 1979, 1982) advance the model of Trowbridge (1952) by incorporating several time-variant and deterministic features concerning the pension plan population and the return on assets. Dufresne (1986, 1988, 1989) and O'Brien (1986) allow these factors to be random so that the pension fund becomes a stochastic process. Haberman (1994b) includes more complicated models for the investment return on the pension fund as well as a number of other features. O'Brien (1987), Benjamin (1989), Boulier *et al.* (1995) and Sung (1997) further regard the pension fund as a dynamic financial system with actuarial input as a form of control. The model in this thesis subsumes aspects common to several of these models and comprises both deterministic and stochastic features as well as static, time-variant and dynamic approaches. Various other theories are in part or in full exploited in this thesis: the classical actuarial theory of pension funding (as developed by McGill (1964) and Trowbridge & Farr (1976), among others) and the financial theory of pension funding [Bagehot (1972), Treynor (1977), Black (1980)] are also relevant.

1.4 Terminology

The historical variations in pension arrangements around the world mean that pension funding has developed in different ways and the associated parlance is diverse. Some consensus on usage is developing in English-speaking countries but confusion may arise even within one country. Competing sets of terminology are favoured by various professional bodies and other authorities. Four features deserve emphasis:

1. The term 'real' is usually taken to mean 'net of salary inflation' in the context of final-salary pension plans in the Anglo-Saxon pension-actuarial jargon. In this thesis, unless otherwise qualified, the term 'real' always means 'net of salary inflation'. *All* mathematical symbols represent amounts that are real (net of salary inflation).

2. 'Cost' means accounting cost and not contribution in the U.K. 'Cost' is used to mean contribution in North America with the term 'pension expense' describing accounting cost. 'Cost' and 'contribution' are distinguished except in established terms such as *normal cost* and *supplemental cost.* Trowbridge & Farr (1976:22) also prefer 'contribution' to 'cost' but defer to common usage in North America.

3. I use the term 'amortization' as it is defined in North American finance and pension work (see below).

4. It is important in pension funding to draw a distinction between actual market values and values that are the results of actuarial calculations. Mathematical symbols in upper-case refer to 'actuarial' values, usually but not exclusively values placed on liabilities; whereas symbols in lower-case refer to market values, usually but not always terms on the asset side of the balance sheet. Terms such as the *unfunded liability* that involve a comparison between assets and liabilities are usually in lower-case because the assumption that values are consistent and comparable is then being made.

This glossary is not comprehensive, but covers some important terms as used in this thesis.

Actuarial basis: see *valuation basis.*

Actuarial cost method: see *pension funding method.*

Actuarial liability: the reserve as defined under various 'individual' *pension funding methods*; also *standard fund.*

Amortization: a schedule of payments (or a particular payment in the schedule) of $S/\ddot{a}_{\overline{n|}}$ at a given interest rate every year over a period of n years to pay off an amount S; in U.S. pension terminology, it is associated with removing actuarial *gains* and *losses*; in U.K. pension terminology, it has come to mean the repayment of a *deficit* or *surplus* spread in a regular manner over time, possibly over a moving term.

Asset valuation method: a procedure for establishing the value of the assets of a pension plan.

Contribution: the total contribution paid by the *plan sponsor* and/or *plan participants*, and comprising the *standard contribution* (or *normal cost*), the *supplementary contribution* (or *supplemental cost*) as well as any *discretionary extra contribution*; sometimes *cost* is used in North America.

Contribution adjustment: the *supplementary contribution.*

Cost: in the U.K. it refers to the accounting cost; in North America it is often used to mean *contribution* while *pension expense* refers to the accounting cost.

Deficit: the excess of *actuarial liability* over assets; also *unfunded liability*, a negative *surplus.*

Employer: a *plan sponsor.*

Discretionary extra contributions: contributions (positive or negative) that a pension fund sponsor may pay into a fund in addition to the actuary's recommended contribution.

Gain: a negative *loss.*

Initial unfunded liability: unfunded liability or *deficit* arising because of a change in valuation assumptions, funding method, amendment to benefit entitlements or because of past service, but not because of unanticipated experience.

Liability valuation method: a procedure for establishing the value of the liabilities of a pension plan.

Loss (actuarial or inter-valuation or experience loss), $l(t)$: unanticipated change in *unfunded liability*; the difference between actual unfunded liability at the end of the year and the unfunded liability as anticipated on the *valuation basis* for the year.

New entrants: plan participants who have joined the *pension plan* at the beginning of the year.

Normal cost: a regular contribution or premium as defined under various 'individual' *pension funding methods*; also *standard contribution.*

Pension: income in retirement.

Pension expense: accounting cost in Financial Accounting Standard No. 87 (FAS87).

Pension fund: the collection of assets owned by a pension plan; 'superannuation fund' in Australia and New Zealand.

Pension funding method: a systematic way of accumulating funds to meet retirement benefits; also *actuarial cost method.*

Plan sponsor: a corporate *employer* who sets up a pension plan.

Projection: a forecasting exercise, using stochastic or scenario-based assumptions, distinct from a *valuation.*

Real: net of *salary inflation* in pension parlance; net of *price inflation* in economic parlance.

Salary inflation: the general increase in levels of wages across the economy, excluding merit and promotional salary increases; also *wage inflation.*

Standard contribution: see *normal cost.*

Standard contribution rate: standard contribution expressed as a percentage of payroll.

Standard fund: see *actuarial liability.*

Supplementary contribution: the portion of *contribution*, which is additional to the *standard contribution* (or *normal cost*), and recommended by an actuary following a *valuation*; also *contribution adjustment* or *supplemental cost*; see also *discretionary extra contribution.*

Supplemental cost: see *supplementary contribution.*

Surplus: a negative *deficit*; also actuarial surplus.

Term structure of interest rates: the relationship between yields to redemption and term to maturity of government securities (also yield curve).

Unfunded liability: a *deficit* or negative *surplus*; see also *initial unfunded liability.*

Valuation: a consolidation of cash-flows into a pension fund performed regularly to establish the financial status of a pension plan and recommend contributions on an ongoing basis; a valuation can be performed for several statutory or regulatory purposes; also 'actuarial investigation'.

Valuation Basis: the set of assumptions made by an actuary for the purpose of a *valuation*; also *actuarial basis.*

Valuation Discount Rate: the rate at which pension fund cash flows are discounted for the purpose of a *valuation.*

Wage Inflation: see *salary inflation.*

1.5 Glossary of Symbols

$\alpha(t)$	random risk premium in §3.6
α	$\mathrm{E}\alpha(t)$ in §3.6
δ_g	rate of growth of membership
$\delta(t)$	logarithmic (or geometric) rate of return
ϵ	$\mathrm{E}\epsilon(t)$
$\epsilon(t)$	random additive perturbation
σ^2	$\mathrm{Var}i(t)$ (except in §3.3 where it is $\mathrm{Var}\delta(t)$)
σ_ϵ^2	$\mathrm{Var}\epsilon(t)$
σ_g^2	$\mathrm{Var}g(t)$
a	entry age
$adj(t)$	supplementary contribution or contribution adjustment
AL	actuarial liability or standard fund
B	benefit outgo
$c(t)$	contribution
d	$i/(1+i)$
d_v	$i_v/(1+i_v)$
f_0	initial fund level (at time 0)
$f(t)$	fund level (at market value)
$F(t)$	actuarial asset value
$i(t)$	arithmetic rate of return
i	mean arithmetic rate of return
i_v	valuation discount rate
g	$\mathrm{E}g(t)$
$g(t)$	population distribution function
k	$1/\ddot{a}_{\overline{m}\rvert}$
K	$1-k$
k_p	parameter in equation (5.32)

k_i	parameter in equation (5.32)
$l(t)$	actuarial loss
λ	smoothing parameter in asset valuation method
m	period over which surpluses and deficits are spread forward
m_a	period over which gains and losses are amortized
m_s	m in §3.5
m_i	$1/k_i$ or integral spreading period
NC	normal cost or standard contribution
$P(t)$	amortization payment for initial unfunded liability
q	$u^2+\sigma^2$
Q	$1-k$ in §3.3
r	retirement age in §3.1; risk-free rate of return in §3.6
$u(t)$	$1+i(t)$
u	$\mathrm{E}u(t)$
$U(t)$	unamortized part of initial unfunded liability
ul_0	initial unfunded liability (at time 0)
$ul(t)$	unfunded liability or actuarial deficit
v	$1/u$
w	last age in life table
$y(t)$	proportion of fund invested in risky asset in §3.6
y	fraction of ul_0 that is *not* amortized separately

Chapter 2

Pension Funding and Actuarial Valuations

2.1 The Nature of Pension Funding

Pensions. For an employee, a pension is income received during retirement. For an employer, pensions are a form of deferred remuneration. A pension plan or scheme is an arrangement through which this remuneration can be organised. A pension fund is a trust fund holding the assets required by a pension plan for the purpose of financing pension benefits. How to finance pensions and ancillary benefits is the subject of pension funding theory.

Benefit Policy. An employer who contemplates remunerating his employees through pensions is faced with a decision as to how to plan financially for this. A starting point is to determine what benefits are to be provided: a 'benefit policy' is required. The objectives of pension provision from the viewpoints of employees and employers must be considered:

1. Employees may expect a pension as part of their remuneration package. The pension must be sufficiently high but also guaranteed.

2. Employers wish to attract and maintain a highly motivated workforce. They must be willing to pay benefits that are competitive and attractive. The pension plan must be cost-effective.

Funding Policy. Clearly these objectives are not complementary. Excellent benefits may prove costly. Decisions as to how to finance these benefits must be made: a 'funding policy' must be developed. A pension fund may be considered to be a savings and investment vehicle that helps and encourages a company to provide pension benefits.

The Pension Funding System. Actuaries are involved in controlling the mechanism through which pensions are financed. This involves determining and valuing cash flows. At its simplest, the pension fund can be reduced to three sources of income (employee and employer contributions and investment earnings) and two outgoing cash flows (benefit payments and expenses). A number of uncertain factors will affect the cash flows and actuarial control is required to balance outgo and income. The pension fund may be managed by controlling the benefit outgo, the funding policy, the investment strategy of the fund and the contributions in order to achieve financial balance. The pension fund may be viewed as a mechanism to be regulated, as described by Trowbridge (1966).

The Substitution of Contributions for Investment Income. The funding and benefit policies are interdependent but distinct. For the plan participants, the benefit policy dictates the level of income they will receive in retirement, whereas the funding policy will determine how secure their pension rights are. Their pension rights will normally become effective a long time ahead in the future. For the sponsoring company and the pension fund, the benefit policy determines the total benefit outgo over time, whereas the funding policy will affect cash flows in order that these benefits are met as and when due. Regular contributions together with investment income from the assets of the fund combine to meet the pension benefits of retired employees. The funding policy is about the substitution of contributions by investment income. Funding is about setting monies aside for an event in advance of its contingency. The earlier this is done, the more investment income is being substituted for contributions.

The Structure of a Pension Plan. The income and outgo of a pension fund must ultimately balance. The risks involved in not achieving this balance can be borne either by the sponsoring company or by the plan membership.

- Contributions may be predetermined and benefits may be somehow targeted without any guarantees as to their eventual levels. Such a pension plan is called a *defined contribution* plan: examples are money purchase schemes or accumulation funds.

- Alternatively, benefits can be defined according to set formulae and contributions can be determined to achieve these levels of benefits. This is a *defined benefit* pension plan.

Defined benefit plans are common in many countries and are the subject of this thesis.

Pension benefits are normally a function of the final salary (or an average of final salaries) earned by the pension plan member. This is because employees should be able to maintain their standard of living after retirement. Pension plans may also be based on career-average salary or indeed be independent of salary altogether, as in social insurance systems. We will only consider final-salary plans, by far the most common in English-speaking jurisdictions.

2.2 The Objectives of Pension Funding

This thesis is therefore concerned with the funding of pension benefits in a defined-benefit final-salary pension plan. First, we explore some of the objectives of pension funding and describe the actuarial input in achieving these objectives. In order to explore this further, we will set up a simple mathematical projection model for the pension fund and incorporate aspects of the valuation process.

2.2.1 Reasons for Funding

The reasons why assets are held in a pension fund to meet pension benefits are as follows:

Benefit security: the benefits to which plan members are entitled are perceived to be more secure if funds are set aside in advance for them;

Contribution stability: the plan sponsor makes contribution payments which may be budgeted for and may be more stable in the long term if a fund is built up, while the contributions can also change more smoothly over time if necessary;

Value: both the plan members and the plan sponsor (and his shareholders) want greater value through

- *flexibility*: advance funding implies a degree of flexibility in the contributions required every year, so that there is less strain on corporate cash flows;
- *tax advantages*: advantage may be taken of tax credits on investment earnings on assets in a pension fund;
- *stable and realistic accounting cost*: a funding arrangement allows the accounting charge to be recognised in the period in which the benefits accrue such that it has the least effect on the financial status of the sponsor;
- *fewer expenses*: if benefits are financed in a regular and organised way, expenses such as investment transaction costs and administrative or legal expenses can be minimised.

2.2.2 Types of Objectives

The objectives of pension funding are closely tied to the reasons why funding is undertaken in the first place. These objectives are broadly of two types:

Special Objectives. Special or 'compliance' (Dyson & Exley, 1995) objectives are normative and are usually externally imposed on pension plans, through various statutory and regulatory obligations placed on them. They are usually independent of the funding policy

Special Objectives	Interested Stakeholder
Solvency (Minimum funding)	Plan members
Surplus (Maximum funding)	Revenue authorities
Appropriate accounting	Company shareholders

Table 2.1: Some special compliance objectives of funding for retirement benefits.

being followed by the pension plan. 'Solvency' (or minimum funding) is one such special objective. From the perspective of plan members, the minimum wind-up benefits that need to be paid if the company ceases to operate is very important and should determine a minimum level of funding. There are also limits to pension fund surpluses (or maximum tax-exempt funding) allowed by revenue authorities, and a critical objective will be to avoid such excessive surpluses. Finally, there are rules regarding pension accounting that need to be followed: as benefits accrue, they must be financed and accounted for in a consistent and realistic manner. Some special objectives, along with the party most concerned by them, are listed in Table 2.1.

Management Objectives. Ongoing or 'management' (Dyson & Exley, 1995) objectives depend both on the benefit and funding policies. Whereas special objectives are imposed by statute or by regulation, management objectives are intimately related to the fundamental reasons that justify advance provision of benefits. Attention is often paid to the special objectives because there may be penalties if they are not satisfied. But management objectives represent the true underlying long-term *motivation* for pension funding and are internally generated as a result of creative tension between the conflicting interests of various stakeholders and as a result of arbitration by actuarial and fund management professionals. Management objectives are based on the 'going-concern' principle, i.e. that the plan will exist for a long time. Whether the long-term benefits promised to members upon retirement in 20 or more years are adequately funded will matter to the plan members: it is of little comfort if the fund is technically solvent but the benefits they are promised many years hence are not being financed. Plan members will be concerned with whether the substitution of contributions by investment income affects the security of their benefits. The plan sponsor will be concerned with the cost of funding pension benefits at a particular pace. From his perspective, the opportunity cost of providing pension benefits must be acceptable and it is essential that pension provision remains economic. The contributions required from him will need to be predictable. All the parties will want to maximise economic value by taking advantage of favourable revenue rules as well as of the inherent flexibility of advance funding. Actuarial involvement with a pension fund is therefore about balancing the various objectives of security, stability, flexibility etc. Some of these management objectives are listed in Table 2.2 on the next page. In the next sections and the rest of this thesis, we consider various issues

Ongoing Objectives	Interested Stakeholder
Benefit security	Plan members
Contribution stability	Plan sponsor
Economic value	Plan members and sponsor

Table 2.2: Some ongoing management objectives of funding for retirement benefits.

pertaining to these management objectives.

2.2.3 Security

Security of Benefits or Adequacy of Fund

The most important objective of pension funding, certainly from the point of view of plan members, is to secure long-term benefit entitlements. This means that the pension fund must hold enough assets to meet pension benefits, as and when they are due.

Solvency and Security

Security, as a funding objective, usually encompasses the technical or regulatory concept of fund solvency, which we have classified as a special objective (Table 2.1). In other words, the process of funding on an ongoing basis usually (but not always) satisfies the criteria for solvency. In addition to statutory minimum funding requirements that exist in various jurisdictions, the benefit policy or rules of a plan determine how much should be reserved for wind-up benefits, i.e. how much assets to hold in order to satisfy the minimum benefit pay-out upon wind-up or discontinuance of a pension plan. On the other hand, *funding* is concerned with the ongoing maintenance of, not only the pension plan, but also the pension rights of plan participants. There is no unique best level of funding, and the level of funding will depend upon the funding strategy chosen, but this will normally imply holding more than the minimum solvency requirements or wind-up benefit reserves:

1. A pension fund is a savings and investment vehicle that encourages and helps sponsors to provide benefits for possibly several generations of employees, in changing economic and demographic conditions. It is not merely a deposit for the current generation.

2. A pension plan is set up not only to secure members' benefits but also to help sponsors by achieving stable and flexible contributions, realistic and suitable pension expensing etc. and this may clearly require more than just funding for the termination benefits.

3. Although a pension plan may terminate, the members of a pension plan may move to another plan, in which case the benefits that they have accrued during membership of their original plan will only be preserved if backed by the transfer of sufficient assets.

As a matter of concept rather than semantics, we must therefore strictly differentiate *benefit security* (or fund adequacy) from *fund solvency.* The complex issue of fund solvency is not directly addressed in this thesis. Fund solvency is assessed during solvency valuations, when the actuary assumes that the plan is being wound up: market values of assets are taken and the discontinuance liability might be the cost of purchasing annuities for each plan member (or some other measure, depending upon local regulations). In general (but not always), a fund is solvent if its ongoing funding level is high. Note that it is also argued, for example by McLeish & Stewart (1987), that security should be almost equivalent to technical solvency.

Over-funding and Security

It may be to the disadvantage of both plan sponsor and members if excessive funds are diverted from normal company operations to the pension fund. The opportunity cost of over-funding may be high and *in extremis* may endanger corporate profits and members' employment. Over-funding must be avoided. In addition, pension funds have special tax privileges in many jurisdictions. Financial penalties may be incurred by the plan sponsor if excessive surpluses are held in the pension fund, since these can be construed as 'tax havens'. For both plan sponsor and plan participants, "third-party leakage" of economic value, as described by Exley *et al.* (1997), should be minimised: the pensions promised are adequately funded and the funds set aside for providing these pensions are secure against heavy taxation and resultant financial loss, if maximum surplus limits are not breached. Security therefore involves a balance between under- and over-funding and in addition is not the sole concern of plan members. Loades (1988) considers the level of security provided by a pension fund to be "a measure of the balance between the interests of members and employers." When pension fund valuations take place, actuarial 'management' of pension funding endeavours to achieve secure pension funding levels so that neither under-funding nor over-funding occurs and a suitable *intermediate* course is steered: see, for example, Farren [Thornton & Wilson (1992a): discussion], Wise [McLeish & Stewart (1987): discussion] and Grubbs [Cronquist & Dreher (1972): discussion].

2.2.4 Contribution Stability

This objective is concerned with what Trowbridge & Farr (1976) term the "budgeting problem" faced by the plan sponsor. In other words, the pension plan must operate in such a way that the sponsor can always allow in advance for contributions to the fund in his overall corporate financial planning. Indeed, a reason for funding benefits in advance is that an employer can spread contributions over time and does not immediately bear the costs of the unpredictability of pension provision. The sponsor will find it advantageous if the contributions *recommended* by the actuary are relatively stable over the long term. "Smooth contribution patterns" (Trowbridge & Farr, 1976) are often sought.

Note that the sponsor's contributions are not necessarily equal to the pension accounting cost. Although in the past the contributions that were made by an employer to the plan were taken as the pension expense in the employer's accounts, this is not so any more. The requirement for contribution stability does not therefore necessarily lead to smoothed and unrepresentative pension accounting costs. Note also that contributions will naturally vary as a result of economic wage inflation as well as varying plan membership size, among other things. To adjust for both influences, the contribution rate (contribution as a percentage of payroll) is often considered and the plan sponsor is said to seek a stable contribution rate.

2.2.5 Contribution Flexibility

Funding pensions in advance is inherently flexible in that it allows contributions to be spread over the long term. This means that sponsors are able to vary their contributions, within limits, if it is necessary or advantageous to do so. If capital projects are being planned and pension contributions are sufficiently flexible, then sponsors can temporarily divert cash and so do not have to rely on expensive external financing. Sponsors may alternatively wish to take advantage of tax credits and contribute more when their cash is available. A good pension funding policy will therefore ensure that contributions are not only stable over the long-term but are also flexible over the short-term. This maximises economic value for the corporate sponsor and also indirectly for the working members. Flexibility in employers' contributions is naturally afforded by the amortization of gains or losses (or spreading of surpluses and deficits) over future periods as well as by discretionary extra contributions which a sponsor may pay into the fund.

Although it is an important objective in the management of pension funds, flexibility is subsidiary to the other objectives, and in particular it cannot override the security of benefits. Indeed, 'flexibility' implies that there is some prior constraint on the sponsor contributions. This constraint is the "will to fund" and plan solvency mentioned by McGill (1964:319). McGill (1964:325) also juxtaposes flexibility with the "responsibility" of the sponsor to provide promised pension benefits. Snelson (1970) likewise regards advance funding as implying "a discipline which in itself imposes limitations on flexibility." Griffin (1966) also contrasts a sponsor's requirements for "long-range contribution stability" against "the safety valve of short-range contribution flexibility".

2.3 Valuations and Control of Pension Funding

2.3.1 Valuations

During regular actuarial valuations of a pension plan,

1. the performance of the plan in respect of the various objectives of funding is measured;

2. recommendations are made to manage or control the pension funding system to attain these objectives.

An ongoing or management valuation monitors whether management objectives (as in Table 2.2) are being achieved and leads the actuary to decide upon appropriate contributions to the fund. Special valuations (discontinuance or solvency valuations, accounting valuations, as well as other statutory valuations) will also be performed in connection with certain special objectives (Table 2.1). In this thesis, we concentrate on ongoing management valuations.

Four distinct parts of a valuation can be identified.

1. The first part of a valuation consists of establishing a set of valuation assumptions (a *valuation basis*) about the various economic, demographic and statistical factors that will influence the future evolution of the pension funding system.

2. Future pension liabilities must also be estimated (the *liability valuation*) based on the valuation assumptions in order to establish

   - the reserve (*actuarial liability* or *standard fund*) that should be held to meet such liabilities,
   - the regular premium or contribution (*normal cost* or *standard contribution*) required to maintain such a reserve;

   and thus the fund and contribution targets implied by a chosen funding method are determined.

3. A suitable value must be placed on the assets held in the pension fund (the *asset valuation*).

4. In the final part (the *actuarial recommendation*), the actuary compares the future asset and liability cash flows for the fund in a consistent way, assesses whether the funding objectives are being attained, and

   - determines a suitable contribution;
   - comments on a suitable investment strategy;
   - comments on possible benefit improvements and cognate matters.

The procedure whereby liabilities are evaluated and contributions are determined is commonly referred to as the *pension funding method* or *actuarial cost method*. An important concept in this thesis is that these four parts are *interdependent* and must be chosen in a *consistent* manner.

2.3.2 Pension Fund Control

The three principal ways of exercising control on the pension fund and ensuring that a reasonable balance between the objectives is achieved [Winklevoss (1993:226–241)] are through:

1. a benefit policy;
2. a strategic investment policy;
3. a funding policy.

Benefit Policy

The benefits promised in a defined-benefit pension plan are fixed by definition. The benefit rules are nevertheless changed occasionally. The pension promise is defined in the sense that it is modified only infrequently and is usually only improved. Actuarial input is required to determine what changes are acceptable. It may be customary for pensions in payment to be augmented on a discretionary basis. Fujiki (1994) describes this as one of the ways in which actuarial control may be effected. When large surpluses emerge as a result of profitable investment strategy, they may be shared by the plan sponsor through reduced contributions and by members through *ad hoc* increases in pensions in payment. Adjusting pensions in payment has been particularly prevalent in inflationary economic circumstances in the past in the U.K. and it is viewed as a 'regulator' by Shucksmith [Thornton & Wilson (1992a): discussion]. Such control may be one-way as it is politically difficult to reduce benefits. Another aspect of control through benefit policy is described by Thornton & Wilson (1992b:§8) who note that, in order to control investment risks, vested plan members in the U.K. who leave service could be encouraged to transfer their pension rights to the plan sponsored by their new employer rather than defer their pension benefits. It is also important to control benefit outgo if the pension plan is in financial difficulties. Guidance Note 461 of the Institute of Actuaries of Australia (1994) suggests various "means of control" of technically insolvent funds, such as paying benefits in instalments or even suspending benefit payment until solvency is restored.

Investment Policy or Asset Allocation Control

Investment earnings usually exceed contribution payments and it is possible to control the performance of the plan if the investment of the fund is successfully managed. It is important to determine a suitable investment strategy that allows the pension fund to meet its various objectives. One possible investment policy is to choose to invest in assets that hedge or match liability cash flows so as to reduce the chances of insolvency and maximise security. As losses may be curtailed, this may also stabilise contributions. On the other hand, it has also been argued that profitable investment opportunities may be missed. It is also said that pension funds ought to invest in such a way as to gain from the favourable taxation rules applying to pension funds. This should benefit both sponsor (and its shareholders) and plan members as the sponsoring company gains financially. This minimises the loss of economic value by giving considerable flexibility to the sponsor, but may not appeal to plan members' sense of security. Many pension funds are also managed so that returns are maximised (and contributions minimised) subject to an acceptable degree of risk. Various investment strategies appear to be followed in practice. Control through asset allocation is dealt with in §§3.6 and 4.3.5.

Funding Policy

Actuaries do not have direct control on setting investment strategy and are sometimes confined to, at most, an advisory role in that respect. In a *defined benefit* plan, the scope for managing a fund through benefit improvement is also limited. A far more important way in which actuaries contribute to the successful management of a pension plan is by developing and implementing a funding policy. The funding policy determines the substitution of contributions by investment income and therefore determines the amount and incidence of payments in order to meet future pension benefits. In particular, the actuary controls the pace of funding, i.e. the speed with which the pension plan builds up funds to a desired funding level. This control occurs through the choice of:

Funding method. Some funding methods require larger contributions earlier on and build up a larger fund. Funding methods are discussed in Chapter 3. The actuarial literature on pension funding contains extensive discussions of the choice of funding methods, to which exclusive attention is sometimes paid.

Valuation basis. It is well known that the use of more conservative assumptions when valuing pension liabilities will speed up funding. The "funding basis" is part of the pension funding "control system" of Benjamin (1989). He focuses on the 'valuation rate of interest' (an arithmetic average of real pension fund returns) as the key control variable. Fujiki (1994) also regards the determination of valuation assumptions as a control decision; he simulates various economic scenarios and investigates how assumptions may be set. The concept of prudence in the choice of one important element of the valuation basis, the discount rate, is examined in Chapter 5.

Asset valuation method. Asset values may be volatile and are often smoothed. This is meant to stabilise the funding process over the long term. Asset valuation is discussed in some detail in Chapter 4.

Contribution policy. If assumptions are not realised, losses need to be defrayed and surpluses need to removed, so that contributions must be appropriately adjusted. The method of adjustment is usually applied consistently and affects the dynamics of the funding process. Contribution control has been investigated by Dufresne (1986), O'Brien (1986, 1987), Haberman & Sung (1994), Boulier *et al.* (1995, 1996) and Cairns (1997). See Chapter 3.

These control mechanisms closely parallel the four parts of an ongoing valuation exercise as identified in §2.3.1. The ongoing valuation is in fact intimately related to the funding policy. During a valuation exercise, the actuary can *implement* a suitable funding policy. Benjamin (1984) regards actuarial valuations as control techniques to achieve long-term financial control.

2.4 Uncertainty in Pension Funding and Projections

A pension plan valuation is an exercise in controlling pension funds to achieve long-term objectives in an uncertain financial and demographic environment. These uncertain variables are often explicitly modelled using projections or forecasting exercises of various types.

Whereas valuations are concerned with determining the financial state of the pension plan at the time of the valuation and the best course of action in the inter-valuation period, the aim of projections is to investigate pension funding over a longer time horizon. There is therefore a distinction between valuation assumptions and projection assumptions [Winklevoss (1977:201), Kemp (1996:¶4.5)]. Various events of a demographic and economic nature will affect the future course of the pension plan and its fund. It is impossible to be certain about the magnitude and timing of these events and yet decisions must be made as to the management of the fund to mitigate or control the effect of these various factors. Valuation assumptions represent simple, and usually deterministic and time-independent estimates of these factors so that their effects can be measured on the asset and liability cash flows of the pension fund. In practice, valuation assumptions will also contain margins for prudence. Projection assumptions, in contrast, need to be as realistic as possible and are often stochastic. McGill *et al.* (1996:574) refers to projections and projection assumptions as "stochastic forecasting" and stochastic "experience assumptions" respectively. Different projection assumptions may be used according to the aim of the projection exercise.

Projections are used for two main purposes. The most common one is to investigate and set strategic investment policy. Projections are also of use when investigating funding policy, such as when deciding on suitable valuation assumptions or funding methods. A popular type of projection exercise is Asset-Liability Modelling (ALM), which requires an integrated model for future investment returns on various asset sectors as well as for liability cash flow. Such a model will usually be stochastic or scenario-based and enable the actuary or fund manager to devise an optimal investment strategy that takes into account future liabilities of the plan. Asset-liability modelling has been employed in the context of pension funding mainly to investigate investment policy for example by Daykin *et al.* (1993) and Kemp (1996). Winklevoss (1982, 1993), McKenna (1982), Kingsland (1982), Loades (1988), MacBeth *et* *al.* (1994), Bilodeau (1995), Alphen *et al.* (1997), Kleynen (1997) and Haberman & Smith (1997) among others, show that ALM can also be used to investigate aspects of funding and benefit policies such as contribution levels, pension costs, benefit indexation, valuation bases, demographic changes and funding methods as well as investment policy.

In the following sections, some of the factors, and corresponding valuation and projection assumptions, that affect pension funding are described.

2.5 Demographic Variables in Pension Funding

Various factors influence the membership of a pension plan and consequently affect pension funding. Demographic factors include: mortality (for both retired and active plan members), retirement (including normal, early and late retirement), new entrants (at multiple entry ages), involuntary withdrawal (redundancy), voluntary withdrawal (resignation), disability (including permanent invalidity, disability recovery, partial disability).

Other non-economic variables (sometimes referred to as 'statistical' factors) in the pension funding process include: salary scale (according to merit, promotion and longevity and *excluding* economic wage inflation), option election rates (e.g. commutation to lump sum; this is material if options are not actuarially neutral and there are opportunities for selection by plan members), marriage rates and family composition and dependency (members' spouses and dependants may be entitled to benefits).

Valuation assumptions made about many non-economic factors are collated in the form of *service tables*. These may be generated from past experience of the pension plan or of similar plans (from the same industrial sector) and from national statistics. Plan-specific information and general trends are also be used by actuaries. Demographic *projection assumptions*, as used in Asset-Liability Modelling for example, do not usually differ from valuation assumptions [Kemp (1996:¶A.7)]. This is because changes in the rates of various decrements are observed to change only slowly over several years. For a large plan, the law of large numbers means that a deterministic approach may well be suitable when projecting the evolution plan. In addition, these changes may well be small and immaterial on funding as compared with economic factors and so demographic projection assumptions are often the same as demographic valuation assumptions. Nevertheless, stochastic models allowing for demographic factors such as redundancy, mortality, disability etc. that affect pension liabilities have been constructed by Shapiro (1979) and De Dominicis *et al.* (1991) among others.

Two demographic factors that may change significantly and that may have considerable financial effects in the long term are new entrants and redundancy numbers (involuntary withdrawal). They are sometimes modelled very explicitly in projection work. Tepper (1977), for example, uses an Asset-Liability Model that integrates the "dynamics of the workforce", including "labour use projections", "turnover considerations" and "personnel policy" along with asset modelling. Demographic projection assumptions are usually deterministic and scenario-based and are common in Asset-Liability Modelling work. They may be useful in representing various scenarios:

- The effect of 'trends' such as the growth and decline of firms and industries and their impact on recruitment policies may be modelled (Snelson, 1970). Trends in the plan participant population may be material to the *durability* of pension plans [Lee (1986:159)].
- Another trend that may be usefully modelled in a deterministic manner might be national demographic shifts. In the case of large, public funded pension plans, demographic shifts, such as an ageing population, falling fertility and immigration, that are projected to occur can be modelled.
- In the case of medium-sized private pension plans, the effect of a merger or consolidation or planned and possibly phased redundancies as a business is 'restructured' may be explored. Kemp (1996:¶A.7) suggests that "a planned acquisition or redundancy exercise" may be modelled deterministically.

It is worth noting that 'open-group' valuation techniques such as the *forecast valuation method* [Fleischer (1975), Schnitzer (1977)] also employ deterministic assumptions about the evolution of new entrants into the plan. Bowers *et al.* (1976, 1979, 1982) and Winklevoss (1993) model pension plans with deterministic stable populations (i.e. with constant rate of exponential growth). Randomness in new entrant numbers may also have consequences for the stability and security of the pension fund. O'Brien (1986) assumes random rates of growth of the pension plan population and Mandl & Mazurová (1996) assume that new entrant numbers vary as a stationary autoregressive process. The effect of variable new entrants is investigated in Chapter 3 (§3.4.4).

2.6 Economic Variables in Pension Funding

The major economic variables that enter the pension funding process include:

1. price inflation;
2. wage inflation (excluding salary scale based on merit, promotion and longevity);
3. investment return on various asset types and on the fund as a whole.

We now consider each of these economic factors in turn and examine the valuation and projection assumptions used as a proxy for them.

2.6.1 Price Inflation

Relevance to Pension Funding

There are many different economic theories as to the sources of price inflation. Inflation appears to stem from a number of factors including imbalance between supply and demand (Keynesian demand-pull inflation), imbalance between wage increases and productivity improvements (cost-push inflation), expansion of money supply (monetarist theory of inflation), economic cycles. The general increase of prices in the economy has important consequences for the funding of retirement benefits. The most important consequence is the erosion of the purchasing power of pensioners who receive benefits that are nominally fixed or that do not allow for inflation. The adequacy of pension benefits can only be maintained if some form of linkage of post-retirement benefits with prices, either by formal indexation or on a discretionary basis, is undertaken. Price inflation is also important because it is usually accompanied by salary inflation, which has obvious consequences for a final-salary plan. Finally, the investment return on real assets held by the fund will be influenced by the past and present levels of price inflation.

Price Inflation Valuation Assumption

The selection of an assumption as to economic price inflation in the future is almost invariably the starting point of the valuation basis. The price inflation assumption matters particularly if benefits that are indexed with consumer or retail price indices are promised (as in most large public-sector schemes in the U.K.). Often discretionary post-retirement pension increases are also informally related to the level of inflation in the economy. In other cases, some form of limited price indexation (as in private defined-benefit plans in the U.K.) also means that the assumption as to price inflation matters. The relative levels of the valuation discount rate (or the investment return assumption) and the price inflation assumption are particularly critical to the valuation of pensioner liabilities. In most cases, if price inflation is assumed to be high relative to investment returns, the value of pensioner liabilities will be overestimated, which is more prudent. The price inflation assumption is also important as price inflation affects various other factors, and it must be chosen consistently with other assumptions.

The price inflation assumption is usually set after consideration of price indices data. It is important that attention be paid to long-term historical experience and not just to the recent past, given the long duration of pension liabilities. It is equally important that a forward outlook be maintained: economic forecasts of inflation and the yields on bonds and government securities are important sources of information. The difference in yields between real return and conventional bonds (e.g. index-linked and conventional gilt-edged securities in the U.K.) gives an indication of market expectations of inflation, notwithstanding the uncertainty concerning the relative risk premiums in the returns from such bonds. It is most common to choose a single term-invariant price inflation assumption, but "select and ultimate" rates are also allowed by the Actuarial Standard of Practice No. 27 of the American Academy of Actuaries (1996:¶3.5.2).

Price Inflation Projection Models

Price inflation is most commonly modelled using ARIMA techniques in pension funding work. Inflation in any one year is projected to be dependent on inflation in the past year, probably because of yearly wage negotiations (these will depend on price increases over the past year) as well as yearly price reviews by firms. Wilkie (1987) models the force (or logarithmic rate) of price inflation in the past few decades in the U.K. as an autoregressive process of order 1 (AR(1)). His model has a 'cascade' structure and the price inflation model 'drives' all other variables. Current price inflation depends therefore only on past inflation. Daykin *et al.* (1994) and Sharp (1992) describe a very similar model using other data. Unconditional AR(1) models for projecting price inflation are used by Cairns (1994), Daykin *et al.* (1993) and Knox (1993), among others, in pension funding work.

Several variations on the basic autoregressive model exist:

1. There is evidence to suggest that the residuals of various price inflation series may not be normally distributed and are negatively skewed and 'fat-tailed'. Daykin *et al.* (1994) use shifted Gamma distributions as an alternative to normal distributions.

2. There is also evidence that the variance of the residuals is not constant over time. The addition of ARCH (autoregressive conditional heteroscedastic) (Engle, 1982) effects allows not just current inflation, but also the variance of current inflation, to depend on the previous year's inflation. The variance of the error term is itself stochastic and so the variance of the inflation series is non-stationary and depends on the previous year's inflation. This approach is followed by Wilkie (1995:§2) for example.

3. Another variation is to include both price and wage inflation in a vector autoregressive model, as also discussed by Wilkie (1995:§2), and used by Kleynen (1997) in pension fund projections. In Thomson's (1996) South African asset model, price inflation is *not* independent of other economic variables and is driven by lagged equity dividend growth.

4. Historic data show that price inflation has become unpredictable over the recent past with irregular and large shocks occurring, for instance, after the 1973 Oil Crisis, along with periods of sustained high and low inflation, possibly in line with economic cycles (Huber, 1996). For these reasons, Clarkson [Geoghegan *et al.* (1992)] suggests a non-linear modification to the AR(1) inflation model comprising positively biased trend and random shock (Bernoulli random variable) components. Parameter estimation for such a model is not straightforward (Wilkie, 1995:§2).

2.6.2 Wage Inflation

Relevance to Pension Funding

Wage or salary inflation is the general increase in wages and salaries of workers across the economy, excluding salary increases for reasons of promotion, seniority or merit. It is generally thought of as resulting from:

Price inflation: workers want to maintain their purchasing power and real standards of living and unions will demand that salaries rise at least with general price levels;

Productivity improvements: firms that remain in business must be making profits that rise over and above price inflation as a result of improved productivity, part of this being transferred to labour as higher wages.

Daykin (1976:294, 1987) and Thornton & Wilson (1992a:244), for example, illustrate with U.K. data that salary levels rise faster than prices. It is also thought that there will usually be a lag between price and salary inflation as the labour market reacts, through varying trade union power, government regulation and competitiveness, to price levels. The dynamics of salary inflation and its relationships to other economic variables are very important to a final-salary plan and will determine its investment strategy.

Wage Inflation Valuation Assumption

The salary or wage inflation assumption is distinct from the salary or merit scale mentioned in §2.5. (The salary scale is regarded as an *economic* assumption in North America as in the Standard of Practice for Valuation of Pension Plans of the Canadian Institute of Actuaries (1994:¶4.03), for example. This is only a matter of classification and I will consider that the wage inflation assumption excludes salary scales.) It is usually easier to determine a suitable assumption as to real salary inflation first, as salaries are more stable when considered relative to prices. The value placed on the liabilities of active employees is particularly sensitive to the assumption of investment return (or valuation discount rate) relative to wage inflation. In a young, immature plan, active liabilities dominate the liabilities of pensioners and members whose benefits are in deferment. Assuming high salary inflation relative to investment return, i.e. assuming that returns net of salary inflation are low, usually increases the value placed on active liabilities, which may again be a prudent measure. National earnings data and data on productivity are the most important sources of information used to determine a wage inflation assumption. Again, inordinate emphasis must not be placed on the recent trend in salaries at the expense of long-term trends.

Wage Inflation Projection Models

Wage inflation is often modelled as a simple additive component to the main price inflation model, i.e. *real* wage inflation is projected because it is more stable. In his mathematical model of asset-liability cash flow matching for a closed pension fund, Wise (1984) projects earnings inflation as price inflation plus a constant level. A similar model is employed by Knox (1993) in the context of defined contribution superannuation funds: constant level plus AR(1) price inflation. Geoghegan *et al.* (1992) propose that the force (or logarithmic rate) of salary inflation net of contemporaneous price inflation be modelled as an independent and identically distributed (IID) normal variate. They also suggest an AR(1) process, instead of the IID component with the normal errors of the price inflation and real salary inflation series being independent. However, the possibility of autoregressive *real* salary inflation series on U.K. data is rejected by Wilkie (1995:§3). He rejects co-integration (i.e. stationarity in a linear combination of two series) in U.K. price and wage inflation series, although Sharp (1992) applies it to Canadian wage and price inflation data.

Wilkie (1995:§3) projects the force of salary inflation in various ways, more particularly as

- a proportion of current price inflation as well as an AR(1) component;
- proportions of current and lagged price inflation as well as an AR(1) component, allowing for the lagged effects of prices on wages;
- a vector autoregressive model with salary and price inflation depending on each other, rather than having a unidirectional 'cascade' structure.

Sherris (1995), using Australian data, also develops a vector autoregressive model of order 1 (VAR(1)) which comprises several factors including equity returns, interest rates as well as salary inflation.

2.6.3 Asset Returns

Pension Fund Constituent Assets

Pension funds invest in a diverse range of assets, including *equity* (or common stock, both overseas and local), *debt* (both government securities (conventional or index-linked) and corporate bonds), *property* (directly or through property trust funds), *cash* (usually held for liquidity purposes), as well as derivative instruments, commodities etc. The selection of a suitable portfolio of assets and securities is not directly addressed in this thesis, except in the theoretical context of dynamic portfolio allocation between a risky and a risk-less asset in §3.6. The issues of valuation and investment are related, particularly through the concept of hedging or matching. Certain issues relevant to asset investment are discussed below and in §4.3.

Asset Return–Inflation Linkage

The well-known Fisher hypothesis suggests that the nominal return on gilts, bonds and the like will reflect investors' expectations of price inflation, and also comprise a risk premium (practically nil for government securities) and pure interest. The risk premium and pure interest constitute real interest.

Price inflation also affects equity returns. Company dividend payouts depend upon their profits and turnover and hence upon general price levels. Dividends are therefore expected to grow at least in line with price inflation. Indeed, because of real productivity improvements, dividends ought to outpace inflation. This is illustrated by Daykin (1976:299, 1987) and Thornton & Wilson (1992a:239). In addition, we expect a time lag (also depicted by Daykin, 1976) between price inflation and its effect on company profits and dividend policy, just like that between price and salary inflation.

Equity returns also arise from capital growth as share prices increase. Share prices represent the present value of dividend income expected by the market, at the market discount rate (market's expectations of interest rates). Share prices are thus expected to increase in advance of an increase in dividends, which is generally well forecast by the market. Share prices are also expected to decrease (and dividend yields increase) if interest rates increase. Daykin's (1976) data shows a positive relationship between discount rates in the U.K. economy (through irredeemable gilt (Consol) yields) and equity dividend yields. Higher price inflation leads to higher interest rates, because of the Fisher relation, and lower share prices. It is generally observed that equity returns are not correlated, or even negatively correlated, with price inflation in the short term [Fama & Schwert (1977), Sharpe *et al.* (1995:374)]. There is also broad consensus that, over the long term, equities have returned much more than government securities (exhibiting a premium for the risk involved in holding equities) and also over and above inflation (indicating that equities are 'real' assets).

Equity dividend growth is also expected to be driven by productivity improvements, besides price inflation. Now, productivity gains are also said to contribute to salary inflation. This gives rise to the proposition that equities are (at least partially) a good match or hedge for salary inflation-related liabilities, such as those of a final-salary pension plan. There is strong acceptance of an economic link between salary inflation and dividend growth particularly among British actuaries. Daykin (1976:¶¶33, 34) and Thornton & Wilson (1992a:§5.2) claim that dividend growth is in line with lagged salary inflation. Wilkie (1995:§5) obtains strong evidence on U.K. data that wages and dividends are co-integrated: the difference between log dividend and log wages is stationary. This is usually justified by the theory that macro-economic Gross Domestic Product (GDP) growth must be shared by labour (in the form of wages) and capital (in the form of dividends) according to stable long-term proportions. Increases in productivity are therefore supposed to be reflected in a balanced way in both wage inflation and dividend growth. It is accordingly held that equities are a suitable investment to meet salary-related pension liabilities in the long term.

This is subject to considerable controversy, however, on the grounds that economic growth may arise from inward investment from foreign corporations for example. Exley *et al.* (1997) contest the link between dividend and salary growth for another reason. They examine the share of Gross Domestic Product between dividend and wages in the U.K. and conclude that small changes in wages, as a proportion of total wages, are balanced by much larger changes in dividends, as a proportion of total dividend, because dividends represent a much smaller share of national income than wages. Exley *et al.* (1997) favour the arguably more stable relationship between salary inflation and price inflation plus an estimated real salary inflation component. This argument has important consequences for the determination of the strategic investment policy for a final-salary plan (see §4.3.4) and specially for the valuation of salary-related liabilities (in the choice of valuation discount rate) (see §5.2.1).

Investment Return and Discount Rate Valuation Assumptions

The custom in many pension fund valuations is not to differentiate between the discount rate and investment return assumptions. A single assumption is made covering both and it is often referred to as a 'valuation rate of interest'. I avoid use of this term as it is confusing on several counts, one of which is that it is only indirectly related to economic interest rates, and another being that it does not distinguish between the discount rate and the investment return assumptions. Strictly, the discount rate refers to the rate used to value liabilities, whereas the investment return assumption is a 'best-estimate' assumption as to the long-term return on current and future pension fund assets. A major portion of this thesis (Chapter 5) is concerned with the distinction, conceptual and numerical, between them. Actuarial Standard of Practice No. 27 of the American Academy of Actuaries (1996:§3.6) also distinguishes between the discount rate and the investment return assumption. The two assumptions relate to the return on different portfolios of assets: a notional or hypothetical portfolio in the case of the valuation discount rate, and the actual and anticipated future portfolio of assets held by the fund. The valuation discount rate also incorporates various risk-adjustments. This is discussed further in Chapter 5.

At this stage, we note that the investment return for any portfolio of assets comprises income from various asset classes as well as capital growth upon trading of these assets. In the U.K., the investment return assumption is usually subdivided into a dividend yield (representing income) and a dividend growth (representing capital growth). The dividend yield is also relevant to the Discounted Income Value placed on assets (see §4.4.1). We also note that any estimation of investment return must be consistent with the price and salary inflation assumptions. Liability values are particularly sensitive to their relative levels as discussed above (§§2.6.1, 2.6.2). The value placed on nominally-fixed liabilities will depend on the absolute discount rate assumption, however. Pension liabilities will be *conservatively* estimated (i.e. overestimated) if the discount rate is low.

Asset Return Projection Models

The traditional economic model for stochastic security price behaviour rests on the Efficient Market Hypothesis (EMH). Market efficiency is defined in various ways, but essentially means that all information is reflected in security prices on a market so that it is impossible to predict future price movements and make abnormal profits. It follows from this that security prices are martingales: changes in prices must be statistically independent over time. Much evidence has been presented, in the financial economics literature, that a random walk model can be fitted to security prices, i.e. that the change in share prices is independent and identically distributed. Log-normal distributions are usually chosen because they are mathematically tractable and easily estimated. For example, Godolphin [Ford *et al.* (1980): Appendix C] fits a Gaussian random walk to the U.S. Standard & Poor Composite rolled-up (i.e. with dividend reinvested) index. Ibbotson & Sinquefield (1993) also examine in the form of random walks the rolled-up returns ("cumulative wealth ratios") on a wide variety of U.S. financial data, including common stock and long-dated bonds. Log-normal distributions are also commonly assumed in Asset-Liability modelling or stochastic forecasting exercises for pension funds [McGill *et al.* (1996:575), Kemp (1996)]. Other distributions have been proposed, including Gamma distributions (Smith, 1996) and the general family of stable distributions (Finkelstein, 1997) particularly because empirical distributions of share returns appear to be more 'fat-tailed' than log-normal distributions.

The random walk model has been used in the modelling of pension funds particularly because it is a tractable model for describing the investment of total fund assets. Both Wise (1984) and O'Brien (1986) use a Gaussian random walk in simplified mathematical models for pension fund investment. Knox (1993) assumes that the rate of return net of *price* inflation is independent and identically distributed over time. Pension fund investment return, net of *salary* inflation, is assumed to be independent and identically distributed over time by Dufresne (1986, 1988, 1989), Cairns (1994), Haberman (1992a, 1993, 1995) and Sung (1997), among others. In Asset-Liability modelling, Kemp (1996) also favours independent and identically distributed rates of return on various (contemporaneously correlated) asset classes, including equities and real-return as well as conventional U.K. government bonds (gilts), mainly on account of market efficiency.

Some evidence has accumulated more recently regarding the inappropriateness of random walks over long periods. Fama & French (1988) report that negative autocorrelation can be detected in returns over long periods, induced by some mean reversion in stock prices. Panjer & Bellhouse (1980) find that a number of financial variables, from Moody's and Standard & Poor's data on interest rates and equity dividend yields in the U.S., may be modelled as stationary autoregressive processes. Godolphin [Ford *et al.* (1980): Appendix C] fits both a stationary autoregressive process of order 7 (AR(7)) and a moving average process of order 2 (MA(2)) to rolled-up returns from the U.K. de Zoete (now BZW) equity index. Vector autoregressive models (VAR(1)) incorporating various economic variables are also fitted by Sherris (1995) and Kleynen (1997) and used in pension fund modelling. Frees (1990) also models the log annual returns of the Salomon Brothers Bond Index as both MA(1) and AR(1). Such auto-correlated models run counter to market efficiency arguments. Frees (1990) justifies consideration of such models because

> "when examining the microstructure of investments, returns will follow a martingale *plus* some corrupting influences. It is posited that the corrupting influences account for the observed autocorrelations of returns."

Wilkie's (1987, 1995) model is perhaps the most commonly used in stochastic pension fund projections in the U.K. It consists essentially of autoregressive time series. Variables are ordered into a cascade structure and defined in terms of lagged values of themselves or a variable of lesser order. The residuals are independent and identically normally distributed. Ordinary Least Squares estimates are obtained for the parameters. Wilkie (1995) also uses vector autoregression (VAR) and co-integration. One of the discerning features of Wilkie's (1987, 1995) model is that equity dividend yields and dividend growth are modelled separately, in contrast to most econometric and earlier actuarial models. This is reasonable since equity investment return consists of the two distinct processes of capital growth and dividend income.

Wilkie's (1987) log dividend yield model consists of an AR(1) component, so that there is dependence on past dividend yields, and a proportion of the current force of price inflation, with an increase in price inflation (leading to higher interest rates) causing the share yield to increase as well. His log dividend growth model comprises three components: a moving average MA(1) component (share dividends are influenced by past dividends); a proportion of current price inflation and an exponentially weighted average of past and current price inflation (an increase in price inflation leads to an equal percentage growth in the share dividend gradually, with the dividend being sensitive to current as well as past price inflation); and a proportion of the lagged residual from the share yield series (increases in share dividends are usually forecast by the market so that they are preceded by an increase in share price and hence the dividend yield decreases). Wilkie (1987, 1995) also models long-term interest rates by using yields on a British irredeemable gilt (Consols 2.5%). His model of long-term interest rates consists of: an exponentially weighted moving average of past and current price inflation, which stands for market expectations of price inflation which, by the Fisher relation, are reflected in gilt yields; a stationary AR(3) (simplified to an AR(1)) component, so that there is dependence on past interest rates; a proportion of the residual from the share yield series, describing the dependence of the equity dividend yields on market discount rates, which depend on investors' borrowing rates (short-term rates) which are themselves reflected in long-term interest rates. Various other asset classes, including short-term gilts and property which are relevant to pension fund investment, are considered by Wilkie (1995).

There are certain problems associated with using Wilkie's (1995) model in pension fund projections:

1. The model comprises scores of parameters and is not parsimonious [Kemp (1996:¶5.6), Huber (1996)] which renders its application, especially in mathematical models, difficult.

2. Economic theory suggests that there may be statistical dependence between dividend growth and salary inflation through improvements in productivity. No measure of productivity, which influences both salary inflation and dividend growth, has been included.

3. Wilkie's (1995) model is meant to give a reasonable fit to the data (while not being devoid of economic content). Huber (1996, 1997) nevertheless criticises the statistical fit of Wilkie's (1995) model to actual data and argues that the dividend growth and long-term interest rate models appear to be over-parameterised.

4. The model seeks to retain a degree of economic realism (and thereby does not fully adhere to statistical considerations of investment data). However, it conforms neither to the efficient market hypothesis, nor to the rational expectations hypothesis, nor to aspects of portfolio theory [Huber (1996:76–81)].

5. The model is concerned with investment performance over the long term. Although short-runs of the model are very similar to random walks [Wilkie (1987, 1995:§§4, 5)], in the longer term mean-reversion means that the model is inconsistent with market efficiency. It is theoretically possible, by virtue of mean reversion, to make abnormally large profits without much risk and dynamic investment strategies cannot be devised (Kemp, 1996).

2.7 Valuation and Projection Assumptions in Our Model

A number of simplifying assumptions are made in the mathematical model used subsequently in this thesis. Many of these assumptions are either implicit or are explicitly stated in the models of Trowbridge (1952), Bowers *et al.* (1976, 1979), Dufresne (1986) and Benjamin (1989).

Modelling Assumptions. A set of simple *modelling* assumptions are required first and foremost.

MODELLING ASSUMPTION 2.1 (INCIDENCE OF CASH FLOWS)
Contributions $c(t)$ are paid into the fund and benefits $B(t)$ are paid out at the start of year $(t, t+1)$. The rate of return (net of salary inflation) during the year, based on market values of assets, is $i(t+1)$.

Modelling Assumption 2.1 concerns the timing of cash flows in and out of the pension fund: discrete-time processes are involved.

Modelling Assumption 2.2 (Initialisation)
The initial fund level is known with certainty, $f(0) = f_0$ *w.p. 1.*

Modelling Assumption 2.3 (Invariance of valuation basis and methods)
All actuarial valuation procedure and assumptions (including the valuation discount rate) are assumed fixed in time.

The assumption of a static valuation method and basis as in Modelling Assumption 2.3 is strictly maintained. Actuarial valuation assumptions change slowly in practice.

Modelling Assumption 2.4 ('Real' monetary quantities)
All economic and financial quantities, including benefits, salaries, asset values, actuarial present values of benefits, rates of return and discount rates, are assumed to be net *of general economic salary inflation.*

In the rare occasions where the above assumption is relaxed, it will be signalled that 'nominal' quantities are being considered.

Modelling Assumption 2.5 (Intervaluation period)
Valuations are effected at regular time intervals of one time unit.

This assumption is enforced throughout.

Modelling Assumption 2.6 (Pension funding method)
The pension funding method or actuarial cost method employed is a consistent method in that it does not generate actuarial gains or losses when all assumptions are exactly realised. It is also an 'individual' method in that actuarial present values for the plan are the sum total of actuarial present values for individual plan participants.

Pension funding methods, including 'aggregate' ones, are discussed in Chapter 3.

Modelling Assumption 2.7 (Asset valuation)
The assets of the pension fund are readily marketable and their market value is $f(t)$.

The use of an actuarial value $F(t)$ is considered in Chapter 4.

Plan Assumptions. A simple model pension plan is hypothesised.

Plan Assumption 2.1 (Decrements)
The plan has a single entry age a *and a single retirement age* r*. There are no decrements other than retirement (at age* r*) and mortality.*

All pensioners are credited with $r - a$ service years.

PLAN ASSUMPTION 2.2 (SALARY)
All active members receive the same salary increasing in accordance with a promotional salary scale and general economic inflation.

See Valuation Assumption 2.3 below concerning the salary scale.

PLAN ASSUMPTION 2.3 (BENEFITS)
Only a retirement benefit, effectively a whole-life annuity, is paid. Benefits accrue at an accrual rate b, such that a pension equal to fraction b of final salary is paid for each year of service.

PLAN ASSUMPTION 2.4 (BENEFIT INDEXATION)
Benefits in payment are indexed with economic wage inflation.

Partial indexation of benefits with *price* inflation is a statutory requirement in a few jurisdictions and a commitment to complete indexation is sometimes made. This is likely to become more common with real-return securities being issued and traded in several countries. Political arguments are made regarding the indexation of certain *state* pensions with wages. Plan Assumption 2.4 is therefore unusual but theoretically plausible and mathematically convenient. A consequence of the provision of final-salary wage-indexed benefits (Plan Assumption 2.3) and the measurement of 'deflated' quantities (Modelling Assumption 2.4) is that inflation may be ignored [Dufresne (1986), Haberman (1994b)].

PLAN ASSUMPTION 2.5 (STRICTLY DEFINED BENEFITS)
Plan benefit rules do not change and no discretionary or ad hoc *benefit improvement is allowed (except for indexation).*

Plan Assumption 2.5 implies that a strictly defined benefit pension plan is posited. This is not always the case in practice. The option-like feature of an enhancement to plan members' benefits is disregarded. Control through 'benefit policy' is therefore disallowed (§2.3.2).

Valuation Assumptions. The following valuation assumptions are made.

VALUATION ASSUMPTION 2.1 (INFLATION)
Inflation on benefits is the same as inflation on salaries. No valuation assumption as to economic price inflation is necessary. No valuation assumption as to absolute level of wage inflation is necessary.

This is a consequence of Plan Assumptions 2.3 and 2.4. The relative levels of investment return and wage inflation, rather than the absolute level of wage inflation, is material.

VALUATION ASSUMPTION 2.2 (VALUATION DISCOUNT RATE)
The valuation discount rate net of salary inflation (i_v) is equal to the long term mean rate of return net of salary inflation, i.e. $i_v = \mathrm{E}i(t) = i$.

The assumption that the valuation discount rate equals the rate of return on the fund is discussed in Chapter 5.

VALUATION ASSUMPTION 2.3 (MORTALITY)
Mortality is assumed to be contingent as per a life table $\{a \leq x \leq w : l_x\}$ *that incorporates a salary scale (i.e. if* l'_x *is the standard survival function and* s_x *represents a salary scale, then* $l_x = s_x l'_x$*).*

As a result of plan assumption 2.1, no valuation assumption regarding withdrawal rates, disability, early retirement etc. is required.

Projection Assumptions. The actual experience will generally differ from the actuarial assumptions chosen at valuation. We simplify considerably by establishing the following projection assumptions for future experience.

PROJECTION ASSUMPTION 2.1 (EXPERIENCE)
Actual experience is in accordance with actuarial valuation assumptions except for the rate of investment return (net of salary inflation).

The only source of unexpected experience and hence of actuarial gains and losses is through pension fund investment. The fact that investment return is the most important factor affecting pension funding is well-known and documented [Thornton & Wilson (1992a), Winklevoss (1993:213), McGill *et al.* (1996:574)] and justifies Projection Assumption 2.1.

In order to look at the variation in membership size and age distribution, a population distribution function $g(t)$ as defined hereunder is needed. It is the discrete-time equivalent of the population density function $g_1(t)$ of Bowers *et al.* (1976). The population distribution function $g(t)$ is defined such that the size of the membership aged x at time t is $g(t+r-x)l_x$. r is the retirement age and the number of members retiring (aged r) at time t is $g(t)l_r$.

PROJECTION ASSUMPTION 2.2 (MEMBERSHIP)
The membership is stationary (deterministic) from the start such that the population distribution function $g(t)$ *is constant* $\forall t$*.*

This assumption is modified in Chapter 3.

Finally, an assumption as to investment return is required.

PROJECTION ASSUMPTION 2.3 (INVESTMENT RATE OF RETURN)
The real rate of return $i(t)$ *is a sequence of independent and identically distributed random variables.* $\mathrm{E}i(t) = i > -100\%$. $\mathrm{Var}i(t) = \sigma^2 < \infty$.

Rates of return are also assumed to be serially dependent when simple mean-reverting rates of return are studied in Chapter 3. Projection Assumption 2.3 may be justified on four levels:

1. It accords with the Efficient Market Hypothesis (§2.6.3).

2. Smith (1996:§3) notes that it is prudent not to model market inefficiencies because if these inefficiencies are absent, or change or disappear over time, then incorrect decisions may be made with potentially damaging consequences.

3. The aim of projecting the state of the pension fund stochastically is not necessarily to forecast the finances of the pension plan, but instead to test and investigate the behaviour of the pension fund system as a consequence of economic variability and investment in volatile capital markets. Projection Assumption 2.3 is a simple way of introducing volatility in actuarial models. Frees (1990) views the assumption of independent and identically distributed rates of return as

   > "a useful modification of the traditional deterministic [rate of interest]. This modification permits volatility of interest rates in the model."

4. It is a simple assumption that allows for mathematical tractability, parsimony and a search for optimality and robustness (§1.2) in the pension fund model. Haberman & Wong (1997) note that

   > "In reality the rate of investment return is not deterministic, nor does it follow any known stochastic model so that future returns can only be partially predicted. [...] The representation of these economic variables is a difficult problem for the actuary to determine on a regular basis."

   Simple models are acceptable because, according to Huber (1996:49),

   > "As all models are controversial, an alternative approach is to select the most general mathematically tractable model that is broadly consistent with financial economic theory. Until an empirically adequate and theoretically consistent model is discovered, these hypothetical models are often the most pragmatic alternative."

2.8 Summary

This section summarises some of the major points made in this chapter. The nature of defined-benefit pension funding is explored. A pension fund is viewed as a financial system that is set up to achieve a range of objectives. Some of these objectives are enforced by statute or regulations. Two long-term objectives stand out and are associated with the motivation for advance funding of pension benefits: benefits must be secured through the accumulation of assets, and the contributions required from sponsors must be stable and predictable. Furthermore, limited flexibility in the timing of these contributions is advantageous.

Regular actuarial valuations are likened to exercises in the control of pension funds to achieve these objectives. Actuaries exercise control by determining a funding policy through the choice of suitable funding methods, valuation bases, valuation methods and contributions as well as by influencing strategic asset allocation. This control is necessary to mitigate the effects of the uncertain economic and demographic environment in which pension plans are set up. Projections or forecasts for pension planning are based on basic economic and demographic theory, some elements of which are discussed in §§2.5 and 2.6. Economic price and wage inflation are particularly relevant in final-salary plans, whether benefits are indexed with inflation or are enhanced on a discretionary basis. The relationship between returns on several classes of assets and (price and wage) inflation is crucial and is also discussed.

Finally, a simplified pension fund model is set up to explore various aspects of the dynamics and actuarial control of pension funds. The model is justified by the Efficient Market Hypothesis, but is primarily designed to achieve parsimony and mathematical tractability.

Chapter 3

Methods of Funding

3.1 Funding Methods

3.1.1 Pension Funding Methods

A "pension funding method" or "actuarial cost method" is a systematic way of accumulating funds to provide retirement benefits. It is essentially a plan for the orderly substitution of contribution by investment income [Trowbridge & Farr (1976:19)]. Trowbridge (1952) also defines a pension funding method very generally as "the budgeting scheme or payment plan under which the benefits are to be financed." Pension funding methods determine the extent to which investment income is required as against contributions, i.e. the degree of advance funding sought. A pension funding method defines a subset of the pension liabilities for which assets should be held. Investment return from these assets together with contributions go to meet the liabilities as and when they are due.

An infinite number of funding methods may be devised but the patterns of contributions and funding levels they generate must be consistent. The actuarial present value of any portion of the pension liabilities may be calculated: this is the discounted present value of the liability cash flow stream, allowing for mortality and other contingencies. At any point in time, over the set of all plan members (i.e. assuming a 'closed group'), the actuarial present value of future contributions together with the fund already built up should equal the actuarial present value of future benefits (as well as expenses). In addition, for any 'individual' funding method (see below), this equation of value, or *actuarial equivalence principle*, should hold for every single plan member.

The classic papers of Trowbridge (1952) and Seal (1952) include classifications and discussions of many of the actuarial cost methods allowed by the U.S. Internal Revenue Service and in use by U.S. pension actuaries at the time. Since then, pension funding methods have been classified and codified several times in several places. For the terminology promoted by the professional bodies, see Actuarial Standard of Practice No. 4 of the American Academy of

Actuaries (1993) and Guidance Note 26 of the Manual of Actuarial Practice of the Institute and Faculty of Actuaries (1997b). McGill *et al.* (1996) have devised a different terminology.

Pension funding methods are also described and discussed by several authors, notably Paquin (1975), Trowbridge & Farr (1976), Bowers *et al.* (1976), Dufresne (1986, 1994), Berin (1989), Anderson (1992), Winklevoss (1993) and Aitken (1994). Colbran (1982), Turner *et al.* (1984) and O'Regan & Weeder (1990) take a U.K. perspective on the subject. Accordingly, I will not describe the various pension funding methods here, but I will emphasise certain relevant overriding features. The classification of pension funding methods in terms of these characteristics is principally due to McGill (1964) [see also McGill *et al.* (1996) and Winklevoss (1993)].

Projected v. Accrued Benefit Methods. The retrospective, or accrued benefit, or benefit allocation, or fund approach first defines the liabilities that should be funded in terms of the benefits (based on projected final salary or current salary) that have *accrued* to members in respect of past service. This defines an actuarial liability or standard fund as a *retrospective* reserve. The normal cost or standard contribution for a given year is then found by means of the *actuarial equivalence principle* as the actuarial present value of future benefits less this reserve, which turns out to be the actuarial present value of benefits earned during the year. Benefit is therefore assigned or *allocated* to each service year. Thus, "an accrued benefit cost method is a method which endeavours systematically to match pension costs [contributions] with the year in which each pension benefit is presumed earned or in which it 'accrues' " (Paquin, 1975). Examples of such funding methods are the *Projected* or *Current Unit Credit* methods. The *Unit Credit* family of methods is described in §3.1.2.

Alternatively, a prospective, or projected benefit, or cost allocation, or contribution approach may be taken. This requires a definition of standard contribution or normal cost based on the benefits which members are *projected* to receive. A level standard contribution (typically as a percentage of payroll) may be determined for each year of service. 'Cost' (or contribution) is therefore *allocated* to each service year. The actuarial liability is then the *prospective* reserve equalling the difference between the actuarial present value of future benefits and the actuarial present value of future standard contributions. Examples of such funding methods are the *Aggregate* and *Entry Age Normal* methods (see §3.1.2).

Individual v. Aggregate Approach. Some pension funding methods compute an actuarial liability and a normal cost for each individual plan participant and they are then added up to give totals for the plan membership. Such 'individual' methods may follow either the accrued benefit or projected benefit approach. Examples are the *Entry Age Normal* and the *Unit Credit* methods. Other methods do not ascribe the normal cost and actuarial liability to individual plan participants and are said to follow an 'aggregate' approach. They follow the 'projected benefit' approach. Examples are the *Aggregate* or the *Frozen Initial Liability* methods.

Actuarial Gains/Losses and Supplementary Contributions. The demographic and financial assumptions that are made to calculate actuarial present values are unlikely to be borne out in reality. Gains or losses will emerge at successive valuations and will need to be adjusted for. In addition, there may be initial unfunded liabilities to be amortized. The 'aggregate' methods adjust contributions implicitly as surpluses or deficits emerge. They are also called 'spread gain' cost methods. The 'individual' methods require separate, explicit adjustment to the contribution as gains/losses or surpluses/deficits arise. They are known as 'immediate gain' cost methods (Berin, 1989). Methods of calculating supplementary contributions to adjust explicitly for gains/losses are a central topic of this chapter.

Initial Unfunded Liability. Some actuarial cost methods distinguish initial unfunded liabilities from surpluses/deficits that arise from unpredicted experience and deal with them separately (e.g. the *Frozen Initial Liability* method), whereas others do not (e.g. the *Aggregate* method). This point is stressed by McGill *et al.* (1996) who classify pension funding methods according to whether they are 'with supplemental liability' (if an explicit adjustment for initial unfunded liabilities as well as experience surpluses/deficits is made) or 'without supplemental liability' (if the adjustment for initial unfunded liabilities as well as experience surpluses/deficits is implicit).

Internal Consistency. Any funding method that does not satisfy the *actuarial equivalence principle* is inconsistent in that the contributions it requires are not compatible with the level of funding sought. Actuarial gains/losses continually emerge and the pension fund is not in financial balance and exhibits a persisting surplus or deficit. The U.S. Internal Revenue Service requires the use of a "reasonable funding method" which it defines as a method that generates no gain or loss and stable contributions if actuarial assumptions turn out to be correct. Any such method is said to satisfy a "zero-gain criterion" by Sharp (1996), and to be "conditionally consistent" and "well-defined" by Shapiro (1983).

3.1.2 Some Pension Funding Methods

Precise definitions of various funding methods are given in Actuarial Standard of Practice No. 4 of the American Academy of Actuaries (1993) and Guidance Note 26 of the Manual of Actuarial Practice of the Institute and Faculty of Actuaries (1997b).

Model Plan

The operation of most pension funding methods can be effectively illustrated using a simple model. The simple model of §2.7 is assumed, with the exception of Projection Assumption 2.2:

suppose that the plan membership aged x at time t is $g(t+r-x)l_x$, where $g(t+r-x)$ may be time-variant. Recall from Plan Assumption 2.1 and Valuation Assumption 2.3 that all plan members enter at age a, retire at age r with the survivorship function $\{l_x\}$ defined according to an actuarial life table incorporating a salary scale (the life table terminates at age w). A pension equal to a fraction b of final salary is paid for each year of service to plan members when they reach age r. We also make a somewhat sweeping assumption concerning salary inflation by operating under Plan Assumptions 2.3 and 2.4. There is no difficulty in introducing inflation when describing pension funding methods, as Dufresne (1986, 1994) and Winklevoss (1993) among others show. All monetary terms hereunder are assumed to be net of salary inflation (Modelling Assumption 2.4). Finally, assume that all working members of the pension plan earn a salary s (net of salary inflation) (Plan Assumption 2.2).

The benefit received by a member aged x is

$$B_x = \begin{cases} 0 & a \leq x \leq r-1 \\ sb(r-a) & x \geq r. \end{cases} \tag{3.1}$$

Total benefits paid out to retired members in year $(t, t+1)$ equal

$$B(t) = \sum_{x=r}^{w} g(t+r-x) l_x sb(r-a). \tag{3.2}$$

The actuarial present value of the benefit that an *active* member aged $x < r$ who joined the plan at age a is *projected* to receive is $sb(r-a)_{r-x|}\ddot{a}_x$. The actuarial present value of the benefit he has *accrued* is $sb(x-a)_{r-x|}\ddot{a}_x$. The actuarial present value of the benefit for a *retired* member aged $x \geq r$ with exactly $r-a$ years of service is $sb(r-a)\ddot{a}_x$. The actuarial present value of benefit outgo from year t onwards for the current plan membership is

$$PVB(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x sb(r-a)_{r-x|}\ddot{a}_x + \sum_{x=r}^{w} g(t+r-x) l_x sb(r-a)\ddot{a}_x \tag{3.3}$$

Unit Credit Methods

The *Unit Credit* family of methods is the most common in use. A mathematical demonstration of the method is given by Trowbridge (1952), Bowers *et al.* (1976) and Dufresne (1986, 1994). The difference between the *Projected* and *Current Unit Credit* methods lies in the treatment of salary inflation, which is 'assumed away' in the simplified model plan.

The actuarial liability for a member aged x is defined under the *Unit Credit* pension funding method as the actuarial present value of the benefit he has accrued to date. Therefore,

$$AL_x^{UC} = \begin{cases} sb(x-a)_{r-x|}\ddot{a}_x, & a \leq x \leq r-1, \\ sb(r-a)\ddot{a}_x, & x \geq r. \end{cases} \tag{3.4}$$

The *actuarial equivalence principle* applies to each plan participant for an 'individual' funding method, i.e. the actuarial present value of benefits less the actuarial present value of future normal contributions must equal the reserve. For any plan member aged x,

$$\sum_{y=x}^{w} \frac{l_y}{l_x} v^{y-x} NC_y + AL_x = \begin{cases} sb(r-a)_{r-x|}\ddot{a}_x, & a \leq x \leq r-1, \\ sb(r-a)\ddot{a}_x, & x \geq r. \end{cases} \tag{3.5}$$

Contributions are not paid for retired members and for a plan member aged $x \geq r$ the normal cost is clearly zero, since $AL_x = sb(r-a)\ddot{a}_x$. For a plan member aged $a \leq x \leq r-1$, it is easy to show that equation (3.5) with $AL_x = sb(x-a)_{r-x|}\ddot{a}_x$ takes $NC_x = sb_{r-x|}\ddot{a}_x$ as solution (noting that ${}_{r-x-1|}\ddot{a}_{x-1} v l_{x+1}/l_x = {}_{r-x|}\ddot{a}_x$). The normal cost for an individual aged x is therefore equal to the actuarial present value of the benefits he accrues during the year:

$$NC_x^{UC} = \begin{cases} sb_{r-x|}\ddot{a}_x, & a \leq x \leq r-1, \\ 0, & x \geq r. \end{cases} \tag{3.6}$$

The actuarial liability and normal cost for the plan at time t is (by summing separately over the set of active and retired members):

$$AL^{UC}(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x sb(x-a)_{r-x|}\ddot{a}_x + \sum_{x=r}^{w} g(t+r-x) l_x sb(r-a)\ddot{a}_x, \tag{3.7}$$

$$NC^{UC}(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x sb_{r-x|}\ddot{a}_x. \tag{3.8}$$

***Entry Age Normal* Method**

This is a projected benefit or cost allocation method. The aim of this method is to generate a stable contribution for the duration of the working lifetime of a plan member. The normal cost or standard contribution for an active plan participant aged $a \leq x \leq r-1$ is such that

$$NC_x^{EA} \times \ddot{a}_{a:\overline{r-a|}} = sb(r-a)_{r-a|}\ddot{a}_a \tag{3.9}$$

and for a retired plan member it is zero.

It is easy to show that the actuarial liability for an active member aged $a \leq x \leq r-1$, from the actuarial equivalence principle in equation (3.5), is a prospective reserve

$$\begin{aligned} AL_x^{EA} &= sb(r-a)_{r-x|}\ddot{a}_x \; - \; sb(r-a)\ {}_{r-a|}\ddot{a}_a\ \ddot{a}_{x:\overline{r-x|}}\ /\ \ddot{a}_{a:\overline{r-a|}} \\ &= sb(r-a)\ {}_{r-x|}\ddot{a}_x\ \ddot{a}_{a:\overline{x-a|}}\ /\ \ddot{a}_{a:\overline{r-a|}} \end{aligned} \tag{3.10}$$

since ${}_{r-a|}\ddot{a}_a / {}_{r-x|}\ddot{a}_x = v^{x-a} l_r / l_x$ and $\ddot{a}_{a:\overline{r-a|}} = \ddot{a}_{a:\overline{x-a|}} + \ddot{a}_{x:\overline{r-x|}} v^{x-a} l_r / l_x$. For a retired member aged $x \geq r$ the actuarial liability is

$$AL_x^{EA} = sb(r-a)\ddot{a}_x. \tag{3.11}$$

The actuarial liability and normal cost for the plan at time t is thus (by summing separately over the set of active and retired members):

$$AL^{EA}(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x s b(r-a)\ {}_{r-x|}\ddot{a}_x\ \ddot{a}_{a:\overline{x-a|}}\ /\ \ddot{a}_{a:\overline{r-a|}} + \sum_{x=r}^{w} g(t+r-x) l_x s b(r-a)\ \ddot{a}_x, \tag{3.12}$$

$$NC^{EA}(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x s b(r-a)\ {}_{r-a|}\ddot{a}_a / \ddot{a}_{a:\overline{r-a|}}. \tag{3.13}$$

Aggregate Method

This method operates in the 'aggregate', i.e. over the total plan membership. No actuarial liability and normal cost is defined and assigned to individual plan members. A stable contribution rate (contribution per payroll) is sought. If the actuarial equivalence principle is to hold, then the actuarial present value of future contributions, obtained by applying the constant contribution rate to the total future payroll for all members, must equal the actuarial present value of benefits less the fund already accumulated.

In the simple model considered here, payroll in year t is

$$S(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x s \tag{3.14}$$

and the actuarial present value of payroll from year t onwards is

$$PVS(t) = \sum_{x=a}^{r-1} g(t+r-x) l_x s\ \ddot{a}_{x:\overline{r-x|}}. \tag{3.15}$$

($S(t)$ and $PVS(t)$ are net of salary inflation in this simple model.) The total contribution $c(t)$ required at the beginning of year $(t,\ t+1)$ under the *Aggregate* method is given by

$$c(t) PVS(t)/S(t) = PVB(t) - f(t). \tag{3.16}$$

Frozen Initial Liability Methods

This method is also an 'aggregate' one and does not require the development of an actuarial liability or of a normal cost. It involves a modification of the previous method. Initial unfunded liabilities arise for different reasons than experience surpluses and deficits and may be quite large. It is sensible to amortize them over different periods. The initial unfunded liability may then be 'frozen' and dealt with separately: it is referred to as an 'initial unfunded frozen actuarial liability'. The unfunded frozen actuarial liability may be calculated in the same way as under the *Entry Age* method (in which case the method is known as the *Frozen*

Entry Age method) or as under the *Unit Credit* method (in which case the method is known as the *Frozen Attained Age* method).

The amortization of the frozen unfunded liability ul_0 may be effected through a separate payment $P(t)$:

$$P(t) = \begin{cases} ul_0/\ddot{a}_{\overline{n}|}, & 0 \leq t \leq n-1, \\ 0, & t \geq n. \end{cases} \tag{3.17}$$

In year $(t, t+1)$, the unfunded frozen liability (i.e. the unamortized part of the initial unfunded frozen liability) is:

$$U(t) = \begin{cases} ul_0 \; \ddot{a}_{\overline{n-t}|}/\ddot{a}_{\overline{n}|}, & 0 \leq t \leq n-1, \\ 0, & t \geq n. \end{cases} \tag{3.18}$$

The objective of the *Frozen Initial Liability* methods is that the contributions required in addition to $P(t)$ be a stable proportion of payroll every year. The *actuarial equivalence principle* therefore dictates that the fund at hand together with the actuarial present value of future contributions in excess of the frozen liability payments be equal to the actuarial present value of benefits less the frozen unfunded liability:

$$[c(t) - P(t)]PVS(t)/S(t) = PVB(t) - U(t) - f(t). \tag{3.19}$$

Equilibrium

All consistent funding methods abide by the *actuarial equivalence principle.* This applies at the level of the individual plan member in 'individual' methods (equation (3.5)). It also applies over the total plan membership, since

$$AL(t) = \sum_{x=a}^{w} g(t+r-x) l_x AL_x, \tag{3.20}$$

$$NC(t) = \sum_{x=a}^{w} g(t+r-x) l_x NC_x, \tag{3.21}$$

$$B(t) = \sum_{x=a}^{w} g(t+r-x) l_x B_x. \tag{3.22}$$

In the simple model assumed above, it follows that

$$AL(t) = \sum_{s=t}^{\infty} (1+i)^{t-s} [B(s) - NC(s)] \tag{3.23}$$

for a plan that never terminates and whose benefits are valued at the valuation discount rate $i(> 0)$. Equation (3.23) may be written as

$$AL(t+1) = (1+i)[AL(t) + NC(t) - B(t)]. \tag{3.24}$$

Bowers *et al.* (1976) refer to this as the "liability growth equation". It may be shown that it holds for 'individual' funding methods such as the *Unit Credit* and *Entry Age Normal* methods [Bowers *et al.* (1976), Dufresne (1986)].

Consider the pension fund model of §2.7: the idealised situation of a stationary pension plan population, as under Projection Assumption 2.2, is now asserted. The size of the membership at any age x remains constant over time. Under the individual funding methods, the actuarial liability and normal cost over the total plan population are now constant, say AL and NC respectively. (Replace $g(t+r-x) = g$ (say) $\forall t$ in equations (3.7) and (3.8) or (3.12) and (3.13) for example.) Equation (3.23) or (3.24) may then be written as

$$AL = u(AL + NC - B), \tag{3.25}$$

where $u = 1 + i$, or as

$$B = dAL + NC, \tag{3.26}$$

where $d = i/(1+i)$. These equations denote a state of financial equilibrium in the pension fund: contribution plus the present value of the return on the reserve exactly meet the benefit outgo. Trowbridge (1952) describes equation (3.25) as an "equation of maturity": both the plan membership and the pension fund have grown or matured and become stationary.

3.1.3 Supplementary Funding Methods

'Aggregate' methods calculate the contribution (or contribution rate) for the pension fund directly based on benefits and assets held. The definition of an 'individual' funding method does not refer to the contributions and fund levels but instead incorporates consistent definitions of actuarial liability (or standard fund) and normal cost (or standard contribution). These two concepts are broadly analogous to the concepts of 'actuarial reserve' and 'premium' in life insurance mathematics. Descriptions of 'individual' funding methods are incomplete without mention of supplementary methods that relate the normal cost and actuarial liability to the actual contribution and fund levels.

Consider the model described in §2.7, except that the pension plan membership is allowed to be time-variant (Projection Assumption 2.2 does not apply). The fund level at the end of year $(t,\, t+1)$ is related to the fund level, contribution income and benefit outgo at the start of the year by

$$f(t+1) = u(t+1)[f(t) + c(t) - B(t)], \tag{3.27}$$

where $u(t+1) = 1 + i(t+1)$, where $i(t+1)$ is the (arithmetic) rate of return earned during the year.

The normal cost can be viewed as the "regular contribution" and the actuarial liability as the "theoretical level of assets" [Trowbridge & Farr (1976:22–24)] required under a pension funding method if all actuarial valuation assumptions are borne out exactly by experience. Since assumptions are not likely to hold exactly, regular valuations are required to determine the deviations in experience and to adjust contributions accordingly. The normal cost is therefore adjusted by an amount $adj(t)$, a function of the experience deviations, to give the contribution

$$c(t) = NC(t) + adj(t). \tag{3.28}$$

The experience deviations give rise to an unfunded liability or actuarial deficit $ul(t)$:

$$ul(t) = AL(t) - f(t), \tag{3.29}$$

a surplus being just a negative deficit. Recurrence relation (3.27) may be written in terms of $ul(t)$ and $adj(t)$:

$$ul(t+1) = AL(t+1) + u(t+1)[ul(t) - adj(t) - AL(t) - NC(t) + B(t)]. \tag{3.30}$$

Actuarial losses or gains emerge as experience differs from actuarial valuation assumptions. McGill *et al.* (1996:522) refers to *experience actuarial gains and losses*:

> "Only by coincidence will the actual experience of the plan conform to the assumptions that underlie the actuarial cost estimates. If the experience of the plan is financially more favorable than the underlying assumptions, *actuarial gains* emerge. If the experience is financially less favorable than that assumed, *actuarial losses* emerge."

The loss (or negative gain) $l(t)$ experienced during year $(t-1, t)$ is the difference between $ul(t)$ and the unfunded liability ($ul^A(t)$) *had all actuarial assumptions held true.* Under Projection Assumption 2.1, all valuation assumptions are borne out by experience, *except* investment returns. $l(t)$ is therefore termed an *asset loss.*

$$ul^A(t) = AL(t-1) + u[ul(t-1) - adj(t-1) - AL(t-1) - NC(t-1) + B(t-1)], \tag{3.31}$$

where the valuation discount rate $i_v = \text{E}i(t) = i$ (and $u = 1+i$), in accordance with Valuation Assumption 2.2. The actuarial loss experienced during year $(t-1, t)$ is therefore

$$l(t) = ul(t) - u[ul(t-1) - adj(t-1) - AL(t-1) - NC(t-1) + B(t-1)] - AL(t-1). \tag{3.32}$$

Projection Assumption 2.2 is now restored. The actuarial liability, normal cost and benefit outgo terms are constant in time and the equation of equilibrium (3.25) of Trowbridge (1952) applies. Hence, the following equations govern the basic model of §2.7:

$$f(t+1) = u(t+1)[f(t) + c(t) - B], \tag{3.33}$$

$$c(t) = NC + adj(t), \tag{3.34}$$

$$ul(t) = AL - f(t), \tag{3.35}$$

$$ul(t+1) = AL + u(t+1)[ul(t) - adj(t) - vAL], \tag{3.36}$$

$$l(t) = ul(t) - u[ul(t-1) - adj(t-1)]. \tag{3.37}$$

($v = 1/u = 1/(1+i)$.)

Actuarial gains and losses emerge as experience deviates from actuarial valuation assumptions, as described in the previous section. Actuarial losses increase the unfunded liability of the pension plan, whereas gains reduce it. But actuarial losses may also arise for reasons other than experience deviations:

Plan inception: Benefits that have accrued before the inception of the pension plan represent an unfunded 'past service liability'.

Plan amendment: If the benefit entitlements of plan members are enhanced (particularly restrospectively), this may increase the pension liability of the plan.

Amendment to valuation assumptions and funding method: Bleakney (1972:127) thus refers to "actuarial revaluation gains and losses" which "arise when actuarial assumptions are changed to reflect a reassessment of anticipated experience."

An 'initial' surplus or deficit in a pension plan, or *initial unfunded liability*, exists for any of the reasons stated above. The initial unfunded liability occurs only at time 0 in our model under Modelling Assumptions 2.2 and 2.3. Suppose the *initial* unfunded liability is $ul_0 = AL - f_0$. It is paid off explicitly under some funding methods, such as the *Frozen Initial Liability* methods (§3.1.2). In other funding methods, it may be paid off separately by explicit contribution payments. The initial unfunded liability is sometimes called a *supplemental liability*, to be amortized by explicit *supplemental costs* or supplementary contributions. Various patterns of payments may be employed but usually ul_0 is amortized over a period n (say): $P(t)$ in equation (3.17) represents the amortization of the initial unfunded liability and $U(t)$ in equation (3.18) is the unamortized part of the initial unfunded liability.

The specification of $adj(t)$ in terms of

- past and present experience deviations, represented either in terms of the actuarial losses $l(t)$ or the actuarial deficit $ul(t)$, and
- amortization of the initial unfunded liability ul_0 (if any)

completes the definition of the 'individual' funding methods. The contribution adjustment $adj(t)$ in equation (3.28) or (3.34) is variously known as the *supplemental cost* (in juxtaposition to the *normal cost* $NC(t)$) or the *supplementary contribution* (relative to the *standard contribution* $NC(t)$). It is also known as the *past-service contribution* as it amortizes unfunded pension liability accruing for service before time t, whereas the normal cost is sometimes termed *future-service contribution* as it amortizes benefits accruing for projected service after time t.

The calculation of the supplementary contribution is crucial to the dynamics of the pension fund. This is often revealed in stochastic projections or simulations of pension funds. Winklevoss (1982:593) finds that the choice of method to fund actuarial deviations is 'critical':

> "The correct treatment of actuarial gains and losses is critical in stochastic simulations because the effect of random fluctuations in salaries and plan assets impact on costs through the funding of such deviations."

There are various ways of adjusting contributions as experience deviates from actuarial assumptions and gains and losses occur. Some of the more common ways of calculating past service contributions include:

Amortizing gains and losses over a fixed term. This is studied mathematically by Dufresne (1986, 1989) among others (see §3.5).

Spreading surpluses and deficits over a moving term. This is also considered in detail by Dufresne (1986, 1988) among others (see §3.2).

Asymmetric spreading of deficits and surpluses. Surpluses and deficits need not be dealt with identically. This is a variation on the method above and is investigated numerically by Haberman & Smith (1997) (see §5.4).

Contribution holidays. 'Contribution holidays' are often taken by sponsors if large surpluses are emerging, no benefit improvement is being planned, and maximum surplus revenue rules are likely to be breached: it was fairly common practice in the U.K. in the 1980's.

Immediate cash injection. In some circumstances, gains/losses or surpluses/deficits may be removed immediately rather than being amortized gradually.

Ignoring gains and losses if they are within a given corridor. The 'corridor' approach is taken in various accounting standards, such as the U.S. Financial Accounting Standard No. 87 and International Accounting Standard No. 19. Small surpluses and deficits are merely carried forward. This is based on the conjecture that gains/losses will cancel out without any adjustment if actuarial valuation assumptions are 'on average' correct: this is investigated by Dufresne (1993, 1995).

Establishing a reserve. The reserving approach is common in jurisdictions where book values are used for assets: this may involve an implicit 'hidden' reserve [Daykin *et al.* (1994)] or an explicit 'investment' reserve. The latter is called 'buffer capital' by Kleynen (1997) who investigates by simulations how to use it to stabilise contributions and deal with surpluses/deficits in a Dutch pension fund. The 'buffer capital' is equal to the market value less the book value of assets upon which contributions are based. The buffer increases by the stock price increase over the year (up to a maximum) when the fund performs better than anticipated, while the buffer decreases (to a minimum of zero) if the fund does not perform as well and an actuarial loss arises.

3.2 Spreading Surpluses & Deficits over a Moving Term

3.2.1 A Supplementary Funding Method

A particular supplementary funding method is examined in more detail in this section. This method requires the unfunded liability at time t to be spread into the future, over a 'spread period' of m thus:

$$adj(t) = k\, ul(t), \quad k = 1/\ddot{a}_{\overline{m}|}. \tag{3.38}$$

The annuity is typically calculated at the valuation discount rate. We assume that the same spread period is used whether there is a surplus or deficit. In practice, this may not be so. Surpluses or deficits are never completely removed except asymptotically as $t \to \infty$ as shown by Bowers *et al.* (1979). This is not a drawback because the random nature of experience deviations is such that no method completely removes gains and losses over a finite term.

When surpluses and deficits are spread forward as in equation (3.38), then from equations (3.28) and (3.29), the contribution paid at the start of year $(t,\, t+1)$ is

$$c(t) = NC(t) + k[AL(t) - f(t)]. \tag{3.39}$$

The following recurrence relations may be set up in terms of the fund level and unfunded liability (given equations (3.27) and (3.30)):

$$f(t+1) = u(t+1)[(1-k)f(t) + NC(t) + kAL(t) - B(t)], \tag{3.40}$$

$$ul(t+1) = AL(t+1) + u(t+1)[(1-k)ul(t) - AL(t) - NC(t) + B(t)]. \tag{3.41}$$

Equation (3.40) may be rewritten as:

$$f(t) = f_0(1-k)^t \prod_{j=1}^{t} u(j) + \sum_{j=0}^{t-1} (1-k)^{t-j-1}[NC(j) + kAL(j) - B(j)] \prod_{\tau=j+1}^{t} u(\tau), \tag{3.42}$$

or in terms of the logarithmic rate of return as:

$$f(t) = f_0(1-k)^t \exp\left[\sum_{j=1}^{t} \delta(j)\right]$$
$$+\sum_{j=0}^{t-1}(1-k)^{t-j-1}[NC(j)+kAL(j)-B(j)]\exp\left[\sum_{\tau=j+1}^{t} \delta(\tau)\right]. \quad (3.43)$$

In the basic model of §2.7, the actuarial liability, normal cost and benefit outgo are fixed and Trowbridge's (1952) equation of equilibrium (3.25) holds. Hence, the following equations apply:

$$c(t) = NC + k(AL - f(t)), \quad (3.44)$$
$$f(t+1) = u(t+1)[(1-k)f(t) + (k-d)AL], \quad (3.45)$$
$$ul(t+1) = AL + u(t+1)[(1-k)ul(t) - vAL], \quad (3.46)$$

Equations (3.42) and (3.43) may also be simplified:

$$f(t) = f_0(1-k)^t \prod_{j=1}^{t} u(j) + AL(k-d)\sum_{j=0}^{t-1}(1-k)^{t-j-1}\prod_{\tau=j+1}^{t} u(\tau) \quad (3.47)$$
$$= f_0(1-k)^t \exp\left[\sum_{j=1}^{t}\delta(j)\right] + AL(k-d)\sum_{j=0}^{t-1}(1-k)^{t-j-1}\exp\left[\sum_{\tau=j+1}^{t}\delta(\tau)\right]. \quad (3.48)$$

The method of adjustment defined by equation (3.38) is commonly used in the United Kingdom. Turner (1984:21–23) gives an account of typical implementations of this method in the U.K. It has been called the 'Spread' method, the terminology being borrowed from Trowbridge (1952). It does not involve actuarial gains and losses directly but is an indirect method or a "*spread* method of dealing with gains and losses" in the sense that "actuarial gains and losses are automatically, and without separate identification, spread" over a future period, as McGill *et al.* (1996:525) explain. In addition, initial unfunded liabilities are not treated separately from experience actuarial gains/losses. Bowers *et al.* (1979) also describe this method as "Normal Cost plus Amortization over a Moving Term".

This supplementary funding method is implicit in the *Aggregate* actuarial cost method, which is itself referred to as a "spread gain valuation" method [Berin (1989:63), Aitken (1994:326)]. Under *Aggregate* funding,

$$f(t+1) = u(t+1)\big[(1 - S(t)/PVS(t))f(t) + PVB(t)\,S(t)/PVS(t) - B(t)\big], \quad (3.49)$$

(by substituting equation (3.16) into equation (3.27)) which somewhat resembles the funding pattern described in equation (3.40) when surpluses and deficits are spread forward. By virtue of the assumptions in the simple model of §2.7, the benefit outgo $B(t)$, the present value of future benefits $PVB(t)$, the total payroll for actives $S(t)$, and the present value of future salaries $PVS(t)$ (in equations (3.2), (3.3), (3.14) and (3.15) respectively) are fixed in time.

$$f(t+1) = u(t+1)\big[(1 - S/PVS)f(t) + PVB\,S/PVS - B\big], \tag{3.50}$$

S/PVS and $PVB\,S/PVS$, when the *Aggregate* funding method is used, correspond precisely to k and $NC{+}kAL$ respectively when surpluses and deficits are spread forward (by comparing equations (3.45) and (3.50)). This point is specifically noted and exploited by Trowbridge & Farr (1976:85), Bowers *et al.* (1979) and Dufresne (1986). Trowbridge (1963) effectively considers S/PVS as a free control parameter for a generalised family of aggregate funding methods. The *Aggregate* method is not pursued here because of this similarity.

It is obvious that the 'pace' of funding depends on the period over which payments for the unfunded liability are spread into the future. It has also been shown by Dufresne (1988) that the choice of this parameter affects the stationarity in the limit of the fund and contribution levels.

Two special and simple cases arise. The following observations follow from equation (3.45).

1. Suppose there is an immediate cash injection to pay off the unfunded liability in its entirety: $m = k = 1$.

$$f(t) = vALu(t), \quad t \geq 1. \tag{3.51}$$

2. Suppose surpluses and deficits are spread in perpetuity: $m = \infty$, $k = d$. Only interest is paid on the surplus or deficit ($k = d$ in equation (3.38)).

$$f(t+1) = f(t)u(t+1)v, \tag{3.52}$$

$$f(t) = f_0 v^t \prod_{\tau=1}^{t} u(\tau), \quad t \geq 1. \tag{3.53}$$

Dufresne (1988) derives the following result based on equations (3.44), (3.45) and Projection Assumption 2.3.

RESULT 3.1 *Let* $u = 1 + i$.

$$\mathrm{E}f(t) = AL - ul_0[u(1-k)]^t, \tag{3.54}$$

$$\mathrm{E}c(t) = NC + ul_0[u(1-k)]^t, \tag{3.55}$$

Provided $|u(1-k)| < 1$, *then*

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL, \tag{3.56}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC. \tag{3.57}$$

If $m = 1$, *then* $k = 1$ *and* $\mathrm{E}f(t) = AL$ *for* $t \geq 1$. *If* $m = \infty$, *then* $k = d$ *and* $\mathrm{E}f(t) = f_0\ \forall t$.

Let $q = \mathrm{E}u(t)^2 = u^2 + \sigma^2$. *Provided* $q(1-k)^2 < 1$, *then*

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \sigma^2 v^2 AL^2/[1 - q(1-k)^2] \tag{3.58}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t) = k^2 \lim \mathrm{Var}f(t). \tag{3.59}$$

Projection Assumption 2.3 requires that $\{i(t)\}$ be a sequence of independent and identically distributed random variables, so that $i(t+1)$ is independent of $i(t)$ and of $f(t)$ in equation (3.45). Solving the difference equation obtained by taking expectations across recurrence relationship (3.45) yields equations (3.54) and (3.56). The first moment of $c(t)$ is obtained from equation (3.44). Equation (3.58) follows by squaring and taking expectations across equation (3.45), and solving the resulting difference equation,

$$\mathrm{Var}f(t+1) = q(1-k)^2 \mathrm{Var}f(t) + \sigma^2 v^2 [\mathrm{E}f(t+1)]^2. \tag{3.60}$$

From equation (3.44), one finds that

$$\mathrm{Var}c(t) = k^2 \mathrm{Var}f(t). \tag{3.61}$$

The solution of equation (3.60) is given by Owadally & Haberman (1999):

RESULT 3.2 *If surpluses and deficits are spread over* m *such that* $\{d \leq k \leq 1,\ k \neq 1 - u/q,$ $k \neq k_{min}\}$ *where* $k_{min} = 1 - 1/\sqrt{q}$, *then*

$$\mathrm{Var}f(t) = \Theta + \Omega[u(1-k)]^t + \Psi[u(1-k)]^{2t} - (\Theta + \Omega + \Psi)[q(1-k)^2]^t, \tag{3.62}$$

$$\mathrm{Var}c(t) = k^2 \mathrm{Var}f(t), \tag{3.63}$$

where $\Theta = \sigma^2 v^2 AL^2/[1 - q(1-k)^2]$, $\Omega = -2\sigma^2 v^2 ALul_0/[1 - vq(1-k)]$, *and* $\Psi = -ul_0^2$. *If* $k > k_{min}$, *then as* $t \to \infty$, $\mathrm{Var}f(t) \to \Theta$. *When* $m = 1$ $(k = 1)$, $\mathrm{Var}f(t) = \sigma^2 v^2 AL^2$ *(which is immediately verified from equation* (3.51)*). When* $m \to \infty$ $(k = d)$, *then* $\mathrm{Var}f(t) = \left[(1 + \sigma^2 v^2)^t - 1\right] f_0^2$ *and* $f(t)$ *is not stationary (as may also be verified from equation* (3.53)*).*

The first moment of the funding process is stable if $|u(1-k)| < 1$. It can be shown that for $i > -100\%$, $0 < u(1-k) < 1$ or $k > d$ (Haberman, 1992a), so that as $t \to \infty$, $\mathrm{E}f(t) \to AL$ and $\mathrm{E}c(t) \to NC$. This implies that the fund eventually accumulates to meet all pension liabilities on average and that the average contribution is eventually equal to the normal cost or standard contribution. For stability in the second moments, there exists a *maximum* spread period, corresponding to k_{min} (i.e. $q(1-k)^2 < 1$ for stability). This condition becomes more constraining as the variance σ^2 of the rate of return process increases.

Dufresne (1986) also investigates the funding process when Valuation Assumption 2.2 does not hold and the valuation discount rate i_v is not equal to the long-term mean rate of investment return on the pension fund. The equation of equilibrium (3.25) or (3.26) may then be written

$$AL = u_v(AL + NC - B), \tag{3.64}$$

$$B = d_v AL + NC \tag{3.65}$$

(where $u_v = 1 + i_v$ and $d_v = i_v/(1 + i_v)$). Equations (3.33), (3.34) and (3.35) apply verbatim while equation (3.36) is modified by application of the equation of equilibrium in (3.64):

$$ul(t+1) = AL + u(t+1)[ul(t) - adj(t) - v_v AL], \tag{3.66}$$

When supplementary contributions are calculated according to equation (3.38), it follows that

$$f(t+1) = u(t+1)[(1-k)f(t) + (k - d_v)AL], \tag{3.67}$$

$$ul(t+1) = AL + u(t+1)[(1-k)ul(t) - v_v AL]. \tag{3.68}$$

It is straightforward to show that the following result applies (see Dufresne (1986)):

RESULT 3.3 *Provided* $|u(1-k)| < 1$, *then*

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL(d_v - k)/(d - k), \tag{3.69}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + ALk(d - d_v)/(d - k). \tag{3.70}$$

Provided $q(1-k)^2 < 1$, *then*

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \sigma^2 v^2 AL^2[1 + (d - d_v)/(k - d)]^2/[1 - q(1-k)^2], \tag{3.71}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t) = k^2 \lim \mathrm{Var}f(t). \tag{3.72}$$

3.2.2 Efficient Spreading Periods

Dufresne (1986, 1988) shows that the choice of spread period m is crucial not only because it affects the pace of funding and the stationarity of the funding process in the limit but also because it influences the *tradeoff* between the security and stability objectives in pension funding. Shorter spreading periods for surpluses and deficits (short m) ought to ensure that the funding objective is reached faster, thereby enhancing the security of benefits. But larger supplementary contributions are then required from the sponsor as and when gains and losses emerge. Dufresne (1986, 1988) shows that, in terms of ultimate variances of the fund and contribution levels, the tradeoff between security and stability breaks down for long spread periods beyond a certain efficient range of spread periods.

Dufresne (1988) proves the following result.

RESULT 3.4 *Let* m^* *be such that* $k^* = 1 - 1/q$. *Assuming* $q > 1$,
if $k_{min} < k < k^*$ *then* $\lim \mathrm{Var}f(t)$ *and* $\lim \mathrm{Var}c(t)$ *decrease with increasing* k;
if $k^* \leq k < 1$ *then* $\lim \mathrm{Var}f(t)$ *decreases and* $\lim \mathrm{Var}c(t)$ *increases with increasing* k.

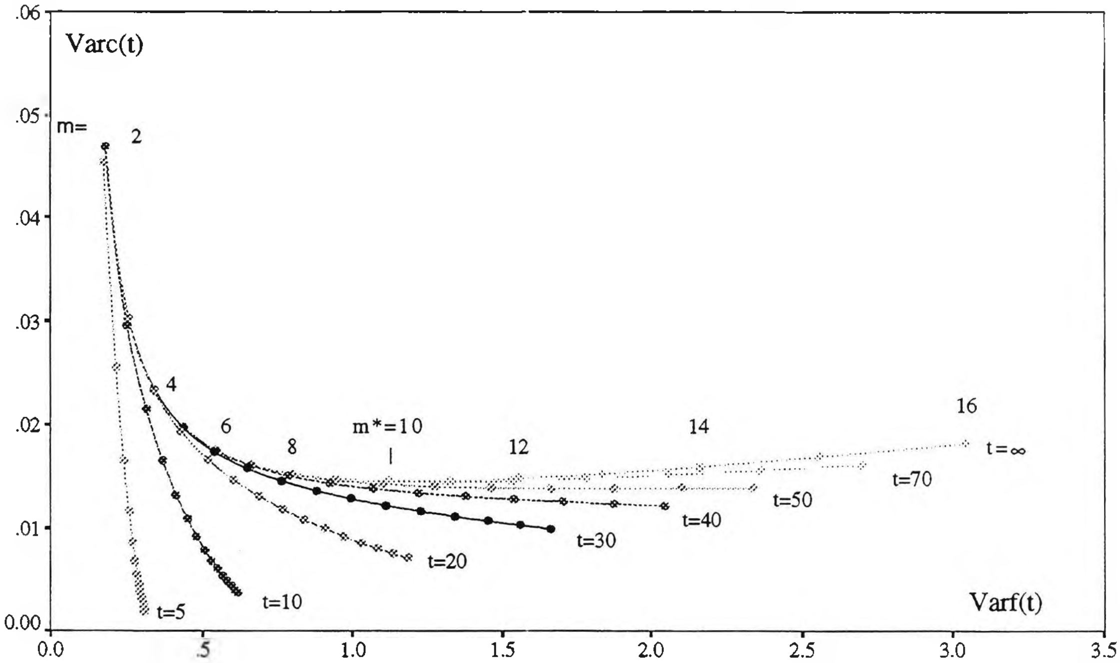


Figure 3.1: Profile of variances over time for a pension fund with $AL = 1.5$, $NC = 0.2$, $i = 3\%$, $\sigma = 0.25$. Surpluses and deficits as well as the initial unfunded liability $ul_0 = 0.5$ are being spread over period m.

m and k have a one-to-one inverse relationship ($k = 1/\ddot{a}_{\overline{m|}}$). A $\lim \text{Varc}(t)$ v. $\lim \text{Var}f(t)$ curve therefore exhibits a minimum: see Figure 3.1, which is obtained from equations (3.62) and (3.63). For spreading periods $m > m^*$ ($k < k^*$), there will always be a shorter spreading period for which both $\lim \text{Var}f(t)$ and $\lim \text{Varc}(t)$ are reduced, and therefore they would be "inadmissible" (Dufresne, 1988). Dufresne (1986) describes the range $1 \leq m \leq m^*$ (corresponding $k^* \leq k \leq 1$) as an "optimal" range of spread periods.

It is clear that the choice of m in the efficient range $[1, m^*]$ depends upon the balance struck among the various stakeholders in the pension planning arrangement between the objectives of *security* (low variance in the fund level) and *stability* (low variance in the contribution level). Dufresne (1986, 1988) lists values of m^* for various combinations of $\{i, \sigma\}$, the first and second moments of the rate of return process. He concludes that under current economic conditions, $m \in [1, 10]$ is an efficient range for the spreading period. Thornton & Wilson (1992a) recommend short spreading periods to be used in the U.K. partly based on Dufresne's (1988) conclusion.

The feature of an efficient range of spreading periods has been reproduced under more realistic investment projection assumptions (see §3.3). Haberman & Smith (1997) use Wilkie's (1987, 1995) model and observe numerically that contribution rate variability decreases and then increases as surpluses and deficits are spread over longer periods. But this feature is not noted by Bilodeau (1995) who also uses Wilkie's (1987) model and reports that contribution variability decreases as the spreading period increases (spreading periods of only up to 20 years were considered). Loades (1992), who uses a deterministic model with cyclic rates of return and more realistic features, such as variable valuation discount rates (exponentially smoothed averages of the rates of return) and a discounted income value of assets, also reports that spreading over longer periods decreases the variability of contributions.

The choice of period over which to spread forward surpluses and deficits thus affects:

- the 'pace' of funding,
- the stationarity of the pension funding process in the limit,
- moments and probability distributions of the fund and contribution levels over time,
- the tradeoff between the objectives of security and stability in pension funding.

Careful consideration should therefore be given to its choice.

3.2.3 Delay, Frequency

Delays in the Valuation Process. For a significant minority of pension plans, especially large ones, there may be a delay of up to one year between data collection and publication of the valuation report. Haberman (1992a) and Zimbidis & Haberman (1993) investigate the effect of such delay by introducing a delay ϕ such that

$$adj(t) = k \; ul(t - \phi) \tag{3.73}$$

It is important to study the effect of delays in the valuation process not just because of the accounting and actuarial effort required during a valuation but also because a number of actuarial methods (e.g. when calculating contribution adjustment or setting asset values) allow the use of delayed information, usually for purposes of smoothing and averaging.

The introduction of delays in the valuation leads to a loss of 'information' and this should make the system less stable. Haberman (1992a), Zimbidis (1992) and Zimbidis & Haberman (1993) demonstrate that the ultimate variances of fund and contribution levels are larger as the delay ϕ increases. They conclude that delays should be avoided as they adversely affect both security and stability in the pension fund. In addition, Owadally & Haberman (1999) illustrate that delays in the funding process cause

1. the (first and second moment) stability conditions for the fund level to be more constrained (i.e. the allowable spread period range is more restricted);
2. the various moments to exhibit possible oscillatory behaviour (on average, there may be a succession of surpluses and then deficits) for certain spread periods, indicating that $f(t)$ and $c(t)$ do not converge 'smoothly' to certain probability distributions in the limit.

If there are long delays in the pension funding process, oscillations in the first and other moments of the fund level mean that the pension system is less 'stable'. The actuary has less control and it becomes more difficult to judge fund solvency and recommend a contribution rate. The fact that delays cause instability in actuarial systems is well-known as Balzer & Benjamin (1980) show in an insurance system with delayed profit/loss-sharing feedback. Balzer (1982) and Taylor (1987) also examine delays in a range of actuarial systems.

Frequency of Valuations. An issue related to delayed information is the frequency of pension fund valuations. Because of the long-term liabilities of a pension plan and the complexities and expense of its valuation, triennial valuations were until recently the norm in the U.K. Valuations are performed every year in the U.S. Even though they may be infrequent, valuations are performed *regularly*, as required by statute. Haberman (1993) investigates the effect of frequency of valuation on the pension fund and finds that less frequent valuations lead to more variable contribution and fund levels. Haberman & Smith (1997:§10) also apply Wilkie's (1995) stochastic asset model and simulate the effect of a change in the frequency of valuations. They also conclude that less frequent valuations increase the variability of the pension funding process. Although it has sometimes been argued (Cronquist & Dreher, 1972) that triennial valuations are more efficient on the basis of expense, actuarial practice tends to more frequent annual valuations. Grubbs' [Cronquist & Dreher (1972): discussion] views confirm that more effective 'control' of pension funds may be achieved with frequent valuations:

> "Actuarial valuations serve as a guide to steer one's course toward the objective of meeting eventual benefit payments. [. . .] The more frequently [valuations are undertaken], the less violent the changes in course that will be required. One objective of employers is to have pension costs which will remain relatively stable as a percentage of compensation if the plan is related to pay, or as a cost per employee if the plan is not related to pay. More frequent valuations serve this end a little better."

3.2.4 Explicit Amortization of Initial Unfunded Liability

When surpluses and deficits are spread forward (as in equation (3.38)), initial unfunded liabilities are not treated separately from experience actuarial gains/losses. This appears to be usual practice in the U.K. but not in North America. There are several reasons why initial unfunded liabilities ought to be explicitly and separately amortized:

1. Initial unfunded liabilities may exist at plan inception or may arise as a result of an amendment to pension plan benefits or to valuation methods and assumptions (§3.1.3). The sources of initial unfunded liabilities are of an entirely different nature from experience gains and losses. They are likely to be 'one-offs' unlike random experience

deviations. There is no reason to treat initial unfunded liabilities like experience gains and losses.

2. Both the U.S. Employee Retirement Income Security Act, 1974 (ERISA) and the Canadian Pension Benefits Standards Act, 1985, allow initial unfunded liabilities to be amortized over different schedules than experience gains and losses. (Some accounting standards, like the Financial Accounting Standards No. 87 (FAS87) also allow separate amortization of "prior service cost" arising from plan amendments etc.)

3. Finally, amortizing initial unfunded liabilities over a different period provides more "funding flexibility", as pointed out by McGill *et al.* (1996:546), since this period can be varied (within set maximum limits). Snelson (1970:¶21) specifies that the 'pace of funding' depends not just on the funding objective defined by a particular funding method, but also on the period over which initial unfunded liabilities are being defrayed. Suitable amortization periods for initial unfunded liabilities are discussed in several papers in the early actuarial literature on pension mathematics (see Shapiro, 1985). (Hagerman & Zmijewski (1979) investigate the choice of periods over which U.S. corporations amortized initial unfunded liabilities for *accounting* purposes at a time when accounting standards allowed flexibility in this respect.)

A simple variation on the method of equation (3.38) that allows explicit amortization of initial unfunded liabilities is now described (Owadally & Haberman, 1999). Initial unfunded liabilities are amortized over a fixed term and the excess of any subsequent actuarial deficit over the unamortized part of the initial unfunded liabilities is spread over a moving term:

$$adj(t) = k[ul(t) - U(t)] \ + \ P(t), \quad k = 1/\ddot{a}_{\overline{m}|}, \tag{3.74}$$

and the contribution recommended by the actuary at the start of year $(t,\, t+1)$ is therefore (from equation (3.34)):

$$c(t) = NC + k[ul(t) - U(t)] \ + \ P(t). \tag{3.75}$$

This method is implicit in *Frozen Initial Liability* methods (§3.1.2) but does not appear to be used with individual actuarial cost methods. The amortization payment $P(t)$ and the unamortized part of the initial unfunded liability $U(t)$ are calculated as in equations (3.17) and (3.18) respectively. The initial unfunded liability is amortized over n years whereas deficits in excess of the unamortized part of the initial unfunded liability are spread forward a rolling term of m years. In the U.S. Employee Retirement Income Security Act, 1974 (ERISA), $P(t)$ in equation (3.17) is termed an *amortization credit* when positive and an *amortization charge* when negative.

A recurrence relation may be set up in terms of the fund level by substituting equation (3.74) into equations (3.33) and (3.34) and using the equation of equilibrium (3.26):

$$f(t+1) = u(t+1)[(1-k)f(t) + (k-d)AL + P(t) - kU(t)]. \tag{3.76}$$

Now consider the *Frozen Initial Liability* pension funding methods. By virtue of the assumptions in the simple model of §2.7, the benefit outgo $B(t)$, the present value of future benefits $PVB(t)$, the total payroll for actives $S(t)$, and the present value of future salaries $PVS(t)$ are time-invariant. The contribution $c(t)$ in equation (3.19) may be substituted into equation (3.33) to give

$$f(t+1) = u(t+1)[(1 - S/PVS)f(t) + (PVB\,S/PVS - B) + P(t) - (S/PVS)U(t)]. \tag{3.77}$$

There is an obvious similarity between equations (3.76) and (3.77) and S/PVS corresponds precisely to k. The *Frozen Initial Liability* methods implicitly employ a form of surplus/deficit spreading with initial unfunded liability amortization.

Owadally & Haberman (1999) obtain the first and second moments of the fund and contribution level when supplementary contributions are calculated according to equation (3.74):

PROPOSITION 3.1 *For* $n < \infty$,

$$\mathrm{E}f(t) = \begin{cases} AL - ul_0\, \ddot{a}_{\overline{n-t}|}/\ddot{a}_{\overline{n}|}, & 0 \le t \le n-1, \\ AL, & t \ge n, \end{cases} \tag{3.78}$$

$$\mathrm{E}c(t) = \begin{cases} NC + ul_0/\ddot{a}_{\overline{n}|}, & 0 \le t \le n-1, \\ NC, & t \ge n. \end{cases} \tag{3.79}$$

If $n = \infty$, *then* $\mathrm{E}f(t) = f_0$, $\mathrm{E}c(t) = NC + d\,ul_0$, *for* $t \ge 0$.

If the initial unfunded liability is amortized over $n < \infty$ *while subsequent surpluses/deficits are spread over* m *such that* $\{d \le k \le 1,\ k \ne k_{min},\ k \ne 1 - \sqrt{(u/q)},\ k \ne 1 - u/q\}$, *then*

$$\mathrm{Var}f(t) = \begin{cases} \Phi + \Lambda u^t + \Upsilon u^{2t} - (\Phi + \Lambda + \Upsilon)[q(1-k)^2]^t, & 1 \le t \le n-1, \\ \Theta - [q(1-k)^2]^{t-n+1}[\Theta - \mathrm{Var}f(n-1)], & t \ge n, \end{cases} \tag{3.80}$$

where

$$\Phi = \sigma^2 v^2 [AL - ul_0/(1-v^n)]^2/[1 - q(1-k)^2], \tag{3.81}$$

$$\Lambda = 2\sigma^2 v^2 [AL - ul_0/(1-v^n)][v^n ul_0/(1-v^n)]/[1 - vq(1-k)^2], \tag{3.82}$$

$$\Upsilon = \sigma^2 v^2 [v^n ul_0/(1-v^n)]^2/[1 - v^2 q(1-k)^2]. \tag{3.83}$$

Θ *and* k_{min} *are defined as in Result 3.2.* $\mathrm{Var}c(t) = k^2\mathrm{Var}f(t)$. *If* $k > k_{min}$, *then as* $t \to \infty$, $\mathrm{Var}f(t) \to \Theta$.

If the initial unfunded liability is amortized in perpetuity ($n = \infty$) *and surpluses/deficits are spread over* m *such that* $\{d \le k \le 1,\ k \ne k_{min}\}$, *then*

$$\mathrm{Var}f(t) = \sigma^2 v^2 f_0^2 [1 - (q(1-k)^2)^t]/[1 - q(1-k)^2]. \tag{3.84}$$

The proof is very straightforward and is only sketched here (see Owadally & Haberman (1999) for more details). The first moment results are easily obtained by taking expectations across equation (3.76) and solving the resulting two-stage difference equation (noting that $U(t) = P(t)+vU(t+1)$ and letting $U(n) = 0$). To obtain the second moments, (1) first square equation (3.76) and then take expectation (noting that $u(t+1)$ is independent of $f(t)$); (2) first take expectation on equation (3.76), then square both sides and finally multiply both sides by q/u^2; (3) deduct (2) from (1) to obtain

$$\mathrm{Var} f(t+1) = \begin{cases} q(1-k)^2 \mathrm{Var} f(t) + \sigma^2 v^2 [AL - U(t+1)]^2, & 0 \le t \le n-2, \\ q(1-k)^2 \mathrm{Var} f(t) + \sigma^2 v^2 AL^2, & t \ge n-1. \end{cases} \tag{3.85}$$

This is readily solved to give equation (3.80).

Remarks:

1. After a finite number n of years, we *expect* the initial unfunded liability to be completely amortized and the fund and contribution levels to reach their target values (equations (3.78) and (3.79)). Liabilities are only funded asymptotically when surpluses and deficits are spread without special amortization of the initial unfunded liability (in Result 3.1).

2. The first-moment evolution of the funding process does not depend on the spread period m (cf. Result 3.1) but the second moments do.

3. Whether the initial unfunded liability is separately amortized or not, the second moments of the funding process are the same in the limit (unsurprisingly): in both Result 3.2 and Proposition 3.1, $\lim \mathrm{Var} f(t) = \Theta$ and $\lim \mathrm{Var} c(t) = k^2\Theta$, and the same stability conditions apply. The efficient range of spreading periods m (see Result 3.4) as obtained by Dufresne (1986, 1988) applies.

3.3 Dependent Rates of Return

3.3.1 Dependence in Pension Fund Investment Rates of Return

The rates of return on the pension fund from year to year are likely to be statistically dependent. This is so for two reasons:

1. There is considerable controversy over the Efficient Market Hypothesis in the long-term (§2.6.3). Nominal returns will incorporate price inflation which is generally observed to be correlated over the long term (§2.6.1). Returns on several asset classes have been found to be correlated over time, as is demonstrated by the statistical analysis of Panjer & Bellhouse (1980), Fama & French (1988), Poterba & Summers (1988), Frees (1990), Wilkie (1995) among others. (See §2.6.3.)

2. Whether or not markets are efficient, not all the securities held by the fund will be traded every year typically and some dependence in the returns from individual securities will occur [Vanderhoof (1973), Dufresne (1994:140)]. This is particularly the case where debt securities are held until maturity, possibly to match certain liability cash flows. Even when the asset portfolio of a pension fund is actively managed, many securities will be held for over a year. (McGill *et al.* (1996:663) report that "in a volatile business environment, a third or a half of a common stock (equity) portfolio may turn over within a one-year period.")

It is of interest to consider to what extent the results of Dufresne (1986, 1988) (Results 3.1 and 3.4) hold when dependent rates of return are assumed. The two most common time series of rates of return applied in mathematical actuarial models are the autoregressive [Pollard (1971), Panjer & Bellhouse (1980), Bellhouse & Panjer (1981), Dhaene (1989)] and moving average processes [Frees (1990), Dufresne (1990a)]. Haberman (1994a) examines pension funding in the context of stationary Gaussian autoregressive logarithmic rates of return of order 1 and 2, AR(1, 2). Wong (1995) and Haberman & Wong (1997) consider stationary Gaussian moving average logarithmic rates of return of order 1 and 2, MA(1, 2).

Haberman (1994a) shows that exact closed-form solutions cannot be obtained for the limiting first and second moments of the pension funding process when autoregressive rates of return (AR(1, 2)) are assumed and surpluses or deficits are spread forward as in equation (3.38). His numerical analysis shows that first-order approximations are accurate when the rate of return process is not strongly autocorrelated. Cairns & Parker (1997) establish some improved upper and lower bounds on the moments for an AR(1) logarithmic rate of return. Moving average processes are somewhat more tractable mathematically (Frees, 1990) and Wong (1995) and Haberman & Wong (1997) show that explicit closed-form solutions may be obtained when rates of return are MA(1, 2). Haberman (1994a) and Haberman & Wong (1997) report two noteworthy sets of results:

1. Conditions for the stability of first and second moments of the pension funding process are obtained for AR(1, 2) and MA(1, 2): these restrict the feasible range of the spreading period m. This range decreases as the variance of the rate of return increases. This is similar to the conclusion in Result 3.1 as obtained by Dufresne (1986, 1988). The allowable range of spread periods also becomes more restricted as the rate of return process is more positively correlated.

2. An 'optimal' spread period range, in the sense of Dufresne (1986), is also obtained for the AR(1, 2) and MA(1, 2) cases when the rate of return process exhibits moderate autocovariance under practical conditions. As surpluses and deficits are spread over a longer period, more stable recommended contributions and more variable fund levels are obtained. But beyond a certain spreading period, both contribution and fund levels

become more variable. As in the random walk model of Dufresne (1986, 1988), the optimal range is more restricted as the variance of the rate of return process increases. The optimal range also decreases for more positively autocorrelated AR(1, 2) and MA(1, 2) rate of return processes. As the variance and/or autocovariance of the rate of return increase beyond some threshold, the optimal range vanishes and minimum contribution and fund level variability are yielded when surpluses and deficits are not spread into the future ($m = 1$).

3.3.2 AR(p) Rates of Return

In this section, the results of Haberman (1994a) are extended to a more general stationary autoregressive process. Projection Assumption 2.3 is replaced by the following for the remainder of §3.3:

PROJECTION ASSUMPTION 3.1 (INVESTMENT RATE OF RETURN)
The logarithmic or geometric rate of return process $\delta(t) = \ln(1 + i(t))$ follows a stationary Gaussian autoregressive process of order p, AR(p), $p < \infty$:

$$\delta(t+1) - \delta = \varphi_1(\delta(t) - \delta) + \varphi_2(\delta(t-1) - \delta) + \cdots + \varphi_p(\delta(t-p+1) - \delta) + e(t+1), \tag{3.86}$$

where $\{e(t)\}$ is a sequence of zero-mean independent and identically normally distributed variables. $\delta(t)$ is stationary and we may let $\mathrm{E}\delta(t) = \delta$ and $\mathrm{Var}\delta(t) = \sigma^2 \; \forall t$.

It may be shown that (Appendix A.1)

$$\mathrm{Cov}[\delta(t), \delta(t-h)] = \sigma^2 \sum_{i=1}^{p} A_i G_i^h, \tag{3.87}$$

for $h \geq 0$ where $\{G_i\}$ is the set of *distinct* roots of

$$z^p - \varphi_1 z^{p-1} - \varphi_2 z^{p-2} - \cdots - \varphi_p = 0 \tag{3.88}$$

and $\{A_i\}$ is defined by $\mathrm{Cov}[\delta(t), \delta(t-h)]$ for $h \in [0, p-1]$ characterising the stationary AR(p) process (3.86). Stationarity implies that

$$|G_i| < 1 \quad \text{for } i \in [1, p]. \tag{3.89}$$

Furthermore,

$$\mathrm{E}\exp\left[\sum_{u=s+1}^{t} \delta(u)\right] = c^{t-s} e^{-\sum_i z_i} \exp\left[\sum_i z_i G_i^{t-s}\right], \tag{3.90}$$

$$\mathrm{E}\exp\left[2\sum_{u=s+1}^{t} \delta(u)\right] = (cw)^{t-s} e^{-4\sum_i z_i} \exp\left[4\sum_i z_i G_i^{t-s}\right]. \tag{3.91}$$

and

$$\mathrm{E}\exp\left[\sum_{u=s+1}^{t}\delta(u)+\sum_{w=\tau+1}^{t}\delta(w)\right]$$
$$=c^{t-\tau}w^{t-s}e^{-3\sum_i z_i}\exp\left\{\sum_i\left[2z_iG_i^{t-s}+2z_iG_i^{t-\tau}-z_iG_i^{s-\tau}\right]\right\}, \quad (3.92)$$

where

$$c=\exp\left[\delta+\frac{\sigma^2}{2}\sum_i A_i\frac{1+G_i}{1-G_i}\right], \quad (3.93)$$

$$w=\exp\left[\delta+\frac{3\sigma^2}{2}\sum_i A_i\frac{1+G_i}{1-G_i}\right], \quad (3.94)$$

$$z_i=\sigma^2A_iG_i(1-G_i)^{-2}. \quad (3.95)$$

See Appendix A (§A.1) for a proof of the above.

3.3.3 Pension Funding under AR(p) Rates of Return

It is assumed that the pension fund rate of return $\delta(t)$ follows a known AR(p) process, so that $\{G_i\}$ may be found from equation (3.88). Wilkie's (1995) model may be linearised into a set of autoregressive series (Kemp, 1996). Rate of return data may also be identified as AR(p) series with estimates of $\{\rho_i\}$. There is an exact correspondence between $\{\rho_i\}$ and $\{\varphi_i\}$ through the Yule-Walker equations (Box *et al.*, 1994:57) and $\{G_i\}$ may be found numerically. (Estimation errors may affect the robustness of the subsequent analysis.) Note also from equation (3.87) that if the terms $|A_iG_i|$ are small, then $\{\delta(t)\}$ exhibits weak autocorrelation.

The first and second moments of the pension fund and contribution levels when investment returns follow an AR(p) process (3.86) may be investigated using equations (3.90)–(3.95). The pension funding process is shown to be stable in the sense that the first and second moments of $f(t)$ and $c(t)$ converge as $t\to\infty$. The fund level at time t is given by equation (3.48) or, in terms of the notation of Haberman (1994a),

$$f(t)=f_0Q^t\exp\left[\sum_{u=1}^{t}\delta(u)\right]+R\sum_{s=0}^{t-1}Q^{t-s-1}\exp\left[\sum_{u=s+1}^{t}\delta(u)\right], \quad (3.96)$$

where $Q=1-k$ and $R=AL(k-d)$.

First Moments

Applying equation (3.90) to equation (3.96),

$$\mathrm{E}f(t)=f_0(Qc)^te^{-\sum_i z_i}\exp\left[\sum_i z_iG_i^t\right]+RQ^{-1}e^{-\sum_i z_i}\sum_{s=0}^{t-1}(Qc)^{t-s}\exp\left[\sum_i z_iG_i^{t-s}\right]. \quad (3.97)$$

Given that $\delta(t)$ is stationary and therefore inequality (3.89) holds, it follows that $\exp(\sum_i z_i G_i^s) \leq \exp(\sum_i |z_i G_i|)$ for $s \in \mathrm{N}$.

$$\mathrm{E}f(t) \leq f_0(Qc)^t e^{-\sum_i z_i} \exp\left[\sum_i |z_i G_i|\right] + RQ^{-1} e^{-\sum_i z_i} \exp\left[\sum_i |z_i G_i|\right] \sum_{s=0}^{t-1} (Qc)^{t-s} \quad (3.98)$$

and the right hand side of the above is convergent as $t \to \infty$, provided that $|Qc| < 1$. $\mathrm{E}f(t)$ is bounded above and since all terms in equation (3.97) are non-negative, $\mathrm{E}f(t)$ is convergent as $t \to \infty$.

A first order approximation to $\lim \mathrm{E}f(t)$ can also be obtained. From equation (3.97),

$$\begin{aligned} \mathrm{E}f(t) &\approx f_0(Qc)^t e^{-\sum_i z_i}\left[1 + \sum_i z_i G_i^t\right] \\ &\quad + RQ^{-1} e^{-\sum_i z_i} \sum_{s=0}^{t-1} (Qc)^{t-s} \left[1 + \sum_i z_i G_i^{t-s}\right] \quad (3.99) \\ &\to e^{-\sum_i z_i} Rc \left[(1 - Qc)^{-1} + \sum_{i=1}^{p} z_i G_i (1 - QcG_i)^{-1}\right], \quad (3.100) \end{aligned}$$

as $t \to \infty$ provided $|Qc| < 1$ (and given $|G_i| < 1$). This is a reasonable approximation when the terms $z_i G_i$ are small in magnitude, i.e. when $|A_i G_i|$ are small, which implies that $\{\delta(t)\}$ exhibits weak autocorrelation (equation (3.87)). Indeed, Haberman (1994a) shows that under such conditions,

$$\lim_{t \to \infty} \mathrm{E}f(t) \approx e^{-\sum_i z_i} Rc/(1 - Qc) \quad (3.101)$$

is an accurate approximation to $\lim \mathrm{E}f(t)$ for AR(1) and AR(2) processes. Lower and upper bounds are also discussed by Cairns & Parker (1997) for the AR(1) process.

Finally, note from equation (3.44) that $\lim \mathrm{E}c(t) = NC + k(AL - \lim \mathrm{E}f(t))$.

Second Moments

When both sides of equation (3.96) are squared and expanded and expectation is taken,

$$\begin{aligned} \mathrm{E}f(t)^2 &= f_0^2 Q^{2t} \mathrm{E}\exp\left[2\sum_{u=1}^{t} \delta(u)\right] + 2f_0 RQ^{t-1} \sum_{s=0}^{t-1} Q^{t-s} \mathrm{E}\exp\left[\sum_{u=1}^{t} \delta(u) + \sum_{w=s+1}^{t} \delta(w)\right] \\ &\quad + 2R^2 Q^{-2} \sum_{s=1}^{t-1} \sum_{\tau=0}^{s-1} Q^{t-s} Q^{t-\tau} \mathrm{E}\exp\left[\sum_{u=s+1}^{t} \delta(u) + \sum_{w=\tau+1}^{t} \delta(w)\right] \\ &\quad + R^2 Q^{-2} \sum_{s=0}^{t-1} Q^{2(t-s)} \mathrm{E}\exp\left[2\sum_{u=s+1}^{t} \delta(u)\right] \quad (3.102) \end{aligned}$$

The third term on the right hand side of equation (3.102) may be simplified, upon application of equation (3.92), to

$$2R^2Q^{-2}\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Qw)^{t-s}(Qc)^{t-\tau}e^{-3\sum_i z_i}\exp\left\{\sum_i\left[2z_iG_i^{t-s}+2z_iG_i^{t-\tau}-z_iG_i^{s-\tau}\right]\right\}$$
$$=2R^2Q^{-2}e^{-3\sum_i z_i}\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Q^2cw)^{t-s}(Qc)^{s-\tau}\exp\left\{\sum_i\left[2z_iG_i^{t-s}+2z_iG_i^{t-\tau}-z_iG_i^{s-\tau}\right]\right\}. \tag{3.103}$$

Again, $\delta(t)$ is stationary and equation (3.89) holds. Consequently,

$$\exp(\sum_i 2z_iG_i^s)\leq\exp(\sum_i 2|z_iG_i|), \tag{3.104}$$
$$\exp(-\sum_i z_iG_i^s)\leq\exp(\sum_i |z_iG_i|) \tag{3.105}$$

for $s\in\mathrm{N}$. Therefore, the third term on the right hand side of equation (3.102) is bounded above by

$$2R^2Q^{-2}e^{-3\sum_i z_i}\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Q^2cw)^{t-s}(Qc)^{s-\tau}\exp\left[\sum_i 5|z_iG_i|\right], \tag{3.106}$$

which, as $t\to\infty$, converges to

$$2R^2Q^{-2}e^{-3\sum_i z_i}\exp\left[\sum_i 5|z_iG_i|\right]\frac{Q^2cw}{1-Q^2cw}\frac{Qc}{1-Qc}, \tag{3.107}$$

provided $|Qc|<1$ and $Q^2cw<1$. Since all terms in (3.103) are non-negative, the third term on the right hand side of equation (3.102) converges as $t\to\infty$.

To first order, the third term on the right hand side of equation (3.102) is

$$2R^2Q^{-2}\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Qw)^{t-s}(Qc)^{t-\tau}e^{-3\sum_i z_i}\left\{1+\sum_i\left[2z_iG_i^{t-s}+2z_iG_i^{t-\tau}-z_iG_i^{s-\tau}\right]\right\}$$
$$=2R^2Q^{-2}e^{-3\sum_i z_i}\left\{\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Q^2cw)^{t-s}(Qc)^{s-\tau}+\sum_i 2z_i\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Q^2cwG_i)^{t-s}(Qc)^{s-\tau}\right.$$
$$\left.+\sum_i 2z_i\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Q^2cwG_i)^{t-s}(QcG_i)^{s-\tau}-\sum_i z_i\sum_{s=1}^{t-1}\sum_{\tau=0}^{s-1}(Q^2cw)^{t-s}(QcG_i)^{s-\tau}\right\}, \tag{3.108}$$

which converges as $t\to\infty$ to

$$e^{-3\sum_i z_i}2R^2Qc^2w\left\{(1-Q^2cw)^{-1}(1-Qc)^{-1}+\sum_i z_iG_i\Big[2(1-Q^2cwG_i)^{-1}(1-Qc)^{-1}\right.$$
$$\left.+2G_i(1-Q^2cwG_i)^{-1}(1-QcG_i)^{-1}-(1-Q^2cw)^{-1}(1-QcG_i)^{-1}\Big]\right\}. \tag{3.109}$$

This is a suitable approximation when the terms $|A_iG_i|$ are small, i.e. when the serial correlation of $\{\delta(t)\}$ is small. Haberman (1994a) shows that the further approximation

$$e^{-3\sum_i z_i} 2R^2Qc^2w(1-Q^2cw)^{-1}(1-Qc)^{-1} \tag{3.110}$$

is accurate for practical purposes for $p = 1, 2$.

As for the last term on the right hand side of equation (3.102), substitution of equation (3.91) gives

$$\begin{aligned} R^2Q^{-2}e^{-4\sum_i z_i} \sum_{s=0}^{t-1} (Q^2cw)^{t-s} \exp\left[4\sum_i z_i G_i^{t-s}\right] \\ \leq R^2Q^{-2}e^{-4\sum_i z_i} \sum_{s=0}^{t-1} (Q^2cw)^{t-s} \exp\left[4\sum_i |z_iG_i|\right] \\ \to e^{-4\sum_i z_i} \exp\left[4\sum_i |z_iG_i|\right] R^2cw/(1-Q^2cw) \end{aligned} \tag{3.111}$$

as $t \to \infty$ provided $Q^2cw < 1$. The fourth term on the right hand side of equation (3.102) thus also converges as $t \to \infty$ and its limit, to first order, may be obtained as follows:

$$\begin{aligned} e^{-4\sum_i z_i} R^2Q^{-2} \sum_{s=0}^{t-1} (Q^2cw)^{t-s} \left[1 + 4\sum_i z_i G_i^{t-s}\right] \\ \to e^{-4\sum_i z_i} R^2Q^{-2} \left[\frac{Q^2cw}{1-Q^2cw} + \sum_i 4z_i \frac{Q^2cwG_i}{1-Q^2cwG_i}\right] \quad \text{as } t \to \infty \\ = e^{-4\sum_i z_i} R^2cw \left[(1-Q^2cw)^{-1} + \sum_i 4z_iG_i(1-Q^2cwG_i)^{-1}\right]. \end{aligned} \tag{3.112}$$

Haberman (1994a) shows that

$$e^{-4\sum_i z_i} R^2cw(1-Q^2cw)^{-1} \tag{3.113}$$

is a reasonably accurate approximation in practice for $p = 1, 2$.

The first two terms on the right hand side of equation (3.102) vanish provided $|Qc| < 1$, $Q^2cw < 1$ as well as $|Q| < 1$. The last condition is redundant since $Q = 1 - k = 1 - 1/\ddot{a}_{\overline{m|}}$. Hence, we have shown that $\mathrm{E}f(t)^2$ is convergent as $t \to \infty$ for $\delta(t)$ following a stationary AR(p) process, $p \in \mathrm{N}$, $p < \infty$, provided $|Qc| < 1$ and $Q^2cw < 1$. Combining equations (3.110) and (3.113),

$$\lim_{t\to\infty} \mathrm{E}f(t)^2 \approx e^{-3\sum_i z_i} \frac{2R^2Qc^2w}{(1-Q^2cw)(1-Qc)} + e^{-4\sum_i z_i} \frac{R^2cw}{1-Q^2cw}. \tag{3.114}$$

The numerical analysis of Haberman (1994a) shows that this is an accurate approximation for AR(1) and AR(2) processes. Cairns & Parker (1997) provide some lower and upper bounds for $\lim \mathrm{E}f(t)^2$ for the AR(1) process.

Finally, an approximation to $\lim \mathrm{Var}f(t)$ may be obtained from equations (3.101) and (3.114). $\lim \mathrm{Var}c(t) = k^2 \lim \mathrm{Var}f(t)$ given equation (3.44).

Some Remarks

1. When surpluses and deficits are not spread and $m = 1$, the fund level is described by equation (3.51) and its moments are straightforward:

$$\mathrm{E}f(t) = vAL\mathrm{E}e^{\delta(t)} = AL, \tag{3.115}$$

$$\mathrm{Var}f(t) = v^2AL^2\mathrm{Var}e^{\delta(t)} = AL^2(e^{\sigma^2} - 1). \tag{3.116}$$

for $t \geq 1$ (since $\mathrm{E}\exp(\delta(t)) = \exp(\delta + \sigma^2/2) = u = v^{-1}$).

2. c, w and z_i (in equations (3.93), (3.94) and (3.95) respectively) correspond to the definitions of Haberman (1994a) for $p = 1, 2$. The forms of the first and second moments in equations (3.101) and (3.114) respectively generalise the forms obtained by Haberman (1994a) for $p = 1, 2$, which justifies Haberman's (1992b) conjecture to this effect.

3. The stability conditions $|Qc| < 1$ and $Q^2cw < 1$ are also obtained by Haberman (1994a) for the cases $p = 1, 2$. They impose a stability constraint on allowable spread periods m, as discussed by Haberman (1994a). These restrictions are quite general and extend to $p \in \mathrm{N}$, $p < \infty$. Furthermore, one observes that both c and w decrease as the mean δ of the logarithmic rate of return process decreases, so that the maximum spread period allowed if the first and second moments of the funding process are to be stable (i.e. if the stability conditions above hold) increases. It is also apparent that when the rate of return process exhibits small autocovariance and the terms $|A_iG_i|$ are small, both c and w are small and the maximum allowable spreading period for stability is larger than it would otherwise be. These observations are demonstrated numerically by Haberman (1994a) and Haberman & Wong (1997) for an AR(1) and MA(1) rate of return process respectively.

4. The second moment of the fund level in the limit depends on the spread period through Q alone and not through c, w and $\{z_i\}$. The structure of the limiting second moments (equation (3.114)) of the pension funding process for $p = 1, 2$ is retained for $p \in \mathrm{N}$, $p < \infty$ and for $\{\delta(t)\}$ with small autocovariance over time. One may therefore surmise that the conclusion from Haberman's (1994a) numerical analysis concerning the existence of an 'optimal' range of spreading periods when rates of return are AR(1, 2) holds generally for $p \in \mathrm{N}$, $p < \infty$, viz. the 'optimal' range reduces as the variance of the rate of return increases and as the rate of return is more positively correlated. The numerical investigations of Haberman & Smith (1997) using Wilkie's (1995) asset model (essentially a linear autoregressive model—see §2.6) do indeed reveal the existence of an 'optimal' range of spreading periods. This, combined with the conclusions of Dufresne (1986, 1988) and Haberman & Wong (1997) for different rate of return models, provides a strong indication that shorter spreading periods for surpluses and deficits are called for when asset returns are very variable and/or very positively correlated over time.

3.4 Perturbations in the Funding Process

3.4.1 Some Random Disturbances

The only source of uncertainty included in the previous sections was the investment return process. Other possible perturbations in the pension system include

- contributions differing from the actuary's recommended contribution;
- accounting or valuation errors at each valuation exercise;
- random variation in benefit outgo resulting from changes in plan membership.

These may be viewed as additive noisy perturbations in the pension funding system. The first two items are dealt with in §3.4.3 while variable pension plan populations are considered in §3.4.4. It is of interest to verify how robust Dufresne's (1986, 1988) conclusions are when perturbations are introduced.

3.4.2 Discretionary Contributions

Of the possible reasons for including additive perturbations in the model, variations in sponsor contributions are the least often considered. The contribution recommended by an actuary after an ongoing or management valuation is not legally binding on the plan sponsor in many jurisdictions. (Matters may be different when statutory compliance valuations are performed.) A legitimate motivation for advance funding of pension benefits is that it provides flexibility to sponsors. Flexibility arises from the sponsor having some limited control on the timing of their contributions to provide retirement benefits (§2.2.5). Indeed, Shapiro (1983) defines flexibility as the ability of the plan sponsor to control the *incidence* of contributions. Flexibility arises, to some extent, from all aspects of the funding and contribution policies adopted by the pension plan. Pension funding methods can themselves be seen as "eminently flexible financing approaches developed to provide pension benefits" (Paquin, 1975). A major source of flexibility is in the choice of periods over which initial unfunded liabilities and gains/losses are amortized [McGill (1964:319), Snelson (1970:¶33), Humphrey *et al.* (1970:§8), Shapiro (1983)]. The (direct or indirect) influence of the sponsor in the choice of funding method and in the choice of valuation basis may also afford him some flexibility (this is discussed further in §5.2.4).

Discretionary extra contributions (positive or negative) from the plan sponsor are a result of the requirement for flexibility. Trowbridge (1966) states that:

> "Another desirable characteristic is an element of flexibility, or employer control [...] such that an additional contribution can be made when a surplus of cash is on hand, and a reduced contribution can be made when money is hard to find."

Snelson (1970:¶6) is of a similar opinion:

> "If funding is well established it is possible to arrange, within limits for the employer's contribution to match his ability to pay. Thus in a year when profits are high he can make an extra contribution and when profits are low the contributions can be reduced."

Bassett [Griffin (1966): discussion] believes that the sponsor should have *greater control* on its contributions according to its financial needs, especially if extraordinary circumstances apply:

> "The actual funding of the plan to meet the objectives must take into account how much the [sponsoring] company may desire to contribute at any particular time. This is usually expressed by the flexibility that the company has in determining the contributions in any year or period of years. [...]
>
> Flexibility in determining the current level of contributions is vitally important for most companies. We have seen it in the casualty insurance companies, where contributions were decreased in years of sizable losses such as occurred when [hurricanes] Hazel and Betsy created such havoc. We have seen it in utility companies who have made extra contributions when the revenues were unusually large and profits greater following cold winters or hot summers. We saw it in the case of steel companies who reduce contributions in the years in which they had prolonged industrial strikes. Flexibility in the amount of company contributions has been an important factor in the companies' financial planning.
>
> It is fine to set forth long-range objectives and goals based upon level of funding and adequacy of the fund, but let it remain with the company to plot its own course to reach these objectives. [...] A company should be left free to establish its own patterns to reach these goals and even to change them as conditions change in the future."

Statute and regulations nevertheless do limit the flexibility allowed in the timing of sponsor contributions, as noted by McGill *et al.* (1996:342):

> "Not only is the overall costs of retirement programs important, but the timing of the costs can be equally important. [...] Defined benefit plans offer sponsors [...] flexibility in that they do not require that maximum contributions be made to the plan every year. Again, contributions can be larger in successful years and smaller in less successful ones, although the range of flexibility in this regard has been considerably restricted in recent years by changes in the laws governing pension funding [...].

Discretionary sponsor contributions are called "contribution variances" by Winklevoss (1993:97), "extra contributions" by Bassett [Griffin (1966): discussion] and Snelson (1970), while Trowbridge (1966) refers to "additional contribution" or "reduced contribution" depending on the sponsor's financial situation. For modelling and projection purposes, it is appropriate to recognise that actual, flexible contributions may differ from the actuary's recommended contributions and explicitly deal with this in the actuarial planning and control of pension funds. Any attempt to model discretionary extra contributions (positive or negative) that a sponsor may make is limited by the fact that it is extremely difficult to model corporate cash flows separately. Indeed, corporate finance theory indicates that it may not be meaningful to distinguish between the cash flows of a defined benefit pension fund and those of the corporate sponsor (see also §5.2.4): the balance sheet of the pension fund can properly be considered as part of the "extended balance sheet" of the corporation (Bagehot, 1972). Nevertheless, it may be worthwhile to characterise discretionary contributions if only for the purpose of examining their effect on the management of pension funds.

3.4.3 Pension Funding under Additive Stochastic Disturbances

Discretionary extra contributions and random valuation errors may be represented in a simple way by an additive 'noisy' input, $\epsilon(t)AL$, at the start of year $(t, t+1)$, so that the net cash flow into the pension fund is

$$NC + adj(t) - B + \epsilon(t)AL. \tag{3.117}$$

The fund grows according to the modified recurrence relationship

$$f(t+1) = u(t+1)[(1-k)f(t) + (k-d)AL + \epsilon(t)AL], \tag{3.118}$$

(cf. equation (3.45)) so that

$$f(t) = f_0(1-k)^t \prod_{j=1}^{t} u(j) + \sum_{j=0}^{t-1} (1-k)^{t-j-1}(k-d+\epsilon(j))AL \prod_{\tau=j+1}^{t} u(\tau) \tag{3.119}$$

(cf. equation (3.47)).

Model for $\epsilon(t)$

A possible model for $\epsilon(t)$ is that of a single input ϵ occurring at time t, termed a 'spike' or 'impulse' input [Balzer & Benjamin (1980), Fujiki (1994), Khorasanee (1997)], so that $\epsilon(t) = \epsilon\delta(t)$, where $\delta(t)$ is the Kronecker delta function. A surplus or deficit is instantaneously created and will be spread into the future. It is trivial that only transient effects are produced. On the other hand, the effect of a constant non-zero input—to which Benjamin (1984, 1989) refers as a 'step' input—or of a random input will not be transient. The additive input $\epsilon(t)$ is projected to be as follows:

Projection Assumption 3.2

$\epsilon(t)$ follows a zero-mean stationary autoregressive process of order p, $AR(p)$,

$$\epsilon(t+1) = \varphi_1\epsilon(t) + \varphi_2\epsilon(t-1) + \cdots + \varphi_p\epsilon(t-p+1) + e(t+1), \tag{3.120}$$

where $\{e(t)\}$ is a sequence of unbiased independent and identically normally distributed variables. $\epsilon(s)$ and $i(t)$ are independent $\forall\ t,\ s$. $\mathrm{E}\epsilon(t) = 0$ and $\mathrm{Var}\epsilon(t) = \sigma_\epsilon^2$

This is motivated by the following:

1. Valuation errors are likely to be random and non-systematic: a zero-mean sequence of independent and identically distributed random variables is a special case of Projection Assumption 3.2.

2. If $\epsilon(t)$ represents volatility in benefit outgo, possibly as a consequence of random demographic changes in the makeup of the pension plan *membership*, it is not clear that $\epsilon(t)$ will be correlated with the rate of return on the pension *fund*. If variation in benefit payments occur because of random price and salary inflationary effects, then this may be correlated with the return on the pension fund, although this is ignored here, as a first approximation. In either case, serial dependence in $\epsilon(t)$ is plausible.

3. $\epsilon(t)AL$ may also represent the discretionary extra contribution (in addition to the actuary's recommended overall contribution) made by the sponsor to the pension fund at the start of year $(t,\ t+1)$. The actual 'flexible' contribution paid by the sponsor at the beginning of year $(t,\ t+1)$ is $c(t) + \epsilon(t)AL$, where $c(t) = NC + adj(t)$ is the actuary's recommended contribution. Corporate finance theory predicts that sponsors will seek to influence the management of the pension fund to their financial advantage and VanDerhei & Joanette (1988) describe the influence of the sponsor as "an interjection, not usual but not infrequent". Discretionary contributions (positive or negative) depend on the finances of the company, on its financial planning, on other capital investment projects arising within the firm as well as on taxation issues. Although the rate of investment return on the pension fund may be correlated with the financial performance of the sponsor, through macroeconomic cyclical effects, there is only an indirect dependence between them. The financial performance of the sponsoring company is likely to depend on its performance in previous years and there may therefore be some dependence between discretionary contributions over time.

It was shown in §§3.3.2 and A.1 that for $\epsilon(t)$ following a stationary AR(p) process (as in equation (3.120)),

$$\mathrm{Cov}[\epsilon(t), \epsilon(t-s)] = \sigma_\epsilon^2 \sum_{i=1}^{p} A_i G_i^s, \tag{3.121}$$

for $s \geq 0$ where $\{G_i\}$ is the set of *distinct* roots of

$$z^p - \varphi_1 z^{p-1} - \varphi_2 z^{p-2} - \cdots - \varphi_p = 0 \tag{3.122}$$

and $\{A_i\}$ is defined by $\mathrm{Cov}[\epsilon(t), \epsilon(t-s)]$ for $s \in [0, p-1]$ characterising the stationary AR(p) process (3.120). Furthermore, $\sum A_i = 1$. Stationarity implies that

$$|G_i| < 1 \quad \text{for } i \in [1, p]. \tag{3.123}$$

First and Second Moments

The following results may be shown to hold when the additive random input $\epsilon(t)$ is as in Projection Assumption 3.2. Provided $|u(1-k)| < 1$,

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL, \tag{3.124}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC. \tag{3.125}$$

Provided also that $q(1-k)^2 < 1$,

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \frac{\sigma^2 v^2 AL^2}{1-q(1-k)^2} + \frac{\sigma_\epsilon^2 AL^2 q}{1-q(1-k)^2} \sum_{i=1}^{p} A_i \frac{1+u(1-k)G_i}{1-u(1-k)G_i}, \tag{3.126}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t) = k^2 \lim \mathrm{Var}f(t). \tag{3.127}$$

The limit first moments in equations (3.124) and (3.125) are obvious and depend crucially on the independence between rates of investment return and $\epsilon(t)$ over time. $f(t)$ depends on $u(s)$ ($s \leq t$) and $\epsilon(j)$ ($j \leq t-1$) as is clear from equation (3.119). $u(t+1)$ is independent of $u(s)$ and $\epsilon(s)$ for $s < t+1$ and thus $u(t+1)$ and $f(t)$ are independent. From equation (3.118),

$$\mathrm{E}f(t+1) = u[(1-k)\mathrm{E}f(t) + (k-d)AL], \tag{3.128}$$

and since $u(k-d)/[1-u(1-k)] = 1$, one obtains the result of equation (3.124). The limit of the first moment of $c(t)$ is immediate upon application of equation (3.44). It may easily be shown that, if $\mathrm{E}\epsilon(t) = \epsilon \neq 0$, then

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL + \epsilon AL/(k-d), \tag{3.129}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC - k\epsilon AL/(k-d), \tag{3.130}$$

which shows (unsurprisingly) that systematic negative discretionary extra contributions from the sponsor or errors in the valuation process ought to be avoided as they result in a persisting deficit. The second moment results in equations (3.126) and (3.127) are also straightforward: refer to Appendix A (§A.2).

Some Remarks

The conditions for stability in the first and second moments of the pension funding process do not change with the introduction of a stationary additive input. The maximum allowable spread periods implied by these conditions remain the same, irrespective of the variability in discretionary extra contributions or benefit outgo.

It is easy to verify that equation (3.126) may be rewritten as

$$\lim_{t\to\infty} \mathrm{Var} f(t) = \frac{\sigma^2 v^2 AL^2}{1-q(1-k)^2} + \frac{\sigma_\epsilon^2 AL^2 u^2}{1-u^2(1-k)^2} \sum_{i=1}^{p} A_i \frac{1+u(1-k)G_i}{1-u(1-k)G_i}$$
$$+ \frac{\sigma^2 \sigma_\epsilon^2 AL^2}{[1-u^2(1-k)^2][1-q(1-k)^2]} \sum_{i=1}^{p} A_i \frac{1+u(1-k)G_i}{1-u(1-k)G_i}, \tag{3.131}$$

where the first term on the right hand side represents variance due to $i(t)$ alone and the second term represents variance due to $\epsilon(t)$ alone. The first term is identical to the result obtained by Dufresne (1988) (equation (3.58)).

If surpluses and deficits are not spread forward ($m = k = 1$), then since $\sum_i A_i = 1$,

$$\lim_{t\to\infty} \mathrm{Var} f(t) = \lim_{t\to\infty} \mathrm{Var} c(t) = [\sigma^2 v^2 + \sigma_\epsilon^2 q] AL^2, \tag{3.132}$$

and the variance of the funding process depends on the variance, but not the lagged autocorrelations, of the perturbations.

If the noisy input $\{\epsilon(t)\}$ is a sequence of unbiased independent and identically distributed random variables, then

$$\lim_{t\to\infty} \mathrm{Var} f(t) = \frac{(\sigma^2 v^2 + \sigma_\epsilon^2 q) AL^2}{1-q(1-k)^2}, \tag{3.133}$$

$$\lim_{t\to\infty} \mathrm{Var} c(t) = k^2 \lim \mathrm{Var} f(t), \tag{3.134}$$

as $G_i \to 0$ and $\sum_i A_i = 1$. A consequence of the random noisy input has been to increase the variability in the fund and contribution levels, as can be anticipated. It is immediately obvious that $\lim \mathrm{Var} f(t)$ increases monotonically as m increases (or k decreases). By comparison with the result of Dufresne (1986, 1988) (Result 3.4), $\lim \mathrm{Var} c(t)$ decreases as m increases, for $1 \le m \le m^*$, and increases as m increases, for $m > m^*$, where m^* is as in Result 3.4. Dufresne's (1986) 'optimal' range of spreading periods therefore exists.

The situation where investment returns are certain and $\{\epsilon(t)\}$ is a sequence of unbiased independent and identically distributed random variables is examined by Dufresne (1994:92): from equations (3.133) and (3.134) with $\sigma = 0$, it is not difficult to observe that the 'optimal' range of spread periods is one such that $d(1+v) \le k \le 1$ (i.e. $\lim \mathrm{Var} c(t)$ has a minimum at $k = 1 - 1/u^2 = d(1+v)$ and $\lim \mathrm{Var} f(t)$ increases monotonically with increasing m). (Dufresne (1994:90) also shows that if the *only* random influence is the additive perturbation *and* if the

φ_ϵ	$\sigma_\epsilon = 0.1$	$\sigma_\epsilon = 0.25$
-0.95	20.073	21.717
-0.8	20.029	21.079
-0.4	19.874	20.195
0	m^* =19.612	m^* =19.612
0.4	19.020	18.828
0.8	15.991	15.530
0.95	5.280	2.810

Table 3.1: Spread period for which $\lim \text{Var}c(t)$ is minimised for various σ_ϵ and φ_ϵ. $i_v = \text{E}i(t) = 3\%$, $\sigma = 0.1$.

pension fund makes no investment return ($u = 1$, w.p. 1) (assuming $k \neq 1/\ddot{a}_{\overline{m}|}$), $\lim \text{Var}f(t)$ increases and $\lim \text{Var}c(t)$ decreases as k decreases over the whole of a certain stable range.)

Result 3.4 [Dufresne (1986, 1988)] concerning an efficient or 'optimal' spread period range does not hold when $\{\epsilon(t)\}$ is autocorrelated: if there does exist an efficient range, then it is different because of the second term on the right hand side of equation (3.126). This may be investigated further by specifying that the additive noisy term $\epsilon(t)$ follows a stationary AR(1) process. Let $\varphi_1 = \varphi_\epsilon$ in equation (3.120). Equation (3.122) has only one root $G = \varphi_\epsilon$ and $A = 1$. Hence,

$$\lim_{t\to\infty} \text{Var}f(t) = \frac{\sigma^2 v^2 AL^2}{1 - q(1-k)^2} + \frac{\sigma_\epsilon^2 AL^2 q[1 + u(1-k)\varphi_\epsilon]}{[1 - q(1-k)^2][1 - u(1-k)\varphi_\epsilon]}, \tag{3.135}$$

$$\lim_{t\to\infty} \text{Var}c(t) = k^2 \lim \text{Var}f(t). \tag{3.136}$$

Assuming stability in the first moment of the funding process and given that $\{\epsilon(t)\}$ is stationary, $|u(1-k)\varphi_\epsilon| < 1$. For $i > -100\%$, Haberman (1992a) shows that $0 < u(1-k) < 1$. $\lim \text{Var}f(t)$ and $\lim \text{Var}c(t)$ (in equations (3.135) and (3.136) respectively) increase as σ_ϵ and/or φ_ϵ increase. The more positively autocorrelated the noisy additive disturbance, the more variable pension fund and contribution levels become. Numerical tests show that, in most practical circumstances, $\lim \text{Var}f(t)$ increases as m increases whereas $\lim \text{Var}c(t)$ exhibits a minimum. For small autocovariance (small $|\varphi_\epsilon|$), the 'optimal' spread period range does not change much: the spread period for which $\lim \text{Var}c(t)$ is minimised remains fairly close to $m^* = 19.612$ years in the example illustrated in Table 3.1. The 'optimal' spread period range appears to reduce as additive perturbations become more positively autocorrelated.

3.4.4 Variable Pension Plan Population

Some demographic pension plan population assumptions were considered in §2.5. In this section, two membership projection models are considered: a deterministic stable membership and a model with random new entrants. We restrict to the simple pension plan described in §2.7 (with suitable modifications). Only one entry age, one retirement age and simple (normal retirement) benefits are allowed. This is similar to the model of Bowers *et al.* (1976). (Several entry and retirement ages as well as grades of employees may be considered, possibly in the manner of Giesecke (1994).)

Deterministic Stable Plan Population

The population distribution function $g(t + r - x)$ as described in §2.7 may be used to describe new entrant numbers into the plan. Suppose that $g(t + r - x)$ is deterministic and time-variant. The actuarial liability $AL(t)$, normal cost $NC(t)$ and benefit outgo $B(t)$ in equations (3.20), (3.21) and (3.22) respectively are time-variant. The "liability growth equation" (3.24) of Bowers *et al.* (1976) applies. For the sake of simplicity, Projection Assumption 2.3 is replaced by the following:

PROJECTION ASSUMPTION 3.3 (INVESTMENT RATE OF RETURN)
The pension fund is invested at a risk-less (certain and constant) logarithmic rate of return δ (net of salary inflation), equal to the valuation discount rate. $1 + i(t) = u(t) = e^{\delta}\ \forall t$.

It is not difficult to see that when the "liability growth equation" (3.24) (with $1 + i = e^{\delta}$) is replaced into equation (3.41) (with $u(t + 1) = e^{\delta}$),

$$ul(t+1) = e^{\delta}(1-k)ul(t) \quad \text{or} \quad ul(t) = [e^{\delta}(1-k)]^{t} ul(0). \tag{3.137}$$

If $|e^{\delta}(1 - k)| < 1$, then $ul(t) \to 0$, $f(t) \to AL(t)$ and $c(t) \to NC(t)$, as $t \to \infty$.

Suppose furthermore that the pension plan is fully funded from the outset, $ul(0) = 0$. No gain or loss emerges and, from equations (3.137) and (3.39),

$$ul(t) = 0, \quad f(t) = AL(t), \quad c(t) = NC(t), \quad \text{for } t \geq 0. \tag{3.138}$$

The pension plan membership is now projected to be *stable*. This model is standard in demographic studies and has been employed by Bowers *et al.* (1976, 1979, 1982) and Winklevoss (1993:58) in pension plan modelling. Such a model allows for an exponentially growing or decaying membership, representing the pension plan in an industry or firm in a state of expansion or contraction. Membership Projection Assumption 2.2 may now be replaced by the following:

PROJECTION ASSUMPTION 3.4 (MEMBERSHIP)
The pension plan membership is stable. *The new entrant function is defined as follows:* $g(t + r - x) = \exp[(t + r - x)\delta_g]$.

The number of plan members aged x therefore grows from one year to the next by a logarithmic growth factor of δ_g. In fact, the pension plan membership as a whole grows by δ_g every year, since the total number of members at time t is

$$\sum_{x=a}^{w} g(t+r-x)l_x = e^{t\delta_g} \sum_{x=a}^{w} e^{(r-x)\delta_g} l_x. \tag{3.139}$$

I shall make a convenient abuse of economic terminology by referring to δ_g as the (logarithmic) "rate of demographic inflation". The stationary membership of Projection Assumption 2.2 is a special case of the stable membership of Projection Assumption 3.4 with no growth, $\delta_g = 0$.

Since $g(t+1+r-x) = e^{\delta_g} g(t+r-x)$, it is readily observed from equations (3.20), (3.21) and (3.22) that $AL(t+1) = e^{\delta_g} AL(t)$, $NC(t+1) = e^{\delta_g} NC(t)$ and $B(t+1) = e^{\delta_g} B(t)$. Pension liabilities are therefore growing at logarithmic rate δ_g. From equation (3.138), it also follows that $f(t+1) = e^{\delta_g} f(t)$. By virtue of the asset growth recurrence relation (3.27),

$$f(t) = e^{\delta - \delta_g}[f(t) + c(t) - B(t)]. \tag{3.140}$$

Recall that δ is the risk-less rate of return on pension fund assets net of salary inflation and δ_g is the constant rate of growth in new entrants. Final-salary pensions are accrued (Plan Assumption 2.3) and pensions in payment are assumed to be linked with salary inflation (Plan Assumption 2.4). Pension liabilities (in nominal terms) therefore grow through both salary inflation and 'demographic inflation'. $\delta - \delta_g$ may be usefully regarded as the rate of return on the fund net of both economic and 'demographic' inflation. Consider the following three scenarios:

1. $\delta > \delta_g$. The rate of return net of salary inflation on the fund exceeds the rate of growth of the pension plan population. It follows from equation (3.140) that $c(t) < B(t)$: contributions and interest payments from the fund meet the benefit outgo and maintain the fund at a *stable* level.

2. $\delta = \delta_g$. The nominal rate of return on pension fund assets equals the rate of growth of pension liabilities (in terms of salary inflation and membership expansion). This implies that $c(t) = B(t)$: the contribution paid every year equals the benefit outgo. This is the situation of a pay-as-you-go or unfunded arrangement. The return on pension fund assets net of economic and 'demographic' inflation is nil.

3. $\delta < \delta_g$. Finally, if the rate of return on pension fund assets net of salary inflation is less than the rate of growth of the pension plan population, it follows from equation (3.140) that $c(t) > B(t)$. The required contribution is then greater than under an unfunded arrangement. It is cheaper not to provide advance funding for benefits. Assets provide negative returns net of economic and 'demographic' inflation. It is necessary to contribute more than the benefit outgo in order to maintain the fund in real (economic and 'demographic') terms.

These results are well-known and sufficiently important in the mathematical theory of funding to warrant their inclusion here:

- Bowers *et al.* (1976) use a continuous-time and deterministic pension plan model to demonstrate the result that funded plans require less contribution than pay-as-you-go plans when real (net of salary inflation) returns are greater than the rate of growth of plan population.

- A related result has been shown in the pension economics literature in the context of state pension schemes by Aaron (1966).

- A corollary is also reported in the actuarial literature by Ammeter (1963): if the real rate of return (net of salary inflation) is zero and the population is stationary (deterministic), then contribution in a funded plan is the same as in a pay-as-you-go (unfunded) situation, independently of the funding method.

- Similar concepts are expressed by Taylor (1987:§3) through the proposition that "if the rate of expansion of the target fund exceeds the rate of investment income" a smaller premium is required in an unfunded situation than in a funded system.

Random New Entrants

The deterministic stable membership model above is a suitable model for immature or young plans or plans associated with declining industries. Most pension plans achieve some form of maturity with on average stationary memberships at some point. It is then more relevant to consider the effect of random numbers of new entrants on the pension fund and contribution levels. Mandl & Mazurová (1996) (in a state pension plan model) allow for new entrant numbers that vary as a Gaussian stationary autoregressive process, which admits *negative* new entrants, albeit with a small probability. The probability distribution of new entrants is not specified in this section and the size of the membership at various ages is not correlated.

It is not clear that there is a direct and immediate statistical dependence between the new membership of a company's pension *plan* and the returns achieved by its pension *fund*. Data concerning the possible correlation of recruitment level and pension benefits (let alone the level of funding for pensions) is sparse. It is possible nevertheless to envisage a dependence of both investment returns and the recruitment of the sponsoring company on the macroeconomic cycle. In addition, trends in the membership (e.g. whether it is ageing) will influence long-term (strategic) investment. Here we assume that there is *no* dependence between the new entrant function $g(t)$ and the rate of return $i(s)$ for all t, s. The model of §2.7 is assumed, including random rates of return (Projection Assumption 2.3). Membership is projected as follows:

PROJECTION ASSUMPTION 3.5 (MEMBERSHIP)
The population distribution function $g(t+r-x)$ is assumed to be a sequence of independent and identically distributed random variables. $\mathrm{E}g(t+r-x) = g$, $\mathrm{Var}g(t+r-x) = \sigma_g^2$.

Furthermore, $\{g(t)\}$ *and* $\{i(s)\}$ *are independent* $\forall\ t, s$.

Since surpluses and deficits are being spread forward, the fund level evolves according to equations (3.40) and (3.43):

$$f(t+1) = u(t+1)[(1-k)f(t) + NC(t) + kAL(t) - B(t)], \tag{3.141}$$

$$f(t) = f_0(1-k)^t \prod_{j=1}^{t} u(j) + \sum_{j=0}^{t-1} (1-k)^{t-j-1}[NC(j) + kAL(j) - B(j)] \prod_{\tau=j+1}^{t} u(\tau). \tag{3.142}$$

It is shown in Appendix A (§A.3) that, provided $|u(1-k)| < 1$,

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL, \tag{3.143}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC, \tag{3.144}$$

where $AL = \mathrm{E}AL(t)$ and is the actuarial liability if the number of new entrants were constant at $\mathrm{E}g(t) = g$. Similarly, $NC = \mathrm{E}NC(t)$. Define

$$S_1(k) = \sum_{x=a}^{w} \sum_{y=a}^{w} [u(1-k)]^{|x-y|} l_x l_y [NC_x + kAL_x - B_x][NC_y + kAL_y - B_y], \tag{3.145}$$

$$S_2(k) = \sum_{x=a+1}^{w} \sum_{y=a}^{x-1} [u(1-k)]^{x-y} l_x l_y [NC_x + kAL_x][NC_y + kAL_y - B_y], \tag{3.146}$$

$$S_3(k) = \sum_{x=a}^{w} l_x^2 [NC_x + kAL_x]^2. \tag{3.147}$$

Provided that $q(1-k)^2 < 1$,

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \frac{\sigma^2 v^2 AL^2 + \sigma_g^2 q S_1(k)}{1 - q(1-k)^2}, \tag{3.148}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t) = k^2 \lim \mathrm{Var}f(t) - 2\sigma_g^2 k(1-k)^{-1} S_2(k) + \sigma_g^2 S_3(k). \tag{3.149}$$

The conditions for stability of the first and second moments of the funding process are unchanged from Result 3.1. These impose a maximum spread period, which is unaffected by the variability of the random membership.

The dependence of the variance of the fund level in equation (3.148) on k (or m) changes as compared to equation (3.58) and to equation (3.133) (where additive uncorrelated perturbations were assumed) even though the numbers of new entrants are uncorrelated over time. This is because the actuarial liability function (as well as the normal cost and benefit outgo) is similar to a weighted moving average process and is autocorrelated in a given interval if any cohort remains part of the plan membership during that interval.

	$\sigma_g = 0$		$\sigma_g = 0.5$	
m	R.S.D. $f(t)$	R.S.D. $c(t)$	R.S.D. $f(t)$	R.S.D. $c(t)$
5	0.0832	0.8743	0.1156	0.8810
10	0.1211	0.6980	0.1453	0.7051
15	0.1557	0.6550	0.1755	0.6623
$m^* = 17$	0.1694	0.6513	0.1879	0.6587
20	0.1902	0.6545	0.2069	0.6619
25	0.2260	0.6765	0.2404	0.6838
30	0.2642	0.7145	0.2769	0.7216
35	0.3058	0.7661	0.3170	0.7730

Table 3.2: Relative standard deviations of $f(t)$ and $c(t)$ in the limit for various spreading periods when the *Unit Credit* funding method is used.

The variance of the fund level in the limit in equation (3.148) may be rewritten as

$$\lim_{t\to\infty} \text{Var}f(t) = \frac{\sigma^2 v^2 AL^2}{1-q(1-k)^2} + \frac{\sigma_g^2 u^2 S_1(k)}{1-u^2(1-k)^2} + \frac{\sigma^2\sigma_g^2 S_1(k)}{[1-u^2(1-k)^2][1-q(1-k)^2]}, \quad (3.150)$$

where the first term on the right hand side represents variance due to random investment returns alone and the second term on the right hand side represents variance due to random membership alone.

The variance of the fund and contribution levels due to random membership depends explicitly on the pension funding method used, since $S_1(k)$ in equation (3.148) depends on the individual AL_x, NC_x and B_x functions, as defined in §3.1.2 for example. It is typical that demographic variation is small compared to investment variability ($\sigma_g \ll \sigma$) in which case it is clear that the first term on the right hand side of equations (3.148) and (3.149) will dominate and the efficient spread period range noted by Dufresne (1988) is likely to emerge. Some numerical results based on the *Unit Credit* method are shown in Table 3.2. (The relative standard deviation or coefficient of variation of $f(t)$ is $\sqrt{(\text{Var}f(t))}/\text{E}f(t)$.) These results indicate that fund and contribution levels become more variable as a result of random new entrants; and that fund levels become increasingly variable as surpluses and deficits are spread over longer periods, whereas the variability of contribution levels is minimised at a certain spreading period. The 'optimal' range of spreading periods, in the sense defined by Dufresne (1986), does not seem to change very much when random new entrants are permitted. Similar results are obtained when the *Entry Age* method is used, as displayed in Table 3.3 on the next page. The numerical assumptions and calculations pertaining to Tables 3.2 and 3.3 are described in Appendix A (§A.3).

m	$\sigma_g = 0$ R.S.D. $f(t)$	R.S.D. $c(t)$	$\sigma_g = 0.5$ R.S.D. $f(t)$	R.S.D. $c(t)$
5	0.0832	1.3097	0.1132	1.1316
10	0.1211	1.0456	0.1434	1.0513
15	0.1557	0.9812	0.1738	0.9869
$m^* = 17$	0.1694	0.8757	0.1863	0.9814
20	0.1902	0.9805	0.2055	0.9861
25	0.2260	1.0135	0.2392	1.0192
30	0.2642	1.0703	0.2757	1.0760
35	0.3058	1.1476	0.3160	1.1534

Table 3.3: Relative standard deviations of $f(t)$ and $c(t)$ in the limit for various spreading periods when the *Entry Age* funding method is used.

3.5 Amortizing Gains/Losses over a Fixed Term

3.5.1 A Supplementary Funding Method

A different supplementary funding method from the one defined in §3.2 and employed in §§3.3 and 3.4 is now considered. In the United States and Canada, the practice is to amortize the actuarial gains and losses directly and over a fixed term. The initial unfunded liability is also separately and explicitly amortized.

$$adj(t)_a = P(t) + \sum_{j=0}^{m-1} l(t-j)/\ddot{a}_{\overline{m}|}, \tag{3.151}$$

where $l(t) = 0$ for $t \leq 0$ and is defined in equation (3.37). $P(t)$ is the amortization payment for any initial unfunded liability (see equation (3.17)) (in our model fund, an initial unfunded liability may only arise at $t = 0$ given Modelling Assumptions 2.2 and 2.3). Both the initial unfunded liability and the set of intervaluation losses are amortized entirely, the former within n years and the latter within m years. McGill *et al.* (1996:525) talk of a "*direct* method of determining and dealing with actuarial gains and losses" in contrast to the "*spread* method of dealing with gains and losses" in §3.2.

REMARK 3.1 *The subscript a hereafter denotes this method and the subscript s refers to the method of spreading surplus and deficits over a moving term, as described in §3.2.*

The U.S. Employee Retirement Income Security Act, 1974 (ERISA) requires that actuarial gains and losses be amortized over not more than 5 years for single-employer pension plans, and 15 years for multi-employer plans (the maximum amortization period for initial unfunded liabilities arising from changes in actuarial assumptions, plan amendments etc. is different). [Source: McGill *et al.* (1996:597)]

The simpler case where $m = n$ is considered by Dufresne (1989), such that we can put $l(0) = ul_0$, $l(t) = 0$ for $t < 0$ and $P(t) = 0\ \forall t$. It can be shown formally (see Dufresne (1994)) that as long as the initial unfunded liability is amortized over a finite period, the ultimate behaviour of our model fund is independent of that amortization period. Clearly, if $m = n = 1$, this method of adjusting the normal cost is identical to the surplus/deficit spreading method above.

3.5.2 Moments of the Funding Process

The first and second moments of $f(t)_a$ and $c(t)_a$ are derived by Dufresne (1989). If the method is applied according to equation (3.151), and in particular if the valuation discount rate is the same as the mean long-term rate of return (Valuation Assumption 2.2), then it can be shown (by comparison with Dufresne (1994:122)) that, for $n < \infty$,

$$\mathrm{E}f(t)_a = \begin{cases} AL - ul_0\ \ddot{a}_{\overline{n-t}|}/\ddot{a}_{\overline{n}|}, & 0 \leq t \leq n-1, \\ AL, & t \geq n, \end{cases} \tag{3.152}$$

$$\mathrm{E}c(t)_a = \begin{cases} NC + ul_0/\ddot{a}_{\overline{n}|}, & 0 \leq t \leq n-1, \\ NC, & t \geq n. \end{cases} \tag{3.153}$$

This is similar to the first moment results in Proposition 3.1, when surpluses/deficits are spread forward over m years and the initial unfunded liability is amortized over a fixed term of n years. If there is no stochastic variation and if the rate of return is the same as the valuation discount rate, then there is no difference between amortizing gains/losses and spreading surpluses/deficits, provided the initial unfunded liability is also amortized. When $m = 1$, the two methods of spreading surpluses/deficits and amortizing gains/losses are in fact identical and

$$\lim \mathrm{Var}f(t)_s = \lim \mathrm{Var}f(t)_a = \lim \mathrm{Var}c(t)_s = \lim \mathrm{Var}c(t)_a = \sigma^2 AL^2 v^2. \tag{3.154}$$

Differences arise when the second moments are considered. Only the asymptotic situation when $t \to \infty$ is explored. For $m > 1$, if $\sigma^2 \sum \beta_j^2 < 1$, Dufresne (1989) gives

$$\lim_{t\to\infty} \mathrm{Var}f(t)_a = \frac{\sigma^2 AL^2 v^2 \sum \lambda_j^2}{1 - \sigma^2 \sum \beta_j^2}, \tag{3.155}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t)_a = \frac{\sigma^2 AL^2 v^2 m}{(1 - \sigma^2 \sum \beta_j^2)\ddot{a}^2_{\overline{m}|}}, \tag{3.156}$$

where $\beta_j = a_{\overline{m-j}|}/\ddot{a}_{\overline{m}|}$ for $j \in [1, m]$ and $\lambda_j = \ddot{a}_{\overline{m-j}|}/\ddot{a}_{\overline{m}|}$ for $j \in [0, m]$. (For equations (3.155)) and (3.156)) to hold for $m = 1$, define $\ddot{a}_{\overline{0}|} = a_{\overline{0}|} = 0$.)

Note also the identity:

$$\sum \beta_j^2 = v^2 \left(\sum \lambda_j^2 - 1\right). \tag{3.157}$$

In the following section, we investigate the existence of an 'optimal' range of *amortization* periods, analogous to the 'optimal' spread period range of Dufresne (1986, 1988) (as in Result 3.4).

3.5.3 Optimal Amortization Periods

Before the tradeoff between the variance of contribution and fund levels is considered, the following are defined, based on equations (3.58), (3.59), (3.155) and (3.156):

$$\alpha_s(m) = 1 - (u^2 + \sigma^2)(1-k)^2 \tag{3.158}$$

$$\alpha_a(m) = \left(1 - \sigma^2 \sum \beta_j^2\right) / \sum \lambda_j^2 \tag{3.159}$$

$$\beta_s(m) = \left(1 - (u^2 + \sigma^2)(1-k)^2\right) / k^2 \tag{3.160}$$

$$\beta_a(m) = \left(1 - \sigma^2 \sum \beta_j^2\right) \ddot{a}^2_{\overline{m}|}/m \tag{3.161}$$

where the αs are proportional to the reciprocal of the variance of the fund level in the limit, and the βs are proportional to the reciprocal of the variance of the contribution rate level in the limit. ($\beta_a(m)$ is distinct from β_j.)

The following two inequalities are also required and are proven in Appendix B (§B.1).

$$\left(\ddot{a}^2_{\overline{m-2}|} + \cdots + \ddot{a}^2_{\overline{0}|}\right) / \ddot{a}^2_{\overline{m-1}|} < \left(\ddot{a}^2_{\overline{m-1}|} + \cdots + \ddot{a}^2_{\overline{0}|}\right) / \ddot{a}^2_{\overline{m}|}, \quad \text{for } m \geq 2. \tag{3.162}$$

$$\left(\ddot{a}^2_{\overline{m}|} + \cdots + \ddot{a}^2_{\overline{1}|}\right)\left(\ddot{a}^2_{\overline{m-1}|} + \cdots + \ddot{a}^2_{\overline{1}|}\right) > \left(\ddot{a}_{\overline{m}|}\ddot{a}_{\overline{m-1}|} + \cdots + \ddot{a}_{\overline{2}|}\ddot{a}_{\overline{1}|}\right)^2, \quad \text{for } m \geq 2. \tag{3.163}$$

From a numerical investigation, Dufresne (1986) draws the sensible conclusion that, when gains/losses are amortized, "greater emphasis is laid on security of benefits" than if surpluses/deficits are spread. This is encapsulated in the following proposition, proven in Appendix B (§B.3).

PROPOSITION 3.2 *For equal amortization and spread periods,*

$$\lim \mathrm{Var} f(t)_a < \lim \mathrm{Var} f(t)_s, \quad m > 1, \tag{3.164}$$

$$\lim \mathrm{Var} f(t)_a = \lim \mathrm{Var} f(t)_s, \quad m = 1. \tag{3.165}$$

The following proposition concerns a range of amortization period for gains and losses that is 'optimal' in the sense of Dufresne (1986). Let m_s^* (resp. m_a^*) be the spread (resp. amortization) period for which $\lim \mathrm{Var} c(t)_s$ (resp. $\lim \mathrm{Var} c(t)_a$) is a minimum. Clearly, m_s^* is related to k^* in Result 3.4. Let m_s^∞ (resp. m_a^∞) be the maximum spread (resp. amortization) period for which $\lim \mathrm{Var} f(t)_s$ (resp. $\lim \mathrm{Var} f(t)_a$) remains unbounded.

PROPOSITION 3.3 *There exists an optimal range of amortization periods* $[1, m_a^*]$*:*

if $1 < m < m_a^*$, $\lim \mathrm{Var} f(t)_a$ *increases and* $\lim \mathrm{Var} c(t)_a$ *decreases with increasing* m*;*

if $m_a^* < m < m_a^\infty$, *both* $\lim \mathrm{Var} f(t)_a$ *and* $\lim \mathrm{Var} c(t)_a$ *increase with increasing* m*;*

if $m > m_a^\infty$, *then* $f(t)_a$ *and* $c(t)_a$ *are not stationary in the limit.*

Furthermore, $m_a^* > m_s^*$ *and* $\lim \mathrm{Var} c(t)_a > \lim \mathrm{Var} c(y)_s$ *at* $m = m_s^*$.

Proof in Appendix B (§B.4).

Proposition 3.3 confirms that amortizing gains and losses over a longer period means that the fund level is more variable and pension benefits are less secure. As in the case where surpluses/deficits are being spread (Result 3.4) and somewhat counterintuitively, the contribution variability does not always decrease as the amortization period increases.

The variation of $\lim \mathrm{Var} f(t)_a$ and $\lim \mathrm{Var} f(t)_s$ v. m as well as $\lim \mathrm{Var} c(t)_a$ and $\lim \mathrm{Var} c(t)_s$ v. m may be 'sketched' using Result 3.4 and Proposition 3.3: see Figure 3.2. The $\lim \mathrm{Var} c(t)_a$ v. $\lim \mathrm{Var} f(t)_a$ curve has a minimum point at m_a^*. For amortization periods in the non-optimal range $(m_a^*, \infty]$, there will always be an amortization period in $[1, m_a^*]$ that yields the same $\lim \mathrm{Var} c(t)_a$ and a lower $\lim \mathrm{Var} f(t)_a$. The tradeoff between fund security and contribution stability breaks down for longer amortization periods.

The existence of an 'optimal' amortization period range, $[1, m_a^*]$, which is larger than the 'optimal' spread period range, has therefore been shown. The numerical test, based on small σ and i, performed by Dufresne (1989) fails to show that the tradeoff breaks down for a large enough amortization period and that an 'optimal' range does therefore exist. The purely numerical work of Cairns (1994) illustrates some of the above propositions.

Based on the numerical investigation of these authors, we deduce that a real rate of return (net of salary inflation) of 1%, with standard deviations 0.025, 0.05 or 0.1, yields a value for m_a^* of over 40 years; whereas a real rate of return of 5%, with standard deviation 0.2, yields m_a^* of about 16 years. We conclude that, under current economic conditions, the common practice of amortizing gains/losses over periods of about 5 years or less lies within the 'optimal' range.

3.5.4 Efficiency

It is clear from Result 3.1 (equation (3.54)) and Proposition 3.1 that amortizing the initial unfunded liability over a fixed term rather than spreading it over a moving term has the advantage of hastening its removal. Typically, the initial unfunded liability is amortized over n =10–30 years in North America and a spread period of about m =10–15 years is used in the U.K. Funding is also expected to be more gradual and 'smoother' if spreading is used rather than amortization in the treatment of the initial unfunded liability. The level of contributions also changes more smoothly, which may be more convenient for the plan sponsor for budgetary reasons.

Spreading surpluses and deficits over a rolling term is also different from amortizing gains and losses when stochastic variation in the pension funding process is considered and the

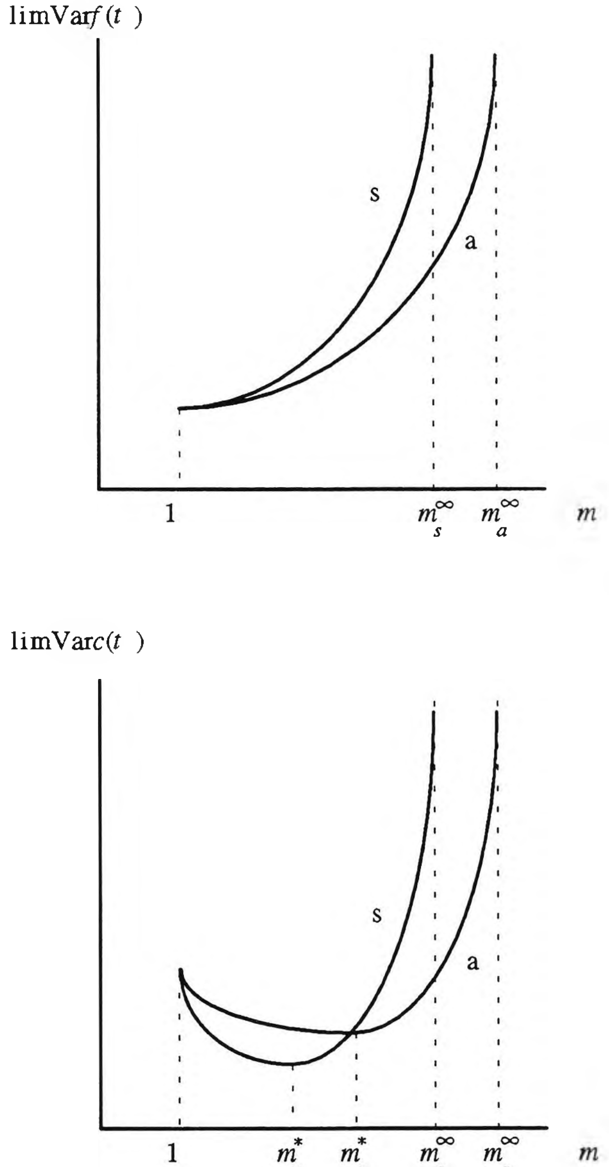


Figure 3.2: Ultimate variances v. m. 's' denotes spreading surpluses/deficits, whereas 'a' denotes the amortization of gains/losses.

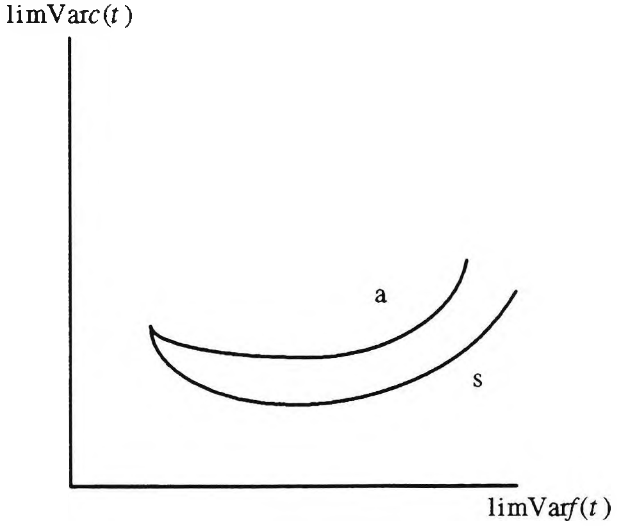


Figure 3.3: Spreading surpluses/deficits ('s') is more efficient than the amortization of gains/losses ('a').

second moments of the funding process are compared. We can now show that spreading surpluses and deficits over a moving term should be preferred to amortizing gains and losses over a fixed term, on the grounds that variability in the fund and contribution levels are minimised.

PROPOSITION 3.4 *According to the objective of minimising ultimate variances of fund and contribution levels, spreading surpluses or deficits is more efficient than amortizing gains and losses since for* $\{m_a, m_s \neq 1\}$ *such that* $\lim \mathrm{Var} f(t)_a = \lim \mathrm{Var} f(t)_s$, *then* $\lim \mathrm{Var} c(t)_a > \lim \mathrm{Var} c(t)_s$.

Proof in Appendix B (§B.5).

The $\lim \mathrm{Var} c(t)_a$ v. $\lim \mathrm{Var} f(t)_a$ curve lies above the $\lim \mathrm{Var} c(t)_s$ v. $\lim \mathrm{Var} f(t)_s$ curve (except at $m = 1$ where they coincide): see Figure 3.3. For a given ultimate variance of the fund level, spreading surpluses and deficits will always yield a lower ultimate variance of the contribution level than amortizing gains and losses (except at $m = 1$). Hence, for equivalent 'security', spreading achieves better 'stability'. This is not surprising since the amortization of gains and losses implies the use of information *delayed* by up to m_a years: the feedback of delayed information was considered in §3.2.3 and was found to lead to higher variances, less stability and less security, in the pension funding process.

Note again that spreading is implicit in the *Aggregate* and *Frozen Initial Liability* pension funding methods (§§3.2.1 and 3.2.4). Trowbridge & Farr (1976:62) describe "the smoothness of contributions, the automatic adjustment for actuarial gain or loss, and the relatively easy computations" as attractive characteristics of such methods. They also argue (p. 85) that the amortization of gains and losses becomes cumbersome if the amortization period is too long. They also state that the apparent drawback of spreading, which is that surpluses and deficits are never completely removed except in the limit, is not a weakness, since the random nature of deviations from assumed experience means that no method completely removes surpluses and deficits over a finite term. For these reasons, Trowbridge & Farr (1976:42) seem to prefer spreading over all other methods. Finally, note that the investment rates of return on the pension fund were assumed to be independent and identically distributed from year to year (Projection Assumption 2.3 is also made by Dufresne (1989)). Gerrard & Haberman (1996) obtain stability conditions for the first and second moments of the funding process when gains and losses are amortized over a fixed term and the (arithmetic) rate of return process follows a stationary autoregressive process of order 1.

3.6 Optimal Contribution and Asset Allocation

3.6.1 The Pension Fund System

Assume that the fund may be invested in two assets: a risk-less asset earning risk-free rate r and a risky asset earning $r + \alpha(t+1)$ in year $(t,\, t+1)$. $\alpha(t+1)$ is a random risk premium. Furthermore, let $y(t)$ be the proportion of the fund invested in the risky asset in year $(t,\, t+1)$, and $1 - y(t)$ be the proportion invested in the risk-less asset. The arithmetic rate of return on the fund in year $(t,\, t+1)$ is

$$r + y(t)\alpha(t+1). \tag{3.166}$$

It is further assumed that $\{\alpha(t)\}$ is a sequence of independent and identically distributed random variables over time, with mean $\alpha > 0$ and variance σ^2.

The pension fund can be considered as a random system (or 'plant'),

$$f(t+1) = [1 + r + y(t)\alpha(t+1)][f(t) + c(t) - B], \tag{3.167}$$

where the fund level $f(t)$ is a state variable and $c(t)$ and $y(t)$ are the contribution and asset allocation control variables respectively.

Optimal contributions $c^*(t)$ and asset allocation decisions $y^*(t)$ may be derived by means of dynamic programming and optimal control [Bellman (1957), Åström (1970), Bertsekas (1976), Whittle (1982)]. They are derived by Boulier *et al.* (1995) in a continuous-time model with the sole objective of minimising (mean-square) contributions. O'Brien (1986, 1987) seeks to minimise mean-square contributions as well as deviations in the fund level and assumes a finite time horizon. He also assumes a single random asset with the only control variable being plan sponsor contributions. In the context of a deterministic model of social security rather than a private pension plan, Vandebroek (1990) derives the optimal contribution with the objective of minimising (mean-square) deviations of both contribution and fund levels from respective targets (presumably normal costs and actuarial liabilities under some chosen funding method). Haberman & Sung (1994) and Sung (1997) also consider only contributions optimised over a similar criterion but in discrete-time (which is less idealised since valuations are performed at discrete intervals). Cairns (1997) extends the method of Boulier *et al.* (1995) and assumes a quadratic performance criterion comprising both contribution and fund levels. Cairns (1997) also assumes two risky and contemporaneously correlated assets. The more realistic set-up of 'incomplete state information' is studied by Sung (1997) in a discrete-time and finite-horizon model: delay and random error is allowed in the valuation process and optimal contribution strategies are derived. Sung (1997) also discusses supplementary performance criteria that incorporate a desirable pace of funding over time.

In the following, it is assumed that both asset allocation (over one risk-less and one risky asset) and contribution controls are available. The objectives of the funding process are to stabilise contributions and remove the unfunded liability over a finite horizon N. A finite horizon is useful for four reasons:

1. The control period N may be perceived as a control parameter, akin to the amortization period m in the foregoing sections, at the disposal of the actuary (Haberman, 1994b).

2. It is typically required that unfunded liabilities be removed over a finite term in most jurisdictions, e.g. under the U.S. Employee Retirement Income Security Act, 1974 (ERISA) and the Canadian Pension Benefits Standards Act, 1985.

3. A fixed control period is important when technical solvency status is material (Sung, 1997).

4. Optimal control over a finite horizon permits consideration of time-variant parameters, such as contribution and fund targets and expected return on the risky asset [Grauer & Hakansson (1985), Sung (1997)].

It is assumed that valuations are performed over discrete and regular intervals, in keeping with the previous sections:

1. It is usually a statutory requirement that valuations be performed regularly. They are not performed continuously in practice and cash flows only occur at discrete intervals.

2. No distributional assumption is placed on the rate of return process. It is only required that the rate of return be independent and identically distributed from year to year, in accordance with the Efficient Market Hypothesis. Distributions other than lognormal have been suggested for asset returns because they are leptokurtic [Daykin *et al.* (1994), Smith (1996), Finkelstein (1997)] (see §2.6.3).

Finally, additive noise in the pension fund system (3.167) is suppressed—the derivation of the optimal controls may be readily adapted to random additive perturbations. O'Brien (1987), Sung (1997) and Cairns (1997) allow for random benefit volatility and show that optimal controls are independent of the variance of additive noise, a consequence of the property of 'certainty equivalence' in additive random inputs. It is well-known that a system with random coefficients such as in equation (3.167) does not exhibit 'certainty equivalence' in terms of the random coefficients [Chow (1975:227), Bertsekas (1976:80), Sung (1997:192)], i.e. the optimal control cannot be replaced by the corresponding deterministic optimal control solution with the random coefficients (the rate of return process in equation (3.167)) being replaced by their expected values. Randomness in asset returns must therefore be characterised.

3.6.2 Optimal Control

Define W_t to be all information available up to time t, i.e. $\{f(0), \ldots, f(t), c(0), \ldots, c(t-1), y(0), \ldots, y(t-1)\}$. A control policy π then specifies $c(t)$ and $y(t)$ in terms of W_t for $1 \leq t \leq N-1$ [Bertsekas (1976:42), Whittle (1996:174)].

The performance of the pension fund may be judged in terms of the proximity of the fund and contribution variables to their desired levels. The 'cost' incurred for any such deviation at time $0 \leq t \leq N-1$ may be defined as

$$C(f(t), c(t), t) = \theta_1(f(t) - FT_t)^2 + \theta_2(c(t) - CT_t)^2 \tag{3.168}$$

In this equation, the target fund and contribution levels are FT_t and CT_t and may be time-variant: they relate to the actuarial liability and normal cost generated by the pension funding method in use. Different weights ($\theta_1 \geq 0$ and $\theta_2 \geq 0$) are placed on the twin long-term objectives of fund security and contribution stability. At the end of the 'control period', a terminal cost is incurred if an unfunded liability still exists:

$$\mathbb{C}_N = \theta_0(f(N) - FT_N)^2. \tag{3.169}$$

The performance of the fund may be given different importance over time. The *discounted* cost of deviation occurring t years ahead is $\beta^t C(f(t), c(t), t)$. $0 < \beta \leq 1$ represents a psychological discount rate [Boulier *et al.* (1995, 1996)] or risk discount rate (Cairns, 1997). It may be equal to the valuation discount rate [Haberman & Sung (1994), Sung (1997)]. The discounted cost-to-go or discounted cost incurred from time t to N is

$$\mathbb{C}_t = \sum_{s=t}^{N-1} \beta^{s-t} C(f(s), c(s), s) + \beta^{N-t}\mathbb{C}_N. \tag{3.170}$$

An objective criterion for the performance of the pension funding system may therefore be defined to be

$$\mathrm{E}[\mathbb{C}_0|W_0] = \mathrm{E}\left[\beta^N \mathbb{C}_N + \sum_{s=0}^{N-1} \beta^s C(f(s), c(s), s) \middle| W_0\right] \tag{3.171}$$

This criterion may be minimised by means of dynamic programming. This can be formulated as follows [Åström (1970), Bertsekas (1976), Whittle (1982)]. A value function may be defined as the minimum over realisable control policies of the expected cost-to-go from time t given information at t,

$$J(W_t) = \min_{\pi} \mathrm{E}[\mathbb{C}_t | W_t], \tag{3.172}$$

and the total value function is then the minimum expected cost-to-go from time 0 given information at t,

$$G(W_t) = \min_{\pi} \mathrm{E}[\mathbb{C}_0 | W_t]. \tag{3.173}$$

The Bellman optimality equation then gives:

$$G(W_t) = \min_{c(t),y(t)} \mathrm{E}[G(W_{t+1}) | W_t], \tag{3.174}$$

with the boundary condition $G(W_n) = \mathrm{E}[\mathbb{C}_0 | W_N]$, and the minimising values of $c(t)$ and $y(t)$ are then optimal controls at time t.

Two important simplifying assumptions have been made in §3.6.1:

1. *Perfect state observation:* It is assumed that the fund level is measured or observed without delay or corrupting noise. Sung (1997) considers the optimal contribution control of pension funds with incomplete information.

2. *Markov property:* Since $\{\alpha(t)\}$ is a sequence of independent and identically distributed random variables over time, with mean α and variance σ^2, $f(t)$ has the Markov property and $Pr[f(t+1)|W_t] = Pr[f(t+1)|f(t)]$.

Given the Markov property,

$$J(W_t) = \min_{\pi_t} \mathrm{E}[\mathbb{C}_t | f(t)] = J(f(t), t), \tag{3.175}$$

where the control policy π_t depends only on $c(t), \ldots, c(N-1), y(t), \ldots, y(N-1)$. Upon minimising on π_t, the value function is a function of $f(t)$ and t alone. As for the total value function,

$$\begin{aligned} G(W_t) &= \min_{\pi_t} \mathrm{E}[\mathbb{C}_0 | f(t)] \\ &= \min_{\pi_t} \mathrm{E}\left[\beta^N \mathbb{C}_N + \sum_{s=t}^{N-1} \beta^s C(f(s), c(s), s) \middle| f(t) \right] + \sum_{s=0}^{t-1} \beta^s C(f(s), c(s), s) \end{aligned}$$

(since we take expectation given information up to time t and minimise on π_t)

$$\begin{aligned} &= \beta^t \min_{\pi_t} \mathrm{E}[\mathbb{C}_t | f(t)] + \sum_{s=0}^{t-1} \beta^s C(f(s), c(s), s) \\ &= \beta^t J(f(t), t) + \sum_{s=0}^{t-1} \beta^s C(f(s), c(s), s). \end{aligned} \tag{3.176}$$

The right hand side of the Bellman equation (3.174) may also be simplified:

$$\min_{c(t),y(t)} \mathrm{E}[G(W_{t+1})|W_t]$$

$$= \min_{c(t),y(t)} \mathrm{E}\left[J(f(t+1),t+1) + \sum_{s=0}^{t} \beta^s C(f(s),c(s),s) \middle| f(t)\right]$$

(given (3.176))

$$= \min_{c(t),y(t)} \left\{\mathrm{E}[\beta^{t+1} J(f(t+1),t+1)|f(t)] + \beta^t C(f(t),c(t),t)\right\} + \sum_{s=0}^{t-1} \beta^s C(f(s),c(s),s). \tag{3.177}$$

Since we take expectation given information up to time t and minimise on $c(t)$, $y(t)$, the sum on the right hand side of equation (3.177) lies outside the expectation and minimum operators.

The Bellman optimality equation may now be simplified. Replacing the left and right hand sides of equation (3.174) by equations (3.176) and (3.177) respectively and dividing by β^t yields

$$\begin{aligned} J(f(t),t) &= \min_{c(t),y(t)} \left\{C(f(t),c(t),t) + \beta\mathrm{E}[J(f(t+1),t+1)|f(t)]\right\} \\ &= \min_{c(t),y(t)} \left\{\theta_1(f(t) - FT_t)^2 + \theta_2(c(t) - CT_t)^2 + \beta\mathrm{E}[J(f(t+1),t+1)|f(t)]\right\} \end{aligned} \tag{3.178}$$

As for the closing condition, note from equation (3.176) that

$$G(W_N) = \beta^N J(f(N),N) + \sum_{s=0}^{N-1} \beta^s C(f(s),c(s),s), \tag{3.179}$$

while

$$\begin{aligned} \mathrm{E}[\mathbb{C}_0|W_N] &= \mathrm{E}\left[\beta^N \mathbb{C}_N + \sum_{s=0}^{N-1} \beta^s C(f(s),c(s),s) \middle| f(N)\right] \\ &= \beta^N \mathbb{C}_N + \sum_{s=0}^{N-1} \beta^s C(f(s),c(s),s). \end{aligned} \tag{3.180}$$

The boundary condition on the recurrence relationship (3.174) is therefore

$$J(f(N),N) = \mathbb{C}_N = \theta_0(f(N) - FT_N)^2. \tag{3.181}$$

The Bellman dynamic programming principle states that the minimising values of $c(t)$ and $y(t)$ in equation (3.178) are the optimal contribution and asset allocation controls.

3.6.3 Optimal Contribution and Asset Allocation Strategies

It may be shown by mathematical induction that the solution to the Bellman equation (3.178) is

$$J(f(t),t) = P_t f(t)^2 - 2Q_t f(t) + R_t, \tag{3.182}$$

where

$$P_t = \theta_1 + \theta_2 \beta \sigma^2 (1+r)^2 \tilde{P}_{t+1} P_{t+1}, \tag{3.183}$$
$$\tilde{P}_{t+1} = [\theta_2(\alpha^2+\sigma^2) + \beta\sigma^2(1+r)^2 P_{t+1}]^{-1}, \tag{3.184}$$

with boundary condition

$$P_N = \theta_0, \tag{3.185}$$

and

$$Q_t = \theta_1 FT_t + \theta_2 \beta \sigma^2 (1+r) \tilde{P}_{t+1}[Q_{t+1} - P_{t+1}(1+r)(CT_t - B)] \tag{3.186}$$

with boundary condition

$$Q_N = \theta_0 FT_N. \tag{3.187}$$

Define

$$\Theta_t = \theta_2(\alpha^2+\sigma^2)\tilde{P}_t. \tag{3.188}$$

The optimal contribution is

$$c^*(t) = \Theta_{t+1} CT_t + (1-\Theta_{t+1})[B - f(t) + Q_{t+1} P_{t+1}^{-1}(1+r)^{-1}] \tag{3.189}$$

and the optimal asset allocation decision is

$$y^*(t) = \frac{\alpha \Theta_{t+1}(1+r)[Q_{t+1} - P_{t+1}(1+r)(f(t)+CT_t-B)]}{(\alpha^2+\sigma^2)[(1-\Theta_{t+1})Q_{t+1} + \Theta_{t+1}P_{t+1}(1+r)(f(t)+CT_t-B)]}. \tag{3.190}$$

These results are proven in Appendix C (§C.1).

Some Properties of the Optimal Controls. First note that since $P_N = \theta_0 > 0$, $P_t > 0$, $\tilde{P}_t > 0$, $0 < \Theta_t < 1$ for $t \in [1, N-1]$ from the Riccati difference equation for P_t (3.183) (see Appendix C, §C.1). It is reasonable to assume that the fund and contribution targets in any year are such that $FT_t > 0$ and $CT_t < B$, as otherwise there is no sense to funding in advance for pension benefits. Then, $Q_t > 0$ for $t \in [1, N]$ from equation (3.186).

The optimal contribution is linear in $f(t)$. From equation (3.189), $c^*(t)$ may be written as $c_0(t) - (1-\Theta_{t+1})f(t)$, where $1 - \Theta_{t+1} > 0$. The optimal contribution is therefore in the same form as a 'proportional controller' (Loades, 1998) and is very similar to the contribution calculated when surpluses or deficits are spread over a moving term [Sung (1997:218), Cairns (1997)].

It is also clear from equation (3.190) that $\partial y^*(t)/\partial f(t)$ is directly proportional to

$$-(1+r)^2(\alpha^2+\sigma^2)\alpha\Theta_{t+1}P_{t+1}Q_{t+1} \quad < \quad 0. \tag{3.191}$$

The proportion invested in the risky asset therefore decreases as the fund level increases. This is a very significant result and is considered further in §3.6.4.

Asymptotic properties. Optimal asset allocation and contribution controls in an infinite horizon, with stationary parameters FT and CT and no terminal cost may also be derived. Since the cost in equation (3.168) is non-negative and $\beta > 0$, it is possible to show that the value function in equation (3.175) is monotonic increasing and that the limit as $N \to \infty$ of the dynamic programming equation (3.178) exists (Bertsekas & Shreve, 1978). As $t \to -\infty$, P_t and Q_t converge to stable values under certain conditions and stationary optimal contribution and asset allocation may be obtained. These are the discrete-time counterparts of the results of Boulier *et al.* (1995) and Cairns (1997). For long control horizons N, the stationary controls may provide reasonable approximations to equations (3.189) and (3.190).

Risk-less Investment. Consider the optimal asset allocation decision and contribution control in year $(N-1,\, N)$. These are easily found by replacing P_N and Q_N from boundary conditions (3.185) and (3.187) respectively into equations (3.189) and (3.190). Also,

$$\Theta_N = \theta_2(\alpha^2+\sigma^2)/[\theta_2(\alpha^2+\sigma^2)+\theta_0\beta\sigma^2(1+r)^2]. \tag{3.192}$$

The optimal contribution payment is

$$c^*(N-1) = \Theta_N CT_{N-1} + (1-\Theta_N)c_f(N-1), \tag{3.193}$$

whereas the optimal proportion of the fund to invest in the risky asset is

$$\begin{aligned} y^*(N-1) &= \frac{\Theta_N\alpha(1+r)[FT_N-(1+r)(f(N-1)+CT_{N-1}-B)]}{(\alpha^2+\sigma^2)[(1-\Theta_N)FT_N+\Theta_N(1+r)(f(N-1)+CT_{N-1}-B)]} \\ &= \frac{\Theta_N\alpha(1+r)^2[c_f(N-1)-CT_{N-1}]}{(\alpha^2+\sigma^2)[(1-\Theta_N)FT_N+\Theta_N(1+r)(f(N-1)+CT_{N-1}-B)]}, \end{aligned} \tag{3.194}$$

where

$$c_f(N-1) = FT_N(1+r)^{-1} + B - f(N-1). \tag{3.195}$$

$c_f(N-1)$ is the 'risk-less contribution' that is required at $N-1$ if the fund is entirely invested in the risk-less asset in year $(N-1,\, N)$ and if the fund level target of FT_N is to be met at the end of the year. $c^*(N-1)$ is a weighted average of the risk-less contribution and the desired or target contribution (equation (3.193)).

If the target contribution at the start of year $(N-1, N)$ is equal to the risk-less contribution, then it is optimal to pay the target risk-less contribution $c^*(N-1) = CT_{N-1} = c_f(N-1)$ (from equation (3.193)) and to invest the fund fully in the risk-less asset, $y^*(N-1) = 0$ (from equation (3.194)). Furthermore, the optimal, target, risk-less contribution will be zero if

$$f(N-1) \;=\; FT_N(1+r)^{-1} + B. \tag{3.196}$$

If, in addition, the fund size at the start of the year is already at the desired level ($f(N-1) = FT_N$), then $f(N-1) = B(1+r)/r = f_m$. f_m is the "maximum necessary wealth" [Boulier *et al.* (1995, 1996)] or the "self-financing risk-free steady state" (Cairns, 1997) or the actuarial liability (standard fund) under the so-called total or complete funding method (Trowbridge, 1952): the risk-free interest payments on the fund are enough to meet all benefit outgo without contributions being required.

Special Cases. Suppose now that there is no security objective and it is required only that contributions remain stable and close to the contribution target (an impractical objective in the long term), i.e. $\theta_0 = \theta_1 = 0$. Then $\tilde{P}_t = 1/[\theta_2(\alpha^2 + \sigma^2)]$ and $\Theta_t = 1$. Since $P_N = 0$ and $Q_N = 0$, then $P_t = Q_t = 0$ $t \in [1, N]$ from recurrence relationships (3.183) and (3.186) respectively. Therefore, $c^*(t) = CT_t$ in equation (3.189) while $y^*(t)$ in equation (3.190) is not defined. In other words, this (trivial) objective is attained by merely paying the target contributions CT_t. A more realistic objective may be to minimise the deviation of the fund level from its target (e.g. the unfunded liability) at the end of the control period, i.e. $\theta_1 = 0$, $\theta_0 \neq 0$.

Suppose now that there is no stability objective and it is required only that deviations of the fund level from target should be minimised within period N, i.e. $\theta_2 = 0$. Then $P_t = \theta_1$ and $Q_t = \theta_1 FT_t$ for $t \in [1, N-1]$ from the recurrence relationships (3.183) and (3.186) respectively, and $P_N = \theta_0$ and $Q_N = \theta_0 FT_N$ from boundary conditions (3.185) and (3.187) respectively. $\Theta_t = 0$ for $t \in [1, N]$. From equations (3.190) and (3.189) respectively, it is clear that $y^*(t) = 0$ and that

$$c^*(t) = B - f(t) + FT_{t+1}/(1+r), \tag{3.197}$$

which we may define as the 'risk-less contribution' at time t, $c_f(t)$. It is optimal to invest the fund fully in the risk-less asset for maximum security and pay $c_f(t)$. At the beginning of year $(t, t+1)$ the fund, of amount $f(t)$, earns a risk-free rate of return r and

$$f(t+1) = (1+r)[f(t) + c_f(t) - B] = FT_{t+1}, \tag{3.198}$$

i.e. the fund meets its target with certainty and $f(t) = FT_t$ for $t \in [1, N]$.

3.6.4 Controlling Risk in Pension Funding

Objectives and Risks of Pension Funding. Various objectives of pension funding were considered in §2.2.2. Two important management or ongoing objectives are to secure promised retirement benefits by pre-funding them and to stabilise contributions so that they are predictable and the plan sponsor may plan in advance for them. These funding objectives are related to the very motivation for advance funding of pensions. Some stakeholders in the pension plan have particular interests in these objectives. The plan members (represented by the pension scheme trustees) will be concerned with the security of their benefits, whereas the plan sponsor is more concerned with the stability of future sponsor contributions. There is often a trade-off between these objectives. It is important for all stakeholders that some balance be achieved and actuarial input serves precisely this purpose.

The risk of not meeting the stability and security objectives must be addressed by the actuary. Pension funding may therefore be construed as a risky activity, where the uncertain outcome for the plan membership is that insufficient assets are accumulated to meet liabilities; and the uncertain outcome for the plan sponsor is unpredictable, variable contributions. The absolute deviation of fund and contribution levels from planned targets is therefore a monetary 'loss' or 'penalty' or 'cost' incurred by the various stakeholders. Note that large surpluses are as undesirable (and possibly costly) outcomes as are excessive deficits (§2.2.3). Actuarial control aims at limiting the risks of *fund inadequacy* and *contribution instability.* (These risks loosely incorporate the risk of insolvency and the contribution rate risk as discussed by Haberman (1994b). Griffin (1966) also discusses concepts of adequacy in pension funding.)

The attitude or aversion to risk of the various stakeholders in the pension funding process may be formulated through their respective utility (or disutility) functions. The optimal contribution and asset allocation strategy may be defined as the one that maximises the utility (or minimises the disutility) over time for the different economic agents involved in pension funding.

Utility and Disutility. The concept of utility (or satisfaction) for an economic agent of a reward or gain is equivalent to the concept of disutility of an incurred loss or penalty. Let x represent 'reward' or 'gain'; y represent 'loss' or 'penalty'; $U(x)$ be the utility of a reward; and $L(y)$ be the disutility of a loss. Since it is easier to measure the loss or penalty in the pension funding system in terms of the deviations from planned targets, the concept of disutility rather than utility is more useful. In order to determine a suitable disutility function, we need to consider the sensitivity of each party to risk. A penalty or loss $y = |f(t) - FT_t|$ is incurred if the fund level deviates from the desired level. This relates to the risk of fund inadequacy. A penalty $y = |c(t) - CT_t|$ is incurred if the contribution level deviates from the desired level, and it relates to the contribution instability risk. Three desirable properties of disutility functions are considered hereunder.

Disutility increases for increasing loss or deviation.

$$L'(y) > 0. \tag{3.199}$$

Equivalently, utility decreases for decreasing gain or reward: $U'(x) > 0$. For example, plan members are less tolerant of large deficits than small deficits; the plan sponsor finds large surpluses less desirable than small surpluses; large supplementary contributions are less welcome than small supplementary contributions.

Risk aversion or intolerance. The marginal increase in disutility increases for increasing loss or deviation:

$$\mathrm{E}L(y) > L(\mathrm{E}y), \tag{3.200}$$

$$L''(y) > 0. \tag{3.201}$$

Equation (3.200) shows that greater disutility is derived by a risk-averse agent from undertaking a risky activity than a non-risky one. The disutility function is therefore convex. (Risk aversion in terms of utility means that the marginal increase in utility decreases for increasing gain or reward: $U''(x) < 0$ and the utility function of a risk-averse agent is concave.) An example of risk-aversion is that plan members will (typically) be more dissatisfied by a unit increase in the deficit when the deficit is large rather than small.

Risk aversion is non-decreasing with increasing loss or deviation. This may be represented in terms of Pratt's (1964) coefficient of risk-aversion (or risk-tolerance parameter) $A(y) = L''(y)/L'(y)$:

$$A'(y) \geq 0. \tag{3.202}$$

(Risk aversion is non-increasing with increasing gain or reward: $A'(x) \leq 0$, where $A(x) = -U''(x)/U'(x)$.) The coefficient of risk aversion is usually derived in terms of utilities and a justification for equation (3.202) in terms of disutilities is given in Appendix C (§C.2).

Quadratic Disutilities. The cost incurred in equation (3.168) is a quadratic function of deviations in the fund and contribution levels. The performance criterion in equation (3.171) is therefore based on a quadratic cost functional or quadratic disutility function (Bertsekas, 1976:18). For a quadratic disutility function, $L(y) = \lambda y^2$—where $\lambda > 0$ and the loss or penalty $y \geq 0$ could be the absolute deviation of either fund or contribution level from target—$L'(y) > 0$ and $L''(y) = 2\lambda > 0$, thereby satisfying the properties represented by equations (3.199) and (3.201) respectively. But $A(y) = 1/y$ and $A'(y) = -1/y^2 < 0$: the quadratic disutility function does not satisfy property (3.202). This argument is usually made in terms of quadratic *utility* functions which are therefore found to be imperfect although they are convenient mathematically (Pratt, 1964) and widely used.

Controlling Risks. Since the 'cost' of deviation in the pension funding process is quadratic in equation (3.168), optimisation over the cost functional or performance criterion (3.171) may therefore be interpreted as a minimisation of the discounted quadratic disutilities arising from deviations of the fund and contribution levels from targets, i.e. the risks of *fund inadequacy* and *contribution instability* are minimised. The plan members and sponsor are assumed to be averse to these risks, as $\theta_1 > 0$ and $\theta_2 > 0$ in equation (3.168). The dynamic optimal contribution and asset allocation strategies developed in equations (3.189) and (3.190) represent *minimum-risk* strategies. These strategies *hedge* against fund inadequacy and contribution instability.

It was observed in §3.6.3 that it is optimal to invest a smaller proportion of the pension fund in the risky asset as the fund level increases. This means that the higher the funding level, the more of the fund should be invested into the risk-less asset and, conversely, the more pension liabilities are underfunded, the riskier pension fund investment ought to be. This feature is also obtained by Boulier *et al.* (1995) and Cairns (1997).

Thus, an immature pension plan with a relatively young membership may choose to invest its fund more aggressively in the early years in order to defray a deficit in respect of past service liabilities. Sponsor contributions may also be increased. The balance between these two strategies depends on the relative values of θ_1 and θ_2 in equation (3.168), i.e. between the relative aversion to the risk of fund inadequacy and contribution instability. Conversely, if a plan is in surplus, it should 'lock' into these surpluses by investing them into less risky assets (Exley *et al.*, 1997) and sponsor contributions may also be reduced.

This result appears to contradict the current actuarial propensity to invest pension fund surpluses into risky assets (at least in the U.K.), as noted by Cairns (1997). It is nevertheless a sensible result in view of the fact that pension funds need to hedge their liabilities so as to minimise both the volatility of surpluses or deficits as well as of contributions. This concept is expressed variously in terms of immunization (Vanderhoof, 1984) and hedging (Exley *et al.*, 1997). Exley *et al.* (1997) and Bezooyen & Mehta (1998) also regard a hedging portfolio to be a 'minimum-risk' portfolio. Sometimes consideration of contribution variability is suppressed and only volatility of surpluses (in various forms) is minimised, as in Wise's (1984) discussion of asset-liability matching. Some of these concepts are examined qualitatively in Chapter 4.

Some Limitations. The model used above is a simplification of reality for a number of reasons:

1. *Inflation.* Uncertain salary inflation leads to uncertain growth in the pension liabilities. (Salary inflation was assumed deterministic and monetary quantities net of salary inflation were considered.) The statistical dependence between inflation and returns in the risky asset is therefore crucial. When benefits are not indexed to price inflation, inflation also triggers demands for benefit enhancement so that the assumption of defined

benefits in the defined-benefit pension plan model does not hold exactly.

2. *Utility function.* The quadratic (dis)utility function is simplistic because it does not admit solvency and full funding (maximum surplus) constraints. A pension fund with large deficits may be technically insolvent and may be prohibited from investing more aggressively; larger contributions may be the only acceptable course of action to restore financial balance. Discontinuous cost functions in the context of pension funding are explored numerically by Boulier *et al.* (1996). Note that the fact that the quadratic utility function is two-sided is not a drawback, but a more realistic utility function would be asymmetric: a deficit may be less tolerable than a surplus of equal magnitude.

3. *Short selling.* The proportion of assets held was not constrained to be non-negative, which allows borrowing. This may conflict with long-term liquidity objectives. Such a constraint also proscribes closed-form solutions. Cairns (1997) discusses numerical solutions with such a constraint.

4. *Markovian dynamics.* It has been assumed that the risky asset returns followed a random walk, which is one consequence of the Efficient Market Hypothesis. It is likely that statistical dependence of returns over time will occur, especially when inflation is taken into account. This severely restricts the application of optimal controls.

5. *Multiple risky assets.* It is not difficult to extend the above approach to multiple assets with contemporaneously correlated returns. Cairns (1997) assumes two risky assets, for example.

3.7 Summary

This section summarises some of the major points made in this chapter. The actuarial methods used to fund retirement benefits on a systematic basis are investigated. The concept of a "pension funding method" or "actuarial cost method" is described in §3.1 and the principal methods are illustrated using a simple model. The characteristics of these methods in terms of how they liquidate initial unfunded liabilities and whether they operate on an 'individual' or 'aggregate' basis are highlighted.

The bulk of the chapter is concerned with "supplementary funding methods" which are used to fund actuarial deviations when experience differs from actuarial valuation assumptions. One such method (§3.2) spreads surpluses and deficits forward over a rolling term (say m_s). This method is examined in detail by Dufresne (1986, 1988) and is implicit in the *Aggregate* pension funding method. Dufresne (1986, 1988) examines the first and second moments of the pension fund and contribution levels when rates of return are random. There is a maximum allowable spreading period beyond which the funding process becomes unstable. It is also shown by Dufresne (1986, 1988) that an efficient or 'optimal' range of spread periods

[1, m_s^*] exists, the tradeoff between security and stability (as measured by the variances of the fund and contribution levels in the limit respectively) being broken outside this range.

The pension funding process and the existence of an efficient range of periods over which to spread surpluses and deficits is investigated when various 'realistic' features are incorporated into the model. Haberman (1992a) and Zimbidis & Haberman (1993) show that delays in the pension funding system increase the variance of both contribution and fund levels and should therefore be avoided. A similar effect occurs when the frequency of valuations is reduced as shown by Haberman (1993). These results are not surprising in that delayed information and less frequent feedback control are known to affect adversely the control performance of various systems.

A modified supplementary funding method that allows for the explicit, separate amortization of initial unfunded liabilities (as prescribed under certain funding regulations) over a fixed term n is also described in §3.2.4. This method is found to be implicit in the *Frozen Initial Liability* pension funding methods. The first and second moments of the fund and contribution levels over time are obtained. Full funding is achieved within n years on average, whereas when initial unfunded liabilities are not dealt with separately, liabilities are funded asymptotically. In the limit, the variance of the funding process (and therefore security and stability) is the same whether initial unfunded liabilities are removed explicitly or not.

The assumption of independent and identically distributed random rates of return in the pension fund is replaced by a more realistic assumption in §3.3: the logarithmic rate of return is projected to follow a general stationary Gaussian autoregressive process, AR(p). The results of Haberman (1994a) regarding the moments of the pension fund and contribution levels when $p = 1$, 2 are generalised. These generalised results also suggest that when the mean and/or autocovariance of the rate of return decreases, the maximum allowable spreading period for stability increases, mirroring the numerical results of Haberman (1992a). It is also possible to surmise that his numerical results concerning the evidence of an efficient spreading period range may hold approximately for a general autoregressive process, i.e. that the efficient spread period range reduces as the autocovariance in the rate of return process increases. These results are also obtained numerically by Haberman & Wong (1997) for moving average rates of investment return of order 1 and 2.

Various additive perturbations may be included in the pension funding system. One such perturbation is discretionary extra contributions from the plan sponsor (§3.4.2). These arise from the sponsor's requirement for contribution flexibility. The moments of the pension fund and contribution levels in the limit when stochastic AR(p) additive perturbations are included are obtained in §3.4.3. Numerical results indicate that an efficient range of spreading periods remains and that it does not change much from the 'optimal' range of Dufresne (1986, 1988) except when perturbations are very strongly autocorrelated. Another perturbation in the pension funding system that is considered in §3.4.4 arises from a variable pension plan population. An important and well-known result is shown to hold in a simple, deterministic context: advance funding is only justified when the rate of investment return on assets exceeds the rate of growth of liabilities through inflation and through growth in the pension plan membership. The variance of the pension fund level is shown to increase when random numbers of new entrants are admitted into the plan. Some numerical results based on both the *Unit Credit* and *Entry Age* methods indicate that the efficient spread period range as obtained by Dufresne (1986, 1988) remains and does not change much. These results indicate that Dufresne's (1986, 1988) conclusion regarding efficient spreading periods is robust and holds in different practical circumstances.

A different supplementary funding method is considered in §3.5: actuarial gains and losses are directly amortized over a finite term m_a, with initial unfunded liabilities being amortized separately. Under the assumption of random rates of return and a stationary (deterministic) pension plan membership, it is proven that there exists a range of amortization periods, $[1, m_a^*]$, that is 'optimal' in the sense defined by Dufresne (1986). The tradeoff between the ultimate variance of the fund and contribution levels (representing security and stability in the pension fund respectively) is maintained when $m_a \in [1, m_a^*]$. This range is shown to be wider than the corresponding spread period range $[1, m_s^*]$. It is shown that the amortization of gains and losses over a fixed term yields greater fund security than when surpluses and deficits are spread over a moving term of equal length. However, it is also proven that spreading surpluses and deficits may be regarded as *more* efficient than amortizing gains and losses, based on the criterion of minimising the variances of both fund and contribution levels in the limit. This is because amortization involves the feedback of delayed information into the pension funding process.

Finally, an optimal pension funding policy through both contribution and asset allocation is considered in §3.6. This requires the strict assertion that rates of return are independent from year to year so that the dynamic programming principle is applicable. A simple two-asset model is postulated—one risk-less and one risky asset. Optimization is based on an expected mean square performance criterion measuring the deviation of sponsor's contribution and pension fund level from target values. This is interpreted as a measure of the quadratic disutility derived by plan members and plan sponsor arising from experience deviations and thus from variable unfunded liabilities and supplementary contributions. The optimal contribution control $c^*(t)$ over a finite horizon resembles the proportional spreading of surpluses and deficits over a moving term. The optimal asset allocation $y^*(t)$ is found to be a portfolio that dynamically hedges against the risk of not meeting security and stability funding objectives: it locks into surpluses by requiring less risky investment when surpluses emerge and conversely requires more risky investment when deficits emerge. The optimal funding policy is thus said to hedge against the risks of fund inadequacy and contribution instability and to be a 'minimum-risk' strategy. Some limitations of this analysis are also discussed.

Chapter 4

Asset and Liability Valuation Methods

4.1 Valuation Methodology

Various methods are used by actuaries to value the assets and liabilities of a pension plan. The choice of method should be consistent with the aim of the valuation. The discussion in this chapter pertains only to the ongoing or management valuation. The aim of the ongoing valuation is to compare assets and liabilities and to recommend contributions on a going-concern basis. Two competing methods are in use: the Discounted Cash Flow (DCF) method and the market method.

4.1.1 The Discounted Cash Flow Method

Cash Flows in a Pension Fund. At the core of this method is the view that a pension fund valuation is a comparison of different cash flows: there is income from contributions and investments, and outgo as benefits are paid. There may then be an overall excess of cash at the end of each year, as incoming cash flows exceed benefit outgo, and this excess of cash will be reinvested. Conversely, there may be an overall shortfall, as outgoing cash flows exceed regular contributions and investment proceeds, and disinvestment from some assets is then required to meet the outgo. Gilley & Funnell (1958) and Daykin [Boden & Kingston (1979): discussion] offer very clear descriptions of the valuation process in these terms.

Projecting Cash Flows. If we could project cash flows over time, we could then obtain the excess or deficiency of cash for each year in the future. These could then be consolidated at the rate of reinvestment return on new money to give us a present value of the surplus (or deficit) in the pension fund.

Discounting Cash Flows. In practice, it is computationally difficult to project all cash flows. It is easier (Gilley & Funnell, 1958) to *discount* asset and liability cash flows separately at the rate of reinvestment return. We then obtain discounted asset and liability 'values'. The difference between them is the present value of the surplus of asset over liability cash flows. We may then determine additional contributions which when reinvested at the assumed rate of reinvestment return will amortize the surplus over a given period.

The Breadth of Pension Liabilities. The assets belonging to a pension plan are usually clearly defined. It is more difficult to establish what we mean by the pension liabilities in the context of a valuation. We could for instance consider benefits only when they become payable, e.g. when a member retires. Or we could evaluate only benefits that would have to be paid if the pension plan were to wind up. Or we could consider the expected benefit that would be paid to all members of the plan as soon as they enter the plan. In other words, we need to consider the breadth of the liabilities that must be pre-funded. This determines the degree of advance funding for liabilities, i.e. the amount of assets that the fund should aim to hold to meet its liabilities. The broader the liabilities that are evaluated, the more assets the pension plan will aim to hold, and the more advanced will be the degree of funding.

Pension Funding Methods. Pension funding is about exchanging investment income for contribution income to meet the benefit outgo. Suppose we can project the benefit outgo in the future; suppose also that we have determined the degree of advance funding required, i.e. the breadth of the liabilities, and so the amount of assets that we aim to hold to meet these liabilities; suppose further that the fund actually holds these assets and that we can project investment income from them; then the level of regular contributions necessary to ensure that the liabilities are funded follows. This is because cash flows must eventually balance: the present value of benefit outgo less investment income should equal the present value of contributions. Various pension funding methods or actuarial cost methods (see §3.1) have been devised precisely to evaluate liabilities (termed 'actuarial liability' or 'standard fund') of varying breadth and to determine suitable regular contributions (termed 'normal cost' or 'standard contribution') accordingly.

Contribution Adjustment. The fund will not actually always hold enough assets to meet the liabilities. The Discounted Cash Flow valuation therefore determines any deficit or surplus that may exist, assuming that the 'standard contribution' is paid. Given the surplus or deficit, we may determine the additional contributions that will be required to defray over a given term any existing surplus or deficit: the present value of these future additional contributions should equal the present value of future benefit outgo less investment income and standard contributions in each future year. The adjustment to contribution that we have described in Chapter 3 as spreading surpluses or deficits over a rolling term thus fits in naturally with the

Discounted Cash Flow valuation method.

The 'Long-term' Portfolio. It was implicitly assumed, in the above, that the portfolio of assets held by the pension fund will remain unchanged from what it is at the point of valuation. Clearly this may not be so: tactical positions may have been taken that will be discontinued; new investment opportunities may arise; economic circumstances may change; pension obligations may change. The value that is placed on the surplus as an outcome of the Discounted Cash Flow valuation depends on the timing and amounts of the cash flows, and hence on the portfolio. It is important, therefore, that the portfolio of assets that is assumed for valuation purposes be the one that will be held on a 'long-term' strategic basis.

The 'Matching' Portfolio. Another implicit assumption that was made is that the single 'surplus' value obtained from the discounted value calculation provides an accurate comparison of the future cash flows. This will be so if the assumption as to the reinvestment return on new money in all future periods turns out to be exactly right. But the reinvestment return is an uncertain time-variant economic quantity, although we may assume it to be constant. The surplus value resulting from the Discounted Cash Flow calculation provides a meaningful and acceptable comparison of cash flows only if the discounted values are insensitive to small changes in the assumed reinvestment return.

A strong form of this argument can be expressed in terms of Redington's (1952) theory of 'immunization'. Springbett [Day & McKelvey (1964): discussion] idealises pension fund cash flow valuation in the light of well-known immunization results. He argues that a cash flow stream must be specified by *both* its discounted amount *and* its mean term, and that it is "only appropriate to add or subtract items which [have] identical mean terms." An even stronger version of this is encapsulated in Haynes' & Kirton's (1952) concept of 'absolute matching'. This is the theoretical notion that if asset and liability cash flows are exactly identical in amount and timing, then the valuation discount rate is irrelevant (the same valuation result is obtained irrespective of the valuation discount rate).

A weaker form of this argument can be phrased in terms of 'matching by type': relative changes in the discounted value of assets and liabilities should be roughly comparable for any given economic variation in investment return, inflation, real interest rates, currency exchange rates etc. Assets and liabilities should therefore have similar characteristics, for example in their link with price or salary inflation.

Hypothetical Switch into a Notional Asset Portfolio. It appears therefore that we can only judge the result of the discounted cash flow calculation as a reasonable comparative measure of asset and liability cash flows if two conditions hold:

1. the asset cash flows that we discount must be those arising from the long-term strategic portfolio;

2. the asset and liability cash flows must be at least 'loosely' matched.

Consequently, if the portfolio of assets held by the fund at the point of valuation coincides with the long-term portfolio that will be held by the fund, and this in turn coincides with the portfolio that somehow matches liabilities, the result of the DCF evaluation will be (broadly) reasonable. Such coincidence is unlikely in general. The 'pragmatic' technique employed to resolve this problem involves two adjustments:

1. assets are valued after assuming a hypothetical switch into a 'notional asset portfolio', and

2. a suitably adjusted discount rate is used.

The concepts of matching and hedging with respect to valuations are discussed in §4.2. The role of matching/hedging in portfolio selection is considered in §4.3.3. The long- term portfolio may be the same as the matching portfolio if a minimum risk strategic investment policy is pursued. But this will generally not be so, not least because it is not always clear what a matching portfolio is, but also because it may be a legitimate investment policy of the pension fund to take a 'mismatch' risk in the hope of higher returns. It is well known that optimal long-term strategic portfolio is usually different from a matching or hedge portfolio. Wise (1987) exhibits this difference mathematically (§4.2).

Hypothetical Switch into a Matching Portfolio. The concept of the notional switch into a *matching* portfolio is described by Springbett (1964; Day & McKelvey, 1964: discussion) as a "mismatching adjustment" through a hypothetical switch of assets. Arthur & Randall (1990:¶3.7) also specify that a matching portfolio must be used. If assets are hypothetically switched into a *matching* portfolio, future asset cash flows are 'adjusted', and if the matching portfolio were a portfolio 'hedging' the liabilities, this would be equivalent to a 'risk-neutral' valuation (Exley *et al.*, 1997). Since the actual asset portfolio does not in general perfectly match or hedge the liabilities, surpluses and deficits will subsequently emerge and will be amortized. The valuation discount rate used is therefore the estimated return on new money being reinvested in the *matching* portfolio, rather than in the actual portfolio. The discount rate will also comprise a prudent margin or risk adjustment to allow for the greater downside risk posed by returns that are less than anticipated.

Hypothetical Switch into a Long-Term Portfolio. The fact that the notional portfolio is not the correct long-term one does not directly affect the *comparability* of asset and liability cash flows. Provided that they are *consistently* measured (requiring at least a 'loose' match) any difference will unwind as gains and losses to be amortized. It is also difficult to estimate the returns from a strategic investment portfolio, especially compared with a matching portfolio which is by definition more stable relative to the liability cash flows. It is also doubtful that there exists a single 'long-term portfolio', given that in practice dynamic decisions are made.

A number of actuaries in the U.K. (where the Discounted Cash Flow method is principally used) take the view that the notional portfolio should be a suitable *long-term* strategic investment portfolio, and not exactly a matching portfolio. Kemp (1996:¶8.16) refers to "two main schools of thought" regarding the notional portfolio. Lee (1986:§24.25) and Thornton & Wilson (1992a:¶9.7) consider the notional portfolio to be a suitable long-term strategic portfolio. In such cases, a 'best estimate' of the future asset cash flows is being made, but there is a need to guard against the risk introduced by the mismatch between asset and liability cash flows. Wise [Colbran (1982): discussion] and McKelvey [Exley *et al.* (1997): discussion] state that this may be done through a mismatch risk adjustment in the valuation discount rate. The valuation discount rate used is therefore the estimated return on new money being reinvested in the *long-term* portfolio, rather than the actual portfolio, along with a prudent margin to adjust both for the greater downside risk and for the effect of a projected mismatch between asset and liability cash flows. The values of assets and liabilities will not be the same if the hypothetical switch is made into different portfolios and if the valuation discount rates used are different. It is not clear whether the calculated surplus (the difference between assets and liabilities) is different if assets are switched into a long-term strategic portfolio rather than a matching portfolio. This would presumably depend on the degree of mismatch risk involved and on the exact amount of risk adjustment in the discount rate. Wise [Exley *et al.* (1997): discussion] considers the practice of switching into a portfolio suitable for long-term investment to be a 'pragmatic' one.

Other Issues. The choice of valuation discount rate is discussed further in Chapter 5. Asset valuation in the DCF method is discussed further in §4.4.1. Because of the emphasis on ensuring that asset cash flows meet liability cash flows, Ezra (1988), Exley *et al.* (1997) and others refer to the DCF valuation method as a 'funding valuation' method.

4.1.2 The Market Method

An Economic Value of Cash Flows. The Discounted Cash Flow method seeks to compare asset and liability cash flows by considering them together in the context of funding. The market method also attempts to compare asset and liability cash flows but does so by determining an economic value for assets and liabilities separately, but consistently. It relies on the market price of a set of cash flows as being the measure of economic value.

The No-Arbitrage Hypothesis. As described by Milgrom (1985), Tilley (1988), Sherris (1994) and Exley *et al.* (1997), consistency in the market method is predicated on the no-arbitrage hypothesis: two cash flows identical in amount and timing must be equally priced by the market at any given point in time, lest an arbitrage opportunity arises. This is sometimes described as the 'Law of One Price'. It follows that the *market* values of any two sets of cash flows (asset or liability) should be consistent.

Market Pricing of Asset Cash Flows. The market value of future asset cash flows into the pension fund is the market value of the assets. For reasons of long-term stability and security, pure market values are not always used, given their volatility. This is discussed further in §4.4.1.

Market Pricing of Liability Cash Flows. Since there is no market in typical pension liabilities, liability cash flows must be priced by comparison with similar asset cash flows. This argument has been diversely expressed in the pension funding and actuarial literature in terms of *immunization* [Milgrom (1985), Vanderhoof (1972)], *matching* [Wise (1984, 1987), Arthur & Randall (1990)] and *hedging* [Tilley (1988), Exley *et al.* (1997)]. In other words, the market value of a set of liabilities is the market value of a portfolio of assets that immunizes or matches or hedges the set of liability cash flows.

Pension Funding Method. The breadth of pension liabilities being valued may again be determined according to one of a variety of pension funding methods. Again, the liabilities do not refer to any accounting or legal liabilities, but relate to the degree of advance funding that one wishes to achieve. An actuarial liability is the value of a particular set of liability cash flows as set out by a pension funding method. If these liability cash flows can be priced by reference to traded assets then, in principle, an actuarial liability can be found using the market method. Likewise, a normal cost is the value of a set of cash flows (e.g. benefits accruing to members during the following year) that may also be priced, and adjustments to the contribution may be effected according to emerging gains/losses or surpluses/deficits.

Liability Valuation Discount Rate. A direct consequence, therefore, of the no-arbitrage hypothesis and of the 'Law of One Price' is that liabilities may be priced at the market value of the portfolio of assets that matches or immunizes or hedges these liabilities. Equivalently, liabilities may be discounted at market discount rates implied in the assets that immunize or match or hedge them.

Term-dependent Discount Rates. Since liability cash flows can be compared with income from suitable bonds, the market value of liabilities is often obtained by discounting cash flow streams at discount rates implied by bond prices, i.e. using the term structure of interest rates. Depending upon the nature of the liabilities, different bonds are used for comparison. If the liabilities depend upon price inflation, then index-linked gilts or real-return bonds are used, but if the liabilities are fixed nominally, conventional gilts or Treasury bonds are more appropriate. This concept can be traced back in the context of pension funding to Bagehot (1972) as well as Treynor (1977) who specifically mentions discounting liability cash flows using term-dependent risk-free discount rates with or without inflation expectations. Milgrom (1985) also considers the economic value of liability cash flows in terms of the yield curve, although he considers only nominally fixed liabilities. Ezra (1988; Arthur & Randall, 1990: discussion) sketches a similar method, with reference to the U.K. index-linked yield curve. Several authors refer to the use of gilt yields in the valuation of nominally fixed pensioner liabilities, but they do not articulate a consistent yield for other liabilities and do not refer to term-dependence: Boden & Kingston (1979) and Colbran (1982) suggest asset values at market and a valuation discount rate for nominal pensioner liabilities that is based on gilt yields; gilt yields are also favoured by Jones [Jones (1993): discussion]. Given that the yield curve is not flat, term-variant discount rates must therefore be used, strictly, although duration-weighted averages of gilt or bond yields may be a useful practical approximation.

Other Issues. The market valuation of assets and the choice of the discount rate for valuing liabilities are discussed further in §4.4.1 and Chapter 5 respectively.

4.1.3 Comparison of Market and Discounted Cash Flow Methods

Both market and Discounted Cash Flow (DCF) methods employ a notional portfolio that matches or hedges liabilities. In the market method, it is assumed that a hedge portfolio can be accurately determined; that the appropriate market discount rates can be applied; and that a correct economic value can be placed on liabilities by valuing them by reference to the notional hedge portfolio. In the DCF method, it is assumed that a suitable portfolio that matches the liabilities can be found; that an approximate discount rate in the form of the matching rate of return can be used; and that both assets and liabilities can be consistently valued by reference to the notional matching portfolio. In both cases, liabilities are valued by reference to the hedging or matching portfolio: liability cash flows are discounted at the minimum-risk (or ideally risk-free) rate implied in the hedging or matching portfolio. Ideally, equal liability values can be obtained. Furthermore, assets are also valued equally in both methods if the market price (which the market arrives at by discounting the risky cash flows from the actual asset portfolio at suitable rates incorporating risk premiums) is equal to the discounted income value of the (hypothetically) 'adjusted' asset cash flows from the notional portfolio using the minimum-risk (or ideally risk-free) matching discount rate. Similar valuation results can in principle be obtained from the two methods. Exley *et al.* (1997:§4.7) arrive at a similar conclusion, but do not explicitly argue that it is consistent within the Discounted Cash Flow method to use a matching portfolio.

The methods will generally yield different results in practice:

1. Various approximations are made in the choice of discount rates in both valuation

methods. The use of a computationally convenient term-independent discount rate, intrinsic in the Discounted Cash Flow method, may lead to considerable loss of accuracy [Cogan (Colbran, 1982: discussion), Exley *et al.* (1997:¶3.2.9)].

2. The choice of the notional portfolio in the DCF method is not always made with rigour. The notional portfolio in DCF valuations is often regarded as a long-term strategic portfolio rather than a matching or hedging portfolio, as indicated earlier. Exley *et al.* (1997:§4.7) suggest that the notional 'long-term' portfolio is likely to be riskier than a hedge portfolio and thus lead to higher liability values.

3. Even if a reasonable matching notional portfolio is selected in the DCF method, it is not clear that the valuation discount rate is suitably (if at all) adjusted from the best estimate investment return assumption on the actual portfolio to reflect the lower risk. An expected long-term return is often used, sometimes based on the past performance of the fund.

A major difference between the two methods lie in the relationship between liability valuation and asset investment. An important consequence of the market method of valuation is that the value placed on liabilities is *independent* of the investment or funding policy followed by the pension fund. Ezra (1988) sums this up as follows:

> "Regardless of the funding payments made and the assumptions on which they are based, the annual accrual of pension obligations must be measured in terms of economic value, explicitly using 'best estimate' assumptions that reflect market conditions."

Likewise, Exley *et al.* (1997:¶6.9.3) state that

> "the economic cost of a pension promise depends particularly on the promised cash flows, not on the pace at which they are funded, nor on the assets in which the fund invests."

This is in direct contradiction with the theory of the Discounted Cash Flow (DCF) method, where projected net cash flows being reinvested every year are discounted. This contradiction is relevant in practice when the notional portfolio used in the DCF method is the long-term strategic portfolio, rather than the matching or hedging portfolio.

4.2 Liability Valuation Methods

Actuarial Present Value. As described in §4.1, ongoing pension liabilities are not in any sense accounting or statutory solvency liabilities, but refer rather to a subset of future benefit outgo which is to be pre-funded. They are traditionally valued using the actuarial present value (APV). This is the discounted value of a cash flow stream (at the valuation discount rate), allowing for mortality and other contingencies. Turner (1984) defines it as "a summation over time of the product of the amount payable (allowing where appropriate for earnings increases), the probability of payment at that time, and a reduction factor for interest" all according to the valuation basis. This is usually calculated over the set of all current plan members (i.e. assuming a 'closed group'), although a projection for new entrants may be made. It is typical that *actuarially neutral* options are offered to plan members (e.g. commutation of benefit into a lump sum at retirement). The various options are designed such that their actuarial present values are equal.

Hedging. A statistical hedge for a set of liability cash flows may be found by regressing liability cash flows on income or dividend from various assets, or by rolling up the liability and regressing the growth of liability on the total returns of various assets. The regression coefficients are then taken as the weights of each asset in the hedging portfolio. In terms of their market values, the hedge portfolio and the set of liabilities are expected to respond to changes in market conditions in the same way. Exley *et al.* (1997:§5.3) state that a suitable hedge portfolio for pension liabilities will be such that the surplus (or deficit) of assets over liabilities (at market values) will be uncorrelated with contemporaneous price changes in particular assets, as well as past and future price changes. If risk-less arbitrage is disallowed, then since the liabilities and their hedge portfolio have identical price behaviour, the theoretical value of their underlying cash flows are equal, and so the market value of the liabilities is equal to the market price of the hedge portfolio.

Matching. The term 'matching' is used in various ways in the actuarial literature. Fujiki (1994) describes the various concepts that have been associated with matching. The quantitative theory provided by Wise (1984, 1987) illustrates its application to the valuation of pension liabilities. The theory focuses on the ultimate surplus that emerges as assets are *run off* against liabilities over a suitably long period. Wise (1984) uses moments of the surplus as measures of investment and mismatch risk ('ruin' probabilities are also mentioned by Wise (1989) and Fujiki (1994)). Wise (1984) also shows that the value of liabilities is the value of a matching asset portfolio. The degree of mismatch between assets and liabilities at any point in time is measured by the expected value of the square of the ultimate surplus that is residual when all liabilities (in respect of all plan members at that point in time) are extinguished. (A notional assumption of 'no new entrants' is therefore made, although this may be relaxed.) Stochastic models for interest, price inflation and dividend growth must be assumed. A 'positive matching' portfolio (negative asset holdings are proscribed) may be found that minimises the mean square surplus.

Wilkie (1985) suggests that the 'price' of assets is also a variable—since the liabilities are not traded and cannot be regarded as negative securities—and therefore considering mere proportions of assets is insufficient. Portfolios are therefore constructed in a so-called P-E-V space (price– expected ultimate surplus–variance of ultimate surplus).

For the investment manager, there is a budget constraint: P is fixed and is the current market value of the fund. Optimality is achieved in E-V space when a portfolio has maximum expected ultimate surplus with the minimum variance. This is a reformulation of MPT in terms of surplus, with fixed liabilities (see §4.3.2 below) and a single-period optimization carried out over a suitably defined longer term instead of just one accounting year. Sherris (1992) casts the theory in a more general portfolio selection framework, with the use of a utility function, and attempts an extension to multi-period optimization.

For the actuary, the expected surplus E should be close to zero, and portfolios that minimize both the price of assets and variance of the ultimate surplus are 'efficient portfolios' in P- V space. Specification of the willingness to trade off greater variation of surplus with a lower price for the portfolio (- $\partial V/\partial P$: a risk tolerance factor or 'degree of risk') yields a particular unique efficient portfolio, which we may call the 'optimal' portfolio. (Wise shows that this is not too different from his 'matching portfolio'.) This portfolio is significant because its price, based on the current market value of its constituent securities, represents the market value of the liabilities of the pension scheme.

By contrast with hedging, which applies market valuation principles to the stochastic valuation of liabilities, Wise's (1984, 1987) matching represents stochastic valuation from a *cash flow* perspective. Cash flows are not discounted but accumulated. Consistency is achieved by accumulating both assets and liabilities using a single set of (stochastic) assumptions. The deterministic economic valuation assumptions of the traditional Discounted Cash Flow method (§4.1.1) are replaced by stochastic assumptions, which may be considered to be more realistic. Conversely, the results of a matching valuation may be very dependent on the stochastic model assumed. Analytical results are only possible for a very simple stochastic 'valuation basis'. In practice, simulations will be necessary.

A deficiency of Wise's (1984, 1987) matching valuation is that it is not clear how one determines suitable contributions. This is after all the output of an ongoing long-term valuation. The optimised match of asset and liability cash flows derived by Wise (1984) either ignores contributions or assumes a prior pattern of contributions. It is not clear which of portfolio selection or valuation should come first, and this arises because of the circular argument concerning contributions. This can be clarified by noting that one can equally perform the matching valuation on any subset of the liability cash flows, i.e. obtain a portfolio of assets that matches any subset of the total liabilities and hence value this subset. (In this respect, note that Wise (1989) postulates a criterion of linear combination and scaling of cash flows.)

A given pension funding method defines the actuarial liability as the value of a particular subset of liability cash flows, which can in principle be matched and 'valued'. Likewise, a normal cost is the value of a set of cash flows (e.g. the portion of liabilities that will accrue in the following year) which can be similarly valued. Any deficit (excess of actuarial liability over fund) also represents a liability. To spread this over m years, one can determine the equivalent of the annuity factor in the DCF method: the matched value of a unit cash flow over m years. Hence, the past-service contribution or adjustment to the normal cost is equal to the deficit divided by the 'annuity factor'. (Note that the spread period m could take any positive value. m could conceivably be longer than the period during which all liabilities due to current members are extinguished, provided that the optimisation is carried out over this longer period and an assumption, possibly stochastic, is made about new entrants over this longer period.)

Once recommended contributions are established, an optimal portfolio, from the viewpoint of the investment manager can be found. Optimisation in E-V space, allowing for the recommended contribution cash flows (for the first m years) and the standard contribution (for the remaining years of the projection during which there are active members), and assuming an investment risk tolerance factor $\partial V/\partial E$, then yields an 'optimal' portfolio for investment.

4.3 Some Issues in Strategic Asset Allocation

This section briefly reviews some issues in the choice of assets for pension funding. This is an area where consensus has not emerged. Different views are held in different jurisdictions. These views cannot always be reconciled by appealing to differing pension liability profiles, or to the different structures of financial markets, or to differences in taxation and statute.

4.3.1 Asset-Liability Modelling

In some countries (notably in the U.K. and the Netherlands), actuaries have sought a role in the integrated management of assets and liabilities. Some are involved in Asset-Liability Modelling (ALM) and hence in strategic asset sector selection. Tactical asset allocation (market timing) and security selection are usually left to the fund manager. In North America, it seems that fund managers have more latitude, and have incorporated liabilities (usually accounting liabilities) into their work (see §4.3.2).

One of the problems of ALM is the choice of suitable objectives. Although they make reference to all long-term funding objectives (see §2.2), most authors concentrate on the choice of a portfolio that optimises aspects of security in the long-term. There are differences as to the choice of underlying stochastic asset model and the definition of 'risk'. Both Loades (1988) and Daykin *et al.* (1993) use versions of the Wilkie (1987) model. Daykin *et al.* (1993) investigate four static portfolios in the context of closed pension funds and measure risk in terms of the frequency with which funding levels are less than 1 after a given period of time. Loades (1988) considers the level of security offered by the strength (or conservatism) of the basis used to value liabilities and measures security in terms of the frequency distribution of funding levels. Kemp (1996) prefers a model that is predicated on market efficiency and measures the riskiness of a portfolio as the covariance between its returns and those of a portfolio that closely matches pension liabilities.

Some authors determine an optimal investment portfolio in terms of the contributions required from the sponsor. This approach is followed by McKenna (1982). Black (1995) favours minimising the present value of future contributions. MacBeth *et al.* (1994) describe projections that are aimed at finding the portfolio which stochastically dominates all others in terms of the frequency of contributions at various levels (a relatively short horizon is used). McKenna & Kim (1986) seek the portfolio which dominates in terms of mean and semi-variance of contributions over a term comparable to the duration of the liabilities, given the sponsor's level of risk aversion. Such modelling seems to assume that the actuary uses a given pension funding method—often the *Aggregate* method, which readily yields a contribution without the construction of an actuarial liability or normal cost.

4.3.2 Modern Portfolio Theory with Unmarketable Liabilities

The Modern Portfolio Theory (MPT) of Markowitz (1952) implicitly assumes that liabilities are tradeable securities and hence 'negative' assets. In the 1980's, U.S. fund managers developed models based on MPT and efficient risk/return frontiers as well as performance measurement and risk-adjusted returns to justify and set investment strategy over short-term horizons for pension funds. These models ignored pension liabilities. In the 1990's, the advent of Financial Accounting Standards No. 87 (FAS87) led fund managers to focus on optimising *surpluses* [Leibowitz (1986c), Sharpe *et al.* (1995:477)]. MPT was adapted to included fixed (i.e. unmarketable) liabilities such as those of a pension fund. The efficient set of portfolios is then based on the mean and variance of a 'return' on the pension fund surplus [Ezra (1991), Leibowitz *et al.* (1992)]. This 'return' is the difference between, on the one hand, the rate of return on assets held and, on the other, the rate of growth of liabilities multiplied by some risk tolerance factor and divided by a fixed asset/liability ratio. The fixed funding level is present precisely because liabilities are unmarketable, so that actual prices and liability values (rather than mere proportions) do matter in the portfolio, as pointed out by Wilkie (1985) (see §4.2).

FAS87 provides two measures of pension liabilities at market discount rates (the accumulated and projected benefit obligations: ABO and PBO). They are used by fund managers as proxies for pension liabilities. Sharpe & Tint (1990) note that these measures need to be accepted and used consistently from year to year. Efficient portfolios may be determined such that period-by-period surplus is optimised: its mean is maximised and its variance minimised. Ezra (1991) investigates efficient portfolios, using both the ABO and the PBO as measures of liability, under various conditions.

This approach is a useful extension of MPT in that it integrates assets with liabilities. In this sense, it satisfies the actuarial perspective of "investing to meet liabilities", as argued by Arthur & Randall (1990) *inter alia.* It seeks to maximise security but disregards the other long-term objectives of funding, in particular the sponsor's requirement for contribution stability and flexibility. FAS87 aims to provide an economic value of liabilities but the market discount rates used may be inconsistent with the term structure of discount rates implied in the assets that hedge the pension liabilities in the long term. This method may therefore suffer from too short-term a perspective.

4.3.3 Asset Allocation by Matching or Hedging Liabilities

Some authors suggest that a pension fund should invest its assets in order to match or hedge at least some of its liabilities. Such investment policies aim to minimise the risk of assets not meeting liabilities:

1. 'Dedicated' bond portfolios, which match current (nominal) pensioner liabilities (Leibowitz, 1986a, b), are used to minimise the risk of not meeting these liabilities. Dedicated or immunized portfolios in the context of 'surplus' optimisation (see §4.3.2 above) are considered to be analogous to a risk-free asset (Leibowitz & Henriksson, 1988).

2. Keintz & Stickney (1980) also consider the desirability of an immunized investment strategy for pension funds. Because of the long duration of pension liabilities, they conclude that a significant investment in equities will be necessary, although less is required as the plan membership matures and liability duration shortens.

3. 'Lifestyle switching' arguments such as Samuelson's (1989) 'age-phased' investment strategy—justifying "folk wisdom" concerning less risk being taken by individuals (i.e. less equity investment) as they grow older—are also used in deciding investment strategy. They are sometimes confirmed by Asset-Liability Modelling studies (§4.3.1) such as by Kingsland (1982). Blake (1996) argues that as a plan matures, it should invest less in equities and long bonds and be able to invest in a "liability immunising portfolio". When the plan is mature, any funds *in excess* of the liability immunising portfolio can be invested to achieve higher returns, given the plan sponsor's risk tolerance.

4. Vanderhoof (1984) also seeks a portfolio of assets that somehow immunizes the liabilities. This is meant to minimise investment gains/losses and thus stabilise contributions.

5. Exley *et al.* (1997) and Bezooyen & Mehta (1998) favour a 'minimum risk' portfolio that hedges liabilities. They also note that this should yield, in addition to greater security, more stability in contributions and in reported costs, thereby enabling the plan sponsor to budget for his contributions and raise capital more effectively. They also point out that the greater stability in the funding process as a result of a matching or hedging policy should make it easier for the plan to remain within minimum and maximum funding limits.

Other authors, such as Arthur & Randall (1990), argue that pension funds should identify a matched investment position but may then take a mismatch risk in order to derive higher returns. This is based on the argument that higher returns may then be shared by the sponsor and its shareholders (through lower contributions) and by plan participants (through eventual benefit improvement). In terms of ultimate run-off surplus, Wise (1987) shows that the set of efficient strategic investment portfolios (maximum mean surplus with minimum surplus variance and a given risk tolerance parameter) is generally different from the matching portfolio (see §4.2).

At one extreme, a policy of hedging or matching enhances benefit security for members, but may nullify one of the purposes of funding, i.e. exchanging more investment return for less contribution. At the other extreme, a mismatched investment policy that is successful will reduce the level of contributions required to finance promised pension benefits, but may also put them at risk. In general, there will be some asset-liability mismatch according to the balance struck between the various pension funding objectives (§2.2), which themselves depend on the plan's demographic maturity, its funding level, the sponsor's risk tolerance and financial situation, his requirements for contribution stability, flexibility etc. This may involve a conflict of interest between trustees, members and sponsor.

4.3.4 Equities v. Bonds

There are considerable differences of opinion over whether pension funds should invest mostly in equities or in fixed-income securities. Many of these differences originate from the lack of consensus on certain features of equity investment returns, as discussed in §2.6.3. The equities v. debt debate relates partly to the argument over whether equities provide a better hedge against salary inflation than index-linked bonds. Whether a matching investment policy is followed or not, supporters of equity investment point to

- higher real returns in the long term, giving rise to surpluses that can be shared between the plan sponsor in the form of reduced contributions and plan members by receiving improved benefits;
- returns that may be at least partially correlated with salary inflation in the long term, so that final-salary pension rights are more secure;
- putative mean reversion in long-term returns corresponding to possibly reduced risk that long-term pension obligations are not met.

Supporters of heavier bond investment, on the other hand, point to

- low default risk on government securities and high-grade corporate bonds signifying greater security of benefits for plan members;

- reduced volatility of bond returns and hence greater stability in contributions and pension expenses which benefits the sponsor;
- tax advantages to the plan sponsor and to its shareholders [Black (1980), Tepper (1981), Black & Dewhurst (1981)].

These differences again stem from the relative importance attached to the various objectives of funding.

4.3.5 Control through Asset Allocation

The three major areas of control of pension funding were noted in §2.3.2 to be benefit policy, funding policy and strategic investment policy. A suitable asset allocation strategy is required to balance the various objectives of funding for retirement benefits. The sheer size of pension funds and of the returns available on capital markets means that asset allocation must be regarded as a major controlling variable. The distinctions between special compliance and ongoing management objectives (see §2.2.2), and between solvency and ongoing actuarial liability, are important in this respect.

Liabilities are sometimes conceptually 'partitioned' (Mennis *et al.*, 1981) for investment purposes on the basis of nature of membership (i.e. active vested, retired etc.), although Wise & Annable (1990) among others consider such 'segmentation' not to yield optimal investment performance. The distinction between solvency and ongoing actuarial liability is primarily based on the breadth of liabilities (see §4.1), while solvency liability may also be defined by statute or regulation and will tend to be related to wind-up benefits as stated in the pension plan document. Assets need to be accumulated initially to meet the solvency liability (this is a 'compliance' objective: see §2.2.2) and should be arranged first to hedge solvency liabilities. *Additional* assets may be accumulated to meet the *additional* actuarial liability (i.e. actuarial liability less solvency liability, for example when salary projection is taken into consideration from a going-concern perspective). The exact amount of the actuarial liability will be rationalised through the pension funding method, and will depend on the balance sought between the various objectives of funding, i.e. the security that members should be offered by advance funding, contribution stability and flexibility required by the sponsor, optimal value through savings on tax and transaction costs etc. (see §2.2). This balance will dictate the degree to which the *additional* assets are needed to match the *additional* actuarial liability. In a sense, two notional sub-funds may be deemed to exist. The first sub-fund consists of the solvency liabilities backed by assets that hedge them. The second consists of the additional actuarial liability backed by 'additional' assets, with the extent of mismatch being related to the balance struck between the various funding objectives.

It is not clear how much influence actuaries have in setting investment strategy and this appears to differ among various jurisdictions. Actuaries are often professionally required to comment on any mismatch between the strategic investment policies and the liabilities of a defined benefit pension plan when they perform an ongoing funding valuation. See for example Professional Standard 400 of the Institute of Actuaries of Australia (1995:¶64) and Guidance Note 9 of the Institute and Faculty of Actuaries (1997a:¶3.5.3). Such comments appear to be reserved for more extreme cases of mismatch. Indeed, the investment policies of pension funds that are technically insolvent or are close to insolvency may be restricted. Actuaries do not continually determine investment strategies (unlike contribution rates) for pension funds. Neither trustees nor fund managers are bound by the actuary's comment on their investment strategy [Freethy (Arthur & Randall, 1990: discussion)]. Indeed, not all fund managers welcome actuarial interference in asset allocation, although some are positive about it [Blake (1997:16), Ezra (Arthur & Randall, 1990: discussion)]. As compared to the sole responsibility they have for setting contribution rates, actuaries do not exert as much effective control on the pension fund through asset allocation.

4.4 Asset Valuation Methods

Asset valuation methods have been described and discussed by several authors. Again, I will not describe the details of the various asset valuation methods here, but I will summarise a few important features.

4.4.1 Some Practical Methods and Issues

Discounted Income Value of Assets. The discounted income value follows naturally from the Discounted Cash Flow method of valuation (§4.1.1). This involves discounting income from assets and is also sometimes referred to as 'Present Value' method. This methodology has originally been advanced by Heywood & Lander (1961) and Day & McKelvey (1964), as an alternative to book or historic cost value. The method may be applied to value individual securities but is usually applied to broad asset classes instead. Its use in valuing fixed-income securities is very straightforward but various additional assumptions are required when valuing variable income assets such as equities and property. Equities are usually valued using the original Dividend Discount Model of Gordon (1962): a constant growth rate in dividends is most often used. Various modifications such as 'select and ultimate' assumptions as to dividend growth are also employed. The dividend growth rate assumption is therefore very important, in particular relative to the valuation discount rate. The Discounted Income Value of equities is criticised (Dyson & Exley, 1995) on the grounds that it is too sensitive to the difference between assumed dividend growth and discount rate (i.e. the dividend yield in Gordon's (1962) model), that the assumption of a fixed, term-independent dividend yield is not a suitable approximation and that it adversely affects the dynamics of pension funding and investment (see also §4.4.5).

Market-related Value of Assets. The pure market value (price) of assets is rarely used because of the volatility in the price of heavily traded securities such as equities and bonds. In such cases, market-related values are used, with the aim of moderating some of the large and impermanent fluctuations in prices. A great variety of methods appear to be used to achieve this aim. The list below is not exhaustive.

Averaging over Security Prices. Some average over time of the price of each security, especially equities, is calculated. A moving average of prices may be taken over a given period—typically 3 to 5 years [Jackson & Hamilton (1968), McGill *et al.* (1996:678), Winklevoss (1993:173)]. An alternative is to use some form of exponential smoothing by recognising only a proportion of the change in prices [McGill *et al.* (1996:678), Winklevoss (1993:173)]. Aitken (1994:289), for example, describes a smoothed actuarial value that is a weighted average of current market value (20%) and previous actuarial value (80%).

Averaging over Fund. A common method is to recognise only a fixed percentage of the capital appreciation in a year or else recognise only a fixed fraction of the return in excess of the assumed return on the valuation basis. Dyson & Exley (1995:§7) propose a related smoothing method, based on the funding level or asset/liability ratio rather than the asset value. These methods put only a certain weight on current market values: Jackson & Hamilton (1968) and Colbran (1982) seem to suggest 10%; Ferris & Welch (1996) report typical fractions of 25% or 30% being used in Australia; Winklevoss (1993:174) suggests that current market values are weighted by 25–33% in the U.S.; finally, Dyson & Exley (1995:¶7.5.3.10) conclude that, in the calculation of actuarial values in the U.K., an effective weight of about 20% is placed on current market value. Among other methods, "spreading capital gains arising in any year over a specified period, typically two to five years" is sometimes used according to Ferris & Welch (1996). For expensing purposes, the U.S. Financial Accounting Standards No. 87 (FAS87) also allows averaging that "recognizes changes in fair [i.e. market] value in a systematic and rational manner over not more than five years." Finally, note that a simple average over time of the total market value of the fund cannot be taken as cash flows must be allowed for [Anderson (1992:§5.2)].

Other Methods. Various 'pragmatic' methods appear to be employed to value assets. Sometimes, different methods are used to value different types of securities, e.g. prices of common stock or equities may be averaged but bonds are taken at discounted income ('amortized') value. Many methods that restrict only extreme fluctuations in market value have also been suggested. Jackson & Hamilton (1968) describe "adjustment accounts" in which all or part of the return in excess of a reasonable investment return in a year is held, while losses are immediately recognised but can be deducted from the adjustment account. Boden & Kingston (1979) appear to suggest a combination of market and Discounted Cash Flow methods: the more extreme fluctuations of market equity values are dampened by limiting the variation of equity dividend yields (on an index) and discounting income at these extremes. If y is the current dividend yield on a suitable equity index, $[y_0, y_1]$ is an 'acceptable' range for dividend yields, and M is market value of equities, then

$$\text{Assessed equity value} = \begin{cases} My/y_0 & \text{if } y < y_0, \\ M & \text{if } y_0 \leq y \leq y_1, \\ My/y_1 & \text{if } y > y_1. \end{cases} \tag{4.1}$$

It is not clear how consistency with the valuation of liabilities would be maintained if such an asset valuation method were used. Methods that only remove extreme fluctuations do not smooth market values continuously and this may make contribution rates more unpredictable.

Proximity to Market Values. There is also concern that actuarial values, while smoothing market values, should not stray too far from them. In the context of the U.S. Employee Retirement Income Security Act, 1974 (ERISA), it is generally regarded that the smoothed actuarial value should reflect current market values [Winklevoss (1993:172)]. The U.S. Internal Revenue Service (IRS) imposes a 20% corridor of market ('fair') value within which the actuarial value must lie [Anderson (1992:108), McGill *et al.* (1996:679)]. Actuarial Standard of Practice No. 4 of the American Academy of Actuaries (1993:¶5.2.6) thus requires that asset values should "generally reflect some function of market value". The requirements of the Canadian Pension Benefits Standards Act 1985 [Office of the Superintendent of Financial Institutions Canada (1987:43)] are that:

> "Averaging techniques and other smoothing methods are permitted, provided the values obtained do not systematically exceed a reasonable market-related value."

4.4.2 Consistency

Since pension fund valuation is a comparison of asset and liability cash flows, it is important that values placed on assets and liabilities are consistent and comparable. As interest and inflation rates became more unstable and equity investment in pension funds rose during the 1960's, the practice of valuing assets at book value or historic cost declined, precisely because such asset values became increasingly inconsistent with pension liabilities. In North America, market-related methods (§4.1.2) gained pre-eminence. Book values were revised in line with market value changes, using various methods as described by Jackson & Hamilton (1968). Market values, usually smoothed somehow, are now in use. The valuation of assets and liabilities using market methods can be consistent, but there are various practical difficulties in the choice of economic assumptions to achieve this. In the U.K., the Discounted Cash Flow valuation method (§4.1.1) became more popular. Since liabilities and assets are discounted using closely related assumptions, it is claimed that this method yields asset values that are more consistent with liability values.

4.4.3 Realism and Objectivity

It is not a primary objective of a funding or management valuation to place an absolute value on assets, although it is sometimes argued that actuarial asset values can be superior to market values in the long term. If consistency is preserved (whether in the market or Discounted Cash Flow method), a consistent value of the unfunded liability will be obtained, regardless of whether intermediate 'values' placed on assets and liabilities are correct or realistic. Asset values must nevertheless be chosen in an objective fashion and must bear some relationship to reality if an artificial valuation result is to be avoided.

When the market method is used, various averaging techniques are applied to market prices so that short-term volatility is removed. Since most assets do have to be sold eventually—not all assets are held to redemption—and the assumption of perpetuity made when discounting income from equities is not always tenable [Anderson (1992:103)], it is said that market methods are more realistic. Nevertheless, the opacity and variety of averaging techniques employed make market-related asset valuations somewhat arbitrary.

Some actuaries (notably in the U.K.) contend that discounting the cash flow from assets allows the actuary to place a superior long-term 'value' or 'worth' on pension fund assets. This 'value' is considered to be superior to the market price: see for example Day & McKelvey (1964) and Pemberton (1998). This is based on the assumption that the actuary can take into account the degree of mismatch between pension fund assets and liabilities and differences between various market investors in terms of their liabilities or their tax position. But the arbitrariness of the Discounted Income Value is also apparent: it has been shown [Atkinson (1994), Dyson & Exley (1995)] that the Discounted Income Value method, when applied to variable income assets such as equities, is very sensitive to the (usually term-independent) assumption as to income growth; in addition, the choice of a notional asset portfolio (see §4.1.1) is often arbitrary.

4.4.4 Smoothness

According to Bleakney (1972:125),

> "The special asset valuation methods in use are [...] designed to strike a balance between two purposes, which are sometimes in opposition:
> A recognition of each security's intrinsic value at the time of valuation;
> An attempt to gain stability of valuation, so as to avoid fluctuating gains and losses which have no long term significance."

For funding purposes (as opposed to compliance purposes, such as for solvency or accounting), it is perhaps less important to find absolute 'values' of the assets and liabilities: comparable values are required. In this respect, the short-term volatility of pure market values may not reflect the true financial situation of the pension fund. Actuarial asset values (whether in the market or Discounted Cash Flow method) then need to be 'smoothed' so that a correct long-term assessment of the security of pension benefits may be made. In addition, if asset values are smoother and more stable, the resulting surplus or deficiency and hence contributions are smoother and more stable. Thus, the Standard of Practice for Valuation of Pension Plans of the Canadian Institute of Actuaries (1994:¶5.01), allows for

> "a market-related value which moderates the effect of short-term volatility of market values"

while the Actuarial Standard of Practice No. 4 of the American Academy of Actuaries (1993:§5.2.6) requires asset valuation methods that

> "smooth out the effects of short-term volatility in market value".

The Discounted Income Value of assets has been shown by a number of British authors to generate a smoother set of contribution rates than pure market values. This has been used as an argument in favour of using the Discounted Cash Flow method. Indeed, in practical situations, it is smoothness rather than consistency (§4.4.2) that may motivate the use of discounted income values [Atkinson (1994), Dyson & Exley (1995)]. It has also been argued, for example by Ross [source: Lochhead (1994)] and Exley *et al.* (1997), that the Discounted Income Value of assets will only be smooth while the dividend growth assumption is left unchanged: if experience turns out to be different in the long term and this assumption needs to be changed, asset values and consequently contributions will need to be realigned. Clearly, the effect of such changes to the valuation basis should be tempered by a suitable amortization of the 'initial unfunded liability' (see §3.2.4) that is then generated, especially if initial unfunded liabilities are amortized separately. This shows that smoothness is related both to the asset valuation method *and* to the way in which gains/losses or surpluses/deficits are dealt with.

4.4.5 Dynamics

The dynamics of any asset valuation method will affect the pension fund cash flows. Asset-liability models of pension fund are sensitive to the asset valuation method, as noted by Kemp (1996:§8) and Kingsland (1982). Pension fund investment, and thus the *ex post* cost of pensions, will depend on the dynamics of the asset valuation method because asset allocation decisions are taken based on assessments of the liabilities and assets. While achieving 'smoothness', an asset valuation method must not distort the plan trustees' or sponsor's perception of economic reality and lead to wrong investment decisions [Ezra (1979:110)]. Dyson

& Exley (1995) are particularly critical of the Discounted Income Value method in the U.K. for this reason.

The asset valuation method will also affect the timing of contributions: the emergence of investment gains and losses, especially unrealised ones, will be different depending upon how assets are valued. If an asset valuation method does not sufficiently even out the short-term fluctuations of market values, large gains and losses may arise which will lead to volatile contribution rates. It is usually considered that the amortization of these gains/losses (or spreading of the surpluses/deficits) will not be powerful enough on its own to stabilise contributions. For example, Anderson (1992:108) suggests that

> "there are situations where the normal damping of [investment gains (or losses)] through amortization of the gains is not enough: If we want year-to-year stability in pension cost we have to apply extra damping, which means we have to use an artificial asset value in place of the actual market value."

On the other hand, some authors feel that volatility in market values may be dampened by the gain/loss adjustment itself, if this is sufficiently powerful. Trowbridge [Jackson & Hamilton (1968): discussion] states that:

> "There is little reason why common stocks cannot be valued at market on asset side of the pension balance sheet, if at the same time a powerful smoothing device is employed in the calculation of contributions to level out actuarial gain or loss."

Hennington [Jackson & Hamilton (1968): discussion] also believes that:

> "The smoothness of the annual contribution is determined not only by the method for determining asset value but also by the actuarial funding method. [...] An actuarial cost method involving a spreading of actuarial gains and losses makes it easier to use some of the market value methods".

It is not meaningful therefore to consider the issue of asset valuation without addressing contribution adjustment methods. Trowbridge & Farr (1976:73) write of "the consistency between the asset valuation and the techniques of actuarial gain or loss adjustment". (Recall that asset and liability values must also be consistent.) In the following section, a simple analysis of a market-related asset valuation method with surpluses/deficits spread forward is presented.

4.5 A Method for Smoothing Asset Values

4.5.1 Motivation

The fundamental reason for smoothing *asset values* is to generate a stable and smooth pattern of *contribution rates*. As is pointed out by Daskais [Jackson & Hamilton (1968):

discussion], Trowbridge & Farr (1976:93) and others, smooth asset values are not necessary for a smooth sequence of contributions: contributions can be smoothed separately. Dyson & Exley (1995) argue that the asset valuation method *per se* should not be required to yield a smooth value. They draw a distinction between the aim of an asset valuation and the overall objective of smoothness in an ongoing valuation and suggest that a separate mechanism be used to smooth asset values.

Another reason for smoothing asset values is to generate an asset value that is more consistent with the long-term assumptions used when valuing liabilities [Ezra (1979:108)], so that a better long-term assessment of the security of benefits is made. This assumes that any smoothed asset value is consistent with the value placed on the liabilities. It is not clear how one can ensure this consistency. The Standard of Practice for Valuation of Pension Plans of the Canadian Institute of Actuaries (1994:¶5.01) suggests that the asset valuation method will be consistent (with the liability valuation) if

> "the asset valuation method, when considered in conjunction with the assumed rate of investment return (before inclusion of a margin for adverse deviations or compensating adjustments to related assumptions), can reasonably be expected to result in gains and losses which will offset each other over the long term."

Actuarial gains and losses are only consistently measured, however, if one supposes that assets and liabilities are being consistently valued in the first place. Consistency can only genuinely be upheld if the same assumptions are used to discount the liabilities and assets (the Discounted Cash Flow method: §4.1.1) or if the liabilities and assets are both measured at market value (the market method: §4.1.2).

In the following, I will assume that the asset valuation method is a mechanism for producing stable and smooth contribution rates, rather than being a superior measure of value. This is in agreement with the ideas expressed by Dyson & Exley (1995) in the U.K. It is also implied by the U.S. Employee Retirement Income Security Act of 1974 (ERISA) and the Canadian Pension Benefits Standards Act of 1985. This assumption is also central to most North American pension funding theory and practice. Berin (1989:28–29) thus considers asset valuation methods to be "adjusted asset systems" which "should include a smoothing device" and must be "related to Market Value". Ezra (1980) regards the main purpose of actuarial asset values to be to smooth contributions: actuarial asset values are not meant to be superior to market values, but are "damped" or "toned down" in an attempt to remove short-term fluctuations. Likewise, Aitken (1994:289) considers that "actuarial values are market values with dampened volatility." See also Anderson (1992:107) and Winklevoss (1993:171) for further discussion of this point.

4.5.2 Definition

A particular asset valuation method, applied to the whole of the pension fund, is now considered. Let

$$f(t) = \text{the } \textit{market} \text{ value of assets,}$$
$$F(t) = \text{the } \textit{actuarial} \text{ value of the pension fund assets,}$$
$$ul(t) = AL - f(t) = \text{the unfunded liability based on market values,}$$
$$UL(t) = AL - F(t) = \text{the `actuarial' unfunded liability.}$$

$F(t)$ is defined as

- the value of assets as anticipated on the valuation basis after allowing for new cash, with recognition of a fraction of the difference between market value and this anticipated value, or

- a weighted average of the market value of assets and the value of assets anticipated on the valuation basis after allowing for new cash.

For $t \geq 1$,

$$F(t) = \lambda u_v \left[F(t-1) + c(t-1) - B\right] + (1-\lambda) f(t) \tag{4.2}$$
$$= u_v \left[F(t-1) + c(t-1) - B\right] + (1-\lambda)\left[f(t) - u_v(F(t-1) + c(t-1) - B)\right], \tag{4.3}$$

and we may define $F(0) = f_0$, the initial market value of assets.

Remarks:

1. Cash flows (contribution and benefit payments) and the time value of money are being allowed for in equations (4.2) and (4.3).

2. The asset valuation method described by equations (4.2) or (4.3) is sufficiently general to encompass many of the methods described by Jackson & Hamilton (1968), Ferris & Welch (1996) and Aitken (1994:289) (see §4.4.1).

3. It is also mathematically akin to Dyson & Exley's (1995:§7) exponential smoothing method (§4.4.1).

Winklevoss (1993:174) describes a commonly used method to which he refers as a 'corridor' variation on the 'write-up method'. Using his notation, if $(AV)_t$ is the actuarial value of assets in year t then the asset value may be written up to $(AV)'_{t+1}$ after a year, in line with the assumed valuation discount rate i, so that

$$(AV)'_t = [(AV)_{t-1} + C_{t-1} - B_{t-1}](1+i). \tag{4.4}$$

(C_{t-1} and B_{t-1} represent cash flows in year t.) Further, if $(MV)_t$ is the market value of assets, k is an "adjustment fraction", and c_1 and c_2 define a range as a proportion of market values within which the actuarial value must lie, then

$$(AV)_t = \begin{cases} (AV)'_t + k[c_1(MV)_t - (AV)'_t], & \text{if } (AV)'_t < c_1(MV)_t, \\ (AV)'_t - k[(AV)'_t - c_2(MV)_t], & \text{if } (AV)'_t > c_2(MV)_t, \\ (AV)'_t, & \text{otherwise.} \end{cases} \tag{4.5}$$

The method described in equation (4.2) or (4.3) is identical to the method described by Winklevoss (1993:174) when $c_1 = c_2 = 1$ and $k = 1 - \lambda$, i.e. the actuarial value is written up based on the return assumed in the valuation basis and is also adjusted relative to the market value *even if* the actuarial value is within a given range (or 'corridor') of the market value. The discontinuity mentioned by Winklevoss (1993:174) is effectively avoided by not setting a corridor. But the method described in equation (4.2) or (4.3) is not compliant with the ERISA 20% corridor requirement (see §4.4.1). However, if the smoothing parameter λ is properly chosen, the probability that asset values breach the corridor may be reduced.

4.5.3 Model

Trowbridge (1952), Bowers *et al.* (1976, 1979, 1982), Benjamin (1989), O'Brien (1986, 1987), Dufresne (1986, 1988, 1989), Haberman (1992a, 1994a, b), Sung (1997), Boulier *et al.* (1995, 1996), Cairns & Parker (1997) and Loades (1998) consider pure market values of assets only. We must assume, for consistency, that the liabilities are being valued at some average market-related valuation discount rate. In their simulation models, Loades (1992) and Fujiki (1994) use a Discounted Income Value for the assets, with a discount rate consistent with the one used to evaluate the liabilities.

It is also assumed that surpluses and deficits are spread forward over a moving term (§3.2). We make this assumption for several reasons:

1. We have shown that spreading over a moving term is more efficient than amortizing over a fixed term (§3.5).

2. Several authors have observed that using the aggregate funding methods or using a moving term to remove surpluses and deficits seems to result in 'smoother' funding: see Hennington [Jackson & Hamilton (1968): discussion] (§4.4.5), Trowbridge & Farr (1976:62) (§3.5.4).

3. The spreading method deals with a fraction (k) of the surplus or deficit in any year. The asset valuation method that we consider here similarly deals with a fraction $(1-\lambda)$ of the unanticipated change in asset value.

We anticipate a relationship between spreading and smoothing, and between k and λ, on account of item 3 above. It turns out to be more convenient to introduce $K = 1-k = 1-1/\ddot{a}_{\overline{m}|}$. As m and K increase, deficits and surpluses are spread further (note also that when $m = 1$, $K = 0$). As λ increases, less weight is placed on the current market value and asset values are more heavily smoothed.

Assume that surpluses/deficits are spread over a moving period m, with a fraction $(1-y)$ $(0 \leq y < 1)$ of the initial unfunded liability being amortized over a fixed term n, so that

$$\begin{aligned} c(t) &= NC + k(UL(t) - U(t)) + P(t), \\ &= NC + (1-K)(UL(t) - U(t)) + P(t) \end{aligned} \tag{4.6}$$

with

$$P(t) = \begin{cases} (1-y)ul_0/\ddot{a}_{\overline{n}|}, & 0 \leq t \leq n-1, \\ 0, & t \geq n, \end{cases} \tag{4.7}$$

$$U(t) = \begin{cases} (1-y)ul_0\ \ddot{a}_{\overline{n-t}|}/\ddot{a}_{\overline{n}|}, & 0 \leq t \leq n-1, \\ 0, & t \geq n. \end{cases} \tag{4.8}$$

Note that

$$P(t) = U(t) - v_v U(t+1), \tag{4.9}$$

where we define $U(n) = U(n+1) = \ldots = 0$.

Rewriting equation (4.2) in terms of $ul(t)$ and $UL(t)$, we obtain

$$UL(t) = \lambda u_v \left[UL(t-1) - c(t-1) + B - d_v AL\right] + (1-\lambda)ul(t), \tag{4.10}$$

$$= \lambda u_v \left[UL(t-1) - c(t-1) + NC\right] + (1-\lambda)ul(t), \tag{4.11}$$

where we have used the equation of equilibrium (3.65). Now, replace $P(t)$ from equation (4.9) into equation (4.6), and substitute $c(t)$ into equation (4.11) to give

$$UL(t) = \lambda u_v \left[K(UL(t-1) - U(t-1)) + v_v U(t)\right] + (1-\lambda)ul(t), \tag{4.12}$$

$$(UL(t) - U(t)) = \lambda K u_v (UL(t-1) - U(t-1)) + (1-\lambda)(ul(t) - U(t)), \tag{4.13}$$

for $t \geq 1$.

Consider first a situation where surpluses and deficits are not spread forward and are repaid immediately ($m = k = 1$, $K = 0$). A simple relationship between the 'actuarial' and market value of the fund follows from equation (4.13). In terms of unfunded liabilities,

$$UL(t) - U(t) = (1-\lambda)(ul(t) - U(t)). \tag{4.14}$$

Furthermore, from equation (4.6), the recommended contribution becomes

$$c(t) = NC + (1-\lambda)(ul(t) - U(t)) + P(t). \tag{4.15}$$

Now, this form of adjustment to the contribution is exactly identical to the one existing when surpluses and deficits are spread and 'initial unfunded liabilities' are amortized. Equation (4.15) is identical to equation (3.75), except that $1-\lambda$ replaces k. We recognise that Proposition 3.1 applies here, with the appropriate substitution of k by $1-\lambda$. This symmetry is not surprising. The common motivation of smoothing underlies both the methods of spreading forward surpluses/deficits and smoothing asset values.

The more general case may now be considered. At $t = 0$, we have defined $UL(0) = ul_0$ and $U(0) = (1-y)ul_0$, and so $(UL(0) - U(0)) = yul_0$. Hence, it is easily verified that equation (4.13) may be rewritten as

$$(UL(t) - U(t)) = (1-\lambda)\sum_{j=0}^{t} p^{t-j}(ul(j) - U(j)) + \lambda p^t yul_0, \tag{4.16}$$

for $t \geq 0$ and where $p = \lambda K u_v$.

The recommended contribution is therefore, from equations (4.6), (4.9) and (4.16),

$$c(t) = NC + (1-K)(1-\lambda)\sum_{j=0}^{t} p^{t-j}(ul(j) - U(j)) + (1-K)\lambda p^t yul_0 + U(t) - v_v U(t+1). \tag{4.17}$$

It depends on an exponentially weighted sum of past and present unfunded liabilities (at market value).

We may now obtain a recurrence relationship for the unfunded liability. Substituting equation (4.17) for $adj(t)$ into equation (3.66), and rearranging,

$$\begin{aligned}(ul(t+1) &- U(t+1)) \\ &= (AL - U(t+1)) + u(t+1)\Bigg[(ul(t) - U(t)) - (1-K)(1-\lambda)\sum_{j=0}^{t} p^{t-j}(ul(j) - U(j)) \\ &\qquad - v_v(AL - U(t+1)) - (1-K)\lambda p^t yul_0\Bigg],\end{aligned} \tag{4.18}$$

for $t \geq 0$.

4.5.4 First Moments

Pure Smoothing. The direct correspondence between pure spreading (with $\lambda = 0$) and pure smoothing (with $m = k = 1$ or $K = 0$), which we noted earlier in §4.5.3, leads to Proposition 4.1 (cf. Result 3.3 on page 59). We note also that the results concerning $F(t)$ follow from taking expectation and then limits (as $t \to \infty$) on equation (4.14), so that $\lim \mathrm{E}F(t) = (1-\lambda)\lim \mathrm{E}f(t)$.

PROPOSITION 4.1 *Suppose that surpluses and deficits are not spread forward and are repaid in full ($m = 1$). Then, provided $|\lambda u| < 1$,*

$$\lim_{t\to\infty} \mathrm{E}ul(t) = AL(d - d_v)/(d - (1 - \lambda)), \tag{4.19}$$

$$\lim_{t\to\infty} \mathrm{E}UL(t) = AL(1 - \lambda)(d - d_v)/(d - (1 - \lambda)), \tag{4.20}$$

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL(d_v - (1 - \lambda))/(d - (1 - \lambda)), \tag{4.21}$$

$$\lim_{t\to\infty} \mathrm{E}F(t) = AL - AL(1 - \lambda)(d - d_v)/(d - (1 - \lambda)), \tag{4.22}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + AL(1 - \lambda)(d - d_v)/(d - (1 - \lambda)). \tag{4.23}$$

Combined Smoothing and Spreading. We may take expectations across equation (4.18), noting that $u(t+1)$ is independent of $u(t)$ and $ul(t)$, $ul(t-1)$ etc., to obtain

$$\begin{aligned}\mathrm{E}(ul(t+1) - U(t+1)) \\ &= (AL - U(t+1))(1 - uv_v) + u\mathrm{E}(ul(t) - U(t)) \\ - u(1-K)(1-\lambda)\sum_{j=0}^{t} p^{t-j}\mathrm{E}(ul(j) - U(j)) - u\lambda(1-K)p^t y ul_0,\end{aligned} \tag{4.24}$$

for $t \geq 0$. We forward-shift equation (4.24) in time (so that it holds for $t \geq -1$) and upon deducting equation (4.24) multiplied by p, we obtain

$$\begin{aligned}\mathrm{E}[ul(t+2) - U(t+2)] \\ - [p + u - u(1-K)(1-\lambda)]\mathrm{E}[ul(t+1) - U(t+1)] + up\mathrm{E}[ul(t) - U(t)] \\ = AL(1-p)(1-uv_v) - [U(t+2) - pU(t+1)](1 - uv_v),\end{aligned} \tag{4.25}$$

which holds for $t \geq 0$ and requires $\mathrm{E}(ul(0) - U(0)) = y\ ul_0$ and an additional initial condition $\mathrm{E}(ul(1) - U(1))$ that may be found from equation (4.24). These initial conditions have only a transient effect, and we will examine the situation in the limit only and so do not need them.

The characteristic equation of the second order difference equation (4.25) is

$$P(z) = z^2 - [p + u - u(1-K)(1-\lambda)]z + up = 0. \tag{4.26}$$

$U(t) \to 0$ as $t \to \infty$, from equation (4.8). $\mathrm{E}ul(t)$ converges as $t \to \infty$ provided the roots of the characteristic equation (4.26) are less than unity in magnitude. This will occur if (Haberman, 1992a):

$$|up| < 1, \tag{4.27}$$

$$P(z=1) > 0 \quad \Leftrightarrow \quad 1 - u - p + up + u(1-K)(1-\lambda) > 0, \tag{4.28}$$

$$P(z=-1) > 0 \quad \Leftrightarrow \quad 1 + u + p + up - u(1-K)(1-\lambda) > 0. \tag{4.29}$$

Two simplifying assumptions are now made: both the valuation discount rate (i_v) and the mean long-term rate of return on assets (i) are greater than -100%. These assumptions are not restrictive in practice and mean respectively that $u_v > 0$ and $u > 0$. If $i_v > -100\%$, then $0 \leq Ku_v < 1$ (Haberman, 1992a), since $K = 1 - 1/\ddot{a}_{\overline{m}|}$. Given that, by definition, $0 \leq \lambda < 1$, it is also clear that $0 \leq \lambda K u_v = p < 1$.

It is convenient to define

$$\theta = (1-K)(1-\lambda)/(1-p) = (1-K)(1-\lambda)/(1-\lambda K u_v). \tag{4.30}$$

Condition (4.27) can be written as

$$\lambda K u u_v < 1. \tag{4.31}$$

As for condition (4.28), it may be written as $[1-p][1-u(1-\theta)] > 0$, which, since $0 \leq p < 1$ and $u > 0$, simplifies to $\theta > d$. Condition (4.29) follows directly from the additional assumptions that $u_v > 0$ and $u > 0$ (since $u - u(1-K)(1-\lambda) > 0$ and $1 + p + up > 0$).

$\lim \mathrm{E}f(t)$ follows immediately, given these stability conditions. Taking expectations across equation (4.13) and then taking limits, we find that $\lim \mathrm{E}UL(t) = \lim \mathrm{E}ul(t)(1-\lambda)/(1-\lambda K u_v)$ provided

$$-1 < p = \lambda K u_v < 1, \tag{4.32}$$

which holds as shown above. Also, $\lim \mathrm{E}F(t) = AL - \lim \mathrm{E}UL(t)$. Finally, taking expectations and limits across equation (4.6) shows that $\lim \mathrm{E}c(t) = NC + k \lim \mathrm{E}UL(t)$.

These results are summarised in the following proposition:

PROPOSITION 4.2 *Let $\theta = (1-K)(1-\lambda)/(1-\lambda K u_v)$. Provided that*

$$i_v > -100\%, \tag{4.33}$$
$$i > -100\%, \tag{4.34}$$
$$\lambda K u u_v < 1, \tag{4.35}$$
$$\theta > d, \tag{4.36}$$

then

$$\lim_{t\to\infty} \mathrm{E}ul(t) = AL(d-d_v)/(d-\theta), \tag{4.37}$$
$$\lim_{t\to\infty} \mathrm{E}UL(t) = AL\theta(d-d_v)/k(d-\theta), \tag{4.38}$$
$$\lim_{t\to\infty} \mathrm{E}f(t) = AL(d_v-\theta)/(d-\theta), \tag{4.39}$$
$$\lim_{t\to\infty} \mathrm{E}F(t) = AL - AL\theta(d-d_v)/k(d-\theta), \tag{4.40}$$
$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + AL\theta(d-d_v)/(d-\theta). \tag{4.41}$$

Remarks:

1. The similarity with Proposition 4.1 on page 131 and Result 3.3 on page 59 is self-evident. Equation (4.30) reveals that $\theta = k = 1 - K$ when $\lambda = 0$ (pure spreading) and $\theta = 1 - \lambda$ when $m = k = 1$ or $K = 0$ (pure smoothing).

2. There is symmetry between K and λ: they can be interchanged without affecting the first moments of the fund and contribution levels (θ remains unchanged). If surpluses and deficits are spread over a period such that $K = K_1$ and if the weight placed on current market value is $1 - \lambda = 1 - \lambda_1$, then we *expect* exactly the same fund and contribution levels in the limit were a spreading period such that $K = \lambda_1$ and a smoothing parameter of $\lambda = K_1$ to be used. This serves to illustrate that the spreading and smoothing processes complement each other.

3. Assuming that there is no initial unfunded liability, or that it has been completely amortized, then the symmetry between K and λ holds throughout the dynamics of the funding process: K and λ are interchangeable in equations (4.17) and (4.18) (with $ul_0 = 0$).

4. Note that this symmetry does not extend to the 'actuarial' asset value, as can be observed in equations (4.16), (4.38) and (4.40).

4.5.5 First Moments with Unbiased Discount Rate

Simpler results follow if we assume that the valuation discount rate equals the mean rate of return ($i_v = i$ or $d_v = d$). Then conditions (4.27) and (4.32) become

$$|\lambda K u^2| < 1, \tag{4.42}$$

$$|\lambda K u| < 1. \tag{4.43}$$

respectively. The right hand side of condition (4.28) factors to $(1 - Ku)\ (1 - \lambda u)$ while the right hand side of condition (4.29) factors to $(1 + Ku)\ (1 + \lambda u)$:

$$(1 - \lambda u)[1 - Ku] > 0, \tag{4.44}$$

$$(1 + \lambda u)[1 + Ku] > 0. \tag{4.45}$$

These stability conditions are not too restrictive under 'normal' economic conditions. If we again make the further assumption that $i > -100\%$, then $0 \leq Ku < 1$ (Haberman, 1992a) or $0 \leq K < v$. Given that $0 \leq \lambda < 1$, then clearly $0 \leq p = \lambda Ku < 1$ so that stability condition (4.43) holds. For condition (4.44) to hold, we require that $\lambda < v$; inequality (4.42) will hold if $\lambda < v < v/(Ku)$; and the left hand side of condition (4.45) is clearly positive. Hence, it is sufficient that $i_v = i > -100\%$, $0 \leq K < v$ and $0 \leq \lambda < v$ for the first moments to be stable as $t \to \infty$. Putting $d = d_v$ in equations (4.37)–(4.41) gives the required moments. This is summarised in Corollary 4.1.

COROLLARY 4.1 *Provided that*

$$i = i_v > -100\%, \tag{4.46}$$

$$0 \leq K < v, \tag{4.47}$$

$$0 \leq \lambda < v, \tag{4.48}$$

then

$$\lim_{t\to\infty} \mathrm{E}ul(t) = \lim_{t\to\infty} \mathrm{E}UL(t) = \lim_{t\to\infty} \mathrm{E}adj(t) = 0, \tag{4.49}$$

$$\lim_{t\to\infty} \mathrm{E}f(t) = \lim_{t\to\infty} \mathrm{E}F(t) = AL, \tag{4.50}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC. \tag{4.51}$$

Remarks:

1. When $\lambda = 0$, an unsmoothed market value is used (from equation (4.2)). The results of Dufresne (1986) (Result 3.1 on page 57) should follow. This is easily verified to be true since $\lambda = 0 \Rightarrow \theta = 1 - K = k$ (equation (4.30)).

2. Note from Proposition 4.2 that when $i \neq i_v$,

$$\lim_{t\to\infty} \mathrm{E}UL(t) = \lim_{t\to\infty} \mathrm{E}ul(t) - AL(k-\theta)(d-d_v)/k(d-\theta), \tag{4.52}$$

$$\lim_{t\to\infty} \mathrm{E}F(t) = \lim_{t\to\infty} \mathrm{E}f(t) + AL(k-\theta)(d-d_v)/k(d-\theta). \tag{4.53}$$

   When asset values are smoothed and the valuation discount rate is not equal to the long-term expected rate of return (when $\lambda > 0$ and $i \neq i_v$),

   (a) the long-term average actuarial and market values of the fund do not coincide (from equation (4.53)): this is because the return anticipated in the valuation basis, and thus the anticipated value of the fund when the actuarial value is smoothed, are not realised on average;

   (b) the fund does not meet liabilities on average (from equations (4.39) and (4.40)): neither the fund level measured at market, nor the smoothed actuarial value of the fund, eventually equals the actuarial liability.

3. When asset values are smoothed and the valuation discount rate *is* equal to the long-term expected rate of return (when $\lambda > 0$ and $i = i_v$; i_v contains no margin), then over the long-term, no gain/loss is expected to emerge (equation (4.49)). This exactly satisfies the criterion for consistency between asset and liability valuations stipulated in the Standard of Practice for Valuation of Pension Plans of the Canadian Institute of Actuaries (1994:¶5.01), as quoted in §4.5.1.

4.5.6 Second Moments

For the rest of this chapter, we make the simplifying assumption that the valuation discount rate is equal to the long-term expected rate of return ($i = i_v$).

Pure Smoothing. Proposition 4.3 is obtained straightaway by virtue of the direct relationship between pure spreading and pure asset smoothing described in §4.5.3. The second moment of $F(t)$ follows from taking the variance and then limits (as $t \to \infty$) on equation (4.14), so that $\lim \mathrm{Var}F(t) = (1-\lambda)^2 \lim \mathrm{Var}f(t)$.

PROPOSITION 4.3 *Suppose that surpluses and deficits are not spread and are paid immediately ($m = 1$). Let $q = u^2 + \sigma^2$. Provided that $|\lambda u| < 1$ and $q\lambda^2 < 1$, then*

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \sigma^2 v^2 AL^2/(1 - q\lambda^2) = V_\infty \quad \textit{(say)}, \tag{4.54}$$

$$\lim_{t\to\infty} \mathrm{Varc}(t) = \lim_{t\to\infty} \mathrm{Var}F(t) = (1-\lambda)^2 V_\infty. \tag{4.55}$$

Combined Smoothing and Spreading. It is more difficult to determine the variability in the fund and contribution processes when, at once, surpluses and deficits are spread *and* asset values for actuarial purposes are smoothed.

PROPOSITION 4.4 *Let $V_\infty = \sigma^2 v^2 AL^2/Q$, where*

$$Q = (1-qK^2)(1-\lambda^2u^2)(1-\lambda K u^2) - \lambda(1-K)\sigma^2[2K(1-\lambda^2u^2) + \lambda(1-K)(1+\lambda K u^2)], \tag{4.56}$$

$$= (1-q\lambda^2)(1-K^2u^2)(1-\lambda K u^2) - K(1-\lambda)\sigma^2[2\lambda(1-K^2u^2) + K(1-\lambda)(1+\lambda K u^2)]. \tag{4.57}$$

Provided that conditions (4.46)–(4.48) *are true and also provided that*

$$Q \;>\; 0 \tag{4.58}$$

and that

$$(1+\lambda^2K^2qu^2)(1+\lambda^3K^3\sigma^2u^2 - \lambda^4K^4qu^6)$$
$$> 2\lambda^4K^4(\lambda+K)q\sigma^2u^4 + \lambda K(\lambda+K)^2qu^2(1-\lambda^2K^2qu^2), \tag{4.59}$$

then

$$\lim_{t\to\infty} \mathrm{Var}f(t) = V_\infty[(1-\lambda K u^2)(1-\lambda^2K^2u^2) + 2\lambda K(1-\lambda)(1-K)u^2], \tag{4.60}$$

$$\lim_{t\to\infty} \mathrm{Var}F(t) = V_\infty(1-\lambda)^2(1+\lambda K u^2), \tag{4.61}$$

$$\lim_{t\to\infty} \mathrm{Varc}(t) = V_\infty(1-K)^2(1-\lambda)^2(1+\lambda K u^2), \tag{4.62}$$

$$\lim_{t\to\infty} \text{Cov}[f(t), F(t)] = V_\infty(1-\lambda)[1+\lambda K(1-K-\lambda)u^2], \tag{4.63}$$

$$\lim_{t\to\infty} \text{Cov}[f(t), c(t)] = -V_\infty(1-K)(1-\lambda)[1+\lambda K(1-K-\lambda)u^2], \tag{4.64}$$

$$\lim_{t\to\infty} \text{Cov}[c(t), F(t)] = -V_\infty(1-K)(1-\lambda)^2(1+\lambda K u^2). \tag{4.65}$$

Proof in Appendix D.

Remarks:

1. Suppose pure market values of assets are used ($\lambda = 0$). Then, from equation (4.56), $Q = 1 - qK^2$, and second moment results identical to those of Dufresne (1986, 1988) (Result 3.1 on page 57) are obtained. Note also that, unsurprisingly, $\lim \text{Var}F(t) = \lim \text{Var}f(t)$ in this case.

2. Suppose now that smoothed asset values are used, but that surpluses and deficits are not spread ($m = 1$). Then, from equation (4.57), $Q = 1 - q\lambda^2$, and the same second moments as in Proposition 4.3 emerge.

3. The second moments, like the first moments, are also completely symmetrical between K and λ. K and λ can be interchanged without changing the various second moments and covariances of the fund and contribution levels. Q remains unchanged. If surpluses and deficits are spread over a period such that $K = K_1$ and if the weight placed on current market value is $\lambda = \lambda_1$, then the fund and contribution levels will exhibit the same variability and co-variability structure in the limit had a spreading period such that $K = \lambda_1$ and a smoothing parameter of $\lambda = K_1$ been used. It is again clear that surplus/deficit spreading and asset smoothing are performing the same function.

4. Assuming that there is no initial unfunded liability, or that it has been completely amortized, then the symmetry between K and λ holds over time in all moments of the funding process: K and λ are interchangeable in equations (4.17) and (4.18) (with $ul_0 = 0$).

5. K and λ are *not* symmetrical in the 'actuarial' asset value, as can be observed in equations (4.16), (4.61), (4.63) and (4.65).

About Stability. Conditions (4.46)–(4.48), (4.58) and (4.59) are *sufficient* for stability in the second moments. Necessary and sufficient conditions are discussed in Appendix D (§D.2). Under 'normal' economic conditions, the sufficient conditions are accurate and the most constraining condition is inequality (4.58). Upper bounds are easily placed on this condition.

It is shown in the proof of Proposition 4.4 that stability conditions (4.46)–(4.48) and (4.58) imply that

$$K < 1/\sqrt{q}, \tag{4.66}$$

$$\lambda < 1/\sqrt{q}, \tag{4.67}$$

(see equations (D.113) and (D.114) in Appendix D). It is instructive to compare inequality (4.66) with the second condition for stability in Result 3.1 in the case of simple spreading, and also to compare inequality (4.67) with the second convergence condition in Proposition 4.3.

Table 4.1 on the next page exhibits the stability constraints in terms of maximum allowable spread periods for various choices of $\{i,\ \sigma,\ \lambda\}$. Table 4.2 on page 139 shows maximum allowable smoothing parameters for various choices of $\{i,\ \sigma,\ m\}$. Both tables are based on stability conditions (4.46)–(4.48), (4.58) and (4.59).

Inequalities (4.66) and (4.67) are easily verified from Tables 4.1 and 4.2. It is clear from Table 4.1 that the more asset smoothing is applied (i.e. the less weight is placed on the current market value of assets), the shorter the spreading period for surpluses and deficits should be to maintain stability in the second moments. Table 4.2 indicates that the longer the period over which surpluses and deficits are spread, the less asset values should be smoothed if the funding process is to remain stationary in the limit. We conclude that:

1. Very long spreading periods must be avoided, as emphasised by Dufresne (1986, 1988), lest the funding process becomes unstable eventually. Likewise, excessive smoothing, especially in combination with surplus/deficit spreading, must be avoided as the funding process may become non-stationary.

2. Asset gain/loss amortization or spreading and asset smoothing perform a complementary actuarial smoothing function. One must have regard to the cumulative amount of smoothing through both asset valuation and gain/loss adjustment.

4.5.7 Stability of 'Actuarial' Asset Value

It has been understood thus far that the smoothed 'actuarial' asset value $F(t)$

– should remain close to the market value $f(t)$ of the assets, and

– should be more stable than the market value.

These are held to be important properties of an acceptable asset valuation method. It is now possible to verify that these properties hold for the smoothing method described in equations (4.2) or (4.3).

σ	i	$\lambda = 0$	$\lambda = 20\%$	$\lambda = 40\%$	$\lambda = 60\%$	$\lambda = 80\%$	$\lambda = 90\%$
0.05	1%	222	222	221	219	214	203
	3%	110	110	109	107	102	89
	5%	78	77	76	74	68	75
	10%	48	47	46	43	37	13
	15%	36	35	34	32	23	
0.1	1%	112	111	110	109	104	94
	3%	67	67	66	64	59	47
	5%	51	50	49	47	42	29
	10%	33	33	32	30	23	5
	15%	25	25	25	22	14	
0.15	1%	65	65	64	62	57	48
	3%	45	45	44	42	37	27
	5%	36	36	35	33	27	16
	10%	25	25	24	22	16	1
	15%	21	20	19	17	9	
0.2	1%	42	41	41	39	34	26
	3%	32	32	31	28	24	20
	5%	27	27	24	22	19	9
	10%	20	20	17	15	11	
	15%	17	16	15	13	7	
0.25	1%	29	29	28	26	22	14
	3%	24	24	23	21	16	9
	5%	21	20	20	18	13	5
	10%	16	16	15	13	8	
	15%	14	13	13	11	5	

Table 4.1: Maximum allowable spread periods for various mean real (net of salary inflation) returns i, standard deviation of return σ and smoothing parameter λ based on stability conditions (4.46)–(4.48), (4.58) and (4.59). Blanks indicate that conditions do not hold.

σ	i	$m = 1$	3	5	10	15	20	25	30	40	50
0.05	1%	98.9	98.9	98.9	98.9	98.8	98.8	98.8	98.8	98.7	98.6
	3%	97.0	97.0	96.9	96.9	96.8	96.7	96.6	96.4	96.1	95.5
	5%	95.1	95.1	95.1	95.0	94.8	94.6	94.2	93.8	92.7	91.1
	10%	90.8	90.8	90.7	90.4	89.7	88.7	87.4	85.5	76.3	
	15%	86.9	86.8	86.7	86.0	84.5	82.3	78.6	70.5		
0.1	1%	98.5	98.5	98.4	98.4	98.3	98.2	98.0	97.8	97.5	97.0
	3%	96.6	96.6	96.5	96.2	95.9	95.5	94.9	94.2	92.4	88.9
	5%	94.8	94.7	94.6	94.1	93.4	92.5	91.2	89.6	82.9	34.7
	10%	90.5	90.4	90.1	88.8	86.7	83.6	78.0	63.2		
	15%	86.6	86.4	85.9	83.6	79.5	71.4	40.8			
0.15	1%	97.9	97.8	97.7	97.4	97.0	96.6	96.0	95.2	93.1	88.9
	3%	96.1	95.9	95.7	95.0	94.0	92.7	90.9	88.3	73.8	
	5%	94.3	94.0	93.7	92.5	90.7	88.1	84.0	75.5		
	10%	90.1	89.6	89.0	86.2	81.5	71.9	31.7			
	15%	86.2	85.6	84.6	79.8	70.3	31.5				
0.2	1%	97.1	96.9	96.6	95.8	94.5	92.9	90.6	86.8	54.9	
	3%	95.3	94.9	94.5	92.8	90.4	86.6	79.4	58.0		
	5%	93.6	93.0	92.4	89.8	85.7	78.2	55.0			
	10%	89.4	88.6	87.4	82.1	71.5	25.2				
	15%	85.7	84.6	82.7	74.2	49.7					
0.25	1%	96.1	95.6	95.0	92.9	89.6	83.9	70.1			
	3%	94.3	93.6	92.7	89.3	83.3	69.4				
	5%	92.6	91.7	90.5	85.5	75.9	40.5				
	10%	88.6	87.3	85.2	75.9	47.5					
	15%	85.0	83.2	80.2	65.7						

Table 4.2: Maximum allowable smoothing parameter (%) for various mean real (net of salary inflation) returns i, standard deviation of return σ and spread period m based on stability conditions (4.46)–(4.48), (4.58) and (4.59). Blanks indicate that conditions do not hold.

PROPOSITION 4.5 *Assuming stability conditions* (4.46)–(4.48), (4.58) *and* (4.59) *of Proposition 4.4,*

$$\lim_{t\to\infty} \mathrm{E}[f(t) - F(t)]^2 < \infty, \tag{4.68}$$

$$\lim_{t\to\infty} \mathrm{Var}F(t) \le \lim_{t\to\infty} \mathrm{Var}f(t). \tag{4.69}$$

Proof in Appendix E (§E.1).

Remarks:

1. The first part of Proposition 4.5 indicates that the deviation between asset value at market and actuarial asset value as defined in equation (4.2) or (4.3) is bounded in the mean-square. The actuarial asset value over the long term remains in the proximity of market value. This is a requirement in several jurisdictions (§4.4.1).

2. The second part of Proposition 4.5 shows that the smoothed actuarial value of assets is less variable than the pure market value of assets. (See §4.5.1.) This, together with the fact that no gain or loss emerges on average in the long term (§4.5.5), qualifies the method as a suitable one for asset valuation.

3. Equality in equation (4.69) should follow if $\lambda = 0$, as can indeed be verified in equation (E.3) in Appendix E.

4. It is also easily verified from previous results that inequality (4.69) holds if pure smoothing (i.e. without spreading deficits and surpluses) is used: compare equations (4.54) and (4.55) ($0 \le \lambda < 1$ by definition).

4.5.8 Effect of Smoothing and Spreading on Fund Level

Pure Spreading. Dufresne (1986, 1988) shows that spreading surpluses and deficits over longer periods leads to more variable fund levels (Chapter 3) which may endanger the security of funded retirement benefits. This conforms with conventional expectations because a slower recognition of gains and losses should affect security.

Pure Smoothing. Suppose that that asset values are smoothed and also that the fund is disbursed immediately to cover deficits and surpluses. An immediate consequence of the congruence of 'pure' asset smoothing and 'pure' spreading, as described earlier (§4.5.3), is that more smoothing means more variable fund levels—the proof, by direct correspondence with Dufresne's (1986, 1988) result, is trivial (see also Proposition 4.7). This result also agrees with intuition: asset smoothing implies deferred recognition of asset gains and losses. We do not anticipate that asset smoothing will stabilise fund levels at market value, as the primary aim of asset smoothing is to stabilise contributions (§4.5.1).

Combined Smoothing and Spreading. We may now show that the combined effect of asset smoothing and surplus/deficit spreading is to increase the variability of the pension fund level.

PROPOSITION 4.6 *For stable* $\{m, \lambda\}$ *satisfying conditions* (4.46)–(4.48), (4.58) *and* (4.59), $\lim \mathrm{Var} f(t)$ *increases monotonically with both* m *and* λ.

Proof in Appendix E (§E.2). This result is illustrated in Figure 4.1 on the next page (K has a direct one-to-one relationship with m).

Remarks:

1. Heavier smoothing of asset values and spreading of surpluses/deficits over longer periods both lead to more variable fund levels at market value. This is not surprising as both cause asset gains/losses not to be recognised immediately. In addition, heavier smoothing results in more weight being placed on delayed asset value data rather than current market value and delay is known to increase the variability of the pension funding process (§3.2.3).

2. The *combined* effect of spreading and smoothing must be considered in pension funding. Care should be taken that spreading and smoothing do not lead together to excessive smoothing and very variable fund levels.

3. It is clear again that spreading and smoothing perform a similar overall smoothing function in pension funding. They are complementary. Figure 4.1 on the following page exhibits a clear symmetry in the plane $K = \lambda$.

4.5.9 Effect of Smoothing and Spreading on Contribution Level

Pure Spreading. A principal aim of spreading surpluses and deficits over longer periods is to smooth out fluctuations in gains and losses and stabilise contributions. However, Dufresne (1986, 1988) shows that spreading over a term beyond m^* (corresponding to $K^* = 1/q$) is counter-effective, as the long-term variability in contribution levels then increases as the spreading period increases (Result 3.4).

Pure Smoothing. The motivation for using a smoothed 'actuarial' asset value rather than market values is to generate more stable contribution requirements, as described in §4.5.1. In exact symmetry to the result of Dufresne (1986, 1988), it is possible to show that smoothing beyond a certain amount ($\lambda > \lambda^* = 1/q$) is counter- effective: contributions become more variable if the weight placed on current market value of assets is reduced to less than $1 - \lambda^*$. This is summarised in the following proposition.

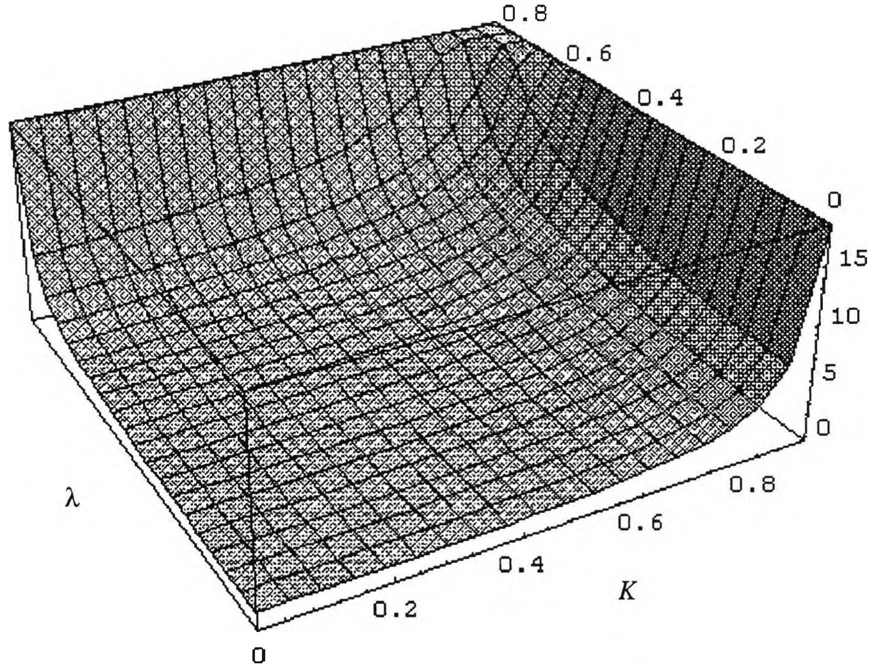


Figure 4.1: $\lim \text{Var}f(t)/AL^2\sigma^2v^2$ against K and λ. $i = 10\%$, $\sigma = 0.05$.

PROPOSITION 4.7 *Suppose that surpluses/deficits are not spread and paid in full immediately ($m = 1$). Then,*

if $0 \leq \lambda < 1/q$, *then* $\lim \text{Var}f(t)$ *increases,* $\lim \text{Var}c(t)$ *and* $\lim \text{Var}F(t)$ *decrease with increasing* λ*;*

if $1/q < \lambda < 1/\sqrt{q}$, *then* $\lim \text{Var}f(t)$, $\lim \text{Var}c(t)$ *and* $\lim \text{Var}F(t)$ *increase with increasing* λ*;*

if $\lambda > 1/\sqrt{q}$, *then* $f(t)$, $c(t)$ *and* $F(t)$ *are non-stationary in the limit.*

Proof. This follows, as a matter of course, by repeating Dufresne's (1986, 1988) proof of Result 3.4, except that $\lim \text{Var}f(t)$ and $\lim \text{Var}c(t) = \lim \text{Var}F(t)$ in equations (4.54) and (4.55) respectively are used and differentiation w.r.t. λ (which exactly replaces $1 - k$) is performed.

Combined Smoothing and Spreading. It is seldom that pure smoothing or pure spreading is used in pension fund valuations. The combined effect of smoothing and spreading must therefore be investigated.

PROPOSITION 4.8 *Suppose* $m > 1$ *and* $\lambda > 0$. *For stable* $\{m, \lambda\}$ *satisfying conditions* (4.46)–(4.48), (4.58) *and* (4.59),

1. *as* m *increases,*

   $\lim \text{Var}c(t)$ *has at least one minimum at some* $m^{*(\lambda)} < m^*$, *provided* $0 < \lambda < \lambda^*$;

$\lim \mathrm{Varc}(t)$ *increases monotonically, provided* either $\lambda \geq \lambda^*$ or $m \geq m^*$;

2. as λ *increases,*

$\lim \mathrm{Varc}(t)$ *has at least one minimum at some* $\lambda^{*(m)} < \lambda^*$, *provided* $1 < m < m^*$;

$\lim \mathrm{Varc}(t)$ *increases monotonically, provided* either $m \geq m^*$ or $\lambda \geq \lambda^*$.

Proof in Appendix E (§E.3). This result is illustrated in Figure 4.3 on the next page (K has a direct one-to-one relationship with m).

Remarks:

1. The first part of Proposition 4.8 is illustrated in Figure 4.2 on the following page. This shows that $\lim \mathrm{Varc}(t)$ against K (for stable $K > 0$ or $m > 1$) has no minimum and increases monotonically when $\lambda = \lambda^* = 0.82$ and $\lambda = 0.85 > \lambda^*$. But when $\lambda < \lambda^*$, a minimum clearly exists. The minimum for $\lambda = 0$ is seen to occur at $K = K^* = 0.82$. The minima for $0 < \lambda < \lambda^*$ clearly occur at some $K^{*(\lambda)} < 0.82$.

2. The two parts of Proposition 4.8 are identical except that m and λ are interchanged. The variation of $\lim \mathrm{Varc}(t)$ with K is similar to its variation with λ. Figure 4.3 on the next page is symmetrical about the plane $K = \lambda$.

3. Smoothing asset values and spreading surpluses/deficits initially reduce the long-term variability of contribution levels. But beyond some critical amount of combined smoothing and spreading, contribution levels become more unstable as a result of increased smoothing or spreading.

4. Once more, the *combined* effect of spreading and smoothing must be considered in pension funding. Excessive spreading and smoothing may result in *more* variable contribution levels, rather than more stable ones.

4.5.10 Optimal Smoothing and Spreading

Pure Spreading. In the situation where market values are used for assets, Dufresne (1986) describes the range of spreading periods $[1, m^*]$ as an 'optimal' one because there always exists a spread period within this range that yields less variable fund levels for the same contribution variability as would occur for any choice of spread period outside this range. This follows from Result 3.4. Dufresne (1988) qualifies the range (m^*, ∞) as 'inadmissible'.

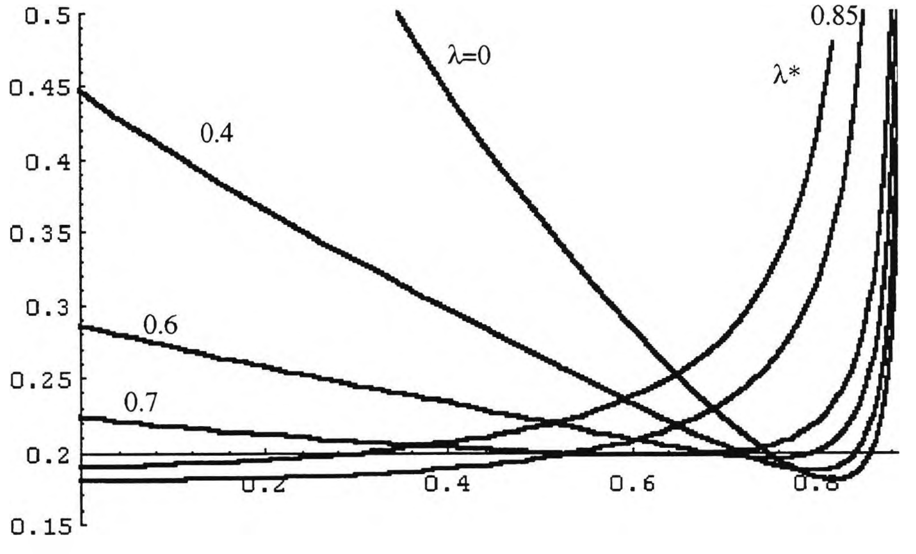


Figure 4.2: $\lim \mathrm{Var}c(t)/AL^2\sigma^2v^2$ against K for various λ. K and λ can be interposed. $i = 10\%$, $\sigma = 0.1$, $\lambda^* = K^* = 0.82$.

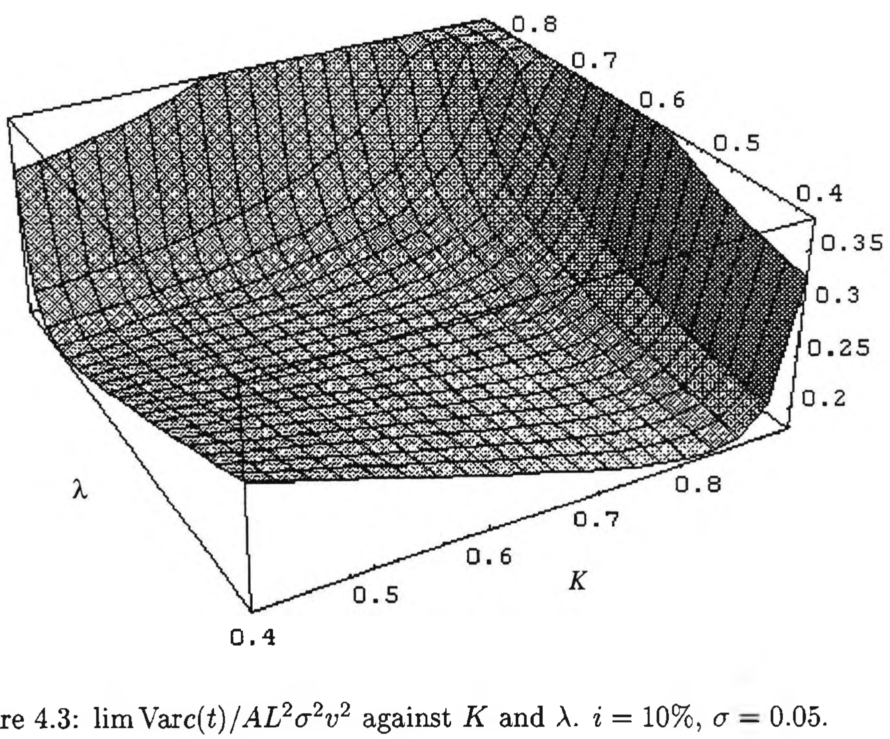


Figure 4.3: $\lim \mathrm{Var}c(t)/AL^2\sigma^2v^2$ against K and λ. $i = 10\%$, $\sigma = 0.05$.

Pure Smoothing. By direct analogy, if surpluses and deficits are not spread forward ($m = 1$), then there exists an 'optimal' smoothing parameter range: $\lambda \in [0, \lambda^*]$. Smoothing within this range will always yield less variable fund levels for the same contribution variability as would occur for any choice of smoothing parameter outside this range. This follows from the fact that $\lim \mathrm{Var} f(t)$ increases with λ but $\lim \mathrm{Var} c(t)$ exhibits a minimum at λ^*: see Proposition 4.7.

Combined Smoothing and Spreading. Propositions 4.6 and 4.8 show that both $\lim \mathrm{Var} f(t)$ and $\lim \mathrm{Var} c(t)$ increase with m (or K) and λ, if *either* $m \geq m^*$ *or* $\lambda \geq \lambda^*$. Since maximising security and minimising contribution stability are rational objectives of pension funding (see §2.2), a situation where *both* $\lim \mathrm{Var} f(t)$ and $\lim \mathrm{Var} c(t)$ increase is *not* admissible. This is summarised in Proposition 4.9.

PROPOSITION 4.9 *It is* not *efficient* either *to spread surpluses/deficits over periods exceeding* m^*, or *to smooth asset values by weighting current market value by less than* $1 - \lambda^*$.

Remarks:

1. The first part of Proposition 4.9 does not conflict with the conclusion of Dufresne (1988) that the range (m^*, ∞) is inadmissible. His conclusion holds, even when smoothed asset values are used.

2. The second part of Proposition 4.9 provides an important lower bound on the weight to be placed on the current market value (i.e. an upper bound on λ). The column for $m = 1$ in Table 4.4 on page 148 shows the upper bound on λ ($\lambda^* = 1/q$) for various choices of $\{i, \sigma\}$.

This section is concluded with some observations from numerical experiments based on the above.

1. There is only one minimum, at $m^{*(\lambda)}$ (say), in the variation of $\lim \mathrm{Var} c(t)$ with m, for $0 < \lambda < \lambda^*$. Table 4.3 on page 147 shows $m^{*(\lambda)}$ for various choices of $\{\lambda, i, \sigma\}$. By comparison with Table 4.2 on page 139, it is clear that $m^{*(\lambda)}$ is less than the maximum allowable spread period for stability in the first and second moments of the pension funding process.

2. There is also only one minimum, at $\lambda^{*(m)}$ (say), in the variation of $\lim \mathrm{Var} c(t)$ with λ, for $1 < m < m^*$. Table 4.3 on page 147 shows $\lambda^{*(m)}$ for various choices of $\{m, i, \sigma\}$. By comparison with Table 4.1 on page 138, it is clear that $\lambda^{*(m)}$ is less than the maximum allowable smoothing parameter for stability in the first and second moments of the pension funding process.

3. From Table 4.3, it is clear that the 'optimal' range of spreading periods $[1, m^{*(\lambda)}]$ shrinks as more asset smoothing is applied. Likewise, in Table 4.4 on page 148 the 'optimal' range of smoothing parameter $[1, m^{*(\lambda)}]$ shrinks as surpluses and deficits are spread over longer periods.

4. Recall from §4.4.1 that the weight placed on current market value is typically in the range 10–33% [Jackson & Hamilton (1968), Colbran (1982), Winklevoss (1993:174), Aitken (1994:289), Dyson & Exley (1995:¶7.5.3.10), Ferris & Welch (1996)]. Table 4.4 indicates that if surpluses and deficits are spread over up to 10 years, then in 'normal' economic conditions, λ should be *at most* 40% if funding is to remain 'optimal', i.e. *the weight placed on current market value should be* at least *60%.* This contrasts markedly with typical practice. For spread periods of around 15 years, the weight placed on current market value should be even higher.

5. It appears that the lowest numerical value of $\lim \text{Var}c(t)$ occurs at $\{(m = m^*, \lambda = 0), (m = 1, \lambda = \lambda^*)\}$. This may be observed from Figure 4.2 on page 144.

6. Figure 4.4 shows contour plots of $\lim \text{Var}f(t)$ and $\lim \text{Var}c(t)$ against K and λ. As K or λ increases, the lighter shading in the first contour plot indicates that $\lim \text{Var}f(t)$ increases.

7. As K or λ increases, the shading in the lower contour plot in Figure 4.4 first darkens and then lightens, indicating a 'trough' where $\lim \text{Var}c(t)$ is lowest. This 'trough' appears to be fairly broad and, in this area, $\lim \text{Var}c(t)$ is fairly insensitive to changes in the degree of smoothing and spreading. This is also a feature of Figure 4.2 on page 144, where the minima in the curves for various λ may be observed to be fairly close for λ between 0 and 0.7.

8. The 'boomerang-shaped' contours of Figure 4.4 on page 149 are a further indication of the complementary function of surplus and deficit spreading and asset smoothing: the same contribution or fund level variability may be achieved by trading off λ and K.

4.6 Summary

This section summarises some of the major points made in this chapter. In §§4.1 to 4.4, the methodology and theory of valuation as applied to the assets and liabilities of a pension fund are explored.

1. Two fundamentally different methods of valuation exist in common practice. In what is termed the Discounted Cash Flow method (§4.1.1), assets are valued by reference to the liabilities using valuation assumptions that are determined by the actuary and

σ	i	$\lambda = 0$	$\lambda = 20\%$	$\lambda = 40\%$	$\lambda = 60\%$	$\lambda = 80\%$	$\lambda = 90\%$
0.05	1%	58	59	59	58	55	49
	3%	23	22	22	21	17	6
	5%	14	14	13	12	8	1
	10%	8	7	7	5	1	†
	15%	5	5	4	3	†	
0.1	1%	42	41	41	39	36	28
	3%	20	19	19	17	14	3
	5%	13	13	12	11	6	1
	10%	7	7	6	5	1	†
	15%	5	5	4	2	†	
0.15	1%	28	27	27	26	22	14
	3%	16	16	15	14	10	1
	5%	11	11	10	9	4	†
	10%	7	6	6	4	2	†
	15%	5	5	4	2	†	
0.2	1%	19	19	18	17	13	5
	3%	13	13	12	11	6	1
	5%	9	9	9	7	3	†
	10%	6	6	5	4	1	
	15%	5	4	4	2	†	
0.25	1%	14	14	13	12	8	1
	3%	10	10	9	8	4	†
	5%	8	8	7	6	2	†
	10%	6	5	5	3	†	
	15%	5	4	3	2	†	

Table 4.3: Spread period $m^{*(\lambda)}$ for various mean real (net of salary inflation) returns $i = i_v$, standard deviation of return σ and smoothing parameter λ at which minimum in $\lim \mathrm{Var}c(t)$ occurs. (†) indicates that $\lim \mathrm{Var}c(t)$ has no minimum and is strictly increasing with its lowest value at $m = 1$. Blanks indicate that stability conditions do not hold.

σ	i	$m = 1$	3	5	10	15	20	25	30	40	50
0.05	1%	97.8	97.7	97.6	97.4	97.1	96.5	95.4	92.7	42.4	14.6
	3%	94.0	93.5	92.8	87.8	41.5	9.5	†	†	†	†
	5%	90.5	88.9	86.0	32.1	†	†	†	†	†	†
	10%	82.5	74.7	39.3	†	†	†	†	†	†	
	15%	75.5	52.6	5.8	†	†	†	†	†		
0.1	1%	97.1	96.9	96.7	96.0	94.8	91.6	73.1	34.5	3.2	†
	3%	93.4	92.6	91.4	80.6	23.9	†	†	†	†	†
	5%	89.9	87.9	83.8	22.8	†	†	†	†	†	†
	10%	82.0	73.0	35.0	†	†	†	†	†		
	15%	75.0	50.4	4.0	†	†	†	†			
0.15	1%	95.9	95.5	94.7	92.6	83.7	37.9	10.1	†	†	†
	3%	92.3	91.0	88.8	55.4	5.7	†	†	†	†	
	5%	88.9	86.0	79.3	11.1	†	†	†	†		
	10%	81.1	70.2	28.0	†	†	†	†			
	15%	74.3	46.8	1.1	†	†	†				
0.2	1%	94.3	93.4	92.0	81.5	27.1	†	†	†	†	
	3%	90.8	88.6	84.2	25.0	†	†	†	†		
	5%	87.5	83.2	70.7	†	†	†	†			
	10%	80.0	66.0	20.2	†	†	†				
	15%	73.4	42.0	†	†	†					
0.25	1%	92.4	90.3	86.8	39.2	†	†	†			
	3%	89.0	85.1	75.4	3.6	†	†				
	5%	85.8	79.2	55.6	†	†	†				
	10%	78.6	59.9	12.0	†	†					
	15%	72.2	36.2	†	†						

Table 4.4: Smoothing parameter $\lambda^{*(m)}$ (%) for various mean real (net of salary inflation) returns i, standard deviation of return σ and spread period m at which minimum in $\lim \mathrm{Var}F(t)$ and $\lim \mathrm{Var}c(t)$ occurs. (†) indicates that $\lim \mathrm{Var}F(t)$ and $\lim \mathrm{Var}c(t)$ have no minimum and are strictly increasing, with their lowest values at $\lambda = 0$. Blanks indicate that stability conditions do not hold.

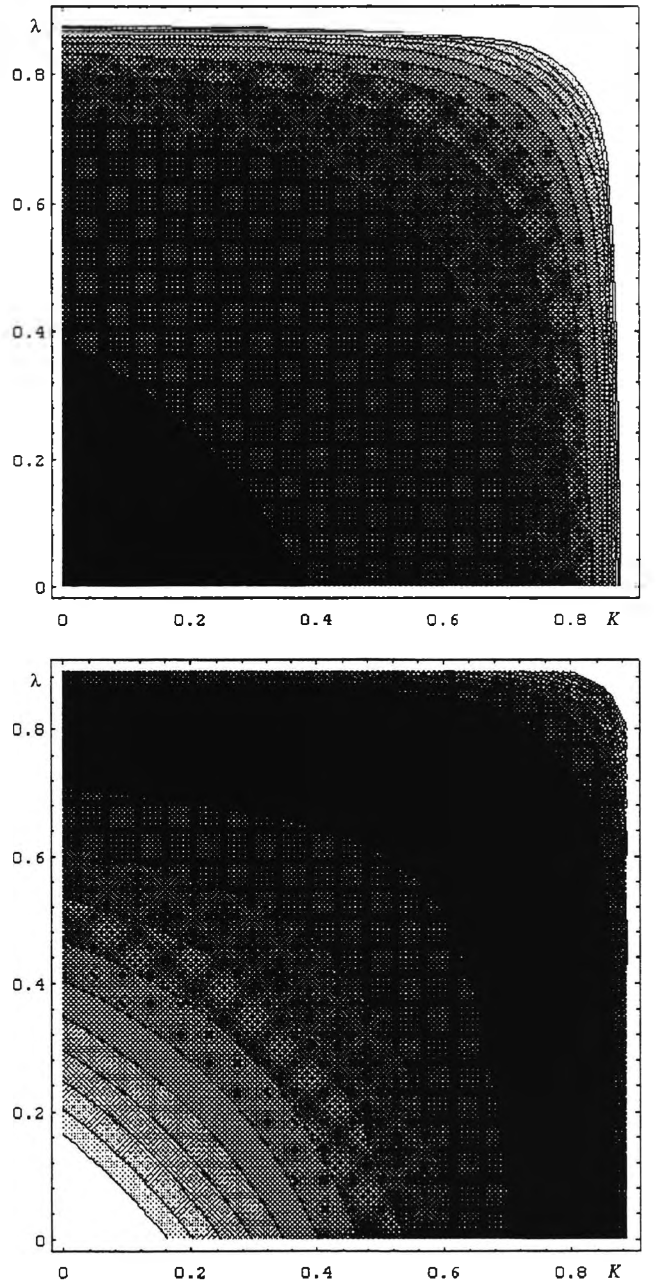


Figure 4.4: Contour plots of $\lim \text{Var} f(t)$ (above) and $\lim \text{Var} c(t)$ (below) against K and λ. $i = 10\%$, $\sigma = 0.05$. Lighter shading represents higher values.

that are applied consistently. In the Market method (§4.1.2), non-marketable liabilities are valued by reference to marketable assets using assumptions that are implied by the market at the time of the valuation. In practice, actuaries often use a pragmatic mix of techniques from both methods.

2. Some form of notional or hypothetical portfolio is required in both methods. This is a portfolio that matches or hedges or immunizes pension liabilities. The use of a long-term strategic portfolio, which may be different from the matching portfolio, in the Discounted Cash Flow method is theoretically imperfect.

3. The same valuation result may in principle be obtained using the Discounted Cash Flow and Market methods, although for various practical reasons, this is unlikely. This conclusion is also drawn by Exley *et al.* (1997), although they do not show that it is self-consistent within the Discounted Cash Flow method for a matching portfolio to be used.

4. Some aspects of the valuation of liabilities, in a traditional, deterministic setting as well as using stochastic methods are reviewed in §4.2.

5. In §4.3, a few issues concerning asset allocation or investment strategy for pension funds are also considered. The issue of hedging or matching is found to be important, even if a deliberate matching policy is not followed. *Valuation* and *asset allocation* are related through the concept of matching or hedging. Asset allocation may be regarded as a means of controlling the pension fund in order to achieve certain funding objectives.

6. Some practical pension fund asset valuation methods are briefly described in §4.4. Their various properties, in terms of consistency with liability valuation, smoothness, objectivity and their impact on the dynamics of pension funding are reviewed.

7. The interrelationship between asset and liability valuation methods and the choice of valuation discount rate is explored. It is also surmised that asset valuation and the techniques of contribution determination and asset gain/loss adjustment should be chosen consistently.

A particular market-related asset valuation method is introduced and studied in §4.5.

1. The method considered is one of various methods used in practice.

2. The evolution of the first and second moments of the pension funding process is derived mathematically, under some simplistic assumptions, notably that pension fund returns perform a random walk.

3. An immediate observation is the congruence between spreading surpluses and deficits when determining contributions and asset smoothing for valuation purposes.

4. Symmetry between the twin processes of spreading and smoothing is striking and complete. They have a complementary function in achieving smoothness in pension funding.

5. Dufresne (1986, 1988) shows that, when pure market values are employed, there is an allowable range of spreading periods for stability (such that $K \in [0,\ 1/\sqrt{q}])$ in the pension funding process. It is found that, when surpluses and deficits are not spread forward and are repaid immediately, there also exists a similar stable range of the smoothing parameter ($\lambda \in [0,\ 1/\sqrt{q}]$) beyond which the second moments of the funding process are unstable. When asset smoothing and surplus/deficit spreading are combined, upper bounds are placed on the possible stable ranges of spread period and smoothing parameter; these stable ranges are also investigated numerically.

6. The 'actuarial' value of assets generated by the smoothing method investigated is shown not to be divergent from the market value of assets. It is also shown to be less variable than the market value of assets. These are regarded as important properties for a suitable asset valuation method.

7. Smoothing and spreading, separately and together, are found to affect adversely the variability of the fund level at market.

8. Smoothing and spreading, separately and together, are also found to improve contribution stability, but only up to a point. The range of spreading periods such that $K \geq 1/q$ and the range of smoothing parameter such that $\lambda \geq 1/q$ are inefficient as shorter spread periods and smaller smoothing parameters may always be chosen to reduce fund variability for equal stability in the contributions.

9. Numerical experiments reveal 'optimal' ranges for the spread period (for a given degree of asset smoothing) and for the smoothing parameter (for a given period over which surpluses and deficits are spread).

10. An important result of this analysis is that the 'optimal' weight to be placed on current market value when smoothing asset values should be upwards of 60% (depending upon the term over which surpluses/deficits are spread). This appears to contrast markedly with the more typical range of 10–33% reported in the literature as being used in practice.

11. A further result is that it is necessary to consider the combined effect of spreading surpluses and deficits as well as smoothing asset values. This emphasises the point that pension fund asset valuation and asset gain/loss adjustment cannot be considered separately.

Chapter 5

Actuarial Prudence in Pension Funding

5.1 Prudent Valuations

Actuarial Prudence. An important feature of actuarial valuations of pension plans and of actuarial control of pension funding is the exercise of prudence in a volatile and possibly unpredictable economic environment. Prudent overestimates of liability cash flows and underestimates of asset cash flows are usually made. Investment returns are the most important source of risk and uncertainty when retirement benefits are funded in advance. A prudent estimate of the rate at which to discount liabilities is therefore essential. In this chapter, some aspects of actuarial prudence in the choice of the valuation discount rate are discussed. Prudence in other assumptions is not considered: Bader (1983) states that the discount rate valuation assumption is "traditionally used as a repository for margins against inadequacies in other assumptions" and Thornton & Wilson (1992a) also recommend best estimates in all assumptions with prudent margins only in the valuation discount rate.

Discount Rate and Valuation Method. The valuation discount rate assumption refers to the rate used to value liabilities. The choice of the discount rate used in a pension fund valuation depends on the *method* of valuation, as discussed in Chapter 4 (§4.1).

When the Discounted Cash Flow valuation method (§4.1.1) is employed, the valuation discount rate is related to the long-term average return on new money being reinvested in a notional portfolio that matches the liabilities (rather than in the actual portfolio of assets held by the fund). Assets and liabilities are both valued using the same discount rate assumption, which ensures consistency. The valuation discount rate in the market method (§4.1.2) is related to the (term-dependent or weighted average) discount rate implied in the assets that hedge or match future liabilities. A market-related value of assets is used, with the liabilities also being valued by reference to the market, so that consistency also follows. Actuaries appear to use a pragmatic mix of both methods.

Best-estimate Investment Return Assumption. The choice of the valuation discount rate depends therefore on the method of valuation employed. Whatever the method of valuation, the discount rate refers to the estimated future return on the portfolio of assets that matches or hedges the liabilities of the pension plan, *rather than* to the estimated return on the actual and future asset portfolio of the fund. In addition to selecting a valuation discount rate for liabilities, an assumption regarding the long-term investment return on current and future pension fund assets may also be made.

Such an investment return assumption is a 'best estimate' of the return on the strategic portfolio of assets that will be held by the fund in the long term. The strategic portfolio may of course be different from the current asset portfolio, as temporary tactical positions may occasionally be taken. Unless the pension fund has a long-term policy of investing in the minimum-risk asset portfolio that hedges or matches its liabilities, there is also a *conceptual* difference between the investment return assumption and the discount rate as they relate to different portfolios. There may or may not be a *numerical* difference in practice given that it is not always possible to determine either the hedging portfolio or the long-term portfolio exactly and given also the imprecision in the estimation of relevant discount rates. A single best estimate of the investment return is usually made whereas multiple liability discount rates may be used.

Terminology. It is customary in most pension fund valuations not to distinguish between the valuation discount rate and the investment return assumption, unlike in product pricing in life assurance. A single assumption is often made and is traditionally termed a 'valuation rate of interest'. This phrase is misleading on several counts:

1. It does not explicitly distinguish between the *valuation discount rate* and the *best-estimate investment return* assumptions.

2. It is possible that *multiple* risk-adjusted discount rates are used in a valuation rather than just one 'valuation rate of interest'.

3. It is possible that *term-dependent* discount rates are used in a pension fund valuation.

4. The discount rate and investment return assumptions are only indirectly related to *economic* interest rates. The valuation discount rate comprises various risk-adjustments, whereas the investment return assumption reflects the risk premiums in the various asset classes in which a pension fund is invested.

For these reasons, I avoid use of the term 'valuation rate of interest'. Ezra (1988) and Smith (1996) also prefer the term 'valuation discount rate' as the rate used to value liabilities for funding purposes. The conceptual difference between the valuation discount rate and the investment return assumption is also made in Actuarial Standard of Practice No. 27 of the American Academy of Actuaries (1996:§3.6).

Prudence Margins. It is a central tenet of this chapter that there exists a conceptual difference between the discount rate and the best-estimate investment return assumption. This difference is in most cases also a numerical one. The valuation discount rate is sometimes referred to as a *prudent* investment return assumption, and the difference between the *prudent* and *best-estimate* assumptions is described as a 'margin' (Thornton & Wilson, 1992a). The term *margin* in the rest of this chapter refers to the difference between the valuation discount rate and the best-estimate investment return assumption. Reasons for the existence of a prudence margin are explored in §5.2 and the effects of such margins on pension funding are considered in the rest of this chapter.

5.2 Prudence Margins in the Valuation Discount Rate

There are various reasons for the difference between the discount rate used to value liabilities and the long-term expected return on assets:

1. Assets are not perfectly matched to liabilities: there exists a mismatch risk margin (§5.2.1).
2. The liability discount rate is adjusted for various risks in the liability cash flows (§5.2.2).
3. Margins seem to be used to control the 'pace' of funding (§5.2.3).
4. Margins are also inserted as a source of flexibility when determining employer contributions (§5.2.4).
5. The choice of discount rate is imprecise (§5.2.5).

5.2.1 Mismatch Risk Margin

Hedging or Matching Portfolio. When the Discounted Cash Flow method is used (with notional switching into a 'long-term' strategic portfolio of assets), the valuation discount rate is the long-term reinvestment yield *less* an adjustment for the mismatch risk taken. When the market method is used, the valuation discount rate is the return on the portfolio of assets that hedges or matches the pension liabilities. The actual and future assets in which the pension fund invests may be different from the matching or hedging portfolio. The long-term estimated return on the actual assets held by the fund may therefore be different from the valuation discount rate, reflecting the mismatch risk taken in the investment of the fund.

McLeish [Thornton & Wilson (1992a): discussion], for example, deals explicitly with the mismatch risk margin:

> "The projection rate is meant to represent the yield which will be obtained on the asset portfolio held by the ongoing fund, whereas the lower settlement rate [the discount rate used to value liabilities] is meant to represent the yield which would be obtained on a portfolio of assets closely matched to accrued liabilities. There would be a difference between these rates, even if both were best estimates of the return from the two different portfolios."

Wise (1984) and Exley *et al.* (1997) make similar arguments, in terms of matching and hedging respectively. Vanderhoof (1972) (see Milgrom (1985:§1)) also relates the liability valuation discount rate to an immunized asset portfolio. Brownlee & Daskais (1991) attempt a similar explanation for the discount rate in terms of the concept of immunization. Any mismatch investment risk taken by the plan sponsor and trustees then results in a (more volatile) surplus, and hence (more volatile) lower contributions.

Investment Return Volatility. It has been shown [Thornton & Wilson (1992b:¶2.2), McGill *et al.* (1996), Winklevoss (1993)] that the variability in asset returns generally tends to dominate any other source of uncertainty in pension funding. The liability discount rate is generally not a risk-free discount rate as it depends upon the minimum-risk asset portfolio that is considered to hedge or match pension liabilities including salary-related ones. (Nevertheless, a 'dedicated' bond portfolio that matches the liabilities has been called a risk-free asset by Leibowitz & Henriksson (1988) among others.) The best-estimate investment return assumption incorporates a *risk premium* for the uncertainty in the return on assets such as equities in the actual pension fund portfolio. The liability discount rate comprises a *risk adjustment* for the uncertain return on assets in the hedging portfolio (which will usually comprise a different proportion of equities or possibly none at all if it is believed that equities are not suitable for hedging pension liabilities). Volatility in equity returns is therefore reflected in a prudence margin between the discount rate and best-estimate investment return assumption. Fixed-interest securities are also not risk-free unless they exactly match liability cash flows until maturity: an adjustment for the risk from the reinvestment of the proceeds from fixed-interest securities is also required.

Hedging Salary-Related Liabilities. There is some debate regarding the best match or hedge for certain pension liabilities and so it is not always clear what the liability valuation discount rate ought to be. Whereas conventional and index bonds hedge nominal and price-related liabilities respectively, there is contention over which assets best hedge salary-related liabilities. It is argued that productivity improvements in the economy are shared between capital (through dividends) and labour (through wages) in roughly stable proportions so that equity dividend growth and wage inflation reflect productivity growth. Equities are therefore held up as a hedge (or at least as a partial hedge) for the final-salary liabilities of active members and so equity returns are at least partially relevant to the valuation of active liabilities. Such an argument ignores the fact that foreign capital may also be invested in the economy. Exley *et al.* (1997) regard the link between equity dividends and salaries as "spurious" for a different reason. They argue that dividends represent a much smaller proportion of national income than wages. Since capital and labour share productivity gains, small variations in wages may be offset by variations in dividends that are, as a proportion of total dividends, much larger. Exley *et al.* (1997:§7) claim that there is therefore considerable instability in the relationship between salary inflation and dividend growth. They seek to establish an alternative link between salary-related liabilities and index-linked gilts (real-return government bonds) by arguing that salary inflation is closely related to price inflation plus an estimated real salary inflation component. (See also §2.6.3.) Various authors including Exley *et al.* (1997) therefore reject any role for equity returns in deciding the valuation discount rate as amounting to taking advance credit for the equity risk premium.

The argument over the role of equities as a hedge for salary-related liabilities is particularly relevant to solvency and accounting valuations, but also to ongoing management valuations. For the purpose of *setting contributions*, some actuaries take the view that the valuation discount rate may allow for a realistic, partial recognition of the mismatch investment risk taken in pension funding: they suggest that the discount rate may then contain a prudent estimate of the mismatch risk premium. In particular, if there is significant equity investment by the pension fund, they argue that it is reasonable to take some advance credit for the excess return on equities versus fixed-income securities, irrespective of whether equities hedge salary-related liabilities. Dyson & Exley (1995:¶6.3.1) call this a "subjective, realistic basis with explicit allowance for both real salary increases and for a rate of expected investment return in excess of that implied by Government securities." This view is taken by Thornton & Wilson (1992a:¶¶6.5.2, 10.11) who suggest that ongoing contributions should be calculated on a realistic basis, "to take credit for best estimate investment returns", whereas funding should be on a prudent basis.

Hedging Nominal and Price-Related Liabilities. When suitable assets exist and are held to meet nominally fixed or price inflation-indexed liabilities, it is much easier to determine a suitable 'matching' valuation discount rate:

1. The "cash flow matching method" of the Actuarial Standard of Practice No. 27 of the American Academy of Actuaries (1996:§3.6) relates the valuation discount rate to the internal rate of return on a hypothetical bond portfolio that generates income that approximately matches future pension liability outgos.

2. When a 'dedicated' bond portfolio (see §4.3.3) is set up to match current pensioner

liabilities, Bader (1983) suggests that the discount rate for the liabilities to which the bonds are dedicated may be set such that the value of these liabilities is equal to the market value of the dedicated bond portfolio. He also suggests as an alternative that the valuation discount rate may be the known future yield on the dedicated bond portfolio.

3. A similar approach for a closed and mature plan whose liabilities are 'immunized' and 'matched' is described by Anderson (1992:162) and Thornton & Wilson (1992a:¶9.5).

4. Taylor (1986:§4.3) makes a related argument in the context of general insurance.

Term-dependent Discount Rates. A further complication in the choice of discount rates is that a strict application of the market valuation method requires the use of term-dependent discount rates (see §4.1.2), since the term structure of interest rates is generally not flat. For many plans (especially large ones, where the expense can be justified), simple sets of term-dependent discount rates are indeed used, especially in North America. Lee (1986:¶11.29) describes this as a "short-term patch". Allison & Winklevoss (1975) contemplate the use of a graded, non-uniform discount rate that varies by calendar-year, with the discount rate interpolated between current and expected long-term returns. Furnish *et al.* (1985) study at length the practical use of "select and ultimate financial assumptions" (analogous to mortality assumptions). Such assumptions are allowed by the Actuarial Standard of Practice No. 27 of the American Academy of Actuaries (1996:¶3.6.4). The main justification for such term-dependent discount rates is, of course, that they conform with a non-flat yield curve and give consistency to market valuations of assets and liabilities. Greater uncertainty in the more distant future may require greater risk-adjustments or larger prudence margins in term-dependent discount rates (this is also mentioned by Cavaye & Springbett (1964:¶4)). Jackson (1984) also suggests that a change in 'ultimate' assumptions may have a less abrupt effect in terms of recommended contribution if term-dependent discount rates are used. The use of discount rates based on the prices of STRIPS (Separate Trading of Registered Interest and Principal securities—see Sharpe *et al.* (1995:119)) of various terms has also been suggested in the context of *accounting* valuations (IFAA, 1997).

In most pension fund valuations, some form of duration-weighted average tends to be used as an approximation to term-dependent discount rates. Dyson & Exley (1995) describe a valuation method that uses spot rates from the yield curve of conventional and index-linked gilts and suggest that an average rate, weighted by the duration of the liabilities of the pension plan, can be used as a first-order approximation to term-dependent discount rates. Thornton [Exley *et al.* (1997): discussion] anticipates that future international accounting standards may also require a single average market discount rate rather than a set of term-dependent discount rates.

Multiple Discount Rates. It is plausible in theory to assume different discount rates for different sets of liabilities. Different discount rates may be used when there exists a set of assets and liabilities that are expressly matched. A practical example occurs in the North American practice of constructing 'dedicated' bond portfolios to match current pensioner liabilities (see §4.3.3). Bader (1983) suggests that the subset of pension assets and liabilities that are matched or 'dedicated' (constituting a "dedicated sub-plan") may be valued differently from the non-dedicated assets and liabilities. Two different discount rates are then used for the two sets of liabilities. These discount rates may both be different from the assumption regarding long-term investment return on the pension fund.

5.2.2 Other Risk Adjustments

The valuation discount rate generally contains adjustments for various risks pertaining to the pension liabilities, whereas the best-estimate investment return assumption does not. This also partly explains the 'prudence margin' between them.

Uncertainty in Cash Flow Amounts. The valuation discount rate is usually adjusted for the uncertainty or risk in the amount of benefit payments (for example demographic risks) and possibly expenses. This may be of particular importance for smaller pension plans. The pension plan sponsor may also improve benefits on a discretionary basis and Wise (1998; Exley *et al.*, 1997: discussion) suggests that discount rates should reflect the uncertainty involved in benefit enhancement. Note also that different liabilities are hedged or matched by different assets. Liability cash flows may therefore be subject to different levels of uncertainty or risk and different risk adjustments may therefore apply for different sets of liabilities [Smith (1996), Exley *et al.* (1997)]. In theory this means using multiple risk-adjusted discounted rates for the various sets of liabilities, although in practice they appear to be combined in some form of approximate weighted average.

Uncertainty in Cash Flow Timing. There is uncertainty not only in respect of the *amount* of benefit and expense cash flows, but also regarding their *timing.* The election of various options may also cause liquidity problems. Actuarial Standard No. 27 of the American Academy of Actuaries (1996) also mentions the risk pertaining to "supplementary benefits triggered by corporate restructuring" (generous benefits being sometimes awarded for voluntary redundancy). Wise (1998; Exley *et al.*, 1997: discussion) also considers the option of the sponsor to wind up the plan at any time (if investment returns are insufficiently high for instance). Such uncertainty in the pension plan arrangement may also warrant further adjustments to the discount rate.

Insolvency and Credit Risk. Solvency considerations may also justify using more prudent discount rates in ongoing funding valuations. Thornton & Wilson (1992b:¶5.3) suggest that there should be "a margin of at least 20% between the discontinuance position and the ongoing funding position" because of the wide fluctuations in investment return experience and potential emergence of deficits coinciding with wind-up of the plan. The valuation discount rate should therefore be a 'prudent' one. McLeish [Thornton & Wilson (1992a): discussion] incorporates "a margin in the settlement rate [the discount rate used to value liabilities] to reduce to an acceptably low level the probability of inadequate funding, which otherwise might be 50% in the event of winding-up." It is not clear how such a margin should be determined. In the classical financial theory of pension plans [Bagehot (1972), Treynor (1977)], a 'risk-free' discount rate is meant to be used in the valuation of pension liabilities as the plan is assumed to be ongoing with benefits 'guaranteed' by the plan sponsor. Pension plans do become insolvent, however, and the discount rate used to value liabilities should reflect this risk. When pension liabilities are valued for accounting purposes, the U.S. Financial Accounting Standards No. 87 (FAS87) and Exposure Draft 54 of the IASC (1996) require the use of discount rates implicit in annuity contracts or corporate bonds to reflect the situation if the plan were wound-up. Exley *et al.* (1997:§8) consider that this does not adequately describe the "credit risk of a pension promise", i.e. the risk of pensions not being paid out if the pension plan is terminated. They argue that this risk depends on the simultaneous events of plan assets defaulting *and* the plan sponsor being bankrupt and unable to make good on the pension promise. These two events are dependent, according to Exley *et al.* (1997:§8), since sponsor bankruptcy may be triggered by having to make large contributions because of minimum funding requirements being breached. Exley *et al.* (1997:§8.5) therefore comment that the risk adjustment in the discount rate should be related to the plan sponsor's creditworthiness; to the funding level, investment policy and maturity of the plan; to the "strength of covenant" between sponsor and members; as well as to the "matching policy". This is sometimes used to vindicate the inclusion of part of the equity risk premium in the discount rate (see §5.2.1), although Exley *et al.* (1997:¶8.5.1) reject this argument.

5.2.3 Margins to Control Funding

'Prudence margins' in the valuation discount rate are also used by actuaries to influence the value placed on liabilities and hence the 'pace' of funding for benefits. Daykin (1976:¶6) suggests that the valuation discount rate can be used

> "as a conscious but undeclared means of speeding up or slowing down the explicit pace of funding."

Some actuaries are uncomfortable with the fact that this control is 'undeclared'. Brownlee & Daskais (1991) accept that the use of margins in the valuation discount rate (which they refer to as "tinkering with the discounting assumption") may be acceptable to speed up funding if there is a risk that the sponsor may become insolvent or that the investment policy being followed is too aggressive, and also provided the prudence margins are *disclosed*. Even with disclosure, it is not always considered to be acceptable practice [Snelson (1970:¶33)].

Valuation Assumptions and Cost of Pension Provision. The justification for such margins in the discount rate seems to be that prudence margins ought not to affect *directly* the ultimate cost of pension funding and therefore do not *directly* matter. It is usually accepted that the *ex post* cost of providing pension benefits depends

- *primarily* upon the benefit policy (i.e. the level of benefits promised) and upon the economic and demographic experience of the plan,
- and only *indirectly* upon the funding and contribution policies (i.e. the funding method and assumptions), as the incidence of contributions influences investment policy or as surpluses entail an improvement in benefits.

Management or ongoing valuations are perceived to be exercises in managing the pension fund and delivering on the various objectives of benefit security, contribution stability etc. Actuarial intervention in the funding of retirement benefits, when the pension fund is being valued, is aimed at managing or controlling the pension plan by setting a funding strategy. Since valuation assumptions are thought not to affect *directly* the ultimate cost of providing pensions, they are considered to be legitimate tools in the control of pension funding. It is argued that it does not *directly* matter that assumptions may be wrong or that prudence margins may be included, provided they are adjusted as experience unfolds.

The valuation discount rate assumption is for this reason rarely referred to as a forecast, particularly since it is commonly held that neither the market nor the actuary can successfully and accurately predict the future economic environment in which a pension plan operates. Some authors thus describe the valuation discount rate assumption as an "educated guess" [Trowbridge (1966), Ezra (1980), Anderson (1992:165)] or an artificial construct [Paquin (Vanderhoof, 1973: discussion)] or a technical tool [Trowbridge (1966), Ezra (1988)] rather than a forecast. Indeed, various authors do use the valuation discount rate as a formal technical tool or control parameter. Cairns & Parker (1997) attempt to maximise contribution stability in a mean-variance context in a simple pension fund model by determining an 'efficient frontier' or range for the valuation discount rate such that the variance of contributions is minimised for a given mean level of contribution. Benjamin (1984, 1989), Loades (1992) and Fujiki (1994) likewise seek to determine a suitable valuation discount rate as some form of average of previous rates of return with or without a prudence margin to stabilise various aspects of pension funding.

For the purposes of funding and setting contributions, liability and asset 'values' are not necessarily economic values (for example in the Discounted Cash Flow method), but should be somehow comparable. An ongoing or management valuation is not therefore strictly an economic valuation (Ezra, 1988) but is concerned with financing the pension benefits [Trowbridge (1952), Paquin (1975)] or regulating the pace of funding [Daykin (1976:¶2)] or budgeting for pension liabilities [Exley *et al.* (1997:¶3.2.5)]. It is therefore argued that, since the relative conservatism of assumptions merely changes the incidence of contributions and advances or delays the financing of pensions, the choice as to the assumptions should be made by balancing the interests of sponsor and members. This view is taken by Trowbridge (1966) for instance:

> "On the one hand, the higher [contribution] is likely to be better from the viewpoint of employee security, and may be fine for the employer as well if he can conveniently concentrate contributions into the early years. On the other hand, the lower initial outlay calculated by less conservative actuarial assumptions or cost methods may well be indicated by any of several business considerations, and the lower employee security well justified by the increased likelihood that over the long haul the plan can be continued."

Smoothing the Emergence of Surplus. There are at least two reasons why prudence margins in the valuation discount rate are used to control funding. The first is that margins may be used to smooth the emergence of surplus. This seems to originate from actuarial practice in conventional life assurance (Fisher & Young, 1971). It prevails in circumstances where continual benefit improvements are the norm. When increases to pensions in payment are made in an *ad hoc* way as surpluses appear, the rate of emergence of surpluses in the fund must be regulated. Daykin (1976:¶6), for example, regards the valuation discount rate assumption as

> "a principal means of control in the management of a fund, for example in funds where benefits are increased as surplus emerges, in which case the rate of increase of benefit may be controlled by a judicious choice of rates of interest in successive valuations."

Smoothing Sponsor's Contributions. Another reason why prudence margins may be used in the valuation discount rate is to stabilise and smooth the sponsor's contributions (§2.2.4). Actuarial assumptions, and in particular the valuation discount rate, are therefore selected to even out the effect of one-off events. Snelson (1970) gives an account of such practice:

> "Sometimes an employer wishes to provide a certain scale of benefits but wishes the initial outlay to be somewhat lower than that originally suggested [. . .] The proper way to deal with this problem is to lengthen the period [of amortization] to achieve the desired degree of funding [. . .] In practice a reduction in outlay is sometimes attempted by making lighter assumptions."

Chambers [Colbran (1982): discussion] believes that due consideration to the sponsor's situation must be paid if valuation assumptions need to be changed in a way that may not have been anticipated by the sponsor:

> "It would often be wrong and damaging to a client to expose him to the traumatic consequences of a material change in the valuation basis if the option is available of a progressive strengthening or weakening over a period. There will be occasions when this option is not available or is inappropriate. Whether the option is available and the extent to which it is available are matters for the actuary to consider. He should discuss these things with his client, but the decision rests with the actuary."

Finally, note that although 'prudence' may be employed in the valuation discount rate to control various aspects of funding, margins do not change the economic value of liabilities. Ezra (1988) states that they simply give rise to a

> "concealed contingency reserve [which] can be used for at least two distinct purposes — to prefund future benefit improvements or to create a reserve against fluctuations in funding contributions."

Smith (1996:§2.3) uses the terminology of Bride & Lomax (1994) and describes margins in the discount rate as creating a "cost of capital adjustment, or COCA" which is used as

> "one way of adjusting the pace of funding, or equivalently, of moving the future capital cost onto and off the balance sheet. The COCA is, in effect, acting as a slush fund to smooth out market fluctuations."

5.2.4 Margins as a Source of Flexibility for the Sponsor

Flexibility in Pension Funding. One of the essential features of any form of advance funding is the flexibility that it provides to the sponsor in terms of the timing of contributions. Indeed, flexibility is often cited as a motivation for advance funding. The fact that money is set aside is taken to imply that sponsors have more freedom in varying their contributions to the pension fund than in a pay-go situation. Flexibility is afforded to sponsors in various ways (§§2.2.5, 3.4.2), notably in the choice of period over which unfunded liabilities or gains/losses are amortized, in the choice of funding method, and in discretionary contributions. Valuation assumptions, and particularly the valuation discount rate, also appear to be varied so as to provide some flexibility to sponsors, although this may be rationalised by alternate means.

Actuaries do take into account the plan sponsor's interests in their recommendations about funding. The reasons why flexibility could be valuable to corporate sponsors are summarised by McGill (1964:319) who states that

> "Sufficient flexibility to permit an employer to take optimum advantage of tax deductions and to adjust his contributions to his earnings experience is regarded as a desirable objective so long as it does not impair the 'will to fund' and hence, the solvency of the plan."

McGill (1964:317) thus explains that there has to be "latitude" as to the choice of assumptions in pension funding since there has to be a "necessary degree of flexibility in projecting the costs and accumulating the funds." Bassett (1972) similarly believes that it is justifiable to vary prudence margins in the various valuation assumptions, including the valuation discount rate:

> "An actuary [...] usually recommends, with considerable justification, that his client take a conservative approach to pension funding, albeit at concomitantly high cost. However, if an employer has enough financial flexibility to increase his contributions, should unfavorable circumstances require it, the actuary is justified in using less conservative, more aggressive assumptions that lead to lower pension costs."

Corporate Finance Theory. The theory of corporate finance in fact suggests that plan sponsors may seek to use pension funds to their financial advantage [Bodie *et al.* (1985), Exley *et al.* (1997)]. Corporations sponsoring defined benefit plans may seek flexibility in their contributions for at least two clear reasons:

1. *To suit their cash flow needs:* Companies may wish to steer assets into the pension fund when times are good, whereas financially distressed sponsors with less spare cash may seek to do the opposite. Bodie *et al.* (1985) refer to plan sponsors maintaining "some financial slack in order to avoid having to rely on external financing at 'unfavorable' times." Exley *et al.* (1997:¶8.15.2) similarly suggest that sponsors may benefit from controlling the timing of contributions in order to reduce "the costs of raising and distributing capital." In particular, Exley *et al.* (1997:¶8.8.7) believe that

   > "Many actuaries, in practice, are not blind to the synergies available to companies who wish to take into account the cash needs of the business when planning their pension fund contributions."

2. *To benefit from tax advantages:* Black (1980), Tepper (1981), Black & Dewhurst (1981) and more recently Exley *et al.* (1997) have described how sponsors and their shareholders would benefit from the pension fund investing in fixed-income securities which are less heavily taxed compared to equities. If sponsors do enjoy a degree of flexibility in their contributions to pension funds and have cash at hand, they may wish to make earlier and larger contributions in order to benefit from the tax privileges of pension funds.

Some Anecdotal Evidence. There is some anecdotal evidence that the sponsor's financial status does influence the choice of actuarial assumptions. Thornton & Wilson (1992a:§10) describe how the financial circumstances of the sponsor may lead to a preference in the level of prudence or in the margins in valuation bases:

> "Some clients would wish to fund on best estimate bases; this is typically in cases where finance is difficult, where pension costs have to be minimised in the company accounts to achieve respectable company results, or where cash is short. [...] In other cases, the sponsoring company itself may be concerned to maintain a strong fund. This might be because the future prospects for the company or its industry may be uncertain, and the risk of having to support an underfunded scheme in later more difficult times is to be avoided."

VanDerhei & Joanette (1988) also describe the interaction between actuary and plan sponsor in the U.S. in anecdotal terms:

> "At times management conveys a cost objective to the actuary, such as the need to reduce pension contributions, and the actuary responds by recommending a change in the actuarial assumptions or methods. But when left to their own devices, most actuaries [...] are conservative by nature. Thus, the economic determinants flowing from management incentives are an interjection, not usual but not infrequent, in the determination by the actuary of the actuarial cost method and assumptions which he or she deems most reasonable for pension cost attribution."

Ezra (1988) gives a subjective account of practice in the early 1980's in North America—when companies were cutting costs as a result of recession—which also shows that actuaries are sensitive to corporate financial requirements when choosing assumptions:

> "[Corporations] discovered that contributions hitherto referred to in actuarial reports as 'required' were in fact quite flexible, and in many cases were not required at all, because the actuarial assumptions on which they were based were so cautious. As corporate financial executives took over control from the actuaries to whom they had implicitly delegated those decisions in the past, they realized that an actuarial 'recommendation' as to contributions was almost meaningless, because it makes little sense to produce a recommendation without considering alternative uses of that money. Intent on survival and improving the bottom line, corporations overruled traditionally recommended contribution rates, which actuaries then gamely changed, revealing unsuspected depths of response to reality."

Sometimes, actuaries may well vary prudence margins in valuation assumptions to suit the interests of plan sponsors but rationalise this by other means. Exley *et al.* (1997:¶8.8.7) point out that actuaries can be "accommodating" in their choice of the assumed equity dividend growth for asset valuation purposes in the U.K.:

> "We do not see why an actuary who, quite reasonably, has taken account of the short term cash flow of the sponsor when reaching a recommendation, should feel obliged to rationalise the process in terms of a change in his view of long term dividend growth."

Parsons [Thornton & Wilson (1992a): discussion] in fact describes the dividend growth valuation assumption as a "device":

> "It is important to remember that, unless one is using 'realistic' assumptions for the whole valuation, a dividend growth assumption used for valuing assets is merely a device for obtaining the appropriate result; it will not be a realistic assumption in its own right. If one does not maintain an appropriate gap [assumed discount rate less dividend growth], one starts dealing in 'funny numbers' and this can only harm the credibility of the actuarial profession."

Some Statistical Evidence. There is little statistical evidence regarding the importance of sponsors' influence in the prudence margins in valuation assumption. In a survey of 63 actuaries who recently valued 106 Australian defined benefit superannuation funds, Ferris & Welch (1996:18) investigate qualitatively the extent of influence by the fund sponsors. Most of the actuaries reported no influence by the sponsors on their assumptions. Approximately 40% of the actuaries stated that they used conservative assumptions, by their own estimates. In four cases the sponsors and actuaries debated the margins incorporated in the basis, with the sponsors desiring *more liberal* assumptions. Bodie *et al.* (1985), Friedman (1983) and Joanette (1985) also claim to find statistical evidence for the positive correlation of plan sponsors' profitability with pension funding levels. (They claim to use some form of normalised funding level to compare different pension plans accurately.) This would appear to confirm the hypothesis that pension funding is not entirely independent of plan sponsors' finances, which may in turn mean that actuarial assumptions are guided or at least influenced by the finances of the corporate plan sponsor.

Some Evidence from Accounting Valuations. There is also anecdotal evidence that the valuation discount rates used for *accounting* purposes in the U.S. prior to Financial Accounting Standards No. 87 (FAS87) were influenced by the sponsor's financial situation. Indeed, the motivation of FAS87 was to standardise financial reporting regarding pension funds and improve comparability between the accounts of companies. Firms in financial difficulty appeared to try to minimise their reported pension liability. VanDerhei & Joanette (1988) suggest that the discount rate used for calculating pension expense and contributions (which were often equal prior to FAS87) was at times not the same as the discount rate used to report pension liability under Financial Accounting Standards No. 36 (FAS36). McGinn [Holcombe *et al.* (1973)] also states that "corporations and boards of directors have pressured actuaries to use realistic assumptions in valuing pension liabilities in order to charge pension costs equitably among generations of common stockholders." Regan (1980) presents some evidence that higher discount rate (net of salary inflation) assumptions were used in the valuation of pension funds whose sponsors were awarded a higher credit rating by Moody's. Bodie *et al.* (1985) find evidence that the valuation discount rates used (and disclosed) in the calculation of FAS36 pension liability was negatively correlated with the profitability of plan sponsors.

Professional Practice. Finally, note that it is customary actuarial practice to discuss the evolution of the pension fund with the sponsor so as to enable him to budget for future contributions or to decide to improve benefits. Many actuaries consider positively the sponsor's financial interests and believe that sponsors should be *explicitly* involved in decisions regarding the pace of funding and the choice of funding method. This view is held by some British actuaries [Colbran (1982:¶10.1), Humphrey *et al.* (1970:¶8.14)] although it finds more favour among North American actuaries (Holcombe *et al.*, 1973). Professional standards do not prohibit input from sponsors and the actuary is not forbidden from considering their financial situation. 'Professional judgement' is apparently called for. Professional Standard 400 of the Institute of Actuaries of Australia (1995:¶¶11, 12) stipulates that actuarial assumptions chosen for an ongoing funding valuation

> "must give proper weight to every aspect of significance, including the operating environment of the employer, and the manner in which discretions are likely to be exercised".

Whereas an actuary in Australia is responsible for the choice of assumptions in consultation with other parties, there may be circumstances where "the actuary may be directed to use particular methods and/or assumptions" and if he disagrees with such assumptions he should discourage their use or decline to perform the valuation.

'Flexible' Margins in Discount Rate. It is apparent therefore that the choice of valuation assumptions, such as the discount rate for valuing liabilities, is not always purely based on the experience of the pension plan. There is circumstantial and anecdotal evidence that the valuation discount rate is not always chosen independently of the sponsor's financial needs. The level of prudence or margin built into the valuation discount rate in comparison with the best-estimate investment return assumption may be influenced by the sponsor's financial circumstances.

5.2.5 Imprecision in the Choice of Discount Rate

The choice of the valuation discount rate is not always made with exactitude. There are various reasons why the discount rate is only an approximation:

Risk-adjustments. It is not always clear *how* various risk-adjustments to the liability discount rate are to be determined. The *size* of such adjustments seems to be chosen in an *ad hoc* manner in practice.

Hedging portfolio. It is also difficult, in most cases, to determine exactly a minimum risk or hedging or matching portfolio for the pension liabilities. This difficulty is particularly apparent in the approximate nature of the notional or hypothetical portfolio used in the valuation of assets in the Discounted Cash Flow method in the U.K.

Normative Aspects. The valuation discount rate employed by actuaries in the valuation of particular pension plans is often observed to change infrequently since it is a 'long-term' assumption. Actuaries are influenced by published surveys of assumptions used by other actuaries. Brownlee & Daskais (1991) describe how actuaries sometimes "consult a survey of what other actuaries are doing and try to avoid being outside the bounds, frequently very wide bounds, of current practice."

Best-estimate Ranges. Actuarial assumptions will not be borne out exactly by experience, except by chance, and cannot be known in advance. It is therefore said that valuation assumptions, especially uncertain economic ones, may be chosen anywhere within a best-estimate *range*. This by itself implies that there is an acceptable margin in any assumption. Given an acceptable range for each assumption, it is sometimes extrapolated that there must be an acceptable range of valuation results. A probably typical illustration of how valuations are carried out in the U.K. is described by Heywood [Colbran (1982): discussion]:

> "[. . .] I believe that there is a range of valuation results. The top limit is the most stringent basis allowable, and the lower limit is that below which it would be dangerous to go. Any result in-between is very much a matter of opinion. My procedure is to write a preliminary summary letter of the results to the client, then talk to him, because I do not believe that the actuary is better qualified than other people to forecast future rates of interest or future rates of inflation. These matters are discussed and eventually a valuation result is reached. The result should be considered against the financial background of the company. If it is doing extremely well, the result might be pushed towards the more stringent end of the range. If it is not doing so well, it might be pushed the other way. However, the ultimate responsibility must rest with the actuary."

Averaging. Multiple discount rates for different liabilities and term-dependent discount rates (§5.2.1) are usually replaced by computationally convenient but approximate single-valued averages in practice.

Stability and Consistency. The importance of contribution stability to the plan sponsor means that asset values are often smoothed (§4.4). Since assets and liabilities must be valued in a consistent way, the assumption that is made concerning the valuation discount rate is also more stable than market discount rates, say. The valuation discount rate is thus said to be a 'long-term' one to correspond to the long-term stable asset value (Ezra, 1980). Anderson (1992:162) states that the valuation discount rate "must reflect not only the type of investments made by the fund, but also how those investments are valued." Trowbridge & Farr (1976:93) stress "the consistency between the approach to asset valuation and the investment earnings assumption." Accordingly, Loades (1992) uses a valuation discount rate that is smoothed or averaged over time. It is not always clear in practice how consistency and stability in the choice of discount rate are achieved.

5.3 Persisting Surpluses/Deficits

A common problem in pension systems is the existence of persisting surpluses or deficits. They occur because of recurring actuarial gains or losses as a consequence of actuarial assumptions differing consistently from experience. In the short term, assumptions are not borne out by experience, except by chance. An 'analysis of surplus' pinpoints systematic deviations from experience over the long term and assumptions may then be rectified. Even if all actuarial valuation assumptions are 'realistic' and 'best estimates', prudence margins are still incorporated in the valuation discount rate, for the various reasons enunciated in §5.2. Such margins inevitably give rise to surpluses. Deficits will arise if, conversely, valuation assumptions are too liberal or optimistic. Large and permanent deficits potentially endanger the security of funded retirement benefits. They may even lead to technical insolvency. Large and permanent surpluses are equally undesirable because over-funding represents an opportunity cost to the plan sponsor as funds are diverted away from productive activity in the firm. This reduces corporate profits, thereby affecting plan members' employment. Excessive surpluses may also incur revenue charges. (See §2.2.3.)

The sensitivity of valuation results to valuation discount rates is well-known. The long duration of pension liabilities means that their (actuarial) present or discounted value is highly sensitive to the choice of discount rate. Various authors [Bizley (1950), Adams (1967)] illustrate this sensitivity mathematically. Since the choice of valuation discount rate influences the value placed on liabilities, it will also influence the contribution recommended, and hence the size of surpluses or deficits that eventually emerge. This will of course depend on the pension funding or actuarial cost method employed. (A common rule of thumb is that a

1/4% change in the discount rate (net of salary inflation) entails a 5% change in long-range recommended contribution when the projected unit credit method is used. Such rules of thumb appear to hold for a range of 'reasonable' discount rates because pension payments occur so far in the future that the duration of pension liabilities is itself fairly insensitive to changes in discount rate, as shown by Keintz & Stickney (1980).)

The size of surpluses and deficits that emerge depends not only on the prudence margin between the valuation discount rate and the long-term mean rate of return on the fund (as well as margins in other assumptions) but also on

1. the method of removing surpluses/deficits or gains/losses, and
2. the period over which they are removed.

This is most clearly illustrated by Dufresne (1986, 1988) in terms of the first and second moments of the pension fund and contribution levels. Assume in the following that all valuation assumptions, *except* for the valuation discount rate, are 'best estimates' and are exactly borne out by experience. The model described in §2.7 is assumed except for Valuation Assumption 2.2: the valuation discount rate is not necessarily equal to the long-term mean rate of return on pension plan assets. All prudence margins are therefore concentrated in the valuation discount rate assumption, as described by Bader (1983) and recommended by Thornton & Wilson (1992a) (see §5.1).

When surpluses/deficits are spread over a rolling term, the second moments of the pension funding process are obtained by Dufresne (1986) (Result 3.3) and depend on $i - i_v$. Let Δ represent the 'margin':

$$\Delta \; = \; d - d_v \; = \; v_v - v. \tag{5.1}$$

Dufresne (1986) then proves the following:

RESULT 5.1 *Provided*

$$i \; > \; -100\%, \tag{5.2}$$

$$d \; < k \; \leq 1, \tag{5.3}$$

then

$$\lim_{t\to\infty} \mathrm{E}ul(t)/AL = \Delta/(d-k), \tag{5.4}$$

$$\lim_{t\to\infty} \mathrm{E}adj(t)/AL = \Delta k/(d-k), \tag{5.5}$$

and provided further that

$$q(1-k)^2 \; < \; 1, \tag{5.6}$$

then

$$\lim_{t\to\infty} \mathrm{E}ul(t)^2/AL^2 = \Delta^2/(d-k)^2 \ + \ \sigma^2 v^2 (k-d+\Delta)^2/[1-q(1-k)^2](k-d)^2 \qquad (5.7)$$

$$= \Delta^2/(d-k)^2 \ + \ \sigma^2 v^2 [1+\Delta/(k-d)]^2/[1-q(1-k)^2], \qquad (5.8)$$

$$\lim_{t\to\infty} \mathrm{E}adj(t)^2/AL^2 = \Delta^2 k^2/(d-k)^2 \ + \ \sigma^2 v^2 k^2 [1+\Delta/(k-d)]^2/[1-q(1-k)^2]. \qquad (5.9)$$

Remarks:

1. The unfunded liability in equations (5.4) and (5.8) should be as small as possible, i.e. full funding is aimed for and the fund level should converge in time to the value placed on the liabilities (the actuarial liability) calculated at the *risk-adjusted, prudent* discount rate i_v.

2. From equation (5.4), as $|\Delta|$ increases, $|\lim \mathrm{E}ul(t)/AL|$ increases. From equation (5.8), as $\Delta \to \pm\infty$, $\lim \mathrm{E}ul(t)^2/AL^2 \to \infty$. (The right hand side of equation (5.8) is quadratic in Δ with the coefficient of Δ^2 being positive, given the stability conditions (5.2), (5.3) and (5.6).)

3. $k = 1/\ddot{a}_{\overline{m|}}$ is a reciprocal annuity factor that is usually calculated at the valuation discount rate i_v. Suppose, however, that it is only a proportional spreading factor which is independent of i_v. From 1 and 2 above, it is clear that large absolute differences between the discount rate and the mean long-term rate of return, i.e. margins of large magnitude cause large surpluses/deficits to emerge and potentially threaten the security of pension benefits.

4. The unfunded liability that ultimately emerges also depends on the spreading period through the spreading factor k in equations (5.4) and (5.8). If surpluses and deficits are spread over a shorter term m (meaning a larger $k-d$ in the denominators of the right hand sides of equations (5.4) and (5.8)), smaller surpluses/deficits will emerge eventually.

5. The expected ultimate surplus such as in equation (5.4) is called a 'bias' by Wise (1984:¶2.16) albeit in a different context (namely, asset-liability matching for a pension fund assumed to be closed to new entrants).

The fact that long-term surpluses and deficits depend on the *period* over which gains and losses are removed as well as on the *margin* between the valuation discount rate and the long-term mean rate of return is noted by a few other authors previously:

1. Trowbridge (1952) shows that the same ultimate fund levels result from the use of the *Aggregate* and *Entry Age Normal* funding methods; he makes the assumption that the valuation discount rate equals the constant (deterministic) rate of return on the

fund. Weaver [Trowbridge (1952): discussion] shows that there is a difference in the ultimate fund levels of the two methods when there is a (prudent or liberal) margin in the discount rate assumption. Trowbridge [Trowbridge (1952): discussion] explains that this difference arises because the two methods spread surpluses/deficits over different periods and Weaver [Trowbridge (1952): discussion] considered Entry Age Normal method with an immediate adjustment for gains/losses only.

2. Adams (1967) investigates numerically the effect of varying discount rates (in a zero-salary inflation setting) on a pension plan funded according to various methods. Pension liabilities have a long duration and so the value placed on them will be sensitive to the valuation discount rate. Adams (1967) shows that methods that build up larger funds result in contributions that are more sensitive to the choice of valuation discount rate, which is assumed to be equal to the actual investment returns that will be made by the fund. Grubbs [Adams (1967): discussion] points out that results will be very different if the valuation discount rate and the actual investment return are not equal. Berin [Adams (1967): discussion] indicates that the amortization of gains/losses as they emerge will also have an effect.

3. Bowers *et al.* (1982) assume that the Aggregate method of funding is used (which is equivalent to spreading surpluses and deficits over a moving term—see §3.2.1) and consider among other things the effect of a difference between the assumed and actual (deterministic and exponential) rates of investment return, salary inflation and population growth. They show that

   > "it is the relationship between the difference of spread between the interest and growth rates in the valuation assumptions and the experience that determines the asymptotic relationship between the size of the fund and the size of the supplementary present value."

   From their results it is obvious that the annuity factor used to spread surpluses or deficits also affects this "asymptotic relationship".

4. Thornton & Wilson (1992a) point out that the cumulative effect of small margins in various valuation assumptions may lead to considerable conservatism in the overall valuation result. They argue that including a 'prudent' (rather than 'cautious') margin in the valuation discount rate assumption, while adopting 'best estimates' for all other assumptions, avoids excessive conservatism. (Various authors including Bassett (1972) and Brownlee & Daskais (1991) also recommend less conservatism in the valuation discount rate assumption.) Thornton & Wilson (1992a:§6.2) show, through a simple static analysis, how prudent/optimistic margins in the valuation discount rate lead to surpluses/deficits (respectively) in the long term. Loades (1992) reaches a similar con-

clusion from more realistic deterministic simulations. Thornton & Wilson (1992a:§6.2) also show how long-term surpluses and deficits depend on the value of the annuity factor used to spread surpluses/deficits and hence urge the use of shorter spreading or amortization periods.

Persisting long-term surpluses emerge as a consequence of the prudence margin in the valuation discount rate (and possibly in other assumptions). If such margins are reduced, then smaller surpluses will arise. For reasons of prudence (and other reasons mentioned in §5.2), a margin in the valuation discount rate cannot always be avoided. The effect of such margins is seen to depend on the method of gain/loss adjustment and on the period over which gains and losses are defrayed. Two methods are used in practice to deal with persisting surpluses or deficits. The first method is to spread surpluses (or deficits) more quickly than deficits (or surpluses). The second is a 'dual-interest' method. We briefly consider these two methods (in §§5.4 and 5.5) and also introduce and analyse a third possible candidate (in §5.6).

5.4 Asymmetric Spreading of Deficits and Surpluses

We have assumed thus far that surpluses and deficits (or gains and losses) are treated similarly. This is not necessarily so in practice. Deficits may be less tolerable than surpluses because of the greater downside risk posed by deficits. Deficits may therefore be spread over a short period or may even be defrayed immediately to maintain the security of benefits, whereas surpluses may be removed over a longer period in an effort to stabilise contributions. The existence of excessive surpluses may alternatively prompt the reverse treatment. If pension fund surpluses persistently emerge over time, then it is plausible to liquidate surpluses faster than deficits.

Pension funds in the U.K. have experienced large surpluses in the past and it has been commonplace to remove surpluses over shorter periods (at times through contribution holidays—Lee (1986:193)) as compared to deficits. Khorasanee (1993) investigates this by using a model plan and simulating economic conditions according to U.K. investment data over the post-second world war period. He concludes that spreading surpluses over a shorter period compared to deficits does reduce the size of surpluses emerging over time—but not by much—while the risk of insolvency increases.

Haberman & Smith (1997:§12) carry out stochastic simulations on a pension fund invested in 70% equities and 30% index-linked gilts, using the Wilkie (1995) model (which is also primarily designed for and fitted to U.K. investment data). Haberman & Smith (1997:66) conclude that when the 'surplus' spread period is shortened (but the period over which deficits are spread is unchanged) the pension fund becomes more efficient at removing surpluses, while deficits are more likely to occur.

Both these studies indicate that a potential solution to the problem of persisting surpluses may be to spread deficits over shorter terms than surpluses, although this may be at the expense of larger and more frequent deficits.

It is difficult to investigate mathematically the asymmetric treatment of deficits and surpluses. Instead, a simple simulation study has been carried out, with the same assumptions as in §2.7 and the following features:

Number of scenarios: 2000.

Time horizon: 150 years.

Funding method: Any of the 'individual' methods with surpluses/deficits being spread over a moving term.

Initial conditions: No initial unfunded liability.

Logarithmic rate of investment return: Independent and identically normally distributed.

(The randomization routine generates the same set of 150×2000 random numbers so that sampling error does not occur when we compare results. 10000 scenarios have also been run in some cases to verify the accuracy of the results.)

If the logarithmic rate of investment return (net of salary inflation) over year $(t-1,\ t)$ is $\delta(t)$, then the arithmetic rate of return (also net of salary inflation) is $i(t) = \exp(\delta(t)) - 1$ with $i = \mathrm{E}i(t)$ and $\sigma^2 = \mathrm{Var}i(t)$.

5.4.1 Valuation Discount Rate without Prudence Margin

The valuation discount rate is first assumed to be the same as the mean arithmetic rate of investment return, so that on average no surplus emerges. The following observations are based on the results displayed in Table 5.1 on the next page and Figure 5.1 on page 176 and Figure 5.2 on page 177.

Average Table 5.1 on the following page shows that when surpluses are being spread over shorter periods than deficits, on average a deficit emerges eventually. This is presumably because deficits are being removed more slowly than surpluses. Contributions are also higher on average. The converse applies when deficits are spread over shorter periods than surpluses.

Dispersion

1. Table 5.1 on the next page shows that fund levels are *less* dispersed (compare histograms A and D of Figure 5.1 on page 176), while contribution levels become *more* dispersed (compare histograms A and D of Figure 5.2 on page 177), as spread periods (equal for surpluses and deficits) are reduced from 20 to 5 years. In this case, Dufresne's (1986, 1988) 'optimal' spread period range is [1, 23] and this result is not surprising.

Surplus/deficit spread periods	20/20	5/20	20/5	5/5
Mean fund level	1.000 (1)	0.9521	1.049	0.9994 (1)
Mean contribution level	0.2000 (0.2)	0.2015	0.1926	0.2000 (0.2)
Variance of fund level	1.184×10^{-2} (1.174×10^{-2})	5.547×10^{-3}	7.844×10^{-3}	2.498×10^{-3} (2.490×10^{-3})
Variance of contribution level	5.035×10^{-5} (4.999×10^{-5})	6.119×10^{-5}	7.074×10^{-5}	1.106×10^{-4} (1.119×10^{-4})

Table 5.1: Sample statistics at time horizon, $i_v = i = 3\%$. (Exact results in the limit are in parentheses.) The top row shows the spreading periods for *surpluses/deficits* respectively. $\sigma = 3\%$, $AL = 1$, $NC = 0.2$, $B = 0.2291262$.

2. From Table 5.1, the variance of the fund level, when surpluses and deficits are not spread over the same periods, is intermediate between:

   - the value of the variance had the 'deficit' spread period been used to spread symmetrically both deficits and surpluses,
   - and the value of the variance had the 'surplus' spread period been used to spread symmetrically both surpluses and deficits.

   Haberman & Smith (1997) report similar results. At least in the cases considered (noting specifically that the 'surplus' and 'deficit' spread periods used are within Dufresne's (1986, 1988) 'optimal' spread period range), we see that shortening the period over which surpluses and deficits are spread reduces the variability in the fund level, but increases the variability in the contributions (Table 5.1).

Frequency Distribution and Skewness

1. Histograms A and D of Figure 5.1 on page 176 show that the distribution of the fund level at the time horizon is positively skewed. Dufresne (1990b) proves that, when the rate of investment return is modelled as a white noise process, the fund level (in continuous time) follows an inverse Gamma distribution. Cairns & Parker (1997) show by recursive methods that the inverse Gamma distribution is a good approximation when the logarithmic investment return is independently and identically normally distributed in a discrete time model. The distribution of the contribution level is negatively skewed (contribution level histograms A and D of Figure 5.2 on page 177). This is also obtained

with more complex asset-liability modelling, as is illustrated by MacBeth *et al.* (1994) and Haberman & Smith (1997).

2. Histograms A and B of Figure 5.1 on the following page illustrate that in the long term large surpluses are less frequent when surpluses and deficits are spread over 5 and 20 years respectively, as opposed to when both are spread over 20 years. Large surpluses can be removed even more drastically by spreading *both* valuation surpluses and deficits over shorter periods (histogram D of Figure 5.1 on the next page). Similarly, large deficits are less frequent when deficits at each valuation are spread over a reduced period (histograms A and C of Figure 5.1 on the following page).

3. Shortening the spread period for surpluses makes the frequency distribution for the fund level more peaked and symmetrical about the 'breakeven' point represented by the actuarial liability (histograms A and B of Figure 5.1 on the next page). Deficits, although not large ones, appear to occur more frequently on the whole.

4. The distribution of contribution levels exhibits a discontinuity since the contribution control applied to the pension funding system is now non-linear (histograms B and C of Figure 5.2 on page 177).

5.4.2 Valuation Discount Rate with Prudence Margin

The preceding results appear to agree with intuition. Now assume that there is a prudence margin in the valuation discount rate (taken to be 3%) so that it is less than the mean arithmetic rate of investment return (4%) on the fund. It is anticipated that a surplus will emerge in the fund in the long term. The following observations are based on the results displayed in Table 5.2 on page 178 and Figure 5.3 on page 180 and Figure 5.4 on page 181.

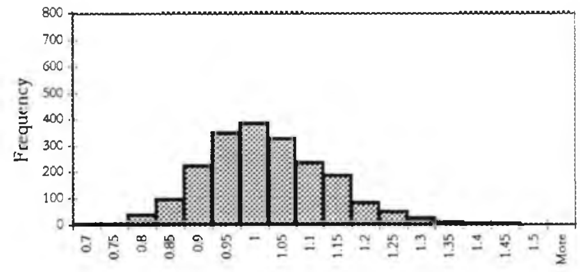
Average

1. When surpluses are spread over a shorter period than deficits, on average a smaller surplus eventually emerges (Table 5.2).

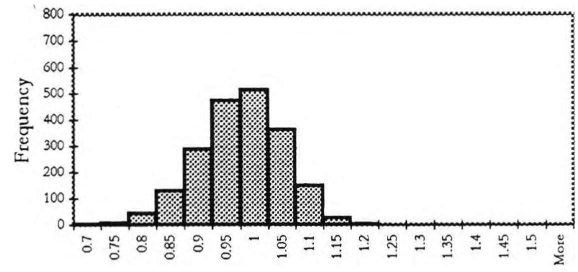
2. A comparison of the mean fund levels in Table 5.1 on the preceding page and Table 5.2 on page 178 shows that the prudence margin of 1% has led to a higher average fund level and a lower average contribution level.

Dispersion

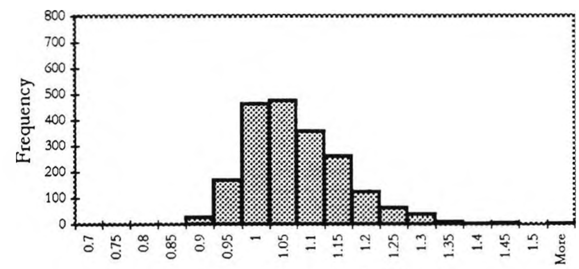
1. It may be observed from Table 5.2 on page 178 that shortening the period over which surpluses are spread reduces the variability in the fund level (just as in §5.4.1 when there was no prudence margin in the discount rate). The effect on contribution variances of



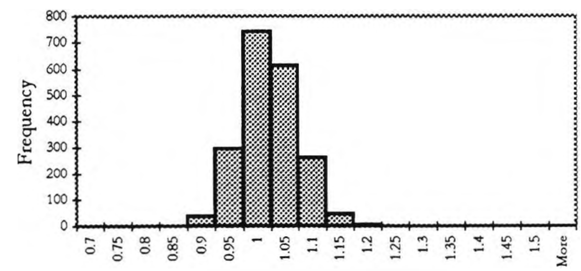
A. Surpluses: 20, Deficits: 20



B. Surpluses: 5, Deficits: 20

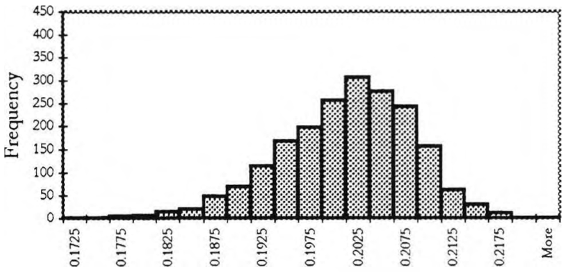


C. Surpluses: 20, Deficits: 5

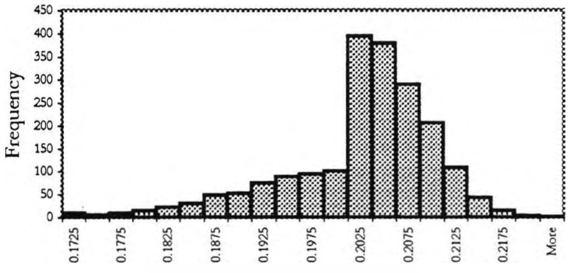


D. Surpluses: 5, Deficits: 5

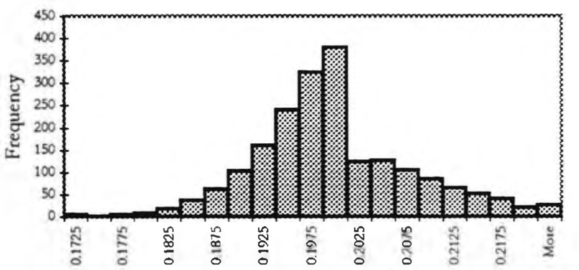
Figure 5.1: Histograms of fund level, with $i_v = i = 3\%$, for various combinations of spreading periods for surpluses and deficits. $\sigma = 0.03$, $AL = 1$, $NC = 0.2$, $B = 0.2291262$.



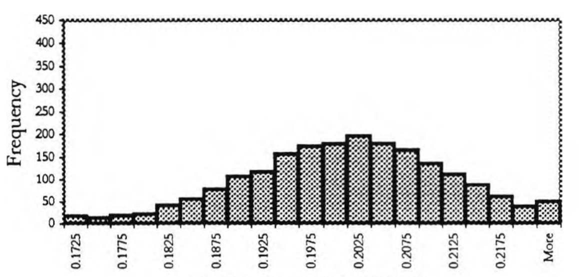
A. Surpluses: 20, Deficits: 20



B. Surpluses: 5, Deficits: 20



C. Surpluses: 20, Deficits: 5



D. Surpluses: 5, Deficits: 5

Figure 5.2: Histograms of contribution level, with $i_v = i = 3\%$, for various combinations of spreading periods for surpluses and deficits. $\sigma = 0.03$, $AL = 1$, $NC = 0.2$, $B = 0.2291262$.

Surplus/deficit spread periods	20/20	10/20	5/20	5/5
Mean fund level	1.341 (1.348)	1.121	1.047	1.053 (1.054)
Mean contribution level	0.1777 (0.1773)	0.1861	0.1889	0.1887 (0.1886)
Variance of fund level	2.777×10^{-2} (2.793×10^{-2})	7.287×10^{-3}	3.390×10^{-3}	2.826×10^{-3} (2.819×10^{-3})
Mean square deviation of fund from AL	0.1443 (0.1429)	2.197×10^{-2}	5.644×10^{-3}	5.652×10^{-3} (5.713×10^{-3})
Variance of contribution level	1.182×10^{-4} (1.189×10^{-4})	8.908×10^{-5}	1.125×10^{-4}	1.252×10^{-4} (1.267×10^{-4})
Mean square deviation of contribution from NC	6.154×10^{-4} (6.358×10^{-4})	2.835×10^{-4}	2.350×10^{-4}	2.537×10^{-4} (2.567×10^{-4})

Table 5.2: Sample statistics at time horizon, $i_v = 3\% < i = 4\%$. (Exact results in the limit are in parentheses.) The top row shows the spreading periods for *surpluses/deficits* respectively. $\sigma = 3\%$, $AL = 1$, $NC = 0.2$, $B = 0.2291262$.

treating surpluses and deficits differently is less clear (because of the discontinuity in the distribution of contributions—see below).

2. The mean square deviation of the fund level from the actuarial liability decreases as surpluses are removed over shorter periods: this is clearly visible when comparing histograms A and B of Figure 5.3 on page 180. This means that the surplus that is expected to emerge in the long run is being effectively removed.

Frequency Distribution and Skewness

1. The fund level is again observed to be positively skewed (histograms A and C of Figure 5.3 on page 180) and the contribution level is again negatively skewed (histograms A and C of Figure 5.4 on page 181).

2. Histograms A and C of Figure 5.3 on page 180 show that hastening the removal of surpluses does reduce the frequency of large surpluses emerging.

3. This does not seem to be at the expense of more frequent large deficits which could threaten the solvency of the fund. When the spread period for surplus is shortened, the frequency distribution for the fund level is more peaked and symmetrical.

4. The frequency distribution and statistics for the long-term fund levels with surpluses spread over 5 years and deficits over 20 years are not very different from those when both surpluses and deficits are spread over a short period of 5 years (compare histograms B and C of Figure 5.3 on the next page).

5. The distribution of contribution levels when surpluses and deficits are not spread equally again exhibits a discontinuity (histogram B of Figure 5.4 on page 181).

5.4.3 Conclusion

These observations broadly agree with the conclusions of Khorasanee (1993) and Haberman & Smith (1997:§12), although they use different (more 'realistic') economic/investment scenarios. It does appear that the incidence of large surpluses is diminished when the period over which surpluses (but not deficits) are spread is shortened. Depending upon the margins left in the actuarial assumptions, this may or may not increase the likelihood of large deficits, i.e. increase the risk of insolvency.

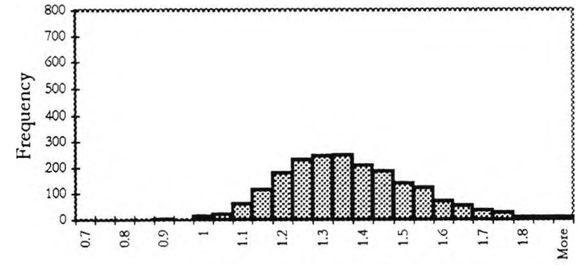
Application of this method causes various problems:

1. The unpredictability of a discontinuous contribution adjustment function may not appeal to the plan sponsor with budgetary requirements for stable contribution cash flows which change *smoothly* from year to year. From one year to the next, a possibly large negative adjustment may follow a negligible adjustment to the contribution, as an actuarial surplus succeeds a deficit when the fund is valued.

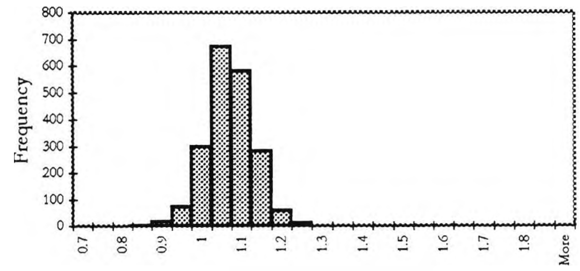
2. The overall consequence on the pace of funding for the plan is unclear and will depend on the prudence margins in the valuation basis. It will be even less clear to plan trustees and to the sponsor.

3. In terms of removing persisting surpluses and reducing the volatility of the fund level, similar results may apparently be obtained more simply by spreading *both* surpluses and deficits over *shorter* periods.

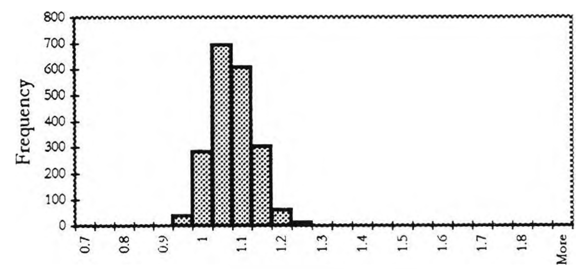
4. Finally, it is difficult to determine exactly what the shorter period for spreading surpluses, as opposed to deficits should be, especially as we do not know *a priori* the size of the margins left in the actuarial assumptions. This problem of parameter uncertainty in the control of the fund is compounded by the inherent non-linearity involved in the method. It is not clear that 'professional judgement' can be successfully applied to determine the different periods over which to spread surpluses and deficits.



A. Surpluses: 20, Deficits: 20

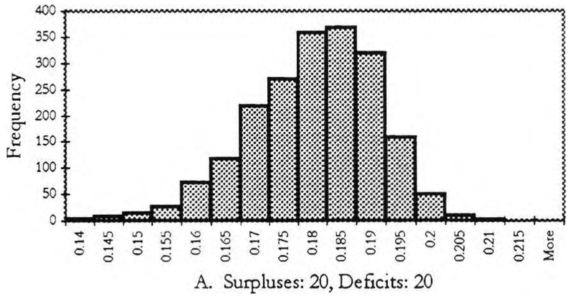


B. Surpluses: 5, Deficits: 20

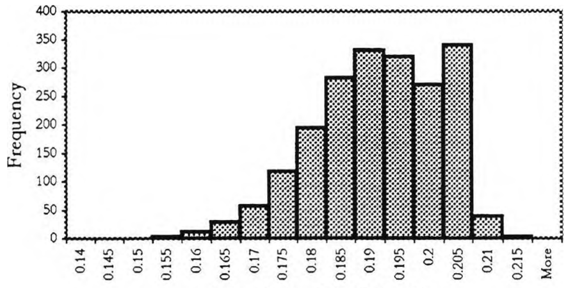


C. Surpluses: 5, Deficits: 5

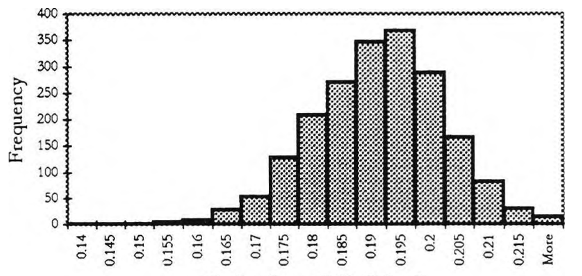
Figure 5.3: Histograms of fund level, with $i_v = 3\% < i = 4\%$, for various combinations of spreading periods for surpluses and deficits. $\sigma = 0.03$, $AL = 1$, $NC = 0.2$, $B = 0.2291262$.



A. Surpluses: 20, Deficits: 20



B. Surpluses: 5, Deficits: 20



C. Surpluses: 5, Deficits: 5

Figure 5.4: Histograms of contribution level, with $i_v = 3\% < i = 4\%$, for various combinations of spreading periods for surpluses and deficits. $\sigma = 0.03$, $AL = 1$, $NC = 0.2$, $B = 0.2291262$.

5.5 The 'Dual-Interest' Method

Another solution to the problem of persisting surpluses or deficits is to recognise explicitly that there is a difference between the valuation discount rate and the long-term rate of return on the fund. An estimate (say i_r) as to the long-term rate of return on assets is then made. It is in general different from the assumption regarding the discount rate used to value liabilities (i_v). A further adjustment to the contribution is then made taking this difference into account. Such a 'dual-interest' method can be studied within a simple model.

No statistical property is ascribed to the assumed rate of return i_r: it is usually a 'best estimate' based on past rates of return and an estimate of future reinvestment yields. Benjamin (1984, 1989), Loades (1992) and Fujiki (1994) consider i_r to be some average of previously observed values of $\{i(t)\}$.

Again, we restrict consideration to the method of spreading surpluses/deficits forward over a moving term as described in §3.2. If contributions and benefits are paid at the beginning of the year but investment income is received at the end of the year (Modelling Assumption 2.1),

$$c(t) = NC + (k + \kappa)(AL - f(t)) + AL(d_v - d_r). \tag{5.10}$$

AL and NC are calculated at the valuation discount rate i_v. (All compound interest symbols deriving from i_v, i_r bear the relevant subscript. $d = i/(1+i)$, $d_v = i_v/(1+i_v)$ and $d_r = i_r/(1+i_r)$.)

REMARK 5.1 *If all cash flows occur at the end of the year, replace d_r and d_v by i_r and i_v respectively; and if they are paid continuously, replace d_r and d_v by δ_r and δ_v respectively.*

Applications. One instance of this method is used in the calculation of pension expense according to the Financial Accounting Standards No. 87 (FAS87) in the U.S. Berin & Lofgren (1987) and Dufresne (1993) give a mathematical development of some aspects of FAS87. FAS87 uses two different rates: a discount rate to value liabilities as well as an "expected long-term rate of return on plan assets". It may be observed that in FAS87, $\kappa = d_r$:

$$c(t) = NC + k(AL - f(t)) + d_v AL - d_r f(t) \tag{5.11}$$

If $c(t)$ in equation (5.11) is analogous to the "pension expense", then NC is the "service cost", $k(AL - f(t))$ represents an "amortization payment", $d_v AL$ is the "interest cost" and $d_r f(t)$ is the "expected return on plan assets". FAS87 requires fixed-term amortization of gains/losses, but this may be modified to allow for the spreading of surpluses and deficits over a moving term.

Thornton & Wilson (1992a:§6.5) describe another instance of the method described in equation (5.10), with $\kappa = 0$. This is apparently adapted from FAS87 and intended for compatibility with the 'best-estimate' assumptions required by the U.K. Statement of Standard

Accounting Practice No. 24 (SSAP24). They refer to it as a 'dual-interest' method and describe it as

> "a variation of the projected unit method whereby the standard fund is calculated by reference to the prudent funding basis, but the contribution rate set takes into account the additional interest which it is expected will be earned on the fund, using a best estimate basis. The intended result of this method is that surpluses should not be built up, that a best estimate rate of interest is used for setting the contribution rate as required for [accounting] purposes, whilst a strong fund is preserved which will satisfy trustees and members."

If $\kappa = 0$ in equation (5.10),

$$c(t) = NC + k(AL - f(t)) + AL(d_v - d_r) \tag{5.12}$$

Rationale. The more general method in equation (5.10) is considered here. The contribution $c(t)$ in equation (5.10) may be replaced in equation (3.33) to yield

$$f(t+1) = u(t+1)\left[f(t) + NC + (k+\kappa)(AL - f(t)) + AL(d_v - d_r) - B\right]. \tag{5.13}$$

Liabilities are valued using the valuation discount rate and so, using the equation of equilibrium (3.65),

$$f(t+1) = u(t+1)[(1-k-\kappa)f(t) + AL(k+\kappa-d_r)]. \tag{5.14}$$

The rationale of this method is immediately obvious if it is assumed that the rates of return earned in future are exactly (and ideally) equal to the assumption about the long-term rate of return on plan assets, i.e. $i(t) = i = i_r \; \forall t$. (The subscript may therefore be dropped.) i may or may not be equal to the liability valuation discount rate i_v. Equation (5.14) may be rewritten in terms of the unfunded liability to give

$$\begin{aligned} ul(t+1) &= AL + u[(1-k-\kappa)ul(t) - vAL], \\ &= u(1-k-\kappa)ul(t), \end{aligned} \tag{5.15}$$

so that

$$ul(t) \;=\; ul_0[u(1-k-\kappa)]^t \;\rightarrow\; 0 \text{ as } t \rightarrow \infty, \tag{5.16}$$

provided that $|u(1-k-\kappa)| < 1$. No long-term surplus or deficit emerges, even though there may exist a margin between the rate of return on assets $i = i_r$ and the liability valuation discount rate i_v.

Random Biased Returns. In order to explore the effect of investment gains and losses, rates of return $\{i(t)\}$ must be assumed to be random. Now, the assumption as to the "long-term rate of return on plan assets" (i_r) may be biased, such that $\mathrm{E}i(t) = i \neq i_r \neq i_v$. Proposition 5.1 summarises some important results. It is assumed that the dual-interest method is used as in equation (5.10) and that $\{i(t)\}$ is identically and independently distributed.

PROPOSITION 5.1 *Suppose* $\mathrm{E}i(t) = i \neq i_r \neq i_v$. *Provided that*

$$|u(1-k-\kappa)| < 1, \tag{5.17}$$

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL(d_r - (k+\kappa))/(d-(k+\kappa)), \tag{5.18}$$

$$\lim_{t\to\infty} \mathrm{E}ul(t) = AL(d-d_r)/(d-(k+\kappa)), \tag{5.19}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + AL(d_v - d_r) + AL(k+\kappa)(d-d_r)/(d-(k+\kappa)). \tag{5.20}$$

Let $q = u^2 + \sigma^2$. *If*

$$k + \kappa > 1 - 1/\sqrt{q}, \tag{5.21}$$

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \sigma^2 v^2 AL^2 [d_r - (k+\kappa)]^2 / [1 - q(1-k-\kappa)^2][d-(k+\kappa)]^2, \tag{5.22}$$

$$\lim_{t\to\infty} \mathrm{E}ul(t)^2 = \lim \mathrm{Var}f(t) + \lim[\mathrm{E}ul(t)]^2, \tag{5.23}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t) = (k+\kappa)^2 \lim \mathrm{Var}f(t). \tag{5.24}$$

The proof, in Appendix F (§F.1), is a straightforward application of the method of Dufresne (1988). Remarks:

1. The conditions for stability in the first two moments (inequalities (5.17) and (5.21)) depend exclusively on the mean and variance of the rate of return process. The limits on $k+\kappa$ depend *neither* on the assumed rate of return (i_r) *nor* on the valuation discount rate (i_v).

2. The 'pace' of funding depends on $k+\kappa$ rather than k alone, as is clear from equations (5.14) and (5.16). For equal spreading periods the method implied by FAS87 ($\kappa = d_r$) will require that deficits are financed faster than in the method described by Thornton & Wilson (1992a:§6.5) ($\kappa = 0$).

3. k is the reciprocal of an annuity that is usually calculated at the valuation discount rate. If we assume that k is only a surplus/deficit spreading factor, then the funding level is independent of the choice of the valuation discount rate but does depend on the assumption made regarding expected future investment returns. As a proportion of the actuarial liability, the first two moments of the ultimate funding level ($f(t)/AL$) in equations (5.18) and (5.22) are independent of the valuation discount rate. The funding level over time is also independent of the valuation discount rate, as may be observed in equation (5.14).

Random Unbiased Returns. More interesting results follow if it is assumed that the assumption as to the long-term return on plan assets is a 'best' estimate, is not biased and is equal to the expectation of the rate of return $i(t)$, i.e. $i_r = \mathrm{E}i(t) = i$. (For economy, the subscript r may be dropped.) Corollary 5.1 immediately follows from Proposition 5.1. It is assumed again that the dual-interest method is used as in equation (5.10), $\{i(t)\}$ is identically and independently distributed.

COROLLARY 5.1 *Suppose* $\mathrm{E}i(t) = i = i_r \neq i_v$. *Provided that*

$$|u(1-k-\kappa)| < 1, \tag{5.25}$$

then,

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL, \tag{5.26}$$

$$\lim_{t\to\infty} \mathrm{E}ul(t) = 0, \tag{5.27}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + AL(d_v - d). \tag{5.28}$$

Furthermore, provided that

$$k + \kappa > 1 - 1/\sqrt{q}, \tag{5.29}$$

then

$$\lim_{t\to\infty} \mathrm{Var}f(t) = \lim_{t\to\infty} \mathrm{E}ul(t)^2 = \sigma^2 v^2 AL^2/[1 - q(1-k-\kappa)^2], \tag{5.30}$$

$$\lim_{t\to\infty} \mathrm{Var}c(t) = (k+\kappa)^2 \lim \mathrm{Var}f(t), \tag{5.31}$$

Remarks:

1. When surpluses and deficits are spread in the normal way, it was found in §5.3 that the ultimate funding level depended on the margin between the valuation discount rate and the mean of the rate of return process: this margin resulted in a persisting average surplus (or deficit) in the long-term. In Corollary 5.1, use of the 'dual-interest' method means that ultimately the surplus is *expected* to be removed (equations (5.26) and (5.27)) *even though* the discount rate used to value liabilities is different from the expected rate of return on assets.

2. The mean square deviation of the ultimate surplus or deficit depends *directly* on $k + \kappa$ (equation (5.30)). The selection of $\kappa = 0$ as in Thornton & Wilson's (1992a:§6.5) method should therefore lead to more stable fund levels and smaller surpluses and deficits emerging in the limit, compared to $\kappa = d$ as according to FAS87. The spreading period m and $k = 1/\ddot{a}_{\overline{m|}}$ have an inverse relationship and, as a consequence, spreading surpluses and deficits over *shorter* periods will reduce the size of surpluses and deficits that arise in the long-term. This is similar to the conclusion from equation (5.4) when normal spreading is used in §5.3.

3. The variances of fund and contribution levels in equations (5.30) and (5.31) are in exactly the same form as the variances in the case when conventional spreading of surpluses and deficits is used (Result 3.1). It is immediately possible to infer that there exists an 'optimal' spread period range (as defined by Dufresne (1986)) such that the choice of any spread period outside this range results in a higher variability of the fund level for the same contribution stability as would be obtained from spreading surpluses and deficits over a term in the 'optimal' range. By direct comparison with the result of Dufresne (1986, 1988), the 'optimal' spread period range is now such that $k \in [1, k^* - \kappa]$, where k^* is as in Result 3.4.

4. The most important difference between Proposition 5.1 and Corollary 5.1 lies in equations (5.19) and (5.27). The expected ultimate surplus or deficit is zero only if $i_r = \mathrm{E}i(t) = i$.

Conclusion. The 'dual-interest' method is therefore effective at avoiding long-term persisting surpluses and deficits, especially when combined with short spreading periods. It depends on an *explicit* distinction being made between the discount rate used to value liabilities and the estimate of the rate of return on present and future plan assets. The 'dual-interest' method only succeeds in removing persisting surpluses and deficits on average if an *unbiased* assumption is made as to future rates of return. The next method does not require any such assumption about future returns to be made.

5.6 An 'Integral' Method

5.6.1 Description and Rationale

It may be difficult to estimate the mean long-term return on the pension fund. The investment returns on certain asset classes are very volatile and difficult to predict. Furthermore, investment objectives and strategic asset allocation may themselves also change. It is easier by comparison to estimate the liability valuation discount rate, especially net of price and salary inflation, as it relates to the portfolio that hedges or matches liabilities and is therefore less volatile. More confidence may thus be attached to the estimate of i_v than i_r in §5.5.

The third method for removing persisting surpluses and deficits that is considered in this chapter does not require any foreknowledge of the mean long-term rate of return on the fund. It is based on the concept of "integral control" as suggested in other contexts by a number of authors, notably Balzer (1982), but it does not appear to be used directly in practice. The recommended contribution is calculated as follows:

$$c(t) = NC + k_p(ul(t) - U(t)) + k_i \sum_{j=0}^{t}(ul(j) - U(j)) + P(t), \tag{5.32}$$

where

$ul(t)$ is the unfunded liability or deficit;

$P(t)$ is a payment towards the amortization of the initial unfunded liability ul_0 (or some fraction $1 - y$ of it) over a fixed schedule of n years (see equation (4.7));

$U(t)$ is the unamortized part of the initial unfunded liability ul_0 (or of the fraction of it being amortized) (see equation (4.8));

k_p and k_i represent actuarial control parameters.

The first term on the right hand side of equation (5.32), NC, is the normal cost or standard contribution that is paid if no experience gain or loss emerges. The second term, $k_p(ul(t) - U(t))$, represents proportional spreading of the current unfunded liability of the pension plan. If a fraction $1-y$ of the initial unfunded liability is being 'frozen' and amortized separately, then the current unfunded liability *in excess of* the unamortized portion of the 'frozen' initial unfunded liability is being spread. This term is therefore akin to spreading forward surpluses or deficits (§3.2), with k_p replacing $k = 1/\ddot{a}_{\overline{m|}}$. It is not essential that the initial unfunded liability be 'frozen' and amortized separately (y may be set to 1), although this may prove to be useful to control the pace of funding for example (see §3.2.4).

The third term on the right hand side of equation (5.32), $k_i \sum(\dots)$, represents historic or integral spreading of cumulative unfunded liabilities of the pension plan. If a fraction $1 - y$ of the initial unfunded liability is being 'frozen' and amortized separately, then this should refer, more precisely, to the unfunded liabilities in excess of the unamortized portions of the 'frozen' initial unfunded liability. This term is zero if the cumulative sum (without interest) of previous and current unfunded liabilities is zero. This would happen if all actuarial gains and losses have 'cancelled out'.

Gain/loss adjustment as in equation (5.32) has been described in various other actuarial systems:

1. Balzer (1982) employs a similar method in the context of a deterministic, discrete-time general insurance system. He refers to the problem of a "persisting surplus or deficit" resulting from a "persisting stream of unpredicted claims". The typical actuarial solution to this, according to him, is to carry out an "*ad hoc* adjustment" to the valuation basis in the valuation process or "predictor". Balzer (1982) suggests instead the use of "integral action" represented by a term similar to the third term on the right hand side of equation (5.32).

2. Taylor (1987) considers a more general funding system, allowing for investment income. Taylor (1987:§6) describes some general formulae to calculate premiums from claim payments and suggests a formula similar to the 'integral' method above. In the event that "deviations of actual from predicted claims experience are too persistent", he suggests

that "ad hoc methods" may be used to change the valuation basis. Taylor (1987:§7) concludes that "estimated claims escalation could be monitored, and the premium formula changed when experience appeared to have departed sufficiently from the expectations implicit in the formula."

3. Loades (1998) also uses an 'integral' method to control pension funding (in a deterministic and continuous-time context) and to remove the "offset" that occurs when there are "interest margins" so that the pension funding system can "home in on the required funding level."

Substitution of the contribution level $c(t)$ from equation (5.32) into equation (3.33) and using the equation of equilibrium (3.64) yields a recurrence relationship in terms of the fund level:

$$f(t+1) = u(t+1)\Bigg[(1-k_p)f(t) - k_i\sum_{j=0}^{t} f(j) + (k_p - d_v)AL + (t+1)k_iAL \\ - k_pU(t) - k_i\sum_{j=0}^{t} U(j) + P(t)\Bigg], \quad (5.33)$$

for $t \geq 0$. This can equally be written in terms of the unfunded liability, by substituting equation (5.32) into equation (3.66), as

$$(ul(t+1) - U(t+1)) - (AL - U(t+1)) \\ = u(t+1)\Bigg[(1-k_p)(ul(t) - U(t)) - k_i\sum_{j=0}^{t}(ul(j) - U(j)) - v_v(AL - U(t+1))\Bigg], \quad (5.34)$$

for $t \geq 0$.

5.6.2 First Moments

It is easier to derive the moments of the fund and contribution levels using $ul(t)$ in equation (5.34). It is more instructive in the first instance to use equation (5.33). We note again that $u(t+1)$ is independent of $u(t)$, $u(t-1)$ etc. and therefore also of $f(t)$, $f(t-1)$ etc. Therefore, when expectations are taken on both sides of equation (5.33),

$$\mathrm{E}f(t+1) = u(1-k_p)\mathrm{E}f(t) - uk_i\sum_{j=0}^{t}\mathrm{E}f(j) + u(k_p - d_v)AL + (t+1)uk_iAL \\ - uk_pU(t) - uk_i\sum_{j=0}^{t} U(j) + uP(t), \quad (5.35)$$

for $t \geq 0$. Forward-shift equation (5.35) in time (so that it holds for $t \geq -1$) and deduct equation (5.35) to obtain

$$\begin{aligned} \mathrm{E}f(t+2) - \mathrm{E}f(t+1) &= u(1-k_p)[\mathrm{E}f(t+1) - \mathrm{E}f(t)] - uk_i\mathrm{E}f(t+1) + uk_iAL \\ &\quad - uk_p(U(t+1) - U(t)) - uk_iU(t+1) + u(P(t+1) - P(t)) \quad (5.36) \end{aligned}$$

which holds for $t \geq 0$ and requires $\mathrm{E}f(0) = f_0$ and an additional initial condition $\mathrm{E}f(1)$ that may be found from equation (5.35). Since these initial conditions have no effect in the limit, and we are interested in limit results only at this stage, we do not require them. Equation (5.36) may be simplified:

$$\begin{aligned} &\mathrm{E}f(t+2) - [1 - u(1-k_p-k_i)]\mathrm{E}f(t+1) + u(1-k_p)\mathrm{E}f(t) \\ &\quad = uk_iAL - u(k_p+k_i)U(t+1) + uk_pU(t) + u(P(t+1) - P(t)). \quad (5.37) \end{aligned}$$

The absolute value of the roots of the characteristic equation

$$P(z) = z^2 - [1 + u(1-k_p-k_i)]z + u(1-k_p) = 0 \quad (5.38)$$

of equation (5.37) must be less than unity for asymptotic stability. Well-known necessary and sufficient conditions for this (as in §4.5.4) are:

$$|u(1-k_p)| < 1 \quad (5.39)$$

$$P(1) = 1 - [1 + u(1-k_p-k_i)] + u(1-k_p) > 0 \;\Leftrightarrow\; uk_i > 0, \quad (5.40)$$

$$P(-1) = 1 + [1 + u(1-k_p-k_i)] + u(1-k_p) > 0 \;\Leftrightarrow\; 1 + u(1-k_p) > uk_i/2, \quad (5.41)$$

Conditions (5.40) and (5.41) may also be combined to give

$$|1 + u(1-k_p-k_i)| \; < \; 1 + u(1-k_p). \quad (5.42)$$

It is necessary and sufficient that these conditions hold for $\mathrm{E}f(t)$ in equation (5.37) to be asymptotically stable:

$$\lim_{t\to\infty} \mathrm{E}f(t) = uk_iAL/[1 - (1 + u(1-k_p-k_i)) + u(1-k_p)] = AL. \quad (5.43)$$

Upon taking expectations on both sides of equation (3.33) and noting that $u(t+1)$ is independent of $f(t)$ and $c(t)$,

$$\mathrm{E}f(t+1) = u(\mathrm{E}f(t) + \mathrm{E}c(t) - B) \quad (5.44)$$

and the limit as $t \to \infty$ of $\mathrm{E}c(t)$ exists if $\mathrm{E}f(t)$ converges as $t \to \infty$. Hence,

$$\begin{aligned} \lim_{t\to\infty} \mathrm{E}c(t) &= B - d \lim_{t\to\infty} \mathrm{E}f(t) \\ &= B - dAL \quad \text{(from equation (5.43))} \\ &= NC + (d_v - d)AL, \quad (5.45) \end{aligned}$$

where the equation of equilibrium (3.65) is used in the last step.

Note that condition (5.40) proscribes $k_i = 0$. This must be examined separately. When $k_i = 0$, the contribution adjustment function in equation (5.32) is exactly identical to the situation where surpluses and deficits are being spread forward over a moving term. The characteristic equation from equation (5.35) is then

$$z - u(1 - k_p) = 0 \tag{5.46}$$

requiring $|u(1 - k_p)| < 1$ for stability so that

$$\lim_{t\to\infty} \mathrm{E}f(t) = ALu(k_p - d_v)/[1 - u(1 - k_p)] = AL(d_v - k_p)/(d - k_p). \tag{5.47}$$

This is the result of Dufresne (1986). If there is no margin between the valuation discount rate and the mean long-term rate of return, i.e. $d = d_v$, then $\lim \mathrm{E}f(t) = AL$. Other first moments are as in Result 5.1 (with k_p replacing k).

These results are summarised in Proposition 5.2.

PROPOSITION 5.2 *If and only if* $|u(1 - k_p)| < 1$ *and* $|1 + u(1 - k_p - k_i)| < 1 + u(1 - k_p)$ ($\Rightarrow$ $k_i \neq 0$, $i \neq -1$)*, then*

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL, \tag{5.48}$$

$$\lim_{t\to\infty} \mathrm{E}ul(t) = 0, \tag{5.49}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + (d_v - d)AL. \tag{5.50}$$

If $k_i = 0$ *and* $|u(1 - k_p)| < 1$*, then*

$$\lim_{t\to\infty} \mathrm{E}f(t) = AL(d_v - k_p)/(d - k_p), \tag{5.51}$$

$$\lim_{t\to\infty} \mathrm{E}ul(t) = AL(d - d_v)/(d - k_p), \tag{5.52}$$

$$\lim_{t\to\infty} \mathrm{E}c(t) = NC + ALk_p(d - d_v)/(d - k_p). \tag{5.53}$$

A shorter proof may be obtained by using $ul(t)$ rather than $f(t)$: see Appendix F (§F.2). Remarks on Proposition 5.2:

1. When $k_i \neq 0$, i.e. when 'integral' spreading of surpluses and deficits is used in the regulation of the contribution, the fund is ultimately *expected* to equal the value placed on the liabilities (the actuarial liability) calculated at the *risk-adjusted* and *prudent* discount rate i_v (equation (5.48)).

2. It was observed in §5.3 that the ultimate pension funding level depends on the margin between the valuation discount rate and the long-term average rate of investment return on the fund when surpluses and deficits are spread over a moving term. A persisting

average surplus emerges if such a prudence margin exists. (Equations (5.4) and (5.52) are identical if $k_p = k$.) It was found in §5.5 that if the mean rate of return on the pension fund is estimated *without bias*, then it is possible to adjust contributions such that long-term average surpluses or deficits are avoided. Equation (5.49) shows that it is possible to remove undesirable permanent surpluses or deficits caused by margins in the valuation discount rate *without* making any prior assumption regarding investment returns on the pension fund.

3. The conditions for stability in the first moments do *not* depend on the valuation discount rate, but only on the mean of the actual rate of return process.

4. The first moments in the limit when $k_i \neq 0$ depend neither on k_p nor on k_i, but when $k_i = 0$ they do depend on k_p (which may be a function of the spreading period).

Sufficient Conditions. Stability conditions (5.39) and (5.42) may be simplified. Sufficient conditions are gathered in Corollary 5.2.

COROLLARY 5.2 *Provided that*

$$i > -100\%, \tag{5.54}$$

$$d < k_p \leq 1, \tag{5.55}$$

$$0 < k_i < 2(1 - d + 1 - k_p). \tag{5.56}$$

the first moments of the funding process in equations (5.48), (5.49) *and* (5.50) *follow.*

Proof. Inequalities (5.54) and (5.55) imply that $0 \leq u(1 - k_p) < 1 \Rightarrow |u(1 - k_p)| < 1$ (condition (5.39)). From inequality (5.56), $k_i > 0$ and given inequality (5.54), condition (5.40) follows. Likewise, from inequality (5.56), $k_i < 2(1 - d + 1 - k_p)$ and given inequality (5.54), $uk_i < 2(1 + u(1 - k_p))$ and condition (5.40) follows. □

Remarks on Corollary 5.2:

1. When integral spreading is used, conditions (5.54) and (5.55) for stability imply that $k_p \neq 0$ and condition (5.56) implies that $k_i \neq 0$.

2. Condition (5.54) is realistic under 'normal' economic conditions. Rates of investment return cannot be negative forever in a growing economy.

3. Conditions (5.54) and (5.55) are similar to conditions (5.2) and (5.3) respectively. k_p is analogous to the reciprocal annuity factor $k = 1/\ddot{a}_{\overline{m|}}$ when surpluses and deficits are spread over a moving term m.

4. A sufficient upper bound may be placed on k_i for stability in the first moments. Since the lowest value of $1 - k_p$ is 0 in inequality (5.55), it follows from condition (5.56) that

$0 < k_i \leq 2v$. This provides for a wide range of values of k_i under usual economic circumstances.

5. The ultimate mean contribution level in equation (5.50) is the same as is in the dual-interest method (equation (5.28)), although no assumption (biased or unbiased) as to the future average rate of return on assets has been made here. The system settles at the same 'equilibrium state' in both cases.

6. Comparison of equations (5.50) and (5.53) reveals that when there is a prudence margin between the valuation discount rate and long-term mean rate of return and when integral spreading is used, as opposed to pure spreading, the long-term average contribution level is closer to the normal cost or standard contribution NC (since $k_p/(k_p - d) > 1$ from stability conditions (5.54) and (5.55)), i.e. a smaller (in magnitude) average supplementary contribution is required from the sponsor.

Integral Spreading. From equation (5.32), as $t \to \infty$, it is clear that

$$\lim \mathrm{E}\left[k_i \sum_{j=0}^{t}(ul(j) - U(j))\right] = \lim \mathrm{E}c(t) - NC - k_p \lim \mathrm{E}(ul(t) - U(t) - \lim P(t)$$

$$= AL(d_v - d), \tag{5.57}$$

$$\lim \sum_{j=0}^{t}(\mathrm{E}ul(j) - U(j)) = AL(d_v - d)/k_i, \tag{5.58}$$

after substituting $\lim \mathrm{E}ul(t)$ from equation (5.49) and $\lim \mathrm{E}c(t)$ from equation (5.50). The right hand side of equation (5.57) is of course identical to the adjustment term in the 'dual-interest' contribution function (the third term on the right hand side of equation (5.10)) when the mean rate of return is estimated without bias ($d_r = d$). The same adjustment is performed 'automatically' when integral spreading is used, without any estimation of the rate of return on the fund. If there is a difference between the valuation discount rate i_v and the long-term mean rate of return on the fund i, actuarial gains and losses will continually emerge. They give rise to the cumulative surplus or deficit represented in equation (5.58) and contributions are accordingly adjusted.

5.6.3 Second Moments

Pension fund gains and losses are volatile and random. The first moment results of §5.6.2 are interesting but it is much more important to consider the variability of surpluses and deficits as they emerge. The second moments of the fund and contribution levels are found in Proposition 5.3.

PROPOSITION 5.3 *Let*

$$V_\infty = \sigma^2 v^2 AL^2/[1-q(1-k_p)^2 - qk_i(k_p-d)/2]. \tag{5.59}$$

Provided that conditions (5.54)–(5.56) *are true and also provided that*

$$0 < k_i < 2u[1-q(1-k_p)^2]/q[1-u(1-k_p)] \tag{5.60}$$

and

$$[1+q(1-k_p)^2][1-qu^2(1-k_p)^4] + u(1-k_p)[1-q(1-k_p)^2][1+q(1-k_p-k_i)^2] > 2q(1-k_p)k_i[1-u^2(1-k_p)^2], \tag{5.61}$$

then

$$\lim_{t\to\infty} \mathrm{Var} f(t) = \lim_{t\to\infty} \mathrm{E}ul(t)^2 = V_\infty, \tag{5.62}$$

$$\lim_{t\to\infty} \mathrm{Var} c(t) = V_\infty[k_p^2 + k_i + k_i(k_p-d)/2], \tag{5.63}$$

$$\lim_{t\to\infty} \mathrm{Cov}[f(t), c(t)] = -V_\infty(k_p + k_i/2). \tag{5.64}$$

The proof of Proposition 5.3, in Appendix F (§F.3), uses some results from Appendix D. Remarks on Proposition 5.3:

1. The conditions for stability (conditions (5.54)–(5.56), (5.60) and (5.61)) in the second moments of the pension funding process are independent of the valuation discount rate i_v and depend only on the moments of the rate of return process $\{i(t)\}$.

2. Assume that k_p and k_i are independent of the valuation discount rate i_v. The variance of the funding level (i.e. fund level as a proportion of the actuarial liability) ultimately is also independent of the valuation discount rate (equation (5.62)).

About Stability. Conditions (5.54)–(5.56), (5.60) and (5.61) are *sufficient* for stability in the second moments. Necessary and sufficient conditions are discussed in Appendix F (§F.3). The sufficient conditions are accurate under 'normal' conditions.

The most constraining condition is found numerically to be inequality (5.60). As $\sigma \to 0$, condition (5.60) tends to condition (5.56) since

$$2u[1-u^2(1-k_p)^2]/u^2[1-u(1-k_p)] = 2v[1+u(1-k_p)] = 2[1-d+1-k_p]. \tag{5.65}$$

A lower bound is easily placed on k_p. It is shown in the proof of Proposition 5.3 (Appendix F) that stability conditions (5.54), (5.55) and (5.60) imply that

$$k_p > 1 - 1/\sqrt{q} \tag{5.66}$$

(see equation (F.56) in §F.3). Compare inequality (5.66) with condition (5.6) when surpluses and deficits are spread forward.

Condition (5.55) is discussed in §5.6.2 and is found to be similar to condition (5.3) when surpluses and deficits are spread forward. Since condition (5.66) is also similar to (5.6), k_p is similar to the reciprocal annuity factor k when conventional spreading is considered.

I will henceforth assume, without loss of generality, that

$$k_p = k = 1/\ddot{a}_{\overline{m|}}, \tag{5.67}$$

$$k_i = 1/m_i. \tag{5.68}$$

The annuity in equation (5.67) is calculated at i_v, as is typical in practice. Balzer (1982) uses a parameter similar to k_i, whereas Loades (1998) uses an "integral time" parameter similar to m_i. Equation (5.68) represents a form of amortization without interest. Interest is ignored in the amortization expense of U.S. Financial Accounting Standards No. 87 (FAS87) and also in the spreading of surpluses and deficits by Dyson & Exley (1995:§7.5.3.12).

Condition (5.60) therefore requires that

$$m_i > m_i^{min} = q(k-d)/2[1-q(1-k)^2]. \tag{5.69}$$

Table 5.3 on the next page shows m_i^{min}, i.e. the *minimum* value of m_i (corresponding to a maximum k_i) according to the stability conditions in Proposition 5.3, for various valuation discount rates (net of salary inflation) i_v, mean rate of return i (also net of salary inflation) and spread period m. It is assumed that the standard deviation of the rate of return is $\sigma = 0.1$. The minimum value of m_i may be seen to increase as the prudence margin $i - i_v$ increases. We envisage that $m_i > 1$ since it is akin to a 'spreading' period. Table 5.3 on the following page shows that for typical spread periods up to 15 years, real valuation discount rates of up to 5% and prudence margins of 3.0% or less, m_i may lie in the range $[1, \infty)$.

5.6.4 Effect on Fund Level

The speed with which retirement benefits are funded depends initially on the period used to amortize any 'frozen' initial unfunded liability. Once these initial deficits or surpluses are amortized, the pace of funding will depend on the choice of k_p and k_i. If the roots of the characteristic equation (5.38) are α and β, then the expected deficit takes the form

$$\mathrm{E}ul(t) = A\alpha^t + B\beta^t, \tag{5.70}$$

from the second order linear difference equation (F.10) (in Appendix F), where initial unfunded liabilities are ignored and it is assumed that α and β are not coincident. Stability conditions (5.54)—(5.56) ensure that $|\alpha| < 1$ and $|\beta| < 1$. In terms of first moments, funding will accelerate (i.e. $\mathrm{E}ul(t)$ vanishes at a faster rate) if $|\alpha|$ and $|\beta|$ are small. $\alpha\beta = u(1-k_p)$

i_v	$i - i_v$	$m = 1$	3	5	10	15	20	25	30	40	50
1%	0.00%	0.510	0.308	0.288	0.280	0.283	0.290	0.299	0.309	0.336	0.372
	0.25%	0.511	0.308	0.288	0.281	0.285	0.292	0.302	0.315	0.350	0.406
	0.50%	0.512	0.309	0.289	0.281	0.286	0.295	0.307	0.322	0.371	0.467
	1.00%	0.515	0.310	0.290	0.283	0.289	0.300	0.318	0.344	0.462	1.487
	1.50%	0.517	0.311	0.290	0.284	0.292	0.307	0.335	0.386	1.287	
3%	0.00%	0.520	0.313	0.292	0.285	0.291	0.301	0.316	0.336	0.402	0.549
	0.25%	0.521	0.313	0.293	0.286	0.292	0.304	0.322	0.348	0.455	0.930
	0.50%	0.522	0.313	0.293	0.287	0.294	0.307	0.330	0.366	0.575	
	1.00%	0.525	0.314	0.294	0.288	0.297	0.316	0.353	0.436		
	1.50%	0.527	0.315	0.295	0.290	0.301	0.330	0.401	0.802		
5%	0.00%	0.530	0.317	0.297	0.290	0.298	0.314	0.339	0.378	0.580	4.628
	0.50%	0.532	0.318	0.298	0.292	0.302	0.324	0.367	0.467		
	1.00%	0.535	0.319	0.298	0.294	0.307	0.340	0.431	1.208		
	2.00%	0.540	0.321	0.300	0.298	0.321	0.427				
	3.00%	0.545	0.323	0.302	0.302	0.352					
10%	0.00%	0.555	0.329	0.308	0.304	0.322	0.363	0.465	0.852		
	0.50%	0.557	0.330	0.309	0.306	0.330	0.406	0.984			
	1.00%	0.560	0.331	0.310	0.309	0.342	0.530				
	2.00%	0.564	0.333	0.311	0.314	0.396					
	3.00%	0.569	0.334	0.313	0.322	1.231					

Table 5.3: m_i^{min}, or *minimum* allowable m_i based on stability conditions (5.54)–(5.56), (5.60) and (5.61), for various $\{i_v, i, m\}$ and for $\sigma = 0.1$. $k_p = 1/\ddot{a}_{\overline{m|}}$, where the annuity is calculated at i_v. $k_i = 1/m_i$. Blanks indicate that stability conditions do not hold.

is likely to be small, or k_p is large. Also, $\alpha + \beta = 1 + u(1 - k_p - k_i)$ is small, or k_i is likely to be large. It appears therefore that large k_p and k_i (or short m and m_i) hasten the pace of funding, although I do not prove this mathematically. This is illustrated in Figure 5.5 on the next page. The parameters k_p and k_i must thus be chosen *in combination* to achieve the requisite 'pace' of funding. Note also that if α and β are conjugate imaginary roots or negative real roots, then oscillations will occur in $\mathrm{E}ul(t)$ in equation (5.70). Conjugate roots will not occur of course if

$$[1 + u(1 - k_p - k_i)]^2 \geq 4u(1 - k_p) \tag{5.71}$$

(based on the determinant of the quadratic in equation (5.38)). Funding without transient cyclic disturbances can be achieved fastest when α and β are coincident real roots, i.e. if the inequality in (5.71) is replaced by an equality. Features such as 'oscillations', 'speed of response', 'critical damping' and 'overshoots' are investigated in some detail by Balzer (1982) and Loades (1998). Balzer (1982) shows that delays in the feedback process in the funding system may lead to oscillations and he uses the 'root locus' method to determine suitable control parameters.

Such considerations may not be particularly relevant to the present model:

1. Oscillations may not be particularly large if the initial unfunded liability (often the largest source of deficit or surplus in the fund) is being amortized separately. For example, if at some time (say $t = 0$) the initial unfunded liability is completely removed ($U(t) = 0$, $t \geq 0$), if there is no surplus or deficit remaining ($\mathrm{E}ul(0) = 0$), and previous gains and losses were negligible ($\sum_{\tau \leq 0} \mathrm{E}ul(\tau) = 0$), equation (F.9) shows that $\mathrm{E}ul(1) = 0$; and from equation (F.10), $\mathrm{E}ul(t) = 0$, $t \geq 0$.

2. Stochastic variation in the funding process (particularly through the investment returns of the pension fund) may dominate and swamp oscillations. (Oscillations may nevertheless be of a large amplitude, as Balzer (1982) and Loades (1998) demonstrate, if k_p is small—which is comparable to a short spreading period—and/or if there are long delays.)

It is more important to consider the variability of the pension fund surpluses and deficits in the long term. This may be measured using the long-term mean-square unfunded liability. A two-sided measure is appropriate since both excessive surpluses and excessive deficits are undesirable.

PROPOSITION 5.4 *Let $\Delta = d - d_v$ represent the prudence margin between the valuation discount rate and the long-term mean rate of return.*

Suppose $\Delta = 0$. Then it is more efficient, in terms of minimising long-term mean square unfunded liability, to use conventional, rather than integral, spreading.

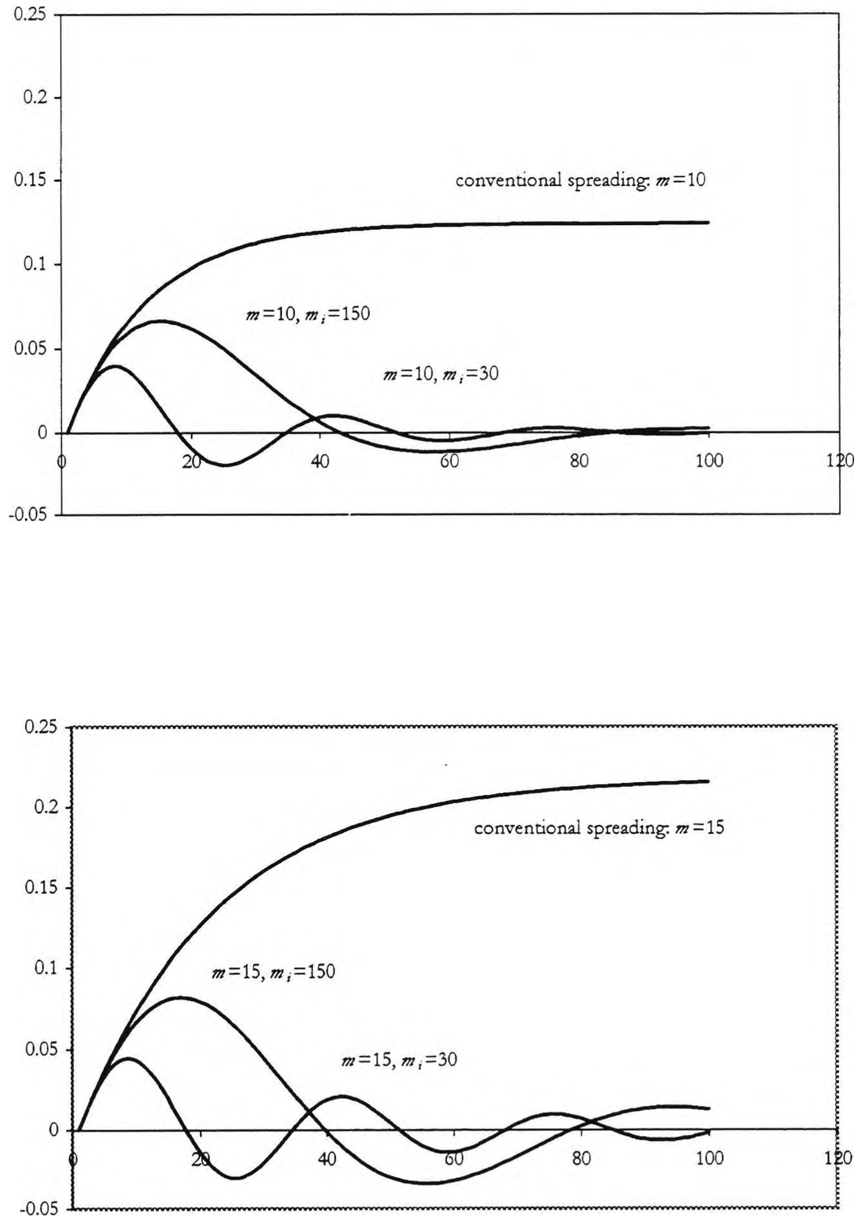


Figure 5.5: Expected surplus against time for various conventional spreading periods m and integral spreading periods m_i. $i_v = 3\%$, $i = 4\%$, $AL = 1$.

Suppose $\Delta > 0$. *Then it is more efficient to use integral, rather than conventional, spreading, provided* $m_i > m_i^f$, *where*

$$m_i^f = \left\{(k-d)q[\sigma^2 v^2 (k-d_v)^2 + (d-d_v)^2(1-q(1-k)^2)]\right\} \Big/ \left\{2(d-d_v)[1-q(1-k)^2][\sigma^2 v^2 (k-d_v+k-d) + (d-d_v)(1-q(1-k)^2)]\right\}. \quad (5.72)$$

For $\Delta \geq 0$, *maximum efficiency (i.e. minimum long-term mean square unfunded liability) is achieved with conventional spreading when* $\Delta = 0$. *The longer* m_i *is, the more efficient integral spreading becomes. As* $m_i \to \infty$, *integral spreading (for any* Δ*) becomes as efficient as conventional spreading as if* $\Delta = 0$.

Proof in Appendix F (§F.4). Remarks on Proposition 5.4:

1. Proposition 5.4 may be interpreted as follows: if there is no prudence margin, smaller long-term surpluses will emerge when conventional spreading is used as opposed to integral spreading; but if a margin does exist, then it is more efficient to use integral spreading, as long as $m_i > m_i^f$.

2. The last part of Proposition 5.4 indicates that a long integral spreading period m_i is more efficient in the sense that it leads to a smaller mean square surplus in the long-term. A very long m_i would of course mean that surpluses are only removed very slowly. A balance must be struck between the pace of funding in the short term and variability in the funding process in the long term.

In order that smaller permanent surpluses and deficits emerge when integral spreading is used ($k_i \neq 0$) as compared to conventional spreading ($k_i = 0$), the parameter m_i should be *greater* than m_i^f. m_i^f is listed in Table 5.4 on the following page for various valuation discount rates (net of salary inflation) i_v, mean real return i (net of salary inflation) and spread period m. It is assumed that the standard deviation of the rate of return is $\sigma = 0.1$.

The following observations are based on Table 5.4 on the next page:

1. It appears that as the prudence margin between the valuation discount rate and the long-term mean rate of return decreases, m_i^f decreases and then increases. For most practical values, Table 5.4 shows that the smaller the prudence margin is, the longer m_i must be in order for integral spreading to be more effective at removing surpluses than conventional spreading. If there is only a small prudence margin, long-term persisting surpluses will be small and it may not be worthwhile to use integral spreading of gains and losses.

2. Very long integral spreading periods are required if current unfunded liabilities are spread forward over short terms (e.g. $m = 1$). It was shown in §5.3 (and it is well-known to several authors including Trowbridge (1952), Dufresne (1986) and Thornton

i_v	$i - i_v$	$m = 1$	3	5	10	15	20	25	30	40	50
1%	0.00%	†	†	†	†	†	†	†	†	†	†
	0.25%	92.04	17.15	9.46	4.54	3.04	2.32	1.91	1.65	1.37	1.26
	0.50%	41.67	7.43	4.11	2.02	1.39	1.10	0.94	0.84	0.76	0.80
	1.00%	17.62	3.01	1.71	0.90	0.67	0.57	0.52	0.51	0.59	1.70
	1.50%	10.26	1.75	1.03	0.60	0.48	0.43	0.43	0.46	1.39	
3%	0.00%	†	†	†	†	†	†	†	†	†	†
	0.25%	95.47	17.34	9.38	4.29	2.76	2.04	1.64	1.41	1.26	1.91
	0.50%	43.14	7.49	4.06	1.91	1.28	0.99	0.84	0.77	0.90	
	1.00%	18.19	3.03	1.69	0.86	0.63	0.54	0.51	0.56		
	1.50%	10.57	1.76	1.02	0.58	0.46	0.43	0.47	0.87		
5%	0.00%	†	†	†	†	†	†	†	†	†	†
	0.50%	44.63	7.55	4.02	1.81	1.17	0.90	0.78	0.80		
	1.00%	18.77	3.04	1.66	0.83	0.60	0.52	0.56	1.38		
	2.00%	7.29	1.22	0.73	0.45	0.40	0.47				
	3.00%	4.10	0.76	0.51	0.37	0.38					
10%	0.00%	†	†	†	†	†	†	†	†		
	0.50%	48.43	7.68	3.90	1.58	0.96	0.78	1.38			
	1.00%	20.23	3.07	1.61	0.75	0.54	0.65				
	2.00%	7.79	1.23	0.72	0.44	0.44					
	3.00%	4.36	0.77	0.50	0.37	1.26					

Table 5.4: m_i^f for various choices of $\{i_v,\ i,\ m\}$ and $\sigma = 0.1$. $k_p = 1/\ddot{a}_{\overline{m|}}$, where the annuity is calculated at i_v. $k_i = 1/m_i$. Blanks indicate that stability conditions do not hold. (†) indicates that it is more efficient to use conventional spreading of surpluses and deficits (i.e. $k_i = 0$).

& Wilson (1992a), although often only in a deterministic context) that shorter spreading periods lead to smaller emergent surpluses in the long term. This is also graphically depicted in Figure 5.3 on page 180 (cf. histograms A and D). It may therefore not be worthwhile to introduce integral spreading of gains and losses when short spreading periods m are being used. Nevertheless, note from Table 5.4 on the page before that $m_i > 20$ is efficient in most cases and will reduce long-term surpluses as compared to conventional spreading.

5.6.5 Effect on Contribution Level

Persisting surpluses or deficits are also accompanied by contributions that are persistently different from the standard contribution or normal cost. Most plan sponsors require long-term stability in their pension fund contributions (§2.2.4). Contribution stability is not only a long-term funding objective, but also a short to medium-term 'investment' objective, according to a survey by Ferris (1997). She states that contribution stability is perhaps more important to sponsors as shorter spreading and amortization periods are now in use, as a result of various funding requirements. The oscillations in the first moments referred to in §5.6.4 also occur in the average contribution (as shown by Loades (1998)) and may not be desirable.

It is more important at this juncture to consider the long-term stochastic variation in the contribution adjustment or supplementary contributions required from the sponsor. This may be measured in terms of the long-term mean-square supplementary contribution. Since contributions may be targeted, a two-sided measure is acceptable.

PROPOSITION 5.5 *Let $\Delta = d - d_v$ represent the prudence margin between the valuation discount rate and the long-term mean rate of return.*

Suppose $\Delta = 0$. Then it is more efficient, in terms of minimising long-term mean square supplementary contribution, to use conventional, rather than integral, spreading.

Suppose $\Delta > 0$. Then it is more efficient to use integral, rather than conventional, spreading, provided $m_i > m_i^c$, where $m_i^c > m_i^f$ (in Proposition 5.4) and

$$m_i^c = \left\{(k-d)q\left[\sigma^2 v^2 (k-d_v)^2 k^2 + (d-d_v)^2(1-q(1-k)^2)d(2k-d)\right] + \sigma^2 v^2 (k-d)^2(2+k-d)[1-q(1-k)^2]\right\} \Big/ \left\{2(d-d_v)[1-q(1-k)^2] \left[\sigma^2 v^2 (k-d_v+k-d)k^2 + (d-d_v)(1-q(1-k)^2)d(2k-d)\right]\right\}. \quad (5.73)$$

For $\Delta \geq 0$, maximum efficiency (i.e. minimum long-term mean square supplementary contribution) is achieved with conventional spreading when $\Delta = 0$. The longer m_i is, the more efficient integral spreading becomes.

Proof in Appendix F (§F.5). Remarks on Proposition 5.5:

1. Suppose there is no prudence margin. Smaller supplementary contributions will be required from the sponsor if conventional as opposed to integral spreading is used. If there is a margin, it is more efficient to use integral spreading, with $m_i > m_i^c$.

2. A long integral spreading period m_i is more efficient in the sense that it leads to smaller and less variable supplementary contributions in the long term. Again, this must be traded off against a slower pace of funding.

In order for smaller supplementary contributions to be required from the sponsor when integral spreading is used ($k_i \neq 0$) as compared to conventional spreading ($k_i = 0$), the parameter m_i should be *greater* than m_i^c. m_i^c is listed in Table 5.5 on the following page for various valuation discount rates (net of salary inflation) i_v, mean rate of return i (net of salary inflation) and spread period m. The standard deviation of the rate of return is $\sigma = 0.1$.

The following observations pertain to Table 5.5 on the next page:

1. The smaller the prudence margin between the valuation discount rate and the long-term mean rate of return, the longer m_i must be in order that integral spreading leads to smaller supplementary contributions than conventional spreading. When a small prudence margin exists, small gains and losses emerge and contributions do not need to be heavily adjusted. Integral spreading of gains and losses may not bring more advantages than conventional spreading.

2. Except for very large margins, a long integral spreading period $m_i > 100$ seems to be required for similar variability of supplementary contributions when conventional and integral spreading are used. Long m_i mean that persisting surpluses will be removed very slowly.

5.6.6 Efficiency

Proposition 5.6 follows directly from Propositions 5.4 and 5.5.

PROPOSITION 5.6 *Suppose there is a prudence margin ($i > i_v$) and the current unfunded liability (with or without separate fixed-term amortization of the initial unfunded liability) is being spread over some (stable) term m.*

1. *In terms of minimising the mean square surplus in the limit, it is more efficient to add integral spreading, with $m_i > m_i^f$, than to use conventional spreading over m alone.*

2. *In terms of minimising mean square surpluses* as well as *mean square supplementary contributions in the limit, it is more efficient to add integral spreading, with $m_i > m_i^c$, than to use conventional spreading over m alone.*

i_v	$i - i_v$	$m = 1$	3	5	10	15	20	25	30	40	50
1%	0.00%	†	†	†	†	†	†	†	†	†	†
	0.25%	404	688	1011	1701	2236	2643	2946	3164	3398	3437
	0.50%	201	338	480	768	967	1100	1183	1227	1232	1157
	1.00%	99	155	211	305	352	371	372	359	306	227
	1.50%	65	94	122	158	169	166	154	137	87	
3%	0.00%	†	†	†	†	†	†	†	†	†	†
	0.25%	410	645	882	1254	1410	1435	1382	1282	1021	747
	0.50%	203	307	406	542	580	567	524	465	329	
	1.00%	99	138	171	203	200	181	153	121		
	1.50%	65	82	95	103	94	78	59	38		
5%	0.00%	†	†	†	†	†	†	†	†	†	†
	0.50%	205	284	349	399	37	315	251	188		
	1.00%	100	124	142	145	124	95	66	39		
	2.00%	47	47	48	41	30	17				
	3.00%	29	24	23	17	10					
10%	0.00%	†	†	†	†	†	†	†	†		
	0.50%	211	238	251	206	139	82	41			
	1.00%	100	99	96	71	43	20				
	2.00%	45	35	31	19	9					
	3.00%	27	18	15	8	3					

Table 5.5: m_i^c for various choices of $\{i_v, i, m\}$ and $\sigma = 0.1$. $k_p = 1/\ddot{a}_{\overline{m|}}$, where the annuity is calculated at i_v. $k_i = 1/m_i$. Blanks indicate that stability conditions do not hold. (†) indicates that it is more efficient to use conventional spreading of surpluses and deficits (i.e. $k_i = 0$).

If there is no prudence margin ($i = i_v$), it is more efficient, in terms of minimising both mean square surpluses and mean square supplementary contributions in the limit, to use conventional spreading over some stable m than to use integral spreading, .

The results of §§5.6.4 and 5.6.5 show that if there is a prudence margin ($i > i_v$), both the mean square unfunded liability and the mean square supplementary contribution are reduced when integral spreading is used and $m_i > m_i^c$. Both security and stability may therefore be improved by the introduction of integral spreading. When there is no prudence margin, integral spreading leads to higher mean square unfunded liability and mean square supplementary contribution than conventional spreading and it is not therefore efficient.

It is important to note that Proposition 5.6 takes no account of the 'pace' of funding. (The 'optimal' range of spread periods of Dufresne (1986, 1988) also ignores intermediate or transient funding objectives.) It may require a long integral spreading period m_i for integral spreading to be more efficient than conventional spreading, in particular in terms of variability in supplementary contributions (as shown in Table 5.5 on the preceding page). In practice some improvement in security may be traded off against higher contribution variability by choosing some $m_i \in [m_i^f, m_i^c]$.

Table 5.6 on the next page shows the amelioration in security and stability, as a percentage reduction in the root mean square of unfunded liability and supplementary contribution respectively. A negative reduction represents a deterioration in security and stability. (The valuation discount rate (net of salary inflation) i_v is assumed to be 3% and the mean of the real (net of salary inflation) rate of return is 4, 5%. The standard deviation of the rate of return is taken to be $\sigma = 0.1$.)

1. Suppose that there is a prudence margin of about 2% between the valuation discount rate and the long-term mean rate of return on the fund. If pension fund contributions are adjusted by about 0.67% of cumulative unfunded liabilities ($m_i = 150$) in addition to a typical 8% of the current unfunded liability ($m = 15$), then smaller surpluses will emerge eventually: the dispersion of fund levels about the actuarial liability reduces by about 52.4%. Smaller supplementary contributions will also be required from the sponsor eventually: the dispersion of contributions about the standard contribution or normal cost decreases by about 26.7%.

2. If the prudence margin is only of the order of 1% and $m_i = 150$ then an improvement in security of about 27.1% is obtained at the expense of a deterioration in stability of some 6%. When $m_i \geq 200$, both security and stability improve with the introduction of integral spreading. In general, for small prudence margins, long integral spreading periods m_i must be used to avoid a significant deterioration in the stability of contributions.

3. Note also from Table 5.6 that the reduction in root mean square surplus is not sensitive to the choice of $m_i > 50$ as m_i^f is generally much less than 50 (see Table 5.4 on page 199).

				$i - i_v = 1\%$		$i - i_v$=2%	
m	k	m_i	k_i	(1)	(2)	(1)	(2)
5	21.2%	50	2%	8.9%	-13.5%	22.4%	-2.4%
		100	1%	9.0%	-4.1%	22.5%	5.1%
		150	0.67%	9.1%	-0.8%	22.5%	7.7%
		200	0.5%	9.1%	0.9%	22.5%	9.0%
10	11.4%	50	2%	18.1%	-34.7%	39.2%	-4.6%
		100	1%	18.2%	-12.8%	39.3%	10.7%
		150	0.67%	18.3%	-4.6%	39.3%	16.4%
		200	0.5%	18.3%	-0.2%	39.4%	19.4%
15	8.1%	50	2%	27.0%	-49.6%	52.3%	-0.3%
		100	1%	27.1%	-18.8%	52.4%	19.1%
		150	0.67%	27.1%	-6.6%	52.4%	26.7%
		200	0.5%	27.1%	0.0%	52.4%	30.8%
20	6.5%	50	2%	35.6%	-56.0%	63.6%	10.9%
		100	1%	35.7%	-20.0%	63.6%	30.8%
		150	0.67%	35.7%	-5.4%	63.6%	38.9%
		200	0.5%	35.7%	2.7%	63.7%	43.3%

Table 5.6: Percentage *reduction* in root mean square of (1) surplus and (2) supplementary contribution when integral spreading is used instead of conventional spreading, for various $\{i,\ m,\ m_i\}$ with $i_v = 3\%$ and $\sigma = 0.1$. $k = 1/\ddot{a}_{\overline{m|}}$, where the annuity is calculated at i_v. $k_i = 1/m_i$.

5.7 Summary

This section summarises some of the major points made in this chapter. Actuarial prudence is a major feature of the actuarial valuation of pension plans. Prudence is particularly necessary when economic factors are estimated because pension funds often operate in volatile economic conditions. The level of prudence in the discount rate when valuing liabilities allows a degree of actuarial control on pension funding (§5.1). The meaning of the valuation discount rate depends on the method of valuation. In the market method, the valuation discount rate is the market discount rate implied in the assets that hedge the pension liabilities. In the Discounted Cash Flow method, it is the estimated rate of return on reinvestment in a notional portfolio of assets that matches the liabilities of the pension plan. The valuation discount rate is conceptually different from the estimated return on the actual and future asset portfolio of the fund. The numerical difference between the valuation discount rate and the best-estimate investment return assumption is referred to as a *prudence margin* and the valuation discount rate is sometimes called a *prudent* valuation rate of interest assumption.

Various factors influence the size of the prudence margin (§5.2):

1. A major determinant is the degree of mismatch between pension fund assets and liabilities. A perfect hedging or matching investment policy is rarely followed. There is a lack of consensus on how to hedge salary-related liabilities, i.e. the pension liability for active members. This has implications for the choice of a suitable discount rate for the valuation of such liabilities. Some actuaries believe that for the purpose of setting contributions, a prudent estimate of the equity risk premium may be incorporated in the discount rate.

2. The liability discount rate is also adjusted to reflect the risk in liability cash flows, both in their amount and timing. The risk of insolvency, both for the pension plan and for the sponsor, should also be taken into account.

3. Margins may also be included as an actuarial control parameter, specially to establish the pace of funding for benefits. The size of the margin may be varied to smooth contributions and stabilise them over the long term in spite of short-term fluctuations in experience. It is believed that margins do not have a direct effect on the ultimate cost of pension provision.

4. A major advantage of advance funding for retirement benefits is that it affords some short-range flexibility to the plan sponsor in terms of the timing of his contributions. It appears that conservative margins are allowed if it is to the advantage of the sponsor to transfer resources to the pension fund. Actuaries may also vary margins (within limits) and hence the pace of funding in order to suit the cash flow needs of the sponsor. This appears to agree with aspects of the theory of corporate finance.

5. The choice of the valuation discount rate is not always very precise. Various approximations are made when choosing a notional hedge or matching portfolio, when averaging over term-dependent discount rates or when averaging over time to avoid unnecessary fluctuations in the valuation result.

The effect of prudence margins in the valuation discount rate is to create persisting surpluses in the long term (§5.3). Large and permanent surpluses may have a deleterious effect on pension funding: plans may be penalised for breaching maximum surplus regulations (full funding limits) by revenue authorities; large surpluses imply that resources are diverted from productive activity and corporate profits and eventually employment in the firm may be affected. The size of such surpluses is seen to depend not only on the size of prudence margins in the valuation discount rate, but also on the method of amortizing gains and losses and particularly on the term over which they are liquidated.

Three methods of removing persisting surpluses while retaining a prudent funding objective are described and their properties analysed. A practical method is to spread surpluses over a shorter period than deficits. This is investigated by means of simple stochastic simulations (§5.4):

1. Undesirable long-term surpluses appear to be removed when this method is applied, although the incidence of deficits may increase.

2. The pattern of contributions generated may not be smooth because gains and losses are not treated identically.

3. It is not clear how the pace of funding is affected if two different spreading periods are used.

4. It is not clear how the different spreading periods for surpluses and deficits should be chosen.

Another method, promoted for compliance with accounting requirements, involves making an estimate of the long-term rate of return on assets, as well as making an assumption regarding the liability discount rate (§5.5). Recommended contributions are then adjusted to take into account the difference between these two assumptions.

1. The first and second moments of the pension funding process in the limit are obtained under the assumption that rates of investment return are independent and identically distributed over time.

2. Some stability conditions constraining the choice of spreading period as well as another parameter are also obtained. An 'optimal' range of spreading period, similar to the one described by Dufresne (1986, 1988), may be derived.

3. The 'dual-interest' method implied by the U.S. Financial Accounting Standards No. 87 (FAS87) is shown to defray deficits and surpluses faster than the comparable 'dual-interest' method of Thornton & Wilson (1992a:§6.5), but the latter leads to less variable surpluses and deficits.

4. It is found that the method is successful at removing extraneous surpluses if an unbiased estimate of the rate of return on the future asset portfolio of the fund can be made. It is argued that much less confidence can be attached to the estimate of future investment returns than to the liability valuation discount rate (both net of price and salary inflation) as the strategic asset portfolio may be mismatched to the liabilities so that returns tend to be volatile.

The final method of avoiding undesirable surpluses does not require an assumption about the future return on pension fund assets and involves a cumulative gain/loss adjustment to the standard contribution or normal cost (§5.6). In addition to spreading the current unfunded liability forward over a rolling term m, cumulative or integral unfunded liabilities are spread over a (different) term m_i.

1. The first and second moments of the pension funding process are derived when this method is applied. An important and restrictive assumption that is made in the analysis is that pension fund investment returns perform a random walk.

2. It is shown that no surplus or deficit emerges in the long term on average, even if there is a margin between the valuation discount rate and the mean rate of return on the assets, without any actuarial assumption being required as to the latter.

3. When integral spreading is used, the contribution that is eventually required from the sponsor is on average closer to the standard contribution or normal cost than with conventional spreading.

4. A wide range of conventional spreading periods m and integral spreading periods m_i are shown to exist for which the first and second moments of the pension funding process are stable. The stability constraints on m and m_i are independent of the valuation discount rate assumption (except that when surpluses and deficits are spread, an annuity factor that is calculated at the valuation discount rate is used) and depend only on the moments of the rate of investment return process.

5. It may be useful, in terms of controlling the pace of funding and avoiding large cyclic disturbances in the funding process, to 'freeze' and amortize initial unfunded liabilities separately over a fixed term. m and m_i must be chosen in combination to achieve the requisite pace of funding. It appears that shorter m and/or m_i hasten the pace of funding on average.

6. Security is an important long-term funding objective in pension planning. If there is a prudence margin between the valuation discount rate and the mean rate of return, the mean square unfunded liability in the limit is minimised when an integral spreading period longer than a certain minimum (m_i^f) is used. The longer the integral spreading period that is used, the smaller the mean square surplus that emerges. Numerical experiments indicate that the minimum m_i^f is fairly short except for very short conventional spread periods m and very small margins: if there is a prudence margin, an improvement in security will usually follow when integral spreading is used.

7. Contribution stability is also an important pension funding objective. It is also shown that if there is a prudence margin, it is more efficient to employ integral spreading (with an integral spreading period longer than some minimum m_i^c) as this results in a lower mean square supplementary contribution being required from the sponsor in the limit.

The longer the integral spreading period that is used, the smaller the mean square supplementary contribution that is required from the sponsor. Numerical experiments show that long integral spreading periods may be required particularly for very short conventional spread periods m and very small prudence margins. High values of m_i may mean that it takes a very long time for surpluses to be removed. In practice, if there is a prudence margin and integral spreading is used, an amelioration in security may have to be traded off against a deterioration in stability.

8. If there is no prudence margin between the valuation discount rate and the long-term mean rate of return, however, it is more efficient to use conventional spreading as this yields a lower mean square surplus as well as a lower mean square supplementary contribution.

Chapter 6

Conclusions

6.1 Summary

This section summarises some of the major points made in this thesis. The aim of the thesis is to investigate the dynamics and actuarial control of defined benefit pension funds.

Chapter 2. The objectives of funding in advance for retirement benefits are considered and the security of benefits and the stability and flexibility of contributions are emphasised. The pension fund is viewed as a system, and actuarial involvement, through regular valuations, is likened to an exercise in the control of the pension system to achieve these objectives. Control is complicated by the uncertain economic and demographic environment in which pension plans operate. Some of the factors that affect final-salary pension funding are discussed in some detail. A simple mathematical model is introduced and described with a view to investigating aspects of the dynamics of funding in the later chapters of the thesis.

Chapter 3. Two aspects of the actuarial funding policy or control are then examined: the determination of a suitable contribution and asset allocation. Pension funding methods—in particular methods that systematically liquidate deviations when economic and demographic experience differs from actuarial valuation assumptions—are discussed. It has been shown that when such deviations are being spread forward over a rolling term and rates of investment return are modelled as simple stochastic processes, a certain range of spreading periods is efficient in terms of achieving a tradeoff between security and stability. The robustness of the efficient range is investigated when different economic and demographic experience is projected, namely when rates of return are general autoregressive processes, when new entrants into the plan vary randomly, and when other stochastic perturbations, such as discretionary contributions, are allowed.

It is proven that there also exists an efficient range of periods over which actuarial gains and losses can be directly amortized. The usual range of amortization periods between 1 and

5 years, as typically used in North America, is found to be within the efficient range. It is shown that the amortization of gains and losses over a fixed term yields greater security than when gains and losses are indirectly spread forward over a rolling term (i.e. when surpluses and deficits are spread). It is nevertheless more efficient to spread surpluses and deficits than to directly amortize gains and losses because there always exists a spreading period such that better security and stability are obtained.

Minimum-risk optimal asset allocation and contribution policies are derived, based on minimising the discounted quadratic disutility to the plan members and sponsor of not achieving security and stability objectives. The optimal contribution control resembles the proportional spreading of surpluses and deficits over a moving term, while the optimal asset allocation dynamically hedges against the risks of fund inadequacy and contribution instability.

Chapter 4. The methodology of actuarial pension plan valuations is discussed. Two distinct market-oriented and cash flow-oriented methods are identified. Both are found to rely on a notional or hypothetical portfolio that matches or hedges the pension liabilities. Some practical issues concerning investment strategy are also considered. Valuation and asset allocation are observed to be related through the concept of hedging or matching liabilities.

Various pension fund asset valuation methods and their properties, in terms of consistency with liability valuations, smoothness, objectivity and dynamics, are briefly discussed. One particular market-related method is analysed in detail. The actuarial asset value is one that recognises a fraction of the unanticipated difference between market and actuarial asset value, or alternatively it is a weighted average of the market value of assets and their actuarial value as anticipated based on the actuarial valuation assumptions. Under this method, symmetry between asset gain/loss amortization and asset valuation is explicitly demonstrated: asset gain/loss amortization and asset valuation have a complementary function in achieving smoothness in the funding process. When gains and losses are paid off immediately, it is shown that it is efficient to weight the current market value upwards of a certain minimum. This range is directly comparable with the efficient range of spreading periods. Smoothing and spreading, separately and together, are found to improve contribution stability, but only up to a point. Numerical experiments show that there are efficient ranges of spreading periods, for a given degree of asset value smoothing, and corresponding and symmetrical efficient ranges of smoothing parameters for a given spreading period. It appears that the efficient weight to be placed on current market value of assets should be upwards of 60%.

Chapter 5. The last aspect of pension plan valuations that is considered concerns actuarial prudence in the determination of liabilities, and in particular prudence margins in the valuation discount rate. The extent of prudence in the discount rate when valuing liabilities allows a degree of actuarial control on pension funding. The valuation discount rate is the market discount rate implied in the assets that hedge the pension liabilities; alternatively, it is the estimated rate of return on reinvestment in a portfolio of assets that matches the liabilities of the pension plan. The valuation discount rate is different from the best-estimate rate of return assumption. This difference is referred to as a 'prudence' margin. The prudence margin depends primarily on the degree of asset-liability mismatch in the pension fund as well as the risk in the amount and timing of liability cash flows. But various other factors appear to influence the size of the 'prudence' margin. Margins are used as a control parameter to establish the pace of funding for benefits, and also to provide some flexibility to the plan sponsor in terms of the timing of his contributions, possibly to suit his cash flow needs, as predicted by corporate finance theory.

The effect of excessive prudence (or 'conservatism') is to create persisting surpluses in the long term. These surpluses have an opportunity cost for the sponsoring firm and may also be expensive if full funding limits (or maximum surplus regulations) are breached. A few methods of avoiding undesirable and volatile surpluses and variable sponsor contributions, while retaining prudent funding objectives, are considered. The first method calls for the defrayal of surpluses over a shorter term than deficits. This is used in practice but stochastic simulations show that the pattern of contributions required from the sponsor is not smooth and in addition it is not clear over what periods surpluses and deficits ought to be spread and how this affects the 'pace' of funding. A second, 'dual-interest', method is promoted for compliance with accounting requirements and originates from the U.S. Financial Accounting Standards No. 87 (FAS87). It is shown to be successful at removing surpluses if an unbiased estimate of the rate of return on the future asset portfolio can be made. But asset returns are likely to be volatile if the pension fund assets are not matched to the liabilities, as a matter of investment policy. The final method does not require any assumption about the future return on plan assets and involves the cumulative or integral spreading of unfunded liabilities. Some suitable parameters to achieve security and stability objectives more efficiently are considered. Substantial reductions in the size and volatility of surpluses and supplementary contributions are possible if this method is applied when typical spreading periods of 10–15 years and prudence margins of 1–2% are used.

6.2 Future Developments

There are various areas in which the analysis undertaken in this thesis may be extended and improved.

Inflation. Price and salary inflation are central economic variables in final-salary defined benefit pension funding. In this thesis, benefits are assumed to be indexed with salary inflation and asset returns net of salary inflation are projected. (Alternatively, like Dufresne (1986), it may be assumed that there is no inflation on salaries and nominal investment returns are then projected.) This is not realistic. Stochastic modelling for salary inflation is necessary.

Benefit indexation is an important issue and price inflation needs to be modelled. Serial and contemporaneous correlation between price and salary inflation must be considered. The linkage between equity returns and salary inflation must also be investigated.

Investment Returns. Very simple models are employed to model the return from investment of the pension fund in this thesis. More 'realistic' modelling is necessary to test for the robustness of the simple conclusions obtained. The model of Wilkie (1995) is a possible candidate for such work. More complicated asset models have been proposed which include ARCH effects and VAR methods.

Stochastic Asset-liability Modelling. The analysis undertaken in this thesis is based on a simplified theoretical pension plan model. The conclusions may nevertheless have a direct practical value. The theoretical result of Dufresne (1986, 1988) regarding an efficient period over which to spread surpluses and deficits appears to have been reproduced in more realistic simulations using a more complicated asset and liability model (Haberman & Smith, 1997). One obvious area of work is to test some of the conclusions of this thesis by employing simulation methods.

Optimal Asset Allocation. Optimal asset allocation strategies may be investigated using several contemporaneously correlated assets, in an infinite horizon stationary context. Discontinuous utility functions, approximating full funding limits and solvency requirements, as well as non-negative asset holding constraints may be required. The robustness of optimal solutions under dependent rates of return may need to be investigated. Finally, realistic delays and random 'measurement' errors in the valuation process should be allowed for.

Asset Valuation. Various asset valuation methods appear to be used in practice and may be investigated in combination with different asset gain/loss amortization methods. One example is arithmetic averaging with consideration of cash flows, as opposed to exponential smoothing. The effect of dependent rates of return, stochastic inflation and a prudent or dynamic valuation basis must be considered. A mathematical comparison of market-related methods with the discounted income asset valuation method may be attempted. The effect of the corridor imposed by the U.S. Employee Retirement Income Security Act, 1974 (ERISA) on actuarial asset values is also of crucial practical importance.

Valuation Methodology. If a market valuation methodology is employed, then market interest rates are required to value pension liabilities. This will require the use of a model for the yield curve. Several such models exist in the Financial Economics literature. The models of Dyson & Exley (1995) and Smith (1996) may be simple enough for mathematical tractability.

Plan Membership. The plan population model that is considered in this thesis is extremely simplified. Further work on random new entrants, with multiple entry and exit ages, deferred benefits and redundancy (withdrawal) is also possible.

Financial Reporting. Finally, this thesis is concerned only with the *funding* of retirement benefits. A mathematical approach may also be used for the development of pension costs, under various accounting standards (including the new international standards). Dufresne (1993, 1994) examines various aspects of the U.S. Financial Accounting Standards No. 87 (FAS87), in particular the treatment of accounting gains and losses and the use of a market discount rate, using both a mathematical and a simulation-based approach.

Appendix A

Stochastic Investment Return, Membership and Other Perturbations

A.1 Stationary Autoregressive Rate of Return

Suppose that the logarithmic or geometric rate of return process $\delta(t)$ is a stationary Gaussian autoregressive process of order p or AR(p), with $p < \infty$, as in equation (3.86). The moments of such a process are well-known [Box *et al.* (1994:54)]. Let $\zeta(t) = \delta(t) - \delta$.

$$\zeta(t+1) = \varphi_1\zeta(t) + \varphi_2\zeta(t-1) + \cdots + \varphi_p\zeta(t-p+1) + e(t+1) \tag{A.1}$$

Since $\delta(t)$ is stationary (unconditional), $E\zeta(t) = 0$ and $E\delta(t) = \delta \ \forall t$. Let $\rho_k = E[\zeta(t)\zeta(t-k)]/E\zeta(t)^2$ be the autocorrelation function. ρ_k obeys a pth order linear difference equation

$$\rho_{k+1} = \varphi_1\rho_k + \varphi_2\rho_{k-1} + \cdots + \varphi_p\rho_{k-p+1}. \tag{A.2}$$

(This follows if equation (A.1) is multiplied by $\zeta(t-k)$, and mathematical expectation is taken. Note that $e(t+1)$ is uncorrelated with $\zeta(t-k)$ and that $\rho_k = \rho_{-k}$.) Difference equation (A.2) requires p 'initial' conditions, namely $\rho_0 = 1$ (by definition), $\rho_{-1} = \rho_1$, $\rho_{-2} = \rho_2$, ..., $\rho_{1-p} = \rho_{p-1}$. There is a correspondence between $\rho_0, \ldots, \rho_{p-1}$ and $\varphi_1, \ldots, \varphi_p$, through the Yule-Walker equations (see Box *et al.* (1994:57)). Assume a given AR(p) process with known $\{\rho_1, \ldots, \rho_{p-1}\}$ or $\{\varphi_1, \ldots, \varphi_p\}$, or alternatively assume that these parameters have been estimated from statistical data.

A general solution to the homogeneous linear difference equation (A.2) is

$$\rho_k = \sum_{i=1}^{p} A_i G_i^k, \tag{A.3}$$

where $\{G_i\}$ is the set of roots of the characteristic equation

$$z^p - \varphi_1 z^{p-1} - \varphi_2 z^{p-2} - \cdots - \varphi_p = 0 \tag{A.4}$$

and $\{A_i\}$ represents arbitrary constants depending upon the 'initial' values $\{\rho_k\}$, $k \in [1, p-1]$. Letting $\text{Var}\delta(t) = \sigma^2$, for a stationary AR($p$) process,

$$\text{Cov}[\delta(t), \delta(s)] = \sigma^2 \sum_{i=1}^{p} A_i G_i^{t-s}. \tag{A.5}$$

Note that by definition, $\rho_0 = \sum_i A_i = 1$. Since $\delta(t)$ is assumed stationary, $|G_i| < 1$ [Box *et al.* (1994:56)]. It has also been assumed, in the manner of Box *et al.* (1994:60), Bellhouse & Panjer (1981) and Haberman (1994a), that all roots G_i are distinct (real or imaginary). This is likely to be the case if the autoregressive process has been estimated from data.

It is required to find the moments of the accumulation of 1 when the logarithmic 'rate of interest' $\delta(t)$ follows an AR(p) process. Such expressions have been obtained by Boyle (1976) for independent and identically normally distributed $\{\delta(t)\}$, by Panjer & Bellhouse (1980) for stationary AR(1) and AR(2) processes, and by Bellhouse & Panjer (1981) for conditional AR(1) and AR(2) processes.

First, consider the variance of a sum of $\delta(t)$ over a term $t - s$.

$$\begin{aligned}\text{Var}\left[\sum_{u=s+1}^{t} \delta(u)\right] &= 2 \sum_{u=s+1}^{t} \sum_{w=s+1}^{u} \text{Cov}[\delta(u), \delta(w)] - \sum_{u=s+1}^{t} \text{Var}\delta(u) \\ &= 2\sigma^2 \sum_i A_i \sum_{u=s+1}^{t} \sum_{w=s+1}^{u} G_i^{u-w} - \sigma^2(t-s),\end{aligned} \tag{A.6}$$

after substituting equation (A.5). This may be expanded to

$$\begin{aligned}&\sum_i A_i \left[\frac{2\sigma^2}{1-G_i}(t-s) + \frac{2\sigma^2 G_i}{(1-G_i)^2} G_i^{t-s} - \frac{2\sigma^2 G_i}{(1-G_i)^2}\right] - \sigma^2(t-s) \\ &\quad = \sum_i \frac{1+G_i}{1-G_i} A_i \sigma^2 (t-s) + \sum_i \frac{2\sigma^2 G_i A_i}{(1-G_i)^2} G_i^{t-s} - \sum_i \frac{2\sigma^2 G_i A_i}{(1-G_i)^2},\end{aligned} \tag{A.7}$$

where use is made of the fact that $\sum_i A_i = 1$.

The covariance of sums of $\delta(t)$ over different terms may also be found.

$$\text{Cov}\left[\sum_{u=s+1}^{t} \delta(u), \sum_{w=\tau+1}^{t} \delta(w)\right] = \text{Var}\left[\sum_{u=s+1}^{t} \delta(u)\right] + \text{Cov}\left[\sum_{u=s+1}^{t} \delta(u), \sum_{w=\tau+1}^{s} \delta(w)\right], \tag{A.8}$$

($s > \tau$) where the second term on the right hand side is

$$\begin{aligned}\sum_{u=s+1}^{t} \sum_{w=\tau+1}^{s} \text{Cov}[\delta(u), \delta(w)] &= \sigma^2 \sum_i A_i \sum_{u=s+1}^{t} \sum_{w=\tau+1}^{s} G_i^{u-w} \\ &= \sum_i \frac{\sigma^2 A_i G_i}{(1-G_i)^2} [1 - G_i^{t-s} - G_i^{s-\tau} + G_i^{t-\tau}],\end{aligned} \tag{A.9}$$

where equation (A.5) has again been used.

Therefore,

$$
\begin{aligned}
&\mathrm{Var}\left[\sum_{u=s+1}^{t}\delta(u)+\sum_{w=\tau+1}^{t}\delta(w)\right] \\
&=\mathrm{Var}\left[\sum_{u=s+1}^{t}\delta(u)\right]+\mathrm{Var}\left[\sum_{w=\tau+1}^{t}\delta(w)\right]+2\mathrm{Cov}\left[\sum_{u=s+1}^{t}\delta(u),\ \sum_{w=\tau+1}^{t}\delta(w)\right] \\
&=3\mathrm{Var}\left[\sum_{u=s+1}^{t}\delta(u)\right]+\mathrm{Var}\left[\sum_{w=\tau+1}^{t}\delta(w)\right]+2\mathrm{Cov}\left[\sum_{u=s+1}^{t}\delta(u),\ \sum_{w=\tau+1}^{s}\delta(w)\right] \qquad \text{(A.10)}
\end{aligned}
$$

and using equations (A.7) and (A.9), this may be simplified to

$$
\sigma^2[3(t-s)+(t-\tau)]\sum_{i=1}^{p}\frac{1+G_i}{1-G_i}A_i-\sum_{i=1}^{p}\frac{2\sigma^2A_iG_i}{(1-G_i)^2}[3-2G_i^{t-s}+G_i^{s-\tau}-2G_i^{t-\tau}]. \qquad \text{(A.11)}
$$

$\{\delta(t)\}$ is a Gaussian process. Since

$$
\mathrm{E}\left[\sum_{u=s+1}^{t}\delta(u)+\sum_{w=\tau+1}^{t}\delta(w)\right]=(t-s)\delta+(t-\tau)\delta, \qquad \text{(A.12)}
$$

and given the variance in equation (A.11), we can write

$$
\begin{aligned}
&\mathrm{E}\exp\left[\sum_{u=s+1}^{t}\delta(u)+\sum_{w=\tau+1}^{t}\delta(w)\right] \\
&\qquad=c^{t-\tau}w^{t-s}e^{-3\sum_i z_i}\exp\left\{\sum_i\left[2z_iG_i^{t-s}+2z_iG_i^{t-\tau}-z_iG_i^{s-\tau}\right]\right\}, \qquad \text{(A.13)}
\end{aligned}
$$

where

$$
\begin{aligned}
c&=\exp\left[\delta+\frac{\sigma^2}{2}\sum_i A_i\frac{1+G_i}{1-G_i}\right], &\text{(A.14)}\\
w&=\exp\left[\delta+\frac{3\sigma^2}{2}\sum_i A_i\frac{1+G_i}{1-G_i}\right], &\text{(A.15)}\\
z_i&=\sigma^2A_iG_i(1-G_i)^{-2}. &\text{(A.16)}
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\mathrm{E}\exp\left[\sum_{u=s+1}^{t}\delta(u)\right]&=c^{t-s}e^{-\sum_i z_i}\exp\left[\sum_i z_iG_i^{t-s}\right], &\text{(A.17)}\\
\mathrm{E}\exp\left[2\sum_{u=s+1}^{t}\delta(u)\right]&=(cw)^{t-s}e^{-4\sum_i z_i}\exp\left[4\sum_i z_iG_i^{t-s}\right]. &\text{(A.18)}
\end{aligned}
$$

A.2 Stationary Autoregressive Additive Perturbations

This proof is concerned with the second moments of the pension funding process with an additive stationary autoregressive input $\epsilon(t)$. $f(t)$ and $\epsilon(t)$ are statistically dependent (equation (3.119)). Multiplying both sides of equation (3.119) by $\epsilon(t)$ and taking expectation yields

$$\text{Cov}[f(t), \epsilon(t)] = (1-k)^{-1}AL\sum_{j=0}^{t-1}[u(1-k)]^{t-j}\text{Cov}[\epsilon(t), \epsilon(j)], \tag{A.19}$$

since $\epsilon(t)$, $u(s)$ are independent $\forall\ t,\ s$, and since $\text{E}\epsilon(t) = 0$, $\text{Cov}[f(t), \epsilon(t)] = \text{E}f(t)\epsilon(t)$ and $\text{Cov}[\epsilon(t), \epsilon(j)] = \text{E}\epsilon(t)\epsilon(j)$. The autocovariance of $\epsilon(t)$ may be replaced from equation (3.121) and

$$\text{Cov}[f(t), \epsilon(t)] = (1-k)^{-1}AL\sigma_\epsilon^2\sum_i A_i\sum_{j=0}^{t-1}[u(1-k)G_i]^{t-j}, \tag{A.20}$$

$$\lim_{t\to\infty}\text{Cov}[f(t), \epsilon(t)] = (1-k)^{-1}AL\sigma_\epsilon^2\sum_i u(1-k)A_iG_i/[1-u(1-k)G_i]. \tag{A.21}$$

The limit in equation (A.21) exists provided $|u(1-k)G_i| < 1$. $|G_i| < 1$ and it is sufficient that $|u(1-k)| < 1$ for the limit to exist.

Both sides of equation (3.118) may be squared and it then follows that

$$\begin{aligned}\text{E}f(t+1)^2 = q\big[&(1-k)^2\text{E}f(t)^2 + AL^2(k-d)^2 + AL^2\text{E}\epsilon(t)^2 \\ &+ 2(1-k)(k-d)AL\text{E}f(t) + 2(1-k)AL\text{E}f(t)\epsilon(t)\big],\end{aligned} \tag{A.22}$$

since $u(t+1)$ is independent of $u(s)$ and $\epsilon(s)$ $(s < t+1)$ and $u(t+1)$ and $f(t)$ are independent. Now square both sides of equation (3.128) and multiply both sides by qv^2 (or $(u^2+\sigma^2)/u^2$); deduct the resulting equation from equation (A.22). This yields a difference equation in $\text{Var}f(t)$:

$$\begin{aligned}\text{Var}f(t+1) = \sigma^2v^2[\text{E}f(t+1)]^2 + q\big[&(1-k)^2\text{Var}f(t) + AL^2\sigma_\epsilon^2 \\ &+ 2(1-k)AL\text{Cov}[f(t), \epsilon(t)]\big].\end{aligned} \tag{A.23}$$

The limits as $t \to \infty$ of $\text{E}f(t+1)$ and $\text{Cov}[f(t), \epsilon(t)]$ are given in equations (3.124) and (A.21) respectively. Provided that $q(1-k)^2 < 1$, the limit of $\text{Var}f(t)$ as $t \to \infty$ exists and is

$$\lim_{t\to\infty}\text{Var}f(t)[1-q(1-k)^2] \tag{A.24}$$

$$= \sigma^2v^2AL^2 + q\left[AL^2\sigma_\epsilon^2 + 2AL^2\sigma_\epsilon^2\sum_i\frac{u(1-k)A_iG_i}{1-u(1-k)G_i}\right]$$

$$= \sigma^2v^2AL^2 + \sigma_\epsilon^2AL^2q\sum_i A_i\frac{1+u(1-k)G_i}{1-u(1-k)G_i}, \tag{A.25}$$

where we use $\sum_i A_i = 1$. As for the second moment of the contribution level, $\lim\text{Var}c(t) = k^2\lim\text{Var}f(t)$ from equation (3.44).

A.3 Random New Entrants

Given $g(t+r-x)$ in Projection Assumption 3.5, the mean actuarial liability, from equation (3.20), is

$$AL = \mathrm{E}AL(t) = g \sum_{x=a}^{w} l_x AL_x, \tag{A.26}$$

and its variance is

$$\mathrm{Var}AL(t) = \sigma_g^2 \sum_{x=a}^{w} l_x^2 AL_x^2. \tag{A.27}$$

Likewise from equations (3.21) and (3.22), the mean and variance of the normal cost and benefit outgo are as in equations (A.26) and (A.27), with AL_x being replaced by NC_x and B_x appropriately. $AL = \mathrm{E}AL(t)$, $NC = \mathrm{E}NC(t)$ and $B = \mathrm{E}B(t)$ represent the actuarial liability, normal cost and benefit outgo when the plan population is stationary, with gl_x members aged x at all times t, so that Trowbridge's (1952) equation of equilibrium (3.25) holds.

It is convenient to define

$$\begin{aligned} R_x &= kAL_x + NC_x - B_x, \\ R(t) &= kAL(t) + NC(t) - B(t) \\ &= \sum_{x=a}^{w} g(t+r-x) l_x R_x, \end{aligned} \tag{A.28}$$

and

$$\begin{aligned} P_x &= kAL_x + NC_x, \\ P(t) &= kAL(t) + NC(t) \\ &= \sum_{x=a}^{w} g(t+r-x) l_x P_x. \end{aligned} \tag{A.29}$$

It follows that

$$R = \mathrm{E}R(t) = kAL + NC - B = (k-d)AL, \tag{A.30}$$

$$P = \mathrm{E}P(t) = kAL + NC, \tag{A.31}$$

$$\mathrm{Var}R(t) = \sigma_g^2 \sum_{x=a}^{w} l_x^2 R_x^2, \tag{A.32}$$

$$\mathrm{Var}P(t) = \sigma_g^2 \sum_{x=a}^{w} l_x^2 P_x^2. \tag{A.33}$$

Note also, from equation (A.28), that $R(t)$ is akin to a 'moving average' process.

$$\begin{aligned}\text{Cov}[R(t), R(t-\tau)] &= \sum_{x=a}^{w}\sum_{y=a}^{w} l_x R_x l_y R_y \text{Cov}[g(t+r-x), g(t-\tau+r-y)] \\ &= \begin{cases} 0 & \text{if } \tau > w-a \\ \sum_{y=a}^{w-\tau} l_y l_{y+\tau} R_y R_{y+\tau} \sigma_g^2 & \text{if } 0 \leq \tau \leq w-a, \end{cases}\end{aligned} \tag{A.34}$$

given that $\{g(t+r-x)\}$ are not correlated over time, and taking into account the correlation of $R(t)$ and $R(t-\tau)$ when the lag τ is short such that cohorts indexed $t-x$ and $t-\tau-y$ 'overlap'. Likewise, the cross-covariance between $P(t)$ and $R(t-\tau)$ is:

$$\text{Cov}[P(t), R(t-\tau)] = \begin{cases} 0 & \text{if } \tau > w-a \\ \sum_{y=a}^{w-\tau} l_y l_{y+\tau} R_y P_{y+\tau} \sigma_g^2 & \text{if } 0 \leq \tau \leq w-a. \end{cases} \tag{A.35}$$

Now, given Projection Assumption 3.5, $\{u(t)\}$ and $\{g(s)\}$ are independent $\forall\ t, s$, and consequently $\{u(t)\}$ and $\{R(s)\}$ are independent $\forall\ t, s$. Therefore, from equation (3.142),

$$\text{E}f(t)R(t) = f_0[u(1-k)]^t \text{E}R(t) + (1-k)^{-1}\sum_{j=0}^{t-1}[u(1-k)]^{t-j}\text{E}R(j)R(t), \tag{A.36}$$

while

$$\text{E}f(t)\text{E}R(t) = f_0[u(1-k)]^t \text{E}R(t) + (1-k)^{-1}\sum_{j=0}^{t-1}[u(1-k)]^{t-j}\text{E}R(j)\text{E}R(t), \tag{A.37}$$

from which it follows that

$$\begin{aligned}\text{Cov}[f(t), R(t)] &= (1-k)^{-1}\sum_{j=0}^{t-1}[u(1-k)]^{t-j}\text{Cov}[R(t), R(j)] \\ &= (1-k)^{-1}\sum_{\tau=1}^{t}[u(1-k)]^{\tau}\text{Cov}[R(t), R(t-\tau)].\end{aligned} \tag{A.38}$$

Substituting from equation (A.34), one obtains

$$\begin{aligned}\text{Cov}[f(t), R(t)] &= (1-k)^{-1}\sum_{\tau=1}^{w-a}[u(1-k)]^{\tau}\sum_{y=a}^{w-\tau} l_y l_{y+\tau} R_y R_{y+\tau}\sigma_g^2 \\ &= (1-k)^{-1}\sigma_g^2\sum_{\tau=1}^{w-a}\sum_{y=a}^{w-\tau}[u(1-k)]^{\tau} l_y l_{y+\tau} R_y R_{y+\tau}.\end{aligned} \tag{A.39}$$

Likewise,

$$\text{Cov}[f(t), P(t)] = (1-k)^{-1}\sum_{\tau=1}^{t}[u(1-k)]^{\tau}\text{Cov}[P(t), R(t-\tau)], \tag{A.40}$$

and upon replacing equation (A.35),

$$\text{Cov}[f(t), P(t)] = (1-k)^{-1}\sigma_g^2 \sum_{\tau=1}^{w-a}\sum_{y=a}^{w-\tau}[u(1-k)]^{\tau} l_y l_{y+\tau} R_y P_{y+\tau}. \tag{A.41}$$

First moment results are obvious from equation (3.141), since $u(t+1)$ is independent of both $f(t)$ and $R(t)$:

$$\text{E}f(t+1) = u[(1-k)\text{E}f(t) + \text{E}R(t)], \tag{A.42}$$

and, using equation (A.30), $\text{E}R(t) = (k-d)AL$, so that

$$\lim_{t\to\infty} \text{E}f(t) = u(k-d)AL/(1-u(1-k)) = AL, \tag{A.43}$$

provided $|u(1-k)| < 1$. $\lim \text{E}c(t)$ is straightforward given equation (3.39) and $\text{E}NC(t) = NC$.

When both sides of equation (3.141) are squared, it follows that

$$\text{E}f(t+1)^2 = q\big[(1-k)^2\text{E}f(t)^2 + \text{E}R(t)^2 + 2(1-k)\text{E}f(t)R(t)\big], \tag{A.44}$$

since $u(t+1)$ is independent of $u(s)$ and $R(s)$ $(s < t+1)$ and $u(t+1)$ and $f(t)$ are independent. Now square both sides of equation (A.42) and multiply both sides by qv^2 (or $(u^2+\sigma^2)/u^2$); deduct the resulting equation from equation (A.44). This yields a difference equation in $\text{Var}f(t)$:

$$\begin{aligned}\text{Var}f(t+1) = \sigma^2 v^2[\text{E}f(t+1)]^2 + q\big[(1-k)^2\text{Var}f(t) + \text{Var}R(t) \\ + 2(1-k)\text{Cov}[f(t), R(t)]\big].\end{aligned} \tag{A.45}$$

The limit as $t \to \infty$ of $\text{E}f(t+1)$ is given in equation (A.43). From equations (A.32) and (A.39), it may be noted that

$$\text{Var}R(t) + 2(1-k)\text{Cov}[f(t), R(t)] = \sigma_g^2 \sum_{x=a}^{w}\sum_{y=a}^{w}[u(1-k)]^{|x-y|} l_x l_y R_x R_y = S_1(k). \tag{A.46}$$

Then, $\lim \text{Var}f(t)$ in equation (A.45) exists provided $q(1-k)^2 < 1$ and is as in equation (3.148).

Finally, it is clear from equation (3.39) that

$$\text{Var}c(t) = \text{Var}P(t) - 2k\text{Cov}[f(t), P(t)] + k^2\text{Var}f(t), \tag{A.47}$$

and substitution of equations (A.33), (A.41) and (3.148), yields the result of equation (3.149).

The results of Table 3.2 on page 84 and Table 3.3 on page 85 are based on equations (3.148) and (3.149). The following data was employed:

Mortality: English Life Table No. 12.

Ages: Single entry age at 20, single retirement age at 65.

	Unit Credit	*Entry Age*
Mean yearly benefit outgo	39866	39866
Mean actuarial liability per mean benefit outgo	16.94	19.16
Mean normal cost per mean benefit outgo	0.3486	0.2630
R.S.D. of benefit	11.60%	11.60%
R.S.D. of actuarial liability	8.000%	7.633%
R.S.D. of normal cost	8.322%	7.474%

Table A.1: Valuation statistics for the *Unit Credit* and *Entry Age* funding methods with random new entrants. R.S.D. = relative standard deviation.

Investment Returns: $i = 4\%$, $\sigma = 5\%$.

New Entrants: $gl_a = 100$, $\sigma_g l_a = 50$.

Funding Methods: *Unit Credit* and *Entry Age* methods, both with surpluses and deficits being spread forward.

The mean and standard deviation of the actuarial liability generated by the two different funding methods (using equations (A.26) and (A.27)) are shown in Table A.1. Statistics for the normal cost and benefit outgo are also shown. The benefit outgo is independent of the funding method, of course. The *Entry Age* method is more conservative than the *Unit Credit* method as the mean actuarial liability for the former is higher than for the latter. The equation of equilibrium (3.26) of Trowbridge (1952) is easily verified to hold for both methods in terms of the mean values:

$$0.03846 \times 16.94 + 0.3486 = 1.000, \quad 0.03486 \times 19.16 + 0.2630 = 1.000.$$

The lower relative standard deviations for the *Entry Age* method indicate its inherent stability to variation in the number of new entrants, as is well known. (The relative standard deviation of benefit outgo $B(t)$ is $\sqrt{(\mathrm{Var}B(t))}/\mathrm{E}B(t)$.)

Appendix B

Efficient Amortization Methods and Periods

B.1 Proof of Inequality (3.162)

First note that

$$\ddot{a}_{\overline{t-1}|}\ddot{a}_{\overline{t+1}|} \quad < \quad \ddot{a}^2_{\overline{t}|} \tag{B.1}$$

since $\ddot{a}_{\overline{t-1}|}\ddot{a}_{\overline{t+1}|}/\ddot{a}^2_{\overline{t}|} \; - \; 1 \; = \; -v^{t-1}(v-1)^2/(1-v^t)^2 \; < \; 0$

Proof by induction. Let $m = 2$; the left hand side of inequality (3.162) is zero, and the right hand side is positive; hence the proposition is true for $m = 2$. (It can also easily be proven for $m = 3$.)

Assume inequality (3.162) is true for $m = t$, $t \in Z^+$. Then, replace the denominator on the left hand side by $\ddot{a}^2_{\overline{t}|}$ and the one on the right hand side by $\ddot{a}^2_{\overline{t+1}|}$. The inequality is maintained as a result of inequality (B.1) and we have:

$$\left(\ddot{a}^2_{\overline{t-2}|} + \cdots + \ddot{a}^2_{\overline{0}|}\right)/\ddot{a}^2_{\overline{t}|} \quad < \quad \left(\ddot{a}^2_{\overline{t-1}|} + \cdots + \ddot{a}^2_{\overline{0}|}\right)/\ddot{a}^2_{\overline{t+1}|}. \tag{B.2}$$

We can then subtract the following from the left hand side to yield the required result (the inequality is maintained again as a result of inequality (B.1)):

$$\ddot{a}^2_{\overline{t}|}/\ddot{a}^2_{\overline{t+1}|} \quad - \quad \ddot{a}^2_{\overline{t-1}|}/\ddot{a}^2_{\overline{t}|} \qquad (>0). \tag{B.3}$$

Since the result holds for $m = 2$, it must hold for $\{ m \in Z^+ : m \geq 2 \}$. □

B.2 Proof of Inequality (3.163)

Proof by induction. Let $m = 2$ and $(\ddot{a}^2_{\overline{2}|} + \ddot{a}^2_{\overline{1}|})\ddot{a}^2_{\overline{1}|} > \ddot{a}^2_{\overline{2}|}\ddot{a}^2_{\overline{1}|}$ and therefore inequality (3.163) holds for $m = 2$. Suppose it also holds for $m = t$, $t \in \mathrm{Z}^+$. Then, we may expand

$$\left[\ddot{a}^2_{\overline{t+1}|} + (\ddot{a}^2_{\overline{t}|} + \cdots + \ddot{a}^2_{\overline{1}|})\right]\left[\ddot{a}^2_{\overline{t}|} + (\ddot{a}^2_{\overline{t-1}|} + \cdots + \ddot{a}^2_{\overline{1}|})\right] - \left[\ddot{a}_{\overline{t+1}|}\ddot{a}_{\overline{t}|} + (\ddot{a}_{\overline{t}|}\ddot{a}_{\overline{t-1}|} + \cdots + \ddot{a}_{\overline{2}|}\ddot{a}_{\overline{1}|})\right]^2, \tag{B.4}$$

cancel out $\ddot{a}^2_{\overline{t+1}|}\ddot{a}^2_{\overline{t}|}$ and obtain $T_1 + T_2$, where

$$T_1 = (\ddot{a}^2_{\overline{t}|} + \cdots + \ddot{a}^2_{\overline{1}|})(\ddot{a}^2_{\overline{t-1}|} + \cdots + \ddot{a}^2_{\overline{1}|}) - (\ddot{a}_{\overline{t}|}\ddot{a}_{\overline{t-1}|} + \cdots + \ddot{a}_{\overline{2}|}\ddot{a}_{\overline{1}|})^2 \;>\; 0, \tag{B.5}$$

and

$$\begin{aligned} T_2 &= \ddot{a}^2_{\overline{t+1}|}(\ddot{a}^2_{\overline{t-1}|} + \cdots + \ddot{a}^2_{\overline{1}|}) + \ddot{a}^2_{\overline{t}|}(\ddot{a}^2_{\overline{t}|} + \cdots + \ddot{a}^2_{\overline{1}|}) - 2\ddot{a}_{\overline{t+1}|}\ddot{a}_{\overline{t}|}(\ddot{a}_{\overline{t}|}\ddot{a}_{\overline{t-1}|} + \cdots + \ddot{a}_{\overline{2}|}\ddot{a}_{\overline{1}|}), \\ &> \left[\ddot{a}_{\overline{t+1}|}(\ddot{a}^2_{\overline{t}|} + \cdots + \ddot{a}^2_{\overline{1}|})^{1/2} - \ddot{a}_{\overline{t}|}(\ddot{a}^2_{\overline{t+1}|} + \cdots + \ddot{a}^2_{\overline{1}|})^{1/2}\right]^2 \;>\; 0. \end{aligned} \tag{B.6}$$

Therefore $T_1 + T_2 > 0$ and the result holds for $m = t + 1$. Hence, the proposition is true for $\{m \in \mathrm{Z}^+ : m \geq 2\}$. □

B.3 Proof of Proposition 3.2

For $m > 1$, we need to show that $\alpha_a(m) > \alpha_s(m)$. Starting from inequality (3.162) and noting that $u(1-k) = \ddot{a}_{\overline{m-1}|}/\ddot{a}_{\overline{m}|}$, it is easily shown that

$$[1 - u^2(1-k)^2] \sum \lambda_j^2 \;<\; 1 \tag{B.7}$$

$$u^2 \sum \lambda_j^2 \cdot \{1 - (u^2 + \sigma^2)(1-k)^2\} \;<\; u^2 + \sigma^2 - \sigma^2 \sum \lambda_j^2 \tag{B.8}$$

where we have multiplied across by $(u^2 + \sigma^2)$. Rearranging and using equation (3.157), the inequality in Proposition 3.2 is proven.

For $m = 1$, it is easy to show identity of the two methods of amortizing gains/losses and spreading surpluses/deficits (equation (3.154)). □

B.4 Proof of Proposition 3.3

About $\lim \mathrm{Var} f(t)_a$. We must first prove that $\lim \mathrm{Var} f(t)_a$ increases monotonically with the amortization period m:

$$\nabla \lim \mathrm{Var} f(t)_a > 0. \tag{B.9}$$

We may equivalently prove that $\nabla \alpha_a(m) < 0$. Using the customary backward difference operator rules,

$$\nabla \alpha_a(m) = \frac{(\ddot{a}^2_{\overline{m}|} + \cdots + \ddot{a}^2_{\overline{0}|})[\ddot{a}^2_{\overline{m}|} - (1+\sigma^2 v^2)\ddot{a}^2_{\overline{m-1}|}] - [\ddot{a}^2_{\overline{m}|} - \sigma^2 v^2(\ddot{a}^2_{\overline{m-1}|} + \cdots + \ddot{a}^2_{\overline{0}|})]\ddot{a}^2_{\overline{m}|}}{(\ddot{a}^2_{\overline{m}|} + \cdots + \ddot{a}^2_{\overline{0}|})(\ddot{a}^2_{\overline{m-1}|} + \cdots + \ddot{a}^2_{\overline{0}|})}. \tag{B.10}$$

The denominator is positive, and the numerator is proved to be negative by rearranging it into

$$(1+\sigma^2 v^2)\ \{(\ddot{a}^2_{\overline{m-2}|} + \cdots + \ddot{a}^2_{\overline{0}|})\ddot{a}^2_{\overline{m}|} \ - \ (\ddot{a}^2_{\overline{m-1}|} + \cdots + \ddot{a}^2_{\overline{0}|})\ddot{a}^2_{\overline{m-1}|}\} \tag{B.11}$$

and using inequality (3.162).

About maximum spread and amortization periods. The stability conditions on equations (3.58) and (3.155) provide maximum spread and amortization periods, which we shall denote by m_s^∞ and m_a^∞ respectively, for the pension funding process to be stationary in the limit. k_{min} in Result 3.4 is clearly related to m_s^∞. For the same i and σ, the maximum spread period allowable for stability is not greater than the maximum allowable amortization period:

$$m_s^\infty \le m_a^\infty. \tag{B.12}$$

This follows from the fact that the $\lim \mathrm{Var} f(t)_s$ v. m curve has a vertical asymptote at $m = m_s^\infty$ (equation (3.58)), whereas the $\lim \mathrm{Var} f(t)_a$ v. m curve has a vertical asymptote at $m = m_a^\infty$ (equation (3.155)). Proposition 3.2 and inequality (B.9) mean that the latter asymptote must occur at an amortization period not less than the former asymptote.

About $\lim \mathrm{Var} c(t)_a$**.** Finally, it may be proven that the $\lim \mathrm{Var} c(t)_a$ v. m curve has only one turning point, which is a minimum point, at which the $\lim \mathrm{Var} c(t)_s$ v. m curve intersects it:

$$m_s^* \quad < \quad m_a^*, \tag{B.13}$$

$$\lim \mathrm{Var} c(t)_s|_{m=m_s^*} \quad < \quad \lim \mathrm{Var} c(t)_a|_{m=m_a^*}. \tag{B.14}$$

First, note that

$$\begin{aligned} \sum_{i=1}^m \beta_s(i) &= \sum_{i=1}^m \left(\ddot{a}^2_{\overline{i}|} - \ddot{a}^2_{\overline{i-1}|}(1+\sigma^2 v^2)\right), \\ &= \ddot{a}^2_{\overline{m}|} - \sigma^2 v^2(\ddot{a}^2_{\overline{m-1}|} + \cdots + \ddot{a}^2_{\overline{0}|}), \\ &= m.\beta_a(m), \end{aligned} \tag{B.15}$$

$$\beta_a(m) = \sum_{i=1}^m \beta_s(i)/m, \tag{B.16}$$

$$\nabla(m\beta_a(m)) \ = \ \beta_s(m). \tag{B.17}$$

A number of alternative methods can be followed.

1. $\beta_a(m)$ is an average of $\beta_s(i)$ over $i \leq m$. By Result 3.4, $\beta_s(m)$ has a maximum; $\beta_s(m_s^\infty) = \beta_a(m_a^\infty) = 0$ with $m_s^\infty < m_a^\infty$ (inequality (B.12)); $\beta_s(m=1) = \beta_a(m=1)$. As $\beta_s(m)$ increases with m, $\beta_a(m)$ increases, but $\beta_a(m) < \beta_s(m)$. When $\beta_s(m)$ decreases and intersects $\beta_a(m)$, $\beta_a(m)$ then starts decreasing.

   There will only be one maximum in $\beta_a(m)$. It occurs where the two curves intersect and the maximum in $\beta_s(m)$ will occur before the maximum in $\beta_a(m)$ as m increases, i.e. $m_s^* < m_a^*$. $\lim \mathrm{Varc}(t)_s$ and $\lim \mathrm{Varc}(t)_a$ are 'reciprocals' of $\beta_s(m)$ and $\beta_a(m)$ respectively.

2. Differencing equation (B.16),

$$\nabla\beta_a(m) = \frac{m\beta_s(m) - \sum_{i=1}^{m}\beta_s(i)}{m(m-1)}$$

$$= \frac{\beta_s(m) - \beta_a(m)}{m-1} \tag{B.18}$$

$$\nabla^2\beta_a(m) = \frac{\nabla\beta_s(m) - 2\nabla\beta_a(m)}{m-2} \tag{B.19}$$

   For a turning point, $\nabla\beta_a(m) \approx 0$, or $\beta_s(m) \approx \beta_a(m)$. Hence, turning points in $\beta_a(m)$ will only occur where it crosses $\beta_s(m)$. Consider the shape of $\beta_s(m)$; $\beta_s(m_s^\infty) = \beta_a(m_a^\infty) = 0$ with $m_s^\infty < m_a^\infty$ (inequality (B.12)); $\beta_s(m=1) = \beta_a(m=1)$. Hence, $\beta_a(m)$ can only intersect $\beta_s(m)$ once when it is decreasing, i.e. $\nabla^2\beta_a(m) < 0$, giving rise to a single maximum in $\beta_a(m)$, i.e. a single minimum point in $\lim \mathrm{Varc}(t)_a$.

Hence, Proposition 3.3 is proven. □

B.5 Proof of Proposition 3.4

Consider a spread period m_s (and $k_s = 1/\ddot{a}_{\overline{m_s}|}$) and an amortization period m_a (with corresponding $\sum \lambda_j^2$ and $\sum \beta_j^2$) such that $\lim \mathrm{Var}f(t)_s = \lim \mathrm{Var}f(t)_a$. Using equations (3.158) to (3.161), we find that

$$\alpha_s(m_s) = \alpha_a(m_a), \tag{B.20}$$

$$\beta_s(m_s)k_s^2\ddot{a}_{\overline{m_a}|}^2 \sum \lambda_j^2 = \beta_a(m_a)\, m_a. \tag{B.21}$$

If $m_a = m_s = 1$, then clearly $\lim \mathrm{Varc}(t)_a = \lim \mathrm{Varc}(t)_s$. For other $\{m_a, m_s\}$, if we show that $k_s^2\ddot{a}_{\overline{m_a}|}^2 \sum \lambda_j^2 < m_a$, then we will show that $\lim \mathrm{Varc}(t)_a > \lim \mathrm{Varc}(t)_s$.

Substituting equations (3.158) and (3.159) into equation (B.20), we have

$$[1 - (u^2 + \sigma^2)(1 - k_s)^2] \sum \lambda_j^2 = 1 - \sigma^2 \sum \beta_j^2. \tag{B.22}$$

Replacing $\sum \beta_j^2$ using equation (3.157), multiplying across by u^2 and rearranging, we obtain $[1 - u^2(1-k_s)^2]\sum \lambda_j^2 = 1$, and so $k_s^2 = [(\sum \lambda_j^2)^{1/2} - v(\sum \lambda_j^2 - 1)^{1/2}]^2 / \sum \lambda_j^2$.

Hence,

$$k_s^2 \ddot{a}^2_{\overline{m_a}|} \sum \lambda_j^2 \tag{B.23}$$

$$= \left[\left(\sum \lambda_j^2\right)^{1/2} - v\left(\sum \lambda_j^2 - 1\right)^{1/2}\right]^2 \ddot{a}^2_{\overline{m_a}|}$$

$$= \left[(\ddot{a}^2_{\overline{m_a}|} + \cdots + \ddot{a}^2_{\overline{1}|})^{1/2} - v(\ddot{a}^2_{\overline{m_a-1}|} + \cdots + \ddot{a}^2_{\overline{1}|})^{1/2}\right]^2 \tag{B.24}$$

$$= (\ddot{a}^2_{\overline{m_a}|} + \cdots + \ddot{a}^2_{\overline{1}|}) + v^2(\ddot{a}^2_{\overline{m_a-1}|} + \cdots + \ddot{a}^2_{\overline{1}|})$$

$$- \; 2v[(\ddot{a}^2_{\overline{m_a}|} + \cdots + \ddot{a}^2_{\overline{1}|})(\ddot{a}^2_{\overline{m_a-1}|} + \cdots + \ddot{a}^2_{\overline{1}|})]^{1/2} \tag{B.25}$$

$$< (\ddot{a}^2_{\overline{m_a}|} + \cdots + \ddot{a}^2_{\overline{1}|}) + v^2(\ddot{a}^2_{\overline{m_a-1}|} + \cdots + \ddot{a}^2_{\overline{1}|}) - 2v(\ddot{a}_{\overline{m_a}|}\ddot{a}_{\overline{m_a-1}|} + \cdots + \ddot{a}_{\overline{2}|}\ddot{a}_{\overline{1}|}) \tag{B.26}$$

where we use inequality (3.163) in the last step. Since $(\ddot{a}_{\overline{t}|} - v\ddot{a}_{\overline{t-1}|})^2 = 1$, $t \in \mathrm{Z}^+$, we find that $k_s^2 \ddot{a}^2_{\overline{m_a}|} \sum \lambda_j^2 < m_a$. From equation (B.21), we have therefore proven that $\lim \mathrm{Var}c(t)_a > \lim \mathrm{Var}c(t)_s$. □

Appendix C

Optimal Controls

C.1 Proof of Optimal Asset Allocation and Contribution Controls

This proof is concerned with the optimal asset allocation and contribution decisions as given in §3.6.3.

Assume that

$$\Phi_t = f(t) + c(t) - B, \tag{C.1}$$

$$\Psi_t = 1 + r + y(t)\alpha. \tag{C.2}$$

and note that

$$y(t)^2 = \alpha^{-2}\Psi_t^2 - 2\alpha^{-2}(1+r)\Psi_t + \alpha^{-2}(1+r)^2, \tag{C.3}$$

$$(c(t) - CT_t)^2 = \Phi_t^2 - 2(f(t) - B + CT_t)\Phi_t + (f(t) - B + CT_t)^2. \tag{C.4}$$

Since $\alpha(t)$ is independent and identically distributed over time, it follows from equation (3.167) that

$$\mathrm{E}[f(t+1)|f(t)] = \Psi_t\Phi_t, \tag{C.5}$$

$$\mathrm{Var}[f(t+1)|f(t)] = \sigma^2\Phi_t^2 y(t)^2, \tag{C.6}$$

and using equation (C.3)

$$\mathrm{E}[f(t+1)^2|f(t)] = (1+\sigma^2\alpha^{-2})\Phi_t^2\Psi_t^2 - 2\sigma^2\alpha^{-2}(1+r)\Phi_t^2\Psi_t + \sigma^2\alpha^{-2}(1+r)^2\Phi_t^2. \tag{C.7}$$

The Bellman optimality equation (equation (3.178)) is

$$J(f(t),t) = \min_{c(t),y(t)} J, \tag{C.8}$$

where

$$J = \theta_1(f(t) - FT_t)^2 + \theta_2(c(t) - CT_t)^2 + \beta \mathrm{E}[J(f(t+1), t+1)|f(t)], \tag{C.9}$$

and with boundary condition, at time $t = N$,

$$J(f(N), N) = \theta_0(f(N) - FT_N)^2. \tag{C.10}$$

A trial solution for equation (C.8) is

$$J(f(t), t) = P_t f(t)^2 - 2Q_t f(t) + R_t. \tag{C.11}$$

The boundary condition in equation (C.10) certainly satisfies the trial solution, with

$$P_N = \theta_0, \tag{C.12}$$

$$Q_N = \theta_0 FT_N. \tag{C.13}$$

We proceed by induction. Suppose that equation (C.11) is a solution of the Bellman equation (C.8) at $t+1$. Then,

$$\begin{aligned}
&\mathrm{E}[J(f(t+1), t+1)|f(t)] \\
&\quad = P_{t+1}\mathrm{E}[f(t+1)^2|f(t)] - 2Q_{t+1}\mathrm{E}[f(t+1)|f(t)] + R_{t+1} \\
&\quad = (1+\sigma^2\alpha^{-2})P_{t+1}\Phi_t^2\Psi_t^2 - 2[Q_{t+1}\Phi_t + \sigma^2\alpha^{-2}(1+r)P_{t+1}\Phi_t^2]\Psi_t \\
&\qquad + \sigma^2\alpha^{-2}(1+r)^2 P_{t+1}\Phi_t^2 + R_{t+1}.
\end{aligned} \tag{C.14}$$

where we make use of equations (C.5) and (C.7).

J may now be written as a quadratic expression in Ψ_t, Φ_t and $f(t)$, by substituting equations (C.4) and (C.14) into equation (C.9):

$$\begin{aligned}
J = &[\beta(1+\sigma^2\alpha^{-2})P_{t+1}\Phi_t^2]\Psi_t^2 - 2\beta[Q_{t+1}\Phi_t + \sigma^2\alpha^{-2}(1+r)P_{t+1}\Phi_t^2]\Psi_t \\
&+ [\theta_2 + \beta\sigma^2\alpha^{-2}(1+r)^2 P_{t+1}]\Phi_t^2 - 2\theta_2(f(t) - B + CT_t)\Phi_t \\
&+ \beta R_{t+1} + \theta_2(f(t) - B + CT_t)^2 + \theta_1(f(t) - FT_t)^2.
\end{aligned} \tag{C.15}$$

The first two terms on the right hand side of equation (C.15) are quadratic in Ψ_t and (upon completing the square) may be written as

$$\Psi_t^A(\Psi_t - \Psi_t^B)^2 + \Psi_t^C, \tag{C.16}$$

where

$$\Psi_t^A = \beta(1+\sigma^2\alpha^{-2})P_{t+1}\Phi_t^2, \tag{C.17}$$

$$\Psi_t^B = \frac{\sigma^2\alpha^{-2}(1+r)P_{t+1}\Phi_t + Q_{t+1}}{(1+\sigma^2\alpha^{-2})P_{t+1}\Phi_t}, \tag{C.18}$$

$$\Psi_t^C = -\frac{\beta[\sigma^2\alpha^{-2}(1+r)P_{t+1}\Phi_t + Q_{t+1}]^2}{(1+\sigma^2\alpha^{-2})P_{t+1}}. \tag{C.19}$$

The third and fourth terms on the right hand side of equation (C.15) as well as the last term (Ψ_t^C) in expression (C.16) are quadratic in Φ_t and may be written as

$$\begin{aligned}&\Phi_t^2 \tilde{P}_{t+1}^{-1}(\alpha^2+\sigma^2)^{-1}\\ &\qquad - \Phi_t[2\theta_2(\alpha^2+\sigma^2)(f(t)-B+CT_t)+2\beta\sigma^2(1+r)Q_{t+1}]/(\alpha^2+\sigma^2)\\ &\qquad\qquad - \beta\alpha^2 Q_{t+1}^2/[(\alpha^2+\sigma^2)P_{t+1}] \end{aligned} \quad (C.20)$$

or, by completing the square, as

$$\Phi_t^A(\Phi_t-\Phi_t^B)^2+\Phi_t^C, \quad (C.21)$$

where

$$\tilde{P}_{t+1} = [\theta_2(\alpha^2+\sigma^2)+\beta\sigma^2(1+r)^2P_{t+1}]^{-1}, \quad (C.22)$$

$$\Phi_t^A = \tilde{P}_{t+1}^{-1}(\alpha^2+\sigma^2)^{-1}, \quad (C.23)$$

$$\Phi_t^B = \theta_2(\alpha^2+\sigma^2)\tilde{P}_{t+1}(f(t)-B+CT_t)+\beta\sigma^2(1+r)\tilde{P}_{t+1}Q_{t+1}], \quad (C.24)$$

$$\begin{aligned}\Phi_t^C &= -\tilde{P}_{t+1}[\theta_2(\alpha^2+\sigma^2)(f(t)-B+CT_t)+\beta\sigma^2(1+r)Q_{t+1}]^2]/(\alpha^2+\sigma^2)\\ &\quad - \beta\alpha^2(\alpha^2+\sigma^2)^{-1}P_{t+1}^{-2}Q_{t+1}^2. \end{aligned} \quad (C.25)$$

Finally, the last three terms on the right hand side of equation (C.15) along with the last term (Φ_t^C) in expression (C.21) are quadratic in $f(t)$ and may be written as

$$\begin{aligned}&f(t)^2\left[\theta_1+\theta_2-\theta_2^2(\alpha^2+\sigma^2)\tilde{P}_{t+1}\right]\\ &\quad - f(t)\left[2\theta_1FT_t+2\theta_2(B-CT_t)-\tilde{P}_{t+1}[2\theta_2^2(\alpha^2+\sigma^2)(B-CT_t)-2\theta_2\beta\sigma^2(1+r)Q_{t+1}]\right]\\ &\qquad + \text{remaining terms independent of } f(t), \end{aligned} \quad (C.26)$$

or

$$A_tf(t)^2-2B_tf(t)+\text{remaining terms independent of } f(t), \quad (C.27)$$

where

$$A_t = \theta_1+\theta_2\beta\sigma^2(1+r)^2\tilde{P}_{t+1}P_{t+1}, \quad (C.28)$$

$$B_t = \theta_1FT_t+\theta_2\beta\sigma^2(1+r)\tilde{P}_{t+1}[Q_{t+1}+(1+r)P_{t+1}(B-CT_t)], \quad (C.29)$$

since, from equation (C.22),

$$1-\theta_2(\alpha^2+\sigma^2)\tilde{P}_{t+1} \;=\; \beta\sigma^2(1+r)^2\tilde{P}_{t+1}P_{t+1}. \quad (C.30)$$

J may therefore be written as

$$\begin{aligned}J \;=\; &\Psi_t^A(\Psi_t-\Psi_t^B)^2+\Phi_t^A(\Phi_t-\Phi_t^B)^2\\ &+A_tf(t)^2-2B_tf(t)+\text{remaining terms independent of } f(t). \end{aligned} \quad (C.31)$$

J has a unique minimum in Ψ_t and Φ_t provided $\Psi_t^A > 0$ and $\Phi_t^A > 0$. It is sufficient that $P_t > 0$ for $t \in [1, N]$ for both these conditions to be satisfied (since $P_t > 0 \Rightarrow \tilde{P}_t > 0$). The minimum occurs when $\Psi_t = \Psi_t^B$ and $\Phi_t = \Phi_t^B$ simultaneously. Now, there is a direct linear relationship between $c(t)$ and Φ_t (equation (C.1)) and between $y(t)$ and Ψ_t ($\alpha > 0$ in equation (C.2)). Therefore,

$$\min_{c(t),y(t)} J \;=\; A_t f(t)^2 - 2B_t f(t) + \text{remaining terms independent of } f(t), \tag{C.32}$$

which is in the form postulated in the trial solution (C.11). Since the solution holds for $t = N$, it holds for $t \in [1, N]$. $P_t = A_t$ and $Q_t = B_t$, by comparing equations (C.11) and (C.32), with respective boundary conditions (C.12) and (C.13). Since $P_N = \theta_0 > 0$, $P_t = A_t > 0$ for $t \in [1, N]$ and the sufficient condition for the existence of a single minimum is satisfied.

Let $\Theta_t = \theta_2(\alpha^2 + \sigma^2)\tilde{P}_t$. Then,

$$1 - \Theta_{t+1} = \beta\sigma^2(1+r)^2\tilde{P}_{t+1}P_{t+1}, \tag{C.33}$$

from equation (C.30).

$$\Phi_t^* = \Phi_t^B = \Theta_{t+1}(f(t) + c(t) - B) + (1 - \Theta_{t+1})Q_{t+1}P_{t+1}^{-1}(1+r)^{-1}. \tag{C.34}$$

Using equation (C.1), $c^*(t) = \Phi_t^* + B - f(t)$ is readily obtained. Replacing $\Phi_t = \Phi_t^*$ in the right hand side of equation $\Psi_t = \Psi_t^B$ gives

$$\Psi_t^* = \frac{Q_{t+1} + \sigma^2\alpha^{-2}(1+r)P_{t+1}\Phi_t^*}{(1+\sigma^2\alpha^{-2})P_{t+1}\Phi_t^*} \tag{C.35}$$

and from equation (C.2),

$$\begin{aligned} \alpha y^*(t) &= \Psi_t^* - (1+r) \\ &= \frac{Q_{t+1} - (1+r)P_{t+1}\Phi_t^*}{(1+\sigma^2\alpha^{-2})P_{t+1}\Phi_t^*}, \end{aligned} \tag{C.36}$$

which readily yields $y^*(t)$ when Φ_t^* from equation (C.34) is substituted.

C.2 Property of Increasing Risk Aversion

This section is concerned with the coefficient of risk aversion and equation (3.202) of §3.6.4. It involves a slight variation (in terms of disutilities and losses) on the usual derivation (based on utilities and gains), as given by Bertsekas (1976:17–18) or Elton & Gruber (1987:198).

Let y be the random loss or penalty incurred by an agent undertaking a risky activity. $\text{E}y = \bar{y}$ and $\text{Var}y = \sigma^2$. Assume that the properties given by equations (3.199) and (3.200) apply. If the agent is risk-averse, then suppose that he is willing to pay an additional penalty or risk premium π to incur the loss $\bar{y}$, and avoid undertaking the risky activity. Then,

$$L(\bar{y} + \pi) \;=\; \text{E}L(y). \tag{C.37}$$

The premium π is an intuitive measure of the aversion to risk of the agent. Now, the left hand side of equation (C.37) may be expanded around $\bar{y}$,

$$L(\bar{y} + \pi) = L(\bar{y}) + \pi L'(\bar{y}) + \cdots , \tag{C.38}$$

and the right hand side of equation (C.37) may be expanded about $\bar{y}$ thus:

$$L(y) = L(\bar{y}) + (y - \bar{y})L'(\bar{y}) + \frac{1}{2}(y - \bar{y})^2 L''(\bar{y}) + \cdots$$
$$EL(y) = L(\bar{y}) + 0 + \frac{\sigma^2}{2} L''(\bar{y}) + \cdots \tag{C.39}$$

Hence, equating (C.38) and (C.39),

$$\pi \approx \frac{\sigma^2}{2} \frac{L''(\bar{y})}{L'(\bar{y})} \tag{C.40}$$

which justifies the use of $A(y) = L''(y)/L'(y)$ as a coefficient of local risk aversion in terms of disutilities. If the agent is (ideally) increasingly risk-averse as a larger loss is incurred, then a greater positive premium π is paid to forego risk and, from equation (C.40), $A'(y) > 0$.

For a quadratic disutility function $L(y) = \lambda y^2$, π is the solution to a quadratic equation and, neglecting the negative solution ($\pi + y > 0$),

$$\pi = \sqrt{(y^2 + \sigma^2)} - y, \tag{C.41}$$

and $\partial\pi/\partial y = y/\sqrt{(y^2 + \sigma^2)} - 1 < 0$. An agent with a quadratic disutility function therefore exhibits decreasing risk aversion with increasing losses.

Note also that any disutility function can be transformed into a utility function by a series of linear operations. Linear operations have no effect on preference. Cost and reward are related by, say, $x + y =$ constant and

$$U(x) = -AL(y) + B \quad (A > 0).$$

then it follows that $-U''(x)/U'(x) = L''(y)/L'(y)$.

Appendix D

Asset Valuation Method

D.1 A General Model to Prove Propositions 4.4 and 5.3

A general model is set up that allows the second moments of the funding process in two special cases to be investigated:

- the method for smoothing asset values as described in §4.5 and
- the method for avoiding persisting surpluses as described in §5.6.

Generalised Recurrence Relation

Consider a generalised recurrence relationship for the unfunded liability:

$$ul(t+1) = AL + u(t+1)[ul(t) - adj(t) - v_v AL], \quad t \geq 0, \tag{D.1}$$

where

$$adj(t) = k_1(ul(t) - U(t)) + k_2 \sum_{j=0}^{t} p^{t-j}(ul(j) - U(j)) + k_3 p^t + P(t). \tag{D.2}$$

By definition,

$$ul(t) = AL - f(t), \tag{D.3}$$

$$c(t) = NC + adj(t). \tag{D.4}$$

If a fraction $(1-y)$ of the initial unfunded liability ul_0 is being amortized over a fixed term n then, the amortization payment at the beginning of year $(t, t+1)$ is

$$P(t) = \begin{cases} (1-y)ul_0/\ddot{a}_{\overline{n|}}, & 0 \leq t \leq n-1, \\ 0, & t \geq n, \end{cases} \tag{D.5}$$

and the unamortized part of $(1-y)ul_0$ at the beginning of year $(t,\ t+1)$ is

$$U(t) = \begin{cases} (1-y)ul_0\ \ddot{a}_{\overline{n-t}|}/\ddot{a}_{\overline{n}|}, & 0 \le t \le n-1, \\ 0, & t \ge n. \end{cases} \tag{D.6}$$

Note that

$$P(t) = U(t) - v_v U(t+1), \tag{D.7}$$

where we define $U(n) = U(n+1) = \ldots = 0$.

By equations (D.2), (D.4) and (D.7), the contribution process in this general model is therefore

$$c(t) = NC + k_1(ul(t) - U(t)) + k_2 \sum_{j=0}^{t} p^{t-j}(ul(j) - U(j)) + k_3 p^t + U(t) - v_v U(t+1). \tag{D.8}$$

For brevity, we also define

$$ul^u(t) = ul(t) - U(t), \tag{D.9}$$

$$a_t = \mathrm{E}ul^u(t), \tag{D.10}$$

$$b_t = \mathrm{E}ul^u(t)^2, \tag{D.11}$$

$$c_t = \mathrm{E}\left[ul^u(t) \sum_{j=0}^{t} p^{t-j} ul^u(j)\right], \tag{D.12}$$

$$d_t = \mathrm{E}\left[\sum_{j=0}^{t} p^{t-j} ul^u(j)\right]^2. \tag{D.13}$$

We may substitute $P(t)$ from equation (D.7) into equations (D.1) and (D.2) and, using $ul^u(t)$ from equation (D.9), rearrange to obtain, for $t \ge 0$,

$$\begin{aligned} &ul^u(t+1) - (AL - U(t+1)) \\ &= u(t+1)\left[ul^u(t)(1-k_1) - v_v(AL - U(t+1)) - k_2 \sum_{j=0}^{t} p^{t-j} ul^u(j) - k_3 p^t\right]. \end{aligned} \tag{D.14}$$

After taking expectations across the following simple identity,

$$\left[\sum_{j=0}^{t} p^{t-j} ul^u(j)\right]^2 \equiv 2\sum_{j=0}^{t} p^{t-j} ul^u(j) \sum_{k=0}^{j} p^{t-k} ul^u(k) - \sum_{j=0}^{t} p^{2(t-j)} ul^u(j)^2, \tag{D.15}$$

and noting that $p^{t-k} = p^{t-j} p^{j-k}$, we obtain:

$$d_t = 2\sum_{j=0}^{t} p^{2(t-j)} c_j - \sum_{j=0}^{t} p^{2(t-j)} b_j. \tag{D.16}$$

Another useful result follows from taking expectations across the identity,

$$ul^u(t+1)p\sum_{j=0}^{t}p^{t-j}ul^u(j)+ul^u(t+1)^2\equiv ul^u(t+1)\sum_{j=0}^{t+1}p^{t+1-j}ul^u(j), \tag{D.17}$$

which gives:

$$\mathrm{E}\left[ul^u(t+1)p\sum_{j=0}^{t}p^{t-j}ul^u(j)\right]+\mathrm{E}ul^u(t+1)^2=c_{t+1}. \tag{D.18}$$

Dynamic System of Equations

A number of simultaneous difference equations may be set up. First, taking expectations across equation (D.14), and noting that $\{u(t)\}$ is a sequence of independent and identically distributed random variables, so that $u(t+1)$ is independent of $u(s)$ and $ul(s)$ for $s\leq t$, we obtain

$$a_{t+1}=a_t u(1-k_1)-uk_2\sum_{j=0}^{t}p^{t-j}a_j+(1-uv_v)(AL-U(t+1))-uk_3p^t, \tag{D.19}$$

for $t\geq 0$.

In order to find the second moments, we proceed to square both sides of equation (D.14) and then take expectations. Note again that $\{u(t)\}$ is a sequence of independent and identically distributed random variables so that $u(t+1)$ is independent of $u(s)$ and $ul(s)$ for $s\leq t$. Upon substituting d_t from equation (D.16) and collecting like terms, we have, for $t\geq 0$,

$$\begin{aligned}b_{t+1}-q(1-k_1)^2b_t+qk_2^2\sum_{j=0}^{t}p^{2(t-j)}b_j+2qk_2(1-k_1)c_t-2qk_2^2\sum_{j=0}^{t}p^{2(t-j)}c_j\\ =2(AL-U(t+1))a_{t+1}-2qv_v(1-k_1)(AL-U(t+1))a_t-2qk_3(1-k_1)p^ta_t\\ +2qv_vk_2(AL-U(t+1))\sum_{j=0}^{t}p^{t-j}a_j+2qk_2k_3\sum_{j=0}^{t}p^{2(t-j)}p^ja_j\\ +(qv_v^2-1)(AL-U(t+1))^2+2qk_3v_v(AL-U(t+1))p^t+qk_3^2p^{2t}.\end{aligned} \tag{D.20}$$

A third equation may be found by multiplying equation (D.14) by $p\sum_{j=0}^{t}p^{t-j}ul^u(j)$ and adding $ul^u(t+1)^2$ on both sides, and then taking expectations. Upon using equation (D.18) and collecting like terms, we find that, for $t\geq 0$,

$$\begin{aligned}c_{t+1}-up(1-k_1)c_t+uk_2pd_t-b_{t+1}\\ =(1-uv_v)(AL-U(t+1))p\sum_{j=0}^{t}p^{t-j}a_j-uk_3p^{t+1}\sum_{j=0}^{t}p^{t-j}a_j.\end{aligned} \tag{D.21}$$

Again, d_t may be replaced from equation (D.16) to give:

$$c_{t+1} - up(1-k_1)c_t + 2uk_2p\sum_{j=0}^{t} p^{2(t-j)}c_j - b_{t+1} - uk_2p\sum_{j=0}^{t} p^{2(t-j)}b_j$$
$$= (1-uv_v)(AL - U(t+1))p\sum_{j=0}^{t} p^{t-j}a_j - uk_3p\sum_{j=0}^{t} p^{2(t-j)}p^j a_j. \quad \text{(D.22)}$$

Alternative Representation of Difference Equations

Equations (D.19), (D.20) and (D.22) constitute a *linear* dynamic system. In the limit (or 'steady-state'), the system is not affected by initial conditions. We are interested only in the ultimate situation and we may therefore ignore initial conditions. From equation (D.6), $U(t) = 0$ for $t \geq n$. We could consider difference equations (D.19), (D.20) and (D.22) for $t \geq n$. Equivalently and more simply, it is convenient to put $U(t) = 0$ in these three difference equations. Similarly, we can assume that $a_s = b_s = c_s = 0$ for $s < 0$.

It is also convenient to use the forward shift operator E (as distinct from the statistical expectation operator E) where $E^m x_t = x_{t+m}$. One could alternatively use z-transforms or generating functions. Note that, if $x_s = 0$ for $s < 0$,

$$\sum_{j=0}^{t} a^{t-j}x_j = \sum_{j=-\infty}^{t} a^{t-j}x_j = \sum_{j=-\infty}^{t} a^{t-j}(E^{-1})^{t-j}x_t = E(E-a)^{-1}x_t. \quad \text{(D.23)}$$

This also follows from Brand's (1966:375) definition of the 'power shift':

$$E(E-a)^{-1}x_t = (E-a)^{-1}x_{t+1} = (E-a)^{-1}a^t a^{-t}x_{t+1} = a^{t-1}\triangle^{-1}a^{-t}x_{t+1}, \quad \text{(D.24)}$$

and since

$$\triangle\sum_{-\infty}^{t-1} a^{-j}x_{j+1} = a^{-t}x_{t+1}, \quad \text{(D.25)}$$

therefore,

$$E(E-a)^{-1}x_t = a^{t-1}\sum_{-\infty}^{t-1} a^{-j}x_{j+1} = \sum_{-\infty}^{t} a^{t-j}x_j = \sum_{0}^{t} a^{t-j}x_j. \quad \text{(D.26)}$$

In terms of the E operator, equation (D.19) may be written as

$$Ea_t = AL(1-uv_v) + a_tu(1-k_1) - E(E-p)^{-1}a_tuk_2 - uk_3p^t. \quad \text{(D.27)}$$

We may multiply across by $(E-p)$. This is equivalent to applying the following operation to equation (D.19): equation (D.19) is forward-shifted in time (so that it holds for $t \geq -1$) and equation (D.19) multiplied by p is then deducted. This yields

$$P(E)a_t = (E-p)\theta_t, \quad \text{(D.28)}$$

which holds for $t \geq 0$ and where

$$P(E) = [E - u(1-k_1)][E-p] + Euk_2, \tag{D.29}$$

$$\theta_t = AL(1-uv_v) - uk_3p^t. \tag{D.30}$$

Using the forward-shift operator, equation (D.20) may be written as:

$$\begin{aligned}
[E - q(1-k_1)^2]b_t + E(E-p^2)^{-1}b_tqk_2^2 + c_t2qk_2(1-k_1) - E(E-p^2)^{-1}c_t2qk_2^2 \\
= [E - qv_v(1-k_1)]a_t2AL - a_tp^t2qk_3(1-k_1) + E(E-p)^{-1}a_t2ALqv_vk_2 \\
+ E(E-p^2)^{-1}p^ta_t2qk_2k_3 + (qv_v^2 - 1)AL^2. \quad \text{(D.31)}
\end{aligned}$$

Multiplying across by $(E-p^2)(E-p)$ gives

$$\psi(E)(E-p)b_t + \gamma(E)(E-p)c_t = \nu(E)a_t + (qv_v^2-1)AL^2(E-p^2)(E-p), \tag{D.32}$$

which holds for $t \geq 0$ and where

$$\psi(E) = [E - q(1-k_1)^2][E-p^2] + Eqk_2^2, \tag{D.33}$$

$$\gamma(E) = 2qk_2(1-k_1)(E-p^2) - 2qk_2^2E, \tag{D.34}$$

$$\begin{aligned}
\nu(E) = 2AL[E - qv_v(1-k_1)](E-p^2)(E-p) - (E-p^2)(E-p)p^t2qk_3(1-k_1) \\
+ 2qv_vk_2ALE(E-p^2) + E(E-p)p^t2qk_2k_3. \quad \text{(D.35)}
\end{aligned}$$

Equation (D.22) may also be expressed in terms of the forward-shift operator as:

$$\begin{aligned}
[E - up(1-k_1)]c_t + E(E-p^2)^{-1}c_t2uk_2p - Eb_t - E(E-p^2)^{-1}b_tuk_2p \\
= E(E-p)^{-1}a_tAL(1-uv_v)p - E(E-p^2)^{-1}a_tp^tuk_3p. \quad \text{(D.36)}
\end{aligned}$$

Multiplying on both sides by $(E-p^2)(E-p)$ gives

$$\varphi(E)(E-p)c_t - \mu(E)(E-p)b_t = \omega(E)a_t, \tag{D.37}$$

which holds for $t \geq 0$ and where

$$\varphi(E) = [E - up(1-k_1)][E-p^2] + 2uk_2pE, \tag{D.38}$$

$$\mu(E) = E(E-p^2) + uk_2pE, \tag{D.39}$$

$$\omega(E) = AL(1-uv_v)pE(E-p^2) - uk_3pE(E-p)p^t. \tag{D.40}$$

Recurrence Relation for b_t

The difference equations (D.28), (D.32) and (D.37) hold simultaneously for $t \geq 0$. We can contrive to cancel out $\{c_t\}$ from equations (D.32) and (D.37) if we perform a series of linear operations on each. Given the coefficient of c_t in equation (D.32), we multiply across equation (D.37) by $\gamma(E)$ and equation (D.32) multiplied by $\varphi(E)$ is then deducted. Thus we have

$$\begin{aligned}(E-p)&[\psi(E)\varphi(E)+\mu(E)\gamma(E)]b_t \\ &= (qv_v^2-1)AL^2(E-p^2)(E-p)\varphi(E)+\nu(E)\varphi(E)a_t-\omega(E)\gamma(E)a_t. \quad \text{(D.41)}\end{aligned}$$

Now,

$$\begin{aligned}\psi(E)&\varphi(E)+\mu(E)\gamma(E) \\ &= [(E-q(1-k_1)^2)(E-p^2)+Eqk_2^2]\varphi(E)+[E(E-p^2)+uk_2pE]\gamma(E) \\ &= (E-q(1-k_1)^2)(E-p^2)\varphi(E)+E(E-p^2)\gamma(E) \\ &\quad + Eqk_2^2(E-up(1-k_1))(E-p^2)+uk_2pE2qk_2(1-k_1)(E-p^2) \\ &= (E-p^2)Q(E), \quad \text{(D.42)}\end{aligned}$$

where

$$Q(E)=[E-q(1-k_1)^2]\varphi(E)+E\gamma(E)+Eqk_2^2[E+up(1-k_1)]. \quad \text{(D.43)}$$

We can now replace a_t from equation (D.28) into equation (D.41) while making use of equations (D.42) and (D.43). If we also multiply across by $P(E)(E-p)^{-1}$, we obtain, for $t \geq 0$,

$$P(E)Q(E)(E-p^2)b_t = (qv_v^2-1)AL^2(E-p^2)\varphi(E)P(E)+\nu(E)\varphi(E)\theta_t-\omega(E)\gamma(E)\theta_t. \quad \text{(D.44)}$$

Equation (D.44) is a linear difference equation satisfied by b_t (ignoring initial conditions).

Recurrence Relation for c_t

We can also cancel $\{a_t\}$ and $\{b_t\}$ from the system of equations to obtain an equation in $\{c_t\}$ alone. Substituting b_t from equation (D.44) and a_t from equation (D.28) into equation (D.37) results in the following:

$$\begin{aligned}c_t&\varphi(E)(E-p) \\ &= \left[(qv_v^2-1)AL^2(E-p^2)\varphi(E)P(E)+\nu(E)\varphi(E)\theta_t-\omega(E)\gamma(E)\theta_t\right] \times \\ &\qquad [\mu(E)(E-p)/P(E)Q(E)(E-p^2)] + (E-p)\omega(E)/P(E)\theta_t. \quad \text{(D.45)}\end{aligned}$$

Multiplying across by $P(E)Q(E)(E-p^2)(E-p)^{-1}$ gives

$$\begin{aligned}\varphi(E)&P(E)Q(E)(E-p^2)c_t \\ &= \mu(E)\left[(qv_v^2-1)AL^2(E-p^2)\varphi(E)P(E)+\nu(E)\varphi(E)\theta_t\right] \\ &\qquad -\omega(E)\gamma(E)\mu(E)\theta_t+\omega(E)Q(E)(E-p^2)\theta_t. \quad \text{(D.46)}\end{aligned}$$

The last two terms on the right hand side of the above simplify, using equation (D.42), to

$$\omega(E)[Q(E)(E-p^2)-\gamma(E)\mu(E)]\theta_t=\omega(E)\psi(E)\varphi(E)\theta_t. \tag{D.47}$$

Substitution of equation (D.47) into equation (D.46) supplies

$$P(E)Q(E)(E-p^2)c_t=(qv_v^2-1)AL^2(E-p^2)P(E)\mu(E)+\nu(E)\mu(E)\theta_t+\omega(E)\psi(E)\theta_t. \tag{D.48}$$

Limits

If we disregard initial conditions, equations (D.28), (D.44) and (D.48) are linear difference equations satisfied by a_t, b_t and c_t respectively. From the left hand side of equation (D.28), it is clear that stability depends on the roots of the equation $P(z)=0$. From the left hand sides of equations (D.44) and (D.48), stability also depends on the polynomial $Q(z)$.

Special Cases

Special cases can be considered at this stage. The second moments for the funding process when asset values are smoothed, as described in Chapter 4, are derived in the next section. The second moments when persisting surpluses are removed using a method described in Chapter 5 are derived separately in Appendix F.

D.2 Proof of Proposition 4.4

A Special Case of §D.1

This proof concerns the second moments of the pension funding process when asset values are smoothed according to the method described by equations (4.2) or (4.3). Comparison of equations (4.17) and (D.8) reveals that this can be considered to be a special case of the general model described in §D.1.

Note that in Proposition 4.4, we assume condition (4.46) from Corollary 4.1, i.e. that

1. the valuation discount rate equals the mean long-term rate of return on assets ($i=i_v$)
2. the mean rate of return on assets $i>-100\%$, from which it follows that $0\le Ku<1$ (Haberman, 1992a) (since $K=1-1/\ddot{a}_{\overline{m|}}$) and therefore that $0\le\lambda Ku=p<1$ (by definition, $0\le\lambda<1$).

In equations (D.2) and (D.8), we may therefore let

$$i=i_v,\quad k_1=0,\quad k_2=(1-K)(1-\lambda),\quad k_3=k\lambda yul_0,\quad p=\lambda Ku,\quad |p|<1. \tag{D.49}$$

In this special case, equation (D.2) now becomes

$$adj(t) = (1 - K)(UL(t) - U(t)) + P(t), \tag{D.50}$$

where

$$UL(t) - U(t) = (1 - \lambda)\sum_{j=0}^{t} p^{t-j} ul^u(j) + \lambda yul_0 p^t. \tag{D.51}$$

By definition,

$$F(t) = AL - UL(t). \tag{D.52}$$

Equation (D.14) can be rewritten as

$$\begin{aligned} &ul^u(t+1) - (AL - U(t+1)) \\ &\quad = u(t+1)\left[ul^u(t) - k(1-\lambda)\sum_{j=0}^{t} p^{t-j} ul^u(j) - k\lambda yul_0 p^t - v(AL - U(t+1))\right]. \end{aligned} \tag{D.53}$$

We have shown in Corollary 4.1 that, provided certain conditions hold,

$$\lim_{t\to\infty} \mathrm{E}ul(t) = \lim_{t\to\infty} \mathrm{E}ul^u(t) = 0, \tag{D.54}$$

$$\lim_{t\to\infty} \mathrm{E}UL(t) = 0, \tag{D.55}$$

$$\lim_{t\to\infty} \mathrm{E}adj(t) = 0, \tag{D.56}$$

and by taking expected values of equation (D.51) and then limits as $t \to \infty$ and using equation (D.55), it is evident that

$$\lim_{t\to\infty} \sum_{j=0}^{t} p^{t-j} \mathrm{E}ul^u(j) = \lim_{t\to\infty} \sum_{j=0}^{t} p^{t-j} a_j = 0. \tag{D.57}$$

Mean Square Deviation of $ul(t)$ and Variance of $f(t)$

In the special case (equations (D.49)) considered in this section,

$$\theta_t = -uk_3p^t, \tag{D.58}$$

$$P(E) = (E-u)(E-p) + Eu(1-K)(1-\lambda), \tag{D.59}$$

$$\psi(E) = (E-q)(E-p^2) + Eq(1-K)^2(1-\lambda)^2, \tag{D.60}$$

$$\gamma(E) = 2q(1-K)(1-\lambda)[E-p^2-(1-K)(1-\lambda)E], \tag{D.61}$$

$$\nu(E) = 2AL(E-qv)(E-p^2)(E-p) - p^t 2qk_3(E-p^2)(E-p), \tag{D.62}$$

$$\varphi(E) = (E-up)(E-p^2) + 2u(1-K)(1-\lambda)pE, \tag{D.63}$$

$$\mu(E) = E(E-p^2) + u(1-K)(1-\lambda)pE, \tag{D.64}$$

$$\omega(E) = -uk_3p^{t+1}E(E-p), \tag{D.65}$$

$$Q(E) = (E-q)\varphi(E) + E\gamma(E) + Eq(1-K)^2(1-\lambda)^2(E+up) \tag{D.66}$$

$$\begin{aligned} &= (E-q)(E-up)(E-p^2) + 2u(1-K)(1-\lambda)pE(E-q) \\ &\quad + 2q(1-K)(1-\lambda)E[E-p^2-(1-K)(1-\lambda)E] + Eq(1-K)^2(1-\lambda)^2(E+up). \end{aligned} \tag{D.67}$$

Hence, equation (D.44) may be written as

$$\begin{aligned} &P(E)Q(E)(E-p^2)b_t \\ &\quad = \sigma^2v^2AL^2(E-p^2)\varphi(E)P(E) \\ &\qquad - \left[2AL(E-qv)(E-p^2)(E-p) - p^t2qk_3(E-p^2)(E-p)\right] \times \\ &\qquad\qquad \left[(E-up)(E-p^2) + 2u(1-K)(1-\lambda)pE\right] uk_3p^t \\ &\qquad - \left[uk_3p^{t+1}E(E-p)\right] 2q(1-\lambda)\left[E-p^2-(1-K)(1-\lambda)E\right] uk_3p^t, \end{aligned} \tag{D.68}$$

while equation (D.48) is

$$\begin{aligned} &P(E)Q(E)(E-p^2)c_t \\ &\quad = \sigma^2v^2AL^2(E-p^2)\mu(E)P(E) \\ &\qquad - \left[2AL(E-qv)(E-p^2)(E-p) - p^t2qk_3(E-p^2)(E-p)\right] \times \\ &\qquad\qquad \left[E(E-p^2) + u(1-K)(1-\lambda)pE\right] uk_3p^t \\ &\qquad + \left[uk_3p^{t+1}E(E-p)\right]\left[(E-q)(E-p^2) + Eq(1-K)^2(1-\lambda)^2\right] uk_3p^t. \end{aligned} \tag{D.69}$$

Note that

$$\varphi(1) = (1 - up)(1 - p^2) + 2u(1 - K)(1 - \lambda)p \tag{D.70}$$
$$= (1 - \lambda K u^2)(1 - \lambda^2 K^2 u^2) + 2\lambda K(1 - \lambda)(1 - K)u^2, \tag{D.71}$$
$$\mu(1) = 1 - p^2 + u(1 - K)(1 - \lambda)p \tag{D.72}$$
$$= 1 + \lambda K(1 - K - \lambda)u^2, \tag{D.73}$$
$$\gamma(1) = 2q(1 - K)(1 - \lambda)[1 - p^2 - (1 - K)(1 - \lambda)]$$
$$= 2q(1 - K)(1 - \lambda)[1 - \lambda^2 K^2 u^2 - (1 - K)(1 - \lambda)], \tag{D.74}$$
$$Q(1) = (1 - q)\varphi(1) + \gamma(1) + q(1 - K)^2(1 - \lambda)^2(1 + up). \tag{D.75}$$

Now, $p = \lambda K u$ (from equation (D.49)) is symmetrical in K and λ, i.e. K and λ may be interchanged. Similarly, $\varphi(1)$, $\mu(1)$, $\gamma(1)$ and $Q(1)$ are symmetrical in K and λ. $Q(1)$ in equation (D.75) may be expanded by replacing $p = \lambda K u$, $\varphi(1)$ (from equation (D.71)) and $\gamma(1)$ (from equation (D.74)), and collecting terms in λ. This yields

$$Q(1) = (1 - qK^2) - \lambda K[u^2(1 - qK^2) + 2(1 - K)\sigma^2]$$
$$- \lambda^2[u^2(1 - qK^2) + \sigma^2(1 - K)^2] + \lambda^3 K u^2[u^2(1 - qK^2) + \sigma^2(1 - K^2)] \tag{D.76}$$
$$Q(1) = (1 - qK^2)[1 - \lambda u^2 K - \lambda^2 u^2 + \lambda^3 u^4 K]$$
$$- \lambda[2K(1 - K)\sigma^2 + \lambda(1 - K)^2\sigma^2 - \lambda^2 K(1 - K^2)\sigma^2 u^2]$$
$$= (1 - qK^2)(1 - \lambda^2 u^2)(1 - \lambda K u^2)$$
$$- \lambda(1 - K)\sigma^2[2K(1 - \lambda^2 u^2) + \lambda(1 - K)(1 + \lambda K u^2)]. \tag{D.77}$$

Because $Q(1)$ is symmetrical in K and λ, $Q(1)$ may also be written as

$$Q(1) = (1 - q\lambda^2)(1 - K^2 u^2)(1 - \lambda K u^2)$$
$$- K(1 - \lambda)\sigma^2[2\lambda(1 - K^2 u^2) + K(1 - \lambda)(1 + \lambda K u^2)]. \tag{D.78}$$

Equations (D.68) and (D.69) are linear difference equations in b_t and c_t respectively. Since $|p| < 1$ (equation (D.49)), the second and third terms on the right hand side of equations (D.68) and (D.69) vanish as $t \to \infty$. *Assume* for now that the limits of b_t and c_t as $t \to \infty$ exist. Then, taking limits as $t \to \infty$ on equation (D.68),

$$P(1)Q(1)(1 - p^2) \lim_{t\to\infty} b_t = \sigma^2 v^2 A L^2 (1 - p^2)\varphi(1)P(1), \tag{D.79}$$
$$\lim_{t\to\infty} b_t = \sigma^2 v^2 A L^2 \varphi(1)/Q(1). \tag{D.80}$$

Define $Q = Q(1)$ (from either equation (D.77) or (D.78)). Also define $V_\infty = \sigma^2 v^2 A L^2/Q$. Then,

$$\lim_{t\to\infty} b_t = V_\infty \varphi(1), \tag{D.81}$$
$$= V_\infty[(1 - \lambda K u^2)(1 - \lambda^2 K^2 u^2) + 2\lambda K(1 - \lambda)(1 - K)u^2], \tag{D.82}$$

where $\varphi(1)$ is substituted from equation (D.71).

It is now possible to find the variance of the fund level.

$$\begin{aligned}\lim_{t\to\infty} \mathrm{Var}f(t) &= \lim_{t\to\infty} \mathrm{Var}ul(t) \quad \text{(given equation (D.3))} \\ &= \lim_{t\to\infty} \mathrm{Var}ul^u(t) \quad \text{(from equation (D.9))} \\ &= \lim_{t\to\infty} \mathrm{E}ul^u(t)^2 \quad \text{(using equation (D.54))} \\ &= \lim_{t\to\infty} b_t \quad \text{(by definition from equation (D.11)).} \end{aligned} \tag{D.83}$$

Hence, using equation (D.82),

$$\lim_{t\to\infty} \mathrm{Var}f(t) = V_\infty[(1-\lambda K u^2)(1-\lambda^2 K^2 u^2) + 2\lambda K(1-\lambda)(1-K)u^2], \tag{D.84}$$

which proves equation (4.60) of Proposition 4.4.

Covariances

The same procedure yields the required covariances. *Assuming* that the limit of c_t as $t \to \infty$ exists and noting that the second and third terms on the right hand side of equation (D.69) vanish as $t \to \infty$ (since $|p| < 1$ (equation (D.49)), then by taking limits on equation (D.69),

$$P(1)Q(1)(1-p^2) \lim_{t\to\infty} c_t = \sigma^2 v^2 A L^2 (1-p^2)\mu(1)P(1), \tag{D.85}$$

$$\lim_{t\to\infty} c_t = \sigma^2 v^2 A L^2 \mu(1)/Q(1),$$

$$= V_\infty \mu(1), \tag{D.86}$$

$$= V_\infty[1 + \lambda K(1 - K - \lambda)u^2], \tag{D.87}$$

where $\mu(1)$ is replaced from equation (D.73), $Q = Q(1)$ (from either equation (D.77) or (D.78)) and $V_\infty = \sigma^2 v^2 A L^2 Q^{-1}$ again.

Hence, assuming that the limit exists,

$$\begin{aligned}\lim_{t\to\infty} \mathrm{Cov}\left[ul^u(t), \sum_{j=0}^{t} p^{t-j} ul^u(j)\right] &= \lim_{t\to\infty} c_t - \lim_{t\to\infty} \mathrm{E}ul^u(t) \lim_{t\to\infty} \sum_{j=0}^{t} p^{t-j} \mathrm{E}ul^u(j) \\ &= V_\infty[1 + \lambda K(1 - K - \lambda)u^2],\end{aligned} \tag{D.88}$$

upon using equations (D.54), (D.57) and (D.87).

Other covariances follow readily:

$$\begin{aligned}\mathrm{Cov}[f(t), F(t)] &= \mathrm{Cov}[AL - ul(t), AL - UL(t)] \quad \text{(from equations (D.3) and (D.52))} \\ &= \mathrm{Cov}[ul(t), UL(t)] \\ &= \mathrm{Cov}[ul(t) - U(t), UL(t) - U(t)] \\ &= \mathrm{Cov}\left[ul^u(t), (1-\lambda)\sum_{j=0}^{t} p^{t-j} ul^u(j) + \lambda y u l_0 p^t\right]\end{aligned}$$

(using equation (D.51))

$$= (1-\lambda)\mathrm{Cov}\left[ul^u(t), \sum_{j=0}^{t} p^{t-j} ul^u(j)\right]. \tag{D.89}$$

Assuming that the limits exist, and using equation (D.88),

$$\lim_{t\to\infty} \mathrm{Cov}[f(t), F(t)] = V_\infty(1-\lambda)[1+\lambda K(1-K-\lambda)u^2]. \tag{D.90}$$

From equations (D.4), (D.50) and (D.52),

$$\mathrm{Cov}[f(t), c(t)] = -(1-K)\mathrm{Cov}[f(t), F(t)], \tag{D.91}$$

and upon using equation (D.90),

$$\lim_{t\to\infty} \mathrm{Cov}[f(t), c(t)] = -V_\infty(1-K)(1-\lambda)[1+\lambda K(1-K-\lambda)u^2]. \tag{D.92}$$

Equations (D.90) and (D.92) are of course identical to equations (4.63) and (4.64) respectively in Proposition 4.4.

Variance of $F(t)$

Other results follow from the above. In the special case defined in equations (D.49) that we are considering, equation (D.21) becomes

$$c_{t+1} - upc_t + u(1-K)(1-\lambda)pd_t - b_{t+1} = -uk_3 p^{t+1} \sum_{j=0}^{t} p^{t-j} a_j. \tag{D.93}$$

Again, assuming that the limits exist and using equations (D.57), (D.81) (replacing $\varphi(1)$ from equation (D.70)) and (D.86) (replacing $\mu(1)$ from equation (D.72)),

$$\begin{aligned}
u(1-K)(1-\lambda)p \lim_{t\to\infty} d_t &\\
&= \lim b_t - (1-up)\lim c_t \\
&= V_\infty[(1-up)(1-p^2) + 2u(1-K)(1-\lambda)p] \\
&\qquad - V_\infty(1-up)[(1-p^2) + u(1-K)(1-\lambda)p] \\
&= V_\infty u(1-K)(1-\lambda)p(1+up).
\end{aligned} \tag{D.94}$$

Hence,

$$\begin{aligned}
\lim_{t\to\infty} d_t &= V_\infty(1+up) \\
&= V_\infty(1+\lambda K u^2).
\end{aligned} \tag{D.95}$$

It is straightforward that

$$\lim_{t\to\infty} \text{Var} \sum_{j=0}^{t} p^{t-j} ul^u(t) = \lim_{t\to\infty} d_t - \lim_{t\to\infty} \left[\sum_{j=0}^{t} p^{t-j} \text{E}ul^u(t) \right]^2$$
$$= V_\infty(1 + \lambda K u^2), \tag{D.96}$$

where we use equations (D.57) and (D.95). Furthermore,

$$\text{Var}F(t) = \text{Var}UL(t) \quad \text{(given equation (D.52))}$$
$$= (1-\lambda)^2 \text{Var} \sum_{j=0}^{t} p^{t-j} ul^u(t) \quad \text{(from equation (D.51))}, \tag{D.97}$$

$$\lim_{t\to\infty} \text{Var}F(t) = V_\infty (1-\lambda)^2 (1 + \lambda K u^2) \quad \text{(using equation (D.96))}, \tag{D.98}$$

which proves equation (4.61) of Proposition 4.4.

Variance of $c(t)$

Finally,

$$\lim_{t\to\infty} \text{Var}c(t) = \lim_{t\to\infty} \text{Var}adj(t) \quad \text{(given equation (D.4))}$$
$$= (1-K)^2 \lim \text{Var}F(t) \quad \text{(from equations (D.50) and (D.52))}$$
$$= V_\infty (1-K)^2 (1-\lambda)^2 (1 + \lambda K u^2) \quad \text{(using equation (D.98))}, \tag{D.99}$$

and

$$\lim_{t\to\infty} \text{Cov}[c(t), F(t)] = -\lim_{t\to\infty} \text{Cov}[adj(t), UL(t)] \quad \text{(given equations (D.4) and (D.52))}$$
$$= -\lim_{t\to\infty} (1-K)\text{Var}UL(t) \quad \text{(from equation (D.50))}$$
$$= -\lim_{t\to\infty} (1-K)\text{Var}F(t) \quad \text{(from equation (D.97))}$$
$$= -V_\infty (1-K)(1-\lambda)^2 (1 + \lambda K u^2) \quad \text{(using equation (D.98))}. \tag{D.100}$$

The last two results (equations (D.99) and (D.100)) correspond respectively to equations (4.62) and (4.65) of Proposition 4.4.

Stability Conditions

We have assumed thus far that these limits exist and must now establish appropriate stability conditions. From Corollary 4.1, for the limits in equations (D.54)–(D.57) to exist, it is sufficient that:

$$i_v = i > -100\%, \tag{D.101}$$
$$0 \le K < v, \tag{D.102}$$
$$0 \le \lambda < v. \tag{D.103}$$

Conditions (D.101), (D.102) and (D.103) imply that $0 \leq p = \lambda K u < 1$, which satisfies the requirement that $|p| < 1$ in equation (D.49).

The limits of b_t and c_t as $t \to \infty$ exist if and only if the magnitude of the roots of $P(z)Q(z)(z - p^2) = 0$ are less than unity, from the right hand sides of equations (D.68) and (D.69) respectively. But $P(z)$ is the characteristic polynomial that determines first moment stability, as seen in equation (D.28) (as well as in equation (4.26)). We need to investigate the roots of $Q(z) = 0$.

From equation (D.67), $Q(z)$ can be expanded into a standard cubic form:

$$Q(z) = z^3 - Az^2 + Bz - C, \tag{D.104}$$

where

$$A = p^2 + q(1 - (1-K)(1-\lambda))^2 + up(1 - 2(1-K)(1-\lambda)), \tag{D.105}$$

$$B = [up^2 + uq(1 - (1-K)(1-\lambda))^2 + pq(1 - 2(1-K)(1-\lambda))]p, \tag{D.106}$$

$$C = up^3 q. \tag{D.107}$$

The coefficient of z^3 in $Q(z)$ is one. Necessary and sufficient conditions for the roots of $Q(z) = 0$ to be less than unity in magnitude are given by, among others, Jury (1962, 1964:136) and Marden (1966):

$$|C| < 1, \tag{D.108}$$

$$|A + C| < 1 + B, \tag{D.109}$$

$$|AC - B| < |1 - C^2|. \tag{D.110}$$

We now show that conditions (D.108), (D.109) and (D.110) are true if conditions (D.101)–(D.103) hold as well as if

$$Q > 0 \tag{D.111}$$

and

$$(1 + \lambda^2 K^2 q u^2)(1 + \lambda^3 K^3 \sigma^2 u^2 - \lambda^4 K^4 q u^6) > 2\lambda^4 K^4 (\lambda + K) q \sigma^2 u^4 + \lambda K (\lambda + K)^2 q u^2 (1 - \lambda^2 K^2 q u^2), \tag{D.112}$$

where $Q = Q(z = 1) = 1 - A + B - C$ (from equation (D.104)) and Q is also expanded in equations (D.77) and (D.78). First note that

$$Q > 0 \quad \Rightarrow \quad qK^2 < 1, \tag{D.113}$$

$$Q > 0 \quad \Rightarrow \quad q\lambda^2 < 1. \tag{D.114}$$

The second term on the right hand side of equation (D.77) is positive, given stability conditions (D.101)–(D.103). In the first term on the right hand side of equation (D.77), $1 - \lambda^2 u^2$

and $1 - \lambda K u^2$ are also positive given the same first moment stability conditions as above. For $Q > 0$ (condition (4.58)), it is necessary, but *not* sufficient, that $1 - qK^2 > 0$. By the symmetry between K and λ (using equation (D.78)), it is also necessary, but not sufficient, that $1 - q\lambda^2 > 0$.

Condition (D.108)

If conditions (D.102), (D.103) and (D.111) are true (and using implication (D.113)), then $|C| = |up^3 q| = (\lambda u)^2 qK^2 |\lambda u||uK| < 1$, i.e. condition (D.108) follows.

Condition (D.109)

The requirement that $A + C < 1 + B$ is of course equivalent to inequality (D.111).

Next, consider $A + C > -(1 + B)$. Assuming that conditions (D.101)–(D.103) are true, then $C \geq 0$. In addition,

$$\begin{aligned} A &= p^2 + qK^2 + q(1-K)^2\lambda^2 + 2qK(1-K)\lambda + up - 2\lambda(1-\lambda)u^2 K(1-K) \\ &= p^2 + qK^2 + q(1-K)^2\lambda^2 + up + 2(1-K)\lambda K(q - (1-\lambda)u^2) \\ &= p^2 + qK^2 + q(1-K)^2\lambda^2 + up + 2(1-K)\lambda K(\sigma^2 + \lambda u^2) \\ &\geq 0, \end{aligned} \tag{D.115}$$

given conditions (D.101), (D.102), (D.103). Further,

$$\begin{aligned} B &= p[up^2 + uq(1 - (1-K)(1-\lambda))^2 + u^2 p(1 - 2(1-K)(1-\lambda)) \\ &\qquad + \sigma^2 p(1 - 2(1-K)(1-\lambda))] \\ &= upA + \sigma^2 p^2 (1 - 2(1-K)(1-\lambda)) \qquad\qquad (D.116) \\ &= up[p^2 + qK^2 + q(1-K)^2\lambda^2 + up + 2(1-K)\lambda^2 K u^2] + \sigma^2 p^2 (1 + 2\lambda(1-K)) \\ &\geq 0, \end{aligned} \tag{D.117}$$

given (D.101), (D.102) and (D.103). Hence, $A + C \geq 0 > -1 \geq -(1 + B)$.

Therefore, if conditions (D.101), (D.102), (D.103) and (D.111) hold, then condition (D.109) also holds.

Condition (D.110)

Provided conditions (D.101), (D.102), (D.103) and condition (D.111) are true (using implication (D.113)), $0 \leq C < 1$. First, consider $AC - B < 1 - C^2$. If condition (D.109) holds, then $A + C < 1 + B$ and so $AC + C^2 < 1 + B$.

Next, consider $AC - B > -(1 - C^2)$. This turns out to be an inequality of the sixth degree in λ. By exploiting the symmetry between K and λ and collecting terms in λK and $\lambda + K$, it may be written as inequality (D.112).

Sufficient Conditions

We have shown that provided conditions (D.101), (D.102), (D.103), (D.111) and (D.112) are true, the limits in equations (D.82)–(D.100) do exist.

This concludes the proof of Proposition 4.4. □

Appendix E

Properties of Asset Valuation Method

E.1 Proof of Proposition 4.5

This proof is concerned with the stability of actuarial asset values (as defined in equation (4.2) or (4.3)) and their proximity to market values.

$$\mathrm{E}[f(t) - F(t)]^2 = \mathrm{Var}f(t) + \mathrm{Var}F(t) - 2\mathrm{Cov}[f(t), F(t)] + (\mathrm{E}[f(t) - F(t)])^2, \tag{E.1}$$

and since $\lim(\mathrm{E}[f(t) - F(t)])^2 = (\lim \mathrm{E}f(t) - \lim \mathrm{E}F(t))^2 = 0$ (using equation (4.50)), it is clear that, in the limit as $t \to \infty$,

$$\lim \mathrm{E}[f(t) - F(t)]^2 = \lim \mathrm{Var}f(t) + \lim \mathrm{Var}F(t) - 2\lim \mathrm{Cov}[f(t), F(t)]. \tag{E.2}$$

Each of the limits on the right hand side of equation (E.2) exists (equations (4.60), (4.61) and (4.63)), given the stability conditions of Proposition 4.4, so that inequality (4.68) is true.

As for inequality (4.69) of Proposition 4.5, we substitute $\lim \mathrm{Var}f(t)$ and $\lim \mathrm{Var}F(t)$ from equations (4.60) and (4.61) respectively to obtain

$$\begin{aligned}
&\lim_{t\to\infty} \mathrm{Var}f(t) - \lim_{t\to\infty} \mathrm{Var}F(t) \\
&\qquad = V_\infty[(1 - \lambda K u^2)(1 - \lambda^2 K^2 u^2) + 2\lambda K(1-\lambda)(1-K)u^2 \\
&\qquad\qquad - (1-\lambda)^2(1 + \lambda K u^2)]
\end{aligned}$$

(and upon expanding and collecting terms in λ)

$$\begin{aligned}
&= V_\infty\big[-\lambda^3 u^2 K(1 - u^2K^2) - \lambda^2(1 - u^2K^2) \\
&\qquad + 2\lambda(1 - u^2K^2)\big] \\
&= V_\infty(1 - u^2K^2)\lambda\left[(1-\lambda) + (1 - \lambda^2 u^2 K)\right] \qquad \text{(E.3)} \\
&\geq 0,
\end{aligned}$$

from inequalities (4.46), (4.47) and (4.48).

This concludes the proof of Proposition 4.5. □

E.2 Proof of Proposition 4.6

This proof is concerned with the variation of $\lim \mathrm{Var} f(t)$ with K and λ. Note the following at the outset:

1. m and $K = 1 - 1/\ddot{a}_{\overline{m|}}$ have a direct one-to-one relationship.
2. $\lim \mathrm{Var} f(t)$, in equation (4.60) in Proposition 4.4, is symmetrical in K and λ: they can be interchanged without changing the value of $\lim \mathrm{Var} f(t)$. Hence, any proof that $\partial[\lim \mathrm{Var} f(t)]/\partial K > 0$ will apply, except with all K and λ interchanged, to show that $\partial[\lim \mathrm{Var} f(t)]/\partial\lambda > 0$.

We use equations (D.80) and (D.83) from Appendix D and express

$$\lim_{t\to\infty} \mathrm{Var} f(t) = \sigma^2 v^2 A L^2 \varphi / Q, \tag{E.4}$$

where $\varphi = \varphi(1)$ in equation (D.71) and

$$Q = Q(1) = (1-q)\varphi + \gamma(1) + q(1-K)^2(1-\lambda)^2(1+up) \tag{E.5}$$

in equation (D.75). By replacing $\gamma(1)$ from equation (D.74), we may simplify

$$\begin{aligned}
&\gamma(1) + q(1-K)^2(1-\lambda)^2(1+up) \\
&= 2q(1-K)(1-\lambda)[1 - \lambda^2K^2u^2 - (1-K)(1-\lambda)] \\
&\quad + q(1-K)^2(1-\lambda)^2(1+\lambda K u^2) \\
&= q(1-K)(1-\lambda)[2 - 2\lambda^2K^2u^2 - 2(1-K)(1-\lambda) \\
&\qquad + (1-K)(1-\lambda)(1+\lambda K u^2)] \\
&= q(1-K)(1-\lambda)[2(1-\lambda^2K^2u^2) - (1-K)(1-\lambda)(1-\lambda K u^2)] \\
&= q(1-K)(1-\lambda)\rho \quad \text{(say)}.
\end{aligned} \tag{E.6}$$

Therefore,

$$Q = Q(1) = (1-q)\varphi + q(1-K)(1-\lambda)\rho, \tag{E.7}$$

and, in equation (E.4), we write

$$\lim_{t\to\infty} \mathrm{Var} f(t) = \sigma^2 v^2 A L^2 \varphi / [(1-q)\varphi + q(1-K)(1-\lambda\rho)]. \tag{E.8}$$

Now, note from equation (D.71) that

$$\varphi = 1 + K\lambda u^2(1-2\lambda) - K^2\lambda u^2(2-\lambda) + K^3\lambda^3u^4, \tag{E.9}$$

$$\partial\varphi/\partial K = \lambda u^2(1-2\lambda) - 2K\lambda u^2(2-\lambda) + 3K^2\lambda^3u^4. \tag{E.10}$$

From equation (E.6), ρ may also be expressed as a polynomial in K and differentiated w.r.t. K:

$$\rho = (1+\lambda) + K(1-\lambda)(1+\lambda u^2) - K^2\lambda(1+\lambda)u^2, \tag{E.11}$$

$$\partial\rho/\partial K = (1-\lambda)(1+\lambda u^2) - 2K\lambda(1+\lambda)u^2. \tag{E.12}$$

$\lim \mathrm{Var} f(t)$ is finite in the stable range imposed by stability conditions (4.46)–(4.48) and (4.58). In particular, $Q > 0$, so that when examining $\partial[\lim \mathrm{Var} f(t)]/\partial K$ we only need consider the 'numerator' of the differential of the right hand side of equation (E.4):

$$\begin{aligned}\partial[\lim \mathrm{Var} f(t)]/\partial K &\propto Q\partial\varphi/\partial K - \varphi\partial Q/\partial K\\ &= [(1-q)\varphi + q(1-K)(1-\lambda)\rho]\partial\varphi/\partial K\\ &\qquad - \varphi[(1-q)\partial\varphi/\partial K - q(1-\lambda)\rho + q(1-K)(1-\lambda)\partial\rho/\partial K]\\ &= q(1-\lambda)[(1-K)\rho\partial\varphi/\partial K + \varphi\rho - (1-K)\varphi\partial\rho/\partial K]\\ &\propto (1-K)\rho\partial\varphi/\partial K + \varphi\rho - (1-K)\varphi\partial\rho/\partial K.\end{aligned} \tag{E.13}$$

This may be expanded by replacing φ, $\partial\varphi/\partial K$, ρ and $\partial\rho/\partial K$ from equations (E.9), (E.10), (E.11) and (E.12) respectively, and by collecting terms in K to give

$$\begin{aligned}&[2\lambda - 2\lambda^3u^2] + K[2 - 2\lambda - 2\lambda^2u^2 + 2\lambda^3u^2] - K^2[2\lambda u^2 + 2\lambda^2u^2 - 2\lambda^3u^4 - 2\lambda^4u^4]\\ &\qquad - K^3[2\lambda^2u^4 - 2\lambda^3u^4 - 2\lambda^4u^6 + 2\lambda^5u^6] + K^4[2\lambda^2u^4 - 2\lambda^4u^6].\end{aligned} \tag{E.14}$$

Now $2(1-\lambda^2u^2)$ is a factor in each 'coefficient' of K above. $1-\lambda^2u^2 > 0$ from condition (4.48). After factoring out $2(1-\lambda^2u^2)$, we are left with

$$\lambda + K(1-\lambda) - K^2\lambda u^2(1+\lambda) - K^3\lambda^2u^4(1-\lambda) + K^4\lambda^2u^4. \tag{E.15}$$

Terms in λ may be collected to give:

$$\begin{aligned}&K + \lambda(1-K-K^2u^2) - \lambda^2K^2u^2(1+Ku^2-K^2u^2) + \lambda^3K^3u^4\\ &= \lambda(1-K) - \lambda^2K^3(1-K)u^4 + K[1-\lambda Ku^2 - \lambda^2Ku^2 + \lambda^3K^2u^4]\\ &= \lambda(1-K)[1-\lambda u^4K^3] + K[1-\lambda Ku^2][1-\lambda^2u^2K]\\ &\geq 0.\end{aligned} \tag{E.16}$$

The last inequality follows by virtue of conditions (4.47) and (4.48) and the definitions of $0 \leq \lambda < 1$ and $0 \leq K < 1$. ($\lambda u^4K^3 = \lambda u(Ku)^3 < 1$, $\lambda Ku^2 = (\lambda u)(Ku) < 1$, $\lambda^2u^2K = (\lambda u)^2K < 1$.) Equality in (E.16) occurs if $K = 0$ (or $m = 1$) *and* $\lambda = 0$.

Hence, we have shown that $\partial[\lim \mathrm{Var} f(t)]/\partial K > 0$ for $K > 0$ and $\lambda > 0$. By virtue of arguments 1. and 2. at the beginning of this section, we have therefore shown that, as m increases or as λ increases, $\lim \mathrm{Var} f(t)$ increases monotonically.

This concludes the proof of Proposition 4.6. □

E.3 Proof of Proposition 4.8

This proof is concerned with the variation of $\lim \mathrm{Varc}(t)$ with m and λ. Note at the outset that m and K have a direct one-to-one relationship ($\partial m/\partial K > 0$) so that we may investigate the variation of $\lim \mathrm{Varc}(t)$ with K rather than with m.

Q from equation (E.7) may be replaced into equation (4.62) from Proposition (4.4) so that

$$\lim_{t\to\infty} \mathrm{Varc}(t) = \sigma^2 v^2 A L^2 (1-K)^2 (1-\lambda)^2 (1+\lambda K u^2)/[(1-q)\varphi + q(1-K)(1-\lambda)\rho]. \quad (E.17)$$

We note that

$$\partial[(1-K)^2(1+\lambda K u^2)]/\partial K = (1-K)[\lambda u^2 - 2 - 3\lambda K u^2], \quad (E.18)$$

and also that

$$\begin{aligned}\partial Q/\partial K &= \partial[(1-q)\varphi + q(1-K)(1-\lambda)\rho]/\partial K \\ &= (1-q)\partial\varphi/\partial K - q(1-\lambda)\rho + q(1-\lambda)(1-K)\partial\rho/\partial K. \quad (E.19)\end{aligned}$$

Since the denominator is required to be positive ($Q > 0$) from stability condition (4.58), we need only consider the 'numerator' of the differential of the right hand side of equation (E.17):

$$\begin{aligned}&\partial[\lim\mathrm{Varc}(t)]/\partial K \\ &\quad\propto Q\partial[(1-K)^2(1+\lambda K u^2)]/\partial K - [(1-K)^2(1+\lambda K u^2)]\partial Q/\partial K \\ &\quad= [(1-q)\varphi + q(1-K)(1-\lambda)\rho](1-K)[\lambda u^2 - 2 - 3\lambda K u^2] \\ &\qquad- [(1-K)^2(1+\lambda K u^2)][(1-q)\partial\varphi/\partial K - q(1-\lambda)\rho + q(1-\lambda)(1-K)\partial\rho/\partial K], \\ &\qquad\qquad (E.20)\end{aligned}$$

by using equations (E.18), (E.19).

$1-K$ may be factored. We may also replace φ, $\partial\varphi/\partial K$, ρ and $\partial\rho/\partial K$ from equations (E.9), (E.10), (E.11) and (E.12) respectively. Upon collecting terms in K, we obtain

$$\begin{aligned}&[-2 + 2\lambda q + 2\lambda^2 u^2 - 2\lambda^3 q u^2] + K[2q - 2\lambda q - 2\lambda^2 q u^2 + 2\lambda^3 q u^2] \\ &\quad+ K^2[-2\lambda q u^2 + 2\lambda^2 q u^2 + 2\lambda^3 q u^4 - 2\lambda^4 q u^4] + K^3[2\lambda^2 q u^2 - 2\lambda^2 q u^2 + 2\lambda^2 u^4 - 2\lambda^2 q u^4 \\ &\qquad - 2\lambda^3 q u^4 + 2\lambda^4 q u^4 - 2\lambda^4 u^6 + 2\lambda^4 q u^6]. \quad (E.21)\end{aligned}$$

$2(1-\lambda^2 u^2) > 0$ from condition (4.48) and may be factored from each 'coefficient' of K in the above to give

$$-1 + \lambda q + K(1-\lambda)q - K^2\lambda(1-\lambda)q u^2 + K^3\lambda u^2[q - \lambda q + \lambda u^2 - \lambda q u^2]. \quad (E.22)$$

This becomes, after terms in λ are collected,

$$\pi(\lambda) = -(1-Kq) + \lambda(1-K)q(1-u^2K^2) + \lambda^2 K^2 u^2[u^2(1-Kq) + (1-K)\sigma^2]. \quad (E.23)$$

We have shown therefore that

$$\partial[\lim \mathrm{Var} c(t)]/\partial K \ \propto \ (1-K)2(1-\lambda^2u^2)\pi(\lambda), \tag{E.24}$$

where $1-K>0$ and $1-\lambda^2u^2>0$ by conditions (4.47) and (4.48). The sign of $\pi(\lambda)$ needs to be investigated.

Proof that $m \geq m^* \Rightarrow \lim \mathrm{Var} c(t)$ increases monotonically with m. It is easily proven that it is sufficient for $m \geq m^*$ (and $\lambda > 0$) for $\pi(\lambda) > 0$. The right hand side of equation (E.23) is easily rearranged into

$$\pi(\lambda) = -(1-Kq)[1-\lambda^2K^2u^4] + \lambda(1-K)q(1-u^2K^2) + \lambda^2K^2(1-K)u^2\sigma^2. \tag{E.25}$$

The last two terms in the right hand side of the above are positive (for $\lambda > 0$ and given condition (4.47)). The sign of the first term depends on $1-Kq$ (note that $1-\lambda^2K^2u^4 = 1-(\lambda u)^2(Ku)^2 > 0$, given conditions (4.47) and (4.48)). Hence, if $\lambda > 0$ and $m \geq m^*$ ($1-Kq \leq 0$) then $\pi(\lambda) > 0$. It follows therefore that $\lim \mathrm{Var} c(t)$ increases with increasing $m \geq m^*$.

Proof that $\lambda \geq \lambda^* \Rightarrow \lim \mathrm{Var} c(t)$ increases monotonically with m. Next, it can also be proven that it is sufficient for $\lambda \geq \lambda^*$ (and $m > 1$) for $\pi(\lambda) > 0$. Note that this has already been proven when $m \geq m^*$ in the previous paragraph. We need only consider $m < m^*$ therefore. Now, from equation (E.23), $\pi(\lambda)$ is quadratic in λ. When $1-Kq > 0$ (i.e. when $m < m^*$), the coefficient of λ^2 is positive, so that $\pi(\lambda)$ has a minimum. Further, $\pi(\lambda = 0) < 0$ and $\pi'(\lambda = 0) > 0$. If we show that $\pi(\lambda = \lambda^* > 0) > 0$, then it follows that $\pi(\lambda \geq \lambda^*) > 0$.

$$\begin{aligned}
&q^2\pi(\lambda = \lambda^* = 1/q)\\
&= u^2K^2[u^2(1-Kq) + (1-K)\sigma^2] + q^2(1-K)(1-u^2K^2) - q^2(1-Kq)\\
&= u^2K^2[u^2(1-Kq) + (1-K)\sigma^2] - q^2K + q^2K[q - u^2K(1-K)]\\
&= K\left[u^2K[u^2(1-Kq) + (1-K)\sigma^2] + q^2[q-1-u^2K(1-K)]\right]\\
&= K\left[u^4K - qu^4K^2 + K(1-K)\sigma^2u^2 + q^2(q-1) - q^2u^2K(1-K)\right]\\
&= K\left[u^2K[u^2 - qu^2K + (1-K)\sigma^2 - q^2(1-K)] + q^2(q-1)\right]\\
&= K\left[u^2K[-u^2(q-1) - (1-K)\sigma^2(q-1)] + q^2(q-1)\right]\\
&= K(q-1)\left[-u^2K(u^2 + (1-K)\sigma^2) + q^2\right]
\end{aligned}$$

$$\begin{aligned}
&> K(q-1)\left[-u^2K(u^2 + (1-K)\sigma^2) + u^2\right] \quad (\text{since } q^2 > u^2)\\
&= K(q-1)u^2\left[1 - K(u^2 + (1-K)\sigma^2)\right]\\
&\geq K(q-1)u^2[1-Kq] \quad (\text{since } u^2 + (1-K)\sigma^2 \leq q)\\
&> 0 \quad (\text{when } m < m^*).
\end{aligned}$$

Hence, we have shown that $\pi(\lambda \geq \lambda^*) > 0$ or that $\lim \text{Var}c(t)$ increases monotonically with increasing m provided $\lambda \geq \lambda^*$.

Proof that $\lim \text{Var}c(t)$ **has at least one minimum, as** m **increases, at some** $m^{*(\lambda)} < m^*$, **provided** $0 < \lambda < \lambda^*$. $\pi(\lambda)$ in equation (E.23) may be rearranged into a cubic polynomial in K:

$$\pi(K) = -(1 - \lambda q) + Kq(1 - \lambda) - K^2\lambda(1 - \lambda)qu^2 + K^3\lambda u^2(q - \lambda qu^2 - \lambda\sigma^2). \tag{E.26}$$

If $\lambda < \lambda^* = 1/q$, then it is clear that $\pi(K = 0) < 0$. In addition, $\pi'(K = 0) > 0$ and $\pi''(K = 0) < 0$. We have shown already that $\pi(K = K^* = 1/q) \geq 0$ from equation (E.25). $\pi(K)$ in equation (E.26) is a cubic in K. Hence, if $\lambda < \lambda^*$, $\pi(K)$ must have at least one root at some $K = K^{*(\lambda)}$, such that $0 < K^{*(\lambda)} < 1/q$ and such that $\pi'(K = K^{*(\lambda)}) > 0$, i.e. $\pi(K)$ must have at least one minimum at $0 < K^{*(\lambda)} < 1/q$.

Hence, we have shown that, if $0 < \lambda < \lambda^*$, as m increases, $\lim \text{Var}c(t)$ has at least one minimum at some $1 < m^{*(\lambda)} < m^*$.

Symmetry between K **and** λ**.** A consequence of the symmetry between K and λ is that $\lim \text{Var}c(t)$ remains unchanged in equation (4.62) if K and λ are interchanged. Therefore, the proof above concerning the variation of $\lim \text{Var}c(t)$ with m (or K) applies when we consider the variation with λ, except that all K and λ are exactly interchanged. The second part of Proposition 4.8 is therefore exactly symmetrical with the first part.

This concludes the proof of Proposition 4.8. □

Appendix F

'Dual-interest' and 'Integral Spreading' Methods

F.1 Proof of Proposition 5.1

This proof is concerned with the 'dual-interest' method of §5.5. It is a straightforward application of the method employed by Dufresne (1986, 1988). Since $\{i(t)\}$ is a sequence of independent identically distributed random variables, it follows that $u(t+1)$ in equation (5.14) is independent of $u(t)$, $u(t-1)$ etc. and also of $ul(t)$, $ul(t-1)$ etc. Upon taking expectations on both sides of equation (5.14), it is clear that

$$\mathrm{E}f(t+1) \;=\; u(1-k_1)\mathrm{E}f(t) + uAL(k_1 - d_r), \tag{F.1}$$

where $k_1 = k + \kappa$. Hence, provided that $|u(1-k_1)| < 1$, taking limits as $t \to \infty$ yields equation (5.18). By definition,

$$ul(t) \;=\; AL - f(t), \tag{F.2}$$

and taking expectations and limits as $t \to \infty$ on equation (F.2) and using equation (5.18), $\lim \mathrm{E}ul(t)$ in equation (5.19) follows readily. $\lim \mathrm{E}c(t)$ in equation (5.20) is obtained by taking expectation and limits as $t \to \infty$ on equation (5.10) and using equation (5.19).

The second moments are obtained most easily by following the method of Dufresne (1986, 1988). When equation (F.1) is subtracted from equation (5.14), it is found that

$$f(t+1) - \mathrm{E}f(t+1) = u(t+1)[f(t) - \mathrm{E}f(t)](1-k_1) + (u(t+1) - u)Y(t), \tag{F.3}$$

where

$$Y(t) = (k_1 - d_r)AL + (1-k_1)\mathrm{E}f(t). \tag{F.4}$$

Further, when both sides of equation (F.3) are squared,

$$[f(t+1) - \mathrm{E}f(t+1)]^2 = u(t+1)^2[f(t) - \mathrm{E}f(t)]^2(1-k_1)^2 + (u(t+1)-u)^2Y(t)^2 + 2u(t+1)(u(t+1)-u)[f(t) - \mathrm{E}f(t)]Y(t)(1-k_1). \quad \text{(F.5)}$$

It is clear that $u(t+1)$ is again independent of $\{u(s)\}$, $\{f(s)\}$ for $s \le t$. Taking expectations as $t \to \infty$ on both sides of equation (F.5) yields

$$\mathrm{Var}f(t+1) = q(1-k_1)^2\mathrm{Var}f(t) + \sigma^2 Y(t)^2. \quad \text{(F.6)}$$

Taking expectations across equation (F.4) and then taking limit as $t \to \infty$ and applying equation (5.18), it follows that

$$\lim_{t\to\infty} Y(t) = (k_1 - d_r)AL + (1-k_1)\lim_{t\to\infty}\mathrm{E}f(t) = ALv(d_r - k_1)/(d - k_1). \quad \text{(F.7)}$$

Thus, if $q(1-k_1)^2 < 1$ in equation (F.6), it follows that

$$\lim_{t\to\infty}\mathrm{Var}f(t) = \sigma^2 \lim Y(t)^2/[1 - q(1-k_1)^2]. \quad \text{(F.8)}$$

Replacing $\lim Y(t)$ in equation (F.7) into equation (F.8), the result in equation (5.22) follows. $\lim \mathrm{Var}c(t)$ is obtained by taking variance and limits as $t \to \infty$ on both sides of equation (5.10) of course.

This concludes the proof of Proposition 5.1. □

F.2 Proof of Proposition 5.2

Define $ul^u(t) = ul(t) - U(t)$. We note that $u(t+1)$ is independent of $\{u(s)\}$ and $\{ul(s)\}$ for $s \le t$. Taking expectations across equation (5.34),

$$\mathrm{E}ul^u(t+1) = u(1-k_p)\mathrm{E}ul^u(t) - uk_i\sum_{j=0}^{t}\mathrm{E}ul^u(j) + (1 - uv_v)(AL - U(t+1)), \quad \text{(F.9)}$$

for $t \ge 0$. Forward-shift equation (F.9) in time (so that it holds for $t \ge -1$) and deduct equation (F.9) to obtain

$$\mathrm{E}ul^u(t+2) - [1 - u(1-k_p-k_i)]\mathrm{E}ul^u(t+1) + u(1-k_p)\mathrm{E}ul^u(t) = (1-uv_v)(U(t+1) - U(t+2)). \quad \text{(F.10)}$$

The characteristic equation and stability conditions for difference equation (F.10) are identical to those of equation (5.37). Assuming these conditions hold, then

$$\lim_{t\to\infty}\mathrm{E}ul(t) = \lim_{t\to\infty}\mathrm{E}ul^u(t) = 0 \quad \Leftrightarrow \quad \lim_{t\to\infty}\mathrm{E}f(t) = AL, \quad \text{(F.11)}$$

since $U(t) \to 0$, as $t \to \infty$. The rests of the proof proceeds as in §5.6.2. □

F.3 Proof of Proposition 5.3

A Special Case of §D.1

This proof is concerned with the second moments of the pension funding process when gains and losses are adjusted according to the method described by equation (5.32). Comparison of equations (5.32) and (D.8) reveals that this can be considered to be a special case of the general model described in §D.1.

We may therefore consider a special case of the general situation described in §D.1 by letting

$$k_1 = k_p, \qquad k_2 = k_i, \qquad k_3 = 0, \qquad p = 1. \tag{F.12}$$

Equation (D.2) is now

$$adj(t) = k_p ul^u(t) + k_i \sum_{j=0}^{t} ul^u(j) + P(t), \tag{F.13}$$

and equation (D.14) becomes

$$\begin{aligned} &ul^u(t+1) - (AL - U(t+1)) \\ &\qquad = u(t+1)\left[ul^u(t)(1-k_p) - k_i \sum_{j=0}^{t} ul^u(j) - v_v(AL - U(t+1))\right]. \end{aligned} \tag{F.14}$$

This is the same as equation (5.34).

We have shown in Proposition 5.2 and Corollary 5.2 that, provided certain conditions hold,

$$\lim_{t\to\infty} \mathrm{E}ul(t) = \lim_{t\to\infty} \mathrm{E}ul^u(t) = 0, \tag{F.15}$$

$$\lim_{t\to\infty} \mathrm{E}adj(t) = AL(d_v - d). \tag{F.16}$$

From equation (5.58),

$$\lim_{t\to\infty} \sum_{j=0}^{t} \mathrm{E}ul^u(j) = AL(1 - uv_v)/uk_i = AL(d_v - d)/k_i. \tag{F.17}$$

Mean Square Deviation of $ul(t)$ and Variance of $f(t)$

In the special case given by equations (F.12),

$$\theta_t = AL(1 - uv_v), \tag{F.18}$$

$$P(E) = (E - u(1 - k_p))(E - 1) + Euk_i, \tag{F.19}$$

$$\psi(E) = (E - q(1 - k_p)^2)(E - 1) + Eqk_i^2, \tag{F.20}$$

$$\gamma(E) = 2qk_i(1 - k_p)(E - 1) - 2qk_i^2 E, \tag{F.21}$$

$$\nu(E) = 2AL(E - qv_v(1 - k_p))(E - 1)^2 + 2qv_v k_i ALE(E - 1), \tag{F.22}$$

$$\varphi(E) = (E - u(1 - k_p))(E - 1) + 2uk_i E, \tag{F.23}$$

$$\mu(E) = E(E - 1) + uk_i E, \tag{F.24}$$

$$\omega(E) = AL(1 - uv_v)E(E - 1), \tag{F.25}$$

$$\begin{aligned} Q(E) = {} & [E - q(1 - k_p^2)][(E - u(1 - k_p))(E - 1) + 2uk_i E] \\ & + E2qk_i[(1 - k_p)(E - 1) - k_i E] + Eqk_i^2[E + u(1 - k_p)]. \end{aligned} \tag{F.26}$$

Hence, equation (D.44) may be written as

$$\begin{aligned} & P(E)Q(E)b_t \\ & = (qv_v^2 - 1)AL^2\varphi(E)P(E) \\ & \quad + 2AL^2(1 - uv_v)\Big[[(E - qv_v(1 - k_p))(E - 1) + qv_v k_i E]\,[(E - u(1 - k_p))(E - 1) + 2uk_i E] \\ & \qquad - (1 - uv_v)Eqk_i\,[(1 - k_p)(E - 1) - k_i E] \Big], \end{aligned} \tag{F.27}$$

while equation (D.48) becomes

$$\begin{aligned} P(E)Q(E)c_t = {} & (qv_v^2 - 1)AL^2\mu(E)P(E) \\ & + AL^2(1 - uv_v)\Big[[2(E - qv_v(1 - k_p))(E - 1) + 2qv_v k_i]\,[E(E - 1) + Euk_i] \\ & \qquad + (1 - uv_v)\,[(E - q(1 - k_p)^2)(E - 1) + Eqk_i^2] \Big]. \end{aligned} \tag{F.28}$$

Equations (F.27) and (F.28) are linear difference equations in b_t and c_t respectively. Note that $\varphi(1) = 2uk_i$, $P(1) = \mu(1) = uk_i$ and

$$Q(1) = 2uk_i(1 - q(1 - k_p)^2) - qk_i^2(1 - u(1 - k_p)). \tag{F.29}$$

Assume that the limit of b_t as $t \to \infty$ exists. Then, from equation (F.27),

$$\lim_{t\to\infty} b_t = \frac{(qv_v^2 - 1)AL^2 2u^2k_i^2 + 2AL^2(1 - uv_v)\left[qv_v k_i 2uk_i + (1 - uv_v)qk_i^2\right]}{uk_i\left[2uk_i(1 - q(1 - k_p)^2) - qk_i^2(1 - u(1 - k_p))\right]} \tag{F.30}$$

$$= \frac{AL^2 2u^2k_i^2\left[qv_v^2 - 1 + qv^2(1 - uv_v)(1 + uv_v)\right]}{2uk_i uk_i\left[1 - q(1 - k_p)^2 - qvk_i(1 - u(1 - k_p))/2\right]}. \tag{F.31}$$

Since $qv_v^2 - 1 + qv^2(1-uv_v)(1+uv_v) = qv^2 - 1 = \sigma^2 v^2$,

$$\lim_{t\to\infty} b_t = \frac{\sigma^2 v^2 AL^2}{1 - q(1-k_p)^2 - qk_i(k_p - d)/2} = V_\infty \quad \text{(say)}. \tag{F.32}$$

Therefore,

$$\begin{aligned}\lim_{t\to\infty} \mathrm{Var} f(t) &= \lim_{t\to\infty} \mathrm{Var} ul(t) \quad \text{(given equation (D.3))} \\ &= \lim_{t\to\infty} \mathrm{Var} ul^u(t) \\ &= \lim_{t\to\infty} \mathrm{E} ul^u(t)^2 \quad \text{(using equation (F.15))} \\ &= V_\infty \quad \text{(from equation (F.32))}. \end{aligned} \tag{F.33}$$

Covariance

Assume also that the limit of c_t as $t \to \infty$ exists, then from equation (F.28),

$$\begin{aligned}\lim_{t\to\infty} c_t &= \frac{(qv_v^2 - 1)AL^2u^2k_i^2 + AL^2(1-uv_v)\left[2qv_vk_iuk_i + (1-uv_v)qk_i^2\right]}{uk_i\left[2uk_i(1-q(1-k_p)^2) - qk_i^2(1-u(1-k_p))\right]} \\ &= \lim_{t\to\infty} b_t/2 \quad \text{(by comparison with equation (F.30))} \\ &= V_\infty/2. \end{aligned} \tag{F.34}$$

Hence, assuming the limit exists,

$$\begin{aligned}\lim_{t\to\infty} \mathrm{Cov}\left[ul^u(t), \sum_{j=0}^{t} ul^u(j)\right] &= \lim_{t\to\infty} c_t - \lim_{t\to\infty} \mathrm{E}ul^u(t) \lim_{t\to\infty} \sum_{j=0}^{t} \mathrm{E}ul^u(j) \\ &= V_\infty/2, \end{aligned} \tag{F.35}$$

upon using equations (F.15), (F.17) and (F.34).

We can now calculate the covariance between contribution and fund levels:

$$\begin{aligned}\mathrm{Cov}[f(t), c(t)] &= -\mathrm{Cov}[ul(t), adj(t)] \quad \text{(from equations (D.3) and (D.4))} \\ &= -\mathrm{Cov}\left[ul^u(t) + U(t), k_p ul^u(t) + k_i \sum_{j=0}^{t} ul^u(j) + P(t)\right], \end{aligned}$$

(where we have used equation (F.13))

$$= -k_p \mathrm{Var} ul^u(t) - k_i \mathrm{Cov}\left[ul^u(t), \sum ul^u(j)\right]. \tag{F.36}$$

Assuming limits exist, and using equations (F.33) and (F.35),

$$\lim_{t\to\infty} \mathrm{Cov}[f(t), c(t)] = -(k_p + k_i/2)V_\infty. \tag{F.37}$$

Variance of $c(t)$

Other results can also be obtained. Under the special case defined in equations (F.12), equation (D.21) becomes

$$c_{t+1} - u(1-k_p)c_t + uk_i d_t - b_{t+1} = (1-uv_v)(AL - U(t+1))\sum_{j=0}^{t} a_j. \tag{F.38}$$

By taking limits (assuming they exist) on equation (F.38) and using equations (F.17), (F.32) and (F.34),

$$\begin{aligned} uk_i \lim_{t\to\infty} d_t &= \lim b_t - (1-u(1-k_p))\lim c_t + AL^2(1-uv_v)^2/uk_i \\ &= [1-(1-u(1-k_p))/2]V_\infty + AL^2(1-uv_v)^2/uk_i \\ &= (1+u(1-k_p))V_\infty/2 + AL^2(1-uv_v)^2/uk_i \qquad \text{(F.39)} \\ k_i^2 \lim_{t\to\infty} d_t &= k_i(2-k_p-d)V_\infty/2 + AL^2(d_v-d)^2. \qquad \text{(F.40)} \end{aligned}$$

Therefore,

$$\begin{aligned} k_i^2 \lim_{t\to\infty} \mathrm{Var}\sum_{j=0}^{t} ul^u(j) &= k_i^2 \lim_{t\to\infty} d_t - k_i^2 \lim_{t\to\infty}\left[\sum_{j=0}^{t} \mathrm{E}ul^u(j)\right]^2 \\ &= k_i(2-k_p-d)V_\infty/2 \quad \text{(from equations (F.17) and (F.40)).} \qquad \text{(F.41)} \end{aligned}$$

Finally,

$$\begin{aligned} \mathrm{Var}c(t) &= \mathrm{Var}adj(t) \\ &= k_p^2\mathrm{Var}ul^u(t) + 2k_pk_i\mathrm{Cov}\left[ul^u(t), \sum_{j=0}^{t} ul^u(j)\right] + k_i^2\mathrm{Var}\sum_{j=0}^{t} ul^u(j), \qquad \text{(F.42)} \end{aligned}$$

where we have used equations (D.4) and (F.13). Assuming limits exist,

$$\lim_{t\to\infty} \mathrm{Var}c(t) = [k_p^2 + k_i + k_i(k_p-d)/2]V_\infty, \tag{F.43}$$

using equations (F.33), (F.35) and (F.41).

Stability Conditions

We now need to investigate the conditions for which these limits exist. From Corollary 5.2, for the limits in equations (F.15)–(F.17) to exist, it is sufficient that:

$$\begin{aligned} i &> -100\%, \qquad \text{(F.44)} \\ 0 \le u(1-k_p) &< 1, \qquad \text{(F.45)} \\ 0 < k_i &< 2(1-d+1-k_p). \qquad \text{(F.46)} \end{aligned}$$

Furthermore, the limits of b_t and c_t as $t \to \infty$ exist if and only if the magnitude of the roots of $P(z)Q(z) = 0$ are less than unity, from the right hand sides of equations (F.27) and (F.28) respectively. But $P(z)$ is the characteristic polynomial that determines first moment stability, as seen in equation (D.28) (as well as in equation (5.38)). It remains to consider the roots of $Q(z) = 0$.

$Q(z)$, in equation (F.26), can be expanded into a standard cubic form:

$$Q(z) = z^3 - Az^2 + Bz - C, \tag{F.47}$$

where

$$A = 1 + q(1 - k_p - k_i)^2 + u(1 - k_p - 2k_i), \tag{F.48}$$

$$B = [u + uq(1 - k_p - k_i)^2 + q(1 - k_p - 2k_i)](1 - k_p), \tag{F.49}$$

$$C = uq(1 - k_p)^3. \tag{F.50}$$

The coefficient of z^3 in $Q(z)$ is unity. Necessary and sufficient conditions for the roots of $Q(z) = 0$ to be less than one in magnitude—using, for example, the criteria of Jury (1962, 1964:136) or Marden (1966)— are that:

$$|C| < 1, \tag{F.51}$$

$$|A + C| < 1 + B, \tag{F.52}$$

$$|AC - B| < |1 - C^2|. \tag{F.53}$$

We now examine each in turn.

We can show that if conditions (F.44)–(F.46) hold as well as if

$$0 < k_i < 2u[1 - q(1 - k_p)^2]/q[1 - u(1 - k_p)] = \kappa_i \quad \text{(say)} \tag{F.54}$$

and

$$[1 + q(1 - k_p)^2][1 - qu^2(1 - k_p)^4] + u(1 - k_p)[1 - q(1 - k_p)^2][1 + q(1 - k_p - k_i)^2] > 2q(1 - k_p)k_i[1 - u^2(1 - k_p)^2], \tag{F.55}$$

then conditions (F.51), (F.52) and (F.53) are true. First observe that conditions (F.44), (F.45) and (F.54) imply that

$$q(1 - k_p)^2 < 1. \tag{F.56}$$

Condition (F.51)

Since $|C| = q(1 - k_p)^2|u(1 - k_p)|$, it is sufficient that inequalities (F.45) and (F.56) (and hence condition (F.54)) hold for condition (F.51) to be true.

Condition (F.52)

First, consider $A + C < 1 + B$. This readily simplifies to

$$qk_i^2 - 2uk_i \; < \; k_i^2 qu(1-k_p) - 2k_i uq(1-k_p)^2, \tag{F.57}$$

$$k_i[qk_i(1-u(1-k_p)) - 2u(1-q(1-k_p)^2)] \; < \; 0. \tag{F.58}$$

It is clear that $0 < k_i < \kappa_i$, as in condition (F.54), implies that $A + C < 1 + B$.

Next, consider $A + C > -(1 + B)$. Upon expanding and collecting terms in k_i, we have

$$k_i^2 q[1 + u(1-k_p)] - k_i[2u(1 + q(1-k_p)^2) + 4q(1-k_p)] \\ + [2 + 2q(1-k_p)^2 + 2u(1-k_p) + 2uq(1-k_p)^3] \; > \; 0, \tag{F.59}$$

and the term independent of k_i can be factorised as $2[1 + q(1-k_p)^2][1 + u(1-k_p)]$. Thus, $A + C > -(1 + B)$ reduces to a quadratic inequality in k_i:

$$k_i^2 q[1 + u(1-k_p)] - k_i[2u(1 + q(1-k_p)^2) + 4q(1-k_p)] \\ + 2[1 + q(1-k_p)^2][1 + u(1-k_p)] \; > \; 0, \tag{F.60}$$

where the coefficient of k_i^2 is positive (given condition (F.45)). Now, the discriminant of the quadratic on the left hand side of inequality (F.60) is

$$\begin{aligned} &4u^2[1 + q(1-k_p)^2]^2 + 16uq(1-k_p)[1 + q(1-k_p)^2] + 16q^2(1-k_p)^2 \\ &\qquad - 8q[1 + q(1-k_p)^2][1 + u(1-k_p)]^2 \\ &= 4[1 + q(1-k_p)^2][u^2 - qu^2(1-k_p)^2 - 2q] + 16q^2(1-k_p)^2 \\ &= 4u^2[1 + q(1-k_p)^2][1 - q(1-k_p)^2] - 8q[1 - q(1-k_p)^2] \\ &= 4[1 - q(1-k_p)^2][u^2 + u^2(1-k_p)^2 q - 2q] \\ &= -4[1 - q(1-k_p)^2][\sigma^2 + q(1 - u^2(1-k_p)^2)] \\ &< 0, \end{aligned}$$

given conditions (F.45) and (F.56) (or condition (F.54)). Therefore the quadratic inequality (F.60) holds.

Hence, conditions (F.45) and (F.54) are sufficient for condition (F.52) to be true.

Condition (F.53)

$AC - B < 1 - C^2$ may be written in the form of condition (F.55).

Next, consider $AC - B > -(1 - C^2)$. Upon expanding and collecting terms in k_i, we obtain

$$uq(1-k_p)^3[q(1-k_p)^2 + u(1-k_p)] - (1-k_p)[u + q(1-k_p)] + 1 - [uq(1-k_p)^3]^2 \\ - 2k_i[uq(1-k_p)^3(q(1-k_p) + u) - q(1-k_p)(1 + u(1-k_p))] \\ - k_i^2 qu(1-k_p)[1 - q(1-k_p)^2] \; > \; 0. \tag{F.61}$$

The term independent of k_i factors to $[1-u(1-k_p)][1-q(1-k_p)^2][1-q(1-k_p)^2u(1-k_p)]$. $AC-B > -(1-C^2)$ therefore reduces to a quadratic inequality in k_i:

$$k_i^2qu(1-k_p)[1-q(1-k_p)^2]-2k_iq(1-k_p)[u(1-k_p)[1-q(1-k_p)^2]+1-u^2(1-k_p)^2] \\ -[1-u(1-k_p)][1-q(1-k_p)^2][1-q(1-k_p)^2u(1-k_p)] \; < \; 0. \quad \text{(F.62)}$$

Let the left hand side of inequality (F.62) be $\pi(k_i)$. The coefficient of k_i^2 is positive if condition (F.45) and inequality (F.56) (or condition (F.54)) hold, and $\pi(k_i)$ must then have a minimum. The constant term (independent of k_i) is negative if condition (F.45) and inequality (F.56) (or condition (F.54)) are true, so that $\pi(k_i=0)<0$. $\pi(k_i)=0$ has two real roots α and β with $\alpha<0<\beta$ and $\pi(k_i)<0$ for $k_i\in(\alpha,\beta)$. κ_i is defined in inequality (F.54) and $\kappa_i>0$ under condition (F.45) and inequality (F.56) (or condition (F.54)). By disregarding the constant term in inequality (F.62),

$$\begin{aligned}
&\pi(k_i=\kappa_i)/\kappa_i \\
&\quad < \kappa_iqu(1-k_p)[1-q(1-k_p)^2]-2q(1-k_p)[u(1-k_p)(1-q(1-k_p)^2)+1-u^2(1-k_p)^2] \\
&\quad < \kappa_iqu(1-k_p)[1-q(1-k_p)^2]-2q(1-k_p)[1+u(1-k_p)][1-q(1-k_p)^2]
\end{aligned}$$

(since $q=u^2+\sigma^2>u^2$ and $1-u^2(1-k_p)^2>1-q(1-k_p)^2$)

$$= 2u^2(1-k_p)[1-q(1-k_p)^2]^2/[1-u(1-k_p)] \; - \; 2q(1-k_p)[1+u(1-k_p)][1-q(1-k_p)^2]$$

(after replacing κ_i from equation (F.54))

$$\begin{aligned}
&= [u^2(1-q(1-k_p)^2)-q(1-u^2(1-k_p)^2)]2(1-k_p)[1-q(1-k_p)^2]/[1-u(1-k_p)] \\
&= -2(1-k_p)[1-q(1-k_p)^2]\sigma^2/[1-u(1-k_p)] \\
&< 0, \qquad \text{(F.63)}
\end{aligned}$$

under condition (F.45) and inequality (F.56) (or condition (F.54)). Therefore, $\alpha<0<\kappa_i<\beta$; and for $0\le k_i\le\kappa_i$, $\pi(k_i)<0$ or $AC-B>-(1-C^2)$.

Hence, condition (F.53) is true provided conditions (F.45) and (F.54) hold.

Sufficient Conditions

We have shown that it is sufficient that conditions (F.44), (F.45), (F.46), (F.54) and (F.55) hold for the limits in equations (F.30)–(F.43) to exist.

This concludes the proof of Proposition 5.3. □

F.4 Proof of Proposition 5.4

This proof is concerned with the ultimate mean square unfunded liability when integral spreading is used according to the method described in equation (5.32).

Define the following:

$$CS = \lim_{t\to\infty} \mathrm{E}ul(t)^2/AL^2 \text{ when Conventional Spreading is used } (k_i = 0), \tag{F.64}$$

$$IS = \lim_{t\to\infty} \mathrm{E}ul(t)^2/AL^2 \text{ when Integral Spreading is used } (k_i \neq 0). \tag{F.65}$$

The second non-central moment of the unfunded liability when surpluses and deficits are spread forward is found in equation (5.8):

$$CS = \Delta^2/(d-k)^2 + \sigma^2 v^2[1 + \Delta/(k-d)]^2/[1 - q(1-k)^2], \tag{F.66}$$

$$\Delta = 0 \quad \Rightarrow \quad CS = \sigma^2 v^2/[1 - q(1-k)^2]. \tag{F.67}$$

From equation (5.62),

$$IS = \sigma^2 v^2/[1 - q(1-k)^2 - qk_i(k-d)/2]. \tag{F.68}$$

First, consider the last part of Proposition 5.4. Since $k > d$ from stability condition (5.55) and for $\Delta \geq 0$, it is clear that CS in equation (F.66) is smallest when $\Delta = 0$. It is also immediately clear from equation (F.68) that as $k_i \to 0$ (or $m_i \to \infty$), $IS \to \sigma^2 v^2/[1-q(1-k)^2]$. Integral spreading (for any Δ) becomes as efficient as conventional spreading *as if* $\Delta = 0$.

Next, note that IS in equation (F.68) decreases monotonically as k_i decreases (or m_i increases), given conditions (5.54) and (5.55), and since the coefficient of k_i in the denominator of the right hand side of equation (F.68) is negative. The longer m_i is, the more efficient integral spreading becomes.

When $\Delta = 0$, CS in equation (F.67) is less than IS in equation (F.68) (except at $m_i \to \infty$). Conventional spreading is therefore more efficient.

When $\Delta > 0$, CS in equation (F.66) is greater than $\sigma^2 v^2/[1-q(1-k)^2]$. But as m_i increases from m_i^{min} tending to $+\infty$, IS decreases monotonically from $+\infty$ tending to $\sigma^2 v^2/[1 - q(1 - k)^2]$. Therefore, there exists some $\{m_i\}$ such that $IS \leq CS$. This inequality is linear in k_i, and can be rewritten, using equations (F.66) and (F.68), as:

$$k_i \leq k_i^f, \tag{F.69}$$

where

$$k_i^f = \left\{2(d-d_v)[1-q(1-k)^2][\sigma^2 v^2(k-d_v+k-d) + (d-d_v)(1-q(1-k)^2)]\right\} \Big/ \left\{(k-d)q[\sigma^2 v^2(k-d_v)^2 + (d-d_v)^2(1-q(1-k)^2)]\right\}. \tag{F.70}$$

m_i^f in equation (5.72) is the reciprocal of k_i^f in equation (F.70).

This concludes the proof of Proposition 5.4. □

F.5 Proof of Proposition 5.5

This proof is concerned with the ultimate mean square supplementary contribution or contribution adjustment when integral spreading is used according to the method described in equation (5.32).

Define the following:

$$CS_1 = \lim_{t\to\infty} \mathrm{E}adj(t)^2/AL^2 \text{ when } \mathit{C}\text{onventional } \mathit{S}\text{preading is used } (k_i = 0), \quad \text{(F.71)}$$

$$IS_1 = \lim_{t\to\infty} \mathrm{E}adj(t)^2/AL^2 \text{ when } \mathit{I}\text{ntegral } \mathit{S}\text{preading is used } (k_i \neq 0). \quad \text{(F.72)}$$

The second non-central moment of the supplementary contribution when surpluses and deficits are spread forward is found in equation (5.9):

$$CS_1 = k^2\Delta^2/(d-k)^2 \ + \ \sigma^2 v^2 k^2 [1 + \Delta/(k-d)]^2/[1-q(1-k)^2], \quad \text{(F.73)}$$

$$\Delta = 0 \quad \Rightarrow \quad CS_1 = \sigma^2 v^2 k^2/[1-q(1-k)^2]. \quad \text{(F.74)}$$

From equations (5.50) and (5.63),

$$IS_1 = \Delta^2 + \sigma^2 v^2 [k^2 + k_i + k_i(k-d)/2]/[1-q(1-k)^2 - qk_i(k-d)/2]. \quad \text{(F.75)}$$

First, consider the last part of Proposition 5.5. Since $k > d$ from stability condition (5.55) and for $\Delta \geq 0$, it is clear that CS_1 in equation (F.73) is smallest when $\Delta = 0$. From equation (F.75), as $k_i \to 0$ (or $m_i \to \infty$), $IS_1 \to \Delta^2 + \sigma^2 v^2 k^2/[1-q(1-k)^2]$.

Next, note that IS_1 in equation (F.75) decreases monotonically as k_i decreases (or m_i increases). Given conditions (5.54) and (5.55), the coefficient of k_i in the denominator of the right hand side of equation (F.75) is negative, while the coefficient of k_i in the numerator is positive. The longer m_i is, the more efficient integral spreading becomes. As m_i increases from m_i^{min} tending to $+\infty$, IS_1 decreases monotonically from $+\infty$ tending to $\Delta^2 + \sigma^2 v^2 k^2/[1-q(1-k)^2]$.

Since $v > 0$ and $k > d$ from stability conditions (5.54) and (5.55) respectively, then for $k_i > 0$ or $m_i < \infty$,

$$\sigma^2 v^2 [k^2 + k_i + k_i(k-d)/2] \ > \ \sigma^2 v^2 k^2, \text{ and}$$
$$1 - q(1-k)^2 - qk_i(k-d)/2 \ < \ 1 - q(1-k)^2,$$

so that $IS_1 > \sigma^2 v^2 k^2/[1-q(1-k)^2]$. When $\Delta = 0$, $IS_1 > CS_1$, and conventional spreading is more efficient.

Suppose $\Delta > 0$. The first term on the right hand side of equation (F.73) is larger than Δ^2, since $k/(k-d) > 1$ (given conditions (5.2) and (5.3)). Furthermore, $[1+\Delta/(k-d)]^2 > 1$ when $\Delta > 0$ (since $k > d$ given condition (5.3)). Therefore,

$$CS_1 \ > \ \sigma^2 v^2 k^2/[1-q(1-k)^2] + \Delta^2 \ = \ \lim_{k_i \to 0+} IS_1. \quad \text{(F.76)}$$

But as m_i increases from m_i^{min} tending to $+\infty$, IS_1 decreases monotonically from $+\infty$ tending to $\Delta^2 + \sigma^2 v^2 k^2/[1-q(1-k)^2]$. Therefore, there exists some $\{m_i\}$ such that $IS_1 \leq CS_1$. This inequality is linear in k_i, and can be rewritten, using equations (F.73) and (F.75), as:

$$k_i \leq k_i^c, \tag{F.77}$$

where

$$\begin{aligned} k_i^c = & \Big\{ 2(d-d_v)[1-q(1-k)^2][\sigma^2 v^2 (k-d_v+k-d)k^2 + (d-d_v)(1-q(1-k)^2)d(2k-d)] \Big\} \Big/ \\ & \Big\{ (k-d)q[\sigma^2 v^2 (k-d_v)^2 k^2 + (d-d_v)^2 (1-q(1-k)^2) d(2k-d)] \\ & + \sigma^2 v^2 (k-d)^2 (2+k-d)[1-q(1-k)^2] \Big\}. \end{aligned} \tag{F.78}$$

m_i^c in equation (5.73) is the reciprocal of k_i^c in equation (F.78).

By comparing equations (F.68) with (F.75) as well as (F.66) with (F.73), it is easily found that

$$IS_1 = [k^2 + k_i + k_i(k-d)/2]IS + \Delta^2, \tag{F.79}$$

$$CS_1 = k^2 CS. \tag{F.80}$$

Now, assume that $m_i = m_i^c$. Then, $IS_1 = CS_1$ or

$$[k_i + k_i(k-d)/2]IS + k^2(IS - CS) + \Delta^2 = 0. \tag{F.81}$$

Since the first and third term on the left hand side of the above are positive, it follows that $IS < CS$ or $m_i = m_i^c > m_i^f$.

This concludes the proof of Proposition 5.5. □

References

Aaron, H.J. (1966). The social insurance paradox. *Canadian Journal of Economics and Political Science*, 32, 371–374.

Adams, W.R. (1967). The effect of interest on pension contributions. *Transactions of the Society of Actuaries*, 19, 170–193.

Aitken, W.H. (1994). *A Problem-Solving Approach to Pension Funding and Valuation*, ACTEX Publications, Winsted, Connecticut.

Allison, G.D. & Winklevoss, H.E. (1975). The interrelationship among inflation rates, salary rates, interest rates and pension costs. *Transactions of the Society of Actuaries*, 27, 197–210.

Alphen, J. van *et al.* (1997). ALM products compared. *Proceedings of the 7th AFIR International Colloquium*, Cairns, Australia, 1, 99–118.

American Academy of Actuaries (1993). *Actuarial Standard of Practice No. 4: Measuring Pensions Obligations.* Pensions Committee of the Actuarial Standards Board, American Academy of Actuaries, Washington, DC.

American Academy of Actuaries (1996). *Actuarial Standard of Practice No. 27: Selection of Economic Assumptions for Measuring Pension Obligations.* Pensions Committee of the Actuarial Standards Board, American Academy of Actuaries, Washington, DC.

Ammeter, H. (1963). Uber die aquivalenz von finanzierungssystemen in offener kasse. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, 63, 85–100.

Anderson, A.W. (1992). *Pension Mathematics for Actuaries*, 2nd ed. ACTEX Publications, Winsted, Connecticut.

Arthur, T.G. & Randall, P.A. (1990). Actuaries, pension funds and investment. *Journal of the Institute of Actuaries*, 117, 1–49.

Åström, K.J. (1970). *Introduction to Stochastic Control Theory.* Academic Press, New York.

Atkinson, M.N. (1994). Another look at the valuation of variable income assets by the discounted income approach. *International Association of Consulting Actuaries' Congress*, Hong Kong, 1, 161–182.

Bader, L.N. (1983). Actuarial implications of dedicated pension funds. *Transactions of the Society of Actuaries*, 35, 563–584.

Bagehot, W. (pseudonym of Treynor, J.L.) (1972). Risk and reward in corporate pension funds. *Financial Analysts' Journal*, 28, January–February 1972, 28, 80–84.

Balzer, L.A. & Benjamin, S. (1980). Dynamic response of insurance systems with delayed profit/loss-sharing feedback to isolated unpredicted claims. *Journal of the Institute of Actuaries*, 107, 513–528.

Balzer, L.A. (1982). Control of insurance systems with delayed profit/loss-sharing feedback and persisting unpredicted claims. *Journal of the Institute of Actuaries*, 109, 285–316.

Bassett, P.C. (1972). Progressive approach to pension funding. *Harvard Business Review*, November–December 1972, 125–135.

Bellhouse, D.R. & Panjer, H.H. (1981). Stochastic modelling of interest rates with applications to life contingencies—Part II. *Journal of Risk and Insurance*, 48, 628–637.

Bellman, R.E. (1957). *Dynamic Programming.* Princeton University Press, Princeton, New Jersey.

Benjamin, S. (1984). An actuarial layman looks at control theory. *Transactions of the 22nd International Congress of Actuaries*, Sydney, 4, 295–309.

Benjamin, S. (1989). Driving the pension fund. *Journal of the Institute of Actuaries*, 116, 717–735.

Berin, B.N. (1989). *The Fundamentals of Pension Mathematics.* Society of Actuaries, Illinois.

Berin, B.N. & Lofgren, E.P. (1987). Financial Accounting Standards No. 87: Recursion formulas and other related matters. *Transactions of the Society of Actuaries*, 39, 9–18.

Bertsekas, D.P. (1976). *Dynamic Programming and Stochastic Control.* Academic Press, New York.

Bertsekas, D.P. & Shreve, S.E. (1978). *Stochastic Optimal Control. The Discrete Time Case.* Academic Press, New York.

Bezooyen, J.T.S. van & Mehta, S.J.B. (1998). Investment strategies for Dutch and UK pension funds. *Transactions of the 26th International Congress of Actuaries*, Birmingham, 5, 185–214.

Bilodeau, C. (1995). Variability of Pension Plan Cost Ratio and Funding Level. Research Report 95-03, Institute of Insurance and Pension Research, University of Waterloo, Ontario.

Bizley, M.T.L. (1950). The effect on pension fund contributions of a change in the rate of interest. *Journal of the Institute of Actuaries Students' Society*, 10, 47–51.

Black, F. (1980). The tax consequences of long-run pension policy. *Financial Analysts' Journal*, July–August 1980, 21–28.

Black, F. (1995). The plan sponsor's goal. *Financial Analysts' Journal*, July–August 1995, 6–7.

Black, F. & Dewhurst, M.P. (1981). A new investment strategy for pension funds. *Journal of Portfolio Management*, 7, Summer 1981, 26–34.

Blake, D. (1996). Pension Schemes as Options on Pension Fund Assets: Implications for Pension Fund Management. Discussion Paper PI-9606. The Pensions Institute, Birkbeck College, University of London.

Blake, D. (1997). Pension Funds and Capital Markets. An Essay on Mutual Dependence. Discussion Paper PI-9706. The Pensions Institute, Birkbeck College, University of London.

Bleakney, T.P. (1972). *Retirement Systems for Public Employees.* Richard D. Irwin, Homewood, Illinois.

Boden, D.E. & Kingston, T.D. (1979). The effect of inflation on pension schemes and their funding. *Transactions of the Faculty of Actuaries*, 36, 399–468.

Bodie, Z., Light, J.O., Morck, R. & Taggart, R.A. (1985). Corporate pension policy: an empirical investigation. *Financial Analysts' Journal*, September–October 1985, 10–16.

Boulier, J.-F., Trussant, E. & Florens, D. (1995). A dynamic model for pension funds management. *Proceedings of the 5th AFIR International Colloquium*, Leuven, 1, 361–384.

Boulier, J.-F., Michel, S. & Wisnia, V. (1996). Optimizing investment and contribution policies of a defined benefit pension fund. *Proceedings of the 6th AFIR International Colloquium*, Nürnberg, 1, 593–607.

Bowers, N.L., Hickman, J.C. & Nesbitt, C.J. (1976). Introduction to the dynamics of pension funding. *Transactions of the Society of Actuaries*, 28, 177–203.

Bowers, N.L., Hickman, J.C. & Nesbitt, C.J. (1979). The dynamics of pension funding: contribution theory. *Transactions of the Society of Actuaries*, 31, 93–122.

Bowers, N.L., Hickman, J.C. & Nesbitt, C.J. (1982). Notes on the dynamics of pension funding. *Insurance: Mathematics and Economics*, 1, 261–270.

Box, G.E.P., Jenkins, G.M. & Reinsel, G.C. (1994). *Time Series Analysis*, 3rd ed. Prentice-Hall, Englewood Cliffs, New Jersey.

Boyle, P.P. (1976). Rates of return as random variables. *Journal of Risk and Insurance*, 43, 693–711.

Brand, L. (1966). *Differential and Difference Equations.* Wiley, New York.

Bride, M. & Lomax, M.W. (1994). Valuation and corporate management in a non-life insurance company. *Journal of the Institute of Actuaries*, 121, 363–440.

Brownlee, H.J. & Daskais, R. (1991). Pension plans. Choosing critical assumptions. *Proceedings of the Second AFIR International Colloquium*, Brighton, 2, 153–162.

Cairns, A.J.G. (1994). An Introduction to Stochastic Pension Plan Modelling. Paper presented at the workshop on Interest Rate Risk, Vancouver, 19–20 August 1994. Also: Discussion Paper PI-9607. The Pensions Institute, Birkbeck College, University of London.

Cairns, A.J.G. (1997). A comparison of optimal and dynamic control strategies for conti nuous-time pension fund models. *Proceedings of the 7th AFIR International Colloquium*, Cairns, Australia, 1, 309–326.

Cairns, A.J.G. & Parker, G. (1997). Stochastic pension fund modelling. *Insurance: Mathematics and Economics*, 21, 43–79.

Canadian Institute of Actuaries (1994). *Standard of Practice for Valuation of Pension Plans.* Ottawa, Ontario.

Cavaye, C.M. & Springbett, T.M. (1964). Actuarial note on the calculation of premium rates using a decreasing rate of interest and allowing for benefit of immunisation. *Transactions of the Faculty of Actuaries*, 28, 308–322.

Chow, G.C. (1975). *Analysis and Control of Dynamic Economic Systems.* John Wiley & Sons, New York.

Colbran, R.B. (1982). Valuation of final-salary pension schemes. *Journal of the Institute of Actuaries*, 109, 359–385.

Cronquist, N.R. & Dreher, W.A. (1972). A new look at the frequency of pension plan actuarial valuation. *Transactions of the Society of Actuaries*, 24, 79–155.

Day, J.G. & McKelvey, K.M. (1964). The treatment of assets in the actuarial valuation of a pension fund. *Journal of the Institute of Actuaries*, 90, 104–147.

Daykin, C.D. (1976). Long-term rates of interest in the valuation of a pension fund. *Journal of the Institute of Actuaries*, 21, 286–340.

Daykin, C.D. (1987). Further thoughts on long-term rates of interest in the valuation of a pension fund. *Journal of the Institute of Actuaries Students' Society*, 30, 117–140.

Daykin, C.D., Ballantine, G. & Anderson, D. (1993). Modelling the assets and liabilities of a pension fund. *Proceedings of the 3rd International AFIR Colloquium*, Rome, 2, 525–537.

Daykin, C.D., Pentikäinen, T. & Pesonen, M. (1994). *Practical Risk Theory for Actuaries.* Chapman & Hall, London.

De Dominicis, R., Manca, R. & Granata, L. (1991). The dynamics of pension funds in a stochastic environment. *Scandinavian Actuarial Journal*, 2, 118–128.

Dhaene, J. (1989). Stochastic interest rates and autoregressive integrated moving average processes. *ASTIN Bulletin*, 19, 131–138.

Dufresne, D. (1986). The Dynamics of Pension Funding. Ph.D. thesis. School of Mathematics, The City University, London.

Dufresne, D. (1988). Moments of pension contributions and fund levels when rates of return are random. *Journal of the Institute of Actuaries*, 115, 535–544.

Dufresne, D. (1989). Stability of pension systems when rates of return are random. *Insurance: Mathematics and Economics*, 8, 71–76.

Dufresne, D. (1990a). Fluctuations of pension contributions and fund level. *ARCH (Actuarial Research Clearing House)*, 1990.1, 111-120.

Dufresne, D. (1990b). The distribution of a perpetuity, with applications to risk theory and pension funding. *Scandinavian Actuarial Journal*, 1990, 39–79.

Dufresne, D. (1993). Some Aspects of Statement of Financial Accounting Standards No. 87. Report of project sponsored by the Actuarial Education Research Fund, Society of Actuaries. Also reproduced in *ARCH (Actuarial Research Clearing House)*, 1993.2, 1–130.

Dufresne, D. (1994). *Mathématiques des Caisses de Retraite.* Éditions Supremum, Montréal.

Dufresne, D. (1995). Actuarial gains and losses: a martingale approach. Report of project sponsored by the Actuarial Education Research Fund, Society of Actuaries.

Dyson, A.C.L. & Exley, C.J. (1995). Pension fund asset valuation and investment. *British Actuarial Journal*, 1, 471–557.

Elton, E.J. & Gruber, M.J. (1987). *Modern Portfolio Theory and Investment Analysis*, 3rd ed. Wiley, New York.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, 50, 987–1007.

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.

Ezra, D.D. (1979). *Understanding Pension Fund Finance and Investment.* Pagurian Press, Toronto.

Ezra, D.D. (1980). How actuaries determine the unfunded pension liability. *Financial Analysts' Journal*, July–August1980, 43–50.

Ezra, D.D. (1988). Economic values: a pension pentateuch. *Financial Analysts' Journal*, March–April 1988, 58–66.

Ezra, D.D. (1991). Asset allocation by surplus optimization. *Financial Analysts' Journal*, January–February 1991, 51–57.

Fama, E.F. & Schwert, G.W. (1977). Asset returns and inflation. *Journal of Financial Economics*, 5, 115–146.

Fama, E.F. & French, K.R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22, 3–25.

Ferris, S. & Welch, D. (1996). A Survey of Valuation Assumptions and Funding Methods used by Australian Actuaries in Defined Benefit Superannuation Fund Valuations. Research Paper No. 004/96. Actuarial Studies and Demography Department, School of Economic and Financial Studies, Macquarie University, Sydney.

Ferris, S. (1997). Investment Objectives of Australian Superannuation Funds. Research Paper No. 005/97. Actuarial Studies and Demography Department, School of Economic and Financial Studies, Macquarie University, Sydney.

Finkelstein, G.S. (1997). Maturity guarantees revisited: allowing for extreme stochastic fluctuations using stable distributions. *British Actuarial Journal*, 3, 411–482.

Fisher, H.F. & Young, J. (1971). *Actuarial Practice of Life Assurance.* Cambridge University Press, Cambridge.

Fleischer, D.R. (1975). The forecast valuation method for pension plans. *Transactions of the Society of Actuaries*, 27, 93–125. Discussion: 127–154.

Ford, A. *et al.* (1980). Report of the Maturity Guarantees Working Party. *Journal of the Institute of Actuaries*, 107, 103–212.

Frees, E.W. (1990). Stochastic life contingencies with solvency considerations. *Transactions of the Society of Actuaries*, 42, 91–148.

Friedman, B.M. (1983). Pension funding, pension asset allocation and corporation finance, Evidence from individual company data. In: Bodie, Z. & Shoven, J.B. (1983). *Financial Aspects of the United States Pension System.* University of Chicago Press, Chicago.

Fujiki, M.-H. (1994). Pension Fund Valuation. Ph.D. thesis. Department of Actuarial Science and Statistics, The City University, London.

Furnish, J. *et al.* (1985). Select and ultimate financial assumptions in pension plan valuations: an analysis of the issues. Society of Actuaries Committee on Pension Principles and Related Research. *Transactions of the Society of Actuaries*, 38, 351–392.

Geoghegan, T J. *et al.* (1992). Report on the Wilkie stochastic investment model. *Journal of the Institute of Actuaries*, 119, 173–228.

Gerrard, R. & Haberman, S. (1996). Stability of Pension Systems when Gains/Losses are Amortized and Rates of Return are Auto-regressive. *Insurance: Mathematics and Economics*, 18, 59–71.

Giesecke, G.L. (1994). The Projected Unit method with benefits apportioned by interest-adjusted salary. *Transactions of the Society of Actuaries*, 46, 193–226.

Gilley, D.F. & Funnell, D. (1958). Valuation of pension fund assets. *Journal of the Institute of Actuaries Students' Society*, 15, 43–68.

Gordon, M. (1962). *The Investment, Financing and Valuation of the Corporation.* Richard D. Irwin, Homewood, Illinois.

Grauer, R.R. & Hakansson, N.H. (1985). Returns on levered, actively managed long run portfolios of stocks bonds and bills: 1934–1983. *Financial Analysts' Journal*, September–October.

Griffin, F.L. (1966). Concepts of adequacy in pension plan funding. *Transactions of the Society of Actuaries*, 18, 46–63.

Haberman, S. (1992a). Pension funding with time delays—a stochastic approach. *Insurance: Mathematics and Economics*, 11, 179–189.

Haberman, S. (1992b) The Effect of Autoregressive Rates of Return on the Variability of Pension Contributions & Fund Levels. Actuarial Research Paper No. 29, Department of Actuarial Science and Statistics, The City University, London.

Haberman, S. (1993). Pension funding: the effect of changing the frequency of valuations. *Insurance: Mathematics and Economics*, 13, 263–270.

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, 219–240.

Haberman, S. (1994b). Pension Funding Modelling & Stochastic Investment Returns. Paper presented on 29 March 1994 at Joint Meeting of the Staple Inn Actuarial Society and Royal Statistical Society, General Applications Section, London. Actuarial Research Paper No. 62, Department of Actuarial Science and Statistics, The City University, London.

Haberman, S. (1995). Pension funding with time delays and the optimal spread period. *ASTIN Bulletin*, 25, 177–187.

Haberman, S. & Sung, J.-H. (1994). Dynamic approaches to pension funding. *Insurance: Mathematics and Economics*, 15, 151–162.

Haberman, S. & Owadally, M.I. (1997). Pension fund dynamics and surpluses/deficits due to random rates of return. *Proceedings of the 7th AFIR International Colloquium*, Cairns, Australia, 1, 379–399.

Haberman, S. & Smith, D. (1997). Stochastic Investment Modelling and Pension Funding: a Simulation Based Analysis. Actuarial Research Paper No. 102. Department of Actuarial Science and Statistics, The City University, London.

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 scheme. *Insurance: Mathematics and Economics*, 20, 115–135.

Hagerman, R.L. & Zmijewski, M.E. (1979). Some economic determinants of accounting policy choice. *Journal of Accounting and Economics*, 1, 141–161.

Haynes, A.T. & Kirton, R.J. (1952). The financial structure of a life office. *Transactions of the Faculty of Actuaries*, 21, 141–197.

Heywood, G. & Lander, M. (1961). Pension fund valuation in modern conditions. *Journal of the Institute of Actuaries*, 87, 314–338.

Holcombe, S.M. *et al.* (1973). Financial planning for pension plans—discussion. *Transactions of the Society of Actuaries*, 25, Part II, D1–D35.

Huber, P.P. (1996). A Study of the Fundamentals of Actuarial Economic Models. Ph.D. thesis. Department of Actuarial Science and Statistics, The City University, London.

Huber, P.P. (1997). A review of Wilkie's stochastic asset model. *British Actuarial Journal*, 3, 181–210

Humphrey, G.T. *et al.* (1970). Pensions and company finance. *Journal of the Institute of Actuaries*, 96, 189–227.

IASC (1996). *Exposure Draft 54: Accounting for Employee Benefits.* International Accounting Standards Committee.

Ibbotson, R.G. & Sinquefield, R.A. (1993). *Stocks, Bonds, Bills and Inflation, 1993 Yearbook.* Ibbotson Associates, Chicago.

IFAA (1997). *Comments on IASC ED54.* International Forum of Actuarial Associations.

Institute and Faculty of Actuaries (1997a). *Guidance Note 9: Retirement Benefit Schemes—Actuarial Reports.* Manual of Actuarial Practice, Institute of Actuaries, London, and Faculty of Actuaries, Edinburgh.

Institute and Faculty of Actuaries (1997b). *Guidance Note 26: Pension Fund Terminology.* Manual of Actuarial Practice, Institute of Actuaries, London, and Faculty of Actuaries, Edinburgh.

Institute of Actuaries of Australia (1994). *Guidance Note 461. Funding and Solvency Certificates.* Sydney.

Institute of Actuaries of Australia (1995). *Professional Standard 400. Investigations of Defined Benefit Superannuation Funds.* Sydney.

Jackson, P.H. (1984). Variable actuarial assumptions and the solvency of private pension plans. *Transactions of the 22nd International Congress of Actuaries*, Sydney, 3, 311–319.

Jackson, P.H. & Hamilton, J.A. (1968). The valuation of pension fund assets. *Transactions of the Society of Actuaries*, 20, 386–436.

Joanette, F.P. (1985). Managing Corporate Pension Funds: A Study of the Determinants of Pension Funding and Assets Allocation Decisions. Ph.D. thesis. Wharton School, University of Pennsylvania.

Jones, P.D. (1993). The expected return on United Kingdom equities and an implication for pension fund valuations. *Journal of the Institute of Actuaries*, 120, 253–291.

Jury, E.I. (1962). A simplified stability criterion for linear discrete systems. *Proceedings of the Institute of Radio Engineers* (now: *Proceedings of the Institute of Electrical and Electronic Engineers*), 50, 1493–1500.

Jury, E.I. (1964). *Theory and Applications of the z-Transform Method.* John Wiley & Sons, New York.

Keintz, R.J. & Stickney, C.P. (1980). Immunization of pension fund and sensitivity to actuarial assumptions. *Journal of Risk and Insurance*, 47, 223–239.

Kemp, M.H.D. (1996). Asset/liability Modelling for Pension Funds. Paper presented at the Staple Inn Actuarial Society, London.

Khorasanee, M.Z. (1993). Survey of Actuarial Practice in the Funding of U.K. Defined Benefit Schemes. Actuarial Research Paper No. 54. Department of Actuarial Science and Statistics, The City University, London.

Khorasanee, M.Z. (1997). Deterministic modelling of defined contribution pension funds. *North American Actuarial Journal*, 1, no. 4.

Kingsland, L. (1982). Projecting the financial condition of a pension plan using simulation analysis. *Journal of Finance*, 37, 577–584.

Kleynen, R. (1997). Asset liability management for pension funds. A case study. *Proceedings of the 7th AFIR International Colloquium*, 2, 529–541.

Knox, D. (1993). A Critique of Defined Contribution Using a Simulation Approach. Research Paper No. 7, Centre for Actuarial Studies, Department of Economics, University of Melbourne.

Lee, E.M. (1986). *An Introduction to Pension Schemes.* Institute and Faculty of Actuaries, London.

Leibowitz, M.L. (1986a). The dedicated bond portfolio in pension funds—Part I: motivations and basics. *Financial Analysts' Journal*, January–February 1986, 68–75.

Leibowitz, M.L. (1986b). The dedicated bond portfolio in pension funds—Part II: immunization, horizon matching and contingent procedures. *Financial Analysts' Journal*, March–April 1986, 47–57.

Leibowitz, M.L. (1986c). Total portfolio duration: a new perspective on asset allocation. *Financial Analysts' Journal*, September–October 1986, 18–29, 77.

Leibowitz, M.L. & Henriksson, R.D. (1988). Portfolio optimization with a surplus framework. *Financial Analysts' Journal*, March-April 1988, 43–51.

Leibowitz, M.L., Kogelman, S. & Bader, L.N. (1992). Asset performance and surplus control: a dual-shortfall approach. *Journal of Portfolio Management*, 18, 2, Winter 1992, 28–37.

Loades, D.H. (1988). Assessing the security of pension fund valuation bases using a stochastic investment model. *Transactions of the 23rd International Congress of Actuaries*, Helsinki, 5, 105–123.

Loades, D.H. (1992). Instability of pension funding. *Transactions of the 24th International Congress of Actuaries*, Montréal, 2, 137–154.

Loades, D.H. (1998). Elementary engineering control and pension funding. *Transactions of the 26th International Congress of Actuaries*, Birmingham, 5, 239–266.

Lochhead, A. (1994). Asset Valuation Methods. Paper presented on 3 March 1994 to the Glasgow Actuarial Society, Glasgow.

MacBeth, J.D., Emanuel, D.C. & Heatter, C.E. (1994). An investment strategy for defined benefit plans. *Financial Analysts' Journal*, May–June 1994, 34–41.

Mandl, P. & Mazurová, L. (1996). Harmonic analysis of pension funding methods. *Insurance: Mathematics and Economics*, 17, 203–214.

Marden, M. (1966). *The Geometry of Polynomials,* 2nd ed. American Mathematical Society, Providence, Rhode Island.

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

McGill, D.M. (1964). *Fundamentals of Private Pensions,* 2nd ed. Richard D. Irwin, Homewood, Illinois.

McGill, D.M., Brown, K.N., Haley, J.J. & Schieber, S.J. (1996). *Fundamentals of Private Pensions,* 7th ed. University of Pennsylvania Press, Philadelphia.

McKenna, F.W. (1982). Pension plan cost risk. *Journal of Risk and Insurance*, 49, 193–217.

McKenna, F.W. & Kim, Y.H. (1986). Managerial risk preferences, real pension costs, and long-run corporate pension fund investment policy. *Journal of Risk and Insurance*, 53, 29–48.

McLeish, D.J.D. & Stewart, C.M. (1987). Objectives and methods of funding defined benefit pension schemes. *Journal of the Institute of Actuaries*, 114, 155–225.

Mennis, E.A., Valentine, J.L. & Mennis, D.L. (1981). New perspectives on pension fund management. *Journal of Portfolio Management*, 7, 3, Spring 1981, 46–50.

Milgrom, P.R. (1985). Measuring the interest rate risk. *Transactions of the Society of Actuaries*, 37, 241–302.

O'Brien, T. (1986). A stochastic dynamic approach to pension funding. *Insurance: Mathematics and Economics*, 5, 141–146.

O'Brien, T. (1987). A two-parameter family of pension contribution functions and stochastic optimization. *Insurance: Mathematics and Economics*, 6, 129–134.

Office of the Superintendent of Financial Institutions Canada (1987). *Pension Benefits Standards Act, 1985.* Reference Manual.

O'Regan, W.S. & Weeder, J. (1990). A dissection of pensions funding. *Journal of the Staple Inn Actuarial Society*, 32, 71–115.

Owadally, M.I. & Haberman, S. (1999). Pension fund dynamics and gains/losses due to random rates of investment return. *North American Actuarial Journal*, To appear.

Panjer, H.H. & Bellhouse, D.R. (1980). Stochastic modelling of interest rates with applications to life contingencies. *Journal of Risk and Insurance*, 47, 91–110.

Paquin, C.Y. (1975). A review of actuarial cost methods for defined benefit pension plans. *Transactions of the Society of Actuaries*, 27, 493–508.

Pemberton, J. (1998). The Value of Actuarial Values. Paper presented on 17 February 1998 at the Staple Inn Actuarial Society, London.

Pollard, J.H. (1971). On fluctuating interest rates. *Bulletin de l'Association Royale des Actuaires Belges*, 66, 68–97.

Poterba, J.M. & Summers, L.H. (1988). Mean reversion in stock prices. Evidence and implications. *Journal of Financial Economics*, 22, 27–59.

Pratt, J.W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32, 122–136.

Redington, F.M. (1952). Review of the principles of life office valuation. *Journal of Institute of Actuaries*, 78, 286–315

Regan, P.J. (1980). Pension fund perspective. Moody's gathers pension data. *Financial Analysts' Journal*, July–August 1980, 16–17.

Samuelson, P.A. (1989). A case at last for age-phased reduction in equity. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 9048–9051.

Schnitzer, R.J. (1977). Characteristics and operation of projection valuation methods for pension plan funding. *Transactions of the Society of Actuaries*, 29, 269–314.

Seal, H.L. (1952). Acceptable funding methods for self-insured pension plans. *Proceedings of the Conference of Actuaries in Public Practice*, 2, 17–44.

Shapiro, A.F. (1979). A Bayesian approach to persistency in the projection of retirement costs. *Transactions of the Society of Actuaries*, 31, 337–365.

Shapiro, A.F. (1983). Modified cost methods for small pension plans. *Transactions of the Society of Actuaries*, 35, 11–36.

Shapiro, A.F. (1985). Contributions to the evolution of pension cost analysis. *Journal of Risk and Insurance*, 52, 81–99.

Sharp, K.P. (1992). Modelling Canadian Price and Wage Inflation. Research Report 92-12. Institute of Insurance and Pension Research, University of Waterloo, Ontario.

Sharp, K.P. (1996). Pension funding by normal costs or amortization of unfunded liabilities. *Journal of Actuarial Practice*, 4, 257–274.

Sharpe, W.F. & Tint, L.G. (1990). Liabilities—a new approach. *Journal of Portfolio Management*, Winter 1990, 16, 5–10.

Sharpe, W.F., Alexander, G.J. & Bailey, J.V. (1995). *Investments*, 5th ed. Prentice-Hall, Englewood Cliffs, New Jersey.

Sherris, M. (1992). Portfolio selection and matching: a synthesis. *Journal of the Institute of Actuaries*, 119, 87–105.

Sherris, M. (1994). The stochastic valuation of superannuation benefits. *Transactions of the Institute of Actuaries of Australia*, , 522–536.

Sherris, M. (1995). Accumulation versus Deferred Benefits Funds: an Analysis using Simulation Models. Research Paper No. 005/95. Actuarial and Demography Department, School of Economic and Financial Studies, Macquarie University, Sydney.

Smith, A.D. (1996). How actuaries can use Financial Economics. *British Actuarial Journal*, 2, 1057–1193.

Snelson, R.E. (1970). Modern methods of costing insured pension schemes. *Journal of the Institute of Actuaries Students' Society*, 19, 1–41.

Springbett, T.M. (1964). Valuation for surplus. *Transactions of the Faculty of Actuaries*, 28, 231–307.

Sung, J.-H. (1997). Dynamic Programming Approaches to Pension Funding. Ph.D. thesis. Department of Actuarial Science and Statistics, The City University, London.

Taylor, G.C. (1986). Determination of the rate of investment return for the discounting of general insurance outstanding claims. *Journal of the Institute of Actuaries*, 113, 61–101.

Taylor G.C. (1987). Control of unfunded and partially funded systems of payments. *Journal of the Institute of Actuaries*, 114, 371–392.

Taylor, J.R. (1967). The generalized family of aggregate actuarial cost methods for pension funding. *Transactions of the Society of Actuaries*, 19, 1–12, 128–136.

Tepper, I. (1977). Risk vs. return in pension fund investment. *Harvard Business Review*, March–April 1977, 100–107.

Tepper, I. (1981). Taxation and corporate pension policy. *Journal of Finance*, 36, 1, 1–13.

Thomson, R.J. (1996). Stochastic investment modelling: the case of South Africa. *British Actuarial Journal*, 2, 765–801.

Thornton, P.N. & Wilson, A.F. (1992a). A realistic approach to pension funding. *Journal of the Institute of Actuaries*, 119, 229–312.

Thornton, P.N. & Wilson, A.F. (1992b). Notes on the Emergence of Surplus in Occupational Schemes in the 1980's. Institute of Actuaries, London. Supplement to: Thornton, P.N. & Wilson, A.F. (1992). A realistic approach to pension funding. *Journal of the Institute of Actuaries*, 119, 229–312.

Tilley, J.A. (1988). The application of modern techniques to the investment of insurance and pension funds. *Transactions of the 23rd International Congress of Actuaries*, Helsinki.

Treynor, J.L. (1977). The principles of corporate pension finance. *Journal of Finance*, 32, 627–638.

Trowbridge, C.L. (1952). Fundamentals of pension funding. *Transactions of the Society of Actuaries*, 4, 17–43, 657–683.

Trowbridge, C.L. (1963). The unfunded present value family of pension funding methods. *Transactions of the Society of Actuaries*, 15, 151–169.

Trowbridge, C.L. (1966). ABC's of pension funding. *Harvard Business Review*, March–April 1966.

Trowbridge, C.L. & Farr, C.E. (1976). *The Theory and Practice of Pension Funding.* Irwin, Homewood, Illinois.

Turner, M.J. *et al.* (1984). Report of a Working Party of the Pensions Standard Joint Committee on Terminology of Pension Funding Methods. Institute and Faculty of Actuaries.

Vandebroek, M. (1990). Pension funding and optimal control. *Mitteilungen der Vereinigung schweizerischer Versichrungsmathematiker*, 313–325.

VanDerhei, J.L. & Joanette, F.P. (1988). Economic determinants for the choice of actuarial cost methods. *Journal of Risk and Insurance*, 55, 59–74.

Vanderhoof, I.T. (1972). The interest rate assumption and the maturity structure of the assets of a life assurance company. *Transactions of the Society of Actuaries*, 24, 157–205.

Vanderhoof, I.T. (1973). Choice and justification of an interest rate. *Transactions of the Society of Actuaries*, 25, 417–458.

Vanderhoof, I.T. (1984). Optimal allocation of assets for pension plans. *Transactions of the 22nd International Congress of Actuaries*, Sydney, 5, 37–42.

Whittle, P. (1982). *Optimization over Time*, Vol. 1. John Wiley & Sons, Chichester.

Whittle, P. (1996). *Optimal Control: Basics and Beyond.* John Wiley & Sons, Chichester.

Wilkie, A.D. (1985). Portfolio selection in the presence of fixed liabilities: a comment on 'The matching of assets to liabilities'. *Journal of the Institute of Actuaries*, 112, 229–277.

Wilkie, A.D. (1987). Stochastic investment models—theory and applications. *Insurance: Mathematics and Economics*, 6, 65–83.

Wilkie, A.D. (1995). More on a stochastic asset model for actuarial use. *British Actuarial Journal*, 1, 777–964.

Winklevoss, H.E. (1977). *Pension Mathematics: With Numerical Illustrations*. Richard D. Irwin, Homewood, Illinois.

Winklevoss, H.E. (1982). PLASM: pension liability and asset simulation model. *Journal of Finance*, 37, 585–594.

Winklevoss, H.E. (1993). *Pension Mathematics with Numerical Illustrations*, 2nd ed. Pension Research Council, Wharton School of the University of Pennsylvania and University of Pennsylvania Press, Philadelphia.

Wise, A.J. (1984). The matching of assets to liabilities. *Journal of the Institute of Actuaries*, 111, 445–486.

Wise, A.J. (1987). Matching and portfolio selection: part 1. *Journal of the Institute of Actuaries*, 114, 113–133.

Wise, A.J. (1989). Matching. *Journal of the Institute of Actuaries*, 116, 529–535.

Wise, A.J. (1998). Pension costs at market value. *The Actuary*, March 1998, 18–19. Published by the Staple Inn Actuarial Society, London. Comments from: Smith, A.D. (1998). Letter to the Editor: pension costs at market value. *The Actuary*, April 1998, 36.

Wise, A.J. & Annable, M.J. (1990). The current state of asset/liability modelling in the U.K. *Proceedings of the First AFIR International Colloquium*, Paris, 1, 27–34.

Wong, L.Y.P. (1995). Moving Average Rates of Return and the Variability of Pension Contributions and Fund Levels for a Defined Benefit Pension Scheme. M.Sc. dissertation. Department of Actuarial Science and Statistics, The City University, London.

Zimbidis, A. (1992). Delay, Feedback and Variability of Pension Contributions and Fund Levels. M.Sc. dissertation. Department of Actuarial Science and Statistics, The City University, London.

Zimbidis, A. & Haberman, S. (1993). Delay, feedback and variability of pension contributions and fund levels. *Insurance: Mathematics and Economics*, 13, 271–285.