Dynamic Programming Approaches

To Pension Funding

by

JOO-HO SUNG

Thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

THE CITY UNIVERSITY, LONDON

DEPARTMENT OF ACTUARIAL SCIENCE AND STATISTICS

FEBRUARY 1997

Contents

Acknowledgements

During the course of the work that led to the thesis, there are many people who have helped make the thesis possible.

I would like to thank Professor Steven Haberman for his valuable comments and suggestions, and linguistic support.

I also express my gratitude to Professor Jae-Joo Kim (Seoul National University) and Dr. Chang-Soo Lee (Soongsil University) for their continued encouragements.

Finally, I wish to give my sincere thanks to my family, without whose support the thesis could never have been completed.

Declaration

Abstract

The thesis describes a dynamic programming approach to pension funding for a defined benefit pension scheme. The primary purpose of the thesis is to search for a pension funding plan balancing optimally the conflicting interests of the sponsoring employer and the trustees. The starting point of this problem is to introduce and define mathematically two types of risk concerned respectively with the stability and security of pension funding: the "solvency risk" and the "contribution rate risk". Next, two distinct linear asset/liability dynamic models are presented, based on specified assumptions: the "modified solvency-level growth equation" and the "zero-input, 100%-target solvency-level growth equation". We then consider the situation of a short-term, winding-up valuation with contribution rates unconstrained by any funding plan, and introduce three distinct finite-horizon control optimisation problems - deterministic, stochastic with complete state information and stochastic with incomplete state information. We then consider the situation of a long-term, going-concern valuation with contribution rates constrained by the spread funding plan, and introduce four distinct infinite-horizon deterministic control optimisation problems - stationary, quasi-stationary, non-stationary and threshold. The thesis derives optimal funding control procedures for the contribution rate by solving each of these seven control problems by means of the optimal control theory of dynamic programming.

Chapter 1 Introduction

1.1 Occupational pension schemes

The occupational pension scheme can be thought of as that part of the remuneration plan (or bonus program) which is set up the sponsoring employer with the aim of providing an adequate income in retirement for his employees. Broadly, the sources of retirement income for an individual consist of private savings (including personal pension provision) and assets, pensions provided by the pension scheme and some benefits provided by the social security system. In many countries, the employer-sponsored occupational pension scheme plays an important role in assuring a stable and adequate living during the individual's retirement: in 1991 in the UK, about 48% of all employees were members of an occupational pension scheme and about 62% of the population over retirement age (65 for men and 60 for women) received an occupational pension [statistical source: HMSO Vol. I (1994; p28)]. Further, from the viewpoint of taxation advantages on income tax, investment income tax and capital gains tax (as currently in the UK), occupational pension schemes can be attractive ways of saving for retirement income.

Occupational pension schemes are classified into two types according to the salient features of the formulae that determine the retirement income offered by such schemes: defined benefit pension schemes (or earnings-related pension schemes, e.g. final salary pension scheme, the most common type in the UK, and average salary pension scheme) and defined contribution pension schemes (or money purchase pension schemes). That is,

- Defined benefit pension schemes: these are a common means of providing for retirement income in a number of countries including the UK, USA., Canada and the Netherlands. The main feature of these schemes is that the trust deed and rules specify the benefits promised in the event of various contingencies by a predetermined formula (based mainly on the salary and duration of scheme membership of the employee at or near retirement), while the contributions to be paid (by the employer and possibly the members) are specified by the scheme actuary as part of the regular valuation process. Normally, the employer bears most of the risk of poor investment performance causing additional contributions to be paid to support the pre-determined amount of benefits, while the employee bears the potential risk of catastrophic changes in the employer's finances (and hence has more interest in securing his promised benefits at any time); and

- Defined contribution pension schemes: these have been in a minority in comparison with defined benefit pension schemes [see Casey (1993; Diagram 1 in p3)], but are expected to become more popular in the UK as a result of increasing statutory regulation following the report of the Goode Committee and, in particular, the introduction of the minimum funding requirement (MFR) contained in the Pensions Act 1995 (- the MFR will be explained later in section 3.2.3). The defined contribution pension scheme is conceptually simpler than the defined benefit pension scheme - the trust deed and rules prescribe the contributions in advance (normally in the form of a fixed percentage of the pensionable payroll) and the benefits to be ultimately payable are unknown but determined (at the time of retirement) directly by the contributions paid into the scheme, subsequent investment returns that they earn and annuity rates provided by the open market in pension policies, so that the level of the benefits is usually largely dependent on the conditions of open (investment and pension) markets prevailing at the time of retirement. Normally, the employee bears most of the risk of poor investment performance, whereas the employer will discharge his obligation by paying the agreed contributions (and hence has no duty of paying additional contributions unlike the defined benefit pension scheme).

So, the pension professionals (who are involved in establishing and managing the defined benefit pension scheme, e.g. actuary, investment manager and auditor) are commonly required to pay considerable attention to the security of pension rights to which the members in the defined benefit pension scheme are entitled and enable the sponsoring employer to finance steadily the defined benefit pension scheme without undue financial burden on his business.

1.2 Conflicts of interest inherent in defined benefit pension schemes

Since the early 1990's, there has been a growing interest in evaluating the financial strength (or security) of occupational pension schemes in the UK, particularly in order to restore public confidence in the security of schemes which has been seriously disturbed by the so-called "Maxwell affair" (revealed soon after Mr Maxwell's death in November 1991). An important lesson arising from the Maxwell affair is that merely keeping the scheme's assets separate from the assets of both employer and employees does not in itself guarantee security, leading to the real issue of how to strengthen the watchdog role of the trustees (who are a group of persons whose responsibility is to act in the best interests of the members of the scheme). Also, there is renewed emphasis on the need for the actuary to check regularly the financial strength of the scheme and the ability of the employer to meet the contribution rates required to maintain the scheme solvent. Without loss of generality, the security of a scheme may be regarded as the extent to which the accumulated assets meet all the benefits to which the members in the scheme are entitled when they arise (some security measures will be introduced and specified in section 2.1.2.7).

As outlined in section 1.1, the security problem is a specific issue in defined benefit pension schemes because different from defined contribution pension schemes, defined benefit pension schemes usually provide a (deferred or immediate) life annuity to each retiring employee, the amount of which is generally related both to uncertain future salaries of the employee at or near retirement and to his uncertain duration of employment. In other words, a defined benefit pension scheme has contingent liabilities in respect of all members in the scheme. In the thesis, we confine our interests to defined benefit pension schemes (afterwards, scheme means defined benefit pension scheme if there is no other mention).

In view of the long-term commitment of the sponsoring employer, the financial strength of the scheme will be affected mainly by the funding plan and investment strategy (or policy) in respect of the scheme's assets. That is, since the two principal sources of scheme income for meeting the accruing liabilities are normally the contribution income and investment income, optimal funding and investment decisions must play a vital role in encouraging the sponsoring employer to continue to provide a high quality of benefits as well as protecting the accruing liabilities against any potentially catastrophic changes in the employer's finances. The funding plan developed by the actuary may then be regarded as an arrangement/mechanism for determining the sequence of future contribution rates with the purpose of building up assets to meet the accruing liabilities without placing undue financial burdens on the employer. In parallel, the investment strategy adopted by the scheme investment manager may be thought of as the (short-term and/or long-term) asset allocation decisions for selecting the asset mix between securities (both fixed-interest and index-linked bonds, equities, property, cash deposits and other investments), with the aim of improving the investment performance, without subjecting the invested assets to undue financial risks. Throughout the thesis, the funding plans for schemes are considered from the viewpoint of the actuary, since he has, in general, responsibilities to both the employer and the trustees. Henceforth, we shall focus on the funding plan and the investment strategy will be mentioned just in connection with the funding plan.

It is generally agreed [see Lee (1986; Ch.8)] that any acceptable pension funding plan is required to provide the characteristics of security, stability, liquidity and durability (these shall be called the key funding characteristics). The relative importance of them will largely rely on both government legislation and the interests of the employer (including the shareholders) and trustees (including the scheme members). So, the employer will especially look for a funding plan which leads to as smooth and predictable as possible a cash flow of future contributions; that is, it is more important for the employer to maintain stability among the key funding characteristics. By contrast, the trustees seek to establish such a funding plan that ensures the promised benefits should be paid as they fall due, even in the event of the scheme being wound up; that is, it is more important for the trustees to maintain security among the key funding characteristics. Summing up, the employer and trustees each have their respective different priorities and viewpoints on setting up a pension funding plan - this is usually referred to as the conflicting interests between the employer and the trustees [see Loades (1992)].

As a result, the actuary will be required to establish a funding plan weighing the conflicting interests of the employer and the trustees, and so he will make an attempt to secure both stability and security as far as is possible according to their relative suitability: in fact, this must be one of the crucial issues in pension funding. In real life, the relative suitability may be determined by the views of the supervisory authorities who may "recommend" the pre-eminent importance of securing the promised benefits with stability taking second place, without discouraging the employer from continuing to support the pension scheme.

1.3 The aims and outline of the thesis

The primary aim of the thesis is to investigate pension funding plans that reconcile the natural desires of employer and trustees, characterised by their conflicting interests discussed in section 1.2, without unnecessary financial distortion from their expectations (or targets) - our approach is to consider an optimal pension funding plan derived within the mathematical framework of optimal control theory. The optimisation instruments/tools in optimal control theory, especially the method of dynamic programming, are adapted to give some useful theoretical and practical answers to this pension funding problem. In Chapter 2, the fundamental concepts of this methodology are introduced in comparison with various aspects associated with the classical actuarial approach to the pension funding problem. Ultimately, we provide insight into the fundamental process of how concepts and instruments of optimal control theory are selected and adapted to combine with the funding of the scheme (- this process is mathematically illustrated in Chapter 4 and 5 by means of the method of dynamic programming).

The starting point of adapting the dynamic programming approach to pension funding is building mathematical models that would be a structured set of simplified and abstracted formulations of the real financial structure inherent in the scheme. A large portion of Chapter 3 is devoted to the linear dynamic model construction on a discrete-time domain. Next, a unified mathematical framework for deriving the optimal pension funding plan (i.e. the so-called control optimisation problem) is constructed and solved in each of Chapters 4 and 5; in particular, the mathematical description and numerical illustrations of the derived funding plan are given in Chapter 4 for a short-term, winding-up perspective on pension funding and in Chapter 5 for a long-term, going-concern perspective on pension funding. In the closing chapter to the thesis, Chapter 6, we summarise the main concepts and results of the study together with providing some suggestions for future research as an extension to the thesis.

All these efforts involved in designing an effective and newly developed pension funding plan are eventually aimed at bridging the gap between actuarial theory in the field of pension funding and optimal control theory in the field of engineering, which would enable the pension professionals to analyse more systematically the (short-term and long-term) financial structure of the scheme and undertake further applied research on issues related to pension funding, in particular making close use of optimal control theory.

Chapter 2 Background - General descriptions of the funding of defined benefit schemes

2.1 Considerations in determining the pension funding plan

2.1.1 Introduction

As mentioned in section 1.2, we are concerned with the defined benefit pension scheme. The funding plan is considered from the viewpoint of the scheme actuary and the investment strategy will be mentioned just in connection with the funding plan.

Prior to discussing the various types of funding plans (these are described in section 2.2), the next section describes the general concepts related to funding plans to clarify our aims (mentioned in section 1.3): in particular, section 2.1.3 deals with two principal concepts in adapting the dynamic programming approach to the funding of the scheme: the 'solvency risk' and the 'contribution rate risk'.

2.1.2 Actuarial valuation

The starting point in building up confidence in the security of the scheme would be the periodical conducting of fully-detailed analyses of the current financial position of the scheme and then checking up and revising, if necessary, the current actuarial assumptions and funding plan.

Such an investigation can be called the actuarial valuation (or actuarial investigation) of the scheme. It plays an essential role in managing the scheme. As a practical explanation for the term 'periodical' introduced in the above, it is common, in the UK, for the valuation to be annual, although current legislation requires that for contracted-out schemes the valuation must be performed at least every three-and-a-half years.

2.1.2.1 Purposes

The purposes of actuarial valuation are threefold:

(i) providing information to members about the valuation result (i.e. the scheme's funding level or solvency level under the valuation basis chosen according to specific legislative purposes (e.g. statutory requirements for the scheme's security, taxation and so on));

(ii) adjusting the current set of actuarial assumptions regarding the future economic and demographic events to recent economic and demographic experience and the updated forecasts of future economic and demographic experience, if necessary; and

(iii) ultimately, setting the recommended contribution rate (funding rate) which is required to be paid (generally by reference to the valuation result).

In short, what is assumed and decided in the actuarial valuation is intended to keep the scheme continuing in a sound financial position in response to any potentially important adverse (demographic and economic) deviations in experience in the future, with the least distortion of the employer's cash-flows. Now, we shall briefly discuss, in turn, the key concepts of actuarial assumptions, recommended contribution rate, valuation basis and funding level vs. solvency level.

2.1.2.2 Actuarial assumptions

The scheme's liability can be expressed as a real-valued function characterised by the economic and demographic parameters used in the actuarial and financial management of the pension scheme (we shall call these collectively the scheme parameters): the key scheme economic parameters are the rate of salary growth {e.g. general salary inflation and increases in individual salaries arising from promotion and scale increase} and the rates related to investment earnings {e.g. rate of inflation, rate of return on fixed-interest stocks (e.g. ordinary gilts), rate of return on equities, rate of return on index-linked stocks (e.g. index-linked gilts) and running yield on All Share Index} and the key scheme demographic parameters are mortality and morbidity rates, withdrawal rates and the age profile of all the members. Since the scheme has a long lasting arrangement for protecting against various future contingencies, the true value of the scheme parameters can not logically be known with accuracy, that is, the prior commitment gives rise to the problem of scheme parameter uncertainty. Therefore, in order to estimate the most likely value of the scheme's liabilities and then fund these liabilities, the scheme actuary needs first to construct a set of most likely assumptions for the scheme parameters. The set of most-likely estimated (or best estimated) parameters, specific for the actuarial valuation, is called the actuarial assumptions. The actuarial assumptions are usually founded on the long-term position for the reason that the provision of pension is a long-term commitment and thus the possible short-term fluctuations of the scheme parameters about the general (demographic and economic) trend are expected to be averaged around the long-term estimates (though this long-term position is not necessarily appropriate over the short-term period). In reality, how to determine reasonably the set is one of the most important problems in actuarial applications.

At this stage, it is worth reviewing two distinctive works in relation to investigating how to set the actuarial assumptions - Thornton & Wilson (1992) and Fujiki (1994).

Thornton & Wilson (1992) suggest that so-called 'best estimated' bases introduced for the Statement of Standard Accounting Practice No 24 (SSAP 24, 1988) purposes should be derived from an analysis of the economic and demographic experience in order that it should be used as a standard of the degree of prudence appropriate for funding purposes. Of course, the analysis needs to be repeated regularly with incoming updated experiences so as to verify what margins might be appropriate for ensuring that a scheme would remain satisfactorily funded.

On the other hand, Fujiki (1994) employs several different scenarios for actual experience (e.g. gradual change in experience and cyclical change in experience) and several possible responses in the actuarial assumptions corresponding to each scenario (e.g. constant actuarial assumptions, gradual change in actuarial assumptions and/or cyclical change in actuarial assumptions) and then checks the stability of contribution rate movements for each pair of actual experience and actuarial assumptions. As a result, he shows that an effective way of selecting the actuarial assumptions is the use of the averaging of past experience and/or the use of a small in delay changing in the actuarial assumptions (in order to check on the permanence of a change in experience).

Although these two approaches produce similar results in the light of using the experienced data, Thornton & Wilson (1992) are concerned with the future forecasts for the scheme parameters using the best estimates and/or margins to protect against future uncertainty, but in contrast Fujiki (1994) is concerned with analysing the pattern of past experience associated with the scheme parameters using the averaging of past experience and/or a time-delay in the adjustment of the assumptions.

Therefore, the actuarial assumptions are necessarily somewhat subjective but have to be determined on the basis of the actual movements in experience so as to tackle the characteristics and uncertainty surrounding the scheme parameters. The setting of the actuarial assumption is of vital importance to designing a funding plan with the aim of controlling optimally the security of the scheme as well as the stability of the contribution rates. This will be discussed further in the following sections.

Prior to describing a variety of pension funding terminologies [see PMI/PRAG (1992)], we must note that the term 'actuarial' refers to a theoretical or notional concept when compared with the term 'actual or realistic' in a legal or auditing sense (this explanation would make it much easier to understand the 'actuarial' related words). The necessity of this term is attributed to the unpredictability of the long-term future as mentioned above. In the actuarial applications, one of the main tasks of the actuary can be thought of as defining two versions of a particular mathematical function that are based on actuarial assumptions and are based on actual experience, comparing the effect of each on the scheme and finally attempting to minimise (or optimise) the possible gap between their effects subject to some desirable restrictions. In this respect, Thornton & Wilson (1992) and Fujiki (1994) have a common theme.

2.1.2.3 Recommended contribution rate

Based on the valuation result, the actuary is able to recommend any necessary changes to the contributions paid by the employer and employees (i.e. to set the recommended contribution rate) in order to prevent the discrepancy between the scheme's assets and liabilities diverging excessively. The funding plan for recommended contribution rates falls normally into two main groups: in our terms, primary funding methods (usually, referred to as funding methods) and supplementary funding methods (usually, referred to as methods of amortisation) [for details, see section 2.2].

Co-ordinating the primary and supplementary funding methods aims to accumulate assets in a prudent and controlled way in advance of the actuarial contingencies specified in the trust deed, in order to meet the defined benefits as they fall due without undue distortion of the employer's cash-flows and also without undue shortfall in the scheme's assets if the scheme prematurely ceases.

Now, we shall describe briefly each of these concepts.

A primary funding method, except the Aggregate method [see section 2.2.1], specifies the actuarial liability (often referred to as the standard fund or the accrued liability (especially in the USA)) as a notional (or theoretical) target for the fund level and the normal cost (often referred to as the standard contribution rate or the regular cost) as a notional (or theoretical) target for the contribution rate, whose calculations are based on a given set of actuarial assumptions. As noted in the report 'Terminology of Pension Funding Methods' (TPFM, 1984), these functions, normal cost and actuarial liability, should not necessarily be regarded as targets. However, the following would be widely admissible on the normal assumption that the pension scheme is an going-concern entity (i.e. a long lasting arrangements) and then the routine valuation of the scheme is carried out making this going-concern assumption: that is, whatever the actuary's targets in the funding plan would be, the employer may not want to accumulate assets above the actuarial liability, particularly as long as the actuarial assumptions are correctly matched with the actual experience.

From this point of view, the term 'notional target' would be interpreted as an ideally desired level of assets. For a clear example for this interpretation, if there were no mismatching between the scheme's assets and its liabilities and a common valuation basis is set for the assets and the liabilities (or the actuarial assumptions are consistently borne out by actual experience and also the scheme starts with no surplus or deficit), then a primary funding method would maintain the following relationships: at each valuation date, 'actuarial liability =

fund level' and 'normal cost = recommended contribution rate'. Under these relationships, the primary funding method itself is enough and suitable for a funding plan, which illustrates the aptness of the name primary funding method.

In reality, the actual future experiences are unlikely to emerge in accordance with the actuarial assumptions, so the value of the assets would not be equal to the amount of the actuarial liability. Therefore, although the normal cost is the regular pension cost, there must also be variations from the regular cost to allow for the differences between the actual experiences and the actuarial assumptions, i.e. the actuarial gain/loss. A supplementary funding method is designed to spread variations from the regular cost with the objective of limiting fluctuations in funding caused by the accumulated actuarial gain/loss. This fact illustrates the aptness of the name supplementary funding method, as a method supplementary to the primary funding method.

As a result, the recommended contribution rate is composed of two parts: the normal cost, determined by the primary funding method, and the adjustment to the normal cost for the actuarial gain/loss, specified by the supplementary funding method. Subsequently, it should be noted that the terms 'primary' and 'supplementary' are our own terms defined in connection with funding plans, so whenever there is no specific mention about funding methods, afterwards these refer to the primary funding methods.

2.1.2.4 Valuation basis - going-concern, run-off and winding-up

In the case of assessing the financial strength of the scheme, a framework of guiding accounting principles should be clearly stated in the scheme's annual report because the level of the financial strength resulting from the actuarial valuation would vary according to the assumptions relating to the methods used for assessing the scheme's assets and liabilities on the valuation date. There are two distinct assumptions commonly used by the actuary in the process of actuarial valuation: the assumption that the scheme will continue in full force indefinitely (i.e. the going-concern valuation basis) and the assumption that the scheme will be discontinued on the valuation date (i.e. the discontinuance valuation basis).

In our view, it is confusing to refer to 'discontinuance' because the meaning of discontinuance in the pension terminology is the cessation of contributions to a pension scheme leading either to winding-up or to the scheme becoming a frozen (or closed) scheme. Indeed, Lee (1986) states in section 26.6 that " In some cases of discontinuance it would not be feasible for the trustees and their successors to carry on the scheme as a closed investment trust for the benefit of the current and prospective beneficiaries. ... In other cases discontinuance might be interpreted in the sense that the trustees would continue the scheme as a closed investment trust."

In practice, if their sponsoring employers went into liquidation (or bankruptcy), large pension schemes would generally continue as closed schemes because there are practical limitations on the possibilities for a large scheme to transfer to another scheme and/or for the purchase of immediate and deferred annuities corresponding to the promised benefits in the open pension market.

In order to avoid these confusions, we shall subdivide the category of discontinuance valuation into run-off valuation and winding-up (or liquidation) valuation and further make a clear distinction between them, where the terms used below 'run-off' and 'winding-up' have been taken from Daykin et al. (1987).

Daykin et al. (1987) identifies three main set of assumptions that need to be made in relation to measuring financial strength expressed in terms of the operations of an insurance company.

These are termed 'going-concern basis', 'run-off basis' and 'winding-up basis', each of which is summarised below:

- Going-concern basis (i.e. a conventional approach for accounts prepared for shareholders, based on the assumption that the insurer will continue in business; hence liabilities shown in the accounts of the insurer are to be a best estimate and the assets are to be based on a book value or market value, respectively);

- Run-off basis (i.e. an approach usually adopted by supervisory authorities, based on the assumption that the insurer will cease to underwrite new business and run off its assets and liabilities; hence, liabilities and assets shown in the accounts of the insurer are to be cautious estimate and market value, respectively); and

- Winding-up basis (i.e. an approach rarely used for an insurance company, based on the assumption that the insurer will be liquidated and the new business will be prohibited almost immediately by a liquidator appointed by the court; hence, liabilities and assets shown for the accounts of the insurer are to be a best estimate of current value and a value realisable as quickly as possible, respectively).

Although there are many differences between an insurance company and a pension scheme, we believe that these bases described for measuring the financial strength in insurance companies (and the underlying concepts) can be applied to a pension scheme.

By combining these concepts and the meaning of discontinuance, we are then able to establish the following three kinds of valuation bases specific for assessing the financial strength of defined benefit pension schemes:

- Going-concern valuation basis is a conventional/classical valuation approach for management accounting purposes, based on the assumption that the sponsoring employer will continue his business (so, the new entrants will normally continue to be admitted) and the scheme will continue in existence with future contributions being received (i.e. the future lifetime of the scheme is infinite). Hence, this valuation would be useful for checking whether or not the current resources of the scheme are adequate to provide at least the best estimate of the scheme's liabilities (which would be the actuarial liability calculated on a given set of the actuarial assumptions) at each valuation date. The description 'optimistic' is used because this valuation excludes the potential possibility that the employer's business may cease to exist and then the scheme may not continue in existence.

We note that the following two scheme valuation bases are commonly founded on the possibility that the sponsoring employer is unable or unwilling to provide the pension scheme at any point in time (because of possible liquidation or bankruptcy). However, as mentioned above, the pension scheme does not necessarily go into immediate winding-up, so in this respect we can classify the discontinuance valuation basis into the run-off and winding-up valuation bases, that is,

- Run-off valuation basis is a valuation approach based on the assumption that the scheme will run on as a closed scheme without any further contribution rates, so the future lifetime of the scheme is finite. Hence, it would be useful for checking whether or not the current resources of the scheme is adequate to provide its liabilities estimated over a relevant, finite time horizon at each valuation date; and

- Winding-up (or liquidation) valuation basis is a valuation approach based on the assumption that the scheme has been discontinued at the valuation date (i.e. the future lifetime of the scheme are zero), so the trustees will wind up the scheme almost immediately. Hence, this valuation would be useful for checking whether or not the current resources of the scheme are adequate to provide the promised benefits assessed at any valuation date.

As a result, the approaches to actuarial valuation can be categorised as going-concern, run-off and winding-up (note that this classification may not be comprehensive).

2.1.2.5 Comments on each valuation basis

We approach the controlling the conflicts of interest between the employer and the trustees by considering first the choice of valuation basis. In practice, the real concern is which of these bases should be used for gaining an appreciation of the scheme's financial position. Before the actuary produces a full valuation on the chosen valuation basis, he will to some extent take into account the future conditions surrounding the scheme, such as the future prospect of the employer's business and the future variability of the scheme's assets and liabilities (though he is required to follow almost completely the current pension-related legislation or regulations).

These future conditions of the scheme would generally be reflected in the choice of methods employed in valuing the assets and liabilities, i.e. asset valuation methods and liability valuation methods. These valuation methods are described separately in the following sections 2.1.2.6 and 2.1.2.7, mainly in the light of the consistency between the valuations of the assets and liabilities and the degree of realism of the valuations of the assets and liabilities.

It should also be noted that whatever the valuation basis might be adopted, the common purpose behind using each valuation basis would be that both employer and trustees continuously support the pension scheme at least over the relevant time horizon (i.e. different time horizon for each type of valuation), while maintaining a sound financial position and stabilising the contribution rates (optimally if possible).

Finally, the distinctive points of each valuation basis can be summarised as follows:

(i) The going-concern valuation basis can be thought of as a regular management valuation approach to the scheme valuation and then both employer and trustees may focus mainly on regulating the speed at which the fund is built up during the remaining working lifetime of active members (which is usually referred to as the pace of funding) taking a long-term perspective. Hence, the actuary can have relatively more freedom in establishing the funding plan and the investment manager can have relatively more flexibility in choosing the investment policy than under the other two bases (although the trustees would provide general investment guidelines);

(ii) The winding-up valuation basis can be recognised as a strict supervisory approach to the scheme valuation and then both employer and trustees would be primarily concerned with meeting the statutory requirements for the security of the scheme on a short-term perspective rather than with controlling the pace of funding on a long-term basis. Hence, there may be a problem of mismatching between these short-term security requirements and the long-term funding plan, particularly in the case that the investment performance is highly volatile. For example, if the scheme has suffered from poor investment performance, the employer will have to make up the investment loss either from an unexpected rise in the regular costs over a permitted short period or from an immediate cash injection in order to satisfy the statutory measures of security. In this respect, the actuary may have relatively less freedom in establishing the long-term funding plan than under the other two bases, and also the investment manager will have relatively more restrictions on the investment policies than the other two bases (this may sometimes place a significant financial burden on the employer). Therefore, this valuation basis would represent a preferable approach in particular from the viewpoint of the scheme's trustees and beneficiaries, for reinforcing the security of the promised benefits; and

(iii) The run-off valuation basis can be regarded as intermediate between the above two extremes but it is much closer to the statutory supervision of the scheme in the sense that the scheme can be expected to continue as a closed fund, and then matching the existing liabilities and the current assets by size and term would be of vital importance to the future security of the scheme. Hence, both the funding plan and investment policy would need to be reviewed in the light of any estimated mismatch.

2.1.2.6 Asset valuation

Different from the liability valuation, the asset valuation is independent of the choice of funding method, so the asset valuation may be largely at the discretion of the actuary (of course, after a full discussion with the trustees and the employer, taking into account the asset valuation regulations). In practice, the actuary commonly employs and/or adjusts one of the following methods: discounted cash flow method, market value method and book value method.

As a preliminary, we note that the discounted cash flow method allows for the future profitability of the current assets, the market value method focuses on the current market value of the assets, while the book value method totally ignores the current and future market situations, and just depends on the original purchase price of each investment.

(i) Discounted cash flow method:

The current assets can be analysed in terms of their future profitability, that is the future sequence of investment receipts, made by way of dividends, interest payments, redemption monies, sale proceeds and tax rebates. The discounted cash flow method leads to a value for practical use that is defined as the sum of the present values of the projected future inflows of cash (i.e. the cash-flows arising from the emerging streams of investment income), at an appropriate discount rate (or set of rates). Accepting the uncertainty of financial markets, the projected cash-flows may not turn out to be the same as the cash-flows that actually arise. That is, in an uncertain real world, future investment policies (for investment incomes) are unlikely to be fixed with certainty. Hence, both the income-occurrence time and the amount of the corresponding incomes are in reality random, i.e. actual future cash-flows could be specified by a two-dimensional stochastic process with a probability distribution of time and amount (simply denote as {(time, amount); over a projected time horizon}).

For the calculations, we need a number of assumptions to be made on the stochastic process {(time, amount)} so as to establish the projected future cash-flows; usually, this projection would be generated in the form of the expected or mean cash-flows from the assumed initial investment structure, future investment policies and future funding plans. Inevitably, any kind of projection must be subjective and inconclusive because any selected projection is dependent upon the authorised investment manager's decisions and choices in respect of the investments and also is just one realisable result from a large number of such realisations that could have been generated by the stochastic process. The subjectivity and unreality in connection with such calculations can then be thought of as a disadvantage of the discounted cash flow method [see Fujiki (1994; p116)].

It should also be noted that this method values the assets under the assumption that the scheme is going to continue at least during the projection, so this method is applicable to a going-concern valuation and a run-off valuation, but not to a winding-up valuation. Furthermore, in the light of the scheme continuing, choosing an appropriate discount rate (or set of rates) is closely related to the nature of the projected cash-flows and the assumptions made. As the cash-flows are subjective, the discount rate (or set of rates) is somewhat subjective but needs to be consistent with the actuarial economic assumptions; in particular, the discount rate (or set of rates) may be chosen to be equal to the valuation discount rate rather than the (volatile) market-related discounted rates, where the valuation discount rate is, in general, estimated in the terms of the long-term expected future rate of investment return.

If the choice of discount rate (or set of rates) is consistent with the assumptions made in projecting the cash-flows as well as the valuation discount rate, then the discounted cash flow method for the assets is consistent with the normal method for the liabilities valuation that projects expected benefits outgoes and discounts at the valuation discounted rate. Therefore, the use of the discounted cash flow method can maintain consistency between assessing the assets and the liabilities in the cases of going-concern and run-off valuation (- this can be thought of as an advantage of the method).

The discounted cash flow method for the assets is usually applied in a deterministic way and then produces clearly and easily interpretable results. But such an approach does not address the problems of subjectivity and unreality. As an alternative, we may use a stochastic approach and adopt computer simulation techniques using different distributional assumptions about future possible investment polices and their resulting cash-flows where we may calculate approximately the mean and variance of the value of the assets that are expected on the stochastic framework. Such an approach would be more complex to operate and the result might be more difficult to interpret [see The Institute and Faculty of Actuaries (1994; pp8~9)]; however, it will enable the concept of variability to be incorporated in the valuation process. Of course, the resulting estimates may not match reality.

(ii) Market value method:

Unlike the discounted cash flow method, the market value method provides an objective and realistic value of the assets because it is designed to value the underlying assets at the market prices on the valuation date. So, this method may be suitable for measuring the financial strength of the scheme no matter what type of valuation basis is adopted - going-concern, run-off or winding-up valuation. Particularly, for a winding-up valuation this method can also maintain consistency between the valuations of the assets and the liabilities because the value of the liabilities under a winding-up valuation would be measured in terms of the market value of the accrued liabilities (i.e. with a transfer value into another pension scheme and/or annuity premiums in the annuity market) on the valuation date.

The general features of the market value method, objectivity and reality, depend on how reasonable it is to define market prices on the valuation date. The theoretical definition of the market prices of the assets on the valuation date would be the value realised when the assets were sold on the valuation date.

However, we may not be able to derive properly audited figures of asset values at a single point of time (such as at a valuation date) because the market prices, particularly in the case of liquid assets (e.g. equities and index-linked stocks), are liable to fluctuate continuously and on the other hand, particularly in the case of illiquid assets (e.g. properties), market prices would depend much on the subjective opinion of the participants in the transaction, and it may take a considerable time to dispose of illiquid assets. Either smoothing and/or adjusting of market prices of the assets could be used to cope with short-term market fluctuations in market value, although there would then be questions about whether the smoothed and/or adjusted market prices are actually realisable in the event of the transaction.

In practical calculations, the market value of the assets on the valuation date would be treated as the smoothed price of the assets over an appropriate period of time, for example some form of moving average of mid-market prices and/or as the adjusted price of the assets to reflect the changes in the historical movements of the open investment market sectors (broadly, gilt market, equity market and property market) where the assets has been invested over a fixed period (i.e. the adjusted price using appropriate indices of the market sectors) [see Dyson & Exley (1995) and TPFM (1984; p8)]. Thus, these variations of the market value method for practical calculations can be thought of as being somewhat subjective but we should note that they do focus on current market prices and then without loss of generality, we can say that the market value method is less subjective than the discounted cash-flow method.

(iii) Book value method:

The book value of an asset is the price at which the investment was originally purchased. From the viewpoint of stability in the asset values, the book value method may be regarded as the best approach to the asset valuation because it does not take into account any appreciation or depreciation of the investments on the valuation date. Furthermore, it is sometimes acceptable to assume that the random fluctuations in market value of an investment are (approximately) symmetric, over a long-term projection, about the book value of the investment; hence, the book value method would be more applicable in a going-concern valuation rather than in a run-off or winding-up valuation.

Relative to the earlier two methods, the book value method is of historical interest and is now no longer employed for actuarial valuations because it is less realistic than the market value method and less consistent than the discounted cash flow method. On the contrary, if we modify the book value method through the market level adjustment to the book value (i.e. allowance for unrealised appreciation/depreciation), then this variation of book value method may be more attractive to the actuary than the other two methods, particularly where the asset valuation process is too complicated for the calculation of a reasonable value of the assets and/or there are substantial delays in collecting accurate information for the valuation of the assets.

(iv) Conclusions:

The three distinct asset valuation methods have their own advantages and disadvantages, and thus the value placed on the current assets will depend largely on the actuary's view and on the purposes of the actuarial valuation. In particular, the valuation balance sheet audited on the basis of a market value valuation of the assets will give an objective and readily interpretable indicator of the security of the scheme. Throughout this thesis, the value of assets (i.e. the fund level) is assumed to be assessed on the basis of the market value method; hence, the mathematical notation for the value of the assets on hand presents the fund level measured in terms of the (smoothed or adjusted) market value of the underlying assets (simply, it shall be called the market-related fund level, where the term 'related' in the name indicates that carrying the assets at actual market value is subject to some practical limitations, as discussed above).

2.1.2.7 Liability valuation

Using the selected actuarial assumptions and funding method specific for the valuation process of the scheme, the actuary will calculate the theoretically desired level of its liabilities against which its assets are accumulated by the funding method (i.e. actuarial liability, which is a concept similar to the net premium reserve in life insurance mathematics). Since the actuarial liability is a function of a chosen funding method for a given set of actuarial assumptions, we can not regard its value as uniquely determined in the actuarial valuation, but must regard its value as one among many possible values that would be produced by numerous funding methods.

Although there are a variety of funding methods in use in the UK, producing a pair of actuarial liability and normal cost, a consideration of their philosophical approaches to the calculation of the actuarial liability falls into two broad categories: accrued benefit methods (e.g. Current Unit method and Projected Unit method) and projected benefit methods (e.g. Entry Age method and Attained Age method). We note that the Aggregate method is another prototype example of projected benefit methods but it does not involve either a normal cost or actuarial liability, so it is excluded here (note that in section 2.2.1, we shall describe the funding methods in common use in the UK, and the inter-relation between the normal cost and the actuarial liability as a function of a selected funding method).

To make a distinction between actuarial liabilities by accrued and projected benefit methods, we shall describe these as follows: for a given set of actuarial assumptions, the actuarial liability measured by the accrued benefit methods shall be called the accrued actuarial liability, while the actuarial liability measured by the projected benefit methods shall be called the projected actuarial liability. For the simplicity of our discussion, we assume as one part of the actuarial assumptions that the members in the scheme are partitioned into active members and retired members (i.e. pensioners), and the trust deed and rules specify only retirement benefits. Thus, the actuarial liability is broadly classified as follows, according to the funding method applied:

(i) Accrued actuarial liability is to be expressed as the sum of {present value of accrued benefits for active members at the valuation date (i.e. past service pension benefits)} and {present value of future pension payments to retired members at the valuation date}, either on the basis of current salaries (i.e. as in the Current Unit method) or on the basis of projected final salaries (i.e. as in the Projected Unit method). Hence, if the actuary is willing to protect the promised benefits (to which the (active and retired) members have been entitled) against the potential risk of insolvency of the employer, the accrued actuarial liability would be preferable as a suitable target of fund level, rather than the projected actuarial liability described below because the accrued actuarial liability principally disregards any future contributions and future increases in salaries (although the Projected Unit method does not work in this way because of making full allowance for projected future salary increases). Furthermore, for each valuation date, the actuarial liability on the Current Unit method would more closely represent the defined benefits than the actuarial liability on the Projected Unit method.

(ii) Projected actuarial liability is to be expressed as the sum of {present value of future pension payments to active members at the valuation date (i.e. total (past and future) service pension benefits)} less {present value of future normal costs for active members at the valuation date} and {present value of future pension payments to retired members at the valuation date}. Thus, the projected actuarial liability makes full allowance for the future accrual of benefits for active members on the basis of projected final salaries, in other words the currently active members are assumed to continue service up to the retirement age specified in the trustee deed and rules. Since the calculation of the projected actuarial liability is certainly affected by the projection of future normal costs and future salary increases, the actuary may focus more on regulating the pace of funding in the future rather than on securing the defined benefits.

As mentioned in Marshall & Reeve (1993), there has been a clear move from projected benefit methods towards accrued benefit methods in recent years in the UK. This trend can be interpreted as reflecting the relative emphasis that the supervisory authorities attach to the solvency or security of the scheme, so under this regime of supervision the accrued actuarial liability would be more preferable to the actuary than the projected actuarial liability as a target of fund level. As an example of this development in the UK, the Pension Scheme Surpluses (Valuation) Regulation 1987 introduced the use of the accrued actuarial liability for eliminating any surplus in excess of a specified upper limit of tax free funding defined in the regulation, where the actuarial liability is based on service up to the valuation date and makes allowance for projected final earnings for active members; here, this actuarial liability would be the same in concept as the actuarial liability based on the Projected Unit method.

Now, from the viewpoint of the security of the promised benefits, we consider the appropriate matching between the target for the fund level and the scheme valuation basis for the different types of valuation introduced in section 2.1.2.4.

Firstly, as mentioned above, the accrued actuarial liability would be suitable for the target for the fund level under a going-concern valuation. Further, allowing for the fact that the going-concern valuation aims mainly to regulate the pattern of funding, the actuarial liability on the Projected Unit method may be more adequate than the actuarial liability on the Current Unit method because of the allowance made for future salary increases. Furthermore, as a result of the UK preservation legislation (requiring preserved pensions to be increased at prescribed rates up to pension age), the Current Unit method is not currently appropriate in the UK [see GN26]. At this point, it should be noted that the possibility of discontinuance of the scheme is assumed to be almost zero under the going-concern valuation and that there would be little advantage in deriving assumptions to re-assess the actuarial liability according to the movements of the open market in pension policies (consisting of pension providers, such as insurance companies, building societies, unit trusts and banks). In other words, even though using an actuarial liability adjusted according to the movements of the open pension market (simply, we shall refer to this as the market value adjustment of the actuarial liability; for mathematical explanation, see section 3.4.1) may increase somewhat the confidence of the members in the process of funding and monitoring the scheme, this valuation adjustment is likely to add more uncertainty to the actuarial liability and then regulating the pattern of funding might be a greater burden to the actuary than before applying this adjustment.

Instead of allowing for the market value adjustment, the actuary may introduce some margins in the scheme parameters (estimated normally on long-term average assumptions) by reference to the way the economic and demographic conditions have generally moved over a period of interest in order to give some buffer against future demographic and economic uncertainties (specially, against short-term fluctuations around the long-term average) and then to maintain an adequate level of funding. Alternatively, he may change in a routine manner the actuarial assumptions by reference to the actual experience emerging, which is consistent with the view of Fujiki (1994) whose main results are described in section 2.1.2.2 (although this approach may, in practice, cause an extra administrative burden to pension professionals).

In contrast to the going-concern valuation, the run-off and winding-up valuations are both more concerned with the protection of the promised benefits against the possible ruin of the employer's business, rather than securing the accrued actuarial liability. So, the actuary would need to derive assumptions to reproduce the promised benefits in connection with an open pension market at a date of interest in order to establish the target for the fund level suitable for a practical safeguard against inadequate funding. For consistency with the assumptions introduced in the early part of this section, we also assume that the open pension market is broadly composed of the deferred annuity market for active members and the immediate annuity market for retired members (though deferred annuities are currently too "expensive" and there may not be a proper, competitive deferred annuity market, particularly for a small or medium size scheme with its assets below £50 million, see Collins (1992)).

In reality, there will inevitably be a difference between the promised benefits and the actual liability because the actual liability should take into account the pension market conditions. Here, the amount of the actual liability would, in an absolute sense, be the sum of the amount estimated in the deferred annuity market as transfer values for the active members and the cost estimated in the immediate annuity market as immediate annuities for the retired members. The exact value of the actual liability is unlikely to be available before carrying out the winding-up procedure used in the Courts because the bulk deferred annuity market value of a pension scheme would, in practice, be a matter of negotiation largely between receivers and liquidators, even though there is a competitive market between pension providers in terms of immediate annuity rates [see Daykin(1987)]. So, the precise calculation of the actual liability can be thought of as being outside the control of the actuary and thus, for the same reasons as mentioned in section 2.1.2.6, the amount of the actual liability would be regarded as a market-related value of the promised benefits. In order to reduce to some extent the gap between the promised benefits and the actual liability, the actuary may thus need to apply the so-called market value adjustment to the promised benefits (of course, this valuation adjustment is going to be carried out according to the manner prescribed in the trustee deed and rules).

Following on from this argument, the actuary could suggest the following variations for establishing a more realistic target of fund level under each of the run-off and winding-up cases:

- For the winding-up valuation basis, one possible variation of the actual liability is what is obtained by means of the market value adjustment applied to the preserved benefits, produced on the basis of the current (transfer and immediate annuity) market situations (simply, we shall call this variation the market-related current liability). That is, if the purpose of the actuarial valuation is to ensure the continuing security of the promised benefits at all times (i.e. the main purpose of a winding-up valuation), the market-related current liability should be exactly computed on the seller-bidding and purchaser-offering basis in the open pension market, i.e. should be equal to the actual liability (though it is not necessarily available). Thus, this kind of valuation adjustment is likely to cause a significant funding burden to the employer, particularly whenever the scheme suffers from poor investment performance and/or the (transfer and immediate annuity) costs in the open pension market are going up; and

- For the run-off valuation, one possible variation of the actual liability is what is re-assessed by means of the market level adjustment, produced on the basis of the cautiously estimated (transfer and immediate annuity) future market trends over a period of interest on the assumption that the scheme continues as a closed investment fund (simply, we shall call this variation as the market-related cautious liability). Thus, this valuation adjustment may be regarded as a liability valuation intermediate between the above two extremes (i.e. valuations for accrued actuarial liability and market-related current liability), since this adjustment can be expected to balance effectively the conflicting interests between employer and trustees.

It should also be noted that the market-related current and cautious liabilities have been discussed on the basis of a given set of actuarial assumptions. As for the going-concern valuation, we can similarly discuss the choice of assumptions. Thus, in order to reduce the difference between the promised benefits and the actual liability, directly adjusting the promised benefits by reference to the current or future market situations would have a similar effect either to introducing into the actuarial assumptions some margins, particularly in the scheme economic parameters, or to introducing in a routine manner the basis which the competitive pension providers are normally employing for the pricing of immediate and deferred annuities contracts. These ideas are consistent with the views of Thornton & Wilson (1992) and Collins (1992), which will be discussed separately in section 3.3.3.

In summary, the asset and liability valuations are important issues in evaluating the security of the scheme. Of course, these valuations will largely be at the discretion of the scheme actuary after full consideration of the views of the employer and the trustees. In accordance with the relative suitability to their conflicting interests, the actuaries (and supervisory authorities) aim to find a balancing point satisfying both employer and trustees at the same time (if possible). We believe that the market-related cautious liability will serve as one proper approach settling the conflicting interests in the liability valuation.

As a result, we believe that the triple combination (run-off valuation, market value method, market-related cautious liability) may be viewed as the balancing point (or centre of gravity) in the range of all possible ordered triple combinations of {going-concern valuation, run-off valuation, winding-up valuation} $\times$ {discounted cash flow method, market value method, book value method} $\times$ {actuarial liability, market-related cautious liability, market-related current liability}, where the notation '$\times$' denotes the Cartesian product.

2.1.2.8 Security measure for the scheme

The results of the actuarial valuation will normally be integrated and clarified by producing the value of an appropriate security measure. In practice, the size of the scheme's assets in relation to its liabilities at the valuation date is of vital importance to its security. So, the security measure would be a real-valued function of two variables, expressing the value of the scheme's assets and the amount of its liabilities. Broadly, there are two types of security measure: one is based on the scheme's funding level and the other is based on its solvency level. In the strict sense of measuring security, the solvency level would be a more rigorous criterion than the funding level. This will be clarified in the discussion below. Following this line of argument, we can define both surplus/deficiency and solvent-capital/insolvent-debt, which are specified in the end of this section.

Up to now, in the UK current pension scheme rules have not had any general requirement that defined benefit pension schemes should be funded to any specific funding level, let alone a specific solvency level. After the Maxwell affair, the UK government, consulted primarily on the minimum solvency requirement (MSR) proposed by the 1993 report of the Pension Law Review Committees (PLRC) "Pension Law Reform", and has been contemplating a new framework with the main purpose of markedly reinforcing the security of defined benefit pension schemes. As a first regulation for the security guidance, the government introduced, in the Pension Act 1995, the minimum funding requirement (MFR) that the value of the assets of the scheme is not less than the amount of the liabilities of the scheme (which will come into force from 6 April 1997) - the methods of valuation of the assets, the liabilities and the actuarial assumptions set out in regulations [see Greenwood & Keogh (1997)]. Further, the Pension Act 1995 establishes The Occupational Pensions Regulatory Authority as a supervisory authority, which will have the function of monitoring the protection of the interests of the members of occupational pension schemes and will take over most of the functions of the existing Occupational Pensions Board (in section 3.2.2, we will discuss the basic ideas about the MSR and the MFR). In this section, we give a general description of the concepts of funding and solvency levels, which is intended to be a preliminary to the more detailed discussion of section 3.2.2.

(i) Funding level:

The funding level (or funding ratio) of the scheme is commonly defined as a real-valued function expressing the proportion (or ratio) at a specified date of the actuarial value of liabilities (i.e. actuarial liability) that is covered by the actuarial value of assets held. The nature of the funding level means that it may inevitably fluctuate up and down according to technical changes either to the liability valuation method and/or to the asset valuation method. If this definition is related to the valuation of the scheme on a long-term, going-concern basis, the funding level specified above may have to be termed the going-concern funding level.

In practice, the choice of valuation basis may have a major effect on the value of the funding level because according to the valuation basis, the corresponding liability may be identified differently as discussed earlier in section 2.1.2.7. Therefore, in a stricter sense, the term funding level should be used in relation to the results of a chosen valuation basis. Further, there is no loss of generality in assuming that the funding level is non-negative.

We recommend that the funding level should be extended to cover all kinds of valuation bases such as going-concern, run-off and winding-up valuation bases in order to operate fairly as a security measure of the scheme. It is worth recalling that in view of the security of the scheme, the value of the assets of the scheme would be taken as the market-related fund level (see section 2.1.2.6); the amount of the scheme's liabilities on a going-concern valuation basis would be taken as the accrued actuarial liabilities rather than the projected actuarial liabilities; and also the amount of the scheme's liabilities on the other two bases would be taken as the market-related value of the promised benefits, that is, the market-related cautious liability on a run-off valuation basis or the market-related current liability on a winding-up valuation basis [see section 2.1.2.7].

In consequence, the funding level is intended to be a reliable indicator of the security of the scheme at each valuation date, but in practice these levels can not be determined uniquely because of their dependency on the choice of assumptions and methods for the asset and liabilities valuations. Following our discussions of sections 2.1.2.6 and 2.1.2.7, we suggest the following funding levels as appropriate for each valuation basis adopted (although (a) and (b) have inconsistency problems, see section 2.1.2.7):

(a) {funding level on going-concern valuation basis} (≡ {going-concern funding level})

= [market-related fund level] / [actuarial liability on going-concern basis];

(b) {funding level on run-off valuation basis} (≡ {run-off funding level})

= [market-related fund level] / [market-related cautious liability]; and

(c) {funding level on winding-up valuation basis} (≡ {winding-up funding level})

= [market-related fund level] / [market-related current liability].

(ii) Solvency level:

Public interest in the solvency of occupational pension schemes first came to prominence in the UK with the property and equities' crash of 1973~4 which led to the real value of the average scheme's assets being cut by about 60% (statistical source: Pension Fund Indicators 1991). At the end of 1974, pension professionals were required to investigate whether or not the resources of the scheme would be sufficient to cover the promised benefits even if this kind of significant financial loss forced the sponsoring employer into the discontinuance of the scheme. As governmental guidance on the security of the promised benefits, the Social Security Pensions

Act 1975 set out the actuary's certificate to the Occupational Pensions Board (so-called Standard Solvency Certificate or Certificate A) for contracted-out schemes, in which the actuary is required to state that in the event of the scheme being wound-up within the five year period following the date of signing of the certificate, the resources of the scheme will be sufficient financially to meet in full the statutory requirements prescribed in Certificate A.

However, Certificate A has some weaknesses in terms of the security on winding-up. The first point is that the certificate relies on the individual opinion of the actuary, so there may be a problem of inconsistency in judgements between one actuary and another. Secondly, the certificate covers a five year period (though it has to be renewed at least eighteen months before its expiry, or if its validity is affected by any changes), which is a rather long period over which to project, with accuracy, the movements of future contributions, liabilities and assets (although five year is becoming a standard period for certification, as noted in the Social Security Pensions Act 1975). Thirdly, in the absolute sense of a winding-up scheme valuation, the financial adequacy of the scheme could be measured by the current financial status of the scheme without allowing for the future cash-flows, as mentioned in section 2.1.2.4. Lastly, the certificate is only for contracted-out schemes. In our opinion, any kind of security regulations should be extended to cover all defined benefit pension schemes. Hence, Certificate A seems to be founded on the basis of a run-off valuation, not on the basis of a winding-up valuation, in the sense that it requires a finite time projection into the future.

Consequently, the impact of the subjective judgement of the actuary, inherent in Certificate A, should be reduced. Thus, for the purpose of the consistency between the solvency assessments of one scheme and another, solvency should be required to be certified on an exactly prescribed framework, such as on the basis of the minimum solvency requirement (as proposed by the PLRC) or the basis of the minimum funding requirement (as enacted by the Pension Act 1995). This would be consistent with the view of Daykin et al. (1987) expressed that "... it may be true to say that a company is solvent if the supervisor says that it satisfies his requirements.

Solvency in this sense is perhaps best described as 'meeting the statutory solvency requirements'...". In this respect, a 100% statutory minimum solvency level target would be the standard supervisory guideline as the minimum safety ratio for protection against less favourable experience in the future.

The solvency level can then be thought of as a supervisory measure by which the supervisor judges whether or not the financial strength of each scheme is sufficient to meet its requirements (simply, a supervisory measure for solvency valuation). If the statutory minimum funding level is not imposed by the supervisor, there would be no relation between funding level and solvency level, and further, there would be no target of fund level for solvency valuation. In other words, the solvency level is subject to the requirement imposed by the supervisor, such as a statutory minimum funding level. Using the funding level and statutory minimum funding level, we can define the solvency level as

{solvency level} = {(going-concern, run-off and winding-up) funding level} / {statutory minimum (going-concern, run-off and winding-up) funding level}, respectively.

In the particular case that the statutory minimum funding level is 100%, then the solvency level is equal to the funding level. Further, from the non-negativity assumption of both funding level and statutory minimum funding level, the solvency level is also non-negative.

Or equivalently, we may, in general, define the solvency level from the standpoint that the assessment of the solvency of pension schemes is a matter of political decisions. In other words, if any kind of statutory requirement is phased in, it will prescribe the broad principles for the valuation of the assets and liabilities through regulations and/or legislations (simply, the broad principles for solvency valuation)

{solvency level} = {fund level for solvency valuation} / {liability level for solvency valuation}.

In the following paragraphs, we shall take our discussion forward in the light of the first definition in order to show some relationship between the funding level and solvency level.

Subsequently, on each (going-concern, run-off and winding-up) valuation basis, we shall refer to the scheme as being 'solvent' where the calculated funding level is greater than the imposed statutory minimum funding level (i.e. solvency level > 100%) and 'insolvent', the opposite of solvent, where the calculated funding level is less than the imposed statutory minimum funding level (i.e. 0% ≤ solvency level < 100%).

Remark 2.1: If we follow the view of Collins (1992) about solvency that "... the term 'solvency' is used in the context that if a scheme were to wound-up there would be sufficient assets to buy-out the liabilities fully before any augmentation, with an insurance company.", then the solvency level would be defined as the level of the premium costs to purchase the (active, deferred and retired) members' accrued pension rights in the prevailing non-profit deferred and immediate annuity market that is covered by the market value of the scheme assets. This can be considered as a strict solvency measure in the view of measuring exactly the continuing security of defined benefit pension schemes.

(iii) surplus/deficiency vs. solvent-capital/insolvent-debt:

In general, the supervisory authorities may be required to use a method of assessing the financial strength of the scheme which is consistent between one scheme and another as well as giving a clear idea of the financial differences between one scheme and another. In this respect, the ratio type measure, either funding level or solvency level, would be effective. On the other hand, the ratio-type measure can be easily one-to-one transformed into the difference-type measure, defined below. The difference-type measure may be useful for accounts and audit of schemes, whereas the ratio-type measure may be useful for supervising consistently schemes.

Now, we shall define two difference-type measures; one is the alternative measure of the funding level (we shall call it as the unfunded liability) and the other is the alternative measure of the solvency level (we shall call it the insolvent liability).

Firstly, the unfunded liability is commonly defined as the difference between the actuarial liability and the fund level (as defined in Trowbridge (1952) as "The portion of the accrued liability not offset by assets is called the unfunded accrued liability."). But, this definition is oriented towards a going-concern valuation. For the same purposes as the extended definition of the funding level, we shall extend the above definition to cover the run-off and winding-up valuations.

The unfunded liability is the real-valued function presenting the difference between the amount of the liabilities of the scheme and the value of its fund level in relation to the results of the valuation basis used. Each valuation basis will contain the proper assessment of the corresponding assets and liabilities at the valuation date.

Subsequently, we can also define surplus and deficiency in line with the new definition of unfunded liability. That is, a negative unfunded liability shall be called a surplus and a positive unfunded liability, i.e. the opposite of surplus, shall be called a deficiency (or deficit).

Note that surplus and deficiency are concepts which are often referred to but which have still caused some trouble to the actuarial profession, in particular in the UK. Even though there have been several comments about the surplus/deficiency, our definition of surplus/deficiency properly reflects the view of Wilkie (1986) mentioned that "surplus has to be measured in relation to some funding target" in the light that since the term 'funding target' (in our term, the target of fund level) is likely to be differently specified in each valuation basis, our definition of surplus/deficiency is flexibly adapted to the choice of the valuation basis. Furthermore, the actuarial gain/loss concept, widely used in the USA and Canada, can also be extended to be in line with our definition of surplus/deficiency [for a detailed gain/loss analysis, see Lynch (1979) and Dufresne (1989)].

Secondly, we shall define the insolvent liability as a counterpart of the unfunded liability. In a similar way to defining the solvency level, the insolvent liability can be specified on the condition that the statutory minimum unfunded liability is determined a priori. The insolvent liability is then the sum of the calculated unfunded liability and the statutory minimum unfunded liability. That is, the insolvent liability can be defined as

{(going-concern, run-off and winding-up) insolvent liability} = {(going-concern, run-off and winding-up) unfunded liability} + {statutory minimum (going-concern, run-off and winding-up) unfunded liability}, respectively.

As the counterpart of surplus/deficiency, we can also define the solvent-capital/insolvent-debt. That is, a negative insolvent liability shall be referred to as the solvent-capital, which means that the scheme is solvent with spare resources equal to the amount of solvent-capital, and a positive insolvent liability, the opposite of solvent-capital, shall be referred to as the insolvent-debt, which means that the scheme is insolvent with a shortfall equal to the amount of insolvent-debt that should be made up by the employer within a permitted period.

It is worth finally noting that in the UK, funded defined benefit pension schemes are entitled to tax relief (i.e. investment income and realised capital gains within the fund accrue free of income and capital gains tax). In such a case, the supervisory authorities (or Government) may need to introduce an upper limit of tax free funding if a scheme were in surplus and the solvent-capital were not reduced or eliminated within an allowed period from the viewpoint of controlling any undue level of systematic pension fund excess that might be generated; for example, the Pension Scheme Surplus (Valuation) Regulations 1987 can be considered as a practical legislation for this kind of tax penalty, in which the allowed period is fixed at within five years. Therefore, the scheme actuary would make the following suggestions for use of pension fund surplus or solvent-capital in order to avoid the tax chargeable on the investment income on the assets in excess of the tax exempt funding level:

(a) Reduction in (employer and/or employee) contributions (for example, spreading the surplus evenly over an allowed period) or implementation of so-called (employer and/or employee) contribution holiday; and/or

(b) Improvement in the quality of defined benefits; and/or

(c) Transfer-back into the sponsoring company; and/or

(d) Adjustment of the actuarial valuation basis, in particular using less conservative actuarial assumptions and/or switching the valuation basis from going-concern to run-off or winding-up; and/or

(e) Switching from a projected benefit method to an accrued benefit method (if appropriate).

Remark 2.2: Suggestion (d) may not be helpful with a period as short as the five-year period prescribed by the supervisory authorities, since it is possible for the scheme's short-term actual experience to be even better/stronger than the newly adjusted actuarial valuation basis; and Suggestion (e) is consistent with the view of McLeish & Stewart (1993) expressed that "... in order to avoid the tax penalty which would now result from their holding 'excessive surplus' in their funds, schemes which had hitherto been using prospective benefit methods of valuation have reduced their funding levels by switching to accrued benefits methods, particularly the Projected Unit Method." (note that the term prospective benefit methods is an alternative, widely used name for projected benefit methods).

To summarise this section, the security measures, funding level, solvency level, unfunded liability and insolvent liability, are defined in relation to the particular valuation basis used. The statutory minimum funding level and statutory minimum unfunded liability each can be regarded as providing additional guidance on the level of security needed to protect against future uncertainties, determined primarily by the solvency policy of the supervisory authorities. In the general case that the statutory minimum funding level = 100% (or equivalently, the corresponding statutory minimum unfunded liability = 0), the solvency level is exactly the same as the funding level (or equivalently, the insolvent liability is exactly the same as the unfunded liability). Under the adopted valuation basis, the members in the scheme will have almost complete confidence in the security of their rights in the case that the valuation result seems to be sufficient to provide protection against the possibility of the employer's insolvency, that is, whenever the resulting funding level or solvency level ≥ 100%, or equivalently the resulting unfunded liability or insolvent liability ≤ 0. Lastly, the concept of solvency can be regarded as equivalent to the concept of security adopted by the supervisory authorities.

2.1.3 Risks in connection with funding plan

Pension provision inevitably faces many different sources of uncertainty about the values of the parameters used in the pension scheme (i.e. the scheme parameters). One of the sources of uncertainty that is most difficult to manage is the investment market returns in the sense that the returns on the assets side have to be taken as the market return actually realised from the asset allocation decision in increasingly volatile investment markets, while the valuation rate of return on the liabilities side will usually be determined on the basis of expected long-term average return on investments in the future.

Although investment decisions may be implemented successfully on the basis of a diversified portfolio, the expected return may not turn out to be the same as the return that actually arises. In this respect, it is worth noting Haberman (1994) expressed that "in recent experience, one of the principal sources of surplus or deficiency has been the rate of investment return on pension scheme assets."

The above discussion is applicable to the other scheme parameters. As a result, the actuary must confront the possibility that the actual outcomes may differ from what was expected in connection with the scheme parameters.

It should finally be noted that the actual outcomes may broadly take one of three different scenarios; most likely (i.e. average), better than most-likely (i.e. optimistic) and worse than most likely (i.e. pessimistic). Now, we shall introduce the term 'risk', which will provide a helpful and clear measure of the future uncertainties involved in the funding plan. Further, we shall interpret the main purpose of funding as managing both the stability of the contribution rate stability and the security of the rights to benefits, as introduced by Haberman (1997) in which he categorises the main risks involved in a defined benefit pension scheme into two groups, the contribution rate risk and the solvency risk, respectively.

2.1.3.1 Definition of risk

The term 'risk' has been defined in various ways. The main concept of risk is related to the inexactness in forecasts and/or projections of the future situations.

In the area of control theory, Dorf (1992; section 12.1) defines risk as "uncertainties embodied in the idea of unintended consequences of the design." From the viewpoint of risk management, Williams & Heins (1971; Ch. 1) define risk as "the variation in the possible outcomes that exists in nature in a given situation", which is consistent with the view of Houston (1964) that "To the insurer, risk is a function of the variation in the pure premium distribution . . . and may be defined by the standard error of the mean of the pure premium distribution." Further, they define risk management as "the minimisation of the adverse effects of risk at minimum cost through its identification, measurement and control." In the light of the scheme valuation, we may define risk as follows, which would, in concept, be in line with the above definitions:

The term 'risk' may be defined as the objective volatility measured by the gap between the actually realised results and the designed potential (or most likely expected) results, encountered by the unfavourable or unanticipated (economic and/or demographic) experience.

In addition, our definition of risk might be suitable for the mathematical assessment of risk because the term 'objective volatility' can be understood as the dispersion or variation of the possible values of the gap caused mainly by the scheme parameter uncertainty, such as a sample variance under certainty or a stochastic variance under uncertainty.

It is finally worth noting that in the area of financial economics, risk is, in general, categorised as unsystematic risk (or diversifiable risk) and systematic risk (or non-diversifiable risk); the unsystematic risks are those that can be almost completely eliminated by holding a well-diversified portfolio of investments, while the systematic risks are those that can not be reduced to zero by a diversified portfolio of investments [see Hull (1993; p70)]. Therefore, it is the systematic risk that the investment manager should pay expend efforts to control.

2.1.3.2 Contribution rate risk

Contribution rates generated from a particular funding plan are used to pay current benefits and/or are invested in a diversified portfolio of investments that are intended to provide the future benefits prescribed in the trust deed and rules. Different from a money purchase pension scheme, a defined benefit pension scheme is, in principle, required to provide the predetermined benefits as they arise, explicitly independent of the fund level of the scheme accumulated up to their due date. So, the employer will inevitably be exposed to the risk of fluctuating cash-flows in contributions, mainly caused by the systematic risk associated with the investment market.

In view of aversion to the contribution rate risk alone, the actuary will have two options for the employer; one is to switch out of a defined benefit scheme into a money purchase scheme and the other is to lessen the exposure of the pension fund to the risk of significant, undesigned variation in contribution rates (i.e. to maintain the contribution rate stability). The first option is a matter for the employer's decision. The second option is a matter of actuarial management that is a complicated problem involving largely the actuarial assumptions, the asset allocation decisions and the funding policy. However, this stability problem may, in aggregate, be identified by introducing the concept of contribution rate risk.

Following the definition of risk, we may define the contribution rate risk as the objective volatility in cash-flows measured by the gap between the actual cash-flows in contributions and the designed potential (or most likely expected) cash-flows in contributions over the time horizon of interest, that might be finite horizon (e.g. five, ten and twenty years ahead) or infinite horizon. For an example of the stochastic formulation for the contribution rate risk, Haberman (1997) defines the contribution rate risk as the variance of the infinite sum of the present value of future contributions discounted by the valuation rate of interest. Note that his mathematical definition is rather similar to the performance index (or cost functional) in control theory, which will be clear in Chapter 4.

In view of being more faithful to our definition of risk, we may mathematically specify the contribution rate risk as follows:

(i) <u>In the deterministic case</u> (i.e. risk under certainty):

$$\{\text{contribution rate risk at time t}\} = \{(\text{actual contribution rate paid into scheme at time t}) - (\text{designed potential contribution rate at time t})\}^2 \quad \text{and}$$

$$\{\text{contribution rate risk over a period [s, T], } s<T<\infty\} = \sum_{t=s}^{T}(\text{contribution rate risk at time t}),$$

in discrete time; and

(ii) In the stochastic case (i.e. risk under uncertainty):

{contribution rate risk at time t} = E {(actual contribution rate paid into scheme at time t)

- (designed potential contribution rate at time t)}^{2} and

{contribution rate risk over a period [s, T], s<T<∞} = $\sum_{t=s}^{T}$(contribution rate risk at time t),

in discrete time, where the notation 'E' denotes the stochastic expectation.

Alternatively, it is possible to express the formulation in continuous time by replacing the summations with integrals.

It is worth noting that the designed potential contribution rate can be regarded as an actuary's target for the contribution rate (e.g. normal cost) and the term 'the objective volatility in cash-flows' in our definition is mathematically replaced by a squared loss function with the aim of inducing a high penalty for large gaps between the actual cash-flows in contributions and designed potential cash-flows in contributions but a relatively low penalty for small gaps.

Moreover, in the specific case that {designed potential contribution rate} = E {actual contribution rate paid into scheme}, then we know that {contribution rate risk at time t} = Var {actual contribution rate paid in scheme at time t}, in which the notation 'Var' denotes the variance.

As a consequence, the criterion suitable for maintaining the contribution rate stability over a particular time horizon will be orientated toward minimising the contribution rate risk over the same time horizon.

2.1.3.3 Solvency risk

As outlined in section 2.1.2.8, making an assessment of the solvency of the scheme is concerned with comparing the value of the scheme's assets with the amount of the scheme's liabilities, in which the valuation method for the scheme's assets and liabilities has to be clearly prescribed by regulation or legislation. Then, it is worth starting with the identification of risks associated with the solvency of the scheme. The described risks below are originally based on the well-known risks of insolvency involved in general insurance [see Bulmer (1994) and the risk matrix identified by Lewin et al. (1994)], but we re-categorise them in view of our focus on defined benefit pension schemes. Broadly, the risks to solvency associated with the scheme can be divided into two distinct types, investment risk and non-investment risk:

(i) <u>Investment risk</u>:

- asset value risk (Ec, Fc and Lc) - - the risk of a substantial fall in the market-related value of the scheme's assets, primarily caused by the failure of the asset allocation decisions;

- asset income risk (Pc) - - the risk that the investment returns will not paid by the default of borrowers; and

- matching risk (Fc and Pc) - - the risk that the nature of the scheme's assets is inappropriate, given the nature of the scheme's liabilities;

(ii) <u>Non-investment risk</u>:

- liability risk (Ec, Fc and Lc) - - the risk that for the promised benefits of the members, the premium costs offered by pension providers or the transfer values offered by another permitted pension scheme will increase abruptly and largely, especially under run-off and winding-up valuation bases;

- regulatory or tax risk (Lc) - - the risk that there will be a significant adverse change in the scheme regulations or in the tax laws, in particular for the income and capital gain tax;

- sponsoring employer risk (Bc and Pc) - - the risk that the sponsoring employer will be subject to a large trading catastrophe (e.g. the European storms of 1987 and 1990) and thus the employer's contribution will not be available or the scheme will be wound up with the resources of the scheme insufficient to cover the promised benefits; and

- management risk (Pc) - - the risk of inappropriate auditing, including fraud and misappropriation (e.g. Maxwell affair), a large excess of administrative costs and inadequate planning of recruitment and salary structure (e.g. improper rise in salary and benefits, and undesirable structure of workforce, in relation to the scale of the sponsoring company);

where, each capital letter in the brackets indicates the underlying main cause of the corresponding risk, resulting in future unpredictability, that is,

Bc = Business cause, such as demand failure, premature obsolescence and competition weakness;

Ec = Economic cause, such as unfavourable retail price inflation, market premium cost inflation and market investment returns;

Fc = Financial cause, such as unbalanced asset allocation decisions in connection with the scheme's liabilities and inadequate assumptions for the scheme parameters;

Lc = Legislation cause, such as the removal or reduction of taxation advantages, restriction on the asset allocation decisions and some extreme provisions enacted by Acts, especially in relation with the asset/liability valuation (e.g. some regulations or legislations established through Social Security Act, Financial Act or Pensions Act); and

Pc = Project cause, such as improper planning and control in employment, human error or incompetence of pension professionals, including personnel manager and employer, and improper labour relations.

The above described risks each affect, both directly and indirectly, the scheme's solvency position; then, we may call collectively these risks the solvency risk of the scheme. Further, employer, trustees and pension professionals will be concerned with the elimination or reduction of the solvency risk. However, their common and ultimate aim will be to satisfy the solvency requirement imposed by the supervisory authorities subject to a reasonably allowed level of risk in the sense that the above risks can not completely be eliminated through risk management.

So, the desired feature of the imposed solvency requirement would principally be that the chosen solvency measure (e.g. solvency level, funding level, insolvent liability or unfunded liability) is bounded by certain allowed limits, together with a recommendation to follow a specified statutory solvency standard; for example, a 100% solvency level or funding level, or zero insolvent liability or unfunded liability. In other words, the published value of the solvency measure at each valuation date will be required to satisfy firstly, the bound condition such that {allowed minimum value of solvency measure} < {value of solvency measure resulting from the scheme valuation at each valuation date} < {allowed maximum value of solvency measure} and secondly, the guiding condition such that {value of solvency measure resulting from the scheme valuation at each valuation date} is preferred to be equal to {statutory solvency standard at each valuation date}.

Therefore, combining our definition of risk and the statutory solvency requirement in the side of supervisor, we may define the solvency risk as follows.

Solvency risk of the scheme may be defined as the objective volatility in cash-flows measured by the gaps between the actual cash-flows in selected solvency measures and the statutory cash-flows in solvency standards over a time horizon of interest, that might be finite or infinite.

As an appropriate mathematical formulation for the solvency risk, we may suggest the following, which is consistent with the suggested mathematical formulation for the contribution risk.

(i) In the deterministic case (i.e. risk under certainty):

{solvency risk at time t} = {(actual solvency measure resulting at time t)

- (statutory solvency standard imposed at time t)}^{2} and

{solvency risk over a period [s, T], s < T < ∞} = $\sum_{t=s}^{T}$(solvency risk at time t), in discrete time;

(ii) In the stochastic case (i.e. risk under uncertainty):

{solvency risk at time t} = E {(actual solvency measure resulting at time t)

- (statutory solvency standard imposed at time t)}^{2} and

{solvency risk over a period [s, T], s < T < ∞} = $\sum_{t=s}^{T}$(solvency risk at time t), in discrete time.

Alternatively, it is also possible to express the formulation in continuous time just by replacing the summations with integrals.

As mentioned in the discussion of contribution rate risk, we can provide the same interpretation about the squared loss function formulation. In addition, the statutory solvency standard (imposed by the supervisor) can be thought of as the supervisor's guideline or target for the potential measures of the solvency of the scheme. Moreover, in the specific case that {statutory solvency standard} = E {actual solvency measure}, then we know that

{solvency risk at time t} = Var {actual solvency measure resulting at time t}.

In conclusion, the criterion suitable for keeping the security of the promised benefits over a particular time horizon will be orientated toward minimising the solvency risk over the same time horizon.

2.1.3.4 Conclusion

Maintaining the stability of contribution rates corresponds conceptually to minimising the contribution rate risk, while maintaining the security of promised benefits corresponds conceptually to minimising the solvency risk; here, the solvency and contribution rate risks each are formulated by introducing heavy penalties for large departures from their respective targets but relatively small penalties for small departures from their respective targets.

Thus, achieving our funding purpose (i.e. coping with both the security and the stability according to their relative importance/weight) over a specified projection period corresponds to minimising the weighted sum of the contribution risk and the solvency risk over the projection period [for a mathematical formulation, see section 4.1.1].

2.1.4 Life-cycle of defined benefit pension scheme

With regard to the stages of the 'life-time' of a defined benefit pension scheme, it may be possible to introduce three distinct types of scheme: Young, Mature and Declining. This classification is virtually based on the exponential progress of a membership in the scheme, which will be clear in the following sections.

To begin with, we need to introduce two important concepts of population theory associated with the exponential progress of the membership; stable population and stationary population. In actuarial applications of population theory, we will use the two terms population and membership interchangeably.

2.1.4.1 Stable membership vs. Stationary membership

We study two fundamental patterns of growth in the number of members in a defined benefit scheme in the light of mathematical demography. Although it is possible to describe them on the basis of both discrete-time and continuous-time, here we adopt the discrete-time approach because it is more understandable in connection with a life table. Also, the mathematical analysis and properties derived in the discrete-time approach will be directly applicable to the continuous-time approach. For a discrete-time approach, we shall employ some assumptions that are based on the study of a deterministic survivorship group or cohort, represented by a given life table with age measured in years.

The assumptions are as follows:

(i) All new entrants join the scheme at age a, where $a > 0$;

(ii) Survivorship in the scheme is determined by the time-independent survivorship function from the given life table derived from best estimates of rates of death, which indicates the expected number, or number, of survivors at an age y (denote as l_y) where $y = a, a+1, \ldots$;

(iii) The radix of the function l_y is chosen so that l_a represents the number of new entrants in the scheme at age a at time 0;

(iv) The number of new entrants grows geometrically over time t; and

(v) The assumptions (i)~(iv) have been applied for a sufficiently long time so that the distribution of each age in the scheme during the intervaluation period (t, t+1) becomes stable.

The above assumptions enable us to establish the following dynamic growth function of order 1, for all $t \in \{0, 1, 2, ...\}$ and $y \in \{a, a+1, a+2, ...\}$. That is, the growth in the number of members during (t, t+1) is to be measured in terms of a dynamic membership growth function (denote as f) satisfying the following time-varying homogeneous recurrence equation of order 1. Since f is dependent on the time of entry into the scheme (i.e. $t+a-y \equiv z$) and l_a plays the role of the radix, then we have

$$f(z+1) = (1+im_{z+1}) \cdot f(z) \quad \text{with the initial condition } f(0) = 1$$

where, im_{z+1} denotes an annual % growth rate of the number of members aged y which is defined as the real growth in f during the intervaluation period (t, t+1), i.e. $im_{z+1} = [f(z+1) - f(z)] / f(z)$.

Without loss of generality, we can assume that $\text{Prob}[1+im_{z+1} > 0, \text{ for all } z] = 1$, and thus there exists α_{z+1} satisfying $(1+im_{z+1}) = \exp(\alpha_{z+1})$ for all z. So, we obtain the solution of the above equation such that

$$f(z) = \begin{cases} \exp(\alpha_z + \alpha_{z-1} + ... + \alpha_1) & \text{if } z \geq 1 \\ 1 & \text{if } z = 0 \\ \exp(\alpha_z + \alpha_{z-1} + ... + \alpha_{-1}) & \text{if } z \leq -1. \end{cases}$$

Then, the number of members aged a at time t is given by $l_a \cdot f(t)$ because $z = t$, and also the number of members aged y at time t is given by $l_y \cdot f(t+a-y)$ for $y = a, a+1, a+2, ...$.

Now, we shall describe the concepts of stable and stationary membership, which deal with special cases of the above time-varying equation (i.e. a case of $\alpha_z = \alpha$, constant for all z).

For a fixed and given α, $f(z) = \exp(\alpha z)$ is given as a solution of $f(z+1) = \exp(\alpha) \cdot f(z)$ with the initial condition $f(0) = 1$. The total number of existing members (or persons living) at time t (denoted as $N(t; \alpha, a \leq y < \infty)$) and the total number of leaving members (or deaths) at time t (denoted as $M(t; \alpha, a \leq y < \infty)$) are then given, respectively, by

$$N(t; \alpha, a \leq y < \infty) = \sum_{y=a}^{\infty} l_y \cdot f(t+a-y) = \exp(\alpha t) \cdot [\sum_{y=a}^{\infty} \exp(\alpha(a-y)) \cdot l_y] \quad \text{and}$$

$$M(t; \alpha, a \leq y < \infty) = \sum_{y=a}^{\infty} l_y \cdot q_y \cdot f(t+a-y) = \exp(\alpha t) \cdot [\sum_{y=a}^{\infty} \exp(\alpha(a-y)) \cdot l_y \cdot q_y]$$

$$= [1 - \exp(\alpha)] \cdot N(t; \alpha, a \leq y < \infty) + \exp(\alpha(t+1)) \cdot l_a,$$

where, $q_y = (l_y - l_{y+1}) / l_y$ (i.e. the effective annual rate of mortality at age y).

From the above formulae for $N(t; \alpha, a \leq y < \infty)$ and $M(t; \alpha, a \leq y < \infty)$, we can derive the these general properties:

(i) The case of $\alpha > 0$: the total number of existing members and leaving members each is increasing exponentially in the fixed ratio α over time t, and also the number of existing members in each age and leaving members at each age are all increasing exponentially in the fixed ratio α over time t;

(ii) The case of $\alpha < 0$: this is the exact opposite to the case of $\alpha > 0$, i.e. we can repeat the comments in (i), replacing 'increasing exponentially' with 'decreasing exponentially';

(iii) The case of $\alpha = 0$: then $N(t; \alpha, a \leq y < \infty) = [\sum_{y=a}^{\infty} \exp(\alpha(a-y)) \cdot l_y]$ and $M(t; \alpha, a \leq y < \infty) = l_a$, that is, both of them are independent of time t (i.e. constant over time t); and

(iv) Any age-related fraction, at time t, of $N(t; \alpha, a \leq y < \infty)$ and $M(t; \alpha, a \leq y < \infty)$ is independent of time t, that is, for any ages y_0 and y_1, $y_0 \leq y_1$, then $N(t; \alpha, y_0 \leq y \leq y_1) / N(t; \alpha, a \leq y < \infty)$ and $M(t; \alpha, y_0 \leq y \leq y_1) / M(t; \alpha, a \leq y < \infty)$ are all independent of time t, which means that the relative age distribution of existing members and leaving members each is constant over time t.

In a demographic sense [see Keyfitz (1985; sections 4.1 & 12.6)], the membership (or population) resulting from the special case of $\alpha_z = \alpha$, constant for all z, is called a stable membership (or population).

As seen in properties (i), (ii) and (iv), 'α = constant $\neq 0$' implies that the scheme has a stable age distribution, that is, a fixed life table with the characteristic that the number of new entrants aged a is increasing (if $\alpha > 0$) or decreasing (if $\alpha < 0$) exponentially at the rate α over time t, and then the number of existing members at each age as well as the number of leaving members in each age are all increasing (if $\alpha > 0$) or decreasing (if $\alpha < 0$) exponentially at the same rate α over time t. We note that that in order to distinguish between the cases $\alpha > 0$ and $\alpha < 0$, we shall refer to the former as the increasing stable membership and the latter as the decreasing stable membership.

Moreover, the membership (or population) resulting from the special case of '$\alpha = 0$' is called a stationary membership (or population) as a special case of a stable membership (or population). In other words, as seen in property (iii), '$\alpha = 0$' implies that the scheme has a stationary age distribution, in which the size of the membership is constant over time t and the number of new entrants is equal to the number of leaving members, which illustrates the aptness of the name stationary membership [see Bowers et al. (1986), section 18.4].

So far, we have mentioned the concepts of stable and stationary membership, subject to a single contingency of death. However, defined benefit schemes normally provide protection against multiple contingencies such as early retirement (including withdrawal from employment and retirement for disability), death during employment and old-age retirement. As an actuarial application of this extension to a pension scheme, we assume that the scheme's trust deeds and rules specify the immediate payment of the total benefits described when the corresponding contingency happens (which is reflected in the following assumption (i)'). In this respect, we generalise the earlier concepts in relation to a given multiple decrement model.

For a discrete time approach based on the multiple decrement model, we assume the following:

(i)' The term 'decrement' means the termination of membership of the scheme due to death, disability, withdrawal or old-age retirement;

(ii)' Assumptions (i), (iii) and (iv), introduced early in this section, are applied;

(iii)' All active members retire at age b, where b > a;

(iv)' Membership in the scheme is determined by the time-independent membership function of the given multiple decrement table derived from best estimates of rates of death, disability, withdrawal and old-age retirement, which indicate the expected number of members at an age y in the scheme (denote as $l_y^{(m)}$, where the superscript (m) is added to distinguish it from l_y used in life table) where y = a, a+1, . . ., b.

Then, $l_y^{(m)}$ is generated in such a way that $l_y^{(m)} = l_{y-1}^{(m)} \cdot [1 - q_{y-1}^{(m)}]$ with the initial condition $l_a^{(m)} = l_a$ and the boundary condition $l_b^{(m)} = 0$ where $q_y^{(m)} = q_y^{(d)} + q_y^{(w)} + q_y^{(e)} + q_y^{(r)}$, in which $q_y^{(d)}$ = the dependent annual rate of mortality, $q_y^{(w)}$ = the dependent annual rate of withdrawal, $q_y^{(e)}$ = the dependent annual rate of disability and $q_y^{(r)}$ = the dependent annual rate of old-age retirement [see Bowers et al. (1986), section 9.6 &10.2].

And also, following the earlier mathematical formulation, we have

$$N^{(m)}(t;\, \alpha,\, a \le y \le b) = \sum_{y=a}^{b} l_y^{(m)} \cdot f(t+a-y) = \exp(\alpha t) \cdot [\sum_{y=a}^{b} \exp(\alpha(a-y)) \cdot l_y^{(m)}] \quad \text{and}$$

$$M^{(m)}(t;\, \alpha,\, a \leq y \leq b) \;=\; [1 - \exp(\alpha)] \cdot N^{(m)}(t;\, \alpha,\, a \leq y \leq b) + \exp(\alpha(t+1)) \cdot l_a.$$

Therefore, we can follow very similar arguments to the case of the single decrement of death; as a conclusion, 'α = constant $\neq 0$' implies that the scheme has a stable age distribution (i.e. a stable membership) and '$\alpha = 0$' implies that the scheme has a stationary age distribution (i.e. a stationary membership).

2.1.4.2 Young, Mature and Declining schemes

In general, the membership progress of a defined benefit pension scheme may be thought of in a similar way to the changes occurring in an industry (i.e. the product life-cycle that generally takes four consecutive stages in the area of economics [see Harvey (1991; section 13.3)] - 'innovation stage', 'growth stage', 'maturity stage', and 'saturation and decline stage').

In theory, a company in the innovation and growth stages may be required to employ new employees increasingly until reaching the maturity stage in view of the expansion of its current business; a company in the maturity stage may not need to increase the flow of new employees but will need to maintain its current employment policy in view of the stabilisation of its current business; but on the other hand, a company in the saturation and decline stage may be required to curtail new employment consistently until closing its current business in view of the adverse progress of economic conditions in its business. For a historical example of the product life-cycle, we may refer to the British shipbuilding or coal mining industries.

In this respect, we attempt to classify a defined benefit pension scheme into Young, Mature and Declining schemes in the light of its long-term running history (here, these three consecutive stages in the development of the scheme shall be called the life-cycle of a defined benefit pension scheme or simply the scheme life-cycle).

Prior to specifying each scheme, we may briefly outline the relation between the scheme life-cycle and the product life-cycle; that is, a Young scheme might correspond to the scheme run by a company in the innovation and growth stages, a Mature scheme to the scheme run by a company in the maturity stage, and a Declining scheme to the scheme run by a company in the saturation and decline stage.

In order to specify the three distinct schemes (composing the scheme life-cycle) particularly in connection with the concept of (increasing or decreasing) stable and stationary memberships, we shall here make the assumptions that the level of new recruitment occurs at a fixed entry age and that it is increasing or decreasing exponentially at a fixed rate, or constant over time.

(i) Young scheme: the membership age structure can be considered to be closely akin to a stable membership as the scheme continues, by the consistent application of a constant exponential rate of increase in new entrants entering at the fixed entry age. Then the scheme will have ultimately an exponentially growing membership. So, the properties (i) and (iv) of the increasing stable membership with $\alpha > 0$, introduced earlier in section 2.1.4.2, can be more closely applied to the scheme, as the scheme continues.

Therefore, each of the total benefit outgoes and actuarial liability will increase with time but the total normal cost will be sufficient enough to pay the total benefit outgoes over time and then the difference will create additional funds for the future liabilities, since both the average age of active members and proportion of beneficiaries are low and decreasing.

(ii) Mature scheme: the scheme can be considered to have developed from a Young scheme through a reasonably long-running operation with the consistent employment of the same number of new entrants at a fixed entry age. The property (iii) of the stationary membership with $\alpha = 0$, introduced earlier in section 2.1.4.2, may be more relevant to the scheme, as it continues.

Therefore, each of the total benefit outgoes and actuarial liability will be constant over time but the total normal cost is normally expected to be less than the total benefit outgoes over time and then the shortage may be largely met either by the earnings from investments or by cash injection made by the employer, since both the average age of active members and the proportion of beneficiaries have been nearly in equilibrium.

(iii) Declining scheme: the scheme can be considered to have been developed from a Mature scheme, mainly throughout the adverse progress of long-term economic conditions in the employer's business or the introduction of modern technology automation, such as office and/or factory automation. So, we may assume that the new entrants at the fixed entry age are declining exponentially over time and then the scheme is more closely akin to the decreasing stable membership with the properties (ii) and (iv) in the case of $\alpha < 0$, introduced in section 2.1.3.2, as the scheme continues. It should also be noted that as time approaches to infinity, the membership in a Declining scheme will tend toward zero. Then, different from the case of winding up a scheme, the Declining scheme will eventually terminate without any legal liability.

Therefore, allowing only for the retirement benefit, both the total benefit outgoes and the actuarial liability will decrease over time but the total normal cost will be less than the total benefit outgoes and then the shortage may be largely met either by the earnings from investments or by cash injection made by the employer, since both the average age of active members and proportion of beneficiaries are high and increasing.

In summarising section 2.1.4, we note that the increasing stable membership is characterised by the property that the membership at each age is growing exponentially in a fixed rate over time, and hence it can be applied to a Young scheme. The stationary membership is characterised by the property that the membership at each age is independent of time and further the total number of new entrants is equal to that of leaving members (which illustrates the appropriateness of the name stationary membership), and hence it can be applied to a Mature scheme. On the other hand, the decreasing stable membership is characterised by the membership at each age is declining exponentially in a fixed rate over time, and hence it can be applied to a Declining scheme.

2.1.5 Funded schemes vs. Pay-as-you-go (PAYG) schemes

The question of who supports the benefit provision for the scheme`s members as well as how it is financed, is of vital importance to the security of payment of their benefit entitlements.

In other words, if the sponsor of a scheme is assured of continuing support, the entitled benefits are likely to be paid directly from the sponsor's resources as they fall due (without funding their payments in advance): a scheme with no arrangements for advance funding made for future liabilities (i.e. with no fund of investment assets for future liabilities which can accumulate investment returns) is usually referred to as a pay-as-you-go (PAYG) or unfunded scheme, in which the financing principle of PAYG schemes is that the amount of contributions at time t is equal to that of benefit outgoes at time t. In contrast, if the sponsor of a scheme faces a risk of cessation of support at some time in the future (e.g. a potential risk of bankruptcy or liquidation), accumulated resources of the scheme for future liabilities play an essential role in securing the payment of the benefit entitlements against any potential risk: a scheme with arrangements for advance funding made for future liabilities (i.e. with a fund of investment assets for future liabilities which can accumulate investment returns) is usually referred to as a funded scheme.

For example, most state pension schemes (providing retirement benefit to retired eligible people when they reach the prescribed state pension age) are generally organised in the form of PAYG schemes. We note that the UK state pension schemes, the basic state pension scheme and the state earnings-related pension scheme, are financed through (compulsory) national insurance contributions (levied on earnings and paid by the working population) which are paid into the National Insurance Fund. In reality, this fund is not a state pension fund, since it is not accumulated to meet the scheme's future liabilities [for practical details of the UK state pension schemes, see Harrison (1995; Ch. 4 & 5)]. We note that many PAYG schemes have a small liquidity fund for efficient operation (i.e. to respond rapidly to minor-short-term fluctuations in the normal benefit outgoes).

In contrast, most occupational pension schemes (i.e. money purchase scheme, defined benefits scheme and hybrid scheme based on both of them) are usually set up as pension trust funds, for the reason that from the viewpoint of the employee, a PAYG occupational pension scheme may cause twofold disaster in the event of failure of the employer's business, the loss of both the job and the pension rights of the employee [see Lee (1986; section 8.4)]; and then, it is necessary to protect them from any financial risk of the employer's business (e.g. bankruptcy or liquidation) in addition to taking advantage of any tax incentives (e.g. investment income and realised capital gains tax relief).

In the following two subsections, we specify the critical issues in financing funded and PAYG schemes and also suggest the solutions available in respect to each issue.

2.1.5.1 Issues in financing funded schemes

Throughout our previous discussions, we have noted that the primary concern of defined benefit and hybrid pension schemes is how to set up optimal arrangements for advance funding in the light of balancing the conflicting interests between the trustees and the sponsor. However, for a money purchase pension scheme, the main interest is in establishing the optimal investment strategy of determining the scheme asset mix. These issues, establishing optimally the funding arrangement and/or investment strategy, are of common interest in funded schemes. As a way for setting up an optimal funding arrangement for such a funded scheme, we can also suggest minimising both the solvency risk and contribution risk over a specified projection period [see, section 2.1.3].

As a subsidiary point, we note that implementing the optimal funding arrangement and investment strategy in the funded scheme would have the effect of reducing (or eliminating) the risk of the sponsor's financial difficulties (and hence supporting continuously the scheme in a healthy financial status). So, it would not be necessary to impose, as a statutory requirement, full funding at any point in its lifetime and/or almost completely risk-free investments on the funded scheme.

2.1.5.2 Issues in financing PAYG schemes

PAYG schemes are, in principle, founded on the framework of the sponsor's commitment to supporting continuously schemes without any plan for advance funding.

In our view, the commitment should be based on two principles, sufficient reciprocity and persistent solvency (- these corresponds conceptually to the 'continued goodwill' and 'continued solvency' of former employers, respectively, expressed in Blake (1992a; p39)): the sufficient reciprocity principle may be defined as the willingness of the sponsor not to take an unduly biased action affecting either current or future beneficiaries, and the persistent solvency principle as the continued promise of the sponsor to guarantee the beneficiaries an acceptable level of benefit whenever benefits need to be paid. In this respect, PAYG schemes are suitable for social security systems (providing normally such social security benefits as retirement, sickness, disability, unemployment and low-income benefits) because the government is often regarded as an ultimate guarantor for the commitment but we note that the government support for modern social security schemes can not be guaranteed because of demographic pressures (for example, the ageing of many western populations, characterised by the low fertility and falling mortality) [see Thane (1989)]. A state pension scheme is commonly operated in most industrialised countries as a basic social security programme and we now consider on such a state pension scheme.

To begin with, a state pension scheme could be termed a national tax-financed pension scheme, since each year's earnings-related tax incomes are, in principle, designated to be equal to the same year's retirement benefit outgoes according to the PAYG financing principle.

For convenience, we introduce and specify the following three statistical ratios which depend on the demographic structure of the population as well as the economic condition of the state (primarily, the level of employment, nature of indexing of benefit and relationship to prevailing level of inflation):

(a) Support ratio - - the ratio of people of working age to that of pension age (which is defined in the UK government White Paper (1994; p27));

As an extension to the support ratio, we can also specify the following:

(b) Member support ratio - - the ratio of working people (i.e. contributors as the future beneficiaries) to eligible retired people (i.e. state pensioners as the current beneficiaries); and

(c) Finance support ratio - - the ratio of the average amount of (earnings-related tax) contribution paid by overall contributors to the average amount of pension paid to overall state pensioners.

It should be noted that each ratio is a time-varying statistic, since the number of active and retired eligible populations each are observable random variables depending explicitly on time. The support ratio is the same as the inverse ratio of the so-called 'old-age dependency ratio' [see Mortensen (1992; Glossary)]. If the persons of working age are all employed and also the persons of pension age are all pensioners, the support ratio is equal to the member support ratio. Moreover, the support ratio could be used as an approximate measure of the member support ratio, particularly in the case of estimating the member support ratio over the next generation (about 40~50 years, which is based on the general working life of an individual), since the development of the adult population over the next generation can be estimated with reasonable accuracy but there will be several practical limitations to forecasting the economy of a state over such a long-term period as the length of the next one or two generations.

The process of financing a state pension scheme at a particular time t over a (short-term or long-term) projection period will rely completely not only on the member support ratio at time t projected by demographic trends but also on the finance support ratio at time t projected by economic trends (particularly, the trends of employment).

According to the PAYG financing principle, the financing equation of the scheme can be written as follows: for each time $t \in$ projection period, say $[0, T]$ where $0 < T < \infty$,

{average amount of contribution at time t}

= {average amount of pension at time t} / {member support ratio at time t}

($\Leftrightarrow$ {finance support ratio at time t} = {number support ratio at time t} $^{-1}$)

; that is, higher pensions imply a heavier tax burden on contributors in inverse proportion to the level of the member support ratio.

Based on the above equation, we will make only three distinct points (i)~(iii) from the standpoint of showing the crucial issues in PAYG financing (of course, these are not exhaustive).

As a preliminary, we need to specify the conflicts of interest between the contributors and the state pensioners caused by the PAYG financing, which may be compared to the conflicts of interest between the trustees and the employer in a defined benefit pension scheme. The main interest of the contributors is commonly to avoid the overriding tax payment for pensioners, whereas the main interest of the pensioners is to secure an expected level of pension, so these interests are mutually contradictory as seen in the above equation. Moreover, the government (or supervisory authorities) will play a vital role in balancing these conflicts of interest since the government has the responsibility for keeping a high degree of sufficient reciprocity and persistent solvency in financing a state pension scheme (note that this government role is comparable to the primary role of the actuary in a defined benefit pension scheme, mentioned in section 2.1.1). Following the fact that PAYG financing is generally based on projections of the future demographic and economic situation [see Government Actuary's National Insurance Fund Long Term Financial Estimates (1990)], the member support ratio is outside the control of the government (although the political employment policy will affect the member support ratio in a long-term view), while the finance support ratio is essentially under the control of the government. Hence, the governmental supervision of the state pension scheme will be directly associated with the adjustment of the level of finance support ratio in order to balance these conflicts of interest.

(i) In the case that in parallel with the rising tend of a state economy, the size of the newly employed workforce is continuously increasing (for example, the case of the stable membership with $\alpha > 0$, introduced in section 2.1.3.1), then the member support ratio is increasing in time t (which implies, from the above equation, that the corresponding finance support ratio is decreasing in time t). Therefore, the government could easily balance the conflicts of interest at an acceptable level of contribution and/or pension. In other words, since the source of earnings-related tax revenue is on the increase, the government has the potential to satisfy both the contributors and the pensioners at the same time, for example, by means of fine-tuning the current level of contribution downwards and the current level of pension upwards - this scenario (i) is not observed today in most western countries;

(ii) In the case that the demographic age distribution and state economy are both almost constant (for example, the case of the stable membership with $\alpha = 0$ (i.e. stationary membership), introduced in section 2.1.3.1), then the member support ratio will be almost constant over time t (which implies, from the above equation, that the corresponding finance support ratio is also almost independent of time t). Therefore, the government could keep the scheme in balance. In other words, It is possible to maintain the current financing process in the light of balancing the conflicts of interest; and

(iii) In the case that the member support ratio shows a clear trend among the demographic and economic changes towards a declining population of contributors and a rising population of pensioners (for example, the case of the stable membership with $\alpha < 0$, introduced in section 2.1.3.1), then this situation would cause a significant burden to the government because different from the previous two cases, the member support ratio is decreasing in time t (which implies, from the above equation, that the corresponding finance support ratio is increasing in time t). Therefore, it is quite difficult to balance the conflicts of interest without any loss to the members' interests. This problem, involving difficulties in government finance with respect to adverse demographic trends and high unemployment, is currently of vital importance to many industrialised countries [see Dilnot et al. (1994; section 3.3)], so we need to discuss it in more detail in the next paragraph.

In reality, balancing the conflicts of interest under circumstances such as case (iii) has been a crucial issue in the operation of a state pension scheme. As mentioned in Heubeck (1992), it has been widely recognised that the general population model of all member states of the European Community (EC) is that of a shrinking and sharply-ageing population as a result of increased life expectancy and/or a reduced birth rate in the past; hence, this development can be considered as a specific example of case (iii). If the government can freely manipulate the levels of contribution and pension under the case (iii), there is no trouble in financing the scheme as seen in the above equation but this kind of financing process is completely unacceptable in view of the sufficient goodwill principle; in this respect, we shall exclude from our consideration such a financing method as simply increasing the contribution level and/or reducing the pension level.

Alternatively, the government may consider the following ways of balancing the conflicts of interest (either singly or in tandem):

(a) Increasing the state pension ages for men and/or women, which leads to increasing the support ratio and member support ratio. This is consistent with the view of Dilnot et al. (1994; section 3.5) expressed that "If the population is ageing because of increased longevity, then individuals will need a longer period in the labour force to obtain a given level of average consumption over their lifetime. This might lead to individuals prolonging their working lives by postponing retirement ... ."

For example, the UK government has announced the raising of the woman state pension age from 60 to 65 (i.e. equal pension age of 65 for both men and women), expected to be phased in over the 10 years starting in the year 2010, where this adjustment effect can be measured in terms of support ratio, that is, by 2030 year the support ratio, predicted under current state scheme, 2.2:1 will be increased to 2.7:1 under the adjustment [statistical source: UK government White Paper Vol. 1 (1994; p27)]. We note that this policy of raising the state pension ages for men and/or women is also being implemented in other EC member states such as Germany, Greece, Italy and Portugal. For example, in Germany: these are currently 63 for men and 60 for women but these will be equalised at 65 over the 10 years starting in the year 2001 and in Italy: these are currently 60 for men and 55 for women but these will be rise to 65 for men and 60 of women over the 10 years starting in the year 1994 [for more details, see Clifford Chance (1993; pp18 & 33)];

(b) Subsidising the state pension scheme from the other resources of government. However, this may cause the other potential problem of increasing the taxes on the working people, so this way would not be appropriate from a long-term viewpoint; and

(c) Transforming the PAYG financing of the state pension scheme into a partially funded state pension scheme (i.e. using a mixed (hybrid) financing principle of PAYG financing and partial funding). However, as introduced in section 2.1.3.3, the partially funded state pension fund will be also subject to solvency risk (particularly caused by investment risk), so its asset allocation strategy has to be, to a large degree, restricted in view of the persistent solvency principle; in this respect, the accumulated assets from partial funding should be invested largely in fixed-interest and index-linked government securities. For example, the PAYG old age, survivors and disability scheme (OASDI) organised by the US has led to a shift towards a partially funded basis for the social security programme after a series of financing crises in the 1980s [see Dilnot et al. (1994; p59)].

As a conclusion, from a long-term viewpoint, either of (a) or (c), or a mixed combination would be considered as appropriate financing methods for state pension schemes in the industrialised countries, which are suffering higher PAYG financing burdens caused by an ageing population.

2.1.6 Conclusions

As for occupational pension schemes, arrangements for advance funding made for future liabilities (i.e. funded schemes) are essential from the viewpoint of security for the member's benefit entitlements because of the possibility of the employer becoming unable or being unwilling to continue his obligations to pay the promised benefits as they fall due. The security problem is even more important in defined benefit schemes than in money purchase schemes, since the individual benefit entitlements in a money purchase scheme will be determined in connection with the value of the individual interest in the fund accumulated.

As discussed in section 2.1.2.8, we can use the funding level or solvency level as an appropriate security measure but these levels should be differently defined according to the valuation basis being used (i.e. going-concern, run-off or winding-up).

Our purpose of funding in defined benefit schemes is to balance the conflicts of interest between the trustees and employer over a projection period (or equivalently, minimise the solvency risk and contribution risk simultaneously over a projection period according to the relative importance for the particular scheme), whether the assets and liabilities are assessed on going-concern, run-off or winding-up valuation bases. Particularly from the viewpoint of the supervisory authorities, security (more exactly, solvency) is to be regarded as a main objective of funding and stability is to be regarded as a subsidiary objective of funding. For this reason, we need to set out the funding targets (i.e. targets of the fund level and contribution rate), which have to be chosen in accordance with the valuation basis applied.

Moreover, since solvency considerations are in principle based on the assumption that the scheme may be terminated at any time in the future (so, based on the short-term perspective), while the funding plan is in principle based on the assumption that the pension provision of a pension is a very long-term commitment (so, based on the long-term perspective), it would be necessary to take into account the imbalance between the short-term solvency considerations and long-term funding plan when determining the funding targets.

Accordingly, we may use the following as appropriate funding targets: at every valuation date t over a projection period, {target of fund level at time t on going-concern basis} = {accrued actuarial liability at time t}, {target of fund level at time t on run-off basis} = {market related cautious liability at time t}, {target of fund level at time t on winding-up basis} = {market related current liability at time t} [see sections 2.1.2.7 & 2.1.2.8] and {target of contribution rate at time t} = {normal cost at time t}.

We note that these recommended funding targets are estimated over a projection period by means of an adopted funding method (e.g. Current Unit or Projected Unit methods) and/or movements of the open pension market because neither the amount of the scheme's liabilities nor the contribution rate required for solvency requirements is known in advance. Furthermore, following the Surpluses Regulation 1987 for tax purposes, the Projected Unit method would be recommended as the fundamental funding method for the estimation of a suitable funding target [see section 2.1.2.6 & paragraph (iii) of section 2.1.2.7]. Of course, there is no unique optimal funding target because this will depend on the particular circumstances of each pension scheme, particularly in relation to the employer's situation, the economic and demographic prospects of the scheme and the statutory regulations imposed by supervisory authorities.

2.2 General pension funding

2.2.1 Introduction

The title of this section 2.2, general funding plan, is here specified as the combination of primary funding methods and supplementary funding methods, mainly based on a going-concern valuation basis (which are described in the following sections 2.2.2 & 2.2.3). In particular, the term 'general' means that these (both primary and supplementary) funding methods are the most commonly used in a number of countries including the UK, the USA and Canada with the aim of determining a recommended contribution rate.

As outlined earlier in section 2.1.2.3, the recommended contribution rate would be expressed in practical terms as follows: at each valuation date t,

{recommended contribution rate at time t (C_t)}

= {normal cost at time t (NC_t)} + {adjustment to the normal cost at time t (ADJ_t)},

in which the primary funding methods provide the normal cost as a regular cost, while the supplementary funding methods determine the adjustment to the normal cost as a supplementary cost (to the regular cost).

We note here that although the Aggregate method does not have a normal cost nor an actuarial liability, it provides a special form of the above formula, which will be shown in section 2.2.4.2.

2.2.2 Net present value vs. Actuarial present value

As a preliminary to the detailed discussion about funding methods, it will be helpful to consider the concept of present value. In the actuarial valuation for the calculation of normal cost and actuarial liability, this concept is specified by reference to the actuarial (economic and demographic) assumptions, the trust deeds and rules and the primary funding methods (which will be clear in the following subsection (ii)). So, it may cause some confusion with the concept of present value generally used by financial economists (so-called, net present value (NPV)); in this respect, we shall employ the concept of present value used by pension professionals involved in the development of pension funding methods, particularly by pension actuaries (so-called, actuarial present value (APV)). These two concepts are separately described in the next two subsections (i) and (ii) with the aim of showing some differences between them, which depend on the choice of assumptions for the discount function and expected cash flows.

Firstly, we will simply illustrate the concept of NPV on a deterministic and discrete time approach, which will provide the basic idea for the APV.

(i) Net present value (NPV):

The net cash flows emerging from an investment project may be determined a priori with certainty or estimated as the most likely expected cash flows (i.e. best estimates of the prospective cash flows), where (net cash flow at time t) = (cash inflow at time t) - (cash outflow at time t) over a specified projection period. We shall here focus on the investment project appraisal of deterministic cash flows. As mentioned in subsection (i) in 2.1.2.6, the actual prospective cash flows could be specified by a two-dimensional stochastic process {(time, amount)} and its appraisal is theoretically available by statistical techniques including simulations (e.g. mean and variance analysis). But it may well be more straightforward for investors to combine a deterministic approach with a sensitivity analysis (i.e. broadly, using expected, optimistic and pessimistic scenarios about the prospective cash flows according to the analyst's perspective of the investment project).

The appraisal of deterministic cash flows involves the conversion to a common point in time (usually, viewed as time 0, the present time) because it is necessary to allow for the time value of money, inflation and other financial factors (e.g. tax). Then, the subject of how to obtain a stable measure with respect to the expected cash flows is of importance in the investment appraisal. The commonly used techniques are the discounted cash flow methods, net present value (NPV) and internal rate of return (IRR) [see, Lumby (1988; Ch. 4)]; we are only concerned with the NPV method.

The NPV method is adapted by discounting the deterministic cash flows backwards through time with a chosen discount function or a sequence of discount functions. The resulting value of the cash flows is usually referred to as the discounted present value or NPV: this is mathematically described below using a discrete time approach. (Also, it is possible to adopt a continuous time approach)

When employing a deterministic approach, the discount function for an NPV investment project appraisal is usually assumed to be constant over the projection period and based on the analyst's perception of the future sequence of actual rates of return occurring on the investment project, except when there are special circumstances so that the NPV is calculated on the basis of a historical/experienced sequence of discount functions. In general, the discount function suitable for NPV calculations would be presented as a function of a best estimate discount rate such as the anticipated mean rate of the prospective rates of return, or as a function of a risk-adjusted discount rate (i.e. risk-free discount rate + risk-premium, see Mehta (1992)). The size of the risk-premium will reflect the analyst's perception of the "riskiness" of the project, for example, represented by "risk-averse" (i.e. greater risk premium than the project's perceived risk), "risk neutral" (i.e. risk-premium equivalent to the project's perceived risk) and "risk-seeking" (i.e. lower risk-premium than the project's perceived risk).

Thus, the net present value (NPV) for a given project can be expressed as the sum of its discounted prospective cash flows:

$$\text{NPV} = \sum_{t=0}^{n} \{ \text{(project's monetary cash flow at time t)} \cdot \text{(project's discount function)}^{t} \}$$

where, "0" denotes the present time when the project starts; "n" denotes the future time when the last cash flow earned on the project occurs; and (project's discount function) is given in the form of $(1+\text{best estimate discount rate})^{-1}$ or $(1+\text{risk adjusted discount rate})^{-1}$, as discussed above.

(ii) Actuarial present value (APV):

In the actuarial application to pension funding and valuation, the concept of APV is founded on two distinct frameworks.

Firstly, we have the valuation interest rate (specified in the actuarial economic assumptions), which identifies the appropriate discount function for the scheme's prospective cash flows (here, this discount function shall be called the actuarial discount function as a counterpart of the project's discount function). Secondly, we have the decrement (service) table (specified in the actuarial demographic assumptions), which identifies the probability distribution of the timing of the scheme's prospective cash flows associated with either benefit outgoes (on the contingencies specified in the trust deeds and rules) or normal cost income (specified by the primary funding methods).

Therefore, the concept of APV is distinguishable from that of NPV because the decrement table makes it possible to interpret the concept of APV in a statistical sense (i.e. it is, in principle, based on a stochastic approach), while the concept of NPV is, in principle, based on a deterministic approach. In this respect, we can understand the definition of APV made by Bowers and et al. (1986; section 4.2.1) expressed as "the expectation of the present value of a set of payments contingent on survival (a contingent annuity)" (note that in Anderson (1990; section 6.4), APV is also called the expected present value). However, as noted by Bowers et al. (1986; section 10.2), the decrement table can also be used to represent the survivorship of the members existing in the scheme subject to given probabilities of the specified contingencies, and hence the APV can also interpreted in a deterministic manner.

We note that, although we may make a time-varying assumption as to the actuarial economic parameters by using an appropriate stochastic model such as that advocated by Wilkie (1995), the actuarial economic parameters are usually determined in a deterministic approach, using a time-invariant best estimate such as a mean rate which is constant over time, by reference to both the actuary's analysis of experienced values of any past period and the actuary's perspective as to the future political and economic developments. Further, following the fact that the valuation interest rate is the assumed rate of investment return on the scheme's fund, both the valuation interest rate and the actual rates of investment return are measured in the same terms, nominal or real (here, 'real' is considered in the sense that it represents the rate of investment return in excess of the effects of general price inflation).

Therefore, the prospective cash flows (e.g. the accrual of future benefits and normal costs) must be considered to be in nominal terms or in real-terms in accordance with the terms expressed for the valuation interest rate. Consequently, the discount function corresponding to the project's discount function for NPV will be identified by the valuation interest rate.

For the pension funding methods, we introduce the following notation:

PVS_t = APV of future salaries of active members existing at time t throughout their expected future working lifetimes, computed at time t;

PVN_t = APV of future normal costs of active members existing at time t throughout their expected future working lifetimes, computed at time t; and

PVB_t = APV of future benefits for all members, including pensioners, existing at time t, computed at time t.

For the convenience of our arguments in the next sections 2.2.2 & 2.2.3, we will divide PVB_t into three distinct components: that is,

$$PVB_t = PVB^p_t + PVB^f_t + PVB^r_t$$

where, PVB^p_t = APV of past service benefits for the active members existing at time t, computed at time t; PVB^f_t = APV of future service benefits for the active members existing at time t, computed at time t; and PVB^r_t = APV of future benefits for all members excluding active members (i.e. retired and deferred members) existing at time t, computed at time t.

The above notation can be thought of as representing the average or expected outcomes of the actual prospective cash flows, calculated by ignoring the risk of variability of outcomes in future. As an extension to the concept of APV, it would be worth considering the present value of a stochastic cash flows reflecting the future uncertainty, but this will be the subject of future work [see Dufresne (1992) for a discussion of probability distributions of discounted stochastic cash flows, characterised by independent and identically distributed cash flows and discount functions; Buhlmann (1992) for some preliminary comments on the present value of stochastic cash flows and the effect of the stochastic time series of discount functions, characterised by the Beta-Binomial distribution, on the present value of stochastic cash flows; and Norberg (1995) for some practical applications of the conditional moments of present values of the stochastic cash flows characterised by a continuous-time Markov process, using stochastic discount indicator functions).

In the next section, the concept of present value will be used separately as APV and NPV.

2.2.3 Primary funding methods

2.2.3.1 Preliminary

The Institute and Faculty of Actuaries (1984, 1988) has listed and provided standard descriptions of the five main funding methods, commonly used by actuaries in the UK, with the aim of facilitating communication and understanding both within and outside the actuarial profession. These are the Current Unit, Projected Unit, Entry Age, Attained Age, and Aggregate methods. In the following presentation, we focus on the important features of these funding methods. In our nomenclature, each of these methods shall be called a primary funding method. However, no mathematical definition of the primary funding methods are given [for detailed mathematical definitions, see Dufresne (1986; Ch. 1) using discrete and continuous time approaches, Anderson (1992; Ch. 2) using a discrete time approach, and Fujiki (1994; Ch. 3) using a continuous time approach, which are all differently formulated according to their prescribed assumptions].

As expressed in the 1993 report of PLRC [see Vol. 1, section 4.3.10] that "Since the funding method is designed to fulfil the funding objective agreed for the particular scheme, there is no single standard actuarial funding method ... . Much depends on the particular circumstances of the employer and the employer's future strategy and desired pace of funding.", there is no unique best funding method covering all circumstances. Thus, there are a variety of funding methods being applied (differently from scheme to scheme according to its own circumstances) which may be considered a variation or modification of the primary funding methods.

The primary funding methods can be categorised in various ways, subject to the properties of each funding method; for example, the widely recognised classification in the UK is the split between accrued benefit methods = {Current Unit method, Projected Unit method} and projected benefit methods = {Entry Age method, Attained Age method, Aggregate method} and next, the classification made by Dufresne (1986) and Haberman (1992) is the split between individual funding methods = {Current Unit method, Projected Unit method, Entry Age method} and aggregate funding methods = {Attained Age method, Aggregate method}.

The former classification is based on the actuary's viewpoint of the scheme's security; in other words, methods in the first category commonly address the security of the members' accrued rights, whereas methods in the second category commonly address the security of the members' prospective rights relying on a particular pattern of future contribution rates. On the contrary, the latter classification is subject to whether the actuarial calculations are performed for the total members or for individual members. In other words, methods in the first category each produce the normal cost and actuarial liability calculated separately for each member and summed up to yield totals for all members at the valuation date. However, the Attained Age method in the second category produces the normal cost and actuarial liability calculated for all members at the valuation date and the Aggregate method in the second category produces a recommended contribution rate calculated for all members at the valuation date. For convenience, the descriptions for each primary funding method here follow the first classification (i.e. accrued and projected).

For consistency with the earlier discussions in section 2.1.2.6 and the general descriptions, the following assumptions are made:

(A1) Valuations are carried out annually; hence, all descriptions are based on a discrete time approach;

(A2) The trust deeds and rules specify only final salary retirement benefits for age and service; hence, the members existing in the scheme are partitioned into active and retired members;

(A3) A suitable multiple decrement model (the service table) is constructed to represent correctly the survivorship of the members existing in the scheme;

(A4) The salary growth rate (including inflation on salaries and promotional salary scale) is fixed and constant for each unit periods (denoted by i_s); hence, the accrual rate of pension benefits is the same as the salary growth rate from (A2) and is deterministic;

(A5) The valuation interest rate is fixed and constant for each unit time period (denoted by i_v); hence, the discount function for NPV or APV is $(1 + i_v)^{-1}$; and

(Notation) The superscript on the left side of each main symbol is used for describing the primary funding method; for example, ${}^{CU}NC_t$ indicates the normal cost at time t calculated by the Current Unit method

2.2.3.2 Accrued benefit methods

As a preliminary to the details of each accrued benefit methods, we make the following comments. The theoretical funding principle of the accrued benefit methods is that the normal cost is calculated to be the level of contribution required to maintain a 100 per cent target funding level at each valuation date: that is, if all actuarial assumptions are exactly realised up to a fixed valuation date then the actuarial liability is the fund level which is the accumulated value of the past normal costs paid when due. Thus, the main aim is to secure the members' accrued rights with stability taking second place. If there have been (actuarial) surpluses or deficiencies (or, (actuarial) gains or losses), adjustment to the normal cost is likely to focus on restoring the target over an amortisation period determined mainly by the scheme's actuary.

(i) Current unit method:

The funding principle is described as follows: (a) the actuarial liability at the valuation date is calculated as if the scheme would be terminated immediately, based on current salaries for each active members (projected salary growth being disregarded) and (b) the each year's normal cost is estimated to meet the sum of that year's accrual of all benefits for each active members, so that in theory this method is designed to build up a 100 per cent funding level based on current salaries.

Thus, the actuarial liability and normal cost at each valuation date t are defined as follows:

{actuarial liability at time t ($^{CU}AL_t$)}

= {sum of the actuarial present values (APVs) at time t of the past service pension benefits for each active members existing at time t, based on his current salary} + {sum of APVs at time t of future pension payments to each retired members existing at time t}

= $^{CU}PVB^{p}_{t}$ + $^{CU}PVB^{r}_{t}$; and

{normal cost at time t ($^{CU}NC_t$)}

= {sum of the accrual of benefits for each active members existing at time t during the intervaluation period (t, t+1), based on his salary projected only as far as the end of (t, t+1)}

= {sum of APVs at time t of benefits accruing for each active membership existing at time t during (t, t+1), based on his projected salary at the end of (t, t+1)} + {sum of the increase in the actuarial present value (APV) at time t of the benefits already accrued for each active members existing at time t, arising from salary growth during (t, t+1)}

= $[^{CU}PVB^{p}_{t+1} / (1 + i_v) - {}^{CU}PVB^{p}_{t}] + i_s \cdot {}^{CU}PVB^{p}_{t}$.

For the stability of normal costs, this method requires a continuing and constant flow of new members entering at a fixed age (i.e. stationary membership from the start) and requires the average past service benefits at each active age to be approximately constant.

In summary, the name Current Unit method can be considered to contain most of the underlying funding characteristics: in our explanation, the term 'Current' means the member's current salary, the term 'Unit' means one unit which can be defined as the member's accrual of benefits over each year of active membership years based on the member's projected salary at the end of that year.

(ii) Projected Unit method:

The principle of this method is the same as for the Current Unit method, except that the actuarial liability and normal cost at each valuation date t are commonly calculated by reference to the member's salary projected to retirement age (i.e. the member's projected final salary) rather than the member's current salary or projected salary over (t, t+1) (i.e. except that future salary growth is fully taken into account in the actuarial calculations). Of course, if there is no inflation on salaries, and no promotional salary scale, then the Projected Unit method is equivalent to the Current Unit method.

Therefore, the actuarial liability and normal cost at each valuation date t are defined as follows:

{actuarial liability at time t (${}^{PU}AL_t$)}

= {sum of APVs at time t of accrued benefits for each active member existing at time t, based on his projected final salary} + {sum of APVs at time t of future pension payments for each retired member existing at time t}

= ${}^{PU}PVB^{p}_{t}$ + ${}^{PU}PVB^{r}_{t}$; and {normal cost at time t ($^{PU}NC_t$)}

= {sum of APVs at time t of benefits accruing for each active member existing at time t during the intervaluation period (t, t+1), based on his projected final salary}

$= {}^{PU}PVB^{P}_{t+1} / (1 + i_v) - {}^{PU}PVB^{P}_{t}$.

That is, each year's normal cost is then estimated to meet the sum of that year's accrual of benefits for each active member, so in theory this method is designed to build up a 100 per cent funding level based on projected final salaries. Indeed, Thornton & Wilson (1992) conclude that "... no strong reasons to use any other method than the projected unit method for funding large schemes expected to have a continuing flow of new entrants. ... the majority of actuaries are now using the projected unit method". Also, they point out three attractive features of this method: in short, the distinction between the level of funding for accrued benefits and the ongoing level of contribution required for accruing benefits, the consistency with the underlying premise of a continuing scheme and the use of the estimated future cost of the benefit promises made. Thus, this method addresses, to a large degree, the general problem of mismatch between the short-term security position and the long-term funding position, which would be a main reason for this method being widely acceptable.

Consequently, we can say that the Projected Unit method is a very suitable funding method for establishing the funding targets appropriate for both security and stability.

As a summary, we note that the name Projected Unit method can be analysed in a similar way to the analysis of the name Current Unit method: that is, the term 'Projected' means the member's projected final salary, the term 'Unit' means one unit which can be defined as the member's accrual of benefits over each year of active membership years based on the member's projected final salary.

2.2.3.3 Projected benefit methods

Different from the accrued benefit methods, the theoretical funding principle of projected benefit methods takes full account of both the future salary growth and the future accrual of benefits for active members (i.e. the total (past and future) service of active members). In theory, they are concerned more with achieving a stable pace of funding in the future (ideally, producing a level normal cost) than with securing the members' accrued rights; that is, their underlying premise is that the pension scheme of interest is an ongoing entity and its active members continue service up to the assumed retirement age subject to a given multiple decrement table. If there have been (actuarial) surpluses or deficiencies, adjustment to the normal cost would be carried out particularly from the viewpoint of the sponsoring employer, that is, it is likely to be determined with the aim of ensuring stability.

(i) Entry Age method:

Historically, this method was first described (but not named) by Porteous (1936). The funding principle is that (a) a (single) normal entry age is assumed to represent the average age of new entrants (here, we shall call a member assumed to enter the scheme at a normal entry age the notional member) and (b) the normal cost for any active member is defined as a uniform level percentage of salary for the notional member from the date of entry to retirement, necessary to finance the future service pension benefits for the notional member over his expected future working lifetime. Thus, each year's normal cost for any active member is calculated as if at each valuation date he was regarded as the notional member, irrespective of his actual age, and thus the resulting normal cost is the same rate for all the active members.

Therefore, it may be appreciated that this method is designed to estimate, on a long-term and going-concern position, the normal cost appropriate for future new entrants with the normal entry age, rather than for the already existing active members. We note that the normal cost for an existing member in mid-career is generally higher than that for a new entrant. To overcome this disadvantage, it is possible to incorporate a range of entry ages; for example, regarding each member's actual entry age into the scheme as his normal entry age, the normal cost for each active member are effectively fixed at his actual entry and remains level throughout his active membership (- this is usually referred to as the Individual Entry Age method).

The actuarial liability and normal cost at each valuation date t are defined as follows:

{normal cost at time t ($^{EA}NC_t$)}

= {normal cost for the notional member} · {number of the active members existing at time t},

in which the normal cost for the notional member is calculated as

{normal cost for the notional member}

= [{APV at time t of future service pension benefits for the notional member, based on his projected final salary} / {APV at time t of future salaries of the notional member throughout his expected future working lifetime}] · {salary of the notional member at time t}, which is independent of the actual age of any active member at time t; and

{actuarial liability at time t ($^{EA}AL_t$)}

= {sum of APVs at time t of future pension payments to each active member existing at time t, based on his projected final salary} - {sum of APVs at time t of future normal costs for each active member existing at time t} + {sum of APVs at time t of future pension payments to each retired member existing at time t}

= $^{EA}PVB_t$ - $^{EA}PVN_t$.

Considering the stability of normal costs, this method requires a continuing flow of new entrants with an entry age equal, on average, to the assumed normal entry age. As for the security of payment of the members' benefit entitlements, the actuarial liability for active members is greater than the accrued liabilities in respect of past pensionable service, since as time goes by, the active members will have a higher average age than the assumed normal entry age (so, the normal cost required for them should be greater than the normal cost on this method); hence, if the funding level on a going-concern basis is maintained at least 100 per cent, the scheme holds assets significantly in excess of its accrued liabilities at all times [see TPFM (1984; p15)].

Consequently, this method would cope with the stability and security problems under the going-concern basis, but it would have a high potential of generating a systematic surplus, measured against its accrued liabilities, which is likely to give rise to a tax penalty imposed by surplus supervising regulations such as the Pension Scheme Surpluses (Valuation) Regulations 1987 introduced earlier in section 2.1.2.7.

As a summary, we note that the name Entry Age method can also be analysed in relation to its principal characteristics: that is, the term 'Entry Age' means that the active member's normal cost each year is the new entrant rate at the chosen normal entry age, which is constant over his expected future active membership years.

(ii) Attained Age method:

The funding principle is that (a) this method makes no allowance for future new entrants to the scheme and (b) the normal cost covers the cost of all future service pension benefits of the existing active members by reference to the expected future working lifetime of the members' pensionable service, where the term 'future service pension benefits' is identified by the current attained age of each member (i.e. total service pension benefits for all active members is exactly divided into past service pension benefits and future service pension benefits by means of the current attained ages of the existing active members).

However, this method places no restrictions on funding the past service pension benefits; for this reason, there may not be a unique determination of a normal cost or actuarial liability, which will be variable according to how the past service pension benefits are funded. In this respect, this method is a member of the aggregate funding methods family. However, assuming that the normal cost does not cover the past service pension benefits (i.e. the past service pension benefits are left to be funded by the supplementary funding methods), the normal cost and actuarial liability at each valuation date t can be defined as for the individual funding methods: that is,

{normal cost at time t (${}^{AA}NC_t$)}

= [{sum of APVs of future service pension benefits for each active member existing at time t, based on projected final salaries} / {sum of APVs of future salaries of each active member existing at time t throughout his expected future working lifetimes (PVS_t}] · {payroll of the active members existing at time t (S_t)}

= ${}^{AA}PVB^f_t \cdot (S_t / PVS_t)$,

which means that NC_t is determined as a uniform level fraction of S_t since the future accrual of service pension benefits is evenly spread over the working lifetime of the active members by means of the PVS_t term (i.e. the future accrual rate of benefits is in parallel with the salary growth rate from assumption A4).

In other words, $NC_t \cdot (PVS_t / S_t)$ implies that a level annuity of NC_t per year is accumulated over the expected future working lifetime to provide the future service pension benefits; this discussion explains the second equality in the following actuarial liability formula, that is, {actuarial liability at time t ($^{AA}AL_t$)}

= {sum of APVs of future pension payments to each active member existing at time t, based on his projected final salaries)} - {sum of APVs of future normal costs of each active member existing at time t} + {sum of APVs of future pension payments to each retired member existing at time t}

(= $^{AA}PVB_t$ - $^{AA}PVN_t$)

= {sum of APVs of all past service benefits for the active members existing at time t, based on projected final salaries} + {sum of APVs of future pension payments to each retired member existing at time t}

= $^{AA}PVB^p_t$ + $^{AA}PVB^r_t$,

which is equivalent to the (accrued) actuarial liability under the Projected Unit method because $^{AA}PVB^p_t = {}^{PU}PVB^p_t$ and $^{AA}PVB^r_t = {}^{PU}PVB^r_t$ and then $^{AA}AL_t = {}^{PU}AL_t$ for any t.

Therefore, comparing the Attained Age method with the Projected Unit method (described in subsection 2.2.2.2) under the general assumption that the scheme remains open to new entrants, younger on average than the currently active members, the normal cost under this method is normally higher than that under the Projected Unit method. This is because the former can be considered as presenting the average future cost, so it is likely to be estimated to be relatively higher for new young entrants than the latter, but as seen in the above formula, the actuarial liability under the two methods are the same. For this reason, the Attained Age method has the potential for generating a systematic surplus, measured against its accrued liabilities, if the initial past service pension benefits have been paid off.

In theory, this method would be suitable for funding a closed pension scheme because as time goes on, S_t is decreasing and approaching zero and so the normal cost tends to zero (i.e. asymptotically stable with the equilibrium level equal to zero).

In summary, we note that the name Attained Age method can also be analysed in relation to the principal characteristics: that is, the 'Attained Age' plays a vital role in funding because the attained ages of the active members existing at a specific time leads to dividing the total service pension benefits into two parts, past service pension benefits and future service pension benefits, and the funding method is designed to fund the future service pension benefits.

(iii) Aggregate method:

The funding principle is that (a) different from the above four funding methods, this method does not define a normal cost nor an actuarial liability and (b) the recommended contribution rate is determined commonly by the formula: at each valuation date t,

$$C_t = [({}^{AG}PVB_t - {}^{AG}F_t) / PVS_t] \cdot S_t$$

where, ${}^{AG}F_t$ represents the fund level at time t accumulated under the Aggregate method (i.e. the value placed on the assets held at time t by the scheme's actuary).

Remark 2.3: (a) In the case of no salary growth, then (PVS_t/S_t) denotes, in particular, the present value of expected future working lifetime of the active membership existing at time t;

(b) The above formula implies that any imbalance between PVB_t and F_t is automatically met evenly over the expected future working lifetime of the active membership existing at time t, effectively by a level temporary annuity value allowing for projected future growth in salaries thereafter (i.e. by PVS_t / S_t); and

(c) If ${}^{AG}PVB_t$ is based on the projected final salaries of active members, then ${}^{AG}PVB_t$ is equivalent to ${}^{EA}PVB_t$ or ${}^{AA}PVB_t$.

This method is not really a separate method, but a variation of the earlier defined methods, specifically the Entry Age method (particularly in view of the scheme being open to new entrants) or the Attained Age method (particularly in view of the scheme being closed to new entrants), subject to surpluses/deficiencies being spread over the remaining working lifetime of the active membership (which is shown mathematically in section 2.2.4.2).

It is worth finally noting that if the same actuarial assumptions are applied to each of the primary funding methods (except the Aggregate method) and the salary growth rate $i_s > 0$, then the Current Unit method will produce the lowest actuarial liability among them because the others take full account of at least either the future salary growth or the future accrual of benefits. Furthermore, under the assumption that $Pro[i_v, i_s > -1] = 1$ and of a stationary membership, Dufresne (1986, section 1.5) compares the transient behaviour of normal costs and actuarial liabilities and illustrates numerically that the Projected Unit and Entry Age methods produce a relatively quick convergence to its limiting value, in which ${}^{EA}AL_t$ is relatively higher than ${}^{PU}AL_t$ for all t, but ${}^{EA}NC_t$ is relatively higher than ${}^{PU}NC_t$ for a short period of time, while ${}^{EA}NC_t$ is relatively less than ${}^{PU}NC_t$ for a long period of time. These results are very similar to those of O'Brien (1984) based on computer simulations allowing for a logistic membership growth function and a range of entry ages.

2.2.4 Supplementary funding methods (to Primary funding methods)

In reality, we have no mathematical way of accurately predicting the future. Therefore, the normal cost will not be sufficient to finance the actuarial liability because of the likely differences between the actual experience and the actuarial assumptions. As mentioned earlier in section 2.1.2.3, a supplementary funding method is designed for spreading variations from the normal cost with the aim of limiting fluctuations in funding caused by these likely differences. Although there may be, in theory, a variety of methods for controlling any undesirable variations, actuaries employ commonly either the Spread method or the

Amortisation of losses method, as a supplementary method to all of the primary funding methods except the Aggregate method.

2.2.4.1 Spread method

This supplementary method is most commonly used in the UK The funding principle is described as follows: at each valuation date t, (a) it is expected that each valuation will show an overall difference between the actuarial liability at time t (i.e. AL_t) and the fund level at time t (i.e. F_t); (b) it causes the necessity of the adjustment (i.e. ADJ_t) to the normal cost at time t (i.e. NC_t) provided by a primary funding method; hence, (c) the difference, $UL_t \equiv AL_t - F_t$ (which is called (actuarial) unfunded liability, see subsection (iii) in section 2.1.2.8), is required, in practice, to be evenly met over an agreed period; in which (d) the term 'evenly met over an agreed period' would characterise the spreading mechanism of the Spread method and its mechanism may be specified mathematically in various ways.

Two forms are commonly used:

- Dividing UL_t by the net present value (NPV) of an annuity certain of 1 per unit period, payable at the beginning of each period for n-unit periods (positive integer $n \geq 1$) with the discount function $(1+ i_v)^{-1}$, $\ddot{a}_n(i_v)$, (i.e. $UL_t / \ddot{a}_n(i_v)$), where 'n' is usually called the amortisation period which refers to the period over which the surpluses/deficiencies at the valuation date are run-off through the actuarial valuation process (typically, n = 20~25 years, corresponding approximately to the average remaining working lifetime of the active members); and

- Dividing UL_t by a level temporary annuity value defined by the formula with the discount function $(1+i_s)/(1+i_v)$, so that $a_t(i_v - i_s) \equiv PVS_t / S_t$, (i.e. $UL_t / a_t(i_v - i_s)$). Of course, we can find some integer n satisfying $a_t(i_v - i_s) \cong \ddot{a}_n(i_v)$.

In view of spreading UL_t evenly over a projected period (i.e. in order to assure that $\ddot{a}_n(i_v)$, $a_t(i_v-i_s) \geq 1$), it is necessary to assume that $Prob(i_v, i_s > -1) = 1$ and thus the recommended contribution rate at time t under the Spread method would be generally formulated in the following form (2.1): for all t,

$$C_t = NC_t + ADJ_t = NC_t + k_t \cdot UL_t, \ 0 \leq k_t \leq 1, \qquad \text{--- (2.1)}$$

in which the value of k_t can be determined by time-invariant formula $1/\ddot{a}_n(i_v)$, time-varying formula S_t / PVS_t or other formulae (that holds the property of spreading evenly UL_t over a decided amortisation period). So, the boundary values in formula (2.1), i.e. $k_t = 0$ and 1, may well be excluded but these values each have a specific meaning such that 'k_t=0' implies taking no action for amortising UL_t, while 'k_t=1' implies immediate and complete amortisation of UL_t without spreading into the future: for this reason, these values are included in the uniform boundedness condition of k_t, as extreme values.

Thus, k_t defines the process for amortising UL_t: in this respect, k_t shall be called the spread parameter at time t and formula (2.1) the spread funding formula. Then, the above uniform boundedness condition $0 \leq k_t \leq 1$ can be considered to be the parameter space of k_t, i.e. $\{k_t: 0 \leq k_t \leq 1\}$.

From the viewpoint of a classical actuarial valuation, it is usually assumed that $i_v > 0$ and $k_t = k = 1/\ddot{a}_n(i_v)$ constant for all t and then the spread funding formula (2.1) is transformed into its specific and restrictive form such that letting $d_v = i_v/(1+i_v)$, we have

$$C_t = NC_t + k \cdot UL_t, \ d_v \leq k \leq 1 \ \text{with} \ i_v > 0 \qquad \text{--- (2.1)'}$$

; the above term 'specific and restrictive' can be interpreted on the grounds that in comparison with formula (2.1) for each time t, k_t is fixed and constant, $\{i_v > 0\} \subset \{i_v > -1\}$ and $\{k: d_v \leq k \leq$

1 with $i_v > 0\} \subset \{k_t: 0 \leq k_t \leq 1\}$ where the parameter space of k is deduced from the fact that for $i_v > 0$, $\ddot{a}_n(i_v)$ is strictly increasing function of n, $\ddot{a}_1(i_v) = 1$ and $\ddot{a}_\infty(i_v) = 1/d_v$.

It should be finally noted that in our later study (in Chapters 3 and 5), even though this is not mathematically essential, we will not necessarily deal with k_t in (2.1) nor k in (2.1)' in connection with their respective amortisation periods, but we will deal with the parameter space, i.e. $\{k_t: 0 \leq k_t \leq 1\}$ (or $\{k: d_v \leq k \leq 1$ with $i_v > 0\}$), in such a way that $\{k_t$: k_t a real number satisfying $0 \leq k_t \leq 1\}$ (or $\{k$: k a real number satisfying $d_v \leq k \leq 1$ with $i_v > 0\}$).

Remark 2.4: (a) In practice, k_t would be required to be determined differently according to whether the scheme is in surplus or in deficit; for example, on a going-concern valuation basis the scheme in deficit would, in general, take the amortisation period at a fixed level (typically 20~25 years), but on the other hand the scheme in surplus should make arrangements to eliminate the excess above the upper limit of tax free funding within five years to follow the Pension Scheme Surpluses (Valuation) Regulations 1987. In this case, the scheme's actuary can make several different suggestions, as mentioned in subsection (iii) of section 2.1.2.8, in which spreading the surplus evenly over five years would be one. On the other hand, the minimum funding requirement (MFR), enacted by the Pensions Act 1995, now puts constraints on funding deficiencies - this will be explained in section 3.2.3;

(b) Following Haberman (1994)'s economic interpretation, the spread parameter k_t, the fraction of UL_t that makes up ADJ_t, can be thought of as a penal rate of interest that is being charged on the unfunded liability UL_t, except for $k_t = 0$;

(c) In the light of optimal control theory, the spread parameter k_t (or corresponding amortisation period) can be thought of as a controlling parameter under control of the scheme's actuary and which allows continued readjustment of the transient behaviour of C_t and F_t, and accordingly controlling the recommended contribution rate can be completed by way of controlling the spread parameter. Thus, it would be desirable to be able to control k_t optimally subject to a formulated control optimisation problem (which is the main subject in Chapter 5); and

(d) As one pioneering approach for optimising the value of the spread parameter in formula (2.1)', we would refer to Dufresne (1986 & 1988) and Haberman (1992, 1993 & 1994), in which using a discrete-time and stochastic approach, they specify and investigate the optimal region/optimal spread period such that $1 \leq n \leq n^*$, where n^* is defined as the minimum value that brings to an end the trade-off relationship between the limiting variance of F_t and the limiting variance of C_t. However, these can be considered as a non-sequential (or single-point-time) optimisation (especially, focused on $t \to \infty$), rather than a sequential (or multi-time-period) optimisation provided by optimal control theory, since its optimal value n^* is admissible by using a numerical analysis (i.e. trial-and-see method) at some point in time: in this respect, their approach is distinguishable from our approach briefly described in (c).

2.2.4.2 Aggregate method vs. General funding plan

We show here that the Aggregate method is a variation of either the Entry Age method plus the Spread method or the Attained Age method plus the Spread method, which will address our earlier comment in section 2.2.1 in which the term 'a special form of the above formula' corresponds to the spread funding formula (2.1).

Under the assumption that $Prob(i_v, i_s > -1) = 1$, the spread funding formula (2.1) can be rewritten for a specific choice of $k_t = S_t / PVS_t \in [0, 1]$ as follows:

Since under the Entry Age method,

$$^{EA}AL_t = {}^{EA}PVB_t - {}^{EA}PVN_t, \quad {}^{EA}PVN_t = {}^{EA}NC_t \cdot (PVS_t/S_t) \quad \text{and} \quad {}^{EA}UL_t = {}^{EA}AL_t - {}^{EA}F_t$$

(because the normal cost for the notional member is independent of the actual age of any active member at time t (see paragraph (i) in section 2.2.3.3), then

$$C_t = {}^{EA}NC + k_t \cdot ({}^{EA}AL_t - {}^{EA}F_t) = \{({}^{EA}PVB_t - {}^{EA}F_t) / PVS_t\} \cdot S_t,$$

which implies that the Aggregate method is in essence a variation of the Entry Age method under which the unfunded liability is run-off over the future expected working lifetime of the active members.

Next, since under the Attained Age method,

$${}^{AA}AL_t = {}^{AA}PVB^p_t + {}^{AA}PVB^f_t, \quad {}^{AA}NC_t = {}^{AA}PVB^f_t \cdot (S_t/PVS_t) \quad \text{and} \quad {}^{AA}UL_t = {}^{AA}AL_t - {}^{AA}F_t,$$

and thus

$$C_t = {}^{AA}NC + k_t \cdot ({}^{AA}AL_t - {}^{AA}F_t) = \{ ({}^{AA}PVB_t - {}^{AA}F_t) / PVS_t \} \cdot S_t,$$

which implies that the Aggregate method is in essence a variation of the Attained Age method under which the unfunded liability is run-off over the future expected working lifetime of the active members.

In a conclusion, we can say that from the above two cases, the Aggregate method provides a special form of the spread funding formula (2.1): in other words, the Aggregate method can be considered as a special funding plan of general pension funding plan (although the Aggregate method does not have a normal cost nor an actuarial liability).

2.2.4.3 Amortisation of losses method

This supplementary method is most commonly used in Canada and U.S.A, but it is not our principal interest, so we shall only discuss briefly the funding principles below [for more details, see Dufresne (1986 & 1989)].

The funding principle is described as follows: (a) this method is based on the (actuarial) gain and loss analysis at each valuation date; (b) the (actuarial) loss experienced during the intervaluation period (t-1, t), L_t, is defined as $L_t = UL_t$ - {value of UL_t if all actuarial assumptions had been realised during (t-1, t)}; (c) at each valuation date t, the informed intervaluation loss, L_t, is spread evenly over a fixed future term, say m (i.e. $L_t / \ddot{a}_m(i_v)$), in which it is necessary to assume that $Prob(i_v > -1) = 1$ in order to assure $\ddot{a}_n(i_v) \geq 1$, as in the Spread method; and thus, (d) the recommended contribution rate is the normal cost plus the sum of liquidated payments which is still in force, that is, at each valuation date t (assuming $t \geq 1$),

$$C_t = NC_t + ADJ_t = NC_t + \sum_{j=0}^{m-1} L_{t-j} / \ddot{a}_m(i_v)$$

; here, m would be chosen by the scheme's actuary in consultation with the trustees and the employer within a period permitted by the related current law or regulation, typically fixed in the range 5~15 years.

Finally, it is worth drawing attention to Owadally & Haberman (1995). They show that using a discrete-time and stochastic approach, the Spread method is superior to the Amortisation of losses method from the viewpoint of minimising limiting variances of the fund and contribution rate levels.

2.2.5 Summary and Conclusion

Primary funding methods each have their own characteristics in generating normal costs, but according to these we can classify the primary funding methods into accrued benefit methods and projected benefit methods, or individual funding methods and aggregate funding methods.

Since the actuarial assumptions are usually based on a deterministic approach, using mean rates for the various demographic and economic factors (i.e. scheme parameters), any primary funding method is unlikely to match exactly emerging experience. Hence, the supplementary funding methods are essential in financing the scheme to a desired target of funding level or solvency level. In view of spreading evenly the undesirable unfunded liability or (actuarial) gains/losses informed at each valuation date, the Spread method (in UK) and the Amortisation of losses method (in Canada and USA) are most commonly used. Thus, the recommended contribution rate is identified by the normal cost plus the adjustment to the normal cost.

Throughout this section 2.2, we describe the general pension funding plan without a prescribed time-invariant relationship (so-called dynamic relationship) between the recommended contribution rate and the actuarial valuation results (e.g. unfunded liability or funding level).

For this reason, the mechanism for determining the recommended contribution rate is unlikely to give a proper indication of the future actuarial valuation results which may be affected by the likely fluctuations in demographic and economic movements.

Thus, the general pension funding plan used in the classical actuarial approach to valuation is interchangeable with the description static pension funding plan (which is in contrast with the dynamic pension funding plan to be considered in the next section 2.3). Here, the term 'static' is adopted from a static (or instantaneous) system in the field of control theory [see McGillem & Cooper (1991; p14)]. In contrast to a dynamic system, a static system is one in which the output response at time t depends only on the input information at time t, and not on any future or past input values (i.e. zero-memory system). In other words, considering the mechanism for producing the recommended contribution rate, the current reference set $\{NC_t, UL_t\}$ available to the actuary through the actuarial valuation at time t (corresponding to the input information at time t) determines instantaneously the current value of C_t according to the actuary's choice of the value of k_t (corresponding to the output response at time t). Hence, this mechanism will be newly carried out at each valuation date (i.e. non-sequential) because the output response C_t does not depend on the past information set $\{k_j, NC_j, UL_j; j < t\}$ but depends only on the current information set $\{k_t, NC_t, UL_t\}$.

As a conclusion, the general pension funding plan used in the classical actuarial approach to valuation would be characterised by the terms 'static' and 'non-sequential'.

2.3 Dynamic pension funding plan

A dynamic pension funding plan is a funding algorithm or function to generate sequentially a recommended contribution rate in the course of time. The full meaning of this title is given in the paragraph (iii) in section 2.3.2.2. Throughout this thesis, there is no conceptual loss of generality in using the pairs of terms dynamic and sequential or static and non-sequential, interchangeably.

2.3.1 Introduction

In an uncertain world, a general pension funding plan used in the classical actuarial valuation (in our terms, a static pension funding plan, see section 2.2.3) can not be said to lead to securing optimally the promised benefits without undue financial burden being placed on the employer - for example the potential of his insolvency/bankruptcy when the scheme is not fully funded (i.e. achieving our funding purpose described previously, particularly in section 2.1.3.4). This is because the static funding plan pays little attention to estimating/examining how the scheme's financial position might appear at each future valuation date (although we are not always able to predict the results of this activity with certainty).

For funding purposes, it is necessary, particularly for the pension actuary, to understand how the various variables (including scheme parameters) composing the mechanism of pension funding interact with one another and evolve with time. This necessity would be an explanation for the recent move from a static (or non-sequential) approach to a dynamic (or sequential) approach (see the following section 2.3.2.4). Indeed, Daykin et al. (1987) state that "Although the EEC and a number of other supervisory authorities adopted a static approach, actuarial opinion has moved in favour of the dynamic concept, whether in respect of general insurance, life insurance or pension funds" Throughout Chapters 4 and 5, we deal with funding from the standpoint of a dynamic approach.

We believe that one of the main branches involved in the development of a dynamic approach is optimal control theory (from the field of engineering). Historically, the starting point in introducing the optimal control theory into a purpose-orientated activity of decision making and/or of analysing the possible misjudgments (e.g. econometrics, operation research and actuarial valuations (whether in respect of general insurance, life insurance and pension schemes)) would be, as mentioned in Benjamin (1984), the publication of Tustin (1953)'s important book.

Let us now restrict our attention to the control of the contribution rate in a defined benefit pension scheme. Most of system-related descriptions are based on Dorf (1992) and for further details, we have consulted other text books, in particular Jacobs (1993).

2.3.2 Actuarial applications of optimal control theory to pension funding plan

This section is intended to provide a conceptual framework for a dynamic (or sequential) pension funding plan rather than to formulate mathematically a dynamic pension funding plan, which will be dealt with in Chapters 4 and 5. Various terminologies involved in optimal control theory are interpreted in the light of the actuarial valuation.

To begin with, we need to distinguish between the terms 'system' and 'control system': the term 'system' is understood as an interconnection of components through the input and output devices for their common purposes, while the term 'control system' is a system subject to control so as to achieve a desired output response by way of one or more controlling variables.

In short, pension funding can be regarded as a control system subject to control mechanisms established by the scheme's actuary.

2.3.2.1 Mathematical model for pension funding system

In order to gain an understanding of how the pension funding system evolves with time and how it is controlled by reference to the information acquired, we must mathematically specify the behaviour of the pension funding system.

In practice, the pension funding system is too complicated to describe mathematically and solve analytically and hence we need to make some simplifying assumptions. Mathematical modelling usually commences from assumptions concerning the system operation, although these assumptions may lead to some criticisms on the grounds of validity/reality of the assumed mathematical model.

Even though mathematical relations between variables of a model, describing the behaviour of a system, can be specified in various forms, for example discrete-time or continuous-time, linear or non-linear, static or dynamic and/or deterministic (nonprobabilistic) or stochastic (probabilistic), our approach is to model a discrete-time linear dynamic system under the following assumptions.

For simplicity, we assume that contribution income and benefit outgo cash flows occur at the start of each unit valuation period (e.g. scheme year), and that the administrative costs of the scheme are paid separately by the employer and then have no effect on the cash-flows affecting the scheme, under which the cash flows of pension fund levels will be well specified by the first-order difference growth equation evolving in discrete time (as an assumed reduced structural form of the actual pension funding system): for $t = 0, 1, 2, \dots,$

$F_{t+1} = (1 + i_{t+1}) \cdot (F_t + C_t - B_t)$ with the given initial condition F_0 --- (2.2)

where

i_{t+1} = rate of investment return earned during the intervaluation period (t, t+1), defined in a manner consistent with the valuation interest rate (i.e. in nominal or real terms); and

B_t = overall actual benefit outgo for the intervaluation period (t, t+1).

We note that the fund growth equation (2.2) is of fundamental importance in our thesis; for convenience, our discussions in section 2.3 are based on this equation. We can also transform this equation into other first-order difference equation under some assumptions, for example, the solvency level growth equation [see section 3.2.4], on which we can base similar discussions without any difficulty.

In practice, B_t and/or i_{t+1} are affected by the demographic and economic environment. If B_t and/or i_{t+1} are modelled to contain any disturbances, representing the influence of random (demographic and economic) environment, the fund growth equation (2.2) becomes a discrete-time linear dynamic stochastic equation. Otherwise, the fund growth equation (2.2) is a discrete-time linear dynamic deterministic equation, in which case the sequence $\{B_t, i_{t+1}; t=0, 1, 2, \ldots\}$ would be mean values estimated by the pension professionals (particularly, by the scheme's actuary).

In summary, the above discussion about pension funding system is condensed by the diagram of Figure 2.1, which is systematically clarified in Figures 2.1.1 and 2.1.2.

$\{\text{Inputs: } C_t, B_t, i_{t+1}\} \longrightarrow F_{t+1} = (1 + i_{t+1}) \cdot (F_t + C_t - B_t) \longrightarrow \{\text{Output: } F_t\}$

Figure 2.1 Pension funding dynamic system identified by equation (2.2).

{Inputs: C_t, B_t, i_{t+1}} $\rightarrow$ $(1 + i_{t+1}) \cdot (F_t + C_t - B_t)$ $\rightarrow$ L (Time delay) $\rightarrow$ {Output: F_t}

↑_ _ _ _ _ _ _ _ _ _ _ _ F_t _ _ _ _ _ _ _ _ _ _ _↓

Fig. 2.1.1 Pension funding deterministic dynamic system: $L\,F_{t+1} = F_t$.

\- Random Environmental Disturbances -

\- - - - - - - - {Inputs: B_t, i_{t+1} } - - - - - - - -

↓

{Input: C_t} $\longrightarrow$ $(1 + i_{t+1}) \cdot (F_t + C_t - B_t)$ $\rightarrow$ L (Time delay) $\longrightarrow$ {Output: F_t}

↑_ _ _ _ _ _ _ _ _ _ _ _ _ F_t _ _ _ _ _ _ _ _ _ _ _ _↓

Figure 2.1.2 Pension funding stochastic dynamic system when B_t and i_{t+1} are all containing some disturbances; $L\,F_{t+1} = F_t$.

Finally, we note that as seen in Figures 2.1.1 and 2.1.2, the pension funding dynamic system has inputs {C_t, B_t, i_{t+1}} and output {F_t} with their own internal transient response mechanism (characterised by the time delay L) such that the output does not respond instantaneously to the inputs, and then it is required to have a one-dimensional memory inside the dynamic system to store the latest output. In this respect, a dynamic system is completely distinguishable from a static system characterised by zero-memory. Next, we consider the mathematical model of the pension funding control system.

2.3.2.2 Mathematical model for pension funding control system

As mentioned in section 2.3.2, the pension funding control system can be regarded as pension funding system subject to control mechanism (so-called control law) established by the scheme's actuary. The control law can be classified into open-loop and feedback (or closed-loop). The open-loop control law determines a recommended contribution rate only on the basis of information about the desired output response. There is conceptually no difference between the general pension funding plan itself and the open-loop control law because the spread parameter in the general funding plan can be thought of as reflecting the desired output response, whereas the feedback control law adjusts a recommended contribution rate on the basis of information about the desired output response together with information fed back about the actual output response; hence, the general pension funding plan subject to regular actuarial valuations is a feedback control system. These is clearly shown in Figures 2.2.1 & 2.2.2 given at the end of this section. That is, the system subject to open-loop control law is called the open-loop control system and the system subject to feedback control law the feedback control system.

Even though each control system has its own advantages and disadvantages [for details, see Dorf (1992; Ch. 3)], we prefer to realize our funding purpose by way of a feedback control law rather than an open-loop control law, particularly for the distinct advantage that a feedback control system provides the ability to adjust the error for both the transient and the steady-state cases.

Remark 2.5: (a) $e(t) \equiv x(t) - y(t)$, where $e(t)$ = [control error at time t in the general control system (e.g. UL_t)], $x(t)$ = [desired output response at time t (e.g. AL_t)] and $y(t)$ = [actual output at time t (e.g. F_t)]; and

(b) The control error $e(t)$ may be considered in two parts: one part of $e(t)$, known as the transient error, which reduces to zero as time increases and the other part of $e(t)$, known as the steady-state error, which remains after transient errors have decayed to zero; hence, as $t \to \infty$, $e(t)$ = [transient error] + [steady-state error] $\to$ [steady-state error], which may be zero, finite, or unbounded. Satisfactory performance of a control system requires that the steady-state error should, at worst, be finite.

Therefore, the control error is of vital importance to the performance of a control system. Here, we shall introduce some typical performance indices of a control system and compare these with our concepts of solvency risk and contribution rate risk described separately in section 2.1.3.2 and 2.1.3.3. Mathematical definitions of each performance index will be given in section 4.1.1.

In order to make control systems behave in some desired way such as reducing control errors, we must first define a quantitative measure of the system's performance representing adequately the control errors (so-called performance index) and then by minimising the performance index, we would be able to obtain the optimal feedback control law. The performance index for control errors is usually formulated in one of three distinct forms on a specified time domain: integral/sum of squares of the control errors, integral/sum of the absolute values of the control errors and integral/sum of the product of time and squares of the control errors (or the absolute values of the control errors) [see Jacobs (1993; section 6.4.1)], where the terms 'integral' and 'sum' means that the index is formulated in a continuous-time domain and discrete-time domain, respectively. So, the performance index for control errors conceptually corresponds to the solvency risk over the same time domain [see section 2.1.3.3].

On the other hand, we may need to measure the amount of control effort expended through the control action errors (so-called cost of control). The performance index for control action errors is usually formulated on a specified time domain as integral/sum of the squares of the control action errors. So, the cost of control conceptually corresponds to the contribution rate risk over the same time domain.

Considering a performance index only for control errors is not a sufficient index for designing a practical, feedback control law because a zero cost of control action means that the decision maker (e.g. actuary) is doing nothing for the improvement of the performance of a control system. It usually appears in a composite index for control errors and control action errors such as in the form of a weighted sum of them. So, the composite index conceptually corresponds to the weighted sum of contribution risk and solvency risk over the same time domain [see section 2.1.3.4].

Thus, analytic control procedures for designing a feedback control law through the minimising of a composite performance index are applicable to the pension funding process (in defined benefit pension schemes) with the aim of realizing our funding purpose (i.e. minimising the weighted sum of contribution risk and solvency risk, see section 2.1.3.4). We note that in optimal control theory, the term 'process' is the device, plant or system under control, so the pension funding process can be regarded as a pension funding control system, which is specifically under the control of the scheme's actuary.

For completeness, we describe how to realize mathematically our funding purpose.

The actuary's control law is oriented towards determining optimally the recommended contribution rate that ensures the realization of the funding purpose. The starting point in establishing the optimal control law would be to specify the controlled object, the control goal subject to control constraints (if necessary) and the available information (produced from the actuarial valuation process).

(i) <u>Controlled object</u>:

The controlled object is that part of a pension scheme which is to be influenced by the control action; as seen in Figure 2.1, the pension funding dynamic system identified by equation (2.2) is the controlled object because it is the main process generating the control error. Then, the inputs $\{C_t, B_t, i_{t+1}\}$ to the controlled object can be regarded as controlling variables and its output $\{F_t\}$ as the controlled variable. However, from the actuary's viewpoint and from our viewpoint, C_t rather than B_t or i_{t+1}, would be treated as a controlling variable.

(ii) Control goal:

The control goal is to realize our funding purpose as described in section 2.1.3.4. If our funding purpose is realizable, then it may be obtained in many ways. Therefore, as a qualitative measure of the efficiency of our funding control system, the performance index is first formulated in accordance with our funding purpose, and based on this we can adjust optimally the controlling variable C_t to realize our funding purpose. Our performance index leads to the weighted sum of the contribution risk and the solvency risk over the projection period [for a mathematical formulation, see section 4.1.1]. Then, the control goal is attained by solving the control problem (i.e. minimising the established performance index, with respect to our controlling variable C_t, subject to the given constraints). Our specific approach to this control problem is to use the method of dynamic programming [see sections 4.2 & 4.3]. The resulting optimal control law is a linear function of the available information (to be discussed in the next paragraph (iii)), which generates an optimal controlling variable (i.e. optimal C_t) sequentially in the course of time as a controlling input to the controlled object; in this respect, the optimal control law is known as the optimal feedback control law. The optimal feedback control law shall be called the dynamic pension funding plan (as mentioned in the title of this section 2.3).

(iii) Available information:

This is a matter for discussion in stochastic control optimisation, and not an issue in deterministic control optimisation. According to the characteristics of available information, control problems are classified into control problem with complete state information and control problem with incomplete state information.

(a) <u>control problem with complete state information</u>:

In an ideal situation, the actuarial valuation process must describe the exact financial status of the scheme at each valuation date. This is the situation of complete state information (or observation). This ideal situation is mathematically modelled by the system equation (2.2) designated to generate outputs in the course of time; that is, for every time t, the financial status of the scheme at time t is exactly equal to the output response from the controlled object at time t. The control problem under this kind of ideal situation is called the control problem with complete state information. Considering a control problem with complete state information, it is not necessary that at each valuation date t, the actuary has knowledge of full information on the past control history $\{F_0, F_1, \ldots, F_{t-1}, C_0, C_1, \ldots, C_{t-1}\}$ in order to determine the current controlling variable C_t, since the current output F_t informed from the actuarial valuation at time t, yields full information on the past control history, sufficient to generate the future state variables without the knowledge of the past control history. For this reason (i.e. representing the state of the funding dynamic system), the output is particularly referred to as the state variable and the system equation as the state equation. Introducing the concept of a state variable can be of benefit to the actuary in the light of memory efficiency because the past control history is increasing with time t as the information gained is retained. Consequently, the accessibility of the current state variable removes the disadvantage of needing to retain all of the past control history (i.e. the current state variable itself is the complete state information).

(b) <u>control problem with incomplete state information</u>:

In a practical situation, the actuarial valuation process can not always provide the exact financial status of the scheme at each valuation date as a result of accounting and auditing work. This is the situation of incomplete state information (or observation). The control problem under this situation is called the control problem with incomplete state information. That is, the state variable F_t is likely to be partial, delayed and/or noise-corrupted, so the current state value is no longer available to the pension actuary. Mathematically, the actuary's valuation process can be specified by a so-called measurement process $\{Z_t, t = 0, 1, 2, \ldots\}$, which is characterised by the measurement equation such that

$$Z_t = p_t \cdot F_{t-b} + \omega_t , \quad t = 0, 1, 2, \ldots \qquad \text{--- (2.3)}$$

where 'p_t' denotes the partial observation parameter at time t, 'b' the time delay parameter in the observation and 'ω_t' the observation noise (or disturbance) with a given probability distribution.

All these features can occur in practice:

- The actuary at time t may be able to observe only certain aspects of the state variable (i.e. in the case of $0 < p_t < 1$, $b = 0$ and $\omega_t = 0$);

- The actuary at time t may consult only on the basis of the situation as it was some time ago (i.e. in the case of $p_t \neq 0$, $b > 0$ and $\omega_t = 0$); and

- The actuary at time t may have only noise-corrupted observation overlaid by an observation disturbance (i.e. in the case of $p_t = 1$, $b = 0$ and ω_t following the given probability distribution).

Of course, in the situation of complete state information (i.e. in a case of p_t, b, $\omega_t = 0$), then $Z_t = F_t$.

The inability of the actuary to observe the exact value of the state variable could be due to the physical inaccessibility of some of the economic scheme parameters (particularly, the real rate of return on equities) and/or to inaccuracies of the procedures used for measurement for the reason that the financial status of the scheme is highly correlated with the random economic and demographic environments and it is difficult to collect the necessary information. In other cases, it may be very costly to obtain the exact value of the state variable even though it may be physically attainable. Thus, delays in knowing the precise financial status may arise because of the time taken to produce the required accounting estimates.

Given that the financial status of the scheme would be reviewed at the next valuation which would be in 1~3 units' time, we believe that it is reasonable to focus on a one-unit time delay in observations. In practice, a one-year time delay would be the most frequent one. Whenever considering a control problem with incomplete state information in this thesis, we assume that the measurement equation is of the form, $Z_t = F_{t-1}$ with the given initial condition $Z_0 = F_{-1}$.

In contrast with the actuarial valuation under the complete state information situation, the actuary at time t does not have direct access to the current state variable F_t and may need to keep the past control history and current valuation result, Z_t; that is, the information at time t available to the actuary is given by $Y_t = \{Z_0, Z_1, \ldots, Z_t, C_{-1}, C_0, C_1, \ldots, C_{t-1}\}$ with the initially given information $Y_0 = \{Z_0, C_{-1}\}$.

Remark 2.6: (a) The current state variable F_t is not observable at time t; for this reason, F_t is usually called the process or conceptual state variable [see Whittle (1983; Ch. 39)]; and

(b) Y_t is increasing with time as one of the characteristic features of a temporal stochastic optimisation problem; hence, in the light of the memory efficiency, it is necessary to find an observable state variable which summarizes all the appropriate current information and is recursively calculable as a best alternative to the current state variable [see Whittle (1983; Ch. 39)]. Then, the control problem with incomplete state information can be reduced to problem of complete state information by means of a reformulation with respect to the obtained observable variable. These discussions are taken further in section 4.3.3.

The above discussion in relation to pension funding control system is summarised by the diagrams in Figures 2.2 ~ 2.4.

To begin with, we interpret the principal factors used in Figures 2.2.1 ~ 2.4, which are all associated with equations (2.2) and/or (2.3):

- Controlled object $^{D \text{ or } S}$ = Pension funding dynamic system with controlling variable C_t and controlled variable F_t; here, the superscript D or S denote the pension funding deterministic or stochastic dynamic system, respectively [see Figures 2.1.1 and 2.1.2];

- Measurement process = Actuarial valuation process specified by measurement equation (2.3) with input $\{F_t\}$ and output $\{Z_t\}$, which is involved in the feedback control, assuming F_t or F_{t-1};

- {Targets (or Reference variables): ct_t, ft_t} = recommended contribution target at time t (representing the desired level of C_t; e.g. NC_t), ct_t, and fund target at time t (representing the desired level of F_t; e.g. AL_t), ft_t, which can be regarded as exogenous variables providing the external reference information which will be decided by the scheme's actuary in consultation with the trustees and sponsoring employer; and

- Estimator of F_t = process for estimating effectively the current value of the conceptual state variable F_t in the light of memory efficiency, under the situation of incomplete state information (i.e. $Z_t = F_{t-1}$), in which the resulting effective point estimate of F_t is denoted as $\hat{F}_t$.

The open-loop control law can be more easily formulated than the feedback control law. It may be largely dependent on the actuary's past experience and capacity for forecasting the future uncertainty because his desired funding action is straightforward but there is no mechanism to fine-tune systematically any misaction (i.e. no feedback path as illustrated below in Figure 2.2.1). On the other hand, most controlled objects are controlled by embedding them in a feedback system with the mechanism to fine-tune systematically any misaction (i.e. feedback path as illustrated below in Figure 2.2.2) because control engineers study the open-loop responses as part of designing and testing the effects of closing the loop to create a proper control system.

Controlled object $^{D \text{ or } S}$ ⟶ {Output: F_t}

{Controlling variable: C_t}↑

{Targets: ct_t, ft_t} ⟶ Open-loop control law

Figure 2.2.1 Pension funding open-loop control system.

Controlled object $^{D \text{ or } S}$ ⟶ {Output: F_t} ⟶

{Controlling variable: C_t}↑ ↓

{Targets: ct_t, ft_t} ⟶ Feedback control law ⟵ Measurement process

Figure 2.3 Pension funding feedback/closed-loop control system.

We note that the above 'Open-loop control law' is unable to be optimally designed in most cases bacause it uses at time t only the target information (not feedback information); for example, a general pension funding plan specified by formula (2.1) in section 2.2.4.1, in which the spread parameter, k_t, can be interpreted as a target level representing the common features of the targets ct_t and ft_t. On the contrary, by applying control optimisation to the pension funding dynamic control system, we can realise our funding purpose by way of designing optimally the above 'Feedback control law'. The resulting optimal feedback control law is defined as dynamic pension funding plan. The control mechanism can be summarised as in Figures 2.3 and 2.4.

Controlled object $^{D \text{ or } S}$ ⟶ {Output: F_t} ⟶ Measurement process

{Controlling variable: C_t}↑ ↓

{Targets: ct_t, ft_t} ⟶ Optimal feedback control law ⟵ {Feedback: $Z_t = F_t$} ↵

Figure 2.3 Optimal pension funding feedback control system under complete state information (i.e. $Z_t = F_t$).

We note that the above 'Optimal feedback control law' is a dynamic pension funding plan under complete state information, designed using control optimisation, which uses at time t not only the complete state information fed back from the measurement process but also the target information to generate a controlling input C_t to the controlled object.

Controlled object ^{S} ⟶ {Output: F_t} ⟶ Measurement process

{Controlling variable: C_t}↑ {Feedback: $Z_t = F_{t-1}$} ↓

↑ → L (Time delay) → {Feedback: C_{t-1}} → ↓

{Targets: ct_t, ft_t} ⟶ Optimal feedback control law ⟵ { $\hat{F}_t$} ⟵ Estimator of F_t

Figure 2.4 Optimal pension funding feedback control system under incomplete state information (i.e. $Z_t = F_{t-1}$).

We note finally that the above 'Optimal feedback control law' is a dynamic pension funding plan under incomplete state information, designed using control optimisation, which uses at time t not only the current information $\hat{F}_t$ transmitted from 'Estimator of F_t' but also the target information to determine the control action C_t. The above 'Estimator of F_t' uses at time t not only the incomplete state information fed back from 'Measurement process' but also the one-unit-time delayed control action C_{t-1} to derive an effective point estimate $\hat{F}_t$ of F_t. Moreover, the combination of 'Optimal feedback control law' plus 'Estimator of F_t' is usually referred to as the optimal feedback controller, which shows the interaction between the separate dual functions of estimation and control [see, Jacobs (1993; section 15.3)].

2.3.2.3 The need for dynamic pension funding plan

This section provides some supporting arguments for our introducing the concept of dynamic pension funding plan and using this in our approach to pension funding. Our dynamic pension funding plan is designed in the form of feedback control in order to cope optimally with both the contribution rate and solvency risks, as mentioned in section 2.3.2.2.

Firstly, we consider why we formulate our funding purpose as minimising the (target-related) contribution and solvency risks at the same time with their relative importance allowing for the specific circumstances of a defined benefit pension scheme.

Benjamin (1989) posed the following pedantic question expressed that "Actuaries have several methods of controlling the funding of pension schemes. There is, however, very little formal comparison between the methods. If it were possible to consider two pension schemes which were identical in all respects except for the method of funding, and we had the complete histories laid before us, what criteria would use to decide which method of funding had done the better job?"

Although so far we may have no definite criteria suitable for answering these questions, we note the views of Loades (1992) expressed that "If the actuarial process is regarded as a control system . . . the criteria for success need to be addressed, taking into account the conflicting interests of the members and sponsors of a defined benefit pension scheme", and also of Haberman (1997) expressed in terms of contribution rate and solvency risks [see section 2.1.3]. Here, 'the conflicting interests of the members and sponsors' can be translated into contribution rate and solvency risks, as described in section 2.1.3.

Consequently, we believe that a reasonable criterion for judging the quality of the pension funding process should be established in view of both stability and security (or solvency) with the scheme's own relative weight placed on them (which exactly corresponds to our pension funding purpose). In our view, the introduction of the somewhat subjective concept of relative weight is essential because, practically, each pension scheme is operating under somewhat different financial and demographic environments, for example Young, Mature and Declining schemes [see section 2.1.4] and on the other hand, stability and security are essentially in conflict with each other to some extent as in the inevitable conflict of interests between the trustees and supporting employer.

Secondly, the following statement can be considered as expressing the necessity of designing a pension funding plan in the form of a feedback control mechanism (i.e. sequentially controlling the contribution and solvency risks at the same time by reference to information fed back about the actual financial status of a scheme).

PLRC (1993; section 4.4.9) argues that "Funding is a necessary condition of security but will be adequate only if the funding process is sufficiently dynamic and flexible to respond to changing circumstances of both assets and liabilities." This supports our intended approach, since the expression 'sufficiently dynamic and flexible to response to changing circumstances' would be well matched with the characteristics of our dynamic pension funding plan, i.e. optimal feedback control law (although the original purpose of this is to address the necessity of introducing the minimum solvency level from the viewpoint of the supervisory authorities, see section 3.2.2).

Lastly, as quoted in Daykin et al. (1987), Humphrys (1984) argued that "what is needed is a forward spread of cash-flow so a clear picture can be represented of what funds will be available from time to time in future and what cash will be needed to meet claims and expenses as they emerge."

In our view, the point of the above statement would be the necessity of having knowledge about the likely future financial position and making arrangements in advance for the future situation, particularly from the words 'a forward spread of cash-flow'. In this respect, mathematical modelling (if well specified) can play a vital role in gaining reliable information about the future financial status. Further, optimal control theory will be effective in this requirement because the feedback control mechanism is originally based on controlling the future situation in advance and sequentially adjusting the controlled object by means of processing the updated information.

2.3.2.4 Review of actuarial applications in pension funding

After Benjamin (1984) advocated the actuarial applications of the concepts and methods of control theory and stated that "It is likely to be useful to apply the concepts and methods of Control Theory to actuarial work . . . Analogy with Control Theory draws attention to the idea of *designing* actuarial control systems", there has been broadly two actuarial applications of control theory, particularly to the pension funding process:

the first is to control effectively the pension funding process with respect to some controllable parameter used in the valuation assumptions; and the other is to control optimally the pension funding process with respect to the recommended contribution rates.

(i) The first application is associated mainly with adjusting effectively some parameter used in the actuarial assumptions (e.g. valuation interest rate, withdrawal rate) and/or the actuarial valuation methods (e.g. frequency of valuations, amortisation period, margin the valuation interest rate) by reference to the actual experience of the scheme parameters (usually, changes in the investment rates of return).

At this point, it is worth drawing attention to three distinct works, Benjamin (1989), Loades (1992) and Fujiki (1994, Ch. 8) (note that Bejamin (1984) provides the basic tool for Benjamin

(1989) but his main results are very similar and even extended in Benjamin (1989), so we focus on Benjamin (1989)).

(a) Benjamin (1989) is based on a discrete-time, deterministic approach. The variability in recommended contribution rates is compared with respect to a time-varying valuation interest rate (specified as the arithmetic mean of real interest rates earned in recent years), subject to a small change in earned real interest rates (i.e. stationary inputs), in which the recommended contribution rate and the valuation interest rate are regarded as a controlled variable (i.e. output) and controlling variable, respectively. One most interesting result is that the best control law for setting the valuation interest rate is averaging over all past experienced rates. Furthermore, it is worth noting that since the actuarial valuation process can be thought of as the mechanism for processing information fed back about the actual value of an output (i.e. measurement process), his funding process corresponds to the feedback control system (not optimal).

(b) Loades (1992) is also based on a discrete-time, deterministic approach, in which he investigated how the recommended contribution rates and (actuarial) surpluses respond to a periodic oscillation in I_t (i.e. interest rate net of salary growth observed at time t), subject to a separate (not concurrent) change in the valuation methods and a time-varying valuation interest rate specified by $i_t = SF \cdot i_{t-1} + (1 - SF) \cdot I_{t-1}$ where $0 \leq SF$ (smoothing factor) ≤ 1.

In his work, the recommended contribution rate and surplus each can be regarded as a controlled variable (i.e. output) and each factor in the valuation method as a controlling variable. One clear and significant result is that the best control law for the stability of outputs with respect to SF is numerically shown as SF = 1, so $i_t = i_{t-1}, \ldots = i_0$, which implies that the stability of outputs depends on the initial value of i_0 (usually, it would be determined as the arithmetic mean of past experienced sufficient cycles I_t, $t < 0$). So, this result is conceptually consistent with that of Benjamin (1989).

On the other hand, Loades's numerical results based on various scenarios for each controlling variable lead to the conclusion that keeping any controlling variable at a fixed value under unstable changing environments is unlikely to be an effective control law from the viewpoint of the stability of both the contributions and surpluses. Thus, Loades's numerical results show a high risk of instability of contributions and/or surpluses, even sometimes insolvency. As a supporting example, Khorasanee (1993) reached the same conclusion, in which he compared several different pension funding methods with a fixed amortisation period (e.g. Projected Unit, Attained Age and Entry Age methods) by means of a simulation model based on unstable historic investment returns net of salary growth. Consequently, we believe that the starting point of both Loades and Khorasanee's approaches are based on regarding the process of funding defined benefit pension schemes as a feedback control system (not optimal) because of the presence of regular actuarial valuations.

(c) Fujiki (1994; Ch. 8) is also based on a discrete-time, deterministic approach. As mentioned earlier in section 2.1.2.2. he investigated how to modify effectively the actuarial assumptions to improve the long-term stability of contributions under separate (not concurrent) changes in real investment rates of return (relative to salary growth and pension increase), withdrawal rates and equity dividend growth rates. The most interesting result has been described in section 2.1.2.2; in particular, the best control law for setting the valuation interest rate is consistent with those of (a) and (b). In the aspect of a control system, his approach assumes that the pension funding process is essentially a feedback control system as in (a) and (b), since the recommended contribution rate is a controlled variable and each parameter to be modified is a controlling variable and further the actuarial valuation process corresponds to the measurement process. Further, his various cash projections of contributions can each be thought of as having been generated according to his resulting control law (not optimal), such as changing immediately or gradually in relation to the actual experience of each parameter.

(d) Summary: The above approaches can be commonly characterised as follows:

- All of the obtained decision rules for setting effectively some parameter used in the actuarial assumptions and/or valuation methods can be considered as being derived from the posterior analysis on the results simulated on each artificial scenario. This exactly corresponds to the feedback control law (not optimal) because each scenario virtually represents the actuary's subjective plan for his desired output responses but there is a mechanism for processing information fed back about the actual output responses (i.e. the actuarial valuation process); and

- The form of recommended contribution rate is a priori given as a general pension funding plan specified by the Spread method [see section 2.2.3.1]: that is, Benjamin (1989) is based on the Projected Unit plus Spread method, Loades (1992) on the Entry Age plus Spread method; and Fujiki (1994) on the Projected Unit plus Spread method. However, they do not try to optimise the value of some parameter, such as a spread parameter, by a mathematical analysis. At this point, it is worth recalling Remark (d) in section 2.2.3.1: that is, Dufresne (1986 & 1988) and Haberman (1992, 1993 & 1994)'s approaches do not provide sequential optimisation provided by optimal control theory, so their approaches are not associated with optimal control theory; in this respect, these are excluded from this section.

Consequently, the above approaches (a)~(c) all consider the development of pension funding from the standpoint that the pension funding process is regarded as a feedback control system which is not optimal.

(ii) Different from the above subsection (i), the principal idea of the following approaches is introduced of specifying a suitable performance index so that we can optimise the pension funding process with respect to the recommended contribution rates by a mathematical analysis; hence, these approaches belong to optimal control theory, whereas the approaches in subsection (i) belong to ordinary control theory. Now, four distinct works are briefly described separately.

(a)' O'Brien (1987) is based on a continuous-time, stochastic approach. He derived the optimal feedback control law in the form of a linear function of the current state variable; hence, he considered a control problem with complete state information (see, subsection (iii) in section 2.3.2.2). The controlled object (i.e. pension funding dynamic system) was linearly formulated on the assumptions such that the benefit outgo is a linear growth function of time, growth rate in membership and salary, and earned rate of return on the fund are mutually independent normal random variables and the scheme is only for active members. The performance index was designated to evaluate the control errors and the cost of control (excluding the contribution rate target level).

(b)' Benjamin (1989) is based on a discrete-time, deterministic approach. Using a given simplified cash projection, he defined the optimal set of future recommended contribution rates as being a set minimising the performance index designed to measure the changes in recommended contribution rates from year to year (i.e. realising the so-called minimum energy control approach in optimal control theory). Hence, different from the other approaches (a)', (c)' and (d)', his approach failed to provide an optimal feedback control law because the controlled object was not given dynamically (i.e. in the form of a difference equation), so the optimal set can be obtained simultaneously by utilizing the Lagrangian-multiplier method, not sequentially.

(c)' Vanderbroek (1990) is based on a continuous-time, deterministic approach. As in (a)', she also derived the optimal feedback control law in the form of a linear function of the current state variable. The controlled object (i.e. pension funding dynamic system) was linearly formulated on the assumptions such that the benefit outgo, total payroll and the actuarial present value of future benefits are each exponential functions of time. The performance index was designated to evaluate the control errors and the cost of control relative to the contribution rate target level. She is particularly concerned with the application to the case of national social security plans.

(d)' Haberman & Sung (1994) is based on a discrete-time, deterministic and stochastic approach. They derived the optimal feedback control law in the form of a linear function of the current state variable; hence, they considered a control problem with complete state information [see subsection (iii) in section 2.3.2.2]. The controlled object (i.e. pension funding dynamic system) was linearly formulated as in equation (2.2) in section 2.3.2.1; particularly, the stochastic controlled object is modelled on the assumption that the investment returns are represented by independent and identically distributed random variables. The performance index was chosen to evaluate the control errors and the cost of control relative to the contribution rate target level. They are in particular concerned with the application to the pension funding plan for a defined benefit pension scheme. This investigation can be regarded as part of our earlier work, and will be extended and modified primarily in Chapters 4 and 5.

(e)' Summary: The above approaches have the following common characteristics, which are distinguishable from the approaches in subsection (i), that is,

- They are start with specifying a suitable performance index;
- The controlled object is dynamically formulated in a linear form, except for (b)';

- The form of recommended contribution rate is not specified a priori and is to be sought to produce the minimisation of an established performance index subject to some constraints by means of a mathematical analysis (not numerical analysis); and

- The resulting control law is a function of the current state variable (i.e. optimal feedback control law under complete state information), which generates the recommended contribution rates sequentially, except for (b)'.

2.3.3 Summary and Conclusion

It would be reasonable to organise pension funding by reference to the fund and contribution targets and assess the difference between the two sequentially over a projection period. This would be realizable by the dynamic pension funding plan derived from the application of optimal control theory to pension funding process. In other words, once the pension funding process is well specified mathematically and considered as a feedback/closed-loop control system, it is likely to be useful to apply the concept of control optimisation.

In this thesis, we shall call the pension funding plan which realizes our pension funding purpose by means of control optimisation a dynamic pension funding plan; that is, our dynamic pension funding plan is a so-called optimal feedback control law. Thus, our dynamic pension funding plan is generally characterised by sequentially and optimally controlling the contribution and solvency risks in the course of time and adjusting the two risks by means of processing information generated sequentially in time by the feedback mechanism.

As a conclusion, we shall consider the pension funding process as a feedback control system (not an open-loop control system) throughout this thesis for the main reason that our pension funding purpose can be realized optimally on a feedback control system by means of mathematical analysis (i.e. dynamic programming approach). We note that it is very difficult, except by chance, to realize our pension funding purpose optimally on a feedback control system by means of numerical analysis (i.e. trial-and-see approach) because generally, it requires a considerable number of simulations and our planned future scenarios may not completely cover the future range of uncertainty.

Chapter 3 Mathematical models for the financial strength of defined benefit pension schemes

3.1 Introduction: Basic mathematical model

In this section, we consider schemes providing age-service retirement benefits, which are defined by the product of a fixed accrual rate of pension benefits, the number of active service years and the (current or projected) salaries. For consistency with our previous discussion in Chapter 2, we shall concentrate on constructing a mathematical model on a discrete-time domain, subject to some simplifying assumptions, from which we can easily understand the behaviour over time of the financial structure of defined benefit pension schemes and the relations between variables in the model. We note that the simplifying assumptions employed may cause some criticisms on the grounds of validity/reality, as mentioned earlier in section 2.3.2.1 and also that it is possible to derive the continuous-time versions of the following models (which are simply given at the end of this section).

3.1.1 Actuarial liability growth equation

Assuming that for a given set of actuarial assumptions, the expected benefit outgoes (EB) and the normal costs (NC) each occur at the beginning of each unit valuation period (e.g. scheme year), so that the actuarial liabilities (AL) are measured prior to either NC or EB, then the following recurrence relation holds over two consecutive valuation dates t, t+1 (corresponding to the fund growth equation (2.2) in section 2.3.2.1): for each valuation date $t = 0, 1, 2, \ldots,$

$$v \cdot \nabla AL_{t+1} = d_v \cdot AL_t + NC_t - EB_t \quad \{ \Leftrightarrow AL_{t+1} = (1+i_v) \cdot (AL_t + NC_t - EB_t) \} \qquad --(3.1)$$

with the initial condition AL_0 specified

where

∇ = backward difference operator defined by $\nabla AL_t = AL_t - AL_{t-1}$, so that $\nabla^2 AL_t = \nabla(\nabla AL_t) = AL_t - 2AL_{t-1} + AL_{t-2}$ and in general $\nabla^n AL_t = \nabla(\nabla^{n-1} AL_t)$, n = 2,3, ...;

i_v = assumed valuation interest rate during the unit intervaluation period (t, t+1);

$d_v = i_v/(1+i_v)$, assumed valuation discount rate during (t, t+1); and

$v = (1+i_v)^{-1} = 1 - d_v$, assumed discount function during (t, t+1)

; we shall call the above first-order linear difference equation (3.1) the actuarial liability growth equation.

The dynamic relationship specified by equation (3.1) provides the fundamentals for understanding the financing of any type of pension scheme, such as Young, Mature and Declining schemes (see section 2.1.4.2).

Remark 3.1: Trowbridge (1952) describes the mature case (i.e. stationary population with no inflation and no other growth of salaries over time), for which AL_t, NC_t and EB_t are all constant and then the equation reduces to $NC + d_v \cdot AL = EB$, known as the equation of equilibrium (or equation of maturity).

3.1.2 Net actuarial liability growth equation

When the present moment is taken as time t+1, the net increase in actuarial liability during the intervaluation period (t, t+1), NIL(t, t+1) ($\equiv \nabla AL_{t+1}$), is known and so the sequence {NIL(t, t+1), t = 0, 1, 2, ...} shows the time-path of one-unit-time changes of AL and gives information on the development of the scheme's actuarial liabilities.

Here, we can also establish the linear first-order difference equation of NIL such that for each valuation date t = 0, 1, 2, ...,

$NIL(t, t+1) = (1+i_v) \cdot \{NIL(t-1, t) + \nabla(NC_t - EB_t)\}$ with the initial condition $NIL(-1, 0) = AL_0$

--- (3.2)

; we shall call this equation the net actuarial liability growth equation (note that $NIL(t, t+1) = 0$ in the Trowbridge mature case).

Let $t_0 \leq t_1 \leq t_2$, where t_0, t_1 and $t_2 \in \{0, 1, 2, ...\}$and consider the change in AL during the period (t_0, t_2). Then,

$$AL_{t_2} - AL_{t_0} = AL_{t_2} - AL_{t_1} + AL_{t_1} - AL_{t_2} \Leftrightarrow NIL(t_0, t_2) = NIL(t_0, t_1) + NIL(t_1, t_2).$$

Accordingly, we can generally formulate by induction as follows: for any $k \in \{0, 1, 2, ...\}$ and $t_0 \leq t_1 \leq, ..., \leq t_k$ where $t_0, t_1, ..., t_k \in \{0, 1, 2, ...\}$,

$$NIL(t_0, t_k) = NIL(t_0, t_1) + NIL(t_1, t_2) + ... + NIL(t_{k-1}, t_k) \text{ with } NIL(t_k, t_k) = 0 \qquad \text{--- (3.3)}$$

; this property shall be called the property of additive consistency in NIL.

3.1.3 Fund growth equation

In section 2.4.2.1, we have already built the fund growth equation (2.2) on a discrete-time domain under the assumptions that contribution incomes (C_t) and actual benefit outgoes (B_t) occur at the start of each unit valuation period, and that the cost of administration is paid separately by the employer: for each valuation date $t = 0, 1, 2, ...$, and for a given initial condition F_0, the valuation-output sequence (or capital growth sequence) $\{F_0, F_1, F_2, ...\}$ can be characterised by the following forward recurrence equation,

$$F_{t+1} = (1 + i_{t+1}) \cdot (F_t + C_t - B_t) \Leftrightarrow v_{t+1} \cdot \nabla F_{t+1} = d_{t+1} \cdot F_t + (C_t - B_t) \qquad \text{--- (3.4)}$$

where $d_{t+1} = i_{t+1} / (1+i_{t+1})$ and $v_{t+1} = 1 - d_{t+1} = 1 / (1+i_{t+1})$.

This linear dynamic fund growth equation can be thought of as representing the actual financial structure for the funding of defined benefit pension schemes.

Remark 3.2: (a) In practice, the asset variability inherent in the asset mix depends on the nature and distribution of the scheme's assets. By denoting t+ as the beginning of each unit valuation period, then $F_{t+} \equiv F_t + C_t - B_t$ (i.e. accumulated fund level at time t+) can be thought of as the sum of fund levels of N types of assets, that is, for integer $N \geq 1$, $F_{t+} = \sum_{k=1}^{N} F_{k,t+}$ where $F_{k,t+}$ = value of asset-type k at time t+. Hence, $i_{t+1} = (F_{t+1} - F_{t+}) / F_{t+} = \sum_{k=1}^{N} i_{k,t+1} \cdot (F_{k,t+} / F_{t+})$, where $i_{k,t+1}$ = rate of return earned on the asset-type k during the intervaluation period (t, t+1): in other words, there is no loss of generality in taking the value of each type asset as non-negative at any time and hence the total rate of return during (t, t+1), i_{t+1}, is represented by a convex combination of a finite number of $\{i_{k,t+1}; k=1, 2, \ldots, N\}$ with its smoothing factor $0 \leq F_{k,t+}/ F_{t+} \leq 1$. A number of models of asset portfolio behaviour have been discussed in the literature: see, for example, Black (1992b; Ch. 2) for a general mean-variance model, Janssen (1994) for a dynamic stochastic asset-liability management model, Wilkie (1995) for a stochastic asset model and Dardis & Huynh (1995) and Kemp (1996) for a simulation-based asset-liability management model;

(b) In the case of stationary funds (i.e. when $F_t = F$ constant for any t), then the above linear dynamic system equation reduces to the linear static system equation, that is, $B_t = C_t + d_{t+1} \cdot F$, in which $d_{t+1} = i_{t+1}/(1+i_{t+1})$, actual discount rate for the period (t, t+1); and

(c) It should be noted that the equation (3.4) would be controlled by C_t (which is, in practice, recommended by the pension actuary) sometimes using B_t, which is adjustable with the approval of trustees, sponsoring-employer and supervisory authorities, or using both of C_t and B_t. Even for a stationary fund, the fund can remain stationary (*i.e.* constant F) mainly by controlling C_t.

3.1.4 Net fund growth equation

As the counterpart of NIL(t, t+1), the net increase in funds during a unit intervaluation period (t, t+1), NIF(t,t+1) ($\equiv \nabla F_{t+1}$), is evaluated by the linear first-order difference equation such that for each valuation date t = 0, 1, 2, ...,

$$NIF(t, t+1) = (1+i_{t+1}) \cdot NIF(t-1, t) + (1+i_{t+1}) \cdot \nabla(C_t - B_t) \text{ with the initial condition } NIF(-1,0) = F_0 \quad \text{--- (3.5)}$$

; we shall call this equation the net fund growth equation (note that NIF(t, t+1)=0 in the case of a stationary fund).

Following the same procedure by which the property of additive consistency in NIL has been driven, we can easily deduce the following property: for any $k \in \{0, 1, 2, ...\}$ and $t_0 \leq t_1 \leq, ..., \leq t_k$ where $t_0, t_1, ..., t_k \in \{0, 1, 2, ...\}$,

$$NIF(t_0, t_k) = NIF(t_0, t_1) + NIF(t_1, t_2) + \ldots + NIF(t_{k-1}, t_k) \text{ with } NIF(t_k, t_k) = 0 \quad \text{--- (3.6)}$$

; this property shall be called the property of additive consistency in NIF.

3.1.5 Net unfunded liability (NUL) growth equation

As mentioned earlier in subsection (iii) in section 2.1.2.7, the unfunded liability (denoted by UL) is usually defined as $UL_t \equiv AL_t - F_t$. In addition to UL_t, the net increase in UL (denoted by NUL) during a unit period (t, t+1) could be also employed as a measure of the financial strength of defined benefit pension schemes, in particular for showing how the financial strength of the scheme is changing during the intervaluation period. So, we define NUL(t, t+1) $\equiv UL_{t+1} - UL_t$, which is identical to NIL(t, t+1) - NIF(t, t+1) because $UL_t = AL_t - F_t$.

Subtracting equation (3.5) from equation (3.2) leads to the following result: for each valuation date t = 0, 1, 2, ...,

$$NUL(t, t+1) = (i_v \cdot AL_t - i_{t+1} \cdot F_t) + [(1+i_v) \cdot NC_t - (1+i_{t+1}) \cdot C_t] - [(1+i_v) \cdot EB_t - (1+i_{t+1}) \cdot B_t] \quad \text{- - - (3.7)}$$

; in particular, NUL(t, t+1) = 0 in the Trowbridge mature and stationary fund cases.

In some situations when valuations are performed every n-unit periods, it is possible only to measure the net increase during n-unit period (t, t+n), NUL(t, t+n). The sequence {NUL(t, t+n), t=0, n, 2n, ... and n = 0, 1, 2, ...} is also completely determined by the relationship NUL(t, t+n) = NIL(t, t+n) - NIF(t, t+n). Furthermore, as established in NIL and NIF, the property of additive consistency in NUL is provided by those in NIL and NIF: that is,

$$\begin{aligned} NUL(t, t+n) &= NIL(t, t+n) - NIF(t, t+n) \\ &= \{NIL(t, t+1) + NIL(t+1, t+2) + \ldots + NIL(t+n-1, t+n)\} \\ &\quad - \{NIF(t, t+1) + NIF(t+1, t+2) + \ldots + NIF(t+n-1, t+n)\} \\ &= NUL(t, t+1) + NUL(t+1, t+2) + \ldots + NUL(t+n-1, t+n). \end{aligned}$$

3.1.6 A specific funding formula based on NUL

In this section, we consider a specific pension funding formula based on NUL. We assume that the valuation starts at time 0 and is performed every unit period (t, t+1), the Trowbridge mature case is exactly realised, the scheme's funds are stationary, and AL_0 and F_0 are initially given.

The above condition is summarised as follows:

for every $t \in \{0, 1, 2, ...\}$, $B_t = EB$, $EB = NC + d_v \cdot AL_0$ and $B_t = C_t + d_{t+1} \cdot F_0$ and hence NUL(t, t+1) = 0 (i.e. $UL_t = UL_0 = AL_0 - F_0$, constant for all t).

Therefore, the corresponding pension funding formula is to be formulated as follows in order to maintain UL_t at the equilibrium state UL_0 for all t subject to $B_t = EB$ for all t:

$$C_t = NC + d_v \cdot AL_0 - d_{t+1} \cdot F_0 \quad \text{for all } t \in \{0, 1, 2, ...\} \qquad \text{---}(3.8)$$

; this funding formula can be regarded as presenting the static funding mechanism for a Mature scheme in a discrete-time domain.

In order to achieve stability of the contributions, the scheme's actuary may have an option to adjust d_v (i.e. i_v) in the on-going valuation process with the intention to reduce the gap between the actually experienced $d_t{\cdot}F_0$ and the most likely expected $d_v{\cdot}AL_0$. In this respect, this parameter can be thought of as a controlling variable under the control of the scheme's actuary.

On the other hand, the fund manager is going to face a certain opportunity set of scheme assets (i.e. asset allocation decision). Thus, he attempts to achieve an investment rate of return of at least $d_{t+1} \geq d_v{\cdot}AL_0 / F_0$ with the purpose of improving the financial strength of the scheme and/or reducing the financing burden on the sponsoring employer. However, he is not free to adjust exactly its value of d_{t+1} through his asset mix/portfolio selection decision, since d_{t+1} (i.e. i_{t+1}) is highly dependent on the prevailing investment markets.

Therefore, the pension actuary may face two extreme cases in respect of reducing the variability of contribution rates:

The first case is that the actual investment rate of return earned during (t, t+1) is realised exactly for all t in accordance with the assumed valuation interest rate (i.e. $d_{t+1} = d_v$ for all t). Then, there is no necessity to modify the assumed valuation interest rate (i_v). If $i_v > 0$, then the funding plan (3.8) can then be transformed into the following form representing stationary funding: that is,

$C_t = NC + (AL_0 - F_0) / \ddot{a}_{\infty}(i_v)$, constant for all $t \in \{0, 1, 2, ...\}$ --- (3.9)

where $\ddot{a}_{\infty}(i_v)$ represents the net present value of the perpetuity-due calculated at i_v, assuming $i_v > 0$.

We note that the funding plan (3.9) is consistent with the theoretical result of Haberman (1993) considering the stabilisation of future contribution rates such that the lower is the variance of investment rates of return, the larger is the amortisation period, that is, as the variance decreases to zero the amortisation period increases to infinity. Moreover, if the initial unfunded liability (UL_0) equals zero, then $C_t = NC$ for all t will result in $UL_t = UL_0 = 0$ (i.e. funding level = 100%) for all t, which represents the most desirable (or ideal) funding mechanism from the viewpoint of stability and security.

In the second case, there is a difference between the valuation interest rate i_v and the actually experienced investment rate of return i_{t+1}. Here, two extreme options are considered (note that the other possible options will lie between these two extremes): (a) modifying i_v immediately to equal i_{t+1} and (b) keeping i_v deliberately different from i_{t+1}.

To begin with, we note that for notational convenience, the superscripts 'a' and 'b' on the left side of each principal algebraic symbols are used to indicate that they concern options (a) and (b), respectively.

Now, we are concerned with the former option (a) such that $i_v \rightarrow i_{t+1}$, while the latter option (b) will be discussed at the end of this section. Accordingly, AL_t (assumed to be AL_0) is changed to ${}^aAL_t$ fixed for the period (t, t+1) and NC_t (assumed to be NC) is also changed to ${}^aNC_t$ fixed for the period (t, t+1), since the valuation basis is continuously modified according to the investment performance. Then, $EB = NC + d_v \cdot AL_0$ is no longer applicable because this is based on a constant valuation basis.

Hence, the new actuarial liability growth equation ${}^{a}AL_{t+1} = (1+i_{t+1}) \cdot ({}^{a}AL_t + {}^{a}NC_t - EB)$ for the period (t, t+1) should be applied for the varying valuation basis and the corresponding fund growth equation is ${}^{a}F_{t+1} = (1+i_{t+1}) \cdot ({}^{a}F_t + {}^{a}C_t - EB)$ during the period (t, t+1). Supposing further that $i_{t+1}>0$, then ${}^{a}C_t = {}^{a}NC_t + ({}^{a}AL_t - {}^{a}F_t) / \ddot{a}_{\infty}(i_{t+1})$ for the period (t, t+1), in which $\ddot{a}_{\infty}(i_{t+1})$ is calculated at the newly determined valuation interest rate i_{t+1} for the period (t, t+1).

Therefore, if the valuation interest rate is adjusted immediately after the change in the actually experienced investment rate of return and to the same degree, then ${}^{a}C_t$ is not constant but varies according to the fluctuation of the investment performance instead of being maintained at the equilibrium state. If it is assumed further that $UL_0 = 0$, then ${}^{a}C_t = {}^{a}NC_t$ will maintain ${}^{a}F_t = {}^{a}AL_t$ (i.e. funding level = 100%) for all t, which implies that after amortising the initial unfunded liability, the policy of adjusting immediately and exactly i_v to the investment performance will provide the fully funded security of the scheme (although this policy may lead to the instability of contribution rates). Consequently, this policy is unlikely to balance the conflicting interests between the trustees and the employer.

It is worth recalling that the analyses of Benjamin (1989), Loades (1992) and Fujiki (1994), described in subsection (i) in section 2.4.2.4, indicate that a satisfactory policy for determining valuation interest rates would be the averaging of the experienced investment rates of return. This would seem to balance the conflicting interests of the employer and trustees.

Now, we consider option (b) (i.e. keeping i_v different from i_{t+1}). Then, there are no changes in NC_t and AL_t (i.e. NC_t = (assumed) NC and AL_t= (assumed) AL_0) and hence $EB = NC + d_v \cdot AL_0$ is true because this is based on a constant valuation basis. The corresponding fund growth equation is ${}^{b}F_{t+1} = (1+i_{t+1}) \cdot ({}^{b}F_t + {}^{b}C_t - EB)$ during the period (t, t+1). According to our funding policy (i.e. NUL(t, t+1) = 0), ${}^{a}C_t$ is of the form

${}^{b}C_t = [(1+i_v)/(1+i_{t+1})]\cdot NC + (i_v\cdot AL_0 - i_{t+1}\cdot {}^{b}F_t)/(1+i_{t+1}) + [(i_{t+1} - i_v)/(1+i_{t+1})]\cdot EB$ for the period (t, t+1).

; hence, this funding formula can be interpreted as a linear combination of three terms, each of which relates to the gap between the actual investment rate of return and actuarial valuation rate.

If it is assumed further that $UL_0 = 0$, then ${}^{b}C_t$ governed by the above formula will maintain ${}^{a}F_t = AL_0$ (i.e. funding level = 100%) for all t. This implies that after amortising the initial unfunded liability, the funding policy of adjusting the level of ${}^{b}C_t$ (instead of adjusting immediately and exactly i_v to the realised investment performance as investigated in option (a)) will also provide the fully funded security of the scheme but this funding policy may lead to the instability of contribution rates. As in option (a), this funding policy is also unlikely to balance the conflicting interests of the employer and trustees - the next section addresses this problem.

3.1.7 A general funding formula based on NUL

In this section, we generalise the specific funding formula (3.8) derived in the above section 3.1.6 and compare it with the spread funding formula (2.1)' introduced in section 2.2.3.1.

The sequence of unfunded actuarial liabilities, $\{UL_t; t = 0, 1, 2, ...\}$, is convertible to that of the net increase in UL, defined as $\{NUL(t, t+1); t = 0, 1, 2, ...\}$. In general, UL_t can be thought of as a point measure giving information on the financial strength of pension schemes at time t, but in contrast, NUL(t, t+1) is a period measure showing how the financial strength of the pension schemes is developing during the inter-valuation period (t, t+1). So, from the viewpoint of the long-term, on-going position, the newly defined measure NUL (and its progress over time) may be thought of as a compromise for measuring the financial strength of the scheme to balance the conflicts of interest between the trustees and the employer. For example, in the case of an underfunded (Young, Mature or Declining) pension scheme, the financial strength of the scheme will gradually approach that of a fully funded (i.e. funding level = 100%) state by setting up the funding policy as $NUL(t, t+1) = N_t < 0$, during a control period (how the value of N_t is reasonably chosen is illustrated below).

Of course, whenever the above-mentioned three kinds of scheme have already been fully funded we can maintain its financial status by keeping $NUL(t, t+1) = 0$ during the control period. In some situations, it may be necessary to fortify the security of pension schemes, and the pension actuary can boost the speed of approaching the desired funding level simply by lowering the value of $NUL(t, t+1)$, such that $NUL(t, t+1) = N_t' < N_t$, during the control period.

Thus, we believe that such a funding policy that $NUL(t, t+1)$ = specified value of N_t for the period (t, t+1) could harmonise the conflicts of interest between the sponsoring employer and the trustees by means of a reasonable choice of N_t: from equation (3.7), $NUL(t, t+1) = N_t$ leads to, for each t,

$$C_t = [(1+i_v)/(1+i_{t+1})]\cdot NC_t + (i_v\cdot AL_t - i_{t+1}\cdot F_t)/(1+i_{t+1}) + [B_t - (1+i_v)/(1+i_{t+1})\cdot EB_t] - N_t/(1+i_{t+1}) \quad (3.10)$$

Here, $N_t \in (-\infty, \infty)$ can be thought of as a speed adjusting controlling variable at time t, since it should be determined mainly by the compromise between the trustees and the employer on the future financial strength of the scheme, and furthermore by any current statutory requirements. The N_t determined will then indicate how quickly the scheme moves towards its new required funding level. In general, a value of $N_t < 0$ is suitable for underfunded schemes (i.e. scaling down the deficit), that of $N_t = 0$ for fully-funded schemes and that of $N_t > 0$ is appropriate to the all kinds of over-funded schemes (i.e. scaling down the surplus). Moreover, the funding formula (3.10) can be interpreted as a linear combination of three terms, each of which relates to the gap between the actual experience and actuarial assumptions: that is, the first can be referred to as a term reflecting the difference between i_v and i_{t+1}, $[(1+i_v)/(1+i_{t+1})] \cdot NC_t$, the second as a term reflecting difference between i_v and i_{t+1}, EB_t and B_t, and AL_t and F_t, i.e. $(i_v \cdot AL_t - i_{t+1} \cdot F_t) / (1+i_{t+1}) + [B_t - (1+i) / (1+i_{t+1}) \cdot EB_t]$, and the last representing the target value of NUL, i.e. $N_t/(1+i_{t+1})$. (These terms can be further analysed, using algebraic identities, to identify the separate contributions to the gap between actual experience and actuarial assumptions). In particular, if i_{t+1} is exactly equal to i_v and B_t is given by EB_t, then

$C_t = NC_t + (AL_t - F_t)/\ddot{a}_\infty(i_v) - N_t/(1+i_v)$.

Consequently, the newly defined funding formula (3.10) can be regarded as another version of the spread funding formula (2.1)', which will be explained at the end of this section.

Now, we shall consider how to determine the control period and how to amortise gradually the positive initial unfunded liability (i.e. $UL_0>0$) and eventually achieve the fully funded scheme, in the light of a fixed funding policy with respect to time t such that $NUL(t, t+1) = N < 0$ for all t.

With $UL_0 > 0$, the definite solution of the first-order difference equation $NUL(t, t+1) = N$ is that $UL_t = UL_0 + N \cdot t$ for $t = 0, 1, 2, \ldots$; hence, the sequence $\{UL_t;\ UL_0 > 0, t = 0, 1, 2, \ldots\}$ is strictly decreasing and there is no equilibrium state. However, we may interpret the time path of UL_t as being a constant deviation (UL_0) from the moving equilibrium state $N \cdot t$ which is decreasing in time t.

Let T be the control period to be taken to achieve a funding level = 100%. Hence, T = 1 + (greatest integer $\leq - UL_0 / N$) by setting $UL_t = UL_0 + N \cdot t = 0$, which shows that the length of T is proportional to the value of UL_0 and inversely proportional to N. Replacing N_t as constant N in equation (3.10) and then keeping these contribution rates over a control period [0, T] will make the unfunded liability decrease and the funding level reach 100% at time T. The time-path of UL_t, $0 \leq t \leq T$, is dependent upon both the size of UL_0 and the scale of N. During this control period, the reduction in volatility of the levels of contribution may be provided by adjusting the current valuation rate in the form of the convex combination of two extreme values as suggested by Loades (1992) [see section 2.3.2.4]. After time T, the pension actuary, if necessary, can reduce the value of N to zero in order to maintain the fund ratio at the level of 100%.

Finally, in order to illustrate the similarity between the spread funding formula (2.1)' and the newly defined funding formula (3.10), we assume that all actuarial assumptions are realised exactly and the valuation interest rate $i_v > 0$. Therefore, formula (3.10) reduces to $C_t = NC_t + UL_t/\ddot{a}_\infty(i_v) - N_t/(1+i_v)$, which is comparable with formula (2.1)', i.e. $C_t = NC_t + UL_t/\ddot{a}_n(i_v)$.

Letting $N_t = (1+i_v)\cdot[1/\ddot{a}_\infty(i_v) - 1/\ddot{a}_n(i_v)]\cdot UL_t = -[1/s_n(i_v)]\cdot UL_t$ makes the above two funding formulae equivalent to each other, where $s_n(i_v)$ = accumulated value of an immediate annuity of 1 per unit time for n-unit periods (integer $n \geq 1$), calculated at i_v.

In conclusion, the spread funding formula (2.1)' can be thought of as a special form of our funding formula (3.10).

3.1.8 Continuous-time version of NUL

As derived by Bowers et al. (1976), the continuous-time version of the actuarial liability growth equation (3.1) is that for δ being the force of valuation interest and any $t \in [0, \infty)$,

$$dAL_t/dt = \delta \cdot AL_t + NC_t - EB_t \text{ with the initial condition } AL_0 \text{ specified} \qquad \text{--- (3.11)}$$

Moreover, in the Trowbridge mature case (i.e. NC_t and EB_t are constant over time t and $dAL_t/dt=0$) then the equation reduces to $NC + \delta \cdot AL = EB$, which can be regarded as the equation of equilibrium on a continuous-time domain.

Accordingly, the continuous-time version of the fund growth equation (3.4) will be given as for any $t \in [0, \infty)$,

$dF_t/dt = \delta_t \cdot F_t + C_t - B_t$ with the initial condition F_0 specified --- (3.12)

where δ_t = force of investment rate of return at time t.

Additionally, in the case of stationary funds, then equation (3.12) reduces to the linear static equation $B_t = C_t + \delta_t \cdot F$, which makes the obvious statement that the actual benefit outgoes must derive from contributions and/or from the investment return earned on the scheme assets.

On a continuous-time domain, the marginal increase in UL (denoted by MUL as the continuous-time version of NUL) at time t can be defined as $MUL_t \equiv dUL_t/dt = dAL_t/dt - dF_t/dt$. Subtraction equation (3.12) from equation (3.11) provides

$MUL_t = (\delta \cdot AL_t - \delta_t \cdot F_t) + (\delta \cdot NC_t - \delta_t \cdot C_t) - (\delta \cdot EB_t - \delta_t \cdot B_t)$ --- (3.13)

; in particular, MUL(t) = 0 in the Trowbridge mature and stationary fund cases and further, we can establish and discuss a general funding plan based on this measure MUL in a very similar manner as in section 3.1.7. Also, assuming that for any given h > 0, NUL(t, t+h) is a continuous function of $t \in \{0, \infty)$, we can define MUL(t) as the instantaneous rate of change of NUL, that is, $MUL(t) = \lim_{h \to 0^+} NUL(t, t+h) / h$, where h>0 and $h \to 0^+$ denotes that h tends to zero from above.

Furthermore, if MUL(t) and NUL(t_0, t) are continuous functions of t for $t_0 \leq t$, t_0 and $t \in [0, \infty)$ with the initial condition NUL(t_0, t_0) = 0 specified, and the principle of additive consistency in NUL holds, then we can derive the relationship between MUL and NUL such that for $t_0 \leq t_1 \leq t_2$, $\mathrm{NUL}(t_1, t_2) = \int_{t_1}^{t_2} \mathrm{MUL}(s)ds$ [for the proof, see Appendix 3 given at the end of this chapter].

Of course, the above relationship is simply obtainable under the assumption that UL_t is differentiable at all t: for $t_0 \leq t_1 \leq t_2$,

$$\mathrm{UL}_{t_2} - \mathrm{UL}_{t_1} = \int_{t_1}^{t_2} d\mathrm{UL}_s \quad \Leftrightarrow \quad \mathrm{NUL}(t_1, t_2) = \int_{t_1}^{t_2} \mathrm{MUL}(s)ds$$

; consequently, NUL can be evaluated from MUL and the reverse is also true.

3.2 Minimum solvency requirement and Minimum funding requirement

3.2.1 Introduction

In section 2.1.2.7, we have already outlined the concepts of the minimum solvency requirement (MSR) proposed by the Pension Law Review Committee (PLRC) and of the minimum funding requirement (MFR) enacted by the Pensions Act 1995; in particular, the MFR will have a significant impact on funding plans in the future (after April 1997) because there is currently no statutory requirement for funding the accruing liabilities in advance.

In this section, we shall give some more detailed descriptions about the recommendations for the MSR because these will provide a rigorous guidance or yardstick for setting a pension funding plan in order to reduce the risk of insolvency of occupational pension schemes. And also, the MFR will be briefly described, which has now taken the place of the MFR.

As discussed earlier in subsection (ii) in section 2.1.2.7, the term 'solvency' can be used in a way that satisfies the MSR or MFR imposed by a supervisory authority (or government) through legislation and/or regulation. Hence, the solvency test is intended to test whether the scheme's assets as assessed by legislation and/or regulation would be sufficient to cover its liabilities as assessed by legislation and/or regulation. The method for valuing the scheme assets and liabilities must be prescribed for both the MSR or MFR.

3.2.2 Minimum solvency requirement (MSR)

In response to the growing interest in and concern about the security of occupational pension schemes following the so-called Maxwell affair (involving the loss of the pension funds through the misappropriation of scheme assets), the UK government set up the PLRC as an independent committee, under the chairmanship of Roy Goode, professor of law at the university of Oxford, in June 1992, soon after Mr. Maxwell's death in November 1991.

This committee was organised to review the present law and regulation for occupational pension schemes and hence ultimately provide useful recommendations for tightening the law and regulation for occupational pension scheme management and the administrative arrangements.

The objective was to provide greater protection of the accrued pension rights of the (active, deferred and retired) members against any kind of Maxwell-type fraud or misappropriation which could result in the fund being in deficiency [see Pension Law Reform Volume I Report (1993; p iii)] and to restore confidence in occupational pension schemes as a whole. Although the PLRC's 218 recommendations published on 29 September 1993 are addressed mainly to the security of occupational pension schemes, we are concerned about the pension funding related recommendations, and the so-called minimum solvency requirement (MSR).

The PLRC recommendations for the MSR are intended primarily to ensure that the schemes will be able to meet their legal commitments, as prescribed in the trust deed, to all the scheme's members at least in the level of 90 per cent, even in the event of the schemes being wound up immediately.

These can be summarised as follows (for more details, see Pension Law Reform Volume I Report, Ch. 4.4):

(R1) MSR should be introduced for all funded schemes, except those that provide benefits in excess of Inland Revenue earning-related limits or earnings caps (e.g. unapproved top-up pension schemes);
(R2) A solvency level should be calculated by dividing the fund level by the liability level, in which the fund level is the market value of the scheme's assets and the liability level is the sum of cash equivalents, calculated on the same basis as for an individual transfer value, for non-pensioners (i.e. active and deferred members) and the cost of immediate annuities for pensioners;
(R3) A solvency band is introduced in the form of a closed interval [90%, 100%] of all solvency levels between a base level of 90% and the minimum solvency standard of 100%;
(R4) Any scheme which falls below the solvency band should restore the solvency level up to 90% by making an immediate injection of funds within three months (unless the proposed Occupational Pensions Regulator allows a longer period than three months); and
(R5) Any scheme which falls within the solvency band should set up a funding plan to restore the solvency level to 100% within three years.

Setting aside the risk of dishonest removal of the scheme's assets (including fraud, theft and some other misappropriation of the scheme's assets which can lead to the fund being in deficiency as, for example, in the Maxwell affair or the collapse of Barings), the MSR would reinforce greatly the security of defined benefit pension schemes, rather than that of money purchase pension schemes.

As mentioned earlier in section 1.1, in the operation of money purchase pension schemes, the members accept most of the risk of poor investment performance and hence this scheme can be thought of as being fully funded so that there would be no need to apply the MSR as a rule, if the other benefits (e.g. death benefits) are funded from an insurance company.

Consequently, the proposed MSR could be an effective and objective supervisory measure of the security of a funded pension scheme, in a particular defined benefit pension scheme. The basic idea of the MSR is based on a winding-up (or liquidation) valuation approach (although there are some deficiencies in terms of protecting all the members' accrued pension rights in comparison with the nature of a winding-up valuation basis, which will be discussed below).

3.2.2.1 Problems of MSR and their alternatives

Firstly, the minimum solvency requirement (MSR) would be unlikely to protect all the members' accrued pension rights in full against the potential risk of dishonest removal of the scheme's assets in cases where the sponsoring employer is insolvent and unable to restore lost assets. For this reason, the PLRC proposes to establish the compensation scheme (for achieving better protection for all the members of occupational pension schemes against the dishonest removal of the scheme assets) and the Pension Compensation Board (for considering the payment of compensation in cases where the sponsoring employer becomes insolvent as a result of the dishonest removal of the scheme's assets). This proposal is largely introduced in the Pension Act 1995 by changing the name of the Pensions Compensation Board, which will be responsible for all aspects of the compensation scheme as an administrative body. From the viewpoint of pension funding, the PLRC protection measures for all of the members' accrued pension rights of occupational pension schemes would be twofold: the introduction of the MSR and the establishment of the compensation scheme.

Secondly, the use of market value of the scheme assets proposed in (R2) suffers from the problem of being affected by temporary fluctuations in the prevailing investment market sectors (especially, in the equity market); in this respect, the proposed market value of the scheme's assets could be adjusted as in the concept 'market related fund level' introduced in subsection (ii) in section 2.1.2.6.

Thirdly, as recognised by the PLRC [see Pension Law Reform Volume I Report, section 4.4.42], the proposed cash equivalents would not be sufficient to ensure the continuing security of a defined benefit pension scheme (even in the case of actual winding-up), since the premium needed to buy-out the promised benefits of non-pensioners in the prevailing non-profit deferred annuity market could be greater than the proposed cash equivalents. An option for guaranteeing almost completely the promised benefits could be either to introduce some deliberately cautious margins into the actuarial basis used for calculating the cash equivalents (i.e. set out a more conservative basis than best estimate) or, in theory, to adapt widely the actuarial basis used for calculating the cash equivalents to the prevailing basis for the non-profit deferred annuity contracts offered by reputable pension providers.

Moreover, along with the proposed pensioner liabilities calculated as the cost of the appropriate immediate annuities, the cash equivalents would also place a great financial burden on the sponsoring employer and eventually the members' benefits could be scaled down. This is because a scheme's investment manager would feel compelled to change most of the portfolio of assets from equity and property-based to fixed-interest or index-linked investments (e.g. gilts) so as to avoid the volatility on short-term equity and property market conditions and then reduce the risk of an unexpected rise in contribution rates (including an immediate cash injection) in order to satisfy the statutory minimum solvency standard over a short period (i.e. at most within three and quarter years). Hence, this type of investment strategy would lose the long-term investment return on equity and property investments and the prospects of capital growth, given that equities and property produce generally higher long-term investment returns than gilts [see Thornton & Wilson (1992)].

Therefore, in terms of balancing the security of the promised benefits and the financial burden placed on the employer, one option could be that the actuarial basis for calculating the (non-pensioner and pensioner) liabilities should allow for an equity-based valuation of the liabilities and, in parallel, the proposed time limits for restoring solvency (i.e. at most three and quarter years) should be extended to some degree by reference to the typical amortisation period of the Spread method (e.g. 20~25 years) or the amortisation period of the Amortisation of losses method (e.g. 5~15 years) [see section 2.2.3]. However, this option would be likely to cause some criticisms in terms of weakening the security of the members' accrued pension rights in comparison with the original intentions of the MSR.

Lastly, we note that for the valuation of a large pension scheme, a run-off valuation approach would be appropriate (see section 2.1.2.4). However, as seen in (R1), the size of the scheme's assets does not matter in the MSR. So, it may be realistic to apply the MSR differentially according to the size of the scheme's assets, for example, for a small pension scheme the basis for calculating the solvency level would be a winding-up valuation, while for a large pension scheme the basis for calculating the solvency level would be a run-off valuation.

In conclusion, the focus of the funding plan for the MSR would be switched from a (long-term) going-concern valuation basis to a (short-term) winding-up valuation basis. As mentioned in section 2.1.2.6, this winding-up valuation basis could cause a general problem of mismatching between these short-term security requirements and the long-term funding plan. For this reason, although the accrued pension rights under a winding-up valuation basis would be better protected than those under a going-concern or run-off valuation basis, the security supervision using a winding-up valuation approach may place a significant financial burden on the employer. This kind of financial burden may make a further shift from the defined benefit pension scheme towards a money purchase pension scheme, as pointed out in Pension Law Reform Volume I Report (1993, section 4.4.43), "The disadvantage of the cash equivalents solution is that it changes the nature of the pension promise from earnings-related to money purchase." This move may not be acceptable to employees because defined benefit pension schemes were originally designed to provide a stable and adequate level of retirement income for the employees but the employees in money purchase pension schemes bear most of the risk of poor investment performance.

3.2.3 Minimum funding requirement (MFR)

As discussed in the above section 3.2.2.1, the MSR intrinsically contains several problems, particularly from the view of the sponsoring employer. For this reason, the MSR proposed has been adjusted mainly in the direction of balancing the security of the promised benefits and the financial burden placed on the employer. Here, we shall briefly explain three distinct balancing procedures, in relation to the government responses to the MSR.

The first government response to the PLRC recommendations was the White paper, 'Security, Equality, Choice: The Future for Pensions', published on 23 June 1994, in which the recommendations were received with some modification. As for the MSR, the recommendations (R1)~(R5) in section 3.2.2, were accepted in the White paper, except that the method of calculating the cash equivalents has been modified according to the proposal made by The

Institute and Faculty of Actuaries: that is, cash equivalents for younger non-pensioners would be calculated on the basis of equity investments, moving gradually to a gilt investment basis for those members approaching retirement (older non-pensioners) and for pensioners. Thus, the degree of balance between the security of all the members' accrued rights and the financial burden placed on the employer will depend on the actuarial basis finally established for the calculation of cash equivalents.

As a second response to the PLRC recommendations, the government introduced the Pensions Bill on 15 December 1994 in the House of Loads. After a period of further discussion since the publication of the White paper, the MSR was adjusted mainly to improve further the harmonisation between security and the employer's financial burden as follows:

- (allowance for equity-based pensioner liability for large schemes): a solvency level should be calculated by dividing the fund level by the liability level, in which the fund level is calculated on the market value of the scheme's assets averaged over a period of months and the liability level is calculated on the basis for the cash equivalents proposed in the White paper, except that for a large pension scheme (with funds over £100 million) about 25% of the pensioner liabilities can be valued by reference to an equity rate of return; and
- (extension to the time limits for restoring solvency): a solvency band [90%, 100%] is the same as the MSR, but any scheme which falls below the solvency band should return to the basic solvency level of 90% within one year and also any scheme which falls within the solvency band should set up a funding plan to regain the minimum solvency solvency standard of 100% within five years.

Finally, the Pensions Bill was enacted on 19 July 1995 by the Pension Act 1995, which leaves the details to regulations. One major feature of the Act is that a series of modifications to deal with the MSR related problems (discussed in section 3.3.2.1) have led to changing the name of the minimum solvency requirement (MSR, proposed by the PLRC) to the minimum funding requirement (MFR, enacted by the Pensions Act 1995) during its progress through the House of Lords. As we might envisage from the name change to the MFR, the structure of the MFR has a different emphasis from the MSR. The MFR can be summarised as follows:

(L1) The MFR applies to all occupational pension schemes which are not money purchase pension schemes or schemes to be prescribed in regulations;

(L2) The MFR requires the value of the scheme assets to be not less than the amount of the scheme liabilities (although the broad principles of how the scheme assets and liabilities are calculated will be set out in regulations); hence, as mentioned in section 3.3.1, the MFR will define a solvency level as that (solvency level) = (amount of the scheme liabilities assessed by regulations to be prescribed) / (value of the scheme assets assessed by regulations to be prescribed);

(L3) The funding plan (which is termed as a schedule of contributions in the Act) is designed to meet the MFR continuously throughout the period to be prescribed by regulations, or to meet it by the end of that period; and

(L4) The case that solvency level is less than 90% is considered as a serious underprovision and in this occasion the employer must secure an increase of the solvency level up to at least 90% before the end of a period to be prescribed by regulations, by making an appropriate payment to the trustees or managers, by a method to be prescribed or by contributions made before the end of that period.

Given (L3) and (L4), the solvency supervision of the MFR would be characterised by the solvency band [90%, 100%], as under the MSR, even though the regulations for the methods of valuing the assets and liabilities, the time limits for restoring (90% and 100%) solvency level, and some other necessary assumptions have not yet been phased in. In particular, how to set out the time limits for the restoration of solvency and the actuarial basis used for calculating the (non-pensioner and pensioner) liabilities for the MFR will be the most important issues in the prescribing regulations.

In addition, as a supervisory authority with overall responsibility for monitoring and supervising occupational pension schemes, the Occupational Pensions Regulator proposed by PLRC is enacted by the Occupational Pensions Regulatory Authority. This new body will take over most of the functions of the existing Occupational Pensions Board, with the power to remove, suspend and replace trustees and to order an occupational pension scheme to be wound up, under prescribed circumstances.

3.3 Review of studies on the actuarial basis for solvency valuations

Establishing an appropriate actuarial basis for the MSR or MFR is one of the most important concerns in the MSR or MFR structure, in particular setting the scheme economic parameters which will determine the valuation interest rate and will be concerned with estimating the cost of the liabilities (e.g. rate of price inflation, rates of return on equities and gilts and salary growth rate).

The basic idea is that funding conservatively on a going-concern valuation improves the security of the accrued pension rights mainly through higher contribution rates than would arise from a best estimate basis. This leads then to the question of what level of margins in the actuarial basis would be appropriate for the purpose of the MSR or MFR (i.e. solvency valuations).

At this point, it is worth reviewing two distinct approaches to setting the realistic actuarial basis for solvency valuations in order to provide a degree of security even in the case of the scheme being wound-up, Thornton & Wilson (1992) and Collins (1992):

(i) Using the continuous-time equation of equilibrium introduced in section 3.1.8, Thornton & Wilson (1992) investigate the effect on pension funds of the investment rates of return deviating from the assumed rate of return. One of the most interesting results is that in order to reduce the frequency of insolvency on a winding-up valuation basis, a minimum 20% margin between the winding-up valuation and the going-concern valuation is suggested as a desirable liability solvency margin, that is, {amount of actuarial liability on the winding-up valuation} $\geq$ 1.20 $\cdot$ {amount of actuarial liability on the going-concern valuation}. This suggestion will be used in modelling the solvency level growth equation [see section 4.1].

(ii) Employing various economic scenarios in the actuarial basis (with respect to the rate of price inflation, equity dividend growth rate and rate of return on gilts, and salary growth rate), Collins (1992) shows numerically that in order to reduce the frequency of insolvency and cash injection for solvency, firstly the assumed equity dividend growth rate should be less than the assumed rate of price inflation, secondly funding should be on a more conservative actuarial basis than best estimate and lastly the proportion of gilt-based investment should be greater than that of equity-based investment. Following his results, the actuarial basis for solvency valuations would be constructed in a manner that each economic scheme parameter includes some margins for caution in the actuarial basis (which provides a conservative actuarial basis). Also, he shows that simply increasing the target solvency level (in excess of 100%) would not reasonably address the problem of maintaining solvency because an additional cash injection is inevitably required both to fund and maintain the higher target level. However, we note that in a mature scheme once a higher funding level is achieved, investment returns on the higher level of assets would reduce the funding burden on the sponsoring employer. Finally, he ends his paper with a clear conclusion expressed that "The higher the contributions paid the lower the number of insolvencies." Thus, Collins' conclusions would suggest that setting a statutory minimum solvency or funding standard of 100% is appropriate and setting a conservative actuarial basis for solvency valuations plays a vital role in easing the mismatching problem between short-term solvency requirement (i.e. security) and long-term funding requirement (i.e. stability) and also in balancing somewhat the security of the promised benefits and the financial burden placed on the employer.

In conclusion, we can follow two distinct approaches in setting the actuarial basis for solvency valuations. The first is that suggested by Thornton & Wilson (1992) - we may introduce a liability solvency margin to be charged 'in aggregate' on the actuarial liability as assessed on a going-concern valuation. The second follows Collins (1992) - we may introduce, into the actuarial basis used in the going-concern valuation, a deliberate margin for the solvency valuation in the choice of each scheme parameter (simply, we shall call this the parameter solvency margin). The relation between these two approaches, involving respectively the liability and parameter solvency margins, will be discussed in relation to modelling the solvency growth equation [see section 3.4.1].

3.4 Mathematical model for solvency level

Throughout this section, we are concerned with the solvency level for the MSR (proposed by the PLRC, see section 3.2.2). In the same line of section 3.1, we shall here focus on building up a mathematical model of the solvency level on a discrete-time domain.

3.4.1 Solvency level growth equation

The (non-pensioner and pensioner) liabilities for the MSR (simply, called the solvency liability, denoted by SL) is defined as the sum of the cash equivalents for non-pensioners and the cost of buying-out immediate annuities for pensioners from a competitive pension provider [see section 3.2.2].

Here, SL is considered as an appropriate volume measure arising from a solvency valuation (i.e. winding-up valuation in the MSR), while AL (i.e. actuarial liability) is used as an appropriate scheme volume measure from a classical actuarial valuation (i.e. going-concern valuation). The mathematical relationship between SL and AL will be specified below. Also, the fund growth equation (3.4) is assumed to reflect appropriately the effect of the MSR on the investment policy.

Now, we define the solvency level (or solvency ratio) at time t (denoted by SR_t) as $SR_t \equiv F_t/SL_t$, in which there is no loss of generality in taking SL_t as non-zero for all $t \in \{0, 1, 2, \ldots\}$.

By dividing both sides of the fund growth equation (3.4) with SL_{t+1}, we obtain the following linear first-order difference equation of SR such that for each valuation date t = 0, 1, 2, ...,

$$F_{t+1} / SL_{t+1} = (1 + i_{t+1}) \cdot [SL_t / SL_{t+1}] \cdot [F_t / SL_t + C_t / SL_t - B_t / SL_t]$$

$$\Leftrightarrow SR_{t+1} = (1 + i_{t+1}) \cdot [SL_t / SL_{t+1}] \cdot [SR_t + CR_t - BR_t] \qquad \text{--- (3.14)}$$

where

SR_t ($\equiv F_t/SL_t$) $\equiv$ solvency level at time t, with given initial condition $SR_0 = F_0/SL_0$,

CR_t ($\equiv C_t/SL_t$) $\equiv$ contribution ratio at time t and

BR_t ($\equiv B_t/SL_t$) $\equiv$ benefit ratio at time t

; we shall call this equation the solvency-level growth equation, which will be modified further in the following sections 3.4.3 and 3.4.4.

In practice, there may be a difference between the actuarial liability AL and the solvency liability SL. The amount of SL, the sum of the transfer values and the cost of buying-out immediate annuities, will be often higher than that of AL. It will be affected by the short-term volatility in the (transfer and immediate annuity) market conditions; in this respect, the amount of AL could be viewed as the value which smoothes out the short-term fluctuations in the amounts of SL. Then, following the view of Thornton & Wilson (1992) introduced in section 3.3, the relationship between SL and AL would be expressible in the following general form:

$$SL_t = (1+m_t) \cdot AL_t, \quad \text{for } t = 0, 1, 2, \ldots \qquad \text{--- (3.15)}$$

where $AL_t \neq 0$ and $m_t \neq -1$ in the line of the assumption $SL_t \neq 0$; m_t is specified by the convex-combination (interpreted as a weighted average) of immediate annuity market cost adjustment and transfer market cost adjustment assessed at time t, that is, m_t = [(AL_t for pensioners) / AL_t] · [immediate annuity market cost adjustment at time t] + [(AL_t for non-pensioners) / AL_t] · [transfer market cost adjustment at time t]; in this respect, we shall describe m_t as the market cost adjustment at time t, which is corresponding to the liability solvency margin at time t introduced in section 3.3) and can be interpreted as the proportionate experience shortfall at time t (relative to AL_t), i.e. $(SL_t - AL_t) / AL_t$.

The above equation (3.15) implies that the gap between the amounts of AL_t and SL_t, attributable intrinsically to the difference between the actuarial basis for AL valuation and the actuarial basis for SL valuation, is specified by the market cost adjustment m_t. Of course, unless carrying out the actual winding-up procedure used in the Courts, the true value of the members' accrued pension rights can not be determined because the true value of their actual liabilities arises only from a negotiation between receivers and liquidators (see section 2.1.2.7). Thus, assuming that AL_t is well presented by the chosen funding method (e.g. Projected Unit method, Current Unit method, see section 2.2.2.2) and the actuarial assumptions for the purposes of SSAP 24 (i.e. best estimated basis, see section 2.1.2.2), SL_t in equation (3.15) is designed to express more closely the true value of the actual liabilities by way of employing m_t and also to indicate that the solvency liability is distinguishable from the actuarial liability.

As mentioned in section 3.3, the need for a realistic and objective actuarial basis for solvency valuations would appear to be paramount. The following arguments are essential:

(i) Like the true value of actual liabilities through the actuarial valuation process, the exact value of m_t is unlikely to be available. The sequence of market cost adjustments $\{m_0, m_1, m_2, \ldots\}$ to be generated through the actuarial valuation process would be derived most likely by reference to prevailing quotations by competitive pension providers with respect to the actual winding-up of pension schemes. In a practical aspect, we note that the suggestion of Thornton & Wilson (1992) (introduced in section 3.3) that $SL_t = (1 + m) \cdot AL_t$ with the liability solvency margin $m \geq 20\%$ may be applicable, where in theory, the value of m has to be an estimate that minimises $Var(m_t)$ subject to $E(m_t) = m$ over a long-time historical market trends (as in a similar manner to setting the actuarial basis for SSAP 24 purposes), or may be determined in the light of balancing, within any prescribed regulations, the security of the promised benefits against the financial burden placed on the employer. Of course, the process for determining the value of m needs to be repeated regularly in the light of updated market experience and new regulations in order to modify/fine-tune m according to whether or not the currently adopted m is appropriate for ensuring that the scheme would remain satisfactorily funded.

(ii) Or alternatively, as discussed in section 3.3, the actuarial basis for SL valuation will be set out in such a manner that introduces some deliberate margins for the solvency valuation in each of the scheme parameters, particularly the scheme economic parameters, so that the resulting basis would be more conservative than the best estimates for SSAP 24 purposes.

(iii) The relation between the market cost adjustment and the conservative basis is as follows: if AL_t is assessed on a more conservative basis than best estimate, then the market rate m_t may be close to zero (i.e. the gap between the amounts of AL and SL could be closer to zero). In theory, the best decision rule for reducing the gap between AL_t and SL_t (i.e. keeping $m_t \approx 0$) would be to introduce the deferred and immediate annuity basis used by competitive pension providers into the actuarial basis.

In conclusion, the above arguments (ii) and (iii) lead ultimately to the problem of the extent to which it may be appropriate to introduce some margins for solvency valuation in the actuarial basis. Even though we may determine separately a suitable degree of margin in respect of each of the scheme parameters, the estimation of these margins would be extremely complicated because most of the scheme economic parameters are highly correlated to each other. The model (3.15) collectively summarises a range of scheme parameter margins for the solvency valuation, by means of the market cost adjustment (i.e. liability solvency margin) m_t. Thus, m_t can be thought of as a time-varying exogenous parameter, which will be determined at each valuation date t by the scheme actuary by reference to the prevailing quotations by reputable pension providers, the imposed regulations, the interests of the trustees and employer, etc.

3.4.2 Assumptions

We are going to make some assumptions that shall be used to build our discrete-time model of solvency level including both the growth in the number of members and the growth in salaries (note that most aspects have already been introduced separately in sections 2.2.2.1 and 2.1.4.1, except for (A8)):

(A1) Valuations are carried out annually; hence, all descriptions are based on a discrete-time approach;

(A2) The trust deeds and rules specify only final salary retirement benefits for age and service; hence, the members existing in the scheme fall into active and retired members;

(A3) All new entrants join at age a, where $a > 0$;

(A4) Retirement is fixed at age r;

(A5) A suitable multiple decrement table (the service table) is constructed to represent correctly the survivorship of the members existing in the scheme, which is specified by the time-invariant survivorship function l_y, where $y = a, a+1, \ldots$;

(A6) The radix of the function l_y is chosen so that l_a represents the number of new entrants in the scheme at age a at time 0;

(A7) The valuation interest rate is fixed and constant for each unit time period (denoted by i_v);

(A8) The number of new entrants and level of salaries each grows geometrically over time t (note that geometric growth is normally used as the discrete-time analogue of the exponential growth in a continuous-time domain but the geometric growth is convertible to the exponential growth subject to some assumptions as shown in section 2.1.4.1; in this respect, we prefer to start with the concept of geometric growth); and

(A9) Assumptions (A1) ~ (A8) have been applied for a sufficiently long time that the distribution of each age in the scheme during the intervaluation period (t, t+1) becomes stable.

The above assumptions enable us to establish the following dynamic growth functions of order 1, for all $t \in \{0, 1, 2, \ldots\}$ and current age $y \in \{a, a+1, a+2, \ldots\}$.

Firstly, the following descriptions have already been given in details in section 2.1.4.1. Here, we shall briefly restate some parts involved in the development of our model.

The general model of the growth in the number of members during the intervaluation period (t, t+1) is to be measured in terms of a dynamic membership growth function (denoted by f_1) satisfying the following recurrence relationship. Since f_1 is dependent on the time of entry into the scheme (denoted by z, which is t+a-y) and l_a plays the role of the radix, then we have

$f_1(z+1) = (1+im_{z+1}) \cdot f_1(z)$ with the given initial condition $f_1(0)=1$

where im_{z+1} denotes an annual % growth rate of the number of members aged y which is defined as the real growth in f_1 during the intervaluation period (t, t+1), i.e. $[f_1(z+1)-f_1(z)]/f_1(z)$.

Considering a stable membership (i.e. a geometric growth in membership), we have $(1+im_{z+1}) = 1+im$, constant for all z, and $Prob[im > -1] = 1$, and then

$f_1(z+1) = (1+im) \cdot f_1(z)$ with the given initial condition $f_1(0)=1$ --- (3.16)

; this indicates that the size of the total covered membership grows geometrically if $im > 0$ or decay geometrically if $-1 < im < 0$, in particular $im = 0$ implies that the size of the total covered membership is constant over time t (i.e. a stationary membership as a special case of a stable membership, see section 2.1.4.1).

Secondly, the general model of the growth in the annual salaries is also represented by a year-of-experience salary growth function (denoted by f_2) satisfying the following recurrence relationship; since f_2 is independent of the age of an individual in the scheme, we have

$f_2(t+1) = [(1+iw_{t+1})] \cdot f_2(t)$ with the given initial condition $f_2(0) = 1$

where iw_{t+1} = annual % growth rate which is defined as the growth in f_2 during the intervaluation period (t, t+1), i.e. $[f_2(t+1) - f_2(t)] / f_2(t)$, defined in a manner consistent with the valuation interest rate (i.e. in nominal or real terms).

Considering a geometric growth in salaries, we have $(1+iw_{t+1}) = 1+iw$, constant for all t, and $Prob[iw > -1] = 1$, and then

$$f_2(t+1) = (1+iw) \cdot f_2(t) \text{ with the given initial condition } f_2(0) = 1 \qquad \text{--- (3.17)}$$

; this indicates that the active members` salaries are growing geometrically (if $iw > 0$), decaying geometrically (if $-1 < iw < 0$) or constant (if $iw = 0$).

Lastly, using the above two equations (3.16) and (3.17), the growth in AL during the intervaluation period (t, t+1) is expressed in the interaction of the stable membership growth increment and the salary growth increment so that the actuarial liability growth equation (3.1) introduced earlier in section 3.1.1 can be rewritten (see the proof given below)

$$AL_{t+1} = (1+i_v) \cdot (AL_t + NC_t - EB_t)$$

$$= [(1+im) \cdot (1+iw)] \cdot AL_t \text{ with the given initial condition } AL_0 = b \cdot W_0 \qquad \text{--- (3.18)}$$

; here, $(1+im)\cdot(1+iw)$ is the incremental factor of the scheme volume AL during (t, t+1) on a going-concern valuation.

The proof of equation (3.18) is as follows:

Letting two consecutive values of the total annual pensionable payroll as W_t and W_{t+1} assessed at time t and t+1 respectively, then

$$W_t = \sum_{x=a}^{r-1} f_1(t+a-x) \cdot f_2(t) \cdot l_x \cdot w_x \quad \text{and hence}$$

$$W_{t+1} = \sum_{x=a}^{r-1} f_1(t+1+a-x) \cdot f_2(t+1) \cdot l_x \cdot w_x$$

$$= (1+iw_{t+1}) \cdot \{ \sum_{x=a}^{r-1} (1+im_{t+1+a-x}) \cdot f_1(t+a-x) \cdot f_2(t) \cdot l_x \cdot w_x \}$$

where, w_x = annual pensionable salary for each member aged x at time 0, defined in a manner consistent with the valuation interest rate (i.e. in nominal or real terms).

Since $im_{t+1+a-x} = im$ for all t+1+a-x from equation (3.16) and $iw_{t+1} = iw$ for all t from equation (3.17), we obtain

$W_{t+1} = [(1 + im) \cdot (1 + iw)] \cdot W(t)$ with the initial condition

$$W_0 = \sum_{x=a}^{r-1} f_1(a-x) \cdot l_x \cdot w_x$$

And also, AL_t can be expressed as $AL_t = b \cdot W_t$, where b is a known constant parameter.

Finally, we have the equation (3.18) as required, that is,

$AL_{t+1} = b \cdot W(t+1) = b \cdot [(1+im) \cdot (1+iw)] \cdot W_t = [(1+im) \cdot (1+iw_{t+1})] \cdot AL_t$. QED

3.4.3 Modified solvency-level growth equation

Utilising equations (3.14), (3.15) and (3.18), we obtain the following linear, time-varying dynamic system model for solvency level: for all $t \in \{0, 1, 2, \ldots\}$,

$SR_{t+1} = (1+ir_{t+1}) \cdot (SR_t + CR_t - BR_t)$ with the initial condition SR_0 specified --- (3.19)

where $(1+ir_{t+1}) = (1+i_{t+1}) / (SL_{t+1}/SL_t) = [(1+i_{t+1})\cdot(1+m_t)] / [(1+im)\cdot(1+iw)\cdot(1+m_{t+1})]$

; this equation shall be called the modified solvency-level growth equation on the ground that this equation is simply derived in a modified form of equation (3.14) from equations (3.15) and (3.18).

Remark 3.3: (a) From the viewpoint of control theory, we can make the same interpretations about the modified solvency growth equation (3.19) as about the fund growth equation (2.2) introduced in section 2.3.2.1; in particular, the modified solvency-level growth equation (3.19) specifies the controlled object. SR_t is the output from the controlled object (i.e. controlled variable which is also called the state variable in the complete state information case or the conceptual state variable in the incomplete state information case) and $\{CR_t, BR_t, ir_{t+1}\}$ are inputs to the controlled object (i.e. controlling variables, in which from the viewpoint of the actuary, only CR_t should be a controlling variable; we note also that ir_{t+1} is a system parameter) [for more details, see section 2.4.2]; and

(b) As in sections 3.1.2 and 3.1.4, we also easily derive the net increase in SR such that $NSR(t, t+1) \equiv \nabla SR_{t+1} = (1+ir_{t+1}) \cdot NSR(t-1, t) + (1+ir_{t+1}) \cdot \nabla (CR_t - BR_t)$.

Lastly, there is no loss of generality in assuming that $Prob[(1+i_{t+1}), (1+m_t) > 0 \text{ for all } t] = 1$, so the above equation (3.15) is simply rewritten as follows:

$$SR_{t-1} = \exp[\phi_{t+1}] \cdot [SR_t + CR_t - BR_t] \qquad \text{--- (3.20)}$$

where $\phi_{t+1} = [\delta_{t+1} - (\tau_{t+1} - \tau_t) - \alpha - \beta]$, force of combined interest corresponding to $(1+ir_{t+1})$;

δ_{t+1} = force of interest corresponding to $(1+i_{t+1})$;

$\tau_{t+1}-\tau_t$ = difference between the forces of market cost growth corresponding to $(1+m_{t+1})$ and $(1+m_t)$, respectively;

α = force of membership growth corresponding to $(1+im)$;

β = force of salary growth corresponding to $(1+iw)$; with both of these forces assumed to be constant over the period $(t, t+1)$ for all t.

This equation (3.20) is often more convenient than equation (3.19), particularly in the analysis of the effect of each system factor on the controlled variable SR.

3.4.4 Modified solvency-level growth equation with given spread funding formula

Let us now develop the modified solvency-level growth equation (3.20) where the form of recommended contribution rate is specified by the spread funding formula modified for a solvency valuation - winding-up valuation in the MSR, as illustrated below.

To begin with, it is worth recalling that the spread funding formula (2.1) is defined as $C_t = NC_t + k_t \cdot (AL_t - F_t)$, $k_t \in \{k_t: 0 \leq k_t \leq 1\}$, and further, if $k_t = k$ constant for all t, then formula (2.1) reduces to $C_t = NC_t + k \cdot (AL_t - F_t)$, $k \in \{k: d_v \leq k \leq 1 \text{ with } i_v > 0\}$ [see, section 2.2.4.1].

Since formula (2.1) is appropriately designed for the classical actuarial valuation (i.e. long-term, going-concern valuation), it needs to be modified for a solvency valuation. Then, if the fund level F_t is appropriately assessed on the solvency valuation basis for all t (as assumed at the early stage of section 3.4.1), we can define the following solvency-version of formula (2.1), since for all t, AL_t for going-concern valuations corresponds to SL_t for solvency valuations: for all t,

$$C_t = NC_t + k_t \cdot (SL_t - F_t), \quad k_t \in \{k_t: 0 \leq k_t \leq 1\}, \qquad \text{--- (3.21)}$$

and further, if $k_t = k$ constant for all t in formula (3.21), then we have

$$C_t = NC_t + k \cdot (SL_t - F_t), \quad k \in \{k: d_v \leq k \leq 1 \text{ with } i_v > 0\}. \qquad \text{--- (3.21)'}$$

Moreover, simply dividing both sides of formula (3.21) with SL_t and letting $NR_t = NC_t/SL_t$, then we have

$$CR_t = NR_t - k_t \cdot (SR_t - 1), \quad k_t \in \{k_t: 0 \leq k_t \leq 1\}, \qquad \text{--- (3.22)}$$

and further, if $k_t = k$ constant for all t in formula (3.22), then we have

$CR_t = NR_t - k \cdot (SR_t - 1), \ k \in \{k: d_v \le k \le 1 \text{ with } i_v > 0\}$. --- (3.22)'

We note that the above spread funding formulae (2.1), (3.21) and (3.22) are commonly based on the Spread method - spreading evenly the undesirable valuation outcomes, such as unfunded liability (i.e. AL_t-F_t) in (2.1), insolvent liability (i.e. SL_t-F_t) in (3.21) and insolvent ratio (i.e. 1-SR_t) in (3.22), over a projected period. In this respect, formula (3.21) and (3.22) each can also be called the spread funding formula but these are especially designed for the solvency valuation (which is distinguishable from the spread funding formula (2.1)). Here, we concentrate on the spread funding formula (3.22).

Now, we develop the modified solvency-level growth equation (3.20) subject to the spread funding formula (3.22). The assumptions and principal equations for deriving the resulting equation (3.25) are as follows:

Assuming that (a) all actuarial assumptions including the assumptions (A1)~(A9) introduced in section 4.3.2, are exactly realised over time, except for the investment rates of return; (b) η = force of valuation interest = $\ln(1+i_v)$ with $Pro[i_v > -1] = 1$; and (c) the spread funding formula (3.22) is adapted to the modified solvency-level equation (3.20), then we have the following principal equations (3.23), (3.24) and (3.25), from which we derive the resulting equation (3.26). These are

[From assumption (a)]: for all t, $B_t = EB_t$ (so, $BR_t = EBR_t \equiv EB_t/SL_t$). --- (3.23)

[From equations (3.1), (3.15) and (3.23) and assumption (b)]: for all t,

$SL_{t+1} = \exp(\tau_{t+1}+\eta) \cdot [\exp(-\tau_t) \cdot SL_t + NC_t - EB_t]$ --- (3.24)

; this equation shall be called the solvency liability growth equation in a similar manner to the actuarial liability growth equation (3.1).

[By dividing the both sides of equation (3.24) with SL_t and using equation (3.18)]: for all t,

$\exp(\alpha+\beta-\tau_t) = \exp(\eta-\tau_t) + \exp(\eta)\cdot(NR_t - EBR_t)$. --- (3.25)

Finally, applying equation (3.25) to equation (3.20) and using assumption (c) produce the resulting equation (3.26): that is, for all $t \in \{0, 1, 2, \ldots\}$,

$$[SR_{t+1} - 1] = [\exp(\delta_{t+1} - (\tau_{t+1} - \tau_t) - \alpha - \beta) \cdot (1-k_t)] \cdot [SR_t - 1] +$$
$$[\exp(\delta_{t+1} - (\tau_{t+1} - \tau_t) - \alpha - \beta)] \cdot [1 - \exp(-\tau_t)] +$$
$$[\exp(\delta_{t+1} - \tau_{t+1} - \eta) - 1], \text{ with the initial condition } SR_0 - 1 \text{ specified} \quad \text{--- (3.26)}$$

where $k_t \in \{k_t: 0 \leq k_t \leq 1\}$.

Here, we shall call the resulting equation (3.26) as the zero-input, 100%-target solvency-level growth equation, except for $k_t=1$, for the reason that the term 'zero-input' means that the controlling variable does not appear in equation (3.26) and the term '100%-target solvency-level' implies that $[SR_t-1]$ can be considered as a new state variable.

Remark 3.4: (a) From the viewpoint of control theory, the solvency-level spread funding formula (3.22) is a linear static (not dynamic) function of the state variable SR_t-1, which is made proportional to the negative state variable by means of k_t (i.e. $-k_t\cdot(SR_t-1)$). For this reason, the spread parameter k_t can be thought of as the proportional negative state-feedback controlling parameter, (except for the case of $k_t=0$). Also, following Haberman (1994)'s economic interpretation (introduced in Remark (b) in section 2.2.4.1), the spread parameter can be interpreted as a penal rate of interest that is being charged on the insolvent ratio (i.e. $1-SR_t$), except for $k_t=0$;

(b) From equation (3.15) (i.e. $AL_t/SL_t=\exp(-\tau_t)$) and equation (3.18) (i.e. $AL_{t+1}/AL_t=\exp(\alpha+\beta)$) , then the solvency liability growth equation (3.24) reduces to

$\exp(-\tau_t) \cdot [1 - \exp(\alpha + \beta - \eta)] \cdot SL_t = EB_t - NC_t$

; for positive SL, this equation shows the necessity of a certain degree of capitalisation (i.e. a certain amount of funds reserved for the accrued benefits), that is, the essential nature of a funding policy (including pay-as-you-go financing policy); and hence,

(c) At each valuation date t,

- in Mature/Declining pension schemes where $\eta > \alpha + \beta$, the difference between EB_t and NC_t (i.e. $EB_t > NC_t$) may be subsidised by the earnings from investments, such as interest, rent and dividend earnings, or by a cash injection;

- in Young pension scheme where $\eta < \alpha + \beta$, the gap (i.e. $EB_t < NC_t$) will create additional funds for the future liabilities; and

- in the special case of $\eta = \alpha + \beta$, the above equation reduces to $EB_t = NC_t$ which implies that there is no funding for the accrued benefits, that is, we have a pay-as- you-go financing policy [for more details of Declining, Mature and Young pension schemes, see section 2.1.4].

3.4.5 Summary

As discussed in section 3.4.1, the actuarial basis for the solvency valuation can be established by two distinct approaches, introducing either the liability solvency margin or parameter solvency margins into the actuarial basis for SSAP 24 purposes (i.e. the long-term and best estimate basis). We have taken the view that the liability solvency margin m_t specified in equation (3.15) can cover the liability valuation effect of a range of parameter solvency margins and would be flexible and simply adjustable in response to changes in (transfer value and immediate annuity) market values, and this approach has been adopted in deriving two distinct solvency-level growth models (3.20) and (3.26).

When applying optimal control theory in order to find a dynamic pension funding plan, the model (3.20) will be used to specify the controlled object without any specific formulae for the controlling variable CR_t (- this will be considered in Chapter 4 dealing with the situation of a short-term, winding-up valuation), while the model (3.26) can be used for the controlled object when the controlling variable CR_t is governed by the solvency-level spread funding formula (3.22) but in our study, this model (3.26) will be modified later in section 5.1 for dealing with the situation of a long-term, going-concern valuation (- the modified model will be considered in Chapter 5). The rigorous procedures for establishing a dynamic pension funding plan in each case will be given in Chapters 4 and 5.

Appendix 3 (proof of the relationship between MUL and NUL described in p149):

If MUL(t) and NUL(t_0, t) are continuous functions of t for $t_0 \le t$, t_0 and $t \in [0, \infty)$ with NUL(t_0, t_0) = 0, and the principle of additive consistency in NUL holds, then for $t_0 \le t_1 \le t_2$,

$$\mathrm{NUL}(t_1, t_2) = \int_{t_1}^{t_2} \mathrm{MUL}(s)\,ds.$$

Proof. For $t_0 \le t$, we have

$$\begin{aligned}
\mathrm{MUL}(t) &= \lim_{h\to 0^+} \mathrm{NUL}(t, t+h) / h \quad \text{(by definition)} \\
&= \lim_{h\to 0^+} \{\mathrm{NUL}(t_0, t) + \mathrm{NUL}(t, t+h) - \mathrm{NUL}(t_0, t)\} / h \\
&= \lim_{h\to 0^+} \{\mathrm{NUL}(t_0, t+h) - \mathrm{NUL}(t_0, t)\} / h \quad \text{(by the principle of additive consistency)} \\
&= \lim_{h\to 0^+} \{g(t+h) - g(t)\} / h \quad \text{(by letting NUL}(t_0, t) \equiv g(t)\text{, continuous function of t)} \\
&= g'_+(t)
\end{aligned}$$

; hence, MUL(t) is differentiable at time t because we are assuming that g(t) and $g'_+(t)$ are all continuous functions of t and then $g'(t) = g'_+(t) = g'_-(t)$, where $g'(t)$ = derivative of g at time t, $g'_+(t)$ = righthand derivative of g at time t and $g'_-(t)$ = lefthand derivative of g at time t.

Therefore, we obtain $g(t) = \int_{t_0}^{t} MUL(s)ds$, since the initial condition $g(t_0) = 0$ --- (*)

By using the principle of additive consistency in NUL, we can then easily derive the following relationship between MUL and NUL: for $t_0 \leq t_1 \leq t_2$,

$NUL(t_1, t_2) = NUL(t_0, t_1) + NUL(t_1, t_2) - NUL(t_0, t_1)$

$= NUL(t_0, t_2) - NUL(t_0, t_1)$

$= g(t_2) - g(t_1)$

$= \int_{t_1}^{t_2} MUL(s)ds$ (from the above equation (*)). QED

Chapter 4 Dynamic pension funding plan with no given form of controlling variable

4.1 Introduction

Chapters 2 and 3 provide the principal bases for the investigation of this chapter: broadly, Chapter 2 provides the conceptual framework for dynamic pension funding plans and Chapter 3 provides the structural models for the controlled object of defined benefit pension schemes. Thus,

- We have discussed various aspects associated with pension funding plans in Chapter 2. In particular, we have explained our contentions that our funding purpose (i.e. control goal) is designed to balance simultaneously the conflicting interests between the trustees and sponsoring employer of a defined benefit pension scheme, which allows simultaneous minimisation of the contribution rate and solvency risks over a defined projection period (for a summary, see section 2.1.3.4) and produces an optimal feedback control law by employing optimal control theory from the field of engineering, in which the resulting optimal feedback control law is defined as our dynamic pension funding plan (see section 2.3).

- As a fundamental framework for specifying the financial structure of the controlled object, we have constructed several growth equations in Chapter 3, such as the basic growth equations (see section 3.1) and the solvency level growth equations (see section 3.4). These equations are considered to represent a reduced structural model of the real financial structure of defined benefit pension schemes.

This chapter is a major part of our research, in which we are concerned with deriving and presenting the dynamic pension funding plan by solving the (deterministic and stochastic) control optimisation problem (which we have set up, subject to the form of controlling variable being not specified) over a finite control time horizon by means of the method of dynamic programming. Our controlled object is assumed to be specified and governed by the modified solvency level growth equation (3.20) (established earlier in section 3.4.3).

In section 4.2, we will construct the principal structure of the control optimisation problem (both deterministic and stochastic). Further, the fundamental principles for solving the established problem will be illustrated. In section 4.3, we will define a dynamic pension funding plan under a deterministic control problem, while the stochastic control problem will be explored in section 4.4.

4.2 The principal structure of control optimisation problems

4.2.1 Classification of pension funding control system variables

As a preliminary to building up mathematically our control problems, it will be worthwhile modifying our previous discussions in section 2.3 in the light of the newly defined controlled object (governed by the modified solvency level growth equation (3.20)). By simply redefining the various variables involved in modelling pension funding control system, we can have a similar discussion to that of section 2.3. Considering time $t \in [0, T-1]$, these are

(a) Information vector (i.e. all current information available to the actuary at the time of taking the control action, composed of both past control history and current valuation result).

Here, the information vector of the complete state information case, including the deterministic case, is given by $H_t = (SR_0, SR_1, \ldots, SR_t, CR_0, CR_1, \ldots, CR_{t-1})$ with the given initial information $H_0 = SR_0$ and the information vector of the incomplete state information case by $\Im_t = (SR_{-1}, SR_0, \ldots, SR_{t-1}, CR_{-1}, CR_0, \ldots, CR_{t-1})$ with the given initial information $\Im_0 = (SR_{-1}, CR_{-1})$. The difference between H_t and $\Im_t$ is identified by the measurement equation $M_t = SR_{t-b}$, $b = 0$ or 1, described below in (f);

(b) Controlling variables (i.e. inputs $\{CR_t, BR_t, \phi_{t+1}\}$ to the controlled object).

In particular, only CR_t is a controlling variable to be manipulated in the view of the actuary and ϕ_{t+1} a system parameter where if either ϕ_{t+1} or BR_t contains some (random) disturbances, the model (3.20) specifies the stochastic controlled object. Otherwise the model (3.20) specifies the deterministic controlled object (simply, denoted by Controlled object $^{D \text{ or } S}$, in which the superscript D or S indicates the deterministic or stochastic controlled object, respectively). Here, we are concerned with feedback control (not feedforward control) using the optimal control theory of dynamic programming, which implies that our control law is restricted by the causality principle (i.e. present controlling inputs should not depends on future controlled outputs) and other practical requirements. Hence, it is assumed that the admissible controlling variable is expressed as a function of the information vector but the form of the function is not specified, i.e. $CR_t = \pi_t(H_t)$ in the deterministic control optimisation problem and $CR_t = {}^C\pi_t(H_t)$ or ${}^I\pi_t(\Im_t)$ in the stochastic control optimisation problem, in which the superscript C or I indicates the complete or incomplete state information case, respectively.

Therefore, our purpose is to determine the function form to produce whatever will be the optimal performance of a given Controlled object $^{D \text{ or } S}$ (denoted by $\pi^*_t(.)$, ${}^C\pi^*_t(.)$ or ${}^I\pi^*_t(.)$, which each corresponds to our optimal control law at time t, see sections 4.3.2, 4.4.1.3 and 4.4.2.3);

(c) Controlled variable (i.e. output SR_t from the controlled object).

In particular, SR_t is called the state variable in the complete state information case or the conceptual state variable in the incomplete state information case [see Remark 2.6 (a) in section 2.3.2.2]. For convenience, the state space of SR_t is assumed to be some countable set S;

(d) Targets (or Reference variables) (i.e. desired level of CR_t, denoted by crt_t, and desired level of SR_t, denoted by srt_t).

These can be thought of as exogenous variables providing the external reference information (e.g. the statutory minimum solvency standard of 100% for the MSR, see section 3.2.2) determined mainly by the actuary;

(e) Observable/Realisable valuation variable (i.e. output M_t from the actuary's valuation process specified by the measurement equation specified by $M_t = SR_{t-b}$, b = 0 or 1).

Here, the complete state information case is specified by $M_t = SR_t$ with the given initial condition $M_0 = SR_0$ and the incomplete state information case by $M_t = SR_{t-1}$ with the given initial condition $M_0 = SR_{-1}$;

(f) Most effective point estimator of the conceptual state variable SR_t (in the incomplete state information case) in the light of memory efficiency and estimation (denoted by $\hat{SR}_t$).

Thus the actuary at time t does not have direct access to the current value of SR_t and he is required to estimate effectively the current value of SR_t from the available information at time t, $\Im_t$. As a result, $\hat{SR}_t = E(SR_t \mid \Im_t)$ is the most effective point estimator of the unknown current state SR_t (which will be discussed in section 4.4.2.1); and

(g) Linear dynamic system.

The pension funding control system for solvency valuation is called a linear dynamic system because the controlled object is specified by the first-order difference growth equation (3.20) and both this growth equation and the measurement equation given in (e) are all linear with respect to SR_t and CR_t.

Therefore, our control mechanism can be summarised as illustrated below in Figure 4.1 and 4.2. We note that there are no differences in comparison with Figures 2.3 and 2.4 (illustrated in section 2.3.2.2), except for the redefinition of the control system variables. Hence, the detailed discussions related to Figures 2.3 and 2.4 are applicable here.

Controlled object $^{D \text{ or } S}$ ⟶ $\{SR_t\}$ ⟶ Measurement process

$\{CR_t\}$↑ ↓

$\{crt_t, srt_t\}$ → Optimal feedback control law $\{\pi_t^*(.)$ or $^C\pi^*_t(.)\}$ ← {Feedback: $M_t=SR_t$}

Figure 4.1 Optimal pension funding feedback control system for the complete state information case (i.e. $M_t = SR_t$).

The above "Optimal feedback control law" is a dynamic pension funding plan for the complete state information case, designed using control optimisation. The rigorous procedure for finding the dynamic pension funding plan will be given separately in section 4.3 for the deterministic case and in section 4.4 for the stochastic case.

Controlled object ^{S} ⟶ $\{SR_t\}$ ⟶ Measurement process

$\{CR_t\}$↑ {Feedback: $M_t = SR_{t-1}$} ↓

↑ → L (Time delay) → {Feedback: CR_{t-1}} →┐ ↓

$\{crt_t, srt_t\}$ → Optimal feedback control law $\{^I\pi^*_t(.)\}$ ← $\{\hat{}SR_t\}$ ← Estimator of SR_t

Figure 4.2 Optimal pension funding feedback control system for the incomplete state information case (i.e. $M_t = SR_{t-1}$); $\hat{}SR_t = E(SR_t \mid \Im_t)$.

The above "Optimal feedback control law" is a dynamic pension funding plan for the incomplete state information case, designed using control optimisation. The combination of "Optimal feedback control law" plus "Estimator of SR_t" is usually called the optimal feedback controller, which is distinguishable from the control problem with complete state information - this optimality structure is assured by the Separation Theorem which will be explained in subsection (iii) of section 4.2.5. The detailed procedure for establishing a dynamic pension funding plan will be given in section 4.4

4.2.2 Performance index

In order to complete our previous discussion about suitable performance indices in section 2.3.2.2, we shall describe their mathematical definitions on a discrete-time domain. We note that as discussed in section 2.3.2.2, the performance indices for control errors and control action errors conceptually correspond to the solvency risk and contribution rate risk, respectively. The principal performance indices would be defined as follows:

(a) $PI_1 \equiv$ {sum of square of control error over a finite control horizon [0, T], $0<T<\infty$}

$$= \sum_{t=0}^{T} (SR_t - srt_t)^2 \text{ in the deterministic case, or}$$

$$= E\{\sum_{t=0}^{T} (SR_t - srt_t)^2\} \text{ in the stochastic case}$$

; this index correctly corresponds to the solvency risk over [0, T] (see section 2.1.3.3), and is designed to penalise more severely larger control errors;

(b) $PI_2 \equiv$ {sum of absolute value of control error over [0, T], $0 < T < \infty$}

$$= \sum_{t=0}^{T} |SR_t - srt_t| \text{ in the deterministic case, or}$$

$$= E\{\sum_{t=0}^{T} |SR_t - srt_t|\} \text{ in the stochastic case;}$$

(c) $PI_3 \equiv$ {sum of the product of time and square of control error over [0, T], $0<T<\infty$}

$$= \sum_{t=0}^{T} t \cdot (SR_t - srt_t)^2 \text{ in the deterministic case, or}$$

$$= E\{\sum_{t=0}^{T} t \cdot (SR_t - srt_t)^2\} \text{ in the stochastic case}$$

; this index is designated to reduce the effect of the large initial control error as well as put an emphasis on the later control errors;

(d) $PI_4 \equiv$ {sum of square of control action error over [0, T], 0 <T<∞}

$$= \sum_{t=0}^{T} (CR_t - crt_t)^2 \quad \text{in the deterministic case, or}$$

$$= E\{\sum_{t=0}^{T} (CR_t - crt_t)^2\} \quad \text{in the stochastic case}$$

; this index correctly corresponds to the contribution rate risk over [0, T] (see section 2.1.3.2), and is designed to penalise more severely larger control action errors; hence,

(e) $PI_5 \equiv$ {general form of the performance sum over [0, T], 0 < T < ∞}

$$= \sum_{t=0}^{T} f(SR_t - srt_t,\ CR_t - crt_t,\ t) \quad \text{in the deterministic case, or}$$

$$= E\{\sum_{t=0}^{T} f(SR_t - srt_t,\ CR_t - crt_t,\ t)\} \quad \text{in the stochastic case}$$

where $f(.,\ .,\ .)$ is a function of the control error (SR_t-srt_t), control action error (CR_t-crt_t) and time t, so we can establish a variety of performance indices based on various combination of these variables, subject to a specific purpose of measuring the quantitative performance of a control system.

Now, we shall define our performance index (denoted by PI_θ) suitable for our funding purpose (i.e. searching for a pension funding plan for balancing the solvency and contribution rate risks at the same time), which is a specific form of PI_5.

$PI_\theta \equiv$ {discounted weighted-average of the solvency risk and contribution rate risk

over [0, T-1], 0 < T < ∞, plus discounted terminal cost at the final time T}

$$= \sum_{t=0}^{T-1} \{e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2]\} + e^{-\eta T}\cdot(SR_T - srt_T)^2 \qquad \text{--- (4.1)}$$

in the deterministic case, or

$$= E\left\{\sum_{t=0}^{T-1} \{e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2]\} + e^{-\eta T}\cdot(SR_T - srt_T)^2\right\} \quad \text{--- (4.2)}$$

in the stochastic case

where θ is a weighting parameter which balances the importance between solvency risk and contribution rate risk over [0, T-1], assuming $0 < \theta < 1$ in view of our funding purpose, and $e^{-\eta t}$ is the performance discount function at time t in which η is the valuation force of interest.

We note that the deterministic PI_θ can be interpreted as the net present value (NPV) of the project's future costs over a fixed projection period [0, T] and the stochastic PI_θ as the expected NPV of the project's future costs over [0, T] [see section 2.2.1]. Here, the term $e^{-\eta T} \cdot (SR_T - srt_T)^2$ represents the NPV of the solvency risk associated with the final state SR_T at the final time T, which will play the role of a boundary condition. This terminal cost is introduced to estimate the terminal behaviour/response of control system to the final control action CR_{T-1} (which will be clarified in the following sections 4.3 and 4.4).

Lastly, in the special case that $srt_t = E(SR_t)$ and $crt_t = E(CR_t)$, then the solvency risk at time t is defined as $Var(SR_t)$ and the contribution rate risk at time t as $Var(CR_t)$, so the stochastic PI_θ would become

$$PI_\theta = \sum_{t=0}^{T-1} \{e^{-\eta t}\cdot[\theta\cdot Var(SR_t) + (1-\theta)\cdot Var(CR_t)\} + e^{-\eta T}\cdot Var(SR_T),$$

which gives one reason why a squared-error performance index, such as PI_1, PI_3, PI_4 or PI_θ, is often considered in evaluating decision processes.

4.2.3 Optimal performance criterion

The performance criterion for realising our funding purpose introduced in section 2.1.3.4 is termed the optimal performance criterion, which can fall into two groups, primary and supplementary.

4.2.3.1 Primary performance criterion

From the viewpoint of the realisation of our funding purpose, our performance criterion can be expressed as follows: for a given θ,

$$\min_{\{CR_t;\ t=0,1,\ldots,T-1\}} PI_\theta \qquad \text{--- (4.3)}$$

; in particular, this criterion shall be called the primary performance criterion.

The actuary could choose the value of θ on a subjective basis; for example, if he acts only in the best interests of the employer, he will recommend a value of θ close to zero (which will lead to reducing only the contribution rate risk), but in contrast if he acts only in the best interests of the trustees, he will recommend a value of θ close to one (which will lead to reducing only the solvency risk).

In practical terms, the value of θ would be affected mainly by some compulsory requirements imposed by the supervising authorities (e.g. the minimum funding requirement (MFR) contained in the Pensions Act 1995, see section 3.2.3), the conflicting interests of the trustees and employer, and the prospects for the future demographic and financial status of the sponsoring company. These factors would exert influence directly on the pace of funding and the progress of solvency level through θ.

However, the primary performance criterion (4.3) does not address how the value of θ should be appropriately determined. In this respect, the actuary needs to supplement this weakness of the primary performance criterion (4.3) with a new decision criterion for the appropriate choice between all possible values of θ. This subject is considered in the next section.

4.2.3.2 Supplementary performance criterion

(i) Preliminary:

Using the primary criterion (4.3), the actuary can obtain the optimal control action at time t (denoted by CR'_t) and accordingly the optimal control response corresponding to CR'_t (denoted by SR'_{t+1}), for an arbitrary value of θ chosen on his subjective basis (here, these resulting control action and response shall be called the pre-optimal control action CR'_t and response SR'_{t+1}, respectively). Both CR'_t and SR'_{t+1} will then be a function dependent on θ, so θ can be thought of as a parameter that needs to be appropriately controlled by the actuary.

For this reason, he would be required to suggest some additional criteria for determining objectively the value of θ. We suggest the use of one of the three distinct supplementary performance criteria (to the primary performance criterion (4.3)), i.e. Suggestions I, II and III described below, for the completion of the optimal performance criterion: of course, he could construct other suitable criteria. This section is intended to give some insights into determining the value of θ from the actuary's point of view.

(ii) Suggestion I:

The actuary may employ the following supplementary performance criterion designed for a recommendable value of θ:

$$\underset{\theta}{\text{Min}} \ \{ \sum_{t=0}^{T-1} [\lambda_1 \cdot (SR'_t - srt_t)^2 + (1-\lambda_1) \cdot (CR'_t - crt_t)^2] + \lambda_1 \cdot (SR'_T - srt_T)^2 \} \qquad \text{--- (4.4.1)}$$

where $0 < \lambda_1 < 1$ (generally, we may fix at $\lambda_1 = ½$).

Here, λ_1 can be regarded as a weighting parameter to adjust the relative importance between the pre-optimal solvency and contribution rate risks at time t, i.e. $(SR'_t - srt_t)^2$ and $(CR'_t - crt_t)^2$, and further λ_1 may be thought of a semi-parameter of θ for the reason that λ_1 can not influence directly these risks but can exert influence indirectly through θ. We note that this criterion (4.4.1) does not allow for discounting as in the primary criterion (4.1) on the grounds that SR'_t and CR'_t are produced by the primary criterion (4.1) and then each of them represents the pre-optimal value at time t discounted to time 0.

This criterion (4.4.1) can then be thought of as the pre-optimal version of the primary performance criterion (4.3), particularly in the case of $\lambda_1 = \theta$, since this criterion is designed to minimise the pre-optimal contribution rate and solvency risks that will be result from the pre-optimal control action, while the primary performance criterion is designed for defining the pre-optimal control action.

However, Suggestion I may cause a severe funding burden on the employer in view of the fact that this criterion does not take into account the smoothness of the pre-optimal sequences, $\{CR'_t;\ t=0, 1, \ldots, T-1\}$ and $\{SR'_t:\ t=0, 1, \ldots, T\}$. A further problem in employing this criterion would be setting an appropriate value for λ_1.

(iii) Suggestion II:

Taking into account the initially given SR_0, the prospects for the future (demographic and economic) status of the sponsoring company and so on, the trustees and employer each would have their own sensible views on the pace of future funding such that the relative percentage growth of pre-optimal contribution ratios (relative to their original value), i.e. $\nabla CR'_{t+1}/CR'_t$, is bounded by their own acceptable limits: that is, for each time t, $trgc_{min} \leq \nabla CR'_{t+1}/CR'_t \leq trgc_{max}$ in view of the trustees and $ergc_{min} \leq \nabla CR'_{t+1}/CR'_t \leq ergc_{max}$ in view of the employer, say.

On the other hand, the trustees and employer each would also have their own sensible views on the progress of future solvency levels such that the relative percentage growth of pre-optimal solvency levels (relative to their original value), i.e. $\nabla SR'_{t+1}/SR'_t$, is bounded by their own acceptable limits: that is, for each time t, $trgs_{min} \leq \nabla SR'_{t+1}/SR'_t \leq trgs_{max}$ in view of the trustees and $ergs_{min} \leq \nabla SR'_{t+1}/SR'_t \leq ergs_{max}$ in view of the employer, say.

Further, there is no loss of generality in assuming that for every t, $\{trgc_{min} \leq \nabla CR'_{t+1}/CR'_t \leq trgc_{max}\} \cap \{ergc_{min} \leq \nabla CR'_{t+1}/CR'_t \leq ergc_{max}\} \neq \emptyset$ and $\{trgs_{min} \leq \nabla SR'_{t+1}/SR'_t \leq trgs_{max}\} \cap \{ergs_{min} \leq \nabla SR'_{t+1}/SR'_t \leq ergs_{max}\} \neq \emptyset$.

From the above inequality equations, the actuary can first derive such a common region of the trustees and employer's views that $rgc_{min} \leq \nabla CR'_{t+1}/CR'_t \leq rgc_{max}$ for the pace of funding and $rgs_{min} \leq \nabla SR'_{t+1}/SR'_t \leq rgs_{max}$ for the progress of solvency levels, where $rgc_{min} = \text{Min}\ \{trgc_{min}, ergc_{min}\}$, $rgc_{max} = \text{Min}\ \{trgc_{max}, ergc_{max}\}$, $rgs_{min} = \text{Min}\ \{trgs_{min}, ergs_{min}\}$ and $rgs_{max} = \text{Min}\ \{trgs_{max}, ergs_{max}\}$. Next, for each common region he could suggest some balancing values (denoted by rgc and rgs), which each should be a value acceptable to both the trustees and the employer: for example, $rgc = (rgc_{min} + rgc_{max})/2$ and $rgs = (rgs_{min} + rgs_{max})/2$.

By using the balancing values 'rgc' and 'rgs' as the target values of the pre-optimal relative percentage growth sequences, i.e. $\{\nabla CR'_{t+1}/CR'_t;\ t=0, 1, \ldots, T-2\}$ and $\{\nabla SR'_{t+1}/SR'_t;\ t=0, 1, \ldots, T-1\}$, he may construct the following supplementary performance criterion:

$$\underset{\theta}{\text{Min}}\ \{ \sum_{t=0}^{T-2} [\lambda_2 \cdot (\nabla SR'_{t+1}/SR'_t - rgs)^2 + (1-\lambda_2) \cdot (\nabla CR'_{t+1}/CR'_t - rgc)^2] + \lambda_2 \cdot (\nabla SR'_T/SR'_{T-1} - rgs)^2 \} \quad \text{--- (4.4.2)}$$

where $0 < \lambda_2 < 1$ (generally, we may fix at $\lambda_2 = ½$).

As for λ_1, λ_2 can be regarded as a weighting parameter to adjust the relative importance between the squared pre-optimal relative percentage growth errors, $(\nabla SR'_{t+1}/SR'_t - rgs)^2$ and $(\nabla CR'_{t+1}/CR'_t - rgc)^2$ and further λ_2 can be thought of as a semi-parameter of θ for the same reason as in λ_1. We note that the success of using this criterion would depend on how its target values, rgs and rgc, are appropriately chosen. As mentioned above, these values should be determined objectively by the actuary, in consultation with the trustees and employer. A further problem in employing this criterion would be choosing a sensible value of λ_2.

(iv) Suggestion III:

As another possible alternative, the actuary may employ the following supplementary performance criterion using differences: for either n = 1 or 2,

$$\underset{\theta}{\text{Min}} \left\{ \sum_{t=0}^{T-n-1} [\lambda_3 \cdot (\nabla^n SR'_{t+n})^2 + (1-\lambda_3) \cdot (\nabla^n CR'_{t+n})^2] + \lambda_3 \cdot (\nabla^n SR'_T)^2 \right\} \quad \text{--- (4.4.3)}$$

where $0 < \lambda_3 < 1$ (generally, we may fix at $\lambda_3 = \frac{1}{2}$).

Hence, this criterion (4.4.3) is designated to maximise the smoothness of the pre-optimal sequences, $\{CR'_0, CR'_1, \ldots, CR'_{T-1}\}$ and $\{SR'_0, SR'_1, \ldots, SR_T\}$, that is, '$\lambda_3 \to 0$' implies that the actuary places great emphasis on stabilising the pace of funding, but in contrast '$\lambda_3 \to 1$' implies that he has more interest in stabilising the progress of solvency levels. In particular, the criterion with n=1 can be thought of as an extension of the minimum energy control approach in optimal control theory [see Benjamin (1989) introduced in section 2.3.2.4 for an early application to pension funding] and the criterion with n=2 is consistent with the approach to optimising maximum smoothness because $\sum [\lambda_3 \cdot (\nabla^2 SR'_{t+2})^2 + (1-\lambda_3) \cdot (\nabla^2 CR'_{t+2})^2] + \lambda_3 \cdot (\nabla^n SR'_T)^2$ can be thought of as a measure of smoothness from graduation theory [see London (1985)].

However, Suggestion III could have a potential risk for underestimating the pre-optimal errors, $CR'_t - crt_t$ and $SR'_t - srt_t$. For example, there may be a situation where the sequence of either CR'_t or SR'_t deviate relatively far from their respective target values, crt_t or srt_t, during a long period, in comparison with the above Suggestions I and II. A further problem would be, of course, determining an appropriate value of λ_3.

(v) Conclusion:

The above criteria (4.4.1)~(4.4.3) have been suggested as additional means for improving the performance of the primary criterion (4.4) in the light of the pace of funding and/or the progress of solvency level. Of course, we can not say that any of these criteria is the unique and best supplementary performance criterion for determining appropriately the value of θ. Further, these criteria are unlikely to deliver their respective optimal values of θ based on numerical iterations for all possible combinations (θ, λ), $0 < \theta$, $\lambda = \lambda_1$, λ_2 or $\lambda_3 < 1$, because of the possible uncountable number of computational iterations being involved. Moreover, it is not easy to derive the optimal value of θ in a mathematical form by solving these supplementary performance criteria because the pre-optimal quantities CR'_t and SR'_t each have a complicated functional form dependent on θ (- this will become clear later in sections 4.3.2, 4.4.13 and 4.4.2.3), but on the other hand we can find the optimal value of θ by means of numerical analysis, subject to some specified finite admissible combinations of (θ, λ).

We recognise that it would be important to deal with the primary performance criterion (4.3) in association with a supplementary performance criterion formulated in a desirable direction of improving both the pace of funding (as a best interest of the employer) and the progress of solvency levels (as a best interest of the trustees) in relation to their respective target values (e.g. a combination form of the merits inherent in the above criteria (4.4.1)~(4.4.3)). However, this formulation is likely to involve new parameters and the diagnosis of its appropriateness would rely on numerical and/or mathematical performance comparisons between all suitable types of supplementary performance criteria. Owing to constraints of time and space, we propose leaving this subject to future research as a possible extension to this thesis.

For simplicity, we do not take into account any mathematical supplementary performance criteria in our control optimisation problems formulated later in section 4.2.4. Alternatively, the employer and trustees in consultation with the actuary are assumed to determine the best (not necessary optimal) value of θ values (denoted by θ^*) by reference to the predictable forward error projections $\{CR'_t-crt_t;\ t=0, 1, \ldots, T-1\}$ and $\{SR'_t-srt_t;\ t=0, 1, \ldots, T\}$, subject to a specified finite admissible set of θ [see sections 4.3.3.3 and 4.4.3.3]. Hence, the value θ^* would be chosen at the level balancing the conflicting interests of the employer and trustees (although this decision can not avoid the criticism of subjectivity). For convenience, this mechanism for deciding θ^* shall be called the θ^*-criterion.

4.2.4 Control optimisation problems

In this section, we construct mathematically our control optimisation problems on a discrete-time domain, which are characterised as linear dynamic systems with quadratic performance criteria (known by the nomenclature 'LQP optimisation problems'). In other words, the controlled object with the controlling variable CR_t and the controlled variable SR_t is governed by the modified solvency level growth equation (3.20),

$SR_{t+1} = \exp[\phi_{t+1}] \cdot [SR_t + CR_t - BR_t]$, where $\phi_{t+1} = [\delta_{t+1} - (\tau_{t+1} - \tau_t) - \alpha - \beta]$,

and the actuary's valuation process provides at each valuation date t an observation of the form $M_t = SR_{t-b}$ where b = 0 or 1: these are all linear with respect to CR_t and SR_t. Further, the optimal performance criterion, both primary and supplementary, is already established in the form of the quadratic performance criteria (4.3) and (4.4), respectively.

As outlined in section 2.3, the LQP optimisation problems in discrete-time are generally classified into

- deterministic control optimisation problem over [0, T] for T>0
- stochastic control optimisation problem over [0, T] for T>0
    - complete state information case
    - incomplete state information case

(i) Deterministic control optimisation problem over [0, T]:

The squared deviations between SR_t and srt_t and between CR_t and crt_t are completely known for all $t \in [0, T]$, where the sequence $\{SR_t; t \in [0, T]\}$ is generated with certainty by the modified solvency level growth equation (3.20) with the given initial SR_0.

Our dynamic pension funding plan would be defined to generate the sequence of optimal control actions $\{CR^*_t; t \in [0, T-1]\}$ satisfying the following deterministic (pension funding) control optimisation problem over [0, T]:

$$\underset{\{CR_t;\, t=0,1,\ldots,T-1\}}{\text{Min}} \left\{ \sum_{t=0}^{T-1} e^{-\eta t} \cdot [\theta \cdot (SR_t - srt_t)^2 + (1-\theta) \cdot (CR_t - crt_t)^2] + e^{-\eta T} \cdot (SR_T - srt_T)^2 \right\}$$

subject to $SR_{t+1} = \exp[\phi_{t+1}] \cdot [SR_t + CR_t - BR_t]$ with given SR_0, $t \in [0, T-1]$ and

θ^* determined by the θ^*-criterion.

--- (4.5)

The procedure for solving the control problem (4.5) will be considered in section 4.3 and some illustrative numerical examples will be given in section 4.3.3.

(ii) Stochastic control optimisation problem over [0, T]:

The above deterministic LQP optimisation problem can only make limited progress in overcoming the effects of external economic and demographic uncertainty, which can disturb the financial status of pension schemes. A satisfactory theory of control problems needs therefore to recognise the random nature of influences like investment rate of return and changes to the membership of the pension scheme.

Here, uncertainty in control problems is assumed to be characterised by the uncertainty about the values of the unknown investment rate of return δ_{t+1} in ϕ_{t+1} and the unknown benefit outgo

B_t in BR_t, which are separately specified below according to whether or not the available state information is complete.

(a) Stochastic control optimisation problem over [0, T] with complete state information:

The stochastic processes, $\{\delta_{t+1};\ t\in[0, T-1]\}$ and $\{B_t;\ t\in[0, T-1]\}$, each are simply assumed to be made up of the following two components:

$$\delta_{t+1} = \eta + {}^{a}\varepsilon_{t+1} \quad \text{and} \quad B_t = EB_t + {}^{b}\varepsilon_{t+1} \qquad \text{- - - (4.6)}$$

where

${}^{a}\varepsilon_{t+1}$ = unpredictable disturbance which follows an independent and identically distributed (IID) $N(0, \sigma_a^2)$ distribution with $\sigma_a^2 < \infty$), defined on each unit control period [t, t+1);

${}^{b}\varepsilon_{t+1}$ = unpredictable disturbance which follows IID $N(0, \sigma_b^2)$ distribution with $\sigma_b^2 < \infty$, defined on each unit control period [t, t+1); and

${}^{a}\varepsilon_{t+1}$ and ${}^{b}\varepsilon_{s+1}$ are mutually independent for all $t, s \in [0, T-1]$.

Thus, this model implies that the (long-term) force of valuation interest is correctly determined in view of the expected value of the actual forces of interest corresponding to the investment return (i.e. $E(\delta_{t+1}) = \eta$ for all t) and the expected benefit outgo is also well designed in view of the expected value of the actual benefit outgo (i.e. $E(B_t) = EB_t$ for all t).

We note that $\{{}^{a}\varepsilon_{t+1};\ t\in[0, T-1]\}$ and $\{{}^{b}\varepsilon_{t+1};\ t\in[0, T-1]\}$ each is a (zero-mean) Gaussian white noise process (simply, (zero-mean) bivariate Gaussian white noise process with the correlation between ${}^{a}\varepsilon_{t+1}$ and ${}^{b}\varepsilon_{t+1}$ being zero, $\{({}^{a}\varepsilon_{t+1}, {}^{b}\varepsilon_{t+1});\ t\in[0, T-1]\}$) with the property of strong stationarity [see Harvey (1990, section 1.5)]. Further, the available information vector at time t is $H_t \equiv (SR_0, SR_1, \ldots, SR_t, CR_0, \ldots, CR_{t-1})$ with the given initial information $H_0 = SR_0$.

From the stochastic model (4.6) specified for each $t\in[0, T-1]$, we have the stochastic controlled object governed by the following stochastic solvency level growth equation: for all $t\in[0, T-1]$,

$$SR_{t+1} = \exp[\phi_{t+1}] \cdot [SR_t + CR_t - BR_t] \text{ with given } SR_0 \qquad \text{- - - (4.7)}$$

where

$\phi_{t+1} = [\eta-(\tau_{t+1}-\tau_t)-\alpha-\beta] + {}^a\varepsilon_{t+1} \equiv \mu_{t+1} + {}^a\varepsilon_{t+1} \sim \text{IID } N(\mu_{t+1}, \sigma_a^2)$ with $\sigma_a^2 < \infty$;

$BR_t = B_t/SL_t = EB_t/SL_t + {}^b\varepsilon_{t+1}/SL_t \equiv EBR_t + {}^b\varepsilon_{t+1}/SL_t \sim \text{IID } N(EBR_t, VBR_t)$ with $VBR_t = \sigma_b^2/SL_t^2 < \infty$; and, ϕ_{t+1} and BR_s are mutually independent for all t, $s\in[0, T-1]$, and SR_0 is independent of ϕ_{t+1} and BR_t.

Our stochastic (pension funding) control optimisation problem over [0, T] with complete state information can be written in the form:

$$\min_{\{CR_t;\, t=0,1,\ldots,T-1\}} E\{\sum_{t=0}^{T-1} \{e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2]\} + e^{-\eta T}\cdot[\theta\cdot(SR_T - srt_T)^2]\}$$

subject to the stochastic controlled object governed by model (4.7) and θ^* determined by the θ^*-criterion.

- - - (4.8)

The procedure for solving the control problem (4.8) will be considered in section 4.4.1 and some illustrative numerical examples will be given in section 4.4.4.

(b) Stochastic control optimisation problem over [0, T] with incomplete state information:

Now, we consider the incomplete state information version of the problem constructed in (a), which involves the measurement equation $M_t = SR_{t-1}$ with one-unit time delay in the state information and then the information available to the actuary at time t is $\Im_t = (M_0, M_1, \ldots, M_t, CR_{-1}, CR_0, \ldots, CR_{t-1})$ with given initial information $\Im_0 = (M_0, CR_{-1})$, and furthermore we assume that the stochastic solvency level growth equation (4.7) is also applicable at time -1 because of the one-unit-time delay in information.

Thus, our stochastic (pension funding) control optimisation problem over [0, T] with incomplete state information can be written in the form:

$$\min_{\{CR_t;\ t=0,1,...,T-1\}} E\{\sum_{t=0}^{T-1} \{e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2]\} + e^{-\eta T}\cdot[\theta\cdot(SR_T - srt_T)^2]\}$$

subject to the stochastic controlled object specified by model (4.7) being applicable for all $t\in[-1, T-1]$ (so, given initial information $\Im_0=(M_0, CR_{-1})$ is independent of ϕ_{t+1} and BR_t), $M_t = SR_{t-1}$ and θ^* being determined by the θ^*-criterion.

--- (4.9)

The procedure for solving the control problem (4.9) will be considered in section 4.4.2 and some illustrative numerical examples will be given in section 4.4.4 in connection with the numerical examples of the complete state information problem (4.8).

4.2.5 Principles for control optimisation

Prior to solving the LQP optimisation problems (4.5), (4.8) and (4.9) specified in section 4.2.4, it would be helpful to make some brief comments on three distinct principles related to their optimal solutions, Bellman's principle of optimality, the Certainty Equivalence Principle and the Separation Theorem. These principles occupy a central position in the development of optimal control theory, with applications particularly in the fields of engineering, economics and operations research.

(i) Bellman's principle of optimality:

Bellman (1957, p83) states the principle of optimality formally as "An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

Comparing with our control terminology, the term 'stage' corresponds to the unit control period (t, t+1), 'optimal policy' to the optimal control law and 'state resulting from the first decision' to the value of the state variable SR_t in the complete state information case or the value of the effective state estimator $\hat{SR}_t$ in the incomplete state information case.

This principle is concerned with the mathematical technique for multi-stage decision processes (i.e. sequences of decisions). Considering our control optimisation problems (4.5), (4.8) and (4.9), each of them can be thought of as a T-stage decision process in which the resulting sequence of T decisions $\{CR^*_t;\ t\in[0, T-1]\}$ are to be made such that the corresponding optimal performance criterion is realised by transforming each control problem from one of making T decisions simultaneously to one of making the decisions one at each stage but sequentially, so that we can overcome the algebraic difficulties in dealing with a multi-stage decision process.

The method of dynamic programming, developed by Bellman (1957, 1961) as a mathematical technique for multi-stage decision processes, originates from Bellman's principle of optimality because it provides a systematic procedure for making a sequence of interrelated decisions by decomposing control problems into multi-stage decision problems. This is the reason why we employ the dynamic programming method for our pension funding control optimisation problems, both deterministic and stochastic.

(ii) <u>Certainty Equivalence Principle</u>:

Although deterministic LQP optimisation problems are not too difficult to solve, stochastic LQP optimisation problems require, in general, more care. If this principle holds, it is not necessary to solve directly stochastic LQP optimisation problems, but enough to solve its deterministic version, obtained by replacing all random variables in the performance index by their expected values. This principle appears in various forms in many (but not all) stochastic LQP optimisation problems.

These arguments can be illustrated by the following two simple examples.

Firstly, consider the stochastic LQP optimisation problem over a one-unit control horizon such that $\underset{x}{\text{Min}}\ E_c\{(ay + bx + c)^2\}$ where a, b are given constants and E_c denotes mathematical expectation with respect to the random disturbance c. The optimal control solution is attained for $x^* = -(a/b)\cdot y - (1/b)\cdot E(c)$, which is equal to the optimal control solution (the so-called certainty equivalence solution) of the deterministic version of the above stochastic problem such that $\underset{x}{\text{Min}}\ \{[ay + bx + E(c)]^2\}$. Further, we can easily make a similar discussion even in the case of multiperiod LQP optimisation problems by way of Bellman's principle of optimality [see Bertsekas (1976, section 3.1)]. Thus, for stochastic LQP optimisation problems specified only by additive random disturbances the Certainty Equivalence Principle holds and the certainty equivalence solution is optimal for the original problem.

Next, consider the stochastic LQP optimisation problem over a one-unit control horizon such that $\underset{x}{\text{Min}}\ E_{a,b,c}\ \{(ay + bx +c)^2\}$ where $E_{a,b,c}$ denotes mathematical expectation with respect to random coefficients a, b and random disturbance c. Differentiating with respect to x yields the optimal control solution $x^* = -\ [E(ab)/E(b^2)]\cdot y\ -\ [E(bc)/E(b^2)]$. However, the deterministic version of the above stochastic problem is given in the form of $\underset{x}{\text{Min}}\ \{[E(a)\cdot y + E(b)\cdot x + E(c)]^2\}$ and hence the certainty equivalence solution is attained for $x^* = -\ [E(a)/E(b)]\cdot y\ -\ E(c)/E(b)$, which is not optimal for the original stochastic LQP optimisation problem with random coefficients a and b. We can make a similar discussion in the case of multiperiod LQP optimisation problems by way of Bellman's principle of optimality [see Bertsekas (1976, section 3.1)]. Thus, for the stochastic LQP optimisation problem with random coefficients, the Certainty Equivalence Principle does not hold, since if we replace the random coefficients with their expected values, then the resulting certainty equivalence solution is not optimal for the original problem [for a more rigorous example in the pension funding area, see Haberman & Sung (1994)]. So, for our stochastic LQP optimisation problems (4.8) and (4.9) this principle will not hold because of the random coefficient ϕ_{t+1}, which appears in sections 4.4 and 4.5, respectively.

Historically, the Certainty Equivalence Principle first introduced by Simon (1956) provides that the optimal control law can be designed without any consideration of stochastic effects whenever we deal with the stochastic LQP optimisation problems specified only by additive random disturbances.

(iii) Separation Theorem:

This principle is limited to stochastic LQP optimisation problems with incomplete state information. As earlier illustrated as Figure 4.2 (or Figure 2.4 in section 2.3.2.2), the optimal feedback controller can be separated into two parts. The first part is the state estimator (as an optimal filter) which produces, assuming no control action takes place, the most effective estimator (e.g. $\hat{SR}_t$ in Figure 4.2) of the conceptual state variable (e.g. SR_t in Figure 4.2) from the available information vector (e.g. $\Im_t$ in Figure 4.2), in the light of memory efficiency. The second part is the optimal feedback control law of the control problem which provides the control action (e.g. CR_t in Figure 4.2) as a linear function of the most effective estimator (fed directly into the optimal feedback control law), so that the optimal control law is independent of the accuracy of the estimation of the current conceptual state.

This interesting property, which shows that the two parts of the optimal feedback controller can be designed independently/separately as the current state estimation and optimal control solution, has been called the Separation Theorem [see Bertsekas (1976; section 4.3)].

Historically, the Separation Theorem first published by Joseph & Tou (1961) provides a connection between filtering theory and optimal stochastic control theory, as illustrated in Figure 4.2 (or Figure 2.4 in section 2.3.2.2).

4.3 Dynamic pension funding plan for deterministic LQP optimisation problem

In this section, we explore the optimal solution of the deterministic LQP optimisation problem (4.5) specified in subsection (i) in section 4.2.4. Firstly, the objective is to find a sequence of pre-optimal control actions $\{CR'_0, CR'_1, \ldots, CR'_{T-1}\}$ for solving the problem (4.5). Secondly, we determine the best (not necessary optimal) value of θ (i.e. θ^*) in accordance with the θ^*-criterion (mentioned in subsection (v) in section 4.2.3.2). Finally, we can then define a sequence of optimal control actions $\{CR^*_0, CR^*_1, \ldots, CR^*_{T-1}\}$.

4.3.1 A functional equation

Consider time $t \in [0, T-1]$. The whole available information vector at time t is given by $(SR_0, SR_1, \ldots, SR_t, CR_0, CR_1, \ldots, CR_{t-1})$, the target inputs at time t to the optimal feedback control law are $\{crt_t, srt_t\}$ and the dynamic system inputs at time t to the deterministic controlled object are $\{BR_t, \phi_{t+1}\}$[see Figure 4.1 of section 4.2.1]. So, the actuary is required to determine the control action CR_t.

<u>Remark 4.1</u>: In any practical situation, the control law (or decision function) at time t admissible to the actuary, i.e. $\pi_t(.)$ introduced in subsection (b) of section 4.2.1, is expressed as a linear function of the current dynamic state SR_t of the controlled object, for the reason that (a) SR_t contains all the information on the scheme's financial status at time t; (b) as the solvency growth process $\{SR_0, SR_1, \ldots, SR_T\}$ is governed by a first-order linear difference equation (3.22) specified in section 3.4.3, the sequence of future states $\{SR_{t+1}, SR_{t+2}, \ldots, SR_T\}$ can be represented by a function of SR_t only; and hence, (c) it is sufficient to choose CR_t as a linear function of SR_t, that is, the general form of feedback control, i.e. $CR_t = \pi_t(H_t)$, is reduced to $CR_t = \pi_t(SR_t)$ linear in SR_t, which is consistent with the causality principle and other

practical requirements. Therefore, our control law at time t is reduced to linear feedback from the current dynamic state SR_t, which would appear to be practical and reasonable.

Our performance index (4.1), specified in section 4.2.2, can then be written as a sum of two parts:

$$PI_\theta = \sum_{s=0}^{T-1} e^{-\eta s}\cdot[\theta\cdot(SR_s - srt_s)^2 + (1-\theta)\cdot(CR_s - crt_s)^2] + e^{-\eta T}\cdot(SR_T - srt_T)^2$$

$$= PIA_\theta + PIB_\theta \qquad \text{--- (4.10)}$$

where

$$PIA_\theta = \sum_{s=0}^{t-1} e^{-\eta s}\cdot[\theta\cdot(SR_s - srt_s)^2 + (1-\theta)\cdot(CR_s - crt_s)^2]; \text{ and}$$

$$PIB_\theta = \sum_{s=t}^{T-1} e^{-\eta s}\cdot[\theta\cdot(SR_s - srt_s)^2 + (1-\theta)\cdot(CR_s - crt_s)^2] + e^{-\eta T}\cdot(SR_T - srt_T)^2$$

Hence, PIA_θ is independent of the decisions to be made $\{CR_t, CR_{t+1}, \ldots, CR_{T-1}\}$, so to minimise PI_θ with respect to $\{CR_t, CR_{t+1}, \ldots, CR_{T-1}\}$ is equivalent to minimising PIB_θ (which we will refer back to in section 4.4).

To produce the backward recursion in time t, we define

$$V(SR_t, t) = \underset{\{CR_s;\, s=t,t+1,\ldots,T-1\}}{\text{Min}} \{\sum_{s=t}^{T-1} \{e^{-\eta s}\cdot[\theta\cdot(SR_s-srt_s)^2 + (1-\theta)\cdot(CR_s-crt_s)^2]\} + e^{-\eta T}\cdot(SR_T - srt_T)^2\}.$$

Applying the backward dynamic programming method (based on Bellman's principle of optimality) for sequential control optimisation, we have

$$V(SR_T, T) = e^{-\eta T}\cdot(SR_T - srt_T)^2 \quad \text{and}$$

$$V(SR_t, t) = \underset{CR_t}{Min} \{ e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2] \ +$$

$$\underset{\{CR_s;\ s=t+1,t+2,\ldots,T-1\}}{Min} \{ \sum_{s=t+1}^{T-1} \{e^{-\eta s}\cdot[\theta\cdot(SR_s-srt_s)^2+(1-\theta)\cdot(CR_s-crt_s)^2]\} + e^{-\eta T}\cdot(SR_T-srt_T)^2\}\}$$

$$= \underset{CR_t}{Min} \{ e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2] + V(SR_{t+1}, t+1) \} \qquad \text{--- (4.11)}$$

; this functional equation for V(., .) is called the Bellman equation, in which $V(SR_t, t)$ can be interpreted to be the minimal future cost discounted at time 0, independent of the control actions before time t, obeying the above recursion in time.

Hence, the terminal cost associated with the terminal state SR_T, $V(SR_T, T) = e^{-\eta T}\cdot(SR_T - srt_T)^2$, plays the role of a boundary condition for the Bellman equation (4.11).

4.3.2 Control optimisation

Following the fact that the control law at time t is a linear function of the current state variable SR_t [see Remark 4.1], the solution of the Bellman equation (4.11) with the boundary condition $V(SR_T, T) = e^{-\eta T}\cdot(SR_T - srt_T)^2$ is uniquely determined in the following quadratic form

$V(SR_t, t) = A_1(t)\cdot SR_t^2 + A_2(t)\cdot SR_t + A_3(t)$ with the boundary condition $A_1(T) = e^{-\eta T}$,

$A_2(T) = -\ 2e^{-\eta T}\cdot srt_T$ and $A_3(T) = e^{-\eta T}\cdot srt_T^2$, --- (4.12)

which can be verified using the mathematical induction argument below.

This form is identical with the form of the boundary condition (at t = T) and then proceeding by induction, we have

$V(SR_{t+1}, t+1) = A_1(t+1)\cdot SR_{t+1}^2 + A_2(t+1)\cdot SR_{t+1} + A_3(t+1)$ with $A_1(T) = e^{-\eta T}$, $A_2(T) = -\ 2e^{-\eta T}\cdot srt_T$ and $A_3(T) = e^{-\eta T}\cdot srt_T^2$.

Introducing the above trial solution into equation (4.11), we can rewrite the Bellman equation (4.11) in the form: for every $t \in [0, T-1]$,

$$V(SR_t, t) = \underset{CR_t}{Min} \ \{ G(CR_t, t) \}$$

where

$$G(CR_t, t) = [e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)] \cdot CR_t^2 + [-2crt_t\cdot e^{-\eta t}\cdot(1-\theta) +$$
$$2\exp(2\phi_{t+1})\cdot A_1(t+1)\cdot(SR_t - BR_t) + \exp(\phi_{t+1})\cdot A_2(t+1)] \cdot CR_t +$$
$$[e^{-\eta t}\cdot\theta\cdot(SR_t - srt_t)^2 + e^{-\eta t}\cdot(1-\theta)\cdot crt_t^2 + \exp(2\phi_{t+1})\cdot A_1(t+1)\cdot(SR_t - BR_t)^2 +$$
$$\exp(\phi_{t+1})\cdot A_2(t+1)\cdot(SR_t - BR_t) + A_3(t+1)].$$

It is sufficient that the coefficient of CR_t^2 in $G(CR_t, t)$ is positive for all $t \in [0, T-1]$ as a condition for a unique sequence of pre-optimal control actions, that is

$$A_1(t+1) > -[e^{-\eta t}\cdot(1-\theta) / \exp(2\phi_{t+1})] \text{ for all } t \in [0, T-1]. \qquad \text{--- (4.13)}$$

Using the fact that $G(CR_t, t)$ is a strictly convex function under condition (4.13), we obtain the pre-optimal control action at time t for all $t \in [0, T-1]$ with an arbitrary value of θ specified, given by

$$CR'_t = -[D_1(t;\theta) / D_3(t;\theta)] \cdot SR_t + [D_2(t;\theta) / D_3(t;\theta)]$$
$$= -[D_1(t;\theta) / D_3(t;\theta)] \cdot [SR_t - 1] + [D_2(t;\theta) - D_1(t;\theta)] / D_3(t;\theta)$$
$$\equiv \pi'_t(SR_t) \qquad \text{--- (4.14)}$$

where

$D_1(t;\theta) = \exp(2\phi_{t+1})\cdot A_1(t+1)$,

$D_2(t;\theta) = crt_t\cdot e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)\cdot BR_t - \exp(\phi_{t+1})\cdot A_2(t+1)/2$, and

$D_3(t;\theta) = e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)$

; in particular, the above equation shall be called the pre-optimal (linear feedback pension) funding formula and accordingly, $\pi^*_t(.)$ the pre-optimal (linear feedback pension funding) control law at time t.

For completion, substituting CR'_t into Min $G(CR_t, t)$ yields

$$V(SR_t, t) = A_1(t)\cdot SR_t^2 + A_2(t)\cdot SR_t + A_3(t)$$

$$= \{[e^{-2\eta t}\cdot\theta\cdot(1-\theta) + e^{-\eta t}\cdot\exp(2\phi_{t+1})\cdot A_1(t+1)] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)]\}\cdot SR_t^2 + \{[2e^{-\eta t}\cdot\exp(2\phi_{t+1})\cdot((1-\theta)\cdot(crt_t - BR_t) - \theta\cdot srt_t)\cdot A_1(t+1) + e^{-\eta t}\cdot\exp(\phi_{t+1})\cdot(1-\theta)\cdot A_2(t+1) - 2e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot srt_t] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)]\}\cdot SR_t + [\text{Remaining part}].$$

Here, we do not need to calculate the [Remaining part] in full because it does not affect our computation for CR'_t, i.e. CR'_t depends only on the functional coefficients $A_1(t+1)$ and $A_2(t+1)$, not on $A_3(t+1)$. So, we need only to solve the following backward recursive equations for all t:

$$A_1(t) = [e^{-2\eta t}\cdot\theta\cdot(1-\theta) + e^{-\eta t}\cdot\exp(2\phi_{t+1})\cdot A_1(t+1)] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)] \text{ and}$$

$$A_2(t) = [2e^{-\eta t}\cdot\exp(2\phi_{t+1})\cdot((1-\theta)\cdot(crt_t - BR_t) - \theta\cdot srt_t)\cdot A_1(t+1) + e^{-\eta t}\cdot\exp(\phi_{t+1})\cdot(1-\theta)\cdot A_2(t+1) - 2e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot srt_t] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\phi_{t+1})\cdot A_1(t+1)], \qquad \text{--- (4.15)}$$

which start with the boundary conditions, $A_1(T) = e^{-\eta T}$ and $A_2(T) = -\,2e^{-\eta T}\cdot srt_T$, and hence the first recursive equation can be solved to give $A_1(t)$ and then after substitution the second can be solved to give $A_2(t)$.

The above recursion for $A_1(.)$ generates sequentially the positive sequence $\{A_1(t);\ t\in[0, T-1]\}$ in the backward course of time t, starting with $A_1(T) = e^{-\eta T} > 0$; hence, the condition (4.13) for uniqueness is redundant. We have thus shown that the Bellman equation (4.11) has a unique solution of the suggested quadratic form (4.12) with $A_1(t)$ and $A_2(t)$ satisfying the above recursions (4.15) - here, we complete the mathematical induction argument.

It is worth noting that, from the fact that $A_1(t) > 0$ for all t implies that $0 < D_1(t;\theta) < D_3(t;\theta)$ for all t, the pre-optimal funding formula has a similar mathematical form to the (ratio-type) spread funding formula (3.22) given in section 3.4.4: that is, $D_1(t;\theta)/D_3(t;\theta)$ can be thought of as corresponding to the spread parameter k_t and $[D_2(t;\theta)-D_1(t;\theta)]/D_3(t;\theta)$ to NR_t.

Next, the pre-optimal control response SR'_{t+1} corresponding to CR'_t is given in the form: for all $t\in[0, T-1]$,

$$SR'_{t+1} = \exp[\phi_{t+1}] \cdot [SR'_t + CR'_t - BR_t] \text{ with given } SR_0 = SR'_0. \qquad \text{--- (4.16)}$$

After determining a best value of θ (i.e. θ^*) based on the θ^*-criterion, we can obtain a unique sequence of optimal control actions $\{CR^*_0, CR^*_1, \ldots, CR^*_{T-1}\}$ and corresponding optimal control responses $\{SR_0, SR^*_1, \ldots, SR^*_T\}$ with θ determined by θ^* in equations (4.14), (4.15) and (4.16).

In conclusion, our dynamic pension funding plan is defined as a sequence of functions $\{\pi^*_t(.);\ t\in[0, T-1]\}$ where $\pi^*_t(.)$ is defined by the equation with θ determined by θ^* in the pre-optimal funding formula (4.14), for each $t\in[0, T-1]$,

$$\pi^*_t(SR_t) = -[D_1(t;\theta^*) / D_3(t;\theta^*)] \cdot [SR_t - 1] + [D_2(t;\theta^*) - D_1(t;\theta^*)] / D_3(t;\theta^*) \qquad \text{--- (4.17)}$$

; in a similar manner to the pre-optimal funding formula (4.14) and control law $\pi'_t(.)$, this equation shall be called the optimal funding formula and function $\pi^*_t(.)$ the optimal control law at time t (and hence, our dynamic pension funding plan is completely governed by the optimal funding formula).

4.3.3 Numerical illustrations

The aim of this section is to illustrate numerically the relationship between the optimal control law $\{\pi^*_t(.); t\in[0, T-1]\}$ and its corresponding solvency level sequence $\{SR^*_t; t\in[0, T]\}$ with respect to the given initial solvency ratio SR_0, the control targets (crt_t, srt_t), the assumed valuation basis (particularly, the force of valuation interest η) and/or the weighting parameter value θ^*. The results obtained will provide a fundamental framework for the stochastic numerical illustrations to be investigated in section 4.4.4. Further, we compare mathematically the spread funding formula (3.22) with our optimal funding formula (4.17).

Although we can set a variety of scenarios, in particular, of those time-varying components which are not under the control of the actuary, such as δ_{t+1} and B_t, in order to test the effects on the controlling variable CR_t and controlled variable SR_{t+1}, we address the above objectives subject to the following simple assumptions. All numerical illustrations are given in Appendix 4A in tabular form, except for Graph 4.1.

4.3.3.1 Assumptions

The assumptions (A1) and (A2) are consistent with our mathematical modelling assumptions used in section 3.4: for all $t\in[0, T-1]$,

(A1) Actuarial assumptions:

(A1.1) Demographic assumptions

- stable membership (with members at age x dependent on English life table No. 14 (Males) and an exponential new entrant growth rate α per year with $\exp(\alpha) = 1.03$);

- Entry age = 25 (in particular, the life table value l_{25} to be used as the radix); and

- Retirement age = 65.

(A1.2) Economic assumptions (based primarily on the economic parameter estimates by Thornton & Wilson (1992))

- force of market cost adjustment for solvency valuations (τ_t), $\exp(\tau_t) = 1.2$ constant for all t and also $\exp(\tau_T) = 1.2$;

- force of salary growth (β), $\exp(\beta) = 1.02$;

- force of valuation interest (η), $\exp(\eta) = 1.06$; and

- benefits = retirement life annuity for age and service (1/60th of final pensionable salary per year of service).

(A2) All actuarial assumptions, both demographic and economic, are consistently realised by experience, except for investment returns.

(A3) Primary funding method for NC_t and AL_t: Projected Unit method.

(A4) Projection assumptions

- control period: T = 12;

- force of investment interest (δ_{t+1}), $\exp(\delta_{t+1}) = 1.06$ or 1.08 constant for all t;

- admissible weighting parameter set for θ: {90%, 50%, 10%};

- control targets: $(crt_t, srt_t) = (NR_t, 100\%)$ and $srt_T = 100\%$; and

- admissible initial solvency ratio: $SR_0 = 100\%$ or 0%.

Given the above assumptions (A1)~(A3), we can obtain the following formulae from equation (3.18) derived in section 3.4.2, for our numerical illustrations:

(a) $AL_{t+1} = \exp(\alpha+\beta) \cdot AL_t = 1.0506 \cdot AL_t$ with $AL_0 = 150.1738\%$ of initial payroll W_0;

(b) $NC_{t+1} = \exp(\alpha+\beta) \cdot NC_t = 1.0506 \cdot NC_t$ with $NC_0 = 5.0470\%$ of W_0;

(c) $EB_{t+1} = \exp(\alpha+\beta) \cdot EB_t = 1.0506 \cdot EB_t$ with $EB_0 = 6.3787\%$ of W_0; and hence

(d) equation (3.24) derived earlier in section 3.4.4, i.e. $\exp(\alpha+\beta-\tau_t) = \exp(\eta-\tau_t) + \exp(\eta)\cdot(NR_t - EBR_t)$, holds clearly.

; hence, any pair of AL_t, NC_t and EB_t is constant for all t.

Accordingly, EBR_t (= EB_t/SL_t) = 0.035396 and NR_t (= NC_t/SL_t) = 0.028006.

In particular, formula (d), i.e. $\exp(\alpha+\beta-\tau_t) = \exp(\eta-\tau_t) + \exp(\eta)\cdot(NR_t - EBR_t)$, enables us to compare more exactly the spread funding formula (3.22) with our pre-optimal funding formula (4.14) and this is considered in the next section 4.3.3.2.

Remark 4.2: It is worth recalling that our funding purpose is to balance long-term funding requirement (i.e. stability) and short-term solvency requirement (i.e. security) [see section 2.1.3.4 and subsection (ii) in section 3.3]: in this respect, the control targets specified in (A4) would be sensible to both the trustees and the employer. Further, the contribution ratio target NR_t is calculated by way of the Projected Unit method, which is appropriate because $\alpha > 0$ in (A1.1) [see subsection (ii) in section 2.2.2.2].

4.3.3.2 Dynamic pension funding plan vs. Spread funding plan

Here, we are concerned with the mathematical comparison between the pre-optimal funding formula (4.14) and the spread funding formula (3.22) derived in section 3.4.4. Applying the control target assumption in (A4) to formula (4.14) leads to the following simplified form, which is distinct from the spread funding formula (3.22): that is, for all $t \in [0, T-1]$,

$$\pi'_t(SR_t) = - [D_1(t;\theta)/D_3(t;\theta)] \cdot [SR_t - 1] + [D_2(t;\theta) - D_1(t;\theta)] / D_3(t;\theta)$$

$$= NR_t - \varphi(t;\theta)\cdot(SR_t - 1) + \xi(t;\theta), \text{ and hence}$$

$SR'_{t+1} = \exp(\phi_{t+1})\cdot[(1-\varphi(t;\theta))\cdot SR'_t + \exp(\alpha+\beta-\tau-\eta) - \exp(-\tau) + \varphi(t;\theta) + \xi(t;\theta)]$ with the initial condition $SR'_0 = SR_0$ (from equations (3.20) and (3.24)) --- (4.18)

where

$\varphi(t;\theta) = [D_1(t;\theta)/D_3(t;\theta)]$, in which $0 < \varphi(t;\theta) < 1$ for all $t\in[0, T-1]$; and

$\xi(t;\theta) = -[D_1(t;\theta)\cdot(\exp(\alpha+\beta-\tau_t-\eta) - \exp(-\tau_t) + 1) + \exp(\phi_{t+1})\cdot A_2(t+1)/2] / D_3(t;\theta)$.

Thus, the above formula (4.18) has mathematically the same form as the spread funding formula (3.22), except for the term $\xi(t;\theta)$, in which $\varphi(t;\theta)$ can be thought of as the proportional state-feedback controlling parameter (or spread parameter) as in k_t in the spread funding formula (3.22), while $\xi(t;\theta)$ can be regarded as the additive controlling parameter, playing the additional role of a cushion against the solvency and contribution rate risks in connection with $\varphi(t;\theta)$ [see Table 4.1].

Furthermore, we can easily check the characteristics of $\varphi(t;\theta)$ and $\xi(t;\theta)$ with respect to θ, so that for all $t\in[0, T-1]$,

(a) $\varphi(t;\theta)$ is a strictly increasing function of θ because $A_1(T) = e^{-\eta T}$ and $A_1(k)$ is a positive and strictly increasing function of θ, in particular, as $\theta \rightarrow 100\%$, then $\varphi(t;\theta) \rightarrow 1$; that is, strengthening the security (i.e. $\theta=\theta_1 > \theta=\theta_2$) implies a reduction in the amortisation period (i.e. $\varphi(t;\theta_1) > \varphi(t;\theta_2)$);

(b) $\xi(t;\theta)$ can not be said generally to be a (strictly) increasing or decreasing function of θ because of the complexity of the function $A_2(t)$ defined in the recursive equation (4.15), but it is true that as $\theta \rightarrow 100\%$, then $\xi(t;\theta) \rightarrow [\exp(-\phi_{t+1}) + \exp(-\tau_t) - \exp(\alpha+\beta-\tau_t-\eta) - 1]$ because $A_1(t) \rightarrow e^{-\eta t}$, $A_2(t) \rightarrow -2e^{-\eta t}$ and $srt_t = 100\%$ by assumption; and thus,

(c) as $\theta \rightarrow 100\%$, then $SR'_{t+1} \rightarrow 1$ and $CR'_{t+1}-NR_{t+1} \rightarrow [\exp(-\phi_{t+1}) + \exp(-\tau_t) - \exp(\alpha+\beta-\tau_t-\eta) - 1]$, whereas $CR'_0-NR_0 \rightarrow [-(SR'_0-1) + \exp(-\phi_1) + \exp(-\tau_0) - \exp(\alpha+\beta-\tau_0-\eta) - 1]$ where $SR'_0 = SR_0$ given.

By using the above assumptions in section 4.3.3.1, we give a numerical example of the movement of the time-varying parameters $\varphi(t;\theta)$ and $\xi(t;\theta)$ involved in formula (4.18) as time t progresses to T-1, which is illustrated in Table 4.1, based on $\theta = 50\%$. To make the patterns of $\{\varphi(t;\theta);\ t\in[0,\ T-1]\}$ and $\{\xi(t;\theta);\ t\in[0,\ T-1]\}$ clearer, the case of the poorest investment performance (i.e. $\exp(\delta) = 1.04 < \exp(\eta)=1.06$) is additionally illustrated.

From this table, it is clear that both $\varphi(t;\theta)$ and $\xi(t;\theta)$ are nearly constant during the initial periods, while, during the last few periods, $\varphi(k;\theta)$ increases but $\xi(k;\theta)$ decreases as time t becomes close to T-1 (- this may imply that some trade-off between $\varphi(t;\theta)$ and $\xi(t;\theta)$ is maintained over the last few periods). Further, $\varphi(t;\theta)$ and $\xi(t;\theta)$ each are sensitive to the changes in the investment performance.

In the next section, we simply illustrate how to determine the best value of θ (denoted by θ^*) according to the θ^*-criterion (introduced in subsection (v) of section 4.2.3.2) and provide an illustrative projection of our dynamic pension funding plan.

4.3.3.3 Numerical illustrations of dynamic pension funding plan

(i) <u>Searching for θ^*</u>:

From the assumptions in section 4.3.3.1, we find the best value of θ (i.e. θ^*) among the admissible set {90%, 50%, 10%} based on the θ^*-criterion.

The simulated pre-optimal error projections $\{SR'_t-1:\ t=0,\ 1,\ \ldots,\ 12\}$ and $\{CR'_t-NR_t:\ t=0,\ 1,\ \ldots,\ 11\}$ for $\exp(\delta) = 1.06$ are given in Graphs 4.1.1 (in the case of $SR_0 = 0\%$) and 4.1.2 (in the case of $SR_0 = 100\%$). The other case of $\exp(\delta) = 1.08$ leads to a similar result to Graphs 4.1.1 and 4.1.2.

These graphs show that firstly, $\theta^* = 50\%$ would be suitable as a balancing point of the conflicting interests of the employer and trustees: in other words, the pace of funding and the progress of solvency level are better balanced at $\theta^* = 50\%$, rather than at $\theta^* = 90\%$ and 10%. The choice of a balancing θ^* would be more crucial in the case of SR_0 being far away from the solvency target 100% than in the case of SR_0 close to the solvency target 100%: for $SR_0 = 0\%$, both the funding burden on the employer and the scheme solvency are heavily affected by the value of θ^* as seen in Graph 4.1.1, whereas for $SR_0 = 100\%$, there is neither a funding burden nor insolvency, and further, the simulated pre-optimal (control and control action) error projections do not change greatly with the variations in θ^*, as seen in Graph 4.1.2.

Finally, Graph 4.1.1 clearly illustrates that when adopting the supplementary performance criterion (4.4.1) designed for removing quickly the pre-optimal (control and control action) errors, the value θ^* is likely to be set at the level of 90% among the admissible set {θ: 90%, 50%, 10%} because this provides the quickest convergence to the specified control targets (i.e. srt_t=100% and crt_t=NR_t), but this policy is linked with a heavy financial burden on the employer at the initial periods; and when employing the supplementary criterion (4.4.3) emphasising smoothness, the value θ^* is likely to be determined at the level of 10% among the admissible set {θ: 90%, 50%, 10%} because this provide the slowest but monotonical convergence to their respective control targets, but this policy is linked with the problem of slowness in progressing to these targets.

In the next subsection, we shall consider the dynamic pension funding plan with the intermediate value $\theta^* = 50\%$ for the numerical illustrations.

(ii) Projections of dynamic pension funding plan:

Our dynamic pension funding plan is governed by the optimal funding formula defined as follows by replacing θ with θ^* in the pre-optimal funding formula (4.18):

$\pi^*_t(SR_t) = NR_t - \varphi(t;\theta^*) \cdot (SR_t - 1) + \xi(t;\theta^*)$ with $\theta^* = 50\%$ specified, and hence,

$SR^*_{t+1} = \exp(\phi_{t+1}) \cdot [(1-\varphi(t;\theta^*)) \cdot SR^*_t + \exp(\alpha+\beta-\tau-\eta) - \exp(-\tau) + \varphi(t;\theta^*) + \xi(t;\theta^*)]$ with $SR^*_0 = SR_0$ and $\theta^* = 50\%$ specified. --- (4.19)

The optimal projections $\{CR^*_t\text{-}NR_t;\ t=0,1,\ldots,11\}$ and $\{SR^*_{t+1};\ t=0,1,\ldots,11\}$, generated from formulae (4.19), are given in Table 4.2.1 (in the case of $\exp(\delta) = 1.06$) and Table 4.2.2 (in the case of $\exp(\delta) = 1.08$).

Comparing these two tables, we can say that if we have a better investment performance (e.g. $\exp(\delta)>\exp(\eta)$ in Table 4.2.2) rather than a best estimate investment performance (e,g, $\exp(\delta)=\exp(\eta)$ in Table 4.2.1), then our dynamic pension funding plan governed by formula (4.19) leads to more reductions in both the financing burden on the employer and the risk of insolvency. In other words, even in using our dynamic pension funding plan, the investment performance of the pension funds plays vital role in enabling the actuary (and trustees) to manage the finances and solvency of the scheme successfully, as is generally well-known in actuarial applications [see Haberman (1994)].

4.3.3.4 Suggestions for reducing the insolvency risk

As illustrated in Table 4.2.1, the best estimate basis with SR_0=0% is seen to be merely funding for a solvency level target 100% due to the market cost adjustment $\exp(\tau) = 1.2$. Thus, the solvency-level projections starting with SR_0=0% are at a lower level than 100% (i.e. insolvent in the view of PLRC) over most of our control period. On the other hand, the best estimate basis with SR_0=100% is shown to be solvent but quite close to the 100% solvency target over the whole control period.

Given the unpredictable and adverse nature of the risk of solvency for pension funds, the trustees would require more protection against this potential risk. The actuary could then suggest the following four options in order to meet their requirement (although these will yield commonly additional contributions to both finance and maintain the higher solvency level, as illustrated below in subsections (i)~(iv)):

(a) Increasing the value θ^*;

(b) Increasing the solvency ratio target srt_k;

(c) Readjusting the valuation basis more conservatively than the best estimate basis, in particular, employing a more conservative valuation interest rate than a best estimate; or

(d) Combining appropriately the above options (a), (b) and/or (c).

The aim of this section is to illuminate the effects of these suggestions in turn through some illustrative numerical examples. Other possible scenarios will produce similar results. To begin with, we note that the numerical illustrations for these options will not consider the θ^*-criterion because it is not necessary in our discussions (although we have to find out a new best value θ^* according to the changes in the related parameters (e.g. srt_t and η)). So, we shall refer to the pre-optimal values as the optimal values. Moreover, the calculation basis used in Table 4.3.1 is considered to be a standard basis for our further arguments, that is, {$\exp(\delta) = \exp(\eta) = 1.06$ (so, $NR_t = 0.02806$), $\theta^* = 50\%$ and $srt_t = 100\%$}; for convenience, we shall indicate only the differences from the standard basis in the numerical illustrations given in Appendix 4A.

(i) <u>Effects of suggestion (a)</u>:

Applying the property (c) investigated in section 4.3.3.2, as $\theta^* \rightarrow 100\%$, we note that $SR^*_{t+1} \rightarrow 1$ and $CR^*_{t+1}\text{-}NR_{t+1} \rightarrow -0.001478$ for all t, whereas $CR^*_0\text{-}NR_0 \rightarrow 0.998522$ for $SR_0 = 0\%$ and $CR^*_0\text{-}NR_0 \rightarrow -0.001478$ for $SR_0 = 100\%$, which is illustrated in Table 4.3.1 for $\theta^* = 99.9\%$.

Therefore, simply putting greater emphasis on the solvency risk (i.e. $\theta^* \rightarrow 1$) may be unacceptable to both the employer and trustees, particularly in the case of the initial solvency level being far lower than the solvency target 100%, because the additional funding burden at the starting point places a heavy burden on the employer and there will be no substantial improvement in solvency level due to convergence to 100%. On the contrary, in the case that the initial solvency level is close to 100% and the funding policy is to manage solvency levels at the level of 100%, the policy of setting θ^* close to 100% would be highly suitable.

(ii) Effects of suggestion (b):

As illustrated in Table 4.3.2, slightly increasing the solvency target (here, switching the solvency target from 100% to 100.5%) shows that the risk of the occurrence of insolvency are increasingly reduced in the case of SR_0 = 0% (relative to the results shown in Table 4.2.1) and the projections of solvency levels starting with SR_0=100% are also improved to around the new solvency target. The resulting additional contributions occurring during first period are commonly of "reasonable" magnitude (considering the modest reductions in contribution rates during the last few periods).

(iii) Effects of suggestion (c):

Funding on a more conservative basis than a best estimate basis leads to an increasingly healthy solvency position (relative to that being shown in Table 4.2.1), as illustrated in Table 4.3.3. Further, this table clearly shows that the lower is the specified initial solvency level, the more conservative a valuation basis is required in order to reduce the occurrences of insolvencies. Further, the additional contributions that are inevitably required to be paid are higher mainly due to the increase in the normal cost NR_t.

(iv) Effects of suggestion (d):

As examined in the above (i)~(iii), much attention should be focused by the actuary on the case of the initial solvency level being low. In Table 4.3.4, we are concerned with examining the effects of the combined strategy of (a), (b) and/or (c) in the case of SR_0=0%.

This table can be said to imply that even though there are a number of ways of improving the solvency level over the whole control period, the actuary may need to set a best combination among all available ways, subject to some constraints (typically, involved in balancing the conflicting interests of the trustees and employer). This subject would be an interesting and potentially useful area of work but constraints of time and space mean that this subject is left to future research as a possible extension to this thesis.

4.3.3.5 Conclusions

We can derive the following conclusions from Graph 4.1.1 and 4.1.2, and Tables 4.1~ 4.3.4:

(a) Our dynamic pension funding plan governed by the optimal funding formula (4.19) can be characterised mathematically by the additive controlling parameter $\xi(t;\theta^*)$, compared with the spread funding plan governed by the spread funding formula (3.22) [see Table 4.1];

(b) The optimal projections, $\{CR^*_t; t\in[0, T-1]\}$ and $\{SR^*_t; t\in[0, T]\}$, illustrate the effects of reacting to the initial solvency level SR_0, that is, a higher SR_0 contributes to improved stability and security [see Tables 4.2.1~ 4.3.3]. Further, these projections are greatly affected by the value of θ specified, particularly in the case of SR_0 being lower than 100%, that is, as $\theta \to 1$, then the solvency level speedily progresses to 100% but the additional funding burden is quite high at the starting point [see Graph 4.1.1 and Table 4.3.1];

(c) For better protection against insolvency, the actuary is likely to employ one or a combination of the three distinct strategies, increasing the value of θ, increasing the solvency target and adjusting the valuation basis in a more conservative direction [compare Table 4.2.1 with 4.3.1~4.3.4]; and

(d) The better protection against insolvency the more the additional financial burden on the employer, particularly in the case of lower initial solvency level [compare Tables 4.2.1 with 4.3.1~4.3.4]. This clear conclusion is consistent with that of Collins (1992) introduced in section 3.3.

Finally, although we have not illustrated the numerical results in full, one of the most interesting results from our numerical experiments is that the force of new entrant growth α and/or the force of salary growth β has a great impact on the optimal projections $\{CR^*_t;\ t\in[0, T-1]\}$ and $\{SR^*_t;\ t\in[0, T]\}$. As expected, this result is related to the exponential actuarial liability growth equation (a) derived in section 4.3.3.1, i.e. $AL_{t+1} = \exp(\alpha+\beta) \cdot AL_t$. In other words, from the viewpoint of stability and solvency, the optimal projections are worsened severely with the increase in $\exp(\alpha+\beta)$, while the optimal projections are greatly improved with the decrease in $\exp(\alpha+\beta)$; hence, the effects of α and/or β would be well-matched with opposing effects of the force of investment interest (i.e. δ).

4.4 Dynamic pension funding plan for stochastic LQP control optimisation problems

The objective of this section is to find the optimal control laws for solving our stochastic LQP control optimisation problems (4.8) and (4.9) formulated in section 4.2.4. These can be considered as the stochastic version of the problem examined in section 4.3. In a manner analogous to the deterministic optimisation of section 4.3, we will then be able to solve the problems (4.8) and (4.9), but this requires more care because we must allow for the effects of uncertainty existing in the system parameter ϕ_{t+1} and benefit ratio BR_t.

We consider firstly the problem (4.8) with complete state information and secondly the problem (4.9) with incomplete state information.

4.4.1 Complete state information

4.4.1.1 Preliminaries

The problem (4.8) is characterised by

(a) The stochastic controlled object specified by for all $t \in [0, T-1]$,

$SR_{t+1} = \exp[\phi_{t+1}] \cdot [SR_t + CR_t - BR_t]$ with given SR_0,

where $\phi_{t+1} \sim IID\ N(\mu_{t+1}, \sigma_a^2)$ with $\sigma_a^2 < \infty$, $BR_t \sim IID\ N(EBR_t, VBR_t)$ with $VBR_t < \infty$ and ϕ_{t+1} and BR_s are independent for all $t, s \in [0, T-1]$;

(b) The given initial information $H_0 = SR_0$ is independent of ϕ_{t+1} and BR_t; and

(c) The measurement equation $M_t = SR_t$ with given $M_0 = SR_0$ (i.e. there is no time delay in the state information).

Consider a situation at time $t \in [0, T-1]$. As mentioned earlier in subsection (iii) in section 2.3.2.2, the value of the current state SR_t (informed from the actuarial valuation at time t) contains all the essential and summarised information on the available information vector at time t, $H_t = (SR_0, SR_1, \ldots, SR_t, CR_0, CR_1, \ldots, CR_{t-1})$. In other words, the sequence of stochastic variables $\{SR_0, SR_1, \ldots, SR_T\}$ has the Markov property (and hence the sequence is a (discrete-time, finite-state) Markov process): the conditional distribution of the future dynamic state SR_{t+1} given H_t depends only on the current dynamic state SR_t because the controlled object is governed by a stochastic difference equation of order one, i.e. $Pr(SR_{t-1} \mid H_t) = Pr(SR_{t+1} \mid SR_t)$ for all $t \in [0, T-1]$, a property which can be extended to the joint conditional distribution of the future dynamic states $SR_{t+1}, SR_{t+2}, \ldots, SR_T$ given H_t, i.e. $Pr(SR_{t+1}, SR_{t+2}, \ldots, SR_T \mid H_t) = Pr(SR_{t+1}, SR_{t+2}, \ldots, SR_T \mid SR_t)$ for all $t \in [0, T-1]$. Hence, it is sufficient to determine CR_t as a linear function of the current dynamic state SR_t of the controlled object (as in the deterministic case examined in Remark 4.1 in section 4.3.1): in other words, the general form of feedback control described in subsection (b) of section 4.2.1, i.e. $CR_t = {}^{C}\pi_t(H_t)$, is reduced to $CR_t = {}^{C}\pi_t(SR_t)$ linear in SR_t. As a result, SR_t itself is a state variable at time t for every $t\in[0, T-1]$ - this is helpful on grounds of memory efficiency because otherwise, the actuary may need to retain the full information H_t (which has the monotonic property of increasing with time t).

Applying the backward dynamic programming method for sequential control optimisation as in section 4.3.1, we can directly obtain the Bellman equation for the problem (4.8). Prior to utilising this method, the following two properties would be helpful.

(i) Property 1 [for proof, see Grimmett and Stirzaker (1992, pp 67 and 106)]):

Let X and Y be two integrable random variables on a probability space $(\Omega, \aleph, Pr)$. The conditional expectation of Y given X, written as $E(Y \mid X)$, has the important property such that $E(Y) = E\{E(Y \mid X)\}$.

In our applications, the mathematical expectation of PIB_θ (described in equation (4.10) in section 4.3.1) has the property that $E(PIB_\theta) = E\{E(PIB_\theta \mid H_t)\} = E\{E(PIB_\theta \mid SR_t)\}$ for all $t\in[0, T-1]$, in which the second equation comes from the Markov property of $\{SR_0, SR_1, \ldots, SR_T\}$.

(ii) Property 2 [for proof, see Astrom (1970, p 260)]:

Let X be a integrable random variable on a probability space (Ω, $\Im$, Pr) and let Y be the controlling or decision variable as a function of X. Let the deterministic performance index be PI (X, Y) as a function of X and Y. Assuming that PI(X, Y) has a unique minimum with respect to Y for all X, then

$$\underset{Y}{\text{Min}}\, E\{PI(X, Y)\} = E\{\underset{Y}{\text{Min}}\, PI(X, Y)\} = E\{PI(X, Y^*)\}$$

where Y^* denotes the value of Y at which the minimum is achieved

; these equations imply that the expected value E{PI(X, Y)} can be minimised by minimising the inner part PI(X, Y) of E{PI(X, Y)}.

In our applications in connection with Property 1, we find that for all $t \in [0, T\text{-}1]$

$$\underset{CR_t, CR_{t+1}, \ldots, CR_{T-1}}{\text{Min}} E(PIB_\theta)$$

$$= \underset{CR_t, CR_{t+1}, \ldots, CR_{T-1}}{\text{Min}} E\{E(PIB_\theta \mid SR_t)\}$$

$$= E\{ \underset{CR_t, CR_{t+1}, \ldots, CR_{T-1}}{\text{Min}} E(PIB_\theta \mid SR_t) \},$$

which implies that $E(PIB_\theta)$ can be minimised by minimising the conditional expectation $E(PIB_\theta \mid SR_t)$ with respect to CR_t for all SR_{t+1}, $t \in [0, T\text{-}1]$.

4.4.1.2 Bellman equation

To begin with, we note that for notational convenience, the superscript 'C' on the left side of each main symbol is used to indicate that it concerns the problem with complete state information.

In view of the above Properties 1 and 2, we first define the following equation in order to produce the backward recursion in time.

$$^{C}V(SR_t, t) = \underset{\{CR_s;\, s=t,t+1,\ldots,T-1\}}{\text{Min}} E\{\sum_{s=t}^{T-1} \{e^{-\eta s}\cdot[\theta\cdot(SR_s - srt_s)^2 + (1-\theta)\cdot(CR_s - crt_s)^2]\} +$$

$$e^{-\eta T}\cdot(SR_T - srt_T)^2 \mid SR_t\}$$

Applying now the backward dynamic programming method (based on Bellman's principle of optimality) for sequential control optimisation as in section 4.3.1, we find the following version of the Bellman equation

$$^{C}V(SR_T, T) = E\{e^{-\eta T}\cdot(SR_T - srt_T)^2 \mid SR_T\} = e^{-\eta T}\cdot(SR_T - srt_T)^2 \quad \text{and}$$

$$^{C}V(SR_t, t) = \underset{CR_t}{\text{Min}}\, E\{ e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2] +$$

$$\underset{\{CR_s;\, s=t+1,t+2,\ldots,T-1\}}{\text{Min}} E\{\{ \sum_{s=t+1}^{T-1} (e^{-\eta s}\cdot[\theta\cdot(SR_s - srt_s)^2 + (1-\theta)\cdot(CR_s - crt_s)^2]) +$$

$$e^{-\eta T}\cdot[\theta\cdot(SR_T - srt_T)^2] \}\mid SR_{t+1}\} \mid SR_t \}$$

$$= \underset{CR_t}{\text{Min}} \{e^{-\eta t}\cdot[\theta\cdot(SR_t-srt_t)^2 + (1-\theta)\cdot(CR_t-crt_t)^2] + E\{^{C}V(SR_{t+1}, t+1) \mid SR_t\}\} \quad --- (4.20)$$

; here, $^{C}V(SR_t, t)$ can be interpreted to be the minimal expected future cost discounted at time 0 in the case of complete state information, given the summarised information up to time t (i.e. SR_t), which is independent of the control actions before time t, and obeys the above recursion in time.

Further, the terminal cost associated with the terminal state SR_T at the terminal time T, $^{C}V(SR_T, T)$, gives the boundary condition for the Bellman equation (4.20)

4.4.1.3 Control optimisation

Using the fact that the control law at time t is a linear function of the dynamic state SR_t [see section 4.4.1.1], we will now show that the solution of the Bellman equation (4.20) with the boundary condition ${}^{C}V(SR_T, T) = e^{-\eta T}\cdot(SR_T - srt_T)$ is uniquely determined in the following quadratic form (which can be verified using a mathematical induction argument, as discussed in section 4.3.2):

$$^{C}V(SR_t, t) = {}^{C}A_1(t)\cdot SR_t^2 + {}^{C}A_2(t)\cdot SR_t + {}^{C}A_3(t) \quad \text{with the boundary conditions}$$

$$^{C}A_1(T) = e^{-\eta T},\ {}^{C}A_2(T) = -2e^{-\eta T}\cdot srt_T \text{ and } {}^{C}A_3(T) = e^{-\eta T}\cdot srt_T^2. \qquad \text{- - - (4.21)}$$

The above form (4.21) is true for t=T and then proceeding by mathematical induction, we have ${}^{C}V(SR_{t+1}, t+1) = {}^{C}A_1(t+1)\cdot SR_{t+1}^2 + {}^{C}A_2(t+1)\cdot SR_{t+1} + {}^{C}A_3(t+1)$ with ${}^{C}A_1(T)=e^{-\eta T}$, ${}^{C}A_2(T)=-2e^{-\eta T}\cdot srt_T$ and ${}^{C}A_3(T)=e^{-\eta T}\cdot srt_T^2$.

To obtain the solution of the Bellman equation (4.20), we firstly determine the conditional first and second moments of SR_{t+1} given SR_t which come from the following equations:

$$SR_{t+1} = \exp[\phi_{t+1}]\cdot[SR_t + CR_t - BR_t] \text{ and}$$

$$^{C}V(SR_t, t) = \underset{CR_t}{\text{Min}}\ \{e^{-\eta t}\cdot[\theta\cdot(SR_t\text{-}srt_t)^2 + (1\text{-}\theta)\cdot(CR_t\text{-}crt_t)^2] + E\{{}^{C}V(SR_{t+1}, t+1) \mid SR_t\}\}.$$

Then, the conditional first and second moments of SR_{t+1}, given SR_t, are obtained as

$$E\{SR_{t+1} \mid SR_t\} = \exp(\mu_{t+1} + \sigma_a^2/2)\cdot[SR_t + CR_t - EBR_t] \text{ and}$$

$$E\{SR_{t+1}^2 \mid SR_t\} = \exp(2\mu_{t+1}+2\sigma_a^2)\cdot[(SR_t + CR_t)^2 - 2EBR_t\cdot(SR_t + CR_t) + VBR_t + EBR_t^2].$$

Given the above known results, we can rewrite the Bellman equation (4.20) in the form: for each $t\in[0, T\text{-}1]$,

$$^{C}V(SR_t, t) = \underset{CR_t}{\text{Min}}\ \{{}^{C}G(CR_t, t)\}$$

where

$$^{C}G(CR_t, t) = \{e^{-\eta t}\cdot(1-\theta) + \exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)\} \cdot CR_t^2 + \{-2crt_t\cdot e^{-\eta t}\cdot(1-\theta) + 2[\exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)]\cdot(SR_t-EBR_t) + \exp(\mu_{t+1}+\sigma_a^2/2)\cdot{}^{C}A_2(t+1)\} \cdot CR_t + \{e^{-\eta t}\cdot(\theta\cdot(SR_t-srt_t)^2 + (1-\theta)\cdot crt_t^2) + {}^{C}A_1(t+1)\cdot[(\exp(2(\mu_{t+1}+\sigma_a^2)) \cdot (SR_t^2 - 2EBR_t\cdot SR_t + VBR_t + EBR_t^2)] + {}^{C}A_2(t+1)\cdot[\exp(\mu_{t+1}+\sigma_a^2/2) \cdot (SR_t-EBR_t)] + {}^{C}A_3(t+1)\}.$$

It is sufficient that the functional coefficient of CR_t^2 in $^{C}G(CR_t, t)$ is positive for all $t\in[0, T-1]$ as a condition for a unique sequence of pre-optimal control actions $\{CR'_t;\ t\in[0, T-1]\}$, which are optimal subject to an arbitrary specified value of θ, that is

$$^{C}A_1(t+1) > -\{e^{-\eta t}\cdot(1-\theta) / \exp[2(\mu_{t+1}+\sigma_a^2)]\} \quad \text{for all } t\in[0, T-1] \qquad \text{--- (4.22)}$$

; hence, this is clearly true for $t = T-1$.

Utilising the fact that $^{C}G(CR_t, t)$ is a strictly convex function under condition (4.22), we obtain the pre-optimal control action at time t, $^{C}CR'_t$, for all $t\in[0, T-1]$

$$\begin{aligned} ^{C}CR'_t &= -[^{C}D_1(t;\theta)/^{C}D_3(t;\theta)] \cdot SR_t + [^{C}D_2(t;\theta)/^{C}D_3(t;\theta)] \\ &= -[^{C}D_1(t;\theta)/^{C}D_3(t;\theta)] \cdot [SR_t - 1] + [^{C}D_2(t;\theta) - {}^{C}D_1(t;\theta)] / {}^{C}D_3(t;\theta) \\ &\equiv {}^{C}\pi'_t(SR_t) \qquad \text{--- (4.23)} \end{aligned}$$

where

$^{C}D_1(t;\theta) = \exp[(2(\mu_{t+1}+\sigma_a^2)]\cdot {}^{C}A_1(t+1)$,

$^{C}D_2(t;\theta) = crt_t\cdot e^{-\eta t}\cdot(1-\theta) + \exp[2(\mu_{t+1}+\sigma_a^2)]\cdot {}^{C}A_1(t+1)\cdot EBR_t - \exp[\mu_{t+1}+\sigma_a^2/2]\cdot {}^{C}A_2(t+1)/2$, and

$^{C}D_3(t;\theta) = e^{-\eta t}\cdot(1-\theta) + \exp[2(\mu_{t+1}+\sigma_a^2)]\cdot {}^{C}A_1(t+1)$

; in a similar manner to the deterministic case, the above equation shall be called the pre-optimal (linear feedback pension) funding formula and accordingly, $^{C}\pi'_t(.)$ the pre-optimal (linear feedback pension funding) control law at time t, expressed as a function of the current state variable SR_t.

For completion, substituting ${}^{C}CR'_t$ into Min ${}^{C}G(CR_t, t)$ leads to

$$
\begin{aligned}
{}^{C}V(SR_t, t) &= {}^{C}A_1(t)\cdot SR_t^2 + {}^{C}A_2(t)\cdot SR_t + {}^{C}A_3(t) \\
&= \{[e^{-2\eta t}\cdot\theta\cdot(1-\theta) + e^{-\eta k}\cdot\exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)] / [e^{-\eta t}\cdot(1-\theta) + \\
&\quad \exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)]\}\cdot SR_t^2 + \{[2e^{-\eta t}\cdot\exp(2(\mu_{t+1}+\sigma_a^2))\cdot((1-\theta)\cdot \\
&\quad (crt_t - EBR_t) - \theta\cdot srt_t)\cdot{}^{C}A_1(t+1) + e^{-\eta k}\cdot\exp(\mu_{t+1}+\sigma_a^2/2)\cdot(1-\theta)\cdot{}^{C}A_2(t+1) - \\
&\quad 2e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot srt_t] / [e^{-\eta t}\cdot(1-\theta) + \exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)]\}\cdot SR_t + \\
&\quad {}^{C}[\text{Remaining part}].
\end{aligned}
$$

We note that it is not necessary to compute fully the ^{C}[Remaining part] because this constant term does not affect our calculation of ${}^{C}CR'_t$, i.e. ${}^{C}CR'_t$ depends on the functions ${}^{C}A_1(t+1)$ and ${}^{C}A_2(t+1)$, not on ${}^{C}A_3(t+1)$. So, we need only to solve the backward recursions below for all t:

$$
\begin{aligned}
{}^{C}A_1(t) &= [e^{-2\eta t}\cdot\theta\cdot(1-\theta)+e^{-\eta t}\cdot\exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)]/[e^{-\eta k}\cdot(1-\theta)+\exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)] \text{ and} \\
{}^{C}A_2(t) &= [2e^{-\eta t}\cdot\exp(2(\mu_{t+1}+\sigma_a^2))\cdot((1-\theta)\cdot(crt_t - EBR_t) - \theta\cdot srt_t)\cdot{}^{C}A_1(t+1) + e^{-\eta t}\cdot\exp(\mu_{t+1}+\sigma_a^2/2)\cdot \\
&\quad (1-\theta)\cdot{}^{C}A_2(t+1) - 2e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot srt_t] / [e^{-\eta t}\cdot(1-\theta) + \exp(2(\mu_{t+1}+\sigma_a^2))\cdot{}^{C}A_1(t+1)], \quad \text{- - - (4.24)}
\end{aligned}
$$

which is soluble by back-tracking step by step, starting from the boundary conditions, ${}^{C}A_1(T) = e^{-\eta T}$ and ${}^{C}A_2(T) = -2e^{-\eta T}\cdot srt_T$, and hence the first recursive equation can be solved to give ${}^{C}A_1(t)$ and then after substitution the second can be solved to give ${}^{C}A_2(t)$.

It should be noted that the above recursive equation for ${}^{C}A_1(.)$ generates sequentially the positive sequence $\{{}^{C}A_1(t);\ t\in[0, T-1]\}$ in the backward course of time t, starting with ${}^{C}A_1(T) = e^{-\eta T} > 0$; hence, the condition (4.22) for uniqueness is redundant. Thus, we find that the

Bellman equation (4.20) has a unique solution of the suggested quadratic form (4.21) with ${}^{C}A_1(t)$ and ${}^{C}A_2(t)$ satisfying the above recursions (4.24) - here, we complete the mathematical induction argument.

From the fact that ${}^{C}A_1(t) > 0$ for all t, we find that $0 < {}^{C}D_1(t;\theta) < {}^{C}D_3(t;\theta)$ for all t. Then, the pre-optimal funding formula (4.23) has a similar mathematical form to the spread funding formula (3.22) specified in section 3.4.4: ${}^{C}D_1(t;\theta)/{}^{C}D_3(t;\theta)$ can be thought of as corresponding to the spread parameter k_t and $[{}^{C}D_2(t;\theta^*)-{}^{C}D_1(t;\theta^*)]/\ {}^{C}D_3(t;\theta^*)$ to NR_t.

Moreover, the pre-optimal control response SR'_{t+1} corresponding to ${}^{C}CR'_t$ is generated in the form: for each $t\in[0, T-1]$,

$$SR'_{t+1} = \exp[\phi_{t+1}] \cdot [SR'_t + {}^{C}CR'_t - BR_t] \quad \text{with given } SR_0 = SR'_0. \qquad \text{--- (4.25)}$$

After determining a best value of θ (denoted by θ^*) by employing the θ^*-criterion, we can obtain a unique sequence of optimal control actions $\{{}^{C}CR^*_0, {}^{C}CR^*_1, \ldots, {}^{C}CR^*_{T-1}\}$ and corresponding optimal control responses $\{SR_0, SR^*_1, \ldots, SR^*_T\}$ with θ determined by θ^* in equations (4.23), (4.24) and (4.25).

In conclusion, our dynamic pension funding plan is defined as a sequence of time-indexed functions $\{{}^{C}\pi^*_t(.); t\in[0, T-1]\}$ in which ${}^{C}\pi^*_t(.)$ is defined by the equation with θ determined by θ^* in equation (4.21), that is, for every $t\in[0, T-1]$,

$${}^{C}\pi^*_k(SR_t) = -\, [{}^{C}D_1(t;\theta^*)\ /\ {}^{C}D_3(t;\theta^*)] \cdot [SR_t - 1] + [{}^{C}D_2(t;\theta^*) - {}^{C}D_1(t;\theta^*)]\ /\ {}^{C}D_3(t;\theta^*) \qquad \text{--- (4.26)}$$

; in a similar manner to the pre-optimal funding formula (4.23) and control law ${}^{C}\pi'_t(.)$, this equation shall be called the optimal funding formula and function ${}^{C}\pi^*_t(.)$ the optimal control law at time t (and hence, our dynamic pension funding plan is governed by the optimal funding formula).

We note finally that the dynamic economic model of the stochastic process of investment returns$\{\delta_{t+1};\ t \in [0, T-1]\}$ could be extended by considering more complex representations of investment returns, for example, using the time series models of the so-called ARIMA type introduced by Box & Jenkins (1976). Thus, based on Panjer & Bellhouse (1980)'s empirical justification of the autoregressive representation of the underlying force of interest, and following the approach of Haberman (1994), we could set δ_{t+1} as following a stationary first-order (unconditional) autoregressive model SAR(1) such that $\delta_{t+1} = \eta + \gamma \cdot (\delta_t - \eta) + {}^{a}\varepsilon_{t+1}$, $|\gamma| < 1$. Applying the SAR(1) model to our stochastic control problem (4.8), the minimal expected cost function is a function of SR_t and δ_t, so the recursive equations of ${}^{C}A_1(t)$ and ${}^{C}A_2(t)$ can not be derived in full, except at time T-1, because expressions like $E\{{}^{C}A_1(t+1)\cdot\exp(\delta_{t+1}) \mid \delta_t\}$ and $E\{{}^{C}A_2(t+1)\cdot\exp(\delta_{t+1}) \mid \delta_t\}$ are not integrable. Only the case of $\gamma = 0$ is soluble and produces the recursive equations of ${}^{C}A_1(t)$ and ${}^{C}A_2(t)$ as given in (4.24).

Remark 4.3: As mentioned in subsection (ii) in section 4.2.5, the Certainty Equivalence Principle does not hold here, since if we replace the random coefficients, $\exp[\phi_{t+1}]$ for all $t \in [0, T-1]$, with their corresponding expected values, $\exp[\mu_{t+1} + \sigma_a^2/2]$ for all $t \in [0, T-1]$, the resulting certainty equivalence solution is not optimal for our stochastic problem (4.8) because $\{\exp[\mu_{t+1} + \sigma_a^2/2]\}^2 \neq \exp[2(\mu_{t+1} + \sigma_a^2)]$.

4.4.2 Incomplete state information

4.4.2.1 Preliminaries

We consider the incomplete state information version (4.9) of the corresponding complete state information problem (4.8) investigated in section 4.4.1, which is characterised as follows:

(a) The stochastic controlled object is specified by for all $t \in [-1, T-1]$,

$$SR_{t+1} = \exp[\phi_{t+1}] \cdot [SR_t + CR_t - BR_t] \text{ with given } SR_{-1} \quad \text{--- (4.27)}$$

where, $\phi_{t+1} \sim$ IID normal (μ_{t+1}, σ_a^2), $BR_t \sim$ IID normal (EBR_t, VBR_t), and ϕ_{t+1} and BR_s are independent for all $t, s \in [-1, T-1]$;

(b) The given initial information $\Im_0 = (SR_{-1}, CR_{-1})$ is independent of ϕ_{t+1} and BR_t; and

(c) The measurement equation $M_t = SR_{t-1}$ with given $M_0 = SR_{-1}$ (i.e. there is a one-unit period time delay in the availability of the state information).

Thus, except that the current value of dynamic state SR_t of the controlled object is no longer available to the actuary, control problem (4.9) is very similar to control problem (4.8).

Consider the situation at time $t \in [0, T-1]$. The current dynamic state SR_t of the controlled object is not observable because $M_t = SR_{t-1}$; for this reason, SR_t is called the conceptual state variable [see Remark (a) in section 2.3.2.2]. Then, the actuary needs to find an observable state variable as a best alternative to the conceptual state variable SR_t in the light of estimation and memory efficiency, which summarises all the information available to the actuary at time t, i.e. $\Im_t = (SR_{-1}, SR_0, \ldots, SR_{t-1}, CR_{-1}, CR_0, \ldots, CR_{t-1})$, and is recursively calculable [see Remark (b) in section 2.3.2.2]. Providing that we define effectively a new state variable, we can solve the problem (4.9) in a similar manner to the approach employed in section 4.4.1.

Now, we shall define the effective state variable for the problem (4.9) as the best alternative to the unknown SR_t.

Although the current state SR_t is not obtainable at time t, its movements are governed by the system equation (4.27). Then, we should define a new system equation whose dynamic state at time t is generated recursively with certainty when the actuary makes his t-th decision, as in the situation of complete state information.

As a result of this argument, we propose that for each $t \in [0, T]$, the conditional mean $\hat{SR}_t = E(SR_t \mid \Im_t)$ is the most effective state variable based on information up to time t (i.e. $\Im_t$), for these reasons that

(a) Supposing that $E(SR_t^2) < \infty$, $\hat{SR}_t = E(SR_t \mid \Im_t)$ is the so-called minimum mean-squared error estimator/predictor (or best estimator/predictor) of SR_t given $\Im_t$. This is because setting $Z_t \equiv Z(\Im_t)$ (i.e. an estimator of SR_t, presenting a function of $\Im_t$) and using the Properties 1 and 2 described in section 4.4.1.1, then $\underset{Z_t}{\text{Min}}\ E[(SR_t - Z_t)^2] = \underset{Z_t}{\text{Min}}\ E\{E[(SR_t - Z_t)^2] \mid \Im_t\}$ leads directly to the result that $Z_t = E(SR_t \mid \Im_t)$, that is, the conditional expectation of SR_t given $\Im_k$ minimises the mean-squared error $E[(SR_t - Z_t)^2]$ over all Z_t [see Grimmett and Stirzaker (1992; section 7.9)];

(b) Using the fact that $SR_t = \exp(\phi_t) \cdot [\exp(-\mu_t-\sigma_a^2/2) \cdot \hat{SR}_t + EBR_{t-1} - BR_{t-1}]$ from the system equation (4.27) where $\hat{SR}_t = E(SR_t \mid \Im_t) = \exp[\mu_t+\sigma_a^2/2] \cdot [SR_{t-1} + CR_{t-1} - EBR_{t-1}]$, then $\Pr(SR_t \mid \Im_t) = \Pr(SR_t \mid \hat{SR}_t)$ which implies that the value of $\hat{SR}_t$ yields full information on the available information vector $\Im_t$;

(c) The sequence of observable states at each valuation date, $\{\hat{SR}_0, \hat{SR}_1, \ldots, \hat{SR}_T\}$ can be generated recursively in the course of time by the following recursion: for every $t \in [0, T-1]$,

$$\hat{SR}_{t-1} = E(SR_{t+1} \mid \Im_{t+1})$$

$$= \exp[\mu_{t+1}+\sigma_a^2/2] \cdot \{\exp[\phi_t-(\mu_t+\sigma_a^2/2)] \cdot \hat{SR}_t + CR_t - [EBR_t - \exp(\phi_t)\cdot(EBR_{t-1} - BR_{t-1})]\}$$

with the (estimated) initial condition $\hat{SR}_0 = \exp[\mu_0+\sigma_a^2/2] \cdot [SR_{-1} + CR_{-1} - EBR_{-1}]$ --- (4.28)

; hence, this new system equation is a stochastic difference equation of order one, which will sequentially generate the state variable $\hat{SR}_t$ with certainty and no time delay as time progresses, like the system equation dealing with the complete state information case (described in (a) in section 4.4.1.1).

(d) In a similar manner to the complete state information case examined in section 4.4.1.1, the sequence of stochastic variables $\{\hat{SR}_0, \hat{SR}_1, \ldots, \hat{SR}_T\}$ is a (discrete-time, finite-state) Markov process because the conditional distribution of the future dynamic state $\hat{SR}_{t+1}$ given $\Im_t$ depends only on the current dynamic state $\hat{SR}_t$, i.e. $\Pr(\hat{SR}_{t+1} \mid \Im_t) = \Pr(\hat{SR}_{t+1} \mid \hat{SR}_t)$ for all $t \in [0, T-1]$, a property which can be extended to the joint conditional distribution of the future dynamic states $\hat{SR}_{t+1}, \hat{SR}_{t+2}, \ldots, \hat{SR}_T$ given $\Im_t$, i.e. $\Pr(\hat{SR}_{t+1}, \hat{SR}_{t+2}, \ldots, \hat{SR}_T \mid \Im_t) = \Pr(\hat{SR}_{t+1}, \hat{SR}_{t+2}, \ldots, \hat{SR}_T \mid \hat{SR}_t)$ for all $t \in [0, T-1]$. Hence, it is sufficient to determine CR_t as a linear function of the current dynamic state $\hat{SR}_t$ of the controlled object governed by the system equation (4.28): in other words, the general form of feedback control described in subsection (b) of section 4.2.1, i.e. $CR_t = {}^{I}\pi_t(\Im_t)$, is reduced to $CR_t = {}^{I}\pi_t(\hat{SR}_t)$ linear in $\hat{SR}_t$. As a result, $\hat{SR}_t$ itself is a state variable at time t and is the best linear predictor for the unknown SR_t, for every $t \in [0, T-1]$ - this is helpful on grounds of memory efficiency because, otherwise, the actuary may need to retain the full information $\Im_t$ (which has the monotonic property of increasing with time t).

In summary, $\hat{SR}_t$ summarises effectively all the information available to the actuary at the time of taking control action CR_t (i.e. $\Im_t$), and is recursively calculable/observable by means of the new system equation (4.28). Therefore, the control problem (4.9) with incomplete state information can be reduced to a problem with complete state information by way of redefining the stochastic controlled object by the new system equation (4.28) instead of the original system equation (4.27). So, for every $t \in [0, T-1]$ the control law at time t admissible to the actuary is described in the form of a linear function of the (currently observable) dynamic state $\hat{SR}_t$ of the new controlled object governed by the system equation (4.28). Therefore, we can solve the problem (4.9) in a similar manner to the approach employed in section 4.4.1 (dealing with the complete state information case). That is, applying the Properties 1 and 2 introduced in section 4.4.1.1, we find that for all $t \in [0, T-1]$,

$$\underset{CR_t, CR_{t+1}, \ldots, CR_{T-1}}{\text{Min}} E(PIB_\theta)$$

$$= \underset{CR_t, CR_{t+1}, \ldots, CR_{T-1}}{\text{Min}} E\{E(PIB_\theta \mid \hat{}SR_t)\}$$

$$= E\{\underset{CR_t, CR_{t+1}, \ldots, CR_{T-1}}{\text{Min}} E(PIB_\theta \mid \hat{}SR_t)\}.$$

4.4.2.2 Bellman equation

We note first that for notational convenience, the superscript 'I' on the left side of each main symbol is used to indicate that it concerns the problem with incomplete state information.

In order to produce the backward recursion in time (as in the situation of complete state information), we define

$$^{I}V(\hat{}SR_t, t) = \underset{\{CR_s;\ s=t,t+1,\ldots,T-1\}}{\text{Min}} E\{\sum_{s=t}^{T-1} \{e^{-\eta s} \cdot [\theta \cdot (SR_s - srt_s)^2 + (1-\theta) \cdot (CR_s - crt_s)^2]\} +$$

$$e^{-\eta T} \cdot (SR_T - srt_T)^2 | \hat{}SR_t\},$$

which implies that since $\hat{}SR_t$ is the state variable (defined as the best predictor of SR_t given $\Im_t$), the right hand side of this equation is expressed as a function of $\hat{}SR_t$.

Applying now the backward dynamic programming method (based on Bellman's principle of optimality) for sequential control optimisation as in section 4.3.1, we can establish the following Bellman equation:

$$^{I}V(\hat{}SR_T, T) = E\{e^{-\eta T} \cdot (SR_T - srt_T)^2 \mid \hat{}SR_T\} \text{ and}$$

$$^{I}V(\hat{SR}_t, t) = \underset{CR_t}{\text{Min}}\, E\{ e^{-\eta t}\cdot[\theta\cdot(SR_t - srt_t)^2 + (1-\theta)\cdot(CR_t - crt_t)^2] \;+$$

$$\underset{\{CR_s;\, s=t+1,t+2,\ldots,T-1\}}{\text{Min}} E\{\{ \sum_{s=t+1}^{T-1} (e^{-\eta s}\cdot[\theta\cdot(SR_s - srt_s)^2 + (1-\theta)\cdot(CR_s - crt_s)^2]) +$$

$$e^{-\eta T}\cdot(SR_T - srt_T)^2 \}| \hat{SR}_{t+1}\} \mid \hat{SR}_t\}$$

$$= \underset{CR_t}{\text{Min}}\, E\{e^{-\eta t}\cdot[\theta\cdot(SR_t-srt_t)^2 + (1-\theta)\cdot(CR_t-crt_t)^2] + {}^{I}V(\hat{SR}_{t+1}, t+1) \mid \hat{SR}_t\} \quad \text{--- (4.29)}$$

; here, $^{I}V(\hat{SR}_t, t)$ presents the minimal expected future cost discounted at time 0 in the situation of incomplete state information, given the summarised information up to time t (i.e. $\hat{SR}_t$), which is independent of the control actions before time t, and obeys the above recursion in time.

Further, the terminal cost associated with the terminal state $\hat{SR}_T$ at the terminal time T, $^{I}V(\hat{SR}_T, T) = E\{e^{-\eta T}\cdot(SR_T - srt_T)^2 \mid \hat{SR}_T\}$, provides the boundary condition for the above the Bellman equation (4.29), where from the relation $SR_t = \exp(\phi_t)\cdot[\exp(-\mu_t-\sigma_a^2/2)\cdot\hat{SR}_t + EBR_{t-1} - BR_{t-1}]$ for all $t\in[0, T]$ [see subsection (ii) in section 4.4.2.1], $^{I}V(\hat{SR}_T, T)$ is computed as $^{I}V(\hat{SR}_T, T) = [e^{-\eta T}\cdot\exp(\sigma_a^2)]\cdot\hat{SR}_T^2 - [2e^{-\eta T}\cdot srt_T]\cdot\hat{SR}_T + [e^{-\eta T}\cdot\exp(2\mu_T+2\sigma_a^2)\cdot VBR_{T-1}+e^{-\eta T}\cdot srt_T^2]$.

4.4.2.3 Control optimisation

Utilising the fact that the control law at time t is a linear function of the (currently observable) dynamic state $\hat{SR}_t$ [see section 4.4.2.1], the solution of the Bellman equation (4.29) with the boundary condition $^{I}V(\hat{SR}_T, T) = E\{e^{-\eta T}\cdot(SR_T - srt_T)^2 \mid \hat{SR}_T\}$ is uniquely determined in the following quadratic form (as in the complete state information case):

$^{I}V(\hat{SR}_t, t) = {}^{I}A_1(t)\cdot\hat{SR}_t^2 + {}^{I}A_2(t)\cdot\hat{SR}_t + {}^{I}A_3(t)$ with the boundary condition $^{I}A_1(T) = e^{-\eta T}\cdot\exp(\sigma_a^2)$, $^{I}A_2(T) = -2e^{-\eta T}\cdot srt_T$ and $^{I}A_3(T) = e^{-\eta T}\cdot\exp(2\mu_T+2\sigma_a^2)\cdot VBR_{T-1} + e^{-\eta T}\cdot srt_T^2$, --- (4.30)

which can be shown using the following mathematical induction argument.

This form holds clearly for $t = T$, and then proceeding by induction, we have

${}^{I}V(\hat{}SR_{t+1}, t+1) = {}^{I}A_1(t+1) \cdot \hat{}SR_{t+1}^2 + {}^{I}A_2(t+1) \cdot \hat{}SR_{t+1} + {}^{I}A_3(t+1)$ with the boundary conditions specified in equation (4.30).

To obtain the solution of the Bellman equation (4.27), we firstly determine the conditional first and second moments of $\hat{}SR_{t+1}$ given $\hat{}SR_t$ which come from the following equations:

$SR_t = \exp(\phi_t) \cdot [\exp(-\mu_t-\sigma_a^2/2) \cdot \hat{}SR_t + EBR_{t-1} - BR_{t-1}]$,

$\hat{}SR_{t+1} = \exp(\mu_{t+1}+\sigma_a^2/2) \cdot \{ \exp[\phi_k-(\mu_t+\sigma_a^2/2)] \cdot \hat{}SR_t + CR_t - [EBR_t - \exp(\phi_t)\cdot(EBR_{t-1} - BR_{t-1})] \}$

and --- (4.31)

${}^{I}V(\hat{}SR_t, t) = \underset{CR_t}{\mathrm{Min}}\, E\{e^{-\eta t}\cdot[\theta\cdot(SR_t-srt_t)^2 + (1-\theta)\cdot(CR_t-crt_t)^2] + {}^{I}V(\hat{}SR_{t+1}, t+1) \mid \hat{}SR_t\}$. --- (4.32)

Then, the conditional first and second moments of SR_t and $\hat{}SR_{t+1}$, given $\hat{}SR_t$, are obtained as follows:

$E\{SR_t \mid \hat{}SR_t\} = E\{SR_t \mid \Im_t\} = \hat{}SR_t$,

$E\{SR_t^2 \mid \hat{}SR_t\} = \exp(\sigma_a^2) \cdot \hat{}SR_t^2 + \exp(2\mu_t+2\sigma_a^2) \cdot VBR_{t-1}$,

$E\{\hat{}SR_{t+1} \mid \hat{}SR_t\} = \exp(\mu_{t+1}+\sigma_a^2/2) \cdot [\hat{}SR_t + CR_t - EBR_t]$ and

$E\{\hat{}SR_{t+1}^2 \mid \hat{}SR_t\} = \exp(2\mu_{t+1}+\sigma_a^2) \cdot [\exp(\sigma_a^2)\cdot\hat{}SR_t^2 + 2(CR_t - EBR_t)\cdot\hat{}SR_t + (CR_t - EBR_t)^2 + \exp(2\mu_t+2\sigma_a^2)\cdot VBR_{t-1}]$.

Given the above results, we can rewrite the Bellman equation (4.29) in the form:

${}^{I}V(\hat{}SR_t, t) = \underset{CR_t}{\mathrm{Min}} \; \{{}^{I}G(CR_t, t)\}$

where

$^{I}G(CR_t, t) = \{e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot {}^{I}A_1(t+1)\}\cdot CR_t^2 + \{-2crt_t\cdot e^{-\eta t}\cdot(1-\theta) + 2[\exp(2\mu_{t+1}+\sigma_a^2)\cdot {}^{I}A_1(t+1)]\cdot(\hat{SR}_t-EBR_t) + \exp(\mu_{t+1}+\sigma_a^2/2)\cdot {}^{I}A_2(t+1)\}\cdot CR_t + \{e^{-\eta t}\cdot\theta\cdot[\exp(\sigma_a^2)\cdot\hat{SR}_t^2 + \exp(2(\mu_t+\sigma_a^2))\cdot VBR_{t-1} - 2srt_t\cdot\hat{SR}_t + srt_t^2] + e^{-\eta t}\cdot(1-\theta)\cdot crt_t^2 + \exp(2\mu_{t+1}+\sigma_a^2)\cdot {}^{I}A_1(t+1)\cdot[\exp(\sigma_a^2)\cdot\hat{SR}_t^2 - 2EBR_t\cdot\hat{SR}_t + EBR_t^2 + \exp(2(\mu_t+\sigma_a^2))\cdot VBR_{t-1})] + \exp(\mu_{t+1}+\sigma_a^2/2)\cdot {}^{I}A_2(t+1)\cdot(\hat{SR}_t-EBR_t) + {}^{I}A_3(t+1)\}.$

It is sufficient that the functional coefficient of CR_t^2 in $^{I}G(CR_t, t)$ is positive for all $t\in[0, T-1]$ as a condition for a unique sequence of pre-optimal control actions $\{CR'_t;\ t\in[0, T-1]\}$, which are optimal subject to an arbitrary value of θ specified, that is

$$^{I}A_1(t+1) > -[e^{-\eta t}\cdot(1-\theta) / \exp(2\mu_{t+1}+\sigma_a^2)] \text{ for all } t\in[0, T-1], \qquad \text{--- (4.33)}$$

which is true for $t = T-1$.

Using the fact that $^{I}G(CR_t, t)$ is a strictly convex function under condition (4.33), we obtain the pre-optimal control action at time t, $^{I}CR'_t$, for all $t\in[0, T-1]$

$$\begin{aligned} ^{I}CR'_t &= -[^{I}D_1(t;\theta) / {}^{I}D_3(t;\theta)]\cdot\hat{SR}_t + [^{I}D_2(t;\theta) / {}^{I}D_3(t;\theta)] \\ &= -[^{I}D_1(t;\theta) / {}^{I}D_3(t;\theta)]\cdot[\hat{SR}_t - 1] + [^{I}D_2(t;\theta) - {}^{I}D_1(t;\theta)] / {}^{I}D_3(t;\theta) \\ &\equiv {}^{I}\pi'_t(\hat{SR}_t) \qquad \text{--- (4.34)} \end{aligned}$$

where

$^{I}D_1(t;\theta) = \exp(2\mu_{t+1}+\sigma_a^2)\cdot {}^{I}A_1(t+1)$,

$^{I}D_2(t;\theta) = crt_t\cdot e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot {}^{I}A_1(t+1)\cdot EBR_t - \exp(\mu_{t+1}+\sigma_a^2/2)\cdot {}^{I}A_2(t+1)/2$ and

$^{I}D_3(t;\theta) = e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot {}^{I}A_1(t+1)$

; in a similar manner to the complete state information case, this equation shall be called the pre-optimal (linear feedback) funding formula and $^{I}\pi'_t(.)$ the pre-optimal (linear feedback pension funding) control law at time t, expressed as a linear function of the current state variable $\hat{SR}_t$.

We note that the pre-optimal control law ${}^I\pi'_t(.)$ is identical to the pre-optimal control law ${}^C\pi'_t(.)$ for the corresponding complete state information problem derived in section 4.4.1, except for the fact that the state SR_t is now replaced by its conditional mean given $\Im_t$, i.e. $\hat{SR}_t = E\{SR_t \mid \Im_t\}$, and the term $\exp[2(\mu_{t+1}+\sigma_a^2)]$ appearing in ${}^C D_1(t;\theta)$, ${}^C D_2(t;\theta)$ and ${}^C D_3(t;\theta)$ is now replaced by $\exp[2\mu_{t+1}+\sigma_a^2]$ in ${}^I D_1(t;\theta)$, ${}^I D_2(t;\theta)$ and ${}^I D_3(t;\theta)$. Thus, we can easily check that if the random coefficients, ϕ_{t+1}'s, are all regarded as a constant, then we can obtain ${}^I\pi'_t(.)$ simply replacing SR_t in ${}^C\pi'_t(.)$ by $\hat{SR}_t$.

For completion, substituting ${}^I CR'_t$ into Min ${}^I G(CR_t, t)$ yields

$$
\begin{aligned}
{}^I V(\hat{SR}_t, t) &= {}^I A_1(t) \cdot \hat{SR}_t^2 + {}^I A_2(t) \cdot \hat{SR}_t + {}^I A_3(t) \\
&= \{[e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot\exp(\sigma_a^2) + e^{-\eta t}\cdot\exp(2\mu_{t+1}+2\sigma_a^2)\cdot{}^I A_1(t+1) + \\
&\quad \exp(4\mu_{t+1}+2\sigma_a^2)\cdot(\exp(\sigma_a^2) - 1)\cdot{}^I A_1(t+1)^2] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot \\
&\quad {}^I A_1(t+1)]\} \cdot \hat{SR}_t^2 + \{[2e^{-\eta t}\cdot\exp(2\mu_{t+1}+2\sigma_a^2)\cdot((1-\theta)\cdot(crt_t - EBR_t) - \\
&\quad \theta\cdot srt_t)\cdot{}^I A_1(t+1) + e^{-\eta t}\cdot\exp(\mu_{t+1}+\sigma_a^2/2)\cdot(1-\theta)\cdot{}^I A_2(t+1) - 2e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot srt_t] / \\
&\quad [e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot{}^I A_1(t+1)]\} \cdot \hat{SR}_t + {}^I[\text{Remaining part}].
\end{aligned}
$$

Here, it is not necessary to calculate the ^{I}[Remaining part] in full because CR'_t depends only on ${}^I A_1(t+1)$ and ${}^I A_2(t+1)$ (not on ${}^I A_3(t+1)$). Then, we need only to solve the backward recursive equations below. For all $t\in[0, T-1]$,

$$
\begin{aligned}
{}^I A_1(t) &= [e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot\exp(\sigma_a^2) + e^{-\eta t}\cdot\exp(2\mu_{t+1}+2\sigma_a^2)\cdot{}^I A_1(t+1) + \exp(4\mu_{t+1}+2\sigma_a^2)\cdot \\
&\quad (\exp(\sigma_a^2) - 1)\cdot{}^I A_1(t+1)^2] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot{}^I A_1(t+1)] \quad \text{and}
\end{aligned}
$$

$^{I}A_2(t) = [2e^{-\eta t}\cdot\exp(2\mu_{t+1}+2\sigma_a^2)\cdot((1-\theta)\cdot(crt_t - EBR_t) - \theta\cdot srt_t)\cdot{}^{I}A_1(t+1) + e^{-\eta t}\cdot \exp(\mu_{t+1}+\sigma_a^2/2)\cdot$

$$(1-\theta)\cdot{}^{I}A_2(t+1) - 2e^{-2\eta t}\cdot\theta\cdot(1-\theta)\cdot srt_t] / [e^{-\eta t}\cdot(1-\theta) + \exp(2\mu_{t+1}+\sigma_a^2)\cdot{}^{I}A_1(t+1)], \quad \text{--- (4.35)}$$

which are soluble by back-tracking step by step, starting from the boundary conditions, $^{I}A_1(T) = e^{-\eta T}\cdot\exp(\sigma_a^2)$ and $^{I}A_2(T) = -2e^{-\eta T}\cdot srt_T$, and hence the first recursive equation can be solved to give $^{I}A_1(t)$ and then after substitution the second can be solved to give $^{I}A_2(t)$.

We can easily check that the above recursion for $^{I}A_1(.)$ generates sequentially the positive sequence $\{^{I}A_1(t);\ t\in[0, T-1]\}$ in the backward course of time t, starting with $^{I}A_1(T) = e^{-\eta T}\cdot\exp(\sigma_a^2) > 0$, since if $^{I}A_1(t+1) > 0$ then $^{I}A_1(t) > 0$; hence, the condition (4.33) for uniqueness is redundant. We thus find that the Bellman equation (4.29) has a solution of suggested quadratic form (4.30) with $^{I}A_1(t)$ and $^{I}A_2(t)$ satisfying the backward recursions (4.35) - here, we complete the mathematical induction argument.

Further, the fact that $^{I}A_1(t) > 0$ for all t implies that $0 < {}^{I}D_1(t;\theta) < {}^{I}D_3(t;\theta)$ for all t so that the pre-optimal funding formula (4.34) has a similar mathematical form to the spread funding formula (3.22) specified in section 3.4.4: that is, $^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta)$ can be thought of as corresponding to the spread parameter k_t and $[^{I}D_2(t;\theta) - {}^{I}D_1(t;\theta)] / {}^{I}D_3(t;\theta)$ to NR_t.

Next, the pre-optimal control response $^{\wedge}SR'_{t+1}$ corresponding to $^{I}CR'_t$ is generated recursively with time t in the form (which is the same mathematical form as equation (4.31), but presents the system equation after obtaining the pre-optimal control law):

$$^{\wedge}SR'_{t+1} = \exp(\mu_{t+1}+\sigma_a^2/2)\cdot\{\exp(\phi_t-\mu_t-\sigma_a^2/2)\cdot{}^{\wedge}SR'_t + {}^{I}CR'_t - [EBR_t - \exp(\phi_t)\cdot(EBR_{t-1} - BR_{t-1})]\}$$

$$\text{with given } ^{\wedge}SR_0 = {}^{\wedge}SR'_0. \quad \text{--- (4.36)}$$

After determining a best value of θ (denoted by θ^*) by applying the θ^*-criterion, we can obtain a unique sequence of optimal control actions $\{^{I}CR^*_0, {}^{I}CR^*_1, \ldots, {}^{I}CR^*_{T-1}\}$ and corresponding optimal control responses $\{{}^{\wedge}SR^*_0, {}^{\wedge}SR^*_1, \ldots, {}^{\wedge}SR^*_T\}$ with θ determined by θ^* in equations (4.34), (4.35) and (4.36).

In conclusion, our dynamic pension funding plan is defined as a sequence of functions $\{{}^I\pi^*_t(.); t\in[0, T-1]\}$ where ${}^I\pi^*_t(.)$ is defined by the equation with θ determined by θ^* in equation (4.34): that is, for every $t\in[0, T-1]$,

$$ {}^I\pi^*_t({}^{\wedge}SR_t) = -[{}^ID_1(t;\theta^*) / {}^ID_3(t;\theta^*)] \cdot [{}^{\wedge}SR_t - 1] + [{}^ID_2(t;\theta^*) - {}^ID_1(t;\theta^*)] / {}^ID_3(t;\theta^*) \quad --- (4.37) $$

; in a similar manner to the pre-optimal funding formula (4.34) and control law ${}^I\pi'_t(.)$, the above equation shall be called the optimal funding formula and function ${}^I\pi^*_t(.)$ the optimal control law at time t and hence our dynamic pension funding plan is governed by this optimal funding formula.

<u>Remark 4.4</u>: As in the dynamic pension funding plan ${}^C\pi^*_t(.)$ for the complete state information problem of section 4.4.1, we can obtain a general solution ${}^I\pi^*_t(.)$ only for the stationary first-order (unconditional) autoregressive model SAR(1) such that $\delta_{t+1} = \eta + \gamma \cdot (\delta_t - \eta) + {}^a\varepsilon_{t+1}$, $\gamma = 0$, since expressions like $E\{{}^IA_1(t+1)\cdot\exp(\delta_t) \mid \delta_{t-1}\}$ and $E\{{}^IA_2(t+1)\cdot\exp(\delta_t) \mid \delta_{t-1}\}$ are not integrable.

4.4.3 Mean and Variance approach

4.4.3.1 Preliminaries

After obtaining the formulae for the control action ${}^CCR'_t$ and its control response SR'_{t+1} in the case of complete state information, or for ${}^ICR'_t$ and ${}^{\wedge}SR'_{t+1}$ in the case of incomplete state information respectively, the actuary may be asked to give the employer and trustees information on the projections of ${}^{C}CR'_t$ and SR'_{t+1}, or ${}^{I}CR'_t$ and ${}^{\wedge}SR'_{t+1}$. The choice of θ^* is of considerable interest for balancing their conflicting interests, so they may want to check the possible movement of the contribution rates and solvency levels under more or less predictable scenarios. In general, two approaches may be suggested for producing the numerical results for determining θ^* in accordance with the θ^*-criterion.

(i) Simulation approach:

It is necessary to repeat the simulation, say n times, according to the idea of the Monte Carlo method. The simulation will consist of the following steps [for further information about simulation technique, see Daykin et. al (1994; Appendix F)]:

Step 1 - Fix up the parameter set $\{\sigma_a, \sigma_b\}$ assumed in the stochastic model (4.6), an arbitrary value of θ and a specified initial condition (SR_0 in the case of complete state information, while ${}^{\wedge}SR_0$ in the case of incomplete state information);

Step 2 - Generate the zero-mean bivariate normal random numbers with variances σ_a^2, σ_b^2 and correlation zero, that is, $\{({}^{a}\varepsilon_1, {}^{b}\varepsilon_1), ({}^{a}\varepsilon_2, {}^{b}\varepsilon_2), \ldots, ({}^{a}\varepsilon_T, {}^{b}\varepsilon_T)\}$ in the case of complete state information, whereas $\{({}^{a}\varepsilon_0, {}^{b}\varepsilon_0), ({}^{a}\varepsilon_1, {}^{b}\varepsilon_1), \ldots, ({}^{a}\varepsilon_{T-1}, {}^{b}\varepsilon_{T-1})\}$ in the case of incomplete state information;

Step 3 - [In the case of complete state information: for each time k=0, 1, ..., T-1, calculate the pre-optimal value of ${}^{C}CR'_k$ defined in formulae (4.23) and then compute the pre-optimal value of SR'_{t+1} generated by equation (4.25) at each simulated $({}^{a}\varepsilon_{t+1}, {}^{b}\varepsilon_{t+1})$] or

- [In the case of incomplete state information: for each time k=0, 1, ..., T-1, calculate the pre-optimal value of ${}^{I}CR'_k$ defined in formulae (4.34) and then compute the pre-optimal value of ${}^{\wedge}SR'_{t+1}$ generated by equation (4.36) at each simulated $({}^{a}\varepsilon_k, {}^{b}(t)$]; and

Step 4 - Return to Step1 and repeat, say n times.

Even though the simulation approach offers a flexible and powerful means of coping with even the most complicated model specifications, there are several specific problems that arise in connection with gathering the data from simulated experiments. Especially, we assume that the experimental output data, $\{({}^a\varepsilon_t, {}^b\varepsilon_t)\}$, are in the form of a collection of distinct and independent random observations from the (zero-mean) bivariate normal distribution with zero correlation.

However, the observations generated from a simulated experiment are likely to be highly correlated with each other because of the artificiality introduced by the starting (or regeneration) point. Another problem is how many realisations are needed to obtain data which are relevant for predicting the steady-state behaviour of the real system, that is, how to determine the iteration number n. Generally, in order to obtain an immediate, visual idea about the character of the real system, such as the range of variation and whether there are any trends, many realisations are required and also a considerable amount of unproductive computer time may be expended [for more tactical problems involved in the simulation approach, see Hiller & Lieberman (1980; pp 663~664)].

Consequently, even though the simulation approach is one of the best techniques for quantifying future uncertainty, both the employer and trustees may find it difficult to derive some clear information for determining θ^* from a number of simulated results.

(ii) Mean and Variance approach:

A potentially more effective alternative would be employing a mean and variance approach for both the employer and trustees for the following reasons. Firstly, the mean and standard deviation can be interpreted to be the main-trend and most-likely-variation from the main-trend, respectively. Secondly, the mean provides a single but clear realisation and then is more economical in computation. Lastly, this approach provides clear information for determining θ^*

to the employer who may find it easier to understand these outputs rather than the outcomes of n simulations.

In conclusion, we prefer the mean and variance approach to the simulation approach not only to determine θ^* but also to make a diagnosis of our optimal future projections for the contribution rate and solvency level.

4.4.3.2 Mean and Variance in the case of complete state information

Substituting ${}^{C}CR'_t$ specified by formulae (4.23) into equation (4.25) yields

$$SR'_{t+1} = \exp(\phi_{t+1}) \cdot \{[({}^{C}D_3(t;\theta)-{}^{C}D_1(t;\theta))/{}^{C}D_3(t;\theta)] \cdot SR'_t + {}^{C}D_2(t;\theta)/{}^{C}D_3(t;\theta) - BR_t\}$$

with the initial condition $SR_0 = SR'_0$. --- (4.38)

By taking the mathematical expectation and variance on the both side of equation (4.38), we can then derive the following mean and variance of SR'_{t+1} for each $t\in[0, T-1]$, each presented as a recurrence relation: that is,

$$E(SR'_{t+1}) = [\exp(\mu_{t+1}+\sigma_a^2/2) \cdot ({}^{C}D_3(t;\theta) - {}^{C}D_1(t;\theta)) / {}^{C}D_3(t;\theta)] \cdot E(SR'_t) + \exp(\mu_{t+1}+\sigma_a^2/2) \cdot [{}^{C}D_2(t;\theta)/{}^{C}D_3(t;\theta) - EBR_t]$$

with the initial condition $E(SR'_0)=SR'_0=SR_0$ -- (4.39)

and

$$\begin{aligned}Var(SR'_{t+1}) = {} & [({}^{C}D_3(t;\theta)-{}^{C}D_1(t;\theta))/{}^{C}D_3(t;\theta)]^2 \cdot Var(\exp(\phi_{t+1})\cdot SR'_t) + [{}^{C}D_2(t;\theta)/{}^{C}D_3(t;\theta)]^2 \cdot \\ & Var(\exp(\phi_{t+1})) + Var(\exp(\phi_{t+1})\cdot BR_t) + 2[({}^{C}D_3(t;\theta)-{}^{C}D_1(t;\theta))/{}^{C}D_3(t;\theta)] \cdot \\ & [{}^{C}D_2(t;\theta)/{}^{C}D_3(t;\theta)] \cdot Cov[\exp(\phi_{t+1})\cdot SR'_t, \exp(\phi_{t+1})] - 2[({}^{C}D_3(t;\theta) - \\ & {}^{C}D_1(t;\theta))/{}^{C}D_3(t;\theta)] \cdot Cov[\exp(\phi_{t+1})\cdot SR'_t, \exp(\phi_{t+1})\cdot BR_t] - 2[{}^{C}D_2(t;\theta)/{}^{C}D_3(t;\theta)] \cdot \\ & Cov[\exp(\phi_{t+1}), \exp(\phi_{t-1})\cdot BR_t]\end{aligned}$$

with the initial condition $Var[SR'_0]=0$ -- (4.40)

where

(a) $Var(\exp(\phi_{t+1})) = \exp(2\mu_{t+1}+2\sigma_a^2) - \exp(2\mu_{t+1}+\sigma_a^2)$,

(b) $Var(\exp(\phi_{t+1})\cdot SR'_t) = [\exp(2\mu_{t+1}+2\sigma_a^2) - \exp(2\mu_{t+1}+\sigma_a^2)] \cdot [E(SR'_t)]^2 + \exp(2\mu_{t+1}+2\sigma_a^2) \cdot Var(SR'_t)$,

(c) $Var(\exp(\phi_{t+1})\cdot BR_t) = [\exp(2\mu_{t+1}+2\sigma_a^2) - \exp(2\mu_{t+1}+\sigma_a^2)] \cdot EBR_t^2 + \exp(2\mu_{t+1}+2\sigma_a^2) \cdot VBR_t$.

(d) $Cov[\exp(\phi_{t+1})\cdot SR'_t, \exp(\phi_{t+1})] = [\exp(2\mu_{t+1}+2\sigma_a^2) - \exp(2\mu_{t+1}+\sigma_a^2)] \cdot E(SR'_t)$,

(e) $Cov[\exp(\phi_{t+1})\cdot SR'_t, \exp(\phi_{t+1})\cdot BR_t] = [\exp(2\mu_{t+1}+2\sigma_a^2) - \exp(2\mu_{t+1}+\sigma_a^2)] \cdot EBR_t \cdot E(SR'_t)$, and

(f) $Cov[\exp(\phi_{t+1}), \exp(\phi_{t+1})\cdot BR_t] = [\exp(2\mu_{t+1}+2\sigma_a^2) - \exp(2\mu_{t+1}+\sigma_a^2)] \cdot EBR_t$.

And also, by replacing SR_t with SR'_t in formula (4.23) and then taking the mathematical expectation and variance on the both sides of formula (4.23), we can easily derive the following mean and variance of ${}^CCR'_t$ for each $t\in[0, T-1]$:

$$E[{}^CCR'_t] = - [{}^CD_1(t;\theta)/{}^CD_3(t;\theta)] \cdot E[SR'_t] + [{}^CD_2(t;\theta)/{}^CD_3(t;\theta)] \quad \text{and} \qquad \text{--- (4.41)}$$

$$Var[{}^CCR'_t] = [{}^CD_1(t;\theta)/{}^CD_3(t;\theta)]^2 \cdot Var[SR'_t] \qquad \text{--- (4.42)}$$

Lastly, the θ^*-criterion (as our supplementary performance criterion) would need to be consistently applied in accordance with the mean and variance approach: that is, the value for θ^* will be determined mainly by reference to the mean-variance pre-optimal error projections such as $\{E(SR'_t-srt_t): t\in[0, T]\}$, $\{E({}^CCR'_t-crt_t): t\in[0, T-1]\}$, $\{Var(SR'_t): t\in[0, T]\}$ and $\{Var({}^CCR'_t): t\in[0, T-1]\}$.

Therefore, after determining θ^*, we can obtain the optimal projections of the contribution ratio and solvency level by simply replacing the arbitrary value of θ with θ^*. Some numerical illustrations will be given in section 4.4.4.

4.4.3.3 Mean and Variance in the case of incomplete state information

In a similar manner to the case of complete state information, we can derive the mean and variance of $\hat{SR}'_{t+1}$ and $^{I}CR'_{t}$, respectively.

Substituting $^{I}CR'_{t}$ specified by formula (4.34) into equation (4.36) leads to

$$\hat{SR}'_{t+1} = \exp(\mu_{t+1}+\sigma_a^2/2) \cdot \{ [\exp(\phi_t-\mu_t-\sigma_a^2/2) - {}^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta)] \cdot \hat{SR}'_t + [({}^{I}D_2(t;\theta))/{}^{I}D_3(t;\theta)) - EBR_t] + \exp(\phi_t) \cdot [EBR_{t-1} - BR_{t-1}] \}$$

$$\text{with the initial condition } \hat{SR}'_0 = \hat{SR}_0 \qquad \text{--- (4.43)}$$

By taking the mathematical expectation and variance on the both sides of equation (4.43), we can then obtain the following recursive equation for the mean and variance of $\hat{SR}'_{t+1}$, respectively: that is, for every $t\in[0, T-1]$,

$$E(\hat{SR}'_{t+1}) = \exp(\mu_{t+1}+\sigma_a^2/2) \cdot \{ [{}^{I}D_3(t;\theta) - {}^{I}D_1(t;\theta)] / {}^{I}D_3(t;\theta) \cdot E(\hat{SR}'_t) + [{}^{I}D_2(t;\theta) / {}^{I}D_3(t;\theta) - EBR_t] \} \text{ with the initial condition } E[\hat{SR}'_0] = \hat{SR}'_0 = \hat{SR}_0 \qquad \text{--- (4.44)}$$

and

$$Var(\hat{SR}'_{t+1}) = \exp(2\mu_{t+1}+\sigma_a^2) \cdot \{Var[(\exp(\phi_t-\mu_t-\sigma_a^2/2) - {}^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta)) \cdot \hat{SR}'_t] + Var[\exp(\phi_t)\cdot(EBR_{t-1} - BR_{t-1})] + 2Cov[(\exp(\phi_t-\mu_t-\sigma_a^2/2) - {}^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta)) \cdot \hat{SR}'_t, \exp(\phi_t)\cdot(EBR_{t-1}-BR_{t-1})]\} \text{ with the initial condition } Var(\hat{SR}'_0)=0 \quad \text{-- (4.45)}$$

where

(a) $Var[(\exp(\phi_t-\mu_t-\sigma_a^2/2) - {}^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta)) \cdot \hat{SR}'_t] = [\exp(\sigma_a^2) - 1] \cdot [E(\hat{SR}'_t)]^2 + [\exp(\sigma_a^2) - 2\,{}^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta) + ({}^{I}D_1(t;\theta)/{}^{I}D_3(t;\theta))^2] \cdot Var(\hat{SR}'_t)$,

(b) $Var[\exp(\phi_t) \cdot (EBR_{t-1} - BR_{t-1})] = \exp(2\mu_t+2\sigma_a^2) \cdot VBR_{t-1}$, and

(c) $Cov[(exp(\phi_t-\mu_t-\sigma_a^2/2)-{}^ID_1(t;\theta)/{}^ID_3(t;\theta)) \cdot {}^{\wedge}SR'_t, exp(\phi_t)\cdot(EBR_{t-1}-BR_{t-1})] = 0$.

Moreover, by replacing ${}^{\wedge}SR_t$ with ${}^{\wedge}SR'_t$ in formulae (4.34) and then taking the mathematical expectation and variance on the both sides of formula (4.34), we can easily derive the following mean and variance of ${}^ICR'_t$ for each $t \in [0, T-1]$, respectively: that is,

$$E[{}^ICR'_t] = -[{}^ID_1(t;\theta)/{}^ID_3(t;\theta)] \cdot E[{}^{\wedge}SR'_t] + [{}^ID_2(t;\theta)/{}^ID_3(t;\theta)] \text{ and} \quad \text{--- (4.46)}$$

$$Var[{}^CCR'_t] = [{}^ID_1(t;\theta)/{}^ID_3(t;\theta)]^2 \cdot Var[{}^{\wedge}SR'_t] \quad \text{--- (4.47)}$$

In a similar manner to the complete state information case, the value for θ^* will be determined mainly by reference to the mean-variance pre-optimal error projections such as $\{E({}^{\wedge}SR'_t-srt_t): t\in[0, T]\}$, $\{E({}^ICR'_t-crt_t): t\in[0, T-1]\}$, $\{Var({}^{\wedge}SR'_t): t\in[0, T]\}$ and $\{Var({}^ICR'_t): t\in[0, T-1]\}$.

After determining θ^*, we can then establish the optimal cash-flow projections of the contribution ratio and solvency level by simply replacing the arbitrary value of θ with θ^*. Some illustrative numerical examples will be given in the next section 4.4.4 in connection with those of the complete state information problem.

4.4.3.4 Performance comparison measures between ${}^C\pi^*_t(.)$ and ${}^I\pi^*_t(.)$

We have shown in section 4.4.2 that the incomplete state LQP control optimisation problem (4.9) can be reduced to another form of the complete state LQP control optimisation problem (4.8) by means of replacing the intrinsically inaccessible state (i.e. SR_t) with the best linear estimator/predictor for SR_t (i.e. ${}^{\wedge}SR_t$, defined as $E(SR_t \mid \Im_t)$ in section 4.4.2.1). So, we need to measure the closeness of our estimator ${}^{\wedge}SR_t$ to the conceptual state variable SR_t, since the estimation error at time t, defined here as $SR_t - {}^{\wedge}SR_t$, will not be zero in general.

Even though we may adopt several measures (commonly termed loss functions, see Berger (1985; section 2.4)) in which two random variables may be said to be close to one another, the so-called mean-squared error is usually used as an appropriate measure in the field of theoretical statistics [see Grimmett & Stirzaker (1992; section 7.9)]: here, the mean-squared error at time t is defined as $E[(SR_t - {}^{\wedge}SR_t)^2]$ for a measure of the precision of the estimator ${}^{\wedge}SR_t$ of the unknown state SR_t.

Prior to further applications of the mean-squared error, it is worth recalling that after setting the optimal control law with chosen θ^*, we can obtain the sequences of optimal control actions, $\{{}^{C}CR^*_0, {}^{C}CR^*_1, \ldots, {}^{C}CR^*_{T-1}\}$ for the complete state information case and $\{{}^{I}CR^*_0, {}^{I}CR^*_1, \ldots, {}^{I}CR^*_{T-1}\}$ for the incomplete state information case, and the sequences of corresponding optimal control responses, $\{SR^*_0, SR^*_1, \ldots, SR^*_T\}$ for the complete state information case and $\{{}^{\wedge}SR^*_0, {}^{\wedge}SR^*_1, \ldots, {}^{\wedge}SR^*_T\}$ for the complete state information case, respectively, in which the initial states, SR^*_0 and ${}^{\wedge}SR^*_0$, each are fixed at time t=0. Moreover, we find that ${}^{C}CR^*_t$ is a function of state SR^*_t and ${}^{I}CR^*_t$ is a function of state ${}^{\wedge}SR^*_t$.

Hence, we can defined the mean-squared error of the optimal control response as follows: for each $t \in [0, T-1]$,

$$^{SR}MSE_{t+1} \equiv E[(SR^*_{t+1} - {}^{\wedge}SR^*_{t+1})^2] \quad \text{--- (4.48)}$$

$$= Var(SR^*_{t+1}) + [E(SR^*_{t+1})]^2 + Var({}^{\wedge}SR^*_{t+1}) + [E({}^{\wedge}SR^*_{t+1})]^2 - 2E(SR^*_{t+1} \cdot {}^{\wedge}SR^*_{t+1})$$

where

$$E(SR^*_{t+1} \cdot {}^{\wedge}SR^*_{t+1}) = \{\exp(2\mu_{t+1}+\sigma_a^2) \cdot [({}^{C}D_3(t;\theta^*) - {}^{C}D_1(t;\theta^*))/{}^{C}D_3(t;\theta^*)] \cdot [({}^{I}D_3(t;\theta^*) - {}^{I}D_1(t;\theta^*))/{}^{I}D_3(t;\theta^*)]\} \cdot E(SR^*_t \cdot {}^{\wedge}SR^*_t) + \{\exp(2\mu_{t+1}+\sigma_a^2) \cdot [({}^{C}D_3(t;\theta^*) - {}^{C}D_1(t;\theta^*))/{}^{C}D_3(t;\theta^*)] \cdot [{}^{I}D_2(t;\theta^*))/{}^{I}D_3(t;\theta^*) - EBR_t]\} \cdot E(SR^*_t) + \{\exp(2\mu_{t+1}+\sigma_a^2) \cdot [({}^{I}D_3(t;\theta^*) - {}^{I}D_1(t;\theta^*))/{}^{I}D_3(t;\theta^*)] \cdot$$

$[{}^{C}D_2(t;\theta^*)/{}^{C}D_3(t;\theta^*)$ - $EBR_t]\}$ · $E(\hat{}SR^*_t)$ + $\exp(2\mu_{t+1}+\sigma_a^2)$ · $\{EBR_t^2$-

$[{}^{C}D_2(t;\theta^*)/{}^{C}D_3(t;\theta^*) + {}^{I}D_2(t;\theta^*)/{}^{I}D_3(t;\theta^*)]$ · EBR_t + $[{}^{C}D_2(t;\theta^*) \cdot {}^{I}D_2(t;\theta^*)]$ /

$[{}^{C}D_3(t;\theta^*)\cdot{}^{I}D_3(t;\theta^*)]\}$ with the initial condition $E(SR^*_0 \cdot \hat{}SR^*_0) = SR^*_0 \cdot \hat{}SR^*_0$

; hence, ${}^{SR}MSE_{t+1}$ is calculable from the above recursive equation and our previous recursive equations (4.39), (4.40), (4.44) and (4.45) with the arbitrary value of θ replaced by the specified θ^*.

Similarly, we can defined the mean-squared error of the optimal control action as follows: since ${}^{C}CR^*_t$ is a linear function of the state SR^*_t (denoted by $f_1(SR^*_t)$) and ${}^{I}CR^*_t$ is a linear function of the state $\hat{}SR^*_t$ (denoted by $f_2(\hat{}SR^*_t)$), the mean-squared error (denoted by ${}^{CR}MSE_t$) can be defined as ${}^{CR}MSE = E[(f_1(SR^*_t) - f_2(\hat{}SR^*_t))^2]$ in a general form of ${}^{SR}MSE$, that is, for each $t\in[0, T-1]$,

$${}^{CR}MSE_t \equiv E[({}^{C}CR^*_t - {}^{I}CR^*_t)^2] \quad \text{--- (4.49)}$$

$$= Var({}^{C}CR^*_t) + [E({}^{C}CR^*_t)]^2 + Var({}^{I}CR^*_t) + [E({}^{I}CR^*_t)]^2 - 2E({}^{C}CR^*_t \cdot {}^{I}CR^*_t)$$

where

$$E({}^{C}CR^*_t \cdot {}^{I}CR^*_t) = \{{}^{C}D_1(t;\theta^*)\cdot{}^{I}D_1(t;\theta^*)\cdot E(SR^*_t \cdot \hat{}SR^*_t) - {}^{C}D_1(t;\theta^*)\cdot{}^{I}D_2(t;\theta^*)\cdot E(SR^*_t) -$$

$${}^{C}D_2(t;\theta^*)\cdot{}^{I}D_1(t;\theta^*)\cdot E(\hat{}SR^*_t) + {}^{C}D_2(t;\theta^*)\cdot{}^{I}D_2(t;\theta^*)\} / \{{}^{C}D_3(t;\theta^*)\cdot{}^{I}D_3(t;\theta^*)\}$$

; hence, this measure is calculable from our previous recursive equations (4.39), (4.41), (4.42), (4.44), (4.46), (4.47) and (4.48) with replacing the arbitrary value of θ with determined θ^*.

Consequently, in order to compare the performance of the optimal control law ${}^{C}\pi^*_t(.)$ for the complete state information with that of the optimal control law ${}^{I}\pi^*_t(.)$ for the incomplete state information at each time $t\in[0, T-1]$, we shall employ the newly defined measures, ${}^{SR}MSE_{t+1}$ defined by the equation (4.48) and ${}^{CR}MSE_t$ defined by the equation (4.49). An illustrative numerical comparison will be made in section 4.4.4.

4.4.4 Numerical illustrations

In this section, we consider the stochastic version of the deterministic numerical illustrations examined in section 4.3.3, by using the mean and variance approach described in section 4.4.3. All numerical illustrations are given in the form of tables in Appendix 4.2, except for Graphs 4.2.1 and 4.2.2.

4.4.4.1 Assumptions

The assumptions are as follows:

(A1)~(A4): the same as the assumptions made in section 4.3.3.1, except that ϕ_{t+1} ~ IID $N(\mu_{t+1}, \sigma_a^2)$ with $\sigma_a^2 < \infty$ and BR_t ~ IID $N(EBR_t, VBR_t)$ with $VBR_t < \infty$, where due to the one-time-unit time delay, this stochastic model is assumed to be applicable at time t = -1, and the standard deviations each are specified as σ_a = 10% or 30% of $|\mu_{t+1}|$ and $\sqrt{VBR_t}$ = 10% or 30% of $|EBR_t|$.

As mentioned in section 4.3.3.1, we note that $E(BR_t) = EBR_t = 0.035396$ and $NR_t = 0.028006$, constant for all t.

4.4.4.2 Dynamic pension funding plan vs. Spread funding plan

We consider the stochastic version of the deterministic comparison with the spread funding formula (3.22) (described in section 3.4.4), investigated in section 4.3.3.2. First of all, it is worth recalling that assumption $crt_t = NR_t$ in (A4) enables us to compare more clearly, in a mathematical form, our pre-optimal funding formula before deciding on the value θ^*, (i.e. equation (4.23) for the complete state information case and equation (4.34) for the incomplete state information case) with the spread funding formula (3.22).

Then, applying the control target assumption in (A4) to the pre-optimal funding formulae (4.23) and (4.34), we can transform each of them into a form distinguishable from the spread funding formula (3.22): for all $t\in[0, T\text{-}1]$,

$$^{C}\pi'_t(SR_t) = -[^{C}D_1(t;\theta)/^{C}D_3(t;\theta)]\cdot[SR_t - 1] + [^{C}D_2(t;\theta) - {}^{C}D_1(t;\theta)] / {}^{C}D_3(t;\theta)$$

$$= NR_t - {}^{C}\varphi(t;\theta)\cdot(SR_k - 1) + {}^{C}\xi(t;\theta), \text{ and hence}$$

$$SR'_{t+1} = \exp(\phi_{t+1})\cdot[(1-{}^{C}\varphi(t;\theta))\cdot SR'_t + \exp(\alpha+\beta-\tau_t-\eta) - \exp(-\tau_t) + {}^{C}\varphi(t;\theta) + {}^{C}\xi(t;\theta)] \text{ with}$$

the initial condition $SR'_0 = SR_0$; and --- (4.50)

$$^{I}\pi'_t(\hat{S}R_t) = -[^{I}D_1(t;\theta)/^{I}D_3(t;\theta)]\cdot[\hat{S}R_t - 1] + [^{I}D_2(t;\theta) - {}^{I}D_1(t;\theta)] / {}^{I}D_3(t;\theta)$$

$$= NR_t - {}^{I}\varphi(t;\theta)\cdot(\hat{S}R_t - 1) + {}^{I}\xi(t;\theta), \text{ and hence}$$

$$\hat{S}R'_{t+1} = \exp(\mu_{t+1}+\sigma_a^2/2)\cdot\{[\exp(\phi_t-\mu_t+\sigma_a^2/2)-{}^{I}\varphi(t;\theta)]\cdot\hat{S}R'_t + \exp(\alpha+\beta-\tau_t-\eta) - \exp(-\tau_t) + {}^{I}\varphi(t;\theta)$$

$+ {}^{I}\xi(t;\theta) + \exp(\phi_t)\cdot(EBR_{t-1}-BR_{t-1})\}$ with the initial condition $\hat{S}R'_0 = \hat{S}R_0$ --- (4.51)

where

$^{C}\varphi(t;\theta) = [^{C}D_1(t;\theta)/^{C}D_3(t;\theta)]$, in which $0 < \varphi(t;\theta) < 1$ for all $t\in[0, T\text{-}1]$;

$^{C}\xi(t;\theta) = -[^{C}D_1(t;\theta)\cdot(\exp(\alpha+\beta-\tau_t-\eta)-\exp(-\tau_t)+1) + \exp(\mu_{t+1}+\sigma_a^2/2)\cdot{}^{C}A_2(t+1)/2]/^{C}D_3(t;\theta)$;

$^{I}\varphi(t;\theta) = [^{I}D_1(t;\theta)/^{I}D_3(t;\theta)]$, in which $0 < {}^{I}\varphi(t;\theta) < 1$ for all $t\in[0, T\text{-}1]$; and

$^{I}\xi(t;\theta) = -[^{I}D_1(t;\theta)\cdot(\exp(\alpha+\beta-\tau_t-\eta)-\exp(-\tau_t)+1) + \exp(\mu_{t+1}+\sigma_a^2/2)\cdot{}^{I}A_2(t+1)/2] / {}^{I}D_3(t;\theta)$.

Here, $^{C\text{ and }I}\varphi(t;\theta)$ and $^{C\text{ and }I}\xi(t;\theta)$ has the same meanings as $\varphi(t;\theta)$ and $\xi(t;\theta)$ mentioned in section 4.3.3.2, that is, the proportional state-feedback controlling parameter as in k_t in the spread funding formula (3.22) and the additive controlling parameter. Thus, formulae (4.50) and (4.51) each have the same mathematical form as the spread funding formula (3.22), except for their additive controlling parameters $^{C\text{ and }I}\xi(t;\theta)$.

Using basic calculus, we can derive the following definitive characteristics about $^{C}\varphi(t;\theta)$ and $^{C}\xi(t;\theta)$: for all $t\in[0, T\text{-}1]$,

(a) ${}^{C}\varphi(t;\theta)$ is a strictly increasing function of both θ and σ_a^2 because ${}^{C}A_1(T) = e^{-\eta T}$ and ${}^{C}A_1(t)$ is a positive and strictly increasing function of both θ and σ_a^2;

(b) ${}^{I}\varphi(t;\theta)$ is a strictly increasing function of both θ and σ_a^2 because ${}^{I}A_1(T) = e^{-\eta T}\cdot\exp(\sigma_a^2)$ and ${}^{I}A_1(t)$ is a positive and strictly increasing function of both θ and σ_a^2; and additionally as investigated in section 4.3.3.2,

as $\theta \to 100\%$, then

(c) ${}^{C}\varphi(t;\theta) \to 1$ and ${}^{C}\xi(t;\theta) \to \exp(-\mu_{t+1}-3/2\sigma_a^2) + \exp(-\tau_t) - \exp(\alpha+\beta-\tau_t-\eta) - 1$ because ${}^{C}A_1(t) \to e^{-\eta t}$ and ${}^{C}A_2(t) \to -2e^{-\eta t}$ and hence,

(d) $SR'_{t+1} \to \exp(\phi_{t+1}-\mu_{t+1}-3/2\sigma_a^2)$ and ${}^{C}CR'_{t+1}-NR_{t+1} \to [\exp(-\mu_{t+1}-3/2\sigma_a^2)\cdot(1-\exp(\phi_{t+1})) + \exp(-\tau_t) - \exp(\alpha+\beta-\tau_t-\eta)]$, whereas $CR'_0-NR_0 \to [-SR'_0 + \exp(-\mu_1-3/2\sigma_a^2) + \exp(-\tau_0) - \exp(\alpha+\beta-\tau_0-\eta)]$ where $SR'_0 = SR_0$ given.

By using the above assumptions in section 4.4.4.1, the movement of the time-varying parameters ${}^{C\text{ and }I}\varphi(t;\theta)$ and ${}^{C\text{ and }I}\xi(t;\theta)$ for k close to T-1 is illustrated numerically in Table 4.5, subject to $\theta = 50\%$ and $\sigma_a = 30\%$ of $|\mu|$.

From this table, it is obvious that both ${}^{C\text{ and }I}\varphi(t;\theta)$ and ${}^{C\text{ and }I}\xi(t;\theta)$ are almost constant during the initial periods, while during the last few periods, ${}^{C\text{ and }I}\varphi(t;\theta)$ is increasing but ${}^{C\text{ and }I}\xi(t;\theta)$ is decreasing as time t progressed to T-1 (- this may imply that some trade-off between ${}^{C\text{ and }I}\varphi(t;\theta)$ and ${}^{C\text{ and }I}\xi(t;\theta)$ is maintained, as in the numerical illustrations in Table 4.1 for $\varphi(t;\theta)$ and $\xi(t;\theta)$).

4.4.4.3 Numerical illustrations of dynamic pension funding plan

For the reasons mentioned earlier in section 4.4.3, we adopt here a mean and variance approach rather than a simulation approach for the numerical illustrations, and further, the performance comparison between the dynamic pension funding plans for complete state information and incomplete state information is based on the mean-squared error introduced in section 4.4.3.4. All numerical calculations are based on the assumptions in section 4.4.4.1.

(i) Searching for θ^*:

We simply illustrate how to determine the balance point θ^* among the admissible set {90%, 50%, 10%} in relation to the conflicting interests of the employer and trustees (i.e. in accordance with the θ^*-criterion described in subsection (v) of section 4.2.3.2). As noted in the deterministic case [see subsection (i) of section 4.3.3.3], the issues in finding the balance point θ^* would be more crucial in the case of the initial solvency level being far away from the solvency target, even in the stochastic case. Here, we shall content ourselves with dealing with the complete state information case with SR_0 = 0%. The resulting pre-optimal mean and variance projections, $\{SR'_0=SR_0$(given initially), $E(SR'_{t+1}-1)$, $E(CR'_t-NR_t)$: t=0, 1, ..., 11}, and $\{SR'_0=SR_0$ (given initially), $Var(SR'_{t+1})$, $Var(CR'_t)$: t=0, 1, ..., 11}, are visualised in Graphs 4.2.1 and 4.2.2, respectively, subject to σ_a = 30% of $|\mu|$ and $\sqrt{VBR_t}$ = 30% of $|EBR|$ (note that Graph 4.2.2 has different profiles for different θ^* as well as different variabilities, as compared with Graph 4.2.1).

These graphs show that θ^* = 50% is likely to be suitable as a compromise value in the light of the pace of funding and the progress of solvency levels evaluated in terms of their means and variances.

(ii) Mean-variance projections of dynamic pension funding plan:

Simply replacing the arbitrary value of θ with a chosen value θ^* in the pre-optimal control laws (4.50) and (4.51), we have the dynamic pension funding plans $\{^C\pi^*_t(.)$: t=0, 1, ..., 11} and $\{^I\pi^*_t(.)$: t=0, 1, ..., 11}, respectively.

We concentrate on the intermediate case $\theta^* = 50\%$ for the subsequent numerical illustrations. For $\theta^* = 50\%$, the mean-variance optimal projections for the complete state information case, $\{E({}^{C}CR^*_t - NR_t), Var({}^{C}CR^*_t); t=0, 1, ..., 11\}$ and $\{SR^*_0=SR_0$ (given initially), $E(SR^*_{t+1}), Var(SR^*_{t+1}); t=0, 1, ..., 11\}$, are calculated from the equations (4.39)~(4.42) (derived in section 4.4.3.2) and their resulting numerical values are given in Table 4.6.1. And also, the mean-variance optimal projections for the incomplete state information case, $\{E({}^{I}CR^*_t-NR_t), Var({}^{I}CR^*_t); t=0, 1, ..., 11\}$ and $\{\hat{SR}^*_0=SR_0$ (given initially), $E(\hat{SR}^*_{t+1}), Var(\hat{SR}^*_{t+1}); t=0, 1, ..., 11\}$, are calculated from equations (4.44)~(4.47) (derived in section 4.4.3.3) and their numerical values are given in Table 4.6.2. Lastly, in order to compare the performances between ${}^{C}\pi^*_t(.)$ and ${}^{I}\pi^*_t(.)$, we adopt the performance comparison measures specified in section 4.4.3.4, i.e. ${}^{SR}MSE_{t+1}$ defined by the equation (4.48) and ${}^{CR}MSE_t$ defined by the equation (4.49), and these are numerically assessed in Table 4.6.3. Further, each of these tables contains a sensitivity analysis for changes to σ_a and $\sqrt{VBR_t}$.

From Tables 4.6.1~4.6.3, we make the following observations:

(a) <u>the influence of the initial solvency level (SR_0)</u>: the case of $SR_0 = 100\%$ provides a better performance than that of $SR_0=0\%$ in the light of the stability of mean, variance and mean square error [see Tables 4.6.1~4.6.3]. In general, we may say that providing that the initial solvency level is very close to a constant solvency target, our dynamic pension funding plan will play an effective role in stabilising both the contribution rate and solvency level, more quickly than otherwise;

(b) <u>sensitivity analysis for the change of σ_a</u>: the variances, $Var({}^{C}CR^*_t)$, $Var({}^{I}CR^*_t)$, $Var(SR^*_{t+1})$ and $Var(\hat{SR}^*_{t+1})$, all increase with increasing σ_a, while the expectations, $E({}^{C}CR^*_t)$, $E({}^{I}CR^*_t)$, $E(SR^*_{t+1})$ and $E(\hat{SR}^*_{t+1})$, decrease all with increasing σ_a, which implies that there is a trade-off between expectations and variances with respect to increasing σ_a [see Tables 4.6.1 and 4.6.2].

Since this result would be unfavourable to both the trustees and the employer, special care should be taken in investing the scheme fund in the direction of minimising the investment risk σ_a^2;

(c) sensitivity analysis for the change of $\sqrt{VBR_t}$: as we know from expectation formulae (4.39), (4.41), (4.44) and (4.46), the expectations, $E({}^{C}CR^{*}_{t})$, $E({}^{I}CR^{*}_{t})$ and $E(SR^{*}_{t+1})$ and $E(\hat{S}R^{*}_{t+1})$, are all independent of varying $\sqrt{VBR_t}$, while the variances, $Var({}^{C}CR^{*}_{t})$, $Var({}^{I}CR^{*}_{t})$, $Var(SR^{*}_{t+1})$ and $Var(\hat{S}R^{*}_{t+1})$, all increase with increasing $\sqrt{VBR_t}$ but the resulting influence is even lower than that of σ_a [see Tables 4.6.1 and 4.6.2]: for this reason, we can identify σ_a as a more sensitive factor than $\sqrt{VBR_t}$, which is consistent with the view of Haberman (1994) mentioned earlier at the beginning of section 2.1.3; and

(d) performance comparison by mean square error: the mean-squared error ${}^{SR}MSE_{t+1}$ is slowly increasing over time t and then decreases during the last few periods. However, the mean-squared error ${}^{CR}MSE_t$ increases with time t. As expected, both ${}^{SR}MSE_t$ and ${}^{CR}MSE_t$ increase with increasing σ_a and/or $\sqrt{VBR_t}$ [see Table 4.6.3]. This numerical result clearly suggests that in the case that we have to accept inevitably the one-unit time delay in the state information, subsequently we should put great emphasis on minimising σ_a^2 and VBR_t so as to reduce the performance difference between ${}^{C}\pi^{*}_{t}(.)$ and ${}^{I}\pi^{*}_{t}(.)$.

4.4.4.4 Suggestions for reducing the expectation of the risk of insolvency

This section corresponds to section 4.3.3.4. As shown in Tables 4.6.1 and 4.6.2, the projection $\{E(SR^{*}_{t+1}), E(\hat{S}R^{*}_{t+1}); t\in[0, T-1]\}$ starting with the given $SR_0 = 0\%$ is seen to be at a lower level than 100% (i.e. insolvent in relation to the solvency level target 100% in terms of expectation) over most of the control period. In order to provide more confidence in the security of pension schemes (i.e. more than 100% in expected value of solvency level), especially to the trustees/members, the actuary could suggest the same options introduced in section 4.3.3.4: that is,

(a) Increasing the value θ^*;

(b) Increasing the solvency ratio target (i.e. $srt_t > 100\%$);

(c) Readjusting the (long-term) force of valuation interest to be more conservative than the expected force of investment interest: that is, from the stochastic model (4.6), $\delta_{t+1} = \eta + \varepsilon_{t+1} \sim$ IID normal (η, σ_a^2), the force of conservative valuation interest (denoted by η') would be $\eta' < \eta$ (i.e. $\exp(\eta') < \exp(\eta) = 1.06$); or

(d) Combining appropriately the above options (a), (b) and/or (c).

As would be expected, the same arguments as in section 4.3.3.4 will be maintained just by changing the concept of deterministic solvency level with that of expected solvency level. So, the effect of the individual options (a), (b) and (c) is here omitted, but instead we shall illustrate the combined effect of suggestion (d) under three illustrative options, i.e. $\{\theta^* = 90\%$ and $srt_t = 100.5\%\}$ as a combination of suggestions (a) and (b), $\{\theta^* = 90\%$ and $\exp(\eta') = 1.04\}$ as a combination of suggestions (a) and (c), and $\{srt_t = 100.5\%$ and $\exp(\eta') = 1.04\}$ as a combination of suggestions (b) and (c).

Prior to consulting the numerical illustrations, given in Table 4.7, for the above three options, we note that firstly, since the incomplete state information case can be reduced to the complete state information case by means of redefining a new state variable (as seen in section 4.4.2), we consider the effect of each combined options only for the complete state information case; secondly, although we need to determine a new best value of θ according to each of the above three options by using the θ^*-criterion, this is not significant in terms of our further discussion, so we shall regard each value of θ as θ^*; thirdly, we concentrate only on the case of SR_0=0%

because the case of SR_0=100% is shown to be solvent over the whole control period in terms of expectation; and lastly, the calculation basis used for the top values in Table 4.6.1 is considered to be a standard basis for our numerical illustrations, that is, {θ^* = 50%, srt_t = 100%, σ_a = 10% of $|\mu|$ and $\sqrt{VBR_t}$ = 10% of EBR}, so we shall indicate only any differences from the standard basis in Table 4.7.

Comparing Table 4.6.1 with Table 4.7, it is clear that even though the expectation and variance of SR^*_t is increasingly improved by each of the defined three options, the corresponding expectation and variance of ${}^{C}CR^*_t$ is worse than before. This result seems to be unfavourable to the employer. As mentioned in subsection (iv) in section 4.3.3.4, identifying a best (not necessary optimal) combination among all available options for improving the solvency of the scheme, such as suggestion (a), (b) and (c), would then be an interesting subject for further research.

4.4.4.5 Conclusions

We have considered the stochastic version of the numerical illustrations examined for deterministic problems in section 4.3.3. As would be expected, we have very similar conclusions to those described in section 4.3.3.5. That is,

(a) Supposing that the contribution rate target is set by the normal cost ratio (i.e. $crt_t = NR_t$), we derive the dynamic pension funding plan in the form that

${}^{C\ or\ I}CR^*_t = NR_t$ + {Proportional controlling parameter at time t (i.e. ${}^{C\ or\ I}\varphi(\theta^*; t)$)} · {1 - (State variable at time t (i.e. SR_t or $\hat{SR}_t$))} + {Additive controlling parameter at time t (i.e. ${}^{C\ or\ I}\xi(\theta^*; t)$)}

; hence, the term ${}^{C\ or\ I}\xi(\theta^*; t)$ is distinguishable from the spread funding formula (3.22) derived in section 3.4.4 [see Table 4.5];

(b) From the viewpoint of the general funding method, we may regard $NR_t + {}^{C\ or\ I}\varphi(\theta^*; t) \cdot [(SR_t$ or $\hat{S}R_t) - 1]$ as the average terms in ${}^{C\ or\ I}CR^*_t$, to which the new term ${}^{C\ or\ I}\xi(\theta^*; t)$ is added with the aim of avoiding large variations in ${}^{C\ or\ I}CR^*_t$; and

(c) Assuming that a scheme has its own solvency level target with an acceptable lower bound and upper bound, it is sensible to take the following funding policy:

- if the scheme has a solvency level between the acceptable (lower and upper) bounds (which means that there is a small deviation from the target), we can control ${}^{C\ or\ I}CR^*_t$ mainly with adjusting ${}^{C\ or\ I}\varphi(\theta^*; t)$ (rather than with adjusting ${}^{C\ or\ I}\xi(\theta^*; t)$) in order to achieve the target in the near (pre-determined) future;

- if the solvency ratio is outside of the acceptable bounds (which may be caused by a large gap between the actual experience and the actuarial assumptions), we need to adjust ${}^{C\ or\ I}\varphi(\theta^*; t)$ and ${}^{C\ or\ I}\xi(\theta^*; t)$ simultaneously to achieve the target in the near future, according to the trade-off between the two controlling parameters, ${}^{C\ or\ I}\varphi(\theta^*; t)$ and ${}^{C\ or\ I}\xi(\theta^*; t)$; and

- using the funding formula, we can treat surpluses and deficiencies in a different fashion, that is, if we keep the policy for ${}^{C\ or\ I}\varphi(\theta^*; t)$, then ${}^{C\ or\ I}\xi(\theta^*; t)$ should be handled in a different way to achieve the target in the near future.

(d) For improving the confidence in the financial soundness of the pension scheme, the actuary would have one or a combination of the three distinct controllable strategies, i.e. increasing the value θ^*, increasing the solvency target and using a more conservative valuation basis than the best estimate valuation basis;

(e) As illustrated in Table 4.7, the common effect of these options is not only to improve the protection against the expected risk of insolvency but also to lead to additional contributions in return for the higher expectation of solvency level; hence

(f) What combination of these controllable strategies or which one of these controllable strategies is optimal for balancing the additional financial burden on the employer and the better protection against insolvency? This important problem must be left for the subject of future research; and

(g) Finally, we have introduced the performance comparison measures, ${}^{SR}MSE_t$ and ${}^{CR}MSE_t$ defined by formulae (4.48) and (4.49) respectively, which are based on the concept of mean-squared error and the numerical comparisons are illustrated in Table 4.6.3. We believe that this measure would be useful for comparing the incomplete state information control problems with the corresponding complete state information control problems.

Appendix 4A: Numerical illustations for deterministic LQP optimisation problem

A1. (Mathematical comparision with the spread funding formula)

Table 4.1

Pre-optimal funding formula (4.18) with given θ = 50% for t close to T

t	$\pi'_t(SR_t)$ {for each cell, (left, middle, right) values each are given, subject to exp(δ) = (1.04, 1.06, 1.08), respectively}
T-1	NR_{T-1} - (0.64899, 0.65762, 0.66598) · (SR_{T-1} - 1) + (0.01141, - 0.00097, - 0.01321)
T-2	NR_{T-2} - (0.60387, 0.61418, 0.62418) · (SR_{T-2} - 1) + (0.01484, - 0.00126, - 0.01719)
T-3	NR_{T-3} - (0.59721, 0.60787, 0.61820) · (SR_{T-3} - 1) + (0.01608, - 0.00137, - 0.01863)
T-4	NR_{T-4} - (0.59621, 0.60694, 0.61733) · (SR_{T-4} - 1) + (0.01655, - 0.00141, - 0.01916)
T-5	NR_{T-5} - (0.59606, 0.60680, 0.61721) · (SR_{T-5} - 1) + (0.01672, - 0.00142, - 0.01935)
T-6	NR_{T-6} - (0.59604, 0.60678, 0.61719) · (SR_{T-6} - 1) + (0.01679, - 0.00143, - 0.01942)
T-7	NR_{T-7} - (0.59604, 0.60678, 0.61719) · (SR_{T-7} - 1) + (0.01681, - 0.00143, - 0.01945)
T-8	NR_{T-8} - (0.59604, 0.60677, 0.61719) · (SR_{T-8} - 1) + (0.01682, - 0.00143, - 0.01946)
T-9	NR_{T-9} - (0.59604, 0.60677, 0.61719) · (SR_{T-9} - 1) + (0.01683, - 0.00143, - 0.01947)
T-10	NR_{T-10} - (0.59604, 0.60677, 0.61719) · (SR_{T-10} - 1) + (0.01683, - 0.00143, - 0.01947)
T-11	NR_{T-11} - (0.59604, 0.60677, 0.61719) · (SR_{T-11} - 1) + (0.01683, - 0.00143, - 0.01947)
T-12	NR_{T-12} - (0.59604, 0.60677, 0.61719) · (SR_{T-12} - 1) + (0.01683, - 0.00143, - 0.01947)

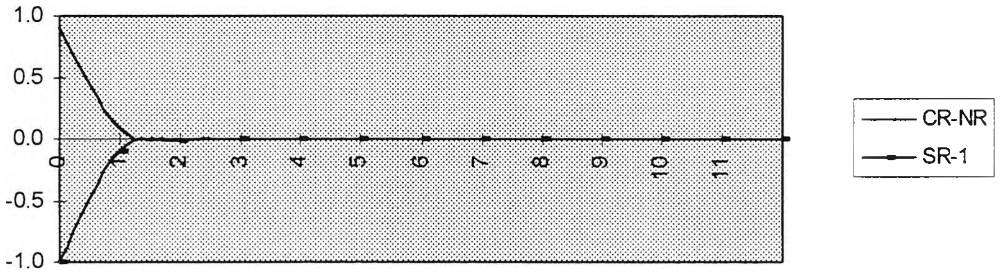
; here, NR_t = 0.02801, constant for all t.

A2. (Illustrating the influence of the value of θ^* on SR^*_t-1 and CR^*_t-NR_t)

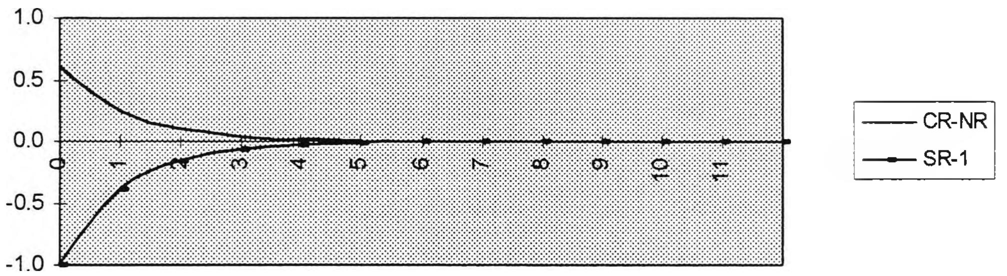
Graph 4.1.1

The time path of SR^*_t-1 and CR^*_t-NR_t for exp(δ) = 1.06 and SR_0 = 0%

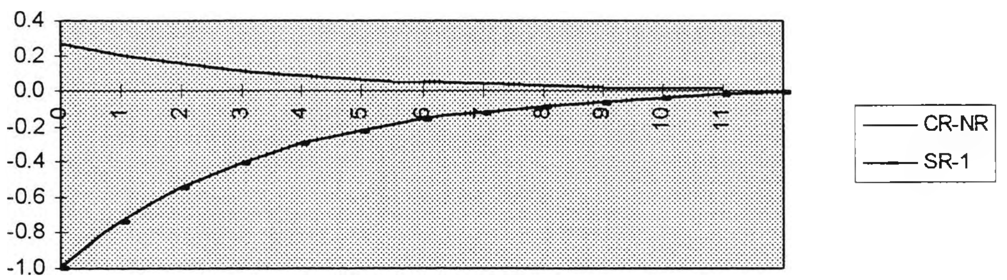
- In a case of θ^* = 90%:



- In a case of θ^* = 50%:



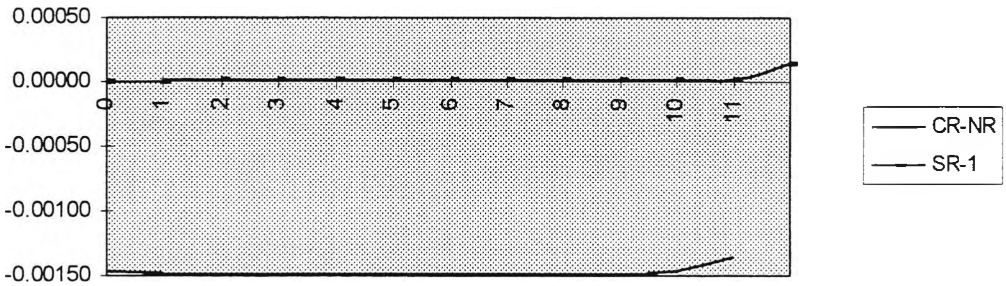
- In a case of θ^* = 10%:



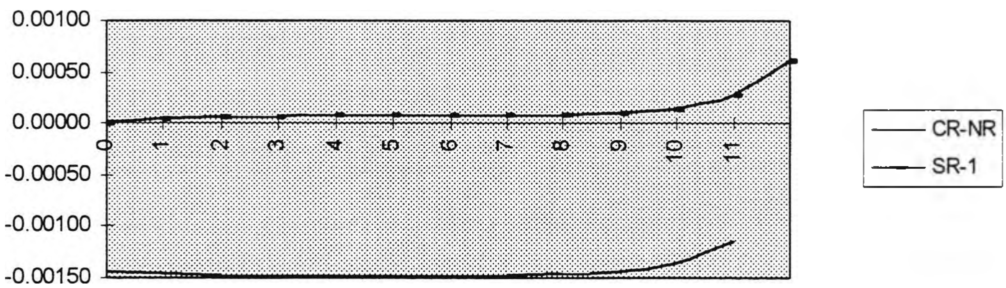
Graph 4.1.2

The time path of SR^*_t-1 and CR^*_t-NR_t for exp(δ) = 1.06 and SR_0 = 100%

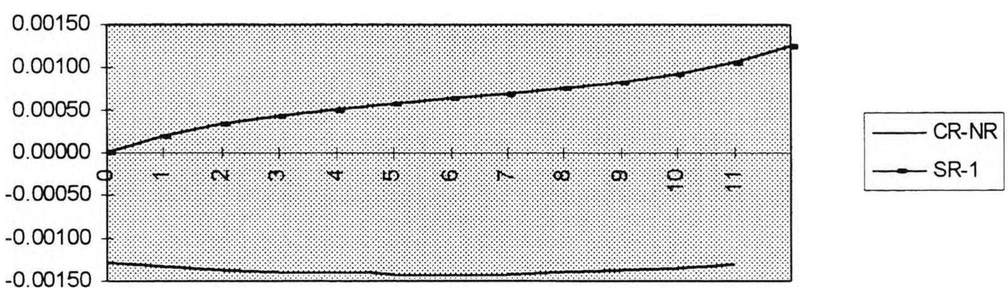
- In a case of θ^* = 90%:



- In a case of θ^* = 50%:



- In a case of θ^* = 10%:



A3. (Comparision in relation to SR_0)

Table 4.2.1

Projections of the dynamic pension funding plan governed by formula (4.19)

t	$\{\theta^* = 50\%,\ \exp(\delta) = 1.06,\ SR_0=100\%\}$		$\{\theta^* = 50\%,\ \exp(\delta) = 1.06,\ SR_0 = 0\%\}$	
	CR^*_t - NR_t	SR^*_t	CR^*_t - NR_t	SR^*_t
0	- 0.001433	1.0	0.605341	0.0
1	- 0.001461	1.000045	0.239273	0.603302
2	- 0.001471	1.000063	0.094038	0.842658
3	- 0.001476	1.000070	0.036417	0.937621
4	- 0.001477	1.000073	0.013557	0.975297
5	- 0.001477	1.000075	0.004487	0.990245
6	- 0.001476	1.000076	0.000890	0.996176
7	- 0.001472	1.000079	- 0.000533	0.998531
8	- 0.001462	1.000085	- 0.001089	0.999471
9	- 0.001433	1.000103	- 0.001285	0.999859
10	- 0.001356	1.000149	- 0.001297	1.000053
11	- 0.001152	1.000273	- 0.001127	1.000236
12	_	1.000605	_	1.000592

; here, NR_t = 0.028006, constant for all t.

A4. (For comparision with Table 4.2.1 in view of better investment performance)

Table 4.2.2

Projections of the dynamic pension funding plan governed by formula (4.19)

t	$\{\theta^* = 50\%,\ \exp(\delta) = 1.08,\ SR_0 = 100\%\}$		$\{\theta^* = 50\%,\ \exp(\delta) = 1.08,\ SR_0 = 0\%\}$	
	CR^*_t - NR_t	SR^*_t	CR^*_t - NR_t	SR^*_t
0	- 0.019467	1.0	0.597718	0.0
1	- 0.019699	1.000375	0.223181	0.606847
2	- 0.019789	1.000523	0.075790	0.845659
3	- 0.019824	1.000581	0.017789	0.939638
4	- 0.019836	1.000606	- 0.005034	0.976623
5	- 0.019834	1.000619	- 0.014009	0.991182
6	- 0.019817	1.000635	- 0.017524	0.996921
7	- 0.019765	1.000669	- 0.018863	0.999207
8	- 0.019623	1.000757	- 0.019268	1.000182
9	- 0.019241	1.000994	- 0.019101	1.000767
10	- 0.018210	1.001630	- 0.018155	1.001541
11	- 0.015434	1.003343	- 0.015411	1.003309
12	_	1.007958	_	1.007946

; here, NR_t = 0.028006, constant for all t.

A5. (Some illustrative effects of suggestion (a) in relation to Table 4.2.1)

Table 4.3.1

Readjusted projections of Table 4.2.1 under suggestion (a)

	In the case of switching from $\theta^* = 50.0\%$ to $\theta^* = 99.9\%$			
t	CR^*_t - NR_t	SR^*_t	CR^*_t - NR_t	SR^*_t
0	- 0.001478	1.0	0.997482	0.0
1	- 0.001478	1.000000	- 0.000430	0.998951
2	- 0.001478	1.000000	- 0.001477	0.999999
3	- 0.001478	1.000000	- 0.001478	1.000000
4	- 0.001478	1.000000	- 0.001478	1.000000
5	- 0.001478	1.000000	- 0.001478	1.000000
6	- 0.001478	1.000000	- 0.001478	1.000000
7	- 0.001478	1.000000	- 0.001478	1.000000
8	- 0.001478	1.000000	- 0.001478	1.000000
9	- 0.001478	1.000000	- 0.001478	1.000000
10	- 0.001478	1.000000	- 0.001478	1.000000
11	- 0.001477	1.000000	- 0.001477	1.000000
12	_	1.000002	_	1.000002

; here, NR_t = 0.028006, constant for all t.

A6. (Some illustrative effects of suggestion (b) in relation to Table 4.2.1)

Table 4.3.2

Readjusted projections of Table 4.2.1 under suggestion (b)

t	In the case of switching from $srt_t = 100.0\%$ to $srt_t = 100.5\%$			
	CR^*_t - NR_t	SR^*_t	CR^*_t - NR_t	SR^*_t
0	0.001558	1.0	0.608332	0.0
1	- 0.000301	1.003063	0.240433	0.606319
2	- 0.001038	1.004278	0.094472	0.846872
3	- 0.001330	1.004760	0.036562	0.942310
4	- 0.001446	1.004951	0.013587	0.980175
5	- 0.001492	1.005028	0.004473	0.995198
6	- 0.001509	1.005059	0.000858	1.001159
7	- 0.001512	1.005073	- 0.000573	1.003526
8	- 0.001503	1.005085	- 0.001131	1.004471
9	- 0.001475	1.005104	- 0.001327	1.004861
10	- 0.001397	1.005153	- 0.001337	1.005057
11	- 0.001186	1.005281	- 0.001161	1.005244
12	_	1.005623	_	1.005610

: here, NR_t = 0.028006, constant for all t.

A7. (<u>Some illustrative effects of suggestion (c) in relation to Table 4.2.1</u>)

Table 4.3.3

Readjusted projections of Table 4.2.1 under suggestion (c)

t	In the case of switching from $\exp(\eta) = 1.06$ to $\exp(\eta) = 1.04$			
t	CR^*_t - NR_t	SR^*_t	CR^*_t - NR_t	SR^*_t
0	- 0.017039	1.0	0.595057	0.0
1	- 0.017238	1.000325	0.222321	0.608950
2	- 0.017316	1.000453	0.076442	0.847278
3	- 0.017345	1.000503	0.019349	0.940555
4	- 0.017354	1.000525	- 0.002993	0.977062
5	- 0.017352	1.000536	- 0.011731	0.991354
6	- 0.017335	1.000551	- 0.015136	0.996957
7	- 0.017287	1.000582	- 0.016426	0.999175
8	- 0.017157	1.000662	- 0.016820	1.000112
9	- 0.016810	1.000875	- 0.016678	1.000660
10	- 0.015888	1.001440	- 0.015836	1.001355
11	- 0.013437	1.002939	- 0.013416	1.002907
12	_	1.006925	_	1.006914

; here, for $\exp(\eta) = 1.06$, $NR_t = 0.028006$ constant for all t; and for for $\exp(\eta) = 1.04$, $NR_t = 0.034506$ constant for all t.

A8. (Some illustrative effects of suggestion (d) in relation to Table 4.2.1)

Table 4.4.4

Readjusted projections of Table 4.2.1 under suggestion (d)

t	Combination of suggestions (a) and (b): switching from (θ^*=50%, srt_t = 100%) to (θ^*=90%, srt_t = 100.5%)		Combination of suggestions (a) and (c): switching from (θ^*=50%, $\exp(\eta)$=1.06) to (θ^*=90%, $\exp(\eta)$=1.04)		Combination of suggestions (b) and (c): switching from (srt_t=100%, $\exp(\eta)$=1.06) to (srt_t=100.5%, $\exp(\eta)$=1.04)	
t	$CR^*_t - NR_t$	SR^*_t	$CR^*_t - NR_t$	SR^*_t	$CR^*_t - NR_t$	SR^*_t
0	0.907882	0.0	0.889205	0.0	0.598073	0.0
1	0.085761	0.908549	0.068148	0.905731	0.223475	0.611994
2	0.006855	0.995750	- 0.009296	0.991162	0.076866	0.851514
3	- 0.000718	1.004120	- 0.016601	0.999220	0.019488	0.945256
4	- 0.001445	1.004923	- 0.017290	0.999980	- 0.002966	0.981946
5	- 0.001515	1.005000	- 0.017355	1.000052	- 0.011748	0.996309
6	- 0.001522	1.005008	- 0.017361	1.000059	- 0.015169	1.001940
7	- 0.001522	1.005008	- 0.017362	1.000059	- 0.016466	1.004170
8	- 0.001522	1.005009	- 0.017361	1.000059	- 0.016862	1.005111
9	- 0.001521	1.005009	- 0.017350	1.000061	- 0.016720	1.005661
10	- 0.001511	1.005010	- 0.017233	1.000072	- 0.015876	1.006359
11	- 0.001398	1.005021	- 0.015936	1.000203	- 0.013450	1.007915
12	_	1.005147	_	1.001643	_	1.011932

; here, for $\exp(\eta)$ = 1.06, NR_t = 0.028006 constant for all t; and for for $\exp(\eta)$ = 1.04, NR_t = 0.034506 constant for all t.

Appendix 4B: Numerical illustations for stochastic LQP optimisation problem

B1. (Mathematical comparision with the spread funding formula)

Table 4.5

Pre-optimal funding formulae (4.50) and (4.51) for t close to T-1

k	$^{C}\pi'_t(SR_t)$ for $\theta = 50\%$ and $\sigma_a = 30\%$ of $\lvert\mu\rvert$	$^{I}\pi'_t(\hat{SR}_t)$ for $\theta = 50\%$ and $\sigma_a = 30\%$ of $\lvert\mu\rvert$
T-1	$NR_{T-1} - 0.657620 \cdot (\hat{SR}_{T-1} - 1) - 0.000979$	$NR_{T-1} - 0.657620 \cdot (\hat{SR}_{T-1} - 1) - 0.000979$
T-2	$NR_{T-2} - 0.614186 \cdot (\hat{SR}_{T-2} - 1) - 0.001274$	$NR_{T-2} - 0.614188 \cdot (\hat{SR}_{T-2} - 1) - 0.001275$
T-3	$NR_{T-3} - 0.607876 \cdot (\hat{SR}_{T-3} - 1) - 0.001380$	$NR_{T-3} - 0.607877 \cdot (\hat{SR}_{T-3} - 1) - 0.001382$
T-4	$NR_{T-4} - 0.606942 \cdot (\hat{SR}_{T-4} - 1) - 0.001420$	$NR_{T-4} - 0.606943 \cdot (\hat{SR}_{T-4} - 1) - 0.001421$
T-5	$NR_{T-5} - 0.606803 \cdot (\hat{SR}_{T-5} - 1) - 0.001435$	$NR_{T-5} - 0.606804 \cdot (\hat{SR}_{T-5} - 1) - 0.001436$
T-6	$NR_{T-6} - 0.606782 \cdot (\hat{SR}_{T-6} - 1) - 0.001440$	$NR_{T-6} - 0.606783 \cdot (\hat{SR}_{T-6} - 1) - 0.001442$
T-7	$NR_{T-7} - 0.606779 \cdot (\hat{SR}_{T-7} - 1) - 0.001442$	$NR_{T-7} - 0.606780 \cdot (\hat{SR}_{T-7} - 1) - 0.001444$
T-8	$NR_{T-8} - 0.606779 \cdot (\hat{SR}_{T-8} - 1) - 0.001443$	$NR_{T-8} - 0.606780 \cdot (\hat{SR}_{T-8} - 1) - 0.001444$
T-9	$NR_{T-9} - 0.606779 \cdot (\hat{SR}_{T-9} - 1) - 0.001443$	$NR_{T-9} - 0.606780 \cdot (\hat{SR}_{T-9} - 1) - 0.001445$
T-10	$NR_{T-10} - 0.606779 \cdot (\hat{SR}_{T-10} - 1) - 0.001443$	$NR_{T-10} - 0.606780 \cdot (\hat{SR}_{T-10} - 1) - 0.001445$
T-11	$NR_{T-11} - 0.606779 \cdot (\hat{SR}_{T-11} - 1) - 0.001444$	$NR_{T-11} - 0.606780 \cdot (\hat{SR}_{T-11} - 1) - 0.001445$
T-12	$NR_{T-12} - 0.606779 \cdot (\hat{SR}_{T-12} -1) - 0.001444$	$NR_{T-12} - 0.60678 \cdot (\hat{SR}_{T-12} - 1) - 0.001445$

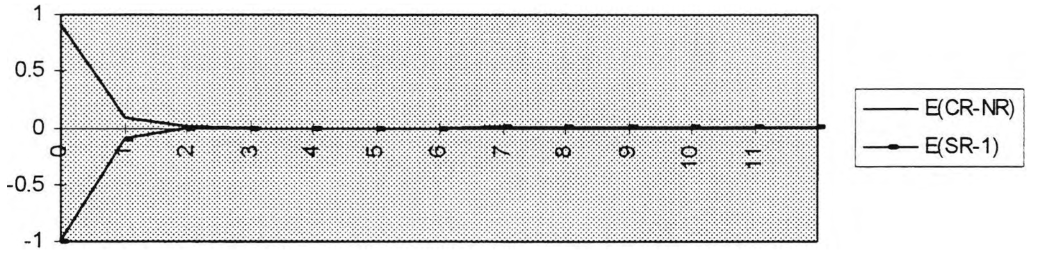
; here, $NR_t = 0.028006$, constant for all t.

B2. (Searching for the value θ^* in accordance with the θ^*-criterion)

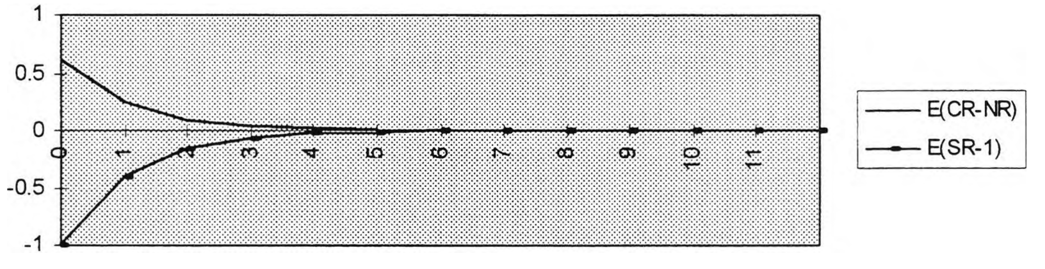
Graph 4.2.1

The time path of $E(SR^*_t\text{-}1)$ and $E(^CCR^*_t\text{-}NR_t)$, subject to $\exp(\delta)=1.06$, σ_a = 30% of $|\mu|$ $\sqrt{VBR_t}$ = 30% of |EBR| and SR_0=0%

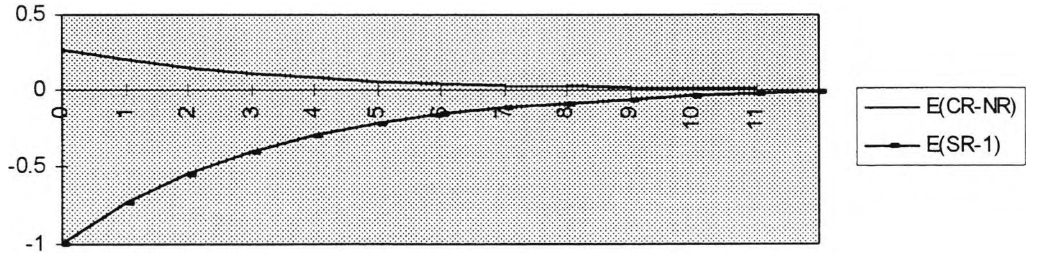
- In a case of θ^* = 90%:



- In a case of θ^* = 50%:



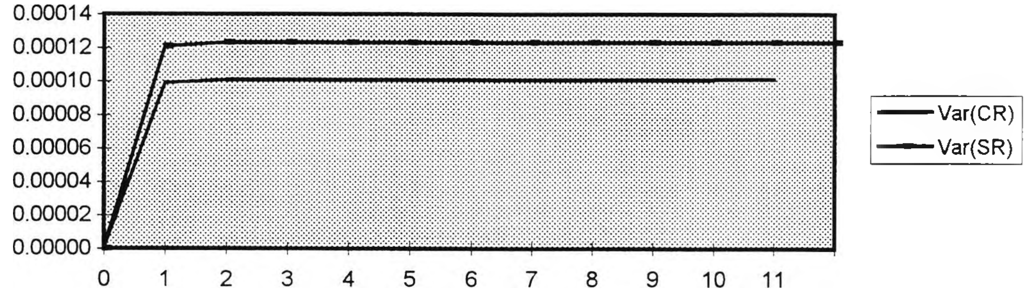
- In a case of θ^* = 10%:



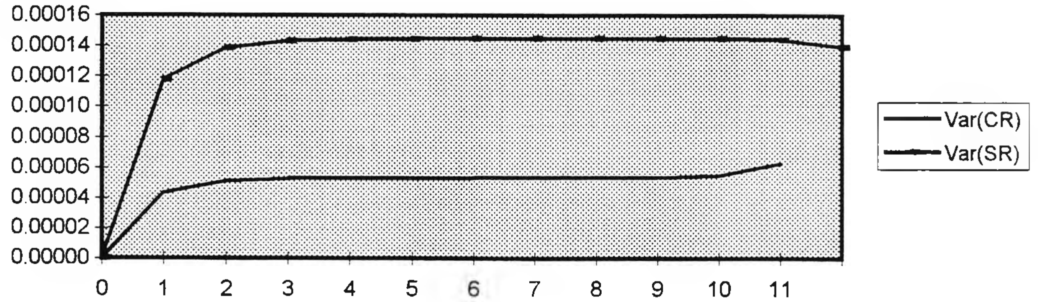
Graph 4.2.2

The time path of $Var(SR^*_t)$ and $Var(^CCR^*_t)$ for $\exp(\delta)=1.06$ σ_a = 30% of $|\mu|$, $\sqrt{VBR_t}$ = 30% of $|EBR|$ and SR_0=0%

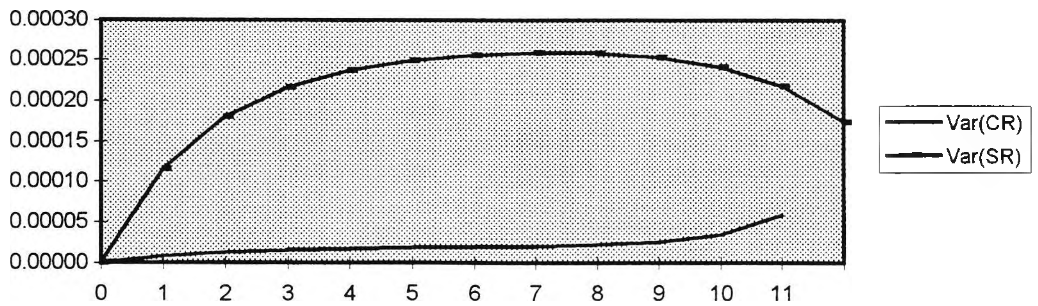
- In a case of θ^* = 90%:



- In a case of θ^* = 50%:



- In a case of θ^* = 10%:



B3. (Sensitivity analysis with respect to σ_a and $\sqrt{VBR_t}$)

Table 4.6.1

Mean and variance of the optimal projections $\{{}^{C}CR^{*}_{t}\}$ and $\{SR^{*}_{t}\}$

t	For each cell, top values subject to { $\theta^* = 50\%$, $\sigma_a = 10\%$ of $\lvert\mu\rvert$ and $\sqrt{VBR_t} = 10\%$ of $\lvert EBR\rvert$ }, middle values subject to { $\theta^* = 50\%$, $\sigma_a = 30\%$ of $\lvert\mu\rvert$ and $\sqrt{VBR_t} = 10\%$ of $\lvert EBR\rvert$ } and bottom values subject to { $\theta^* = 50\%$, $\sigma_a = 30\%$ of $\lvert\mu\rvert$ and $\sqrt{VBR_t} = 30\%$ of $\lvert EBR\rvert$ }							
	$E({}^{c}CR^{*}_{t} - NR_t)$	$(Var({}^{c}CR^{*}_{t}))$	$E(SR^{*}_{t})$	$(Var(SR^{*}_{t}))$	$E({}^{c}CR^{*}_{t} - NR_t)$	$(Var({}^{c}CR^{*}_{t}))$	$E(SR^{*}_{t})$	$(Var(SR^{*}_{t}))$
0	-.001434	(0.0)	1.0	(0.0)	0.605341	(0.0)	0.0	(0.0)
	-.001444	(0.0)	1.0	(0.0)	0.605335	(0.0)	0.0	(0.0)
	-.001444	(0.0)	1.0	(0.0)	0.605335	(0.0)	0.0	(0.0)
1	-.001461	(4.99E-06)	1.000044	(1.35E-05)	0.239273	(4.80E-06)	0.603301	(1.30E-05)
	-.001467	(7.33E-06)	1.000038	(1.99E-05)	0.239267	(5.65E-06)	0.603297	(1.54E-05)
	-.001467	(4.49E-05)	1.000038	(12.2E-05)	0.239267	(4.32E-05)	0.603297	(11.7E-05)
2	-.001472	(5.77E-06)	1.000062	(1.57E-05)	0.094038	(5.66E-06)	0.842657	(1.54E-05)
	-.001476	(8.48E-06)	1.000054	(2.30E-05)	0.094033	(7.45E-06)	0.842650	(2.02E-05)
	-.001476	(5.20E-05)	1.000054	(14.1E-05)	0.094033	(5.09E-05)	0.842650	(13.8E-05)
3	-.001476	(5.90E-06)	1.000069	(1.60E-05)	0.036417	(5.84E-06)	0.937620	(1.59E-05)
	-.001479	(8.66E-06)	1.000060	(2.35E-05)	0.036413	(8.18E-06)	0.937611	(2.22E-05)
	-.001479	(5.31E-05)	1.000060	(14.4E-05)	0.036413	(5.26E-05)	0.937611	(14.3E-05)
4	-.001478	(5.92E-06)	1.000072	(1.61E-05)	0.013556	(5.89E-06)	0.975296	(1.60E-05)
	-.001481	(8.69E-06)	1.000062	(2.36E-05)	0.013553	(8.48E-06)	0.975286	(2.30E-05)
	-.001481	(5.32E-05)	1.000062	(14.5E-05)	0.013553	(5.30E-05)	0.975286	(14.4E-05)
5	-.001478	(5.92E-06)	1.000073	(1.61E-05)	0.004487	(5.91E-06)	0.990243	(1.61E-05)
	-.001481	(8.68E-06)	1.000063	(2.36E-06)	0.004484	(8.61E-06)	0.990234	(2.34E-05)
	-.001481	(5.33E-05)	1.000063	(14.5E-05)	0.004484	(5.32E-05)	0.990234	(14.4E-05)
6	-.001476	(5.92E-06)	1.000075	(1.61E-05)	0.000890	(5.92E-06)	0.996175	(1.61E-05)
	-.001479	(8.69E-06)	1.000065	(2.36E-05)	0.000887	(8.66E-06)	0.996165	(2.35E-05)
	-.001479	(5.33E-05)	1.000065	(14.5E-05)	0.000887	(5.32E-05)	0.996165	(14.5E-05)
7	-.001472	(5.92E-06)	1.000077	(1.61E-05)	-.000534	(5.92E-06)	0.998530	(1.61E-05)
	-.001476	(8.69E-06)	1.000067	(2.36E-05)	-.000537	(8.68E-06)	0.998520	(2.36E-05)
	-.001476	(5.33E-05)	1.000067	(14.5E-05)	-.000537	(5.33E-05)	0.998520	(14.5E-05)

8	-.001462 (5.92E-06)	1.000084 (1.61E-05)	-.001089 (5.92E-06)	0.999470 (1.61E-05)
	-.001465 (8.70E-06)	1.000074 (2.36E-05)	-.001092 (8.69E-06)	0.999460 (2.36E-05)
	-.001465 (5.33E-05)	1.000074 (14.5E-05)	-.001092 (5.33E-05)	0.999460 (14.5E-05)
9	-.001433 (5.94E-06)	1.000101 (1.61E-05)	-.001285 (5.94E-06)	0.999858 (1.61E-05)
	-.001436 (8.72E-06)	1.000091 (2.36E-05)	-.001288 (8.72E-06)	0.999848 (2.36E-05)
	-.001436 (5.33E-05)	1.000091 (14.5E-05)	-.001288 (5.35E-05)	0.999848 (14.5E-05)
10	-.001356 (6.06E-06)	1.000148 (1.61E-05)	-.001297 (6.06E-06)	1.000052 (1.61E-05)
	-.001359 (8.90E-06)	1.000138 (2.36E-05)	-.001300 (8.90E-06)	1.000042 (2.36E-05)
	-.001359 (5.45E-05)	1.000138 (14.5E-05)	-.001300 (5.45E-05)	1.000042 (14.5E-05)
11	-.001152 (6.91E-06)	1.000272 (1.60E-05)	-.001127 (6.91E-06)	1.000235 (1.60E-05)
	-.001152 (1.02E-05)	1.000263 (2.35E-05)	-.001127 (1.02E-05)	1.000226 (2.35E-05)
	-.001152 (6.22E-05)	1.000263 (14.4E-05)	-.001127 (6.22E-05)	1.000226 (14.4E-05)
12	–	1.000604 (1.55E-05)	–	1.000591 (1.55E-05)
		1.000598 (2.27E-05)		1.000585 (2.27E-05)
		1.000598 (13.9E-05)		1.000585 (13.9E-05)

; here, NR_t = 0.028006, constant for all t.

B4. (Sensitivity analysis with respect to σ_a and $\sqrt{VBR_t}$)

Table 4.6.2

Mean and variance of the optimal projections $\{{}^{I}CR^{*}_{t}\}$ and $\{\hat{SR}^{*}_{t}\}$

t	For each cell, top values subject to { θ^{*} = 50%, σ_a = 10% of $\|\mu\|$ and $\sqrt{VBR_t}$ = 10% of $\|EBR\|$ }, middle values subject to { θ^{*} = 50%, σ_a = 30% of $\|\mu\|$ and $\sqrt{VBR_t}$ = 10% of $\|EBR\|$ } and bottom values subject to { θ^{*} = 50%, σ_a = 30% of $\|\mu\|$ and $\sqrt{VBR_t}$ = 30% of $\|EBR\|$ }							
	$E({}^{I}CR^{*}_{t}-NR_t)$	$(Var({}^{I}CR^{*}_{t}))$	$E(\hat{SR}^{*}_{t})$	$(Var(\hat{SR}^{*}_{t}))$	$E({}^{I}CR^{*}_{t}-NR_t)$	$(Var({}^{I}CR^{*}_{t}))$	$E(\hat{SR}^{*}_{t})$	$(Var(\hat{SR}^{*}_{t}))$
0	-.001435	(0.0)	1.0	(0.0)	0.605341	(0.0)	0.0	(0.0)
	-.001445	(0.0)	1.0	(0.0)	0.605335	(0.0)	0.0	(0.0)
	-.001445	(0.0)	1.0	(0.0)	0.605335	(0.0)	0.0	(0.0)
1	-.001461	(5.08E-06)	1.000044	(1.38E-05)	0.239273	(4.78E-06)	0.603301	(1.30E-05)
	-.001467	(7.46E-06)	1.000037	(2.03E-05)	0.239266	(4.78E-06)	0.603297	(1.30E-05)
	-.001467	(45.7E-06)	1.000037	(12.4E-05)	0.239266	(43.0E-06)	0.603297	(11.7E-05)
2	-.001472	(5.88E-06)	1.000062	(1.60E-05)	0.094038	(5.64E-06)	0.842657	(1.53E-05)
	-.001476	(8.63E-06)	1.000052	(2.34E-05)	0.094032	(6.51E-06)	0.842649	(1.77E-05)
	-.001476	(52.9E-06)	1.000052	(14.4E-05)	0.094032	(50.8E-06)	0.842649	(13.8E-05)
3	-.001476	(6.00E-06)	1.000069	(1.63E-05)	0.036417	(5.88E-06)	0.937619	(1.60E-05)
	-.001480	(8.82E-06)	1.000057	(2.39E-05)	0.036412	(7.70E-06)	0.937610	(2.09E-05)
	-.001480	(54.0E-06)	1.000057	(14.7E-05)	0.036412	(52.9E-06)	0.937610	(14.4E-05)
4	-.001478	(6.02E-06)	1.000072	(1.64E-05)	0.013556	(5.97E-06)	0.975295	(1.62E-05)
	-.001481	(8.84E-06)	1.000060	(2.40E-05)	0.013553	(8.35E-06)	0.975284	(2.27E-05)
	-.001481	(54.2E-06)	1.000060	(14.7E-05)	0.013553	(53.7E-06)	0.975284	(14.6E-05)
5	-.001478	(6.03E-06)	1.000073	(1.64E-05)	0.004487	(6.00E-06)	0.990243	(1.63E-05)
	-.001481	(8.85E-06)	1.000061	(2.40E-05)	0.004484	(8.64E-06)	0.990232	(2.35E-05)
	-.001481	(54.2E-06)	1.000061	(14.7E-05)	0.004484	(54.0E-06)	0.990232	(14.7E-05)
6	-.001476	(6.03E-06)	1.000074	(1.64E-05)	0.000890	(6.02E-06)	0.996175	(1.63E-05)
	-.001479	(8.85E-06)	1.000062	(2.40E-05)	0.000887	(8.76E-06)	0.996163	(2.38E-05)
	-.001479	(54.2E-06)	1.000062	(14.7E-05)	0.000887	(54.2E-06)	0.996163	(14.7E-05)
7	-.001473	(6.03E-06)	1.000077	(1.64E-05)	-.000534	(6.02E-06)	0.998530	(1.64E-05)
	-.001476	(8.85E-06)	1.000065	(2.40E-05)	-.000537	(8.82E-06)	0.998518	(2.39E-05)
	-.001476	(54.2E-06)	1.000065	(14.7E-05)	-.000537	(54.2E-06)	0.998518	(14.7E-05)

8	-.001462 (6.03E-06)	1.000084 (1.64E-05)	-.001089 (6.03E-06)	0.999470 (1.64E-05)
	-.001465 (8.85E-06)	1.000072 (2.40E-05)	-.001092 (8.84E-06)	0.999458 (2.40E-05)
	-.001465 (54.3E-06)	1.000072 (14.7E-05)	-.001092 (54.3E-06)	0.999458 (14.7E-05)
9	-.001433 (6.05E-06)	1.000101 (1.64E-05)	-.001285 (6.05E-06)	0.999858 (1.64E-05)
	-.001436 (8.88E-06)	1.000089 (2.40E-05)	-.001288 (8.88E-06)	0.999846 (2.40E-05)
	-.001436 (54.4E-06)	1.000089 (14.7E-05)	-.001288 (54.4E-06)	0.998846 (14.7E-05)
10	-.001357 (6.17E-06)	1.000148 (1.64E-05)	-.001297 (6.17E-06)	1.000051 (1.64E-05)
	-.001359 (9.06E-06)	1.000136 (2.40E-05)	-.001300 (9.06E-06)	1.000039 (2.40E-05)
	-.001359 (55.5E-06)	1.000136 (14.7E-05)	-.001300 (55.5E-06)	1.000039 (14.7E-05)
11	-.001151 (7.04E-06)	1.000272 (1.63E-05)	-.001127 (7.04E-06)	1.000234 (1.63E-05)
	-.001150 (10.3E-06)	1.000261 (2.39E-05)	-.001126 (10.3E-06)	1.000223 (2.39E-05)
	-.001150 (63.3E-06)	1.000261 (14.6E-05)	-.001126 (63.3E-06)	1.000223 (14.6E-05)
12	–	1.000604 (1.57E-05)	–	1.000591 (1.57E-05)
		1.000597 (2.31E-05)		1.000584 (2.31E-05)
		1.000597 (14.2E-05)		1.000584 (14.2E-05)

; here, NR_t = 0.028006, constant for all t.

B5. (Sensitivity analysis with respect to σ_a and $\sqrt{VBR_t}$)

Table 4.6.3

Performance comparision between ${}^{C}\pi^{*}_{t}(.)$ and ${}^{I}\pi^{*}_{t}(.)$ by means of mean-squared error

t	For each cell, top values subject to { $\theta^* = 50\%$, $\sigma_a = 10\%$ of $\lvert\mu\rvert$ and $\sqrt{VBR_t} = 10\%$ of $\lvert EBR\rvert$ }, middle values subject to { $\theta^* = 50\%$, $\sigma_a = 30\%$ of $\lvert\mu\rvert$ and $\sqrt{VBR_t} = 10\%$ of $\lvert EBR\rvert$ } and bottom values subject to { $\theta^* = 50\%$, $\sigma_a = 30\%$ of $\lvert\mu\rvert$ and $\sqrt{VBR_t} = 30\%$ of $\lvert EBR\rvert$ }			
	$E[({}^{C}CR^{*}_{t} - {}^{I}CR_{t})^2]$	$E[(SR^{*}_{t} - \hat{S}R^{*}_{t})^2]$	$E[({}^{C}CR^{*}_{t} - {}^{I}CR^{*}_{t})^2]$	$E[(SR^{*}_{t} - \hat{S}R^{*}_{t})^2]$
0	2.39E-14 1.93E-12 1.93E-12	0.0 ($SR^*_0 = \hat{S}R^*_0 = 1.0$) 0.0 ($SR^*_0 = \hat{S}R^*_0 = 1.0$) 0.0 ($SR^*_0 = \hat{S}R^*_0 = 1.0$)	6.66E-16 4.95E-14 4.95E-14	0.0 ($SR^*_0 = \hat{S}R^*_0 = 0.0$) 0.0 ($SR^*_0 = \hat{S}R^*_0 = 0.0$) 0.0 ($SR^*_0 = \hat{S}R^*_0 = 0.0$)
1	1.01E-05 1.48E-05 9.06E-05	2.73E-05 4.01E-05 2.46E-04	9.58E-06 1.04E-05 8.62E-05	2.60E-05 2.83E-05 2.34E-04
2	1.16E-05 1.71E-05 1.05E-04	3.16E-05 4.65E-05 2.85E-04	1.13E-05 1.40E-05 1.02E-04	3.07E-05 3.79E-05 2.76E-04
3	1.19E-05 1.75E-05 1.07E-04	3.23E-05 4.75E-05 2.91E-04	1.17E-05 1.59E-05 1.06E-04	3.18E-05 4.31E-05 2.87E-04
4	1.19E-05 1.75E-05 1.07E-04	3.24E-05 4.76E-05 2.92E-04	1.19E-05 1.68E-05 1.07E-04	3.22E-05 4.57E-05 2.90E-04
5	1.19E-05 1.75E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04	1.19E-05 1.72E-05 1.07E-04	3.24E-05 4.68E-05 2.91E-04
6	1.19E-05 1.75E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04	1.19E-05 1.74E-05 1.07E-04	3.24E-05 4.73E-05 2.92E-04
7	1.19E-05 1.75E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04	1.19E-05 1.75E-05 1.07E-04	3.24E-05 4.75E-05 2.92E-04

8	1.20E-05 1.76E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04	1.19E-05 1.75E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04
9	1.20E-05 1.76E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04	1.20E-05 1.76E-05 1.08E-04	3.24E-05 4.76E-05 2.92E-04
10	1.22E-05 1.80E-05 1.10E-04	3.24E-05 4.76E-05 2.92E-04	1.22E-05 1.80E-05 1.10E-04	1.61E-05 4.76E-05 2.92E-04
11	1.39E-05 2.05E-05 1.26E-04	3.23E-05 4.74E-05 2.90E-04	1.39E-05 2.05E-05 1.26E-04	3.23E-05 4.74E-05 2.90E-04
12	–	3.12E-05 4.58E-05 2.81E-04	–	3.12E-05 4.58E-05 2.81E-04

B6. <u>(Some illustrative effects of suggestion (d) in relation to Table 4.6.1)</u>

Table 4.7

Mean and variance projections under suggestion (d)

t	$\{\theta^*=90\%,\ srt_t=100.5\%\}$		$\{\theta^*=90\%,\ \exp(\eta^*)=1.04\}$		$\{\exp(\eta')=1.04,\ srt_t=100.5\%\}$	
	$E({}^{C}CR^*_t - NR_t)$ ($Var({}^{C}CR^*_t)$)	$E(SR^*_t)$ ($Var(SR^*_t)$)	$E({}^{C}CR^*_t - NR_t)$ ($Var({}^{C}CR^*_t)$)	$E(SR^*_t)$ ($Var(SR^*_t)$)	$E({}^{C}CR^*_t - NR_t)$ ($Var({}^{C}CR^*_t)$)	$E(SR^*_t)$ ($Var(SR^*_t)$)
0	0.907881	0.0	0.889204	0.0	0.598073	0.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
1	0.085761	0.908548	0.068147	0.905730	0.223474	0.611994
	(1.10E-05)	(1.34E-05)	(6.20E-06)	(7.54E-06)	(2.69E-06)	(7.19E-06)
2	0.006855	0.995750	-.009297	0.991161	0.076866	0.851513
	(1.12E-05)	(1.37E-05)	(6.36E-06)	(7.73E-06)	(3.21E-06)	(8.56E-06)
3	-.000719	1.004119	-.016602	0.999219	0.019488	0.945255
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.34E-06)	(8.91E-06)
4	-.001446	1.004922	-.017291	0.999979	-.002966	0.981945
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.38E-06)	(9.02E-06)
5	-.001515	1.004999	-.017356	1.000051	-.011748	0.996308
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.39E-06)	(9.06E-06)
6	-.001522	1.005007	-.017362	1.000058	-.015170	1.001939
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.40E-06)	(9.07E-06)
7	-.001523	1.005008	-.017363	1.000058	-.016467	1.004169
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.40E-06)	(9.08E-06)
8	-.001523	1.005008	-.017362	1.000059	-.016863	1.005110
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.40E-06)	(9.08E-06)

9	-.0015222	1.005008	-.017351	1.000060	-.016721	1.005660
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.41E-06)	(9.08E-06)
10	-.001511	1.005009	-.017234	1.000072	-.015877	1.006358
	(1.12E-05)	(1.37E-05)	(6.37E-06)	(7.75E-06)	(3.48E-06)	(9.07E-06)
11	-.001398	1.005020	-.015937	1.000202	-.013450	1.007914
	(1.12E-05)	(1.37E-05)	(6.38E-06)	(7.75E-06)	(3.96E-06)	(9.03E-06)
12	–	1.005146	–	1.001642	–	1.011931
		(1.37E-05)		(7.75E-06)		(8.75E-06)

; here, the standard calculation basis is {$\theta^* = 50\%$, $srt_t = 100\%$, $\exp(\eta) = 1.06$, $\sigma_a = 10\%$ of $|\mu|$ and $\sqrt{VBR_t} = 10\%$ of $|EBR|$}, which is equivalent to the basis for top values in Table 4.7.1; and $NR_t = 0.028006$ for $\exp(\eta) = 1.06$ and $NR_t = 0.034506$ for $\exp(\eta') = 1.04$, constant for all t.

Chapter 5 Dynamic pension funding plan with a given form of controlling variable

5.1 Introduction

In Chapter 4, we have discussed various aspects of dynamic pension funding plans derived from the LQP control optimisation problems over a finite control horizon, both deterministic and stochastic, where the mathematical form of the controlling variable CR_t is unconstrained. In sections 4.3.3.2 and 4.4.4.2, it was noted that the derived dynamic pension funding plan could be distinguished from the spread funding plan governed by formula (3.22) in section 3.4.4 by the addition of an additive controlling parameter.

Nevertheless, the actuary may have a strong view about controlling the spread parameter [see Remark 2.5 in section 2.2.4.1]. In this respect, different from Chapter 4 where we considered the situation of a solvency valuation (i.e. short-term, winding-up valuation), this chapter considers the situation of a classical actuarial valuation (i.e. long-term, going-concern valuation) in recognition of the fact that the spread funding plan is normally applied to a classical actuarial valuation.

The objective of this chapter is to gain an understanding of how the spread funding plan can be optimally designed in the light of optimal control theory when the controlling variable is constrained by the spread funding formula. It is required to optimise the value of the spread parameter by solving control optimisation problems (formulated later) with respect to the spread parameter, unknown/undetermined at the time of making a decision: in this respect, the unknown spread parameter can be considered to be a controlling parameter in control optimisation problems. This chapter is then devoted to LQP optimisation problems of the type investigated in Chapter 4 but with three different projection assumptions modified for the classical actuarial valuation (first of all, we note that the symbols adopted below, i.e. C_t, NC_t, AL_t, k_t, d_v, i_v, δ_{t+1}, α, β and η all have the same meaning as given throughout Chapter 3 but CR^f_t, NR^f_t and FR_t are newly introduced and defined for the classical actuarial valuation in order to avoid any notational confusion with CR_t, NR_t and SR_t used in Chapter 4 for the solvency valuation):

(a) Control horizon is assumed to be infinite, i.e. $t\in[0, \infty)$;

(b) Controlling variable CR^f_t is assumed to be specified by the following funding-level spread funding formula: that is, for all t,

$$CR^f_t = NR^f_t - k_t \cdot (FR_t-1) \equiv \mu(FR_t-1; k_t) \qquad \text{--- (5.1)}$$

where

$CR^f_t \equiv C_t/AL_t$ (denoting the contribution ratio at time t on the classical actuarial valuation), $NR^f_t = NC_t/AL_t$ (denoting the normal cost ratio at time t on the classical actuarial valuation), $FR_t \equiv F_t/AL_t$ (denoting the funding level at time t on the classical actuarial valuation), and $\mu(FR_t-1; k_t)$ indicates a linear function of state variable FR_t-1 which has the unknown spread parameter k_t having values in the spread parameter space $\{k_t: 0 \le k_t \le 1\}$ (particularly, $\{k: d_v \le k \le 1 \text{ with } i_v > 0\}$ in the case that $k_t = k$ constant for all t, see section 2.2.4.1) and NR^f_t is considered to be the so-called CR^f_t intercept which is pre-computable on the current actuarial assumptions for the classical actuarial valuation. Here, the linear function $\mu(.; k_t)$ shall be called the spread funding plan (characterised by the unknown spread parameter k_t); and

(c) Controlled object is assumed to be deterministic and governed by the following zero-input, 100%-target funding-level growth equation with the state variable FR_t-1 and the unknown spread parameter k_t (which can be thought of as the funding-level version of the zero-input, 100%-target solvency-level growth equation (3.26) with $\tau_t = 0$, i.e. market cost adjustment $m_t =$

0, for all t and using the relationship (3.15), i.e. $SL_t = (1+m_t) \cdot AL_t$ for all t, lead to the resulting equation),

$$[FR_{t+1} - 1] = [\exp(\delta_{t+1} - \alpha - \beta) \cdot (1-k_t)] \cdot [FR_t - 1] + [\exp(\delta_{t+1} - \eta)] - 1 \text{ with initially given } FR_0\text{-}1$$

$$\equiv q_1(t) \cdot (1-k_t) \cdot [FR_t - 1] + q_2(t) \text{ with initially given } FR_0\text{-}1. \qquad \text{- - - (5.2)}$$

Assumption (a) may appear unrealistic but has been made on the grounds that dealing with infinite-horizon control problems could provide an analytically convenient approximation for optimising the value of the spread parameter for control problems over a finite but long time horizon. Further, we have some benefits of comparing, on a long-term, going-concern valuation basis, our dynamic approach over an infinite control horizon with the static approach of Dufresne (1986 & 1988) and Haberman (1992, 1993 & 1994) [for a brief review of their approach, see Remark 2.5 in section 2.2.4.1].

Although dealing with a stochastic approach may be a more proper way of coping with the uncertain real world, we are here concerned with a deterministic approach, which, we believe, is sufficient to illustrate how the spread parameter can be optimally determined. Further, the deterministic approach has some distinctive advantages such that firstly, diagnosing a deterministic controlled object is somewhat easier and facilitates our understanding of the principal results, as we can easily check by comparing the deterministic results of section 4.3 with the stochastic results of section 4.4. Secondly, owing to the complexity of the pension fund system, there are some restrictions in considering a stochastic controlled object (as discussed at the end of section 4.4.1.3). Lastly, testing the sensitivity of funding levels to variations in the principal factors of interest will provide an approximate but clear description of the future behaviour of solvency levels and the future funding policies securing the desired solvency level [see section 4.3.3].

In this chapter, we consider four distinct control optimisation problems - stationary (defined in section 5.3), quasi-stationary (defined in section 5.4), non-stationary and threshold (defined in section 5.5). In section 5.2, we consider the optimal performance criterion appropriate for the infinite-horizon control optimisation problem: in particular, we make some important comments on the supplementary performance criterion, as in section 4.2.3.2. In section 5.3, we consider a stationary controlled object (i.e. its motions are invariant under a translation of time) and then construct our stationary LQP optimisation problem [for the meaning of LQP, see the beginning stage of section 4.2.4]. In section 5.4, we define and consider a quasi-stationary controlled object (as a variation of the stationary controlled object) and then construct our quasi-stationary LQP optimisation problem. Lastly, in section 5.5, we model and consider a non-stationary controlled object and then construct our non-stationary LQP optimisation problem.

In sections 5.3, 5.4 and 5.5 each, we explore their respective optimal values of the spread parameter by solving their respective control problems and further we give some numerical illustrations at the end of each section.

When considering infinite-horizon control problems, the stability properties of the controlled object are important in deciding the value of the spread parameter and analysing the motions of the controlled object. The general concepts of stability mentioned in sections 5.2 ~ 5.5 are discussed briefly in Appendix 5B.1.

5.2 Optimal performance criterion

As in section 4.2.3 for finite-horizon optimal performance criterion, the infinite-horizon optimal performance criterion would be composed of primary and supplementary, which are commonly designed to realise our funding purpose (introduced in section 2.1.3.4).

5.2.1 Primary performance criterion

Considering the characteristics of our controlling variable (specified by formula (5.1)) and controlled object (specified by equation (5.2)) such as 100%-target related state variable FR_t-1 and spreading CR^f_t around NR^f_t, it would be appropriate to set the funding level target at time t at the level of 100% and the contribution ratio target at time t at the level of NR^f_t; hence, the solvency and contribution rate risks at time t for the classical actuarial valuation are defined as $(FR_t - 1)^2$ and $(CR^f_t - NR^f_t)^2$, respectively [see section 2.1.3].

Therefore, the following infinite-horizon performance index (denoted by IPI_θ, which is distinguishable from PI_θ denoting the finite-horizon performance index for the solvency valuation in section 4.2.3.1) would be suitable for a classical actuarial valuation: that is, for an arbitrary value of θ (chosen on the actuary's subjective basis) where $0 < \theta < 1$,

$$IPI_\theta = \sum_{t=0}^{\infty} \{ e^{-\eta t} \cdot [\theta \cdot (FR_t - 1)^2 + (1-\theta) \cdot (CR^f_t - NR^f_t)^2] \} \quad \text{(by formula (5.1))}$$

$$= \sum_{t=0}^{\infty} \{ e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k_t^2] \cdot [FR_t - 1]^2 \}, \qquad \text{--- (5.3)}$$

which illustrates that the contribution rate risk is completely determined by the solvency risk through the spread parameter.

The above IPI_θ can be interpreted as the net present value (NPV) of the project's future costs over an infinite projection period $[0, \infty)$, identified by the convex combination of the solvency and contribution rate risks to be potentially caused by the operation of the spread funding formula (5.1), where $e^{-\eta}$ is the project's discount function. It should be noted that different from the finite-horizon control optimisation problems examined in chapter 4, the infinite-horizon control optimisation problems should be constructed under the guarantee that the value of performance index IPI_θ should be finite from the mathematical point of view (i.e. defined as a real-valued IPI_θ). This requirement can be met provided that the discount factor η in IPI_θ is positive and the performance index per unit control period [t, t+1), i.e. $[\theta + (1-\theta)\cdot k_t^2]\cdot[FR_t - 1]^2$, is bounded for all t.

Thus, the function value IPI_θ is well defined on such a space that $\{k_t$: $\eta>0$ and $0 \leq [\theta+(1-\theta)\cdot k_t^2]\cdot[FR_t-1]^2 \leq u$ (i.e. some positive real number)$\}$ for all t. Further, taking into account the parameter space of k_t in formula (5.1), the feasible region of k_t should be the intersection of $\{k_t: 0 \leq k_t \leq 1\}$ and $\{k_t: \eta>0 \text{ and } 0 \leq [\theta+(1-\theta)\cdot k_t^2]\cdot[FR_t-1]^2 \leq u\}$: that is, for all t,

$$\{k_t: 0 \leq k_t \leq 1, \eta > 0 \text{ and } 0 \leq [\theta+(1-\theta)\cdot k_t^2] \cdot [FR_t-1]^2 \leq u\} \quad \text{- - - (5.4)}$$

(in the special case of $k_t = k$ constant for all t, then

$\{k: 1-\exp(-\eta) \leq k \leq 1 \text{ with } \eta > 0 \text{ and } 0 \leq [\theta+(1-\theta)\cdot k^2] \cdot [FR_t-1]^2 \leq u\})$

; this will be regarded as the general controlling parameter space of k_t in our infinite-horizon control problems (formulated later).

Our purpose of optimising the performance index (5.3) with respect to the spread parameter, our primary performance criterion can be defined as follows: for a given θ,

$$\underset{\{k_t;\, t=0,\,1,\,\ldots\}}{\text{Min}} \quad IPI_\theta \quad \text{- - - (5.5)}$$

: that is, when the controlling variable CR_t^f is constrained by the given form of the spread funding formula (5.1), controlling both the solvency and contribution rate risks with respect to CR_t^f is equivalent to controlling only the solvency risk with respect to the spread parameter k_t (and hence, controlling CR_t^f is to be completed by controlling k_t) - this is quite different from our previous analysis in Chapter 4. In this respect, k_t is considered to be the controlling parameter at time t (specifying the controlling variable CR_t^f at time t).

5.2.2 Supplementary performance criterion

Following the arguments to (i)~(v) made in section 4.2.3.2, we may consider the infinite-horizon version of the finite-horizon supplementary performance criteria, Suggestion I, II and III, proposed in section 4.2.3.2. As a simple example, if we adopt the infinite-horizon version of Suggestion I as our supplementary performance criterion to the primary performance criterion (5.5), it will be given by

$$\underset{\theta}{\mathrm{Min}} \ \{ \sum_{t=0}^{\infty} \ \lambda\cdot(FR_t'-1)^2 + (1-\lambda)\cdot(CR^f{}_t'-NR^f_t)^2 \} \Leftrightarrow \underset{\theta}{\mathrm{Min}} \ \{ \sum_{t=0}^{\infty} \ (\lambda+(1-\lambda)\cdot k_t')\cdot(FR_t'-1)^2 \}$$

where

$CR^f_t{}'$, FR_t' and λ each have the same meaning as given in section 4.2.3.2; and

k_t' = pre-optimal controlling parameter at time t with the relationship $CR^f_t{}' = NR^f_t - k_t'\cdot(FR_t'-1)$, $0 \le k_t' \le 1$ (particularly, $d_v \le k' \le 1$ for $k'_t=k'$ constant for all t).

As mentioned earlier in section 4.2.3.2, constructing a supplementary performance criterion suitable for improving the pace of funding and the behaviour of solvency levels simultaneously, would involve severe computational problems, even more so than in the finite-horizon case.

Throughout this chapter, we prefer alternatively to illustrate how the pre-optimal value of k_t' responds to changes in the values of θ from a specified finite admissible set of θ, instead of considering formally a supplementary performance criterion.

Remark 5.1: At this point, it would be worth illustrating how the methodology of primary and supplementary performance criterion can be applied to the single-point-time static approach. Here, we give an illustration with respect to the work of Dufresne (1988).

(a) Dufresne's primary performance criterion: using a discrete-time and stochastic model under some simplification assumptions, he illustrated numerically how a best value of the constant spread parameter k at some point of time t (denoted by k^*) can be determined, since generally there is no closed form expression for k^*. His decision criterion at time t is $\underset{k \in [0, 1]}{\text{Min}} \text{Var}(C_t)$ and then the interval $[k^*, 1]$ is defined as his admissible region for k in view of the fact that the tradeoff between $\text{Var}(C_t)$ and $\text{Var}(F_t)$ is maintained over $[k^*, 1]$. In this respect, $\underset{k \in [0, 1]}{\text{Min}} \text{Var}(C_t)$ can be called Dufresne's primary performance criterion (for the admissible region for k) at time t. We note that, more recently, these results have been formalised and extended by Owadally & Haberman (1995). In addition, taking into account the scale of the tradeoff between $\text{Var}(C_t)$ and $\text{Var}(F_t)$, we could make a more condensed admissible region for k than that provided by Dufresne's primary performance criterion. To begin with, assuming that $\text{Var}(C_t)$ and $\text{Var}(F_t)$ are both differentiable with respect to k, then we have the instantaneous rate of growth of $\text{Var}(C_t)$ and of $\text{Var}(F_t)$ with respect to k, defined as $[d\text{Var}(C_t)/dk] \ / \ \text{Var}(C_t) \equiv \text{RVC}_t$ and $[d\text{Var}(F_t)/dk] \ / \ \text{Var}(F_t) \equiv \text{RVF}_t$, respectively. Consequently, the following criterion (b) is intended to focus on the movement of $\text{Var}(C_t)$ with respect to k as well as that of $\text{Var}(F_t)$ with respect to k, at a fixed time t; and

(b) Supplementary performance criterion to Dufresne's primary performance criterion: for any $k \in [k^*, 1]$, $\text{RVC}_t > 0$ and $\text{RVF}_t < 0$ (since $\text{Var}(C_t)$ increases and $\text{Var}(F_t)$ decreases with increasing k), so it is appropriate to define a measure of the tradeoff between RVC_t and RVF_t in the form: $\text{MS}_t \equiv \text{RVC}_t + \text{RVF}_t$. We believe that the new measure MS_t is suited to the balancing the movement of $\text{Var}(C_t)$ against that of $\text{Var}(F_t)$. There will be a more condensed admissible region for k such that $[k_L , k_u] \subset [k^*, 1]$, in which '$k_L$' is defined as the value of k satisfying $\underset{k \in [k^*, 1]}{\text{Min}} \{\text{MS}_t(k) < 0\}$ and also 'k_U' is defined as the value of k satisfying $\underset{k \in [k^*, 1]}{\text{Max}} \{\text{MS}_t(k) < 0\}$. Therefore, $\underset{k \in [k^*, 1]}{\text{Min}} \{\text{MS}_t(k) < 0\}$ and $\underset{k \in [k^*, 1]}{\text{Max}} \{\text{MS}_t(k) < 0\}$ could be considered the supplementary performance criterion (to Dufresne's primary performance criterion). For example, adopting the same numerical bases as given in his Fig. 2 (i.e. the investment rate of return i_t follows an IID with $E(i_t) = i_v = 0.03$ and $Var(i_t) = 0.01$), we can give an illustrative numerical comparison as follows:

$$[k_L, k_u] \text{ vs. } [k^*, 1] = \begin{cases} [0.12,\ 0.38] \text{ vs. } [0.00,\ 1.00] & \text{for } t = 10 \\ [0.15,\ 0.38] \text{ vs. } [0.15,\ 1.00] & \text{for } t = 30 \\ [0.16,\ 0.38] \text{ vs. } [0.16,\ 1.00] & \text{for } t = \infty \end{cases}$$

This indicates that we are able to redefine the admissible region for k in a more restrictive way by employing a supplementary performance criterion appropriately designed for our additional needs and interest.

5.3 Dynamic pension funding plan for stationary LQP optimisation problem

In order to avoid any notational confusion with the other two control problems to be considered later - quasi-stationary and non-stationary, we shall put the superscript 'S' on the left side of each main symbol (introduced in this section), which indicates that it concerns the stationary LQP optimisation problem formulated under the stationary assumptions (given in the following section 5.3.1).

5.3.1 Preliminary

We make first the stationary assumptions such that all system parameters (involved in the system equation (5.2)) are constants for all t based on best estimates, that is, $\delta_{t+1} = \delta$ and $k_t = k$ are all time-invariant (i.e. stationary), as in the classical approach to actuarial valuation.

Applying these parametric assumptions to the spread funding formula (5.1) and the system equation (5.2) and using the fact that $1\text{-}\exp(-\eta) = d_v$, then for all $t \in [0, \infty)$, we have

$$^{S}CR^{f}_{t} = {}^{S}NR^{f}_{t} - k\cdot({}^{S}FR_{t}-1) \equiv {}^{S}\mu({}^{S}FR_{t}-1;\, k),\ \ k\in\{k\colon 1-\exp(-\eta) \leq k \leq 1 \text{ with } \eta > 0\} \quad \text{---}(5.6)$$

; k shall be called the stationary spread parameter, formula (5.6) the stationary spread funding formula, $^{S}\mu(.; k)$ the stationary spread funding plan; and,

$$[{}^{S}FR_{t+1} - 1] = [q_1\cdot (1-k)] \cdot [{}^{S}FR_{t} - 1] + q_2, \quad \text{with given } {}^{S}FR_{0} - 1 \quad \text{---}(5.7)$$

where $q_1 = \exp(\delta\text{-}\alpha\text{-}\beta)$; and $q_2 = \exp(\delta\text{-}\eta) - 1$

; hence, this system equation is autonomous (i.e. zero-input and stationary), and accordingly the controlled object is called the autonomous controlled object.

Further, considering the stationary system equation (5.7), it would be reasonable to impose some stabilising condition on the control error sequence $\{{}^{S}FR_{t}\text{-}1;\, t\in[0, \infty)\}$ because it would be quite unacceptable to the trustees and employer if $\{{}^{S}FR_{t}\text{-}1;\, t\in[0, \infty)\}$ were any type of divergent sequence. In the next section 5.3.2, we give a brief discussion of stability problems of stationary controlled objects [for a further discussion, see Appendix 5A].

In section 5.3.3, we construct a stationary LQP optimisation problem by reference to the stability problem and then find the optimal value of k by solving the formulated control problem in section 5.3.4. Numerical illustrations are given in section 5.3.5.

5.3.2 Stability problems of autonomous controlled object

The purpose of this section is to derive a convergence condition for k from the limiting behaviour of the control error sequence $\{{}^{S}SR_{t}\text{-}1;\, t\in[0, \infty)\}$ and then establish an admissible condition for k.

(i) Definite solution to system equation (5.7):

The general solution to equation (5.7) consists of the sum of two solutions: a particular solution (denote, PS_t) which represents the inter-temporal equilibrium level of ${}^{S}FR_t$-1 (i.e. equilibrium state or moving equilibrium state, see Comments (1.1)~(1.4) in A5.5.2) and is any solution of the complete equation (5.7) with $q_2 \neq 0$, and a complementary function solution (denote, CF_t) which represents the deviation of the inter-temporal equilibrium and is the general solution of the reduced equation (5.7) with $q_2 = 0$.

The initial condition ${}^{S}FR_0$-1 enables us to determine completely the general solution to (5.7) (i.e. leads to the definite solution to (5.7)):

$$ {}^{S}FR_t - 1 = CF_t + PS_t \qquad \text{--- (5.8)} $$

where

- [in the case of $q_1 \cdot (1-k) \neq 1$]:

$CF_t = \{({}^{S}FR_0-1) - q_2 / [1 - q_1 \cdot (1-k)]\} \cdot \{q_1 \cdot (1-k)\}^t$, and

$PS_t = q_2 / [1 - q_1 \cdot (1-k)]$

; hence, we can easily check from Definition 1 in Appendix 5A that PS_t reaches an equilibrium state (here, PS_t is a constant function value satisfying the system equation (5.7), PS); and

- [in the case of $q_1 \cdot (1-k) = 1$]: $CF_t = {}^{S}FR_0 - 1$ and $PS_t = q_2 \cdot t$

; hence, PS_t becomes a moving equilibrium state, since PS_t is a time-varying function value satisfying equation (5.7).

(ii) Stability:

If $|q_1 \cdot (1-k)| < 1$ (for convergence), then the equilibrium state PS is asymptotically stable [see Definition 3 in Appendix 5A], since CF_t decays to zero as t tends to infinity, and then the control error sequence $\{{}^{S}FR_t\text{-}1;\ t \in [0, \infty)\}$ is convergent with the limiting value PS. Thus,

$|q_1 \cdot (1\text{-}k)| < 1$ is the asymptotic stability condition: in other words, as $t \rightarrow \infty$, [control error FR_t - 1 (with given initial error FR_0-1)] $\rightarrow$ (after transient error CF_t has decayed to zero) $\rightarrow$ [steady-state error PS] [see Remark 2.5 in section 2.3.2.2].

Moreover, we have the following inequality from the properties of the absolute values

$|^{S}FR_t\text{-}1\text{-}PS| = |[^{S}FR_0 - 1 - PS] \cdot [q_1 \cdot (1\text{-}k)]^t| \le \{|^{S}FR_0\text{-}1| + |PS|\} \cdot |[q_1 \cdot (1\text{-}k)]|^t$.

From Definition 4 in Appendix 5A, we know that $|q_1 \cdot (1\text{-}k)|$ corresponds to $(1 - {}^{S}\zeta) \in [0, 1)$ (here, the constant convergence rate ${}^{S}\zeta$ is sometimes called the geometric damping rate as the discontinuous analogue of the exponential damping rate defined normally on a continuous-time domain, as mentioned in Comment 4 in Appendix 5A) and that the equilibrium state PS is geometrically stable. As a result, the asymptotic stability condition, $|q_1 \cdot (1\text{-}k)| < 1$, is also a geometric stability condition. Thus, if we can continuously manage our pension scheme subject to $|q_1 \cdot (1\text{-}k)| < 1$, then the scheme becomes stable and the solution sequence $\{^{S}FR_t\text{-}1;\ t\in[0, \infty)\}$ is uniformly convergent to $\lim_{t\rightarrow\infty}(^{S}FR_t - 1) \equiv {}^{S}FR_\infty\text{-}1 = PS$. In other words, if $0 < q_1 \cdot (1\text{-}k) < 1$, the solvency level decreases steadily with limiting value ${}^{S}FR_\infty$, but on the other hand if $-1 < q_1 \cdot (1\text{-}k) < 0$, the solvency level experiences damped oscillations and tends to the limiting value ${}^{S}FR_\infty$, with each successive cycle of smaller amplitude than the preceding one.

As a result, the admissible space for k that guarantees the convergence of the control error sequence $\{^{S}FR_t\text{-}1;\ t\in[0, \infty)\}$ is $\{k: |q_1 \cdot (1\text{-}k)| < 1\}$ and the convergence speed of $\{^{S}FR_t\text{-}1;\ t\in[0, \infty)\}$ to PS is increasing with k: in other words, the convergence rate ${}^{S}\zeta$ is an increasing function of k with the maximum damping rate 1 when k = 1.

(iii) <u>Stationary parameter space of k</u>:

Considering the convergence of the control error sequence together with the finiteness of ${}^{S}IPI_\theta$, the feasible region of k should be the intersection of $\{k: 1\text{-}\exp(-\eta) \le k \le 1 \text{ with } \eta > 0\}$ (i.e. general parameter space of k given in (5.1)) and $\{k: |q_1 \cdot (1\text{-}k)| < 1\}$. Thus, we have

$\{k: 1-\exp(-\eta) \leq k \leq 1$ with $\eta > 0$ and $|q_1 \cdot (1-k)| < 1\}$ --- (5.9)

; this shall be called the stationary controlling parameter space of k (as a specific and restrictive form of the general controlling parameter space (5.4)), which will be used in the formulation of the stationary control problem in the next section 5.3.3.

5.3.3 Stationary LQP optimisation problem

From the discussions made in section 5.3.2, we have found an admissible controlling parameter space (5.9) that guarantees the finiteness of ${}^{S}IPI_{\theta}$. Then, the optimal value of k would be defined as the value of k satisfying the following stationary LQP optimisation problem:

$$\underset{k}{\text{Min}} \left\{ \sum_{t=0}^{\infty} \left[e^{-\eta t} \cdot (\theta + (1-\theta) \cdot k^2) \cdot ({}^{S}FR_t - 1)^2 \right] \right\}$$

subject to given $\theta \in (0, 1)$; stationary system equation (5.7); and $k \in$ {stationary controlling parameter space (5.9)}.

--- (5.10)

The optimisation procedure for solving the above control problem (5.10) will be considered in the next section 5.3.4.

5.3.4 Optimisation procedure with respect to the stationary spread parameter

The objective of this section is to find the optimal value of k (denoted by k^*) by solving the stationary LQP optimisation problem (5.10) and then we define the optimal stationary spread funding plan (specified by k^*) as our dynamic pension funding plan [see, section 5.3.4.1]. We will discuss some advantages and disadvantages of the algebraic approach adopted for the mathematical solution [see, section 5.3.4.2]. Finally, we give a block diagram illustrating our control mechanism as a summary of this section [see, section 5.3.4.3].

5.3.4.1 Optimal stationary spread funding plan

The optimisation procedure with respect to k is carried out as follows using an algebraic approach:

For convenience, letting $q_1 \cdot (1-k) = f(k)$ (as a linear function of k; hence, $k = 1 - f(k)/q_1$) and ${}^{S}FR_0 - 1 = s$. From the relation $k = 1 - f(k)/q_1$, ${}^{S}IPI_\theta$ can be expressed as a function of f(k), so we deal with ${}^{S}IPI_\theta$ as a function of f(k) (denoted by ${}^{S}IPI_\theta(f(k))$).

Then, the stationary controlling parameter space (5.9) yields a more condensed feasible region of values of f(k), that is,

$$\{f(k): 0 \le f(k) < 1 \text{ if } \delta-\alpha-\beta-\eta \ge 0, \text{ otherwise } 0 \le f(k) \le \exp(\delta-\alpha-\beta-\eta)\} \quad \text{--- (5.11)}$$

Then, the infinite series ${}^{S}IPI_\theta$ is a convergent geometric series and reduces to

$${}^{S}IPI_\theta(f(k)) = [\theta + (1-\theta) \cdot (1 - f(k)/q_1)^2] \cdot [(s - s \cdot f(k) - q_2)^2 \cdot (1-e^{-\eta} \cdot f(k)^2)^{-1} + 2(s \cdot q_2 - s \cdot q_2 \cdot f(k) - q_2^2) \cdot (1 - e^{-\eta} \cdot f(k))^{-1} + q_2^2 \cdot (1 - e^{-\eta})^{-1}] \; / \; [1 - f(k)]^2.$$

For convenience, we rewrite the above equation as

$${}^{S}IPI_\theta(f(k)) = [\theta + (1-\theta) \cdot (1 - f(k)/q_1)^2] \cdot G(f(k)) \; / \; [1 - f(k)]^2$$

where

$$G(f(k)) = \{[(e^{-2\eta} \cdot (s - q_2)^2 - e^{-\eta} \cdot s \cdot (s - 2q_2)] \cdot f(k)^3 + [(s^2 + e^{-\eta} \cdot (s^2 - 4s \cdot q_2 + q_2^2) - 2e^{-2\eta} \cdot (s - q_2)^2] \cdot f(k)^2 + [-2s^2 + e^{-\eta} \cdot (s^2 + 2s \cdot q_2 - 2q_2^2) + e^{-2\eta} \cdot (s - q_2)^2] \cdot f(k) + [s^2 + e^{-\eta} \cdot (s^2 - q_2^2)]\} \; / \; \{(1 - e^{-\eta} \cdot f(k)^2) \cdot (1 - e^{-\eta} \cdot f(k)) \cdot (1 - e^{-\eta})\}$$

$$= \{[1 - f(k)]^2 \cdot [(e^{-2\eta} \cdot (s - q_2)^2 - e^{-\eta} \cdot s \cdot (s - 2 \cdot q_2)) \cdot f(k) + e^{-\eta} \cdot (q_2^2 - s^2) + s^2]\} \; / \; \{(1 - e^{-\eta} \cdot f(k)^2) \cdot (1 - e^{-\eta} \cdot f(k)) \cdot (1 - e^{-\eta})\}.$$

Finally, we have a more simplified form with

$$^{S}IPI_{\theta}(f(k)) = [\theta + (1-\theta)\cdot(1 - f(k)/q_1)^2] \cdot [(e^{-2\eta}\cdot(s - q_2)^2 - e^{-\eta}\cdot s\cdot(s - 2\cdot q_2)) \cdot f(k) + e^{-\eta}\cdot(q_2^2 - s^2) + s^2]$$

$$/ [(1 - e^{-\eta}\cdot f(k)^2)\cdot (1 - e^{-\eta}\cdot f(k)) \cdot (1-e^{-\eta})]. \qquad \text{--- (5.12)}$$

As a first step to searching for the value of k to minimise $^{S}IPI_{\theta}(f(k))$ in (5.12), we need to solve the polynomial equation, $d^{S}IPI_{\theta}(f(k)) / dk = [d^{S}IPI_{\theta}(f(k) / df(k)] \cdot [df(k) / dk] = 0$, which leads to a biquadratic equation in f(k) (note that the solutions to the following equation (5.13) may lead to a maximum rather than minimum of $^{S}IPI_{\theta}(f(k))$. After some simplification, we obtain

$$d^{S}IPI_{\theta}(f(k)) / dk = [a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6] \cdot f(k)^4 + [2a_1a_4a_6 - 2a_1a_3 - 2a_2a_3a_7] \cdot f(k)^3 +$$

$$[3a_1a_5a_6 - 3a_2a_3 + a_1a_4a_7 + a_2a_4a_6 - a_1a_4 - a_2a_4a_7] \cdot f(k)^2 +$$

$$[2a_1a_5a_7 + 2a_2a_5a_6 - 2a_2a_4] \cdot f(k) + [a_1a_5 + a_2a_5a_7 - a_2a_4] = 0 \qquad \text{--- (5.13)}$$

where

$a_1 = e^{-2\eta}\cdot(s - q_2)^2 - e^{-\eta}\cdot s\cdot(s - 2\cdot q_2)$; $a_2 = e^{-\eta}\cdot(q_2^2 - s^2) + s^2$; $a_3 = e^{-2\eta} - e^{-3\eta}$; $a_4 = e^{-2\eta} - e^{-\eta}$;

$a_5 = 1 - e^{-\eta}$; $a_6 = (1-\theta)/q_1^2$; and $a_7 = -2(1-\theta)/q_1$.

Moreover, if $[a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6] \neq 0$, it is convenient to reparameterise equation (5.13) by dividing each coefficient by $[a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6]$ and then we have the form

$$d^{S}IPI_{\theta}(f(k)) / dk = f(k)^4 + c_1 \cdot f(k)^3 + c_2 \cdot f(k)^2 + c_3 \cdot f(k) + c_4 = 0 \qquad \text{--- (5.14)}$$

where

$c_1 = [2a_1a_4a_6 - 2a_1a_3 - 2a_2a_3a_7] / [a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6]$;

$c_2 = [3a_1a_5a_6 - 3a_2a_3 + a_1a_4a_7 + a_2a_4a_6 - a_1a_4 - a_2a_4a_7] / [a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6]$;

$c_3 = [2a_1a_5a_7 + 2a_2a_5a_6 - 2a_2a_4] / [a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6]$; and

$c_4 = [a_1a_5 + a_2a_5a_7 - a_2a_4] / [a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6]$.

The algebraic solution of biquadratic equations is generally given by Ferrari's method [see, Upensky (1948, Ch. 5)]. Using this method, the procedure for obtaining the solutions is given in Appendix 5B.1.

On the other hand, in the case of $[a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6] = 0$, equation (5.13) is reduced to a cubic equation. In general, the algebraic solutions of cubic equations are given by Cardan's method [see, Upensky (1948, Ch. 5)]. As seen in Appendix 5B.1, Ferrari's method is completed by using Cardan's method, so we can also refer to Appendix 5B.1 to find the solutions for the case of $[a_1a_4a_6 - a_1a_3a_7 - a_2a_3a_6] = 0$.

Now, we shall build up the following optimisation procedures for determining the optimal value of k (denoted by k^*) because it is too complicated to derive a closed mathematical expression for k^*, on the evidence of Appendix 5B.1.

To begin with, for notational convenience we define x^- for a real number x such that $\lim_{h \to 0^+}(x - h) = x^-$, in which $h \to 0^+$ indicates that the limit is considered as h tends to zero from above: that is, x^- is approximately equal to x but is smaller than x.

Procedure 1: If $f(k) = - [e^{-\eta} \cdot (q_2^2 - s^2) + s^2] / [e^{-2\eta} \cdot (s - q_2)^2 - e^{-\eta} \cdot s \cdot (s-2 \cdot q_2)] \in$ {feasible region of f(k) derived in (5.11)} from formula (5.12), then k^* is given by $k^* = 1 - f(k)/q_1$, which leads to the value of ${}^S IPI_\theta(f(k))$ in (5.12) being zero, so k^* provides the most ideal result in control optimisation because ${}^S IPI_\theta(f(k)) \geq 0$ for any k, otherwise

Procedure 2: If $\delta-\alpha-\beta-\eta \geq 0$, then k^* is the value corresponding to Min { ${}^S IPI_\theta(0)$, ${}^S IPI_\theta(f_j(k))$, ${}^S IPI_\theta(1^-)$ }, where $f_j(k)$ are the real-valued solutions of the polynomial equation (5.13) satisfying $f_j(k) \in [0, 1)$; but on the other hand, Procedure 3: If δ-α-β-η < 0, then k^* is the value corresponding to Min $\{{}^S IPI_\theta(0), {}^S IPI_\theta(f_j(k)), {}^S IPI_\theta(\exp(\delta-\alpha-\beta-\eta))\}$, where $f_j(k)$ are the real-valued solutions of the polynomial equation (5.13) satisfying $f_j(k) \in [0, \exp(\delta-\alpha-\beta-\eta)]$.

In conclusion, we can identify the optimal stationary control action at time t, ${}^S CR^{f*}_t$, after k^* has been determined through Procedures 1, 2 or 3 as follows: that is, for every $t \in [0, \infty)$,

$${}^S CR^{f*}_t = {}^S NR^f_t - k^* \cdot ({}^S FR_t - 1) = {}^S\mu^*({}^S FR_t - 1; k^*) \qquad \text{--- (5.15)}$$

where ${}^S\mu^*(.; k^*)$ denotes the optimal stationary spread funding plan, which is our dynamic pension funding plan (i.e. optimal linear stationary feedback control law for the autonomous controlled object governed by equation (5.7)).

Further, the optimal stationary control response ${}^S FR^*_{t+1}$, corresponding to ${}^S CR^{f*}_t$, is generated sequentially with time t by the following optimal autonomous system equation (5.16):

$$({}^S FR^*_{t+1} - 1) = \exp(\delta-\alpha-\beta) \cdot (1-k^*) \cdot ({}^S FR^*_t - 1) + \exp(\delta-\eta) - 1 \text{ with given } {}^S FR_0 - 1 = {}^S FR^*_0 - 1. \qquad \text{-- (5.16)}$$

We note finally that the above formulae (5.15) and (5.16) are completely characterised by k^* obtained through Procedures 1, 2 or 3.

5.3.4.2 Essential requirements for the algebraic approach

The essential requirements for applying the algebraic approach to the stationary LQP optimisation problem (5.8) are summarised as follows, in their order of priority:

Requirement A1: The unknown spread parameter k_t is fixed over an infinite control horizon, i.e. $k_t = k$, constant for all t;

Requirement A2: The geometric stability condition (i.e. $|\exp(\delta-\alpha-\beta)\cdot(1-k)| < 1$) and the positive discount factor (i.e. $\eta>0$) (for convergent geometric series); and

Requirement A3: The polynomial equation (5.13) obtained has to be an equation of degree less than fifth (for mathematical solutions to (5.13), we note that in general, polynomial equations of degree higher than the fourth are not soluble, see Tignol (1988; Ch. 13) for the proof).

For a practical point of view, these requirements would be a great limitation for using the algebraic approach as a mathematical solution tool of the stationary LQP optimisation problem (5.10).

It is worth noting that the rational motivation of using the algebraic approach is to expect that repeated use of the optimal value of k (denoted by k^*) decided at the initial time, no matter what further information is encountered, will reduce the solvency risk as well as the contribution rate risk over an infinite control horizon. Consequently, the decision of k^* occurs at the initial time (i.e. t=0) and is to be permanently maintained thereafter. Thus, the spread parameter k_t is viewed as an unknown but fixed parameter (to be estimated at the initial time). In the light of statistical decision theory, the viewpoint of algebraic approach on the spread parameter k_t is similar to the Frequentist Perspective in the light of repeated use of decision k^* at the starting time, irrespective of any information updated with time t [see, Berger (1985; section 1.6)].

5.3.4.3 Summary

Our control mechanism for autonomous (i.e. zero-input, stationary) controlled object can be summarised, as illustrated below in Figure 5.1 (which will also provide a useful comparison with Figures 4.1 and 4.2 given in section 4.2.1).

[Autonomous controlled object] $\longrightarrow$ $\{{}^{S}FR_t\text{-}1\}$ $\longrightarrow$

$\{k=k^*\}\uparrow$ $\downarrow$

$\{{}^{S}NR_t\}$ $\longrightarrow$ [Optimal control law ${}^{S}\mu^*(.; k=k^*)$] $\longleftarrow$ {Feedback: ${}^{S}FR_t\text{-}1$}

Figure 5.1 Optimal stationary spread funding control system.

The above Figure 5.1 shows that both the autonomous controlled object and the stationary spread funding plan are optimally designed by way of optimising the value of the unknown but fixed spread parameter k at the level of k^*.

5.3.4.4 Numerical illustrations

In this section, we will simply illustrate numerically the relationships of the optimal value of k (i.e. k^*) with respect to the force of investment interest (i.e. δ), the relative weighting parameter (i.e. θ) and/or the given initial solvency level (i.e. ${}^{S}FR_0$). Owing to the inaccessibility of the closed mathematical form of k^*, it is not possible to derive these relationships analytically. All the numerical illustrations are given in Appendix 5B.

(i) Assumptions:

(A1) All actuarial assumptions are the same as given in section 4.3.3.1, except for $\exp(\tau)=1$ (in short, $\exp(\alpha)=1.03$, $\exp(\beta)=1.02$ and $\exp(\eta)=1.06$); and

(A2) Projection assumptions:

- infinite control horizon: $t \in [0, \infty)$;
- admissible values of θ: {90%, 80%, ..., 10%};
- admissible values of ${}^{S}FR_0$: {100%, 50%, 0%}; and
- force of investment interest (δ): $\exp(\delta) = 1.06$ or 1.08.

Hence, the formulae (a)~(d) and Remark 4.2 introduced in section 4.3.3.1 hold simply by replacing NR_t (=NC_t/SL_t) and EBR_t (=EB_t/SL_t) with their respective ${}^{S}NR^{f}_{t}$ (=NC_t/AL_t) and ${}^{S}EBR^{f}_{t}$ (=EB_t/AL_t).

(ii) Illustrative numerical results:

From the above assumption (i), we have the relationship $\delta < \alpha+\beta+\eta$. So, the optimal value of k (denoted by k^*) will be provided by Procedure 3 (established in section 5.3.4.1) in relation to the admissible combinations of δ, θ and ${}^{S}FR_0$. The illustrative numerical results are given in Table 5.1 in Appendix 5B.

This table illustrates that

(a) k^* increases monotonically with ${}^{S}FR_0$. This implies that under a stationary economic and demographic circumstances, maintaining the solvency level at each valuation date at the level of around 100% reduces both the solvency and contribution rate risks. This follows because as the value of k becomes closer to one, so the control error sequence $\{{}^{S}FR_t\text{-}1;\ t \in [0, \infty)\}$ converges more speedily to the steady-state error $\lim_{t\to\infty}({}^{S}FR_t - 1) = PS$ and the transient error CF_t decays more quickly to zero [see subsection (ii) of section 5.2.2.1];

(b) Weighting the solvency risk more than the contribution rate risk (i.e. $\theta \to 1$) leads to a value of k^* closer to one, irrespective of the initial solvency level, and hence $\{{}^{S}FR_t\text{-}1;\ t\in[0, \infty)\}$ will be stabilised more quickly (and the control action error sequence $\{{}^{S}CR^{f}_{t} - {}^{S}NR^{f}_{t};\ t\in[0, \infty)\}$ as well). This result is as expected, since with θ=1 the immediate and complete spread method is recommended: k^* = 1 means paying-off the control error informed at each valuation date without any spreading into the future;

(c) On the other hand, concentrating on controlling the contribution rate risk provides k^* closer to 1-exp(-η), irrespective of the initial solvency level, (i.e. the lowest available value of general parameter space of k given in (5.1)). This result is also as expected, since spreading the control error informed at each valuation date into an infinite period is recommended (i.e. $1\text{-}\exp(-\eta) = 1/\ddot{a}_{\infty}(i_v)$), which is consistent with the conclusion of Haberman (1997); and

(d) The better investment performance (i.e. $\delta > \eta$) than the most likely expected performance (i.e. $\delta = \eta$) provides a larger k^*, which implies that both $\{{}^{S}FR_t\text{-}1;\ t\in[0, \infty)\}$ and $\{{}^{S}CR^f_t - {}^{S}NR^f_t;\ t\in[0, \infty)\}$ are stabilised more quickly, as discussed in (b) above.

5.4 Dynamic pension funding plan for quasi-stationary LQP optimisation problem

Throughout this section, we put the superscript 'Q' on the left side of each main symbol (to be introduced in this section), which is used to indicates that it concerns the quasi-stationary LQP optimisation problem (formulated later in section 5.4.3).

5.4.1 Quasi-stationary assumptions and Preliminaries

As for the stationary LQP optimisation problem (5.10) in section 5.3.3, the stationary spread parameter k is considered to be unknown but fixed. In practice, the value of the spread parameter would generally be decided and adjusted through the regular valuation process, it would be more realistic to treat k as an unknown quantity depending on the currently available information: in other words, denoting the available information vector at time t by ${}^{Q}H_t = ({}^{Q}FR_0, {}^{Q}FR_1, \ldots, {}^{Q}FR_t, {}^{Q}CR^f_0, {}^{Q}CR^f_1, \ldots, {}^{Q}CR^f_{t-1})$ with the given initial information ${}^{Q}H_0 = {}^{Q}FR_0$, then it is proposed that k be of the form of $k({}^{Q}H_t)$, a time-invariant function of ${}^{Q}H_t$ (i.e. independent explicitly of time t but not constant for all t). In this respect, interpreting the stationary spread parameter k (involved in the stationary control problem in the previous section 5.3), k itself could be thought of as a function of all possible and exhaustive prior information ${}^{Q}H_{\infty} = ({}^{Q}FR_0, {}^{Q}FR_1, {}^{Q}FR_2, \ldots, {}^{Q}CR^f_0, {}^{Q}CR^f_1, {}^{Q}CR^f_2, \ldots)$, say $k = k({}^{Q}H_{\infty})$ constant for all t (i.e. initially dependent on the full history of the controlled object, ${}^{Q}H_{\infty}$), which shows that assuming a stationary spread parameter k is not reasonable as well as not practical.

We make here the same assumptions as the stationary assumptions in section 5.3.1, except for replacing k with $k({}^{Q}H_t)$: that is, $\delta_{t+1} = \delta$, $\tau_t = \tau$ and $k_t = k({}^{Q}H_t)$ for every $t \in [0, \infty)$: in this respect, these assumptions shall be called the quasi-stationary assumptions as a variation of the stationary assumptions (here, the term 'quasi-stationary' is added to reflect the fact that $k({}^{Q}H_t)$ does not depend explicitly on time t but is not constant for all t).

Applying the quasi-stationary assumptions to the spread funding formula (5.1) and the system equation (5.2), then we obtain the following quasi-stationary versions (5.17) and (5.18) of the stationary formula (5.6) and autonomous equation (5.7) derived in section 5.3.1: that is, for all $t \in [0, \infty)$,

$${}^{Q}CR^{f}_{t} = {}^{Q}NR^{f}_{t} - k({}^{Q}H_t) \cdot ({}^{Q}FR_t - 1) \equiv {}^{Q}\mu({}^{Q}FR_t - 1;\, k({}^{Q}H_t)), \qquad \text{--- (5.17)}$$

in which $k({}^{Q}H_t) \in \{k({}^{Q}H_t): 0 \leq k({}^{Q}H_t) \leq 1\}$

; in a similar manner to the stationary formula (5.6), $k({}^{Q}H_t)$ shall be called the quasi-stationary spread parameter, formula (5.17) the quasi-stationary spread funding formula and ${}^{Q}\mu(.;\, k({}^{Q}H_t))$ the quasi-stationary spread funding plan, and further, we note that ${}^{Q}NR^{f}_{t} = {}^{S}NR^{f}_{t}$; and hence,

$$[{}^{Q}FR_{t+1} - 1] = q_1 \cdot (1 - k({}^{Q}H_t)) \cdot ({}^{Q}FR_t - 1) + q_2 \quad \text{with given } {}^{Q}FR_0 - 1 \qquad \text{--- (5.18)}$$

where q_1 and q_2 are the same as earlier specified in the autonomous system equation (5.7).

In particular, this equation (5.18) shall be called the quasi-autonomous (i.e. quasi-stationary and zero-input) system equation, and accordingly the controlled object governed by this equation shall be called the quasi-autonomous controlled object. Here, the term 'quasi-autonomous' is added on the grounds that this equation seems not to be autonomous in view of $k(^QH_t)$ not being constant for all t, but provided that the unknown $k(^QH_t)$ is specified as a function of the current state variable ^QFR_t-1 (i.e. $k(^QH_t) = k(^QFR_t-1)$), it turns out to be an autonomous system equation, i.e. $^QFR_{t+1}-1 = f(^QFR_t-1)$ where $f(^QFR_t-1)$ is stationary and zero-input function of ^QFR_t-1. Hence, whether or not equation (5.18) is autonomous depends on how $k(^QH_t)$ is specified. In fact, the optimal control theory of dynamic programming specifies $k(^QH_t)$ as a function of the current state variable, i.e. $k(^QFR_t-1)$ (as illustrated in the following section 5.4.3.3).

Therefore, $k(^QH_t)$ violates Requirement A1 in section 5.3.4.2, so the algebraic approach is inapplicable. In this case, dynamic programming (DP) approach of optimal control theory is probably the best alternative since unlike the algebraic approach, it provides a systematic procedure for making a sequence of interrelated decisions at a sequence of times by decomposing control problems into multistage decision processes [see, section 4.2.5] (- this will be made clear in section 5.4.4).

In section 5.4.2, we construct two distinct quasi-stationary LQP optimisation problems and then we explore their solution in section 5.4.3. Further, numerical examples are illustrated in section 5.4.4.

5.4.2 Quasi-stationary LQP optimisation problems

Prior to constructing our control problem, we note that using the dynamic programming (DP) approach, the stability problem of the quasi-autonomous controlled object would not be serious because the DP approach is based on Bellman's principle of optimality (introduced earlier in section 4.2.5), by which the control error sequence $\{^QFR_t-1;\ t\in[0, \infty)\}$ is controllable (- this will be clarified in subsection (iii) of section 5.4.5.1). Consequently, different from using the algebraic approach [see Requirement A2 in section 5.3.4.2], we do not need here to impose any stabilising condition on $\{{}^{Q}FR_t\text{-}1;\ t\in[0, \infty)\}$.

However, we have to guarantee the finiteness of ${}^{Q}IPI_{\theta}$ (i.e. quasi-stationary version of IPI_{θ}, see section 5.2.1.1) where ${}^{Q}IPI_{\theta} = \sum_{t=0}^{\infty} \{ e^{-\eta t} \cdot [\theta + (1\text{-}\theta) \cdot k({}^{Q}H_t)^2] \cdot [{}^{Q}FR_t - 1]^2 \}$.

(a) Quasi-stationary controlling parameter space I:

The quasi-stationary version of the general controlling parameter space (5.4) is given as follows: let u be some positive real number, then for all t,

$$\{k({}^{Q}H_t): 0 \le k({}^{Q}H_t) \le 1,\ \eta > 0 \text{ and } 0 \le [\theta + (1\text{-}\theta) \cdot k({}^{Q}H_t)^2] \cdot [{}^{Q}FR_t\text{-}1]^2 \le u\}, \qquad \text{--- (5.19)}$$

on which ${}^{Q}IPI_{\theta}$ is well defined with a positive function value $< \infty$ for all t.

At this point, it is worth noting that even though constraint '$0 \le k({}^{Q}H_t) \le 1$' imposes a somewhat strong restriction on each controlling parameter $k({}^{Q}H_t)$, it is not necessary for securing a real-valued function ${}^{Q}IPI_{\theta}$, so '$0 \le k({}^{Q}H_t) \le 1$' can be regarded as a dummy (or additional) constraint for our control optimisation with respect to $k({}^{Q}H_t)$. For this reason, the next controlling parameter space is specified only to guarantee of the finiteness of ${}^{Q}IPI_{\theta}$.

(b) Quasi-stationary controlling parameter space II:

A possible quasi-stationary version of the general controlling parameter space (5.4) can be defined as follows: let u be some positive real number, then for all t,

$$\{k({}^{Q}H_t): \eta > 0 \text{ and } 0 \le [\theta + (1\text{-}\theta) \cdot k({}^{Q}H_t)^2] \cdot [{}^{Q}FR_t\text{-}1]^2 \le u\} \qquad \text{--- (5.20)}$$

; this space ensures the finiteness of ${}^{Q}IPI_{\theta}$, without imposing restrictions on the controlling parameters because if ${}^{Q}FR_t\text{-}1=0$, then $\{k({}^{Q}H_t): |k({}^{Q}H_t)| < \infty\}$, otherwise $\{k({}^{Q}H_t): |k({}^{Q}H_t)| \le u_1\}$, $u_1 = \sqrt{\{(u - \theta)/[(1\text{-}\theta)\cdot({}^{Q}FR_t\text{-}1)^2]\}}$.

Therefore, these two controlling parameter spaces are quite distinct from one another. So, it would be worth investigating the impact of each space on the control optimisation procedure with respect to the controlling parameters. The following quasi-stationary LQP optimisation problem is formulated in order to consider this matter separately.

$$\underset{\{k(^{Q}H_t);\ t = 0, 1, 2, \ldots\}}{\text{Min}} \left\{ \sum_{t=0}^{\infty} e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k(^{Q}H_t)^2] \cdot [^{Q}FR_t - 1]^2 \right\}$$

[Constraints set I]: given $\theta \in (0, 1)$; quasi-stationary system equation (5.18); and

$k(^{Q}H_t) \in$ {quasi-stationary controlling parameter space I (5.19)}; or

[Constraints set II]: given $\theta \in (0, 1)$; quasi-stationary system equation (5.18); and

$k(^{Q}H_t) \in$ {quasi-stationary controlling parameter space II (5.20)}.

--- (5.21)

Two distinct optimisation procedures, dealing with the control problem (5.21) subject to [Constraints set I] and [Constraints set II] separately, will be considered in sections 5.4.4 and 5.4.5, respectively.

5.4.3 Introduction to control optimisation

We elaborate upon the dynamic programming (DP) approach (based on Bellman's principle of optimality), as a mathematical solution tool of the control problem (5.21).

As a preliminary to investigating the rigorous optimisation procedure for the control problem (5.21), the following two subsections 5.4.3.1 and 5.4.3.2 will provide a framework for control optimisation with respect to quasi-stationary spread parameters.

5.4.3.1 Purposes

First of all, it would be helpful to notice the purposes of dealing with the quasi-stationary control problem (5.21) subject to two distinct constraint sets, separately. They are

- to demonstrate the insolubility of the control problem (5.21) subject to [Constraints set I] (- as an illustration of the disadvantages of the DP approach, see section 5.4.4);

- to illustrate how the DP approach to the control problem (5.21) subject to [Constraints set II] optimises the value of the controlling parameter $k({}^{Q}H_t)$ at each time t (- as an illustration of the advantages of the DP approach, see section 5.4.5); and

- to give some results useful for the threshold LQP control problem (to be considered later in 5.5.5).

Accordingly, an objective is to provide an introduction to spread parameter control optimisation (in the light of optimal control theory) as well as to illustrate some aspects related to the DP approach and the associated mathematical concepts.

5.4.3.2 Forward dynamic programming approach

Throughout Chapter 4, we have been accustomed to the backward dynamic programming (BDP) approach but we are here concerned with the forward dynamic programming (FDP) approach. The following descriptions (i)~(iii) provide a fundamental framework for solving the control problem (5.21).

(i) General discussion of the FDP approach:

The DP approach is classified into two distinct groups - BDP and FDP. The BDP approach, using mathematical induction with a boundary condition, is generally convenient in finite-horizon control problems, while it is necessary to employ the FDP approach, using mathematical deduction with a starting condition in infinite-horizon control problems for the reason that in general, we are unable to determine the boundary condition for the BDP approach to a infinite-horizon control problem [see Bertsekas (1987; pp 24~25 and 180~181)]. Accordingly, we adapt the FDP approach to our infinite-horizon control problem (5.21).

Considering the control optimisation at time $t \in [0, \infty)$ and applying the FDP approach to our control optimisation problem (5.21), the dynamic states ${}^{Q}FR_0$-1, ${}^{Q}FR_1$-1, ..., ${}^{Q}FR_t$-1 have been observed and the controlling parameters $k({}^{Q}H_0)$, $k({}^{Q}H_1)$, ..., $k({}^{Q}H_{t-1})$ have been optimally determined; hence, the controlling parameter $k({}^{Q}H_t)$ is required to be optimally determined (i.e. search for the optimal value of $k({}^{Q}H_t)$, denoted by $k^*({}^{Q}H_t)$).

We can then rewrite our performance index ${}^{Q}IPI_\theta$ as the sum of two parts as in the control optimisation for finite deterministic LQP optimisation problem examined in section 4.3: that is,

$$
\begin{aligned}
{}^{Q}IPI_\theta &= \sum_{s=0}^{\infty} \{ e^{-\eta s} \cdot (\theta + (1-\theta) \cdot k({}^{Q}H_s)^2) \cdot ({}^{Q}FR_s - 1)^2 \} \\
&= {}^{Q}IPIA_\theta + {}^{Q}IPIB_\theta \qquad \text{--- (5.22)}
\end{aligned}
$$

where

$$
{}^{Q}IPIA_\theta \equiv \sum_{s=0}^{t-1} \{ e^{-\eta s} \cdot (\theta + (1-\theta) \cdot k({}^{Q}H_s)^2) \cdot ({}^{Q}FR_s - 1)^2 \} \text{ and}
$$

$$
{}^{Q}IPIB_\theta \equiv \sum_{s=t}^{\infty} \{ e^{-\eta s} \cdot (\theta + (1-\theta) \cdot k({}^{Q}H_s)^2) \cdot ({}^{Q}FR_s - 1)^2 \}.
$$

We note that by letting ${}^{Q}IPIC_\theta = e^{\eta t} \cdot {}^{Q}IPIB_\theta$, ${}^{Q}IPIC_\theta$ can be interpreted to be the future cost at time t discounted to time t, while ${}^{Q}IPIB_\theta$ can be interpreted to be the future cost at time t discounted to time 0 (- this transform will be convenient in the later mathematical discussions).

The first part ${}^{Q}IPIA_{\theta}$ does not depend on the decisions to be made, i.e. $k({}^{Q}H_t)$, $k({}^{Q}H_{t+1})$, ..., so minimising ${}^{Q}IPI_{\theta}$ with respect to these controlling parameters is equivalent to minimising ${}^{Q}IPIC_{\theta}$. Furthermore, the second part ${}^{Q}IPIB_{\theta}$ can be expressed as a function of the current dynamic state ${}^{Q}FR_t$-1, since the sequence of future dynamic states $\{{}^{Q}FR_{t+1}-1, {}^{Q}FR_{t+2}-1, \ldots\}$ is recursively generated from the current dynamic state ${}^{Q}FR_t$-1 by the first-order system equation (5.18). Also, the decision at time t, $k({}^{Q}H_t)$, will be a function of the current state variable ${}^{Q}FR_t$-1 because using the dynamic programming approach based on Bellman's principle of optimality, the knowledge of the current dynamic state is enough to substitute for H_t, i.e. $k({}^{Q}H_t) = k({}^{Q}FR_t-1)$ for all t.

(ii) Possible spread rule for determining the value of the spread parameter in the case of 100% funding level:

Prior to constructing an optimisation procedure at time t with respect to $k({}^{Q}H_t)$, we need to check whether or not the value of the starting/current dynamic state ${}^{Q}FR_t$-1 is zero. That is, if the value of ${}^{Q}FR_t$-1 is zero, we do not need to consider the optimisation procedure at time t because the controlling parameter $k({}^{Q}H_t)$ appears as $k({}^{Q}H_t)^2 \cdot ({}^{Q}FR_t-1)^2$ in ${}^{Q}IPIB_{\theta}$, so the value of ${}^{Q}FR_t$-1 being zero implies that we can choose the optimal value of $k({}^{Q}H_t)$ arbitrarily. This decision would be meaningless in view of our seeking to set up a systematic and unique optimisation procedure with respect to the controlling parameters. In the light of optimal control theory, we can say that the case of a 100% funding level reflects the weakness in the mathematical formulation within the framework of the Spread method. In order to avoid this kind of undesirable situation, it is necessary to make such a possible spread rule that if the funding level is 100% at some time in a process of sequential optimisation, then we set the optimal value of the corresponding controlling parameter at the level of zero; here, the term 'possible' is added to reflect that the rule is consistent with the conceptual purpose of the Spread method, that is, as mentioned in section 3.4.4, the mathematical formulation of the

Spread method is basically designed for penalising the unfunded ratio (i.e. the funding level that is not 100%). So, whenever the funding level reaches the level of 100% we should not place any penalty on 100% funding level. Without loss of generality, this rule will be applied to all the remaining sections of this chapter.

Simply applying the possible spread rule to the optimisation procedure at time t, we will consider the optimisation procedure at time t+1 for the new starting state ${}^{Q}FR_{t+1}$-1 and so on; in particular, if $\delta = \eta$ and ${}^{Q}FR_t$=100% (i.e. initially fully funded and no difference between the actual experience and actuarial assumptions), the optimisation procedure over $[t, \infty)$ is completely determined by the possible spread rule due to ${}^{Q}FR_j-1 = 0$ for all $j \geq t$ from the system equation (5.18).

(iii) Notation for the FDP approach to the control optimisation problem (5.21):

Consider time $t \in [0, \infty)$ and for convenience, assume ${}^{Q}FR_t-1 \in R^1-\{0\}$. The following notations are useful for constructing our optimisation procedure at time t with respect to $k({}^{Q}H_t)$ by using the FDP approach to the control optimisation problem (5.21) - decomposing the control problem (5.21) into an infinite-stage decision process and applying mathematical deduction based on Bellman's principle of optimality.

These mathematical notations will be used in sections 5.4.4 and 5.4.5 and their respective mathematical interpretations will be more apparent later:

n = stage index with non-negative integer, for example, the term 'stage n' corresponds to the unit control period [t+n-1, t+n); in particular, 'stage 0' is the starting stage corresponding to the starting control time $t \in [0, \infty)$;

${}^{Q}FR_t$-1 = starting/current state variable at time t, whose value is assumed to be observable at time t and to be non-zero, i.e. ${}^{Q}FR_t-1 \in R^1-\{0\}$ at time t;

${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$ = sub-optimal function value at time t of control problem (5.21) only processed from stage 0 up to stage n for any ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$ at time t, defined by

$${}^{Q}J_n({}^{Q}FR_t\text{-}1, t) = e^{\eta t} \cdot \underset{\{k({}^{Q}H_s;\ s=t,\ t+1,\ \ldots,\ t+n-1\}}{\text{Min}} \left\{ \sum_{s=t}^{t+n-1} [e^{-\eta s} \cdot (\theta + (1\text{-}\theta) \cdot k({}^{Q}H_s)^2) \cdot ({}^{Q}FR_s\text{-}1)^2] \right\},$$ which

can be interpreted as the minimal future cost over [t, t+n) discounted to time t. In particular, we can let ${}^{Q}J_0({}^{Q}FR_t\text{-}1, t) = 0$ as a starting condition of ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$, since ${}^{Q}FR_t\text{-}1$ is not involved in the past cost ${}^{Q}IPIA_\theta$;

${}^{Q}J^*({}^{Q}FR_t\text{-}1, t)$ = optimal function value at time t of control problem (5.21) for any ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$, defined by the equation ${}^{Q}J^*({}^{Q}FR_t\text{-}1, t) = \underset{\{k({}^{Q}H_s);\ s=t,\ t+1,\ \ldots\}}{\text{Min}} {}^{Q}IPIC_\theta$, which can be interpreted as the minimal future cost discounted at time t over [t, ∞). We will prove in Proposition 5.1 of section 5.4.5.3 that for any ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$, $J^*({}^{Q}FR_t\text{-}1, t)$ is equivalent to the limit function value of the sequence $\{J_n({}^{Q}FR_t\text{-}1, t);\ n=0, 1, 2, \ldots\}$ (denoted by $\lim_{n\to\infty} J_n({}^{Q}FR_t\text{-}1, t)$);

$k_n({}^{Q}FR_t\text{-}1)$ = sub-optimal value of the controlling parameter $k({}^{Q}H_t)$ determining ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$, which can be thought of as a sub-optimal decision value processed from stage 0 up to stage n for any ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$; and

$k^*({}^{Q}FR_t\text{-}1)$ = optimal value of the controlling parameter $k({}^{Q}H_t)$ determining ${}^{Q}J^*({}^{Q}FR_t\text{-}1)$, that is, $k^*({}^{Q}FR_t\text{-}1) = k^*({}^{Q}H_t)$ owing to Bellman's principle of optimality, which can be thought of as an optimal decision value processed from stage 0 up to state ∞ for any ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$. From the relation ${}^{Q}J^*({}^{Q}FR_t\text{-}1, t) = \lim_{n\to\infty} {}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$, we can derive the equation, for each ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$,

$$k^*({}^{Q}FR_t\text{-}1) = \lim_{n\to\infty} k_n({}^{Q}FR_t\text{-}1).$$

We note finally that the term 'sub-optimal' is added to emphasise that both $J_n(.,.)$ and $k_n(.)$ are not our final optimal function but intermediate optimal functions processed up to stage n.

(iv) A brief sketch of the FDP approach to control problem (5.21):

Even though all these descriptions are clearly verified in the following section 5.4.5.1, we shall here sketch out the main points of our optimisation procedure with respect to the controlling parameter $k({}^{Q}H_t)$.

We first construct a N-multistage optimisation procedure for any ${}^{Q}FR_t\text{-}1 \in R^1\text{-}\{0\}$ at time $t\in[0, \infty)$: by using the mathematical deduction - proceeding from the starting condition ${}^{Q}J_0({}^{Q}FR_t\text{-}1, t)=0$ to higher values of the index 'n' in ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$, using the recurrence relationship between ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$ and ${}^{Q}J_{n-1}({}^{Q}FR_t\text{-}1, t)$ for n = 1, 2, ..., N. Then, we obtain $k_n({}^{Q}FR_t\text{-}1)$ corresponding to ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$.

Next, we would extend this optimisation procedure to the infinite-stage optimisation procedure by using the uniform boundedness condition, i.e. $0 \leq [\theta+(1-\theta)\cdot k({}^{Q}H_j)^2] \cdot [{}^{Q}FR_j\text{-}1]^2 \leq u$ for all j=t+1, t+2, ..., and $\eta > 0$: in other words, as $n \rightarrow \infty$, then ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t) \rightarrow {}^{Q}J^*({}^{Q}FR_t\text{-}1, t)$, so $k_n({}^{Q}FR_t\text{-}1) \rightarrow k^*({}^{Q}FR_t\text{-}1) = k^*({}^{Q}H_t)$, expressed as a function of non-zero ${}^{Q}FR_t\text{-}1$.

Further, if the value of ${}^{Q}FR_t\text{-}1$ is zero, then we apply the possible spread rule such that $k^*({}^{Q}FR_t\text{-}1) = 0$, as described in the above subsection (ii).

5.4.4 Control optimisation under [Constraints set I]

- an illustration of the inappropriateness of using the FDP approach

The practical usefulness of the FDP approach is occasionally limited by computational intractability, in particular when concerned with an infinite-horizon control problem with some strong constraints on the controlling variables (or parameters). As an illustration of this kind of weakness of the FDP approach, we shall now show that the FDP approach to the control problem (5.21) subject to [Constraints set I] is incapable of providing a complete control optimisation procedure. To begin with, it is worth recalling that for all t, $k({}^{Q}H_t)$ should satisfy the condition $0 \leq k({}^{Q}H_t) \leq 1$ specified in [Constraints set I].

5.4.4.1 Control optimisation procedure

Consider time $t \in [0, \infty)$. We employ the notations and follow the control optimisation procedure introduced previously in section 5.4.3.2. We are then able to derive the following recurrence relationship between ${}^{Q}J_n({}^{Q}FR_t\text{-}1,\ t)$ and ${}^{Q}J_{n-1}({}^{Q}FR_t\text{-}1,\ t)$ for n = 1, 2, 3, ... with given starting condition ${}^{Q}J_0({}^{Q}FR_t\text{-}1,\ t)=0$.

Firstly, if proceeding from stage 0 up to stage 1, then

$${}^{Q}J_1({}^{Q}FR_t-1,\ t) = \min_{0\leq k({}^{Q}H_t)\leq 1} [(\theta+(1-\theta)\cdot k({}^{Q}H_t)^2)\cdot({}^{Q}FR_t-1)^2 + {}^{Q}J_0({}^{Q}FR_t-1,\ t)] = \theta\cdot({}^{Q}FR_t-1)^2$$

$$\equiv Q_{11}\cdot({}^{Q}FR_t-1)^2 + Q_{12}\cdot({}^{Q}FR_t-1) + Q_{13} \text{ with } Q_{11}=\theta \text{ and } Q_{12}, Q_{13}=0$$

; hence, the sub-optimal decision value proceeding from stage 0 up to stage 1 was uniquely given by $k_1({}^{Q}FR_t\text{-}1) = 0$, since $(1\text{-}\theta)\cdot({}^{Q}FR_t\text{-}1)^2 > 0$.

Secondly, if proceeding from stage 0 up to state 2, then

$${}^{Q}J_2({}^{Q}FR_t-1)$$

$$= \min_{0\leq k({}^{Q}H_t),\ k({}^{Q}H_{t+1})\leq 1} \{ (\theta+(1-\theta)\cdot k({}^{Q}H_t)^2)\cdot({}^{Q}FR_t-1)^2 + e^{-\eta}\cdot[\theta+(1-\theta)\cdot k({}^{Q}H_{t+1})^2]\cdot({}^{Q}FR_{t+1}-1)^2 \}$$

$$= \min_{0\leq k({}^{Q}H_t)\leq 1} \{(\theta+(1-\theta)\cdot k({}^{Q}H_t)^2)\cdot({}^{Q}FR_t-1)^2 + e^{-\eta}\cdot\{ \min_{0\leq k({}^{Q}H_{t+1})\leq 1} [(\theta+(1-\theta)\cdot k({}^{Q}H_{t+1})^2)\cdot({}^{Q}FR_{t+1}-1)^2]\}\}$$

$$= \min_{0\leq k({}^{Q}H_t)\leq 1} \{ (\theta + (1-\theta)\cdot k({}^{Q}H_t)^2)\cdot({}^{Q}FR_t-1)^2 + e^{-\eta}\cdot {}^{Q}J_1(q_1\cdot(1-k({}^{Q}H_t))\cdot({}^{Q}FR_t-1)+q_2,\ t) \}$$

$$= \min_{0\leq k({}^{Q}H_t)\leq 1} \{ [((1-\theta) + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\cdot({}^{Q}FR_t-1)^2]\cdot k({}^{Q}H_t)^2 - 2[(e^{-\eta}\cdot Q_{11}\cdot q_1^2)\cdot({}^{Q}FR_t-1)^2 +$$

$$e^{-\eta}\cdot Q_{11}\cdot q_1\cdot q_2\cdot({}^{Q}FR_t-1)]\cdot k({}^{Q}H_t) + [(\theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\cdot({}^{Q}FR_t-1)^2 +$$

$$2e^{-\eta}\cdot Q_{11}\cdot q_1\cdot q_2\cdot({}^{Q}FR_t-1) + e^{-\eta}\cdot Q_{11}\cdot q_2^2] \}.$$

Then, the sub-optimal decision value proceeding from stage 0 up to stage 2, i.e. $k_2({}^{Q}FR_t\text{-}1)$, will be conditionally determined because $k({}^{Q}H_t) \in \{k({}^{Q}H_t): 0 \leq k({}^{Q}H_t) \leq 1\}$. So, letting kk be the unrestricted sub-optimal decision value in ${}^{Q}J_2({}^{Q}FR_t\text{-}1, t)$, we should consider three disjoint events separately, say Event1 $\equiv$ {kk: $0 \leq kk \leq 1$}, Event2 $\equiv$ {kk: $kk < 0$} and Event3 $\equiv$ {kk: $kk>1$}, in which kk is uniquely determined by the form,

$$kk = [(e^{-\eta}\cdot Q_{11}\cdot q_1^2)\cdot({}^{Q}FR_t\text{-}1) + e^{-\eta}\cdot Q_{11}\cdot q_1\cdot q_2] \,/\, [((1\text{-}\theta) + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\cdot({}^{Q}FR_t\text{-}1)],$$

since $((1\text{-}\theta) + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\cdot({}^{Q}FR_t\text{-}1)^2 > 0$.

Hence, $k_2({}^{Q}FR_t\text{-}1)$ would be determined subject to Event1, Event2 or Event3 (each are denoted by $k_2({}^{Q}FR_t\text{-}1$; Event1), $k_2({}^{Q}FR_t\text{-}1$; Event2) or $k_2({}^{Q}FR_t\text{-}1$; Event3)):

(a) In the case of Event1, $k_2({}^{Q}FR_t\text{-}1$; Event1) = kk, and further, the corresponding sub-optimal function value ${}^{Q}J_2({}^{Q}FR_t\text{-}1$, t; Event1) is given by

$$\begin{aligned} {}^{Q}J_2({}^{Q}FR_t\text{-}1, t;\ \text{Event1}) = {} & \{(\theta\cdot(1\text{-}\theta) + e^{-\eta}\cdot Q_{11}\cdot q_1^2) / (1 - \theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\} \cdot ({}^{Q}FR_t\text{-}1)^2 + \\ & \{(2(1\text{-}\theta)\cdot e^{-\eta}\cdot Q_{11}\cdot q_1\cdot q_2) / (1 - \theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\} \cdot ({}^{Q}FR_t\text{-}1) + \\ & \{((1\text{-}\theta)\cdot e^{-\eta}\cdot Q_{11}\cdot q_2^2) / (1 - \theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\}; \end{aligned}$$

(b) In the case of Event2, $k_2({}^{Q}FR_t\text{-}1$; Event2) = 0 and then ${}^{Q}J_2({}^{Q}FR_t\text{-}1$, t; Event2) $\neq$ ${}^{Q}J_2({}^{Q}FR_t\text{-}1$, t; Event1); otherwise

(c) In the case of Event3, $k_2({}^{Q}FR_t\text{-}1$; Event3)=1 and then ${}^{Q}J_2({}^{Q}FR_t\text{-}1$, t; Event3) $\neq$ ${}^{Q}J_2({}^{Q}FR_t\text{-}1$, t; Event2) $\neq$ ${}^{Q}J_2({}^{Q}FR_t\text{-}1$, t; Event1).

Without any difficulty, we can envisage that the total number of possible sub-optimal decision value per stage will be three. So, by the recurrence relationship between ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$ and ${}^{Q}J_{n-1}({}^{Q}FR_t\text{-}1, t)$, it should be necessary to consider up to 3^{n-1} combinations until reaching stage n in order to find $k_n({}^{Q}FR_t\text{-}1)$. Accordingly, the number of possible combinations is increasing geometrically as $n \to \infty$. Consequently, we can not obtain the optimal value of $k(H_t)$ (denoted by $k^*(H_t)$) because $k^*({}^QH_t)$ is determined by $\lim_{n\to\infty} k_n({}^QFR_t\text{-}1)$. Since the FDP approach produces a sequence of interrelated decisions, we can not obtain the optimal control sequence such as $\{k^*(H_0), k^*(H_1), k^*(H_2), \ldots\}$ because of this computational infeasibility.

5.4.4.2 Conclusion

The resulting basic method of enumeration to find $k_n({}^QFR_t\text{-}1)$ can then be described diagrammatically, by a decision tree (as given in Appendix 5B.3), and we can clearly demonstrate the insolubility of control problem (5.21) subject to [Constraints set I]. In general, the problem of computational dimensionality may be a prototype illustration of the limitation of the DP approach, both FDP and BDP, when considering infinite-horizon (including a large-finite-horizon) control problems with strong boundary constraints on the controlling variables (or parameters), e.g. $0 \leq k(H_t) \leq 1$ for each time t. In this respect, we can conclude in respect of adapting the DP approach to control problems, that the larger the number of strong boundary constraints on the controlling variables (or parameters) and the longer the control horizon, the higher the potential risk of insolubility of the corresponding control problems.

5.4.5 Control optimisation under [Constraints set II]

- an illustration of the applicability of using the FDP approach

Adopting [Constraints set II] instead of [Constraints set I], we shall here show an advantage of using the FDP approach. In general, one of the great strengths of the DP approach, both FDP and BDP, is that a sequence of interrelated decisions can be set up which lead to the optimal control law with considerable computational savings, once an appropriate basic recurrence relationship for the DP calculations has been found and solved.

To begin with, we note that different from the previous discussion in section 5.4.4, here it is not necessary to take into account the constraint imposed on $k(H_t)$ because under [Constraints set II], we can set a sufficiently large real number u_1 so that the constraint that, for all t, $k(^QH_t) \in \{k(^QH_t): |k(^QH_t)| \leq u_1 < \infty)\}$, makes no impact on our control optimisation procedure with respect to $k(^QH_t)$ (as mentioned earlier in section 5.4.2).

5.4.5.1 Control optimisation procedure

In a similar manner as in section 5.4.4.1, we can derive the consecutive recurrence relationship between $^QJ_n(^QFR_t\text{-}1, t)$ and $^QJ_{n-1}(^QFR_t\text{-}1, t)$ for each n=0, 1, 2, ..., and in the case of the value of $^QFR_t\text{-}1$ being zero, we simply set the optimal value of $k(^QH_t)$ at the level of zero in accordance with our possible spread rule introduced in subsection (ii) of section 5.4.3.2. We note that the following algebraic results processed up to stage 2 are simply restated from those derived in the previous section 5.4.4.1.

Firstly, if proceeding from stage 0 up to stage 1, then

$$^QJ_1(^QFR_t\text{-}1, t) = \underset{k(^QH_t)}{\text{Min}} \{ (\theta+(1\text{-}\theta)\cdot k(^QH_t)^2)\cdot(^QFR_t\text{-}1)^2 + {^QJ_0}(^QFR_t\text{-}1, t) \}$$

$$= Q_{11}\cdot(^QFR_t\text{-}1)^2 + Q_{12}\cdot(^QFR_t\text{-}1) + Q_{13} \quad \text{with } Q_{11} = \theta \text{ and } Q_{12}, Q_{13} = 0; \text{ and}$$

$k_1(^QFR_t\text{-}1) = 0.$

Secondly, if proceeding from stage 0 up to stage 2, then

$$^QJ_2(^QFR_t\text{-}1, t) = \underset{k(^QH_t)}{\text{Min}} \{(\theta+(1\text{-}\theta)\cdot k(^QH_t)^2) \cdot (^QFR_t\text{-}1)^2 + e^{-\eta}\cdot {^QJ_1}(q_1\cdot(1\text{-}k(^QH_t))\cdot(^QFR_t\text{-}1)+q_2, t)\}$$

$$= \{(\theta\cdot(1\text{-}\theta) + e^{-\eta}\cdot Q_{11}\cdot q_1^2) / (1 - \theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\} \cdot (^QFR_t\text{-}1)^2 +$$

$$\{(2(1\text{-}\theta)\cdot e^{-\eta}\cdot Q_{11}\cdot q_1\cdot q_2) / (1 - \theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\} \cdot (^QFR_t\text{-}1) +$$

$$\{((1\text{-}\theta)\cdot e^{-\eta}\cdot Q_{11}\cdot q_2^2) / (1 - \theta + e^{-\eta}\cdot Q_{11}\cdot q_1^2)\}$$

$$\equiv Q_{21} \cdot ({}^{Q}FR_t\text{-}1)^2 + Q_{22} \cdot ({}^{Q}FR_t\text{-}1) + Q_{23};$$ and

$$k_2({}^{Q}FR_t\text{-}1) = \{(e^{-\eta} \cdot Q_{11} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1) + e^{-\eta} \cdot Q_{11} \cdot q_1 \cdot q_2\} / \{((1\text{-}\theta) + e^{-\eta} \cdot Q_{11} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1)\}.$$

Thirdly, if proceeding from stage 0 up to state 3, then

$${}^{Q}J_3({}^{Q}FR_t\text{-}1, t) = \underset{k({}^{Q}H_t),\, k({}^{Q}H_{t+1}),\, k({}^{Q}H_{t+2})}{\text{Min}} \{ (\theta+(1\text{-}\theta) \cdot k({}^{Q}H_t)^2) \cdot ({}^{Q}FR_t\text{-}1)^2 + e^{-\eta} \cdot [\theta+(1\text{-}\theta) \cdot k({}^{Q}H_{t+1})^2] \cdot$$

$$({}^{Q}FR_{t+1}\text{-}1)^2 + e^{-2\eta} \cdot [\theta+(1\text{-}\theta) \cdot k({}^{Q}H_{t+2})^2] \cdot ({}^{Q}FR_{t+2}\text{-}1)^2 \}$$

$$= \underset{k({}^{Q}H_t)}{\text{Min}} \{ (\theta+(1\text{-}\theta) \cdot k({}^{Q}H_t)^2) \cdot ({}^{Q}FR_t\text{-}1)^2 + e^{-\eta} \cdot \{ \underset{k({}^{Q}H_{t+1})}{\text{Min}} [(\theta+(1\text{-}\theta) \cdot k({}^{Q}H_{t+1})^2) \cdot$$

$$({}^{Q}FR_{t+1}\text{-}1)^2 + e^{-\eta} \cdot \underset{k({}^{Q}H_{t+2})}{\text{Min}} [(\theta+(1\text{-}\theta) \cdot k({}^{Q}H_{t+2})^2) \cdot ({}^{Q}FR_{t+2}\text{-}1)^2]] \} \}$$

$$= \underset{k({}^{Q}H_t)}{\text{Min}} \{ (\theta + (1\text{-}\theta) \cdot k({}^{Q}H_t)^2) \cdot ({}^{Q}FR_t\text{-}1)^2 + e^{-\eta} \cdot {}^{Q}J_2(q_1 \cdot (1\text{-}k({}^{Q}H_t)) \cdot ({}^{Q}FR_t\text{-}1)+q_2, t) \}$$

$$= \underset{k({}^{Q}H_t)}{\text{Min}} \{[((1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1)^2] \cdot k({}^{Q}H_t)^2 - 2[(e^{-\eta} \cdot Q_{21} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1)^2 +$$

$$e^{-\eta} \cdot q_1 \cdot (Q_{21} \cdot q_2 + Q_{22}/2) \cdot ({}^{Q}FR_t\text{-}1)] \cdot k({}^{Q}H_t) + [(\theta + e^{-\eta} \cdot Q_{21} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1)^2 +$$

$$2e^{-\eta} \cdot q_1 \cdot (Q_{21} \cdot q_2 + Q_{22}/2) \cdot ({}^{Q}FR_t\text{-}1) + e^{-\eta} \cdot (Q_{21} \cdot q_2^2 + Q_{22} \cdot q_2 + Q_{23})]$$

$$= \{[\theta \cdot (1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2] / [(1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2]\} \cdot ({}^{Q}FR_t\text{-}1)^2 +$$

$$\{[2(1\text{-}\theta) \cdot e^{-\eta} \cdot q_1 \cdot (Q_{21} \cdot q_2 + Q_{22}/2)] / [(1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2]\} \cdot ({}^{Q}FR_t\text{-}1) +$$

$$\{[(1\text{-}\theta) \cdot e^{-\eta} \cdot (Q_{21} \cdot q_2^2+Q_{22} \cdot q_2+Q_{23}) + e^{-2\eta} \cdot q_1^2 \cdot (Q_{21} \cdot Q_{23}\text{-}Q_{22}^2/4)] / [(1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2]\}$$

$$\equiv Q_{31} \cdot ({}^{Q}FR_t\text{-}1)^2 + Q_{32} \cdot ({}^{Q}FR_t\text{-}1) + Q_{33}$$

; hence, the sub-optimal decision value proceeding from stage 0 up to stage 3 is uniquely determined by the form

$$k_3({}^{Q}FR_t\text{-}1) = \{(e^{-\eta} \cdot Q_{21} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1) + e^{-\eta} \cdot q_1 \cdot (Q_{21} \cdot q_2+Q_{22}/2)\}$$

$$/ \{((1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1)\},$$ since $((1\text{-}\theta)+e^{-\eta} \cdot Q_{21} \cdot q_1^2) \cdot ({}^{Q}FR_t\text{-}1)^2 > 0$.

Proceeding by mathematical deduction, we can easily deduce the consecutive recurrence relationship between ${}^{Q}J_n({}^{Q}FR_t\text{-}1, t)$ and ${}^{Q}J_{n\text{-}1}({}^{Q}FR_t\text{-}1, t)$ for each n = 1, 2, 3, ... As described in section 5.4.3.2, we first confine our interests to establishing a general N-multistage optimisation procedure and then it will be extended to an infinite-stage optimisation procedure.

(i) N-multistage optimisation procedure (N positive integer):

Proceeding recursively in a similar manner to the above, we obtain the following sub-optimal functional equation from stage 0 up to stage n

${}^{Q}J_0({}^{Q}FR_t\text{-}1, t) = 0$ (given as a starting condition); and

$$ {}^{Q}J_n({}^{Q}FR_t\text{-}1, t) = \min_{k({}^{Q}H_t)} \{(\theta+(1\text{-}\theta)\cdot k(H_t)^2)\cdot({}^{Q}FR_t\text{-}1)^2 + e^{-\eta}\cdot {}^{Q}J_{n-1}(q_1\cdot(1\text{-}k({}^{Q}H_t))\cdot({}^{Q}FR_t\text{-}1)+q_2, t)\} $$

$$ \equiv Q_{n1}\cdot({}^{Q}FR_t\text{-}1)^2 + Q_{n2}\cdot({}^{Q}FR_t\text{-}1) + Q_{n3} \text{ for } n = 1, 2, \ldots, N \qquad \text{--- (5.23)} $$

where

the coefficients of $({}^{Q}FR_t\text{-}1)^2$, $({}^{Q}FR_t\text{-}1)$ and $({}^{Q}FR_t\text{-}1)^0$ each is given recursively by the three simultaneous recursions

$$ Q_{n\,1} = [\theta\cdot(1\text{-}\theta) + e^{-\eta}\cdot Q_{n-1\,1}\cdot q_1^2] / [(1\text{-}\theta) + e^{-\eta}\cdot Q_{n-1\,1}\cdot q_1^2] $$

with $Q_{11} = \theta$ (initially obtained from ${}^{Q}J_1({}^{Q}FR_t\text{-}1, t)$); --- (5.24)

$$ Q_{n\,2} = [2(1\text{-}\theta)\cdot e^{-\eta}\cdot q_1\cdot(Q_{n-1\,1}\cdot q_2 + Q_{n-1\,2}/2)] / [(1\text{-}\theta) + e^{-\eta}\cdot Q_{n-1\,1}\cdot q_1^2] $$

with $Q_{12} = 0$ (initially obtained from ${}^{Q}J_1({}^{Q}FR_t\text{-}1, t)$); and --- (5.25)

$$ Q_{n\,3} = [(1\text{-}\theta)\cdot e^{-\eta}\cdot(Q_{n-1\,1}\cdot q_2^2 + Q_{n-1\,2}\cdot q_2 + Q_{n-1\,3}) + e^{-2\eta}\cdot q_1^2\cdot(Q_{n-1\,1}\cdot Q_{n-1\,3} - Q_{n-1\,2}^{\;2}/4)] $$

$/ [(1\text{-}\theta) + e^{-\eta}\cdot Q_{n-1\,1}\cdot q_1^2]$ with $Q_{13} = 0$ (initially obtained from $J_1({}^{Q}FR_t\text{-}1)$). --- (5.26)

Hence, the above result implies that the sub-optimal decision value proceeding from stage 0 up to stage n, determining ${}^{Q}J_n({}^{Q}FR_0\text{-}1, t)$, is uniquely given (i.e. single-valued) in the form:

$$ k_n({}^{Q}FR_t\text{-}1) = [(e^{-\eta}\cdot q_1^2\cdot Q_{n-1\,1})\cdot({}^{Q}FR_t\text{-}1) + e^{-\eta}\cdot q_1\cdot(q_2\cdot Q_{n-1\,1} + Q_{n-1\,2}/2)] / $$

$$ [((1\text{-}\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{n-1\,1})\cdot({}^{Q}FR_t\text{-}1)], \qquad \text{--- (5.27)} $$

since $((1-\theta)+e^{-\eta}\cdot Q_{n-1\,1}\cdot q_1^{2})\cdot({}^{Q}FR_t-1)^2] > 0$.

At this point, we should note that the function value ${}^{Q}J_n({}^{Q}FR_t-1, t)$ does not depend explicitly on time t, as shown in equation (5.23), which could be expected from the fact that the FDP approach specifies $k({}^{Q}H_t)$ as a function of ${}^{Q}FR_t-1$, so the quasi-autonomous system equation (5.18) turns out to be autonomous (i.e. its motions are invariant for a translation of time, see section 5.4.1). Also, the time-indexed discount function $e^{-\eta s}$ in the performance index per stage, i.e. $e^{-\eta s}\cdot[\theta+(1-\theta)\cdot k({}^{Q}H_s)^2]\cdot({}^{Q}FR_s-1)^2$, s = t, t+1, ..., appears in the form of $e^{-\eta}$ in the consecutive recurrence relationship between ${}^{Q}J_n({}^{Q}FR_t-1, t)$ and ${}^{Q}J_{n-1}({}^{Q}FR_t-1, t)$ for each n = 1, 2, ... N. We shall henceforth re-denote ${}^{Q}J_n({}^{Q}FR_t, t)$ as ${}^{Q}J_n({}^{Q}FR_t)$ and accordingly, ${}^{Q}J^*({}^{Q}FR_t, t)$ as ${}^{Q}J^*({}^{Q}FR_t)$ because the time variable t turns out to be a dummy variable in the application of the FDP approach.

Consequently, for each time t, we do not need to derive and extend the above sub-optimal functional equation (5.23) to the case of an infinite number of stages for its corresponding optimal functional equation (5.28) in subsection (ii), due to the time-invariant nature of the function ${}^{Q}J_n(.)$: otherwise, we need to derive and solve an infinite number of optimal functional equations, which will lead to the insolubility of the control problem (5.21) - this subject will be considered later in section 5.5 in relation to the non-stationary control problem.

Next, we consider the extension of the N-multistage optimisation procedure, specified by formulae (5.23) ~ (5.27), to an infinite-stage optimisation procedure.

(ii) <u>Extending to infinite-stage optimisation procedure</u>:

In order to extend the N-multistage to the infinite-stage, a natural question that arises now is whether or not the optimal function value proceeding from stage 0 up to stage n, i.e. ${}^{Q}J_n({}^{Q}FR_t-1)$, converges monotonically to ${}^{Q}J^*({}^{Q}FR_t-1)$ as $n \to \infty$. The following proposition provides the answer for this requirement [see Bertsekas (1976; section 6.1)].

Proposition 5.1 (Convergence of N-multistage optimisation procedure): for any non-zero ${}^{Q}FR_t$ - 1 at time $t \in [0, \infty)$,

P1. Uniform boundedness of ${}^{Q}J^*({}^{Q}FR_t\text{-}1)$: $0 < {}^{Q}J^*({}^{Q}FR_t\text{-}1) \leq u / (1\text{-}e^{-\eta})$;

P2. Increasing sequence $\{{}^{Q}J_n({}^{Q}FR_t\text{-}1);\ n = 0, 1, 2, \ldots\}$: for each n,

$${}^{Q}J_0({}^{Q}FR_t\text{-}1) \leq {}^{Q}J_1({}^{Q}FR_t\text{-}1) \leq \ldots \leq {}^{Q}J_n({}^{Q}FR_t\text{-}1) \leq \ldots \leq \lim_{n\to\infty} {}^{Q}J_n({}^{Q}FR_t\text{-}1); \text{ and}$$

P3. Unique convergence of $\{{}^{Q}J_n({}^{Q}FR_t\text{-}1);\ n = 0, 1, 2, \ldots\}$ to ${}^{Q}J^*({}^{Q}FR_t\text{-}1)$:

$$\lim_{n\to\infty} {}^{Q}J_n({}^{Q}FR_t\text{-}1) = {}^{Q}J^*({}^{Q}FR_t\text{-}1).$$

Proof (note that these proofs are largely adapted from Bertsekas (1976; section 6.1))

By combining the assumption specified in [Constraints set II], i.e. $0 \leq (\theta+(1\text{-}\theta)\cdot k(H_s)^2)\cdot(FR_s\text{-}1)^2 \leq u$ for all s, $\theta \in (0, 1)$ and ${}^{Q}FR_t\text{-}1 \neq 0$, then we have $0 < (\theta+(1\text{-}\theta)\cdot k(H_t)^2)\cdot(FR_t\text{-}1)^2 \leq u$ at time t and $0 \leq (\theta+(1\text{-}\theta)\cdot k(H_j)^2)\cdot(FR_j\text{-}1)^2 \leq u$ for all j = t+1, t+2, t+3, ..., so the first part P1 can be proved as follows:

$$0 < \sum_{s=t}^{\infty} e^{-\eta\cdot(s-t)} \cdot (\theta + (1\text{-}\theta)\cdot k({}^{Q}H_s)^2) \cdot ({}^{Q}FR_s\text{-}1)^2$$

$$\leq \left\{ \sum_{s=t}^{t+n-1} e^{-\eta\cdot(s-t)} \cdot (\theta + (1\text{-}\theta)\cdot k({}^{Q}H_s)^2) \cdot ({}^{Q}FR_s\text{-}1)^2 \right\} + u \cdot \sum_{s=t+n}^{\infty} e^{-\eta\cdot(s-t)}$$

By taking the minimum for each $k({}^{Q}H_s)$ of both sides and using $\eta > 0$, we obtain the inequality: $0 < {}^{Q}J^*({}^{Q}FR_t\text{-}1) \leq {}^{Q}J_n({}^{Q}FR_t\text{-}1) + u\cdot e^{-\eta\cdot n} / (1\text{-}e^{-\eta})$, in which ${}^{Q}J_n({}^{Q}FR_t\text{-}1)$ takes non-negative value for any n = 0, 1, 2, ... Hence, putting n = 0, then we have the required result because ${}^{Q}J_0({}^{Q}FR_t\text{-}1) = 0$, i.e. $0 < {}^{Q}J^*({}^{Q}FR_t\text{-}1) \leq u / (1\text{-}e^{-\eta})$.

Next, the second part P2 can be proved as follows: using $0 \leq (\theta+(1\text{-}\theta)\cdot k({}^{Q}H_j)^2)\cdot(FR_j\text{-}1)^2 \leq u$ for all j=t+1, t+2, ... and the consecutive recurrence relationship (5.23), then P2 is clear.

The existence of the limit in P3 can be proved as follows: by combining P1 and P2, i.e. $J_n(^{Q}FR_t\text{-}1) \leq J^*(^{Q}FR_t\text{-}1) \leq J_n(^{Q}FR_t\text{-}1) + u \cdot e^{-\eta \cdot n} / (1-e^{-\eta})$, and then taking the limit for n in both sides, we clearly have the required result, i.e. $J^*(^{Q}FR_t\text{-}1) = \lim_{n \to \infty} J_n(^{Q}FR_t\text{-}1)$.

Finally, the uniqueness of the limit in P3 of the increasing sequence $\{J_n(^{Q}FR_t\text{-}1);\ n = 0, 1, 2, \ldots\}$ comes from the absolute value properties, especially the inequality property: that is, assuming that $J_n(^{Q}FR_t\text{-}1) \to J1(^{Q}FR_t\text{-}1)$ and $J_n(^{Q}FR_t\text{-}1) \to J2(^{Q}FR_t\text{-}1)$, we will prove $J_\infty(^{Q}FR_t\text{-}1) = J1(^{Q}FR_t\text{-}1) = J2(^{Q}FR_t\text{-}1)$. By the inequality property for absolute values, we have $0 \leq |J1(^{Q}FR_t\text{-}1) - J2(^{Q}FR_t\text{-}1)| \leq |J1(^{Q}FR_t\text{-}1) - J_n(^{Q}FR_t\text{-}1)| + |J_n(^{Q}FR_t\text{-}1) - J2(^{Q}FR_t\text{-}1)|$. Then taking the limit for n of both sides, we have the required result: $J_\infty(^{Q}FR_t\text{-}1) = J1(^{Q}FR_t\text{-}1) = J2(^{Q}FR_t\text{-}1)$, since as $n \to \infty$, $|J1(^{Q}FR_t\text{-}1) - J_n(^{Q}FR_t\text{-}1)| \to 0$ and $|J_n(^{Q}FR_t\text{-}1) - J2(^{Q}FR_t\text{-}1)| \to 0$. Q.E.D.

Using the results proven in Proposition 5.1 that for any non-zero $^{Q}FR_t\text{-}1$ at time t, the sequence $\{J_0(^{Q}FR_t\text{-}1), J_1(^{Q}FR_t\text{-}1), J_2(^{Q}FR_t\text{-}1), \ldots\}$ is positive, increasing but convergent monotonically to $J^*(^{Q}FR_t\text{-}1)$, it follows that each sequences of the functional coefficients, i.e. $\{Q_{11}, Q_{21}, Q_{31}, \ldots\}$, $\{Q_{12}, Q_{22}, Q_{32}, \ldots\}$ and $\{Q_{13}, Q_{23}, Q_{33}, \ldots\}$, converge monotonically to some real number for any non-zero $^{Q}FR_0\text{-}1$, say $Q_{n1} \to Q_1$, $Q_{n2} \to Q_2$ and $Q_{n3} \to Q_3$, as $n \to \infty$. The optimal functional equation for $^{Q}J^*(.)$ is then given as follows (- this equation shows clearly that this functional equation does not depend explicitly on time t, i.e. time-invariant optimal functional equation): for any non-zero $^{Q}FR_t\text{-}1$ at time t,

$$\lim_{n \to \infty} {}^{Q}J_n(^{Q}FR_t\text{-}1) = \lim_{n \to \infty} \{ \operatorname*{Min}_{k(^{Q}H_t)} \{ (\theta + (1\text{-}\theta)\cdot k(^{Q}H_t)^2) \cdot (^{Q}FR_t\text{-}1)^2 + e^{-\eta} \cdot {}^{Q}J_{n\text{-}1}(q_1 \cdot (1\text{-}k(^{Q}H_t)) \cdot (^{Q}FR_t\text{-}1) + q_2) \} \}$$

$$\Leftrightarrow {}^{Q}J^*(^{Q}FR_t\text{-}1) = \operatorname*{Min}_{k(^{Q}H_t)} \{ (\theta + (1\text{-}\theta)\cdot k(^{Q}H_t)^2) \cdot (^{Q}FR_t\text{-}1)^2 + e^{-\eta} \cdot {}^{Q}J^*(q_1 \cdot (1\text{-}k(^{Q}H_t)) \cdot (^{Q}FR_t\text{-}1) + q_2) \},$$

which leads to a time-invariant quadratic function $J^*(.)$ defined by the equation

$^{Q}J^{*}(^{Q}FR_t\text{-}1) = Q_1 \cdot (^{Q}FR_t\text{-}1)^2 + Q_2 \cdot (^{Q}FR_t\text{-}1) + Q_3$ --- (5.28)

where Q_1, Q_2 and Q_3 each are obtained by taking the limit for n in both sides of their respective recursive equations (5.24), (5.25) and (5.26), that is, Q_1, Q_2 and Q_3 each are the steady-state value (or limiting value) satisfying the three simultaneous equations

$Q_1 = [\theta\cdot(1\text{-}\theta) + e^{-\eta}\cdot q_1^{2}\cdot Q_1] / [(1\text{-}\theta) + e^{-\eta}\cdot q_1^{2}\cdot Q_1]$; --- (5.29)

$Q_2 = [2(1\text{-}\theta)\cdot e^{-\eta}\cdot q_1\cdot(q_2\cdot Q_1 + Q_2/2)] / [(1\text{-}\theta) + e^{-\eta}\cdot q_1^{2}\cdot Q_1]$; and --- (5.30)

$Q_3 = [(1\text{-}\theta)\cdot e^{-\eta}\cdot(q_2^{2}\cdot Q_1 + q_2\cdot Q_2 + Q_3) + e^{-2\eta}\cdot q_1^{2}\cdot(Q_1\cdot Q_3 - Q_2^{2}/4)] / [(1\text{-}\theta) + e^{-\eta}\cdot q_1^{2}\cdot Q_1]$. --- (5.31)

The above result implies that $k_n(^{Q}FR_t\text{-}1)$ in (5.27) converges to a finite limiting value $k^{*}(^{Q}FR_t\text{-}1)$ as $n \to \infty$, that is,

$\lim_{n\to\infty} k_{n\text{-}1}(^{Q}FR_t\text{-}1) = k^{*}(^{Q}FR_t\text{-}1) = k^{*}(^{Q}H_t)$

$= [(e^{-\eta}\cdot q_1^{2}\cdot Q_1)\cdot(^{Q}FR_t\text{-}1) + e^{-\eta}\cdot q_1\cdot(q_2\cdot Q_1 + Q_2/2)] / [((1\text{-}\theta) + e^{-\eta}\cdot q_1^{2}\cdot Q_1)\cdot(^{Q}FR_t\text{-}1)]$, --- (5.32)

expressed as a time-invariant function of the current non-zero state variable $^{Q}FR_t$-1.

Here, the above function $k^{*}(.)$ presents the optimal stationary spread policy that specifies the spreading mechanism of the unfunded ratio (i.e. $1 - {}^{Q}FR_t \in R^{1}\text{-}\{0\}$) informed through the actuarial valuation process, whereas if the informed value of $^{Q}FR_t$ is 100%, we apply an possible spread rule (i.e. $k^{*}(^{Q}FR_t\text{-}1) = 0$, as described in subsection (ii) of section 5.4.3.2).

The uniqueness of our optimal spread policy $k^{*}(.)$ is determined by that of the limiting value Q_1 because $k^{*}(.)$ involves Q_1 and Q_2 but Q_2 is a function of Q_1. So, we concentrate on the property of the sequence $\{Q_{11}, Q_{21}, Q_{31}, ...\}$.

Proposition 5.2 (Property of $\{Q_{11}, Q_{21}, Q_{31}, ...\}$)

The sequence $\{Q_{11}, Q_{21}, Q_{31}, ...\}$ generated by the recursive equation (5.24) is positive and converges monotonically to the limiting value Q_1, which is the unique positive solution of (5.29) given in the form

$$Q_1 = \{ (e^{-\eta}\cdot q_1^2-(1-\theta)) + \sqrt{[(e^{-\eta}\cdot q_1^2-(1-\theta))^2 + 4\cdot\theta\cdot(1-\theta)\cdot e^{-\eta}\cdot q_1^2]} \} / \{2\cdot e^{-\eta}\cdot q_1^2\}.$$

Proof Firstly, the sequence $\{Q_{11}, Q_{21}, Q_{31}, ...\}$ is positive and strictly increasing, which can be proven by proceeding by deduction such that since $Q_{11} = \theta \in (0, 1)$ and $q_1 > 0$, we have

$$Q_{21} - Q_{11} = [(1-\theta)\cdot e^{-\eta}\cdot q_1^2\cdot Q_{11}] / [(1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{11}] > 0,$$

$$Q_{31} - Q_{21} = [(1-\theta)^2\cdot e^{-\eta}\cdot q_1^2\cdot(Q_{21} - Q_{11})] / [((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{21})\cdot((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{11})] > 0,$$

. . . ,

$$Q_{n1} - Q_{n-11} = [(1-\theta)^2\cdot e^{-\eta}\cdot q_1^2\cdot(Q_{n-11} - Q_{n-21})] / [((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{n-11})\cdot((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{n-21})] > 0,$$

. . . > 0.

Secondly, the sequence of consecutive slopes $\{(Q_{31}-Q_{21})/(Q_{21}-Q_{11}), (Q_{41}-Q_{31})/(Q_{31}-Q_{21}), (Q_{51}-Q_{41})/(Q_{41}-Q_{31}), ...\}$ is positive and strictly decreasing, which can be easily proven as follows: using the above results, then we have the following inequality:

$$(Q_{31} - Q_{21}) / (Q_{21} - Q_{11}) = [(1-\theta)^2\cdot e^{-\eta}\cdot q_1^2] / [((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{21})\cdot((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{11})]$$

$$> (Q_{41} - Q_{31}) / (Q_{31} - Q_{21}) = [(1-\theta)^2\cdot e^{-\eta}\cdot q_1^2] / [((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{31})\cdot((1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_{21})],$$

. . . ,

$$> (Q_{n1} - Q_{n-11}) / (Q_{n-11} - Q_{n-21}) = [(1-\theta)^2\cdot e^{-\eta}\cdot q_1^2] / [((1-\theta)+e^{-\eta}\cdot q_1^2\cdot Q_{n-11})\cdot((1-\theta)+e^{-\eta}\cdot q_1^2\cdot Q_{n-21})],$$

. . . > 0

;hence, we have the required result.

Therefore, from the first and second results, the sequence $\{Q_{11}, Q_{21}, Q_{31}, ...\}$ is positive, strictly increasing and convergent monotonically to the limiting value Q_1, i.e. $Q_{n1} \to Q_1$ as $n \to \infty$

(as like the well-known cobweb moving toward an equilibrium point, see Sandefur (1990; section 1.5).

Thirdly, the property of Q_1 being unique can be proven as follows:

assuming that $Q_{n1} \to Q1$ and $Q_{n1} \to Q2$, then by applying the absolute value properties we have $0 \leq |Q1\text{-}Q2| \leq |Q1\text{-}Q_{n1}| + |Q_{n1}\text{-}Q2|$ and then taking limit in both sides provides $0 \leq |Q1 - Q2| \leq \lim_{n\to\infty} \{ |Q1 - Q_{n1}| + |Q_{n1} - Q2| \} = 0$, which implies that $Q_1 = Q1 = Q2$.

Finally, now by taking the limit for n in the both sides of equation (5.24) it follows that the positive limiting value Q_1 satisfies equation (5.29). By multiplying $[(1\text{-}\theta) + e^{-\eta}\cdot q_1^{2}\cdot Q_{11}]$ on the both sides of (5.29), we obtain a positive root of the quadratic equation. Q.E.D.

Remark 5.2: In the light of control theory, equation (5.24) is particularly called the discrete-time Riccati equation (in the scalar case), i.e. $Q_{n\,1} = f(Q_{n-1\,1})$, an equation of the first-order and first-degree where $f(Q_{n-1\,1})$ is rational in $Q_{n-1,\,1}$. Accordingly, equation (5.29) is called the equilibrium Riccati equation (in the scalar case), i.e. $Q_1 = f(Q_1)$ [see, Whittle (1982; sections 5.3 and 5.5)].

(iii) Conclusion:

The control optimisation procedure with respect to $k(^QH_t)$ is specified by the time-invariant equations (5.28)~(5.32) together with the possible spread rule established in subsection (ii) of section 5.4.3.2. The optimal quasi-stationary spread funding formula is then given by

$$^QCR^{f*}_t = {}^QNR^f_t - k^*(^QH_t)\cdot(^QFR_t\text{-}1) = {}^Q\mu^*(^QFR_t\text{-}1;\,k^*(^QH_t)),\ \text{for each } t\in[0,\infty) \qquad \text{---}\ (5.33)$$

where $^Q\mu^*(.;\,k^*(^QH_t))$ denotes the optimal quasi-stationary spread funding plan, which is our dynamic pension funding plan (i.e. optimal control law). The value of the control action $^QCR^{f*}_t$ is produced from the information fed back about the current state (i.e. the actual value of

${}^{Q}FR_t$-1) and the information calculated on the current actuarial assumptions (i.e. the computed value of NR^f_t).

Alternatively, substituting $k^*({}^{Q}H_t)$ defined in (5.32) into equation (5.33) yields

$${}^{Q}CR^*_t = {}^{Q}NR_t - {}^{Q}\varphi(\theta) \cdot ({}^{Q}FR_t\text{-}1) + {}^{Q}\xi(\theta), \text{ for any non-zero } {}^{Q}FR_t\text{-}1,\ t\in[0, \infty) \qquad \text{- - - (5.34)}$$

where

${}^{Q}\varphi(\theta) = [e^{-\eta}\cdot q_1^2\cdot Q_1] / [(1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_1]$, depending on θ and $0 < {}^{Q}\varphi(\theta) < 1$; and

${}^{Q}\xi(\theta) = - [e^{-\eta}\cdot q_1\cdot(q_2\cdot Q_1 + Q_2/2)] / [(1-\theta) + e^{-\eta}\cdot q_1^2\cdot Q_1]$, depending on θ.

; hence, this expression takes mathematically a similar form to the stationary spread funding formula (5.6), except for the term ${}^{Q}\xi(\theta)$ and we notice that ${}^{Q}CR^{f*}_t = {}^{Q}NR^f_t$ for zero-valued ${}^{Q}FR_t$-1, $t\in[0, \infty)$ in accordance with our possible spread rule.

Referring back to the funding formula (4.18) derived in section 4.3.3.2, we can easily check that there is a similarity between formulae (5.34) and (4.18) for non-100% funding level (although formula (5.34) is derived from the form of CR^f_t constrained by the Spread method in the situation of a long-term, going-concern valuation, i.e. from formula (5.17), while formula (4.18) is derived from the unconstrained form of CR_t for the situation of a short-term, winding-up valuation). In short, ${}^{Q}\varphi(\theta)$ in (5.34) can be thought of as the proportional state-feedback controlling parameter as in $\varphi(t; \theta)$ in (4.18) and ${}^{Q}\xi(\theta)$ in (5.34) as the additive controlling parameter as in $\xi(t; \theta)$ in (4.18).

However, there is a fundamental difference in dealing with 100% funding level: that is, for 100% funding level, the unconstrained CR^f_t is to be optimally designed in the form of NR^f_t + ${}^{Q}\xi(\theta)$ but the constrained CR^f_t is best (not necessary optimally) designed in the form of NR^f_t according to our possible spread rule.

Moreover, the optimal quasi-stationary control response ${}^{Q}FR^{*}_{t+1}$ corresponding to ${}^{Q}CR^{*}_{t}$ is recursively generated with time t by the following optimal quasi-autonomous system equation: for each $t \in [0, \infty)$,

$${}^{Q}FR^{*}_{t+1}-1 = q_1 \cdot (1-k^{*}({}^{Q}H_t)) \cdot ({}^{Q}FR^{*}_{t}-1) + q_2 \text{ with given } {}^{Q}FR_0-1 = {}^{Q}FR^{*}_{0}-1. \qquad \text{--- (5.35)}$$

Substituting $k^{*}({}^{Q}H_t)$ into equation (5.34), we have an alternative expression of equation (5.35) such that for each $t \in [0, \infty)$ and given ${}^{Q}FR_0-1 = {}^{Q}FR^{*}_{0}-1$,

$${}^{Q}FR^{*}_{t+1} - 1 = [q_1 \cdot (1 - {}^{Q}\varphi(\theta))] \cdot ({}^{Q}FR^{*}_{t} - 1) + [q_1 \cdot {}^{Q}\xi(\theta) + q_2] \qquad \text{--- (5.36)}$$

; hence, the optimal quasi-autonomous system equation (5.35) turns out to be autonomous (i.e. stationary and zero-input), so if the value of θ is controllable in a direction to satisfying $|q_1 \cdot (1- {}^{Q}\varphi(\theta))| < 1$ (i.e. geometric stability condition of the optimal system equation (5.36)), then the optimal control error sequence, $\{{}^{Q}FR^{*}_{t}-1;\ t \in [0, \infty)\}$, will be geometrically stable about the steady-state error $[q_1 \cdot {}^{Q}\xi(\theta) + q_2] / [1 - q_1 \cdot (1 - {}^{Q}\varphi(\theta))]$. In other words, different from the algebraic approach, the DP approach provides that the control error sequence is controllable by adjusting the values of θ without imposing directly any stability constraints on the controlling parameter.

In conclusion, our optimal spread control sequence, $\{k^{*}({}^{Q}H_0), k^{*}({}^{Q}H_1), k^{*}({}^{Q}H_2), \ldots\}$, will be uniquely generated with time t by the optimal stationary spread policy $k^{*}(.)$ defined in (5.32) together with the possible spread rule.

<u>Remark 5.3</u> (geometric stability condition $|q_1 \cdot (1- {}^{Q}\varphi(\theta))| < 1$): from the fact that $q_1 > 0$ and $0 < {}^{Q}\varphi(\theta) < 1$, the geometric stability condition reduces to $0 < q_1 \cdot (1- {}^{Q}\varphi(\theta)) < 1$ and hence, there exists a constant convergence rate ${}^{Q}\zeta(\theta) \in (0, 1]$ such that $q_1 \cdot (1- {}^{Q}\varphi(\theta)) = (1 - {}^{Q}\zeta(\theta))$ for any

$q_1 \cdot (1 - {}^Q\varphi(\theta)) \in [0, 1)$, $0 < \theta < 1$ (which is the so-called geometric damping rate, see Comment 4 in Appendix 5A). In this respect, $q_1 \cdot (1 - {}^Q\varphi(\theta))$ can be interpreted in terms of the geometric damping rate $\zeta(\theta)$. We can easily check that the convergence rate ${}^Q\zeta(\theta)$ is strictly increasing in $\theta \in (0, 1)$: in other words, as putting more emphasis on reducing the solvency risk (i.e. $\theta \to 1$), $\{{}^QFR^*_t - 1;\ t \in [0, \infty)\}$ converges more quickly (and monotonically) to the steady-state error $[q_1 \cdot {}^Q\xi(\theta) + q_2] / [1 - q_1 \cdot (1 - {}^Q\varphi(\theta))]$.

5.4.5.2 Essential requirements for the FDP approach

The essential requirements for applying the FDP approach to the quasi-stationary LQP optimisation problem (5.21) are summarised as follows (- these will provide a useful comparison with those for the algebraic approach described in section 5.3.4.2):

Requirement B1: The unknown spread parameter should be adapted as a time-invariant function of currently available information vector ${}^QH_t$, i.e. $k_t = k({}^QH_t)$;

Requirement B2: The uniform boundedness condition imposed on the control cost per unit control period (i.e. $0 \le [\theta + (1-\theta) \cdot k({}^QH_t)^2] \cdot [{}^QFR_t - 1]^2 \le u$, for all $t \in [0, \infty)$) and the positive discount factor (i.e. $\eta > 0$) (for uniform convergence of $\{J_n({}^QFR_t - 1);\ n = 0, 1, 2, \dots\}$ in connection with Requirement B3 described below, see Proposition 5.1 in section 5.4.5.1); and

Requirement B3: The mathematical structure of the recursive relationship for FDP calculations is to be stationary (for the extension of the N-multistage optimisation procedure to an infinite stage optimisation procedure in connection with Requirement B2, see Proposition 5.1 in section 5.4.5.1).

We can then check the conceptual difference in mathematical methods between the algebraic approach and the FDP approach by comparing Requirements A1, A2 and A3 with

Requirements B1, B2 and B3 respectively. It is worth noticing that the fundamental differences relate to our viewpoint of the unknown spread parameter k_t: in short, $k_t = k$, constant for all t in the algebraic approach, while $k_t = k({}^QH_t)$, a time-invariant function of H_t in the FDP approach.

The rational motivation for using the FDP approach is to expect that adjusting sequentially the decision for k_t according to the information updated with time t will reduce the solvency risk as well as the contribution rate risk over an infinite control horizon. Consequently, k_t is viewed as an unknown but variable parameter, denoted by $k({}^QH_t)$. In this respect, we may say that this viewpoint is similar to the Bayesian Perspective in the field of statistical decision theory in the light of using the actual observed data as prior information about the unknown quantity k_t [see Berger (1985; section 1.6)]. This viewpoint is distinct from the viewpoint of using the algebraic approach on the spread parameter k_t which is similar to the Frequentist Perspective [see section 5.3.4.2].

5.4.5.3 Summary

Our control mechanism for the quasi-autonomous controlled object, specified by equations (5.32)~(5.36), can be visualised by the following block diagram (which will also provide a useful comparison with Figures 5.1 given in section 5.3.4.3).

[Quasi-autonomous controlled object] ⟶ $\{{}^QFR_t\text{-}1\}$ ⟶

$\{k({}^QH_t) = k^*({}^QFR_t\text{-}1)\}$↑ ↓

$\{{}^QNR_t\}$ ⟶ [Optimal control law ${}^Q\mu^*(.;\, k({}^QH_t)=k^*({}^QFR_t\text{-}1)$] ⟵ {Feedback: ${}^QFR_t\text{-}1$}

Figure 5.2 Optimal quasi-stationary spread funding control system: $k^*({}^QFR_t\text{-}1)$ denotes the optimal value of $k({}^QH_t)$ defined in (5.32).

The above Figure 5.2 illustrates that both the quasi-autonomous controlled object and quasi-stationary spread funding plan are optimally designed by way of optimising sequentially the value of the unknown but variable spread parameter $k(H_t)$ in the course of time t, i.e. by setting $k(^QH_t) = k^*(^QFR_t\text{-}1)$ for each t.

5.4.5.4 Numerical illustrations

The objective of this section is to illustrate numerically the relationship between the optimal value of $k(^QH_t)$ (i.e. $k^*(^QH_t)$) and its resulting control action error $^QCR^{f*}_t\text{-}^QNR^f_t$ and funding level $^QFR^*_{t+1}$ with respect to the force of investment interest (i.e. δ) and the relative weighting parameter (i.e. θ). All the numerical illustrations are given in graphical form in Appendix 5B.4.

(i) <u>Comments on a suitable value of θ:</u>

Controlling only the contribution rate risk with respect to $k(^QH_t)$ would not make any sense to our control problem (5.22) in view of sequential optimisation because minimising $\sum e^{-\eta t}\cdot(^QCR_t - {}^QNR_t)$ leads clearly to $k^*(^QH_t) = 0$ for all t, irrespective of the information fed back about the actual value of current dynamic state $^QFR_t\text{-}1$ (and hence, the corresponding performance index is ideally zero). However, we can say from this ideal result that optimising the value of the unknown spread parameter $k(^QH_t)$ with time t is much more orientated toward controlling the contribution rate risk rather than controlling the solvency risk. This argument is consistent with the view of Dufresne (1988) [see Remark 5.1]. Following this ideal result, we propose some suggestions about how to set up the value of θ involved in our control problem (5.22) without constructing a proper supplementary performance criterion. That is, one possible policy for determining the value of θ would be setting the value of θ close to zero. For this reason, we shall content ourselves with illustrating the influence of θ, subject to $\{\theta: 10\%, 1\%\}$, on the optimal sequences $\{k^*(^QH_t); t\in[0, \infty)\}$, $\{^QCR^{f*}_t\text{-}^QNR^f_t; t\in[0, \infty)\}$ and $\{^QFR^*_{t+1}; t\in[0, \infty)\}$.

(ii) <u>Assumptions</u>:

(A1) All actuarial assumptions are the same as given in section 5.3.4.4 (in short, $\exp(\alpha)=1.03$, $\exp(\beta)=1.02$ and $\exp(\eta)=1.06$); and

(A2) Projection assumptions:

- infinite control horizon: $t \in [0, \infty)$;
- admissible value set of θ: {10%, 1%};
- given initial value ${}^{Q}FR_0 = 50\%$; and
- force of investment interest (δ): $\exp(\delta) = 1.06$ or 1.08.

Hence, the formulae (a)~(d) and Remark 4.2 introduced in section 4.3.3.1 hold simply by replacing NR_t ($=NC_t/SL_t$) and EBR_t ($=EB_t/SL_t$) with their respective ${}^{Q}NR^{f}_{t}$ ($=NC_t/AL_t$) and ${}^{Q}EBR^{f}_{t}$ ($=EB_t/AL_t$). Further, the geometric stability condition of the optimal system equation (5.36), i.e. $|q_1 \cdot (1- {}^{Q}\varphi(\theta))| < 1$, is satisfied and hence the funding level is convergent to an equilibrium state $1 + [q_1 \cdot {}^{Q}\xi(\theta)+q_2] / [1-q_1 \cdot (1-{}^{Q}\varphi(\theta))]$.

In particular, we note that if $\eta = \delta$ (representing the case that all actuarial assumptions are exactly realised), then $q_2 = 0$ (and hence $Q_2 = 0$ from equation (5.30) and ${}^{Q}\xi(\theta) = 0$ from equation (5.34)). Thus, the equilibrium state is equivalent to the 100% funding target and further the optimal value of the spread parameter is constant for all t from equation (5.32), that is, $k^*({}^{Q}H_t) = {}^{Q}\varphi(\theta)$, $0<{}^{Q}\varphi(\theta)<1$, constant for all t.

(iii) <u>Illustrative numerical results</u>:

The projections of optimal quasi-stationary spread funding plan governed by formula (5.33) are visualised in Graph 5.1 (composed of Graphs 5.1.1~5.1.3, produced on the assumption that $\eta = \delta$) and Graph 5.2 (composed of Graphs 5.2.1~5.2.3, produced on the assumption that $\eta < \delta$) of Appendix 5B.4.

Exploring and comparing Graphs 5.1 and 5.2, we can derive the important observations below:

(a) Graphs 5.1 illustrate that the most likely expected investment performance (i.e. $\eta = \delta$) provides that the optimal sequence$\{k^*(^QH_t);\ t\in[0, \infty)\}$ is constant, while the corresponding sequences $\{^QCR^{f*}_t - {}^QNR^f_t;\ t\in[0, \infty)\}$ and $\{^QFR^*_{t+1};\ t\in[0, \infty)\}$ are monotonically convergent to their respective target values 0 and 1 , respectively. Comparing $k^*(^QH_t)$ with k^* (obtained from the stationary control problem using the algebraic approach) subject to $\theta = 10\%$, $\eta = \delta$ and $^SFR={}^QFR=50\%$, then

- these are commonly constant for all t but different from one another, i.e. $k^*(^QH_t) = 0.265734 < k^* = 0.905243$, constant for all t (note that the large difference between these two values can be attributed to the different philosophies employed in finding their respective optimal solutions, as mentioned in sections 5.3.4.2 and 5.4.5.2);

- the optimal sequences $\{^QFR^*_t\}$ and $\{^SFR^*_t\}$ have the same equilibrium state (i.e. 100% funding target) but their respective convergence speeds are different, i.e. [convergence rate of $\{^QFR^*_t\}$ to 100% funding target] = $1 - q_1(1-{}^Q\varphi(\theta)) = 0.259165$ < [convergence rate of $\{^SFR^*_t\}$ to 100% funding target] = $1 - q_1(1-k^*) = 0.904395$; and

- consequently, the algebraic approach provides a quicker convergence to a 100% funding target than the DP approach and accordingly, the contribution rates are more quickly stabilised than when using the DP approach, but we note that the funding burden on the sponsoring employer is even larger than when using the DP approach due to $k^*(^QH_t) << k^*$.

(b) As seen in Graph 5.2.1, the optimal quasi-stationary spread control mechanism under the better investment performance (i.e. $\eta < \delta$) can be broken into four consecutive stages in relation to the control responses of $^QFR^*_t$ (note that the added examples are based on the case of k1 (i.e. θ=10%) in Graph 5.2.1): broadly,

- introduction stage: providing a fast improvement in the initially given unfunded ${}^{Q}FR_0$ in the direction of the 100% funding target but there is no abrupt change in the values of the spread parameter (e.g. $\{k^*({}^{Q}H_t)$: $0 \leq t \leq 6\}$);

- turmoil stage: providing a break in the fast growth of the funding level in the introduction stage (and hence a slowing down in its rate of increase), which causes an abrupt change in the values of the spread parameter, particularly when crossing the 100% funding target (e.g. $\{k^*({}^{Q}H_t)$: $7 \leq t \leq 17\}$);

-stabilising stage: enabling the funding level to approach gradually and steadily to the equilibrium state for the funding level, i.e. $1 + [q_1 \cdot {}^{Q}\xi(\theta)+q_2] / [1-q_1 \cdot (1-{}^{Q}\varphi(\theta))]$ after the turmoil stage (e.g. $\{k^*({}^{Q}H_t)$: $18 \leq t \leq 45\}$); and finally,

- maturity stage: enabling the funding level to be attained at the level of its equilibrium state as a result of the stabilising stage (e.g. $\{k^*({}^{Q}H_t)$: $t \geq 46\}$)

; hence, these four stages are quite distinguishable from the optimal stationary spread control mechanism illustrated in 5.3.4.4, except for the introduction stage: in other words, the turmoil, stabilising and maturity stages shows clearly the difference between the algebraic and dynamic approaches to optimising the value of the spread parameter;

(c) In particular, although we do not show fully the influence of θ, the turmoil stage is very sensitive to the change in values of θ (which is thought to be caused by the weakness in the mathematical formulation of the Spread method in the case of a 100% funding level) and we note also that this stage is essential for the future stability of both funding levels and contribution rates within the framework of the Spread method, as illustrated in Graphs 5.2.1~5.2.3;

(d) One of the most interesting results is that the convergence of the control action errors is the fastest, next the funding levels and lastly the spread controls, which would be consistent with the comments made in the above subsection (i);

(e) The results for the better investment performance (i.e. $\eta < \delta$) provides a faster convergence, better security and less financial burden, as expected than the most likely expected performance (i.e. $\eta = \delta$); and finally, we would like to end our numerical analysis with the following suggestion (f).

(f) Even though this suggestion is likely to be unacceptable in view of the classical actuarial approach, we may suggest from the above results (b)~(d) that if the value of the funding level produced from the valuation process is approximately 100% but less than 100%, the value of the spread parameter is allowed to be negative for the future stability of both funding levels and contribution rates; but on the other hand, if the value of the funding level produced from the valuation process is approximately 100% but larger than 100%, the value of the spread parameter is allowed to be larger than one for the future stability of both funding levels and contribution rates: in other words, it may be necessary to adopt more flexibility and allow the amortisation period not to be constrained within the interval $0 \leq k_t \leq 1$.

5.5 Dynamic pension funding plan for non-stationary LQP optimisation problem

We note first that for notational distinction from the previous symbols used in sections 5.3 and 5.4, the superscript 'N' on the left side of each main symbol (to be introduced in this section) is used to indicate that it concerns the non-stationary LQP optimisation problem (to be formulated later in section 5.5.3).

5.5.1 Introduction

In sections 5.3 and 5.4, we have studied the stationary and quasi-stationary control problems respectively under a common assumption that the investment market-related parameter, i.e. δ_{t+1} (or i_{t+1}) involved in the system equation (5.2), is constant for all t. From a practical point of view, the constancy assumption for δ_{t+1} fails to address the physical reality associated with the time-varying economic situation. In this respect, we allow some variations in δ_{t+1} with time t (-the variation will be mathematically modelled in the next section 5.5.2); hence, the motions of the controlled object governed by the system equation (5.2) is variant for a translation of time, so the controlled object with this property shall be is called the non-autonomous (i.e. non-stationary and zero-input) controlled object. This section is aimed to illustrate how the spread parameter k_t is determined to produce the optimal performance of the non-autonomous controlled object and thus k_t has to be considered as a time-varying controlling parameter.

In general, modelling the time-varying δ_{t+1} would be based on the forecasting of its respective mathematical trend curve through the analysis of a related historical data series. Further, their respective projections would be somewhat dependent on the perspective of the pension experts (especially, the actuary and investment manager) regarding the future investment market movements, showing optimistic, expected and pessimistic scenarios.

In section 5.5.3, we construct our non-stationary LQP optimisation problem in a similar manner to the quasi-stationary LQP optimisation problem in section 5.4.3. We try to search for its optimisation procedure in section 5.5.4 but it turns out to be insoluble. In section 5.5.5, we propose an approximate optimisation procedure for this non-stationary control problem as a best (nor necessary optimal) approach for its solution and then give some illustrative numerical examples. Throughout these sections, we note that the arguments used are an general extension of those related to the quasi-stationary LQP control problem investigated in section 5.4.

5.5.2 Non-stationary assumptions and Model construction

We make here the non-stationary assumptions such that δ_{t+1} and k_t each in the system equation (5.2) depend on time t. Their respective mathematical specifications are given separately below.

(i) Non-stationary controlling parameter:

In order to respond successfully to the time-varying situations of investment markets, the controlling parameter applying over the unit control period [t, t+1), k_t, will depends on the currently available information as well as on the non-stationality characteristics of our controlled object. So, k_t can be expressed as a time-varying function of the currently available information vector, say ${}^{N}H_t$, where ${}^{N}H_t = ({}^{N}FR_0, {}^{N}FR_1, \ldots, {}^{N}FR_t, {}^{N}CR_0, {}^{N}CR_1, \ldots, {}^{N}CR_{t-1})$:

$$k_t = k_t({}^{N}H_t) \quad \text{for all } t \in [0, \infty), \qquad \text{- - - (5.37)}$$

which can be regarded as the non-stationary version of the quasi-stationary spread parameter $k({}^{Q}H_t)$ introduced in section 5.4.1.

(ii) Damped harmonic motion of investment rates of return:

As investigated and fitted by Loades (1992), the historical trend curve of investment rates of return could be characterised by a harmonic (or periodic) curve related to a series of business/economic cycles. So, the future motions of investment rates of return are here modelled as follows: for all $t \in [0, \infty)$,

$$\exp(\delta_{t+1}) = (1+i_{t+1}) = 1 + i_e + \text{sinc}(t+1), \quad i_e > 0 \qquad \text{--- (5.38)}$$

where

i_e = expected long-term rate of return on future investment projected from historical linear trend (from the viewpoint of classical actuarial valuations, i_e would be often used as the valuation interest rate i_v); and

sinc(t+1) = sinc-function to apply over a unit control period [t, t+1), defined by the equation $\text{sinc}(t+1) = \sin[\omega\cdot(t+1) + \varnothing] / [\omega\cdot(t+1) + \varnothing]$, in which ω = the angular frequency = $2\pi/T_b$, $T_b > 0$ denoting the period of the business cycle, $\varnothing$ = the initial phase shift of business cycle and $[\omega\cdot(t+1) + \varnothing]^{-1}$ = damped amplitude of oscillations in i_{t+1} [for more details of sinc-function, see McGillem & Cooper (1991, section 3.8)]

; hence, this mathematical model presents the damped harmonic variations in i_{t+1} around the projected linear trend $1+i_e$ away from the initial time t=0.

It should be noted that the projected linear trend $1+i_e$ coupled with its deviations sinc(t+1) can be justified on the optimistic assumptions that firstly, the fixed mean i_e is a correct assessment of the average future investment performance and next, the pension professionals will continuously reduce the predictive errors around $1+i_e$, in particular by means of periodically conducting fully-detailed post-mortems of their investment performance once all data are collected and then securely selecting a broad asset mix to match a series of business cycles. In addition, comparing Young and Mature pension schemes (as introduced earlier in section 2.1.4.2), the latter would have more capacity to cope with the investment risk associated with a business cycle than the former. In our model, this difference is modelled by the parameter $\varnothing$ (i.e. the initial phase shift): in other words, the Mature pension scheme would generally have a relatively larger $\varnothing$ than the Young pension scheme.

Therefore, adapting the above mathematical models (5.37) and (5.38) to the spread funding formula (5.1) and the system equation (5.2), we derive their corresponding equations (5. 39) and (5.40): that is,

$^{N}CR_t = {}^{N}NR_t - k_t({}^{N}H_t) \cdot ({}^{N}FR_t\text{-}1) \equiv {}^{N}\mu({}^{Q}FR_t\text{-}1;\, k_t({}^{Q}H_t))$, in which

$$k_t({}^{N}H_t) \in \{k_t({}^{N}H_t) \colon 0 \le k_t({}^{N}H_t) \le 1\} \text{ for all } t \in [0, \infty) \qquad \text{---} (5.39)$$

; in particular, $k_t({}^{N}H_t)$ shall be called the non-stationary spread parameter, this formula (5.39) the non-stationary spread funding formula and ${}^{N}\mu(.;\, k_t({}^{N}H_t))$ the non-stationary spread funding plan (we note that the term 'non-stationary' is added to emphasise the fact that $k_t({}^{N}H_t)$ depends explicitly on time t and ${}^{N}NR_t = {}^{Q}NR_t = {}^{S}NR_t$); and hence,

$$FR_{t+1}\text{-}1 = {}^{N}q_1(t) \cdot (1\text{-}k_t({}^{N}H_t)) \cdot ({}^{N}FR_t\text{-}1) + {}^{N}q_2(t) \quad \text{with given } {}^{N}FR_0 - 1 \qquad \text{---} (5.40)$$

where ${}^{N}q_1(t) = (1+i_e+\text{sinc}(t+1)) / \exp(\alpha+\beta)$; and ${}^{N}q_2(t) = [(1+i_e+\text{sinc}(t+1)) / \exp(\eta)] - 1$

; here, this equation shall be called the non-autonomous (i.e. non-stationary and zero-input) system equation, which governs our controlled object.

For the same reason as described at the end of section 5.4.1, the algebraic approach is here inapplicable (i.e. $k_t(H_t)$ violates Requirement A1 described in section 5.3.4.2); hence, we employ the FDP approach as a best alternative mathematical method for our control optimisation problem formulated in the next section, as in the quasi-stationary LQP optimisation problem.

5.5.3 Non-stationary LQP optimisation problem

Prior to formulating our control problem, it is worth referring back to a clear result of the quasi-stationary LQP optimisation problem (5.22) illustrated in section 5.4.4.1 that the FDP approach can not provide any solutions to the control problem (5.22) with the spread parameter being constrained by the boundary condition, i.e. $0 \leq k({}^{Q}H_t) \leq 1$ for all $t \in [0, \infty)$. Similarly, we can demonstrate that this insolubility problem will also apply in the non-stationary infinite-horizon control problem with this same boundary constraint. For this reason, we are here concerned only with the constraints for ensuring the finiteness of ${}^{N}IPI_{\theta}$ (i.e. non-stationary version of IPI_{θ}, see section 5.2.1.1) defined by

$${}^{N}IPI_{\theta} = \sum_{t=0}^{\infty} \{ e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k_t({}^{N}H_t)^2] \cdot [{}^{N}FR_t - 1]^2 \}.$$

Let u be some positive real number, then ${}^{N}IPI_{\theta}$ will be well-defined as a function value $< \infty$ for all t on the following non-stationary version of the quasi-stationary controlling parameter space II given in (5.20): for all $t \in [0, \infty)$,

$$\{k_t({}^{N}H): \ \eta > 0 \ \text{ and } \ 0 \leq [\theta + (1-\theta) \cdot k_t({}^{N}H_t)^2] \cdot [{}^{N}FR_t - 1]^2 \leq u\} \qquad \text{--- (5.41)}$$

; here, this shall be called the non-stationary controlling parameter space, and further, we note that in fact, this space does not contain any specific boundary condition for $k_t({}^{N}H_t)$ [see section 5.4.2].

In accordance with our previous discussions summarised in (5.39)~(5.41), we can construct the following non-stationary version of the quasi-stationary LQP optimisation problem (5.22) formulated in section 5.4.2:

$$\min_{\{k_t({}^{N}H_t);\, t = 0, 1, 2, \ldots\}} \left\{ \sum_{t=0}^{\infty} e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k_t({}^{N}H_t)^2] \cdot [{}^{N}FR_t - 1]^2 \right\}$$

subject to given $\theta \in (0, 1)$; controlled object is governed by the non-stationary system equation (5.40); and $k_t({}^{N}H_t) \in$ {non-stationary controlling parameter space (5.41)}.

--- (5.42)

Section 5.5.4 demonstrates that the optimisation procedure with respect to $k_0(H_0)$, $k_1(H_0)$, $k_2(H_2)$, ... can not be solved. In order to obtain a physically available optimisation procedures, we reformulate this control problem (5.42) according to the approximate model proposed for the non-autonomous system equation (5.40) (- this will be considered in section 5.5.5).

5.5.4 Control optimisation procedure - an illustration of insolubility

The objective of this section is to provide an insight into the difference between optimising our non-stationary control problem (5.42) and optimising the quasi-stationary control problem investigated in section 5.4.5. As a result, the optimal spread policy for the non-stationary control problem will be non-stationary from the fact that the controlled object governed by equation (5.40) and the performance index per unit control period [t, t+1), i.e. $e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k_t({}^N H_t)^2] \cdot [{}^N FR_t - 1]^2)$ depend explicitly on time t, while as illustrated in section 5.4, the quasi-stationary control problem (5.22) has an optimal stationary spread policy defined by the equation (5.32). In the practical aspect of computation, the infinite sequence $\{k^*_t(H_t);\ t=0, 1, 2, \ldots\}$ is not realisable (although we apply the possible spread rule of setting $k^*_t(H_t) = 0$ for ${}^N FR_t - 1 = 0$, see subsection (ii) of section 5.4.3.2), since an infinite number of functions $k^*_t(.)$ are involved because the number of non-zero-valued dynamic states will be infinite because $\text{Prob}[{}^N q_2(t) = 0 \text{ for all } t] = 0$.

We shall focus on proving the insolubility of non-stationary control problem (5.42). Consider time $t \in [0, \infty)$. Then, we can rewrite our performance index ${}^N IPI_\theta$ as the sum of two parts in a similar manner to ${}^Q IPI_\theta$ expressed in equation (5.22),

$$ {}^N IPI_\theta = \sum_{s=0}^{\infty} \{ e^{-\eta s} \cdot (\theta + (1-\theta) \cdot k_s({}^N H_s)^2) \cdot ({}^N FR_s - 1)^2 \} = {}^N IPIA_\theta + {}^N IPIB_\theta \qquad \text{--- (5.43)} $$

where

$$^{N}IPIA_{\theta} \equiv \sum_{s=0}^{t-1} \{e^{-\eta s} \cdot (\theta+(1-\theta) \cdot k_s(^{N}H_s)^2) \cdot (^{N}FR_s-1)^2\} \text{ and}$$

$$^{N}IPIB_{\theta} \equiv \sum_{s=t}^{\infty} \{e^{-\eta s} \cdot (\theta+(1-\theta) \cdot k_s(^{N}H_s)^2) \cdot (^{N}FR_s-1)^2\}.$$

So, $^{N}IPIA_{\theta}$ does not depend on the decisions to be made, i.e. $k_t(^{N}H_t)$, $k_{t+1}(^{N}H_{t+1})$, ..., and then minimising $^{N}IPI_{\theta}$ with respect to these controlling parameters is equivalent to minimising $^{Q}IPIB_{\theta}$. Furthermore, the second part $^{Q}IPIB_{\theta}$ can be expressed as a time-varying function of the current dynamic state $^{N}FR_t$-1 in view of the fact that the sequence of future dynamic states $\{^{N}FR_{t+1}-1,\ ^{N}FR_{t+2}-1,\ \ldots\}$ is recursively generated from the current dynamic state $^{N}FR_t$-1 by way of the non-stationary system equation (5.41) and following the same mathematical deduction steps as in the quasi-stationary control problem examined in section 5.4.5, we can verify that the dynamic programming approach based on Bellman's principle of optimality specifies $k_t(^{N}H_t)$ as a time-varying function of current state variable $^{N}FR_t$-1 (i.e. the knowledge of the current dynamic state is enough to substitute for the information history up to time t, $^{N}H_t$); that is, $k_t(^{N}H_t) = k_t(^{N}FR_t-1)$ for all t.

Next, the optimal spread control sequence $\{k^*_t(^{N}H_t);\ t=0, 1, 2, \ldots\}$ has to satisfy the following infinite number of optimal functional equations: that is, for every $t \in [0, \infty)$,

$$^{N}I^*(^{N}FR_t-1, t) \equiv \underset{\{k_s(^{N}H_s);\ s=t, t+1, t+2, \ldots\}}{\text{Min}} \{ \sum_{s=t}^{\infty} e^{-\eta s} \cdot [\theta + (1-\theta) \cdot k_s(^{N}H_s)^2] \cdot [^{N}FR_s - 1]^2 \}$$

$$= \underset{k_t(^{N}H_t)}{\text{Min}} \{ e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k_t(^{N}H_t)^2] \cdot [^{N}FR_t - 1]^2 +$$

$$\underset{\{k_s(^{N}H_s);\ s=t+1, t+2, \ldots\}}{\text{Min}} \{ \sum_{s=t+1}^{\infty} e^{-\eta s} \cdot [\theta + (1-\theta) \cdot k_s(^{N}H_s)^2] \cdot [^{N}FR_s - 1]^2 \}$$

$$= \underset{k_t(^{N}H_t)}{\text{Min}} \{ e^{-\eta t} \cdot [\theta + (1-\theta) \cdot k_t(^{N}H_t)^2] \cdot [^{N}FR_t - 1]^2 + {}^{N}I^*(^{N}FR_{t+1}, t+1) \},$$

or alternatively, by letting $e^{\eta t} \cdot {}^{N}I^{*}({}^{N}FR_t\text{-}1, t) = {}^{N}J^{*}({}^{N}FR_t\text{-}1, t)$ (i.e. converting the minimal future cost discounted at time 0 to the minimal future cost discounted at time t), the alternative form of the above equation is given, for every $t \in [0, \infty)$, by

$$ {}^{N}J^{*}({}^{N}FR_t\text{-}1, t) = \underset{k_t({}^{N}H_t)}{\text{Min}} \{[\theta + (1\text{-}\theta) \cdot k_t({}^{N}H_t)^2] \cdot [{}^{N}FR_t - 1]^2 + e^{-\eta} \cdot {}^{N}J^{*}({}^{N}FR_{t+1}, t+1)\} \quad \text{---} (5.44) $$

; this has a similar mathematical form to the time-invariant optimal functional equation (5.28) derived in section 5.4.5.1, except for the time variable t; in this respect, we shall call this equation the time-varying optimal functional equation.

Consequently, the optimal spread control sequence $\{k^{*}_t({}^{N}H_t);\ t=0, 1, 2, \ldots\}$ is not available from a practical point of view because $\{k^{*}_t({}^{N}H_t);\ t=0, 1, 2, \ldots\}$ requires an infinite number of optimising computations because as mentioned at the early stage of this section, Prob[the total number of non-zero-valued states $< \infty$, over control period $[0, \infty)$] = 0 due to Prob[${}^{N}q_2(t)$ =0 for all t] = 0. So, we need to develop an approximate model for the non-autonomous system equation by reference to the structural characteristics of the sequence $\{{}^{N}FR_t\text{-}1;\ t=0, 1, 2, \ldots\}$ in order to find a best (not necessary optimal) spread control sequence available in the light of the criterion of computational feasibility (- this subject will be considered in the next section 5.5.5). The next section consider an approximate optimisation procedure for the non-stationary control problem (5.42).

5.5.5 Dynamic pension funding plan for threshold LQP optimisation problem

5.5.5.1 Introduction

As seen in the above section 5.5.4, we noted that the insolubility of our control problem (5.42) arose from the non-stationarity property of equation (5.40). In section 5.5.5.2, we propose the threshold controlled object as an approximate model to the non-stationary controlled object in the light of the criterion of computational feasibility for optimal solutions. Accordingly, we construct the threshold LQP optimisation problem as an approximate LQP optimisation problem to the non-stationary LQP optimisation problem (5.42) in section 5.5.5.3 and then consider its control optimisation procedure in section 5.5.5.4. Finally, we provide a summary and numerical illustrations in sections 5.5.5.5 and 5.5.5.6, respectively.

To begin with, we note that the in order to emphasise that each main symbol to be introduced in this section concerns the threshold LQP optimisation problem, we shall put the superscript 'T' on its left side.

5.5.5.2 Threshold controlled object

As a best alternative to the non-autonomous system equation (5.40), we here propose the threshold system equation specified as a switching system equation from non-autonomous to quasi-autonomous at some fixed point of time. Accordingly, the controlled object governed by this threshold system equation shall be called the threshold controlled object; here, the term 'threshold' is adopted from 'threshold autoregressive models' in the area of time series analysis [for details and examples, see Tong & Lim (1980)]. Its mathematical specification takes the following approximation steps and finally we derive the threshold system equation (5.49).

<u>Step 1. Analysis of time series $\{\exp(\delta_{t+1})\colon t\in[0, \infty)\}$</u>: The non-stationarity property of equation (5.41) is completely identified by the time-varying parameter δ_{t+1}. As was shown in section 5.5.2, the time series $\{\exp(\delta_{t+1})\colon t\in[0, \infty)\}$ governed by model (5.38) is characterised to be damped and eventually convergent to its limit value $1+i_e$. Since the value of t required for this limit value to be reached will be infinite, we need to find, from an approximation point of view, a point of time (denoted by t^*, which shall be called the threshold time value) after which the

time series, $\{\exp(\delta_{t+1}): t \in [0, \infty)\}$, remains relatively unchanged. In this respect, we can set the approximation criterion for determining t^*;

Step 2. Approximation criterion for the threshold time value t^*: For a chosen allowable approximation error value based on the analysis of the time series $\{\exp(\delta_{t+1}): t \in [0, \infty)\}$ (denoted by $\Delta_\delta > 0$), find $t^* \in [0, \infty)$ satisfying $|\exp(\delta_{t+1}) - \exp(\delta_{t^*+1})| < \Delta_\delta$ for any $t = t^*+1, t^*+2, t^*+3, \ldots$. In general, the approximation error value, Δ_δ, would be required to be as small as possible to represent the appropriateness of the approximation. Further, our infinite control horizon can be classified into the two control regimes, i.e. $[0, \infty) = \text{Regime1} \cup \text{Regime2}$, where $\text{Regime1} = \{0, 1, \ldots, t^*-1\}$ and $\text{Regime2} = \{t^*, t^*+1, t^*+2, \ldots\}$; and then,

Step 3. Parametric approximation assumptions: From an approximation point of view, it would be possible to consider dividing the model (5.38) into different models for the two control regimes Regime1 and Regime2, switching from one to the other when time t has crossed the threshold time value t^*. For simplicity, we assume that the sinc-function, i.e. sinc(t+1), is identically equal to its limit value of $1+i_e$ over Regime2: in other words,

$$\exp(\delta_{t+1}) = (1+i_{t+1}) = \begin{cases} 1 + i_e + \text{sinc}(t+1),\ \ i_e > 0 & \text{for each } t \in \text{Regime1} \\ 1 + i_e,\ \ i_e > 0 & \text{for each } t \in \text{Regime2} \end{cases} \quad \text{--- (5.45)}$$

Thus, it would be consistent with the above approximation model (5.45) to reformulate the spread parameter function (5.37) in a way that letting the currently available information vector at time $t \in [0, \infty)$ under the threshold controlled object as ${}^{T}H_t = ({}^{T}FR_0, {}^{T}FR_1, \ldots, {}^{T}FR_t, {}^{T}CR_0, {}^{T}CR_1, \ldots, {}^{T}CR_{t-1})$ with given ${}^{T}H_0 = {}^{T}FR_0$ and denoting the indicator function of set S by 1_S, then for each $t \in [0, \infty)$,

$$k_t = k_t(^TH_t) \cdot 1_{Regime1} + k(^TH_t) \cdot 1_{Regime2}. \qquad \text{--- (5.46)}$$

Therefore, adapting the above parametric approximation model (5.45) and (5.46) to the spread funding formula (5.1) and the system equation (5.2), we derive their respective threshold versions of the non-stationary spread funding formula (5.39) and non-autonomous system equation (5.40): that is, for each $t \in [0, \infty)$,

$$^TCR_t = {}^TNR_t - [k_t(^TH_t) \cdot 1_{Regime1} + k(^TH_t) \cdot 1_{Regime2}] \cdot [^TFR_t - 1]$$
$$\equiv {}^T\mu(^TFR_t\text{-}1;\ k_t(^TH_t)\cdot 1_{Regime1} + k(^TH_t)\cdot 1_{Regime2}) \qquad \text{--- (5.47)}$$

where $^TNR_t = {}^NNR_t = {}^QNR_t = {}^SNR_t$

; in particular, $k_t(^TH_t)\cdot 1_{Regime1} + k(^TH_t)\cdot 1_{Regime2}$ shall be called the threshold spread parameter, this equation the threshold spread funding formula and $^T\mu(.;\ k_t(^TH_t)\cdot 1_{Regime1} + k(^TH_t)\cdot 1_{Regime2})$ the threshold spread funding plan (here, the term 'threshold' is added to indicate that this model allows the structural change according to switching from Regime1 to Regime2); and accordingly, the following system equation shall be called the threshold system equation, which governs our threshold controlled object: that is, for each $t \in [0, \infty)$,

$$^TFR_{t+1} - 1 = [^Tq_1(t) \cdot (1\text{-}k_t(^TH_t)) \cdot 1_{Regime1} + {}^Tq_1 \cdot (1\text{-}k(^TH_t)) \cdot 1_{Regime2}] \cdot [^TFR_t - 1] +$$
$$[^Tq_2(t) \cdot 1_{Regime1} + {}^Tq_2 \cdot 1_{Regime2}] \quad \text{with given } {}^TFR_0 - 1 \qquad \text{--- (5.48)}$$

where $^Tq_1(t) = {}^Nq_1(t)$ and $^Tq_2(t) = {}^Nq_2(t)$ for each $t \in$ Regime1, in which both $^Nq_1(t)$ and $^Nq_2(t)$ are defined in non-autonomous system equation (5.41), while $^Tq_1$ and $^Tq_2$ are their respective constant estimates such that $^Tq_1 = (1+i_e) \cdot \exp(-\alpha-\beta)$ and $^Tq_2 = (1+i_e) \cdot \exp(-\eta) - 1$ for each $t \in$ Regime2; and further, there is no loss of generality in assuming that the value of the starting state $^TFR_{t*}$-1 in Regime2 is non-zero because if $^TFR_{t*}\text{-}1 = 0$, then we can simply redefine the threshold time value as t^*+1 from the viewpoint of the parametric approximation.

Hence, the threshold controlled object with parameter set $\{{}^Tq_1(t), {}^Tq_1, {}^Tq_2(t), {}^Tq_2, k_t({}^TH_t), k({}^TH_t)\}$ has a structural change from the non-stationary controlled object with parameter set $\{{}^Tq_1(t), {}^Tq_2(t), k_t({}^TH_t)\}$ to the quasi-stationary controlled object with parameter set $\{{}^Tq_1, {}^Tq_2, k({}^TH_t)\}$ when the threshold time value t^* is encountered. As a best (not necessary optimal) approximation to the non-stationary LQP optimisation problem (5.22) in view of the requirements of physical solubility, we construct the threshold LQP optimisation problem designed to control optimally the threshold controlled object in the next section 5.5.5.3.

5.5.5.3 Threshold LQP optimisation problem

In a similar manner to sections 5.4.2 and 5.5.3, we need to ensure the finiteness of the threshold performance index (denoted by ${}^TIPI_\theta$) defined by the equation

$${}^TIPI_\theta = \sum_{t=0}^{\infty} \{ e^{-\eta t} \cdot [\theta + (1-\theta) \cdot (k_t({}^TH_t) \cdot 1_{Regime1} + k({}^TH_t) \cdot 1_{Regime2})^2] \cdot [{}^TFR_t - 1]^2 \}.$$

So, ${}^TIPI_\theta$ will be well-defined as a function value $< \infty$ for all t on the following space, since Regime1 is a finite period but Regime2 is an infinite period, so we do not need to put any constraints on a performance index over Regime1: that is, letting u be some positive real number, then for all $t \in$ Regime2,

$$\{k({}^TH):\ \eta > 0 \ \text{and}\ 0 \leq [\theta + (1-\theta) \cdot k({}^TH_t)^2] \cdot [{}^TFR_t - 1]^2 \leq u\} \qquad \text{--- (5.49)}$$

; here, this shall be called the threshold controlling parameter space, which has the same form as the quasi-stationary controlling parameter space II (5.20).

Finally, we can construct the following approximation version of the non-stationary LQP optimisation problem (5.43) (in particular, this shall be called the threshold LQP optimisation problem):

$$\min_{\{k_t(^TH_t);\ t=0,1,\ldots,t^*-1\}} \left\{ \left[\sum_{t=0}^{t^*-1} e^{-\eta t} \cdot (\theta + (1-\theta) \cdot k_t(^TH_t)^2) \cdot (^TFR_t - 1)^2 \right] + \min_{\{k(^TH_t);\ t=t^*, t^*+1, t^*+2, \ldots\}} \left[\sum_{t=t^*}^{\infty} e^{-\eta t} \cdot (\theta + (1-\theta) \cdot k(^TH_t)^2) \cdot (^TFR_t - 1)^2 \right] \right\}$$

subject to given $\theta \in (0, 1)$; controlled object is governed by the threshold system equation (5.48); and $k(^TH_t) \in$ {threshold controlling parameter space (5.49)}.

--- (5.50)

This control problem would be identified by the characteristics of breaks in control structure and switching regimes from Regime1 to Regime2. The optimisation procedure will then be taken separately over the two periods, i.e. Regime1 and Regime2, in the next section 5.5.5.4.

5.5.5.4 Control optimisation procedure

As noted earlier in section 5.5.4, the computational insolubility of the non-stationary LQP optimisation problem (5.42) comes from the infinite number of optimal functional equations established in (5.44). However, the threshold control problem (5.50) can be solved because the optimal spread control sequence, say $\{k^*_0(H_0), k^*_1(H_1), \ldots, k^*_{t^*-1}(H_{t^*-1}), k^*(H_{t^*}), k^*(H_{t^*+1}), k^*(H_{t^*-2}), \ldots\}$, will be produced by solving a finite number of $(t^* + 1)$ optimal functional equations, as illustrated below in subsection (i); The rigorous optimisation procedure will be fully considered in subsection (ii).

(i) Optimal functional equations:

As could be expected from the fact that the threshold control problem can be decomposed into two distinct problems - non-stationary control problem over Regime1 and quasi-stationary control problem over Regime2, the derivation of our optimal functional equations involves the principal results obtained previously, such as the time-invariant optimal functional equation

(5.28) (for quasi-stationary control optimisation derived in section 5.4.5) and the time-varying optimal functional equation (5.44) (for non-stationary control optimisation derived in section 5.5.4). Using these results, we shall verify the computational solubility of the threshold control problem (5.50) by providing the following form of the optimal functional equations (5.51) and (5.52) [see Bertsekas (1976; pp274~275)]. In a similar manner to the derivation of equation (5.44), we have

$$^{T}I^{*}(^{T}FR_{t}\text{-}1, t) \equiv \underset{\{k_s(^{T}H_s);\, s=t,\, t+1,\, \ldots,\, t^*-1\}}{\text{Min}} \{ [\sum_{s=t}^{t^*-1} e^{-\eta s} \cdot (\theta + (1\text{-}\theta)\cdot k_s(^{T}H_s)^2) \cdot (^{T}FR_s - 1)^2] +$$

$$\underset{\{k(^{T}H_s);\, s=t^*,\, t^*+1,\, t^*+2,\, \ldots\}}{\text{Min}} [\sum_{s=t^*}^{\infty} e^{-\eta t} \cdot (\theta + (1\text{-}\theta)\cdot k(^{T}H_t)^2) \cdot (^{T}FR_t - 1)^2] \}.$$

$$= \underset{k_t(^{T}H_t)}{\text{Min}} \{ e^{-\eta t} \cdot (\theta + (1\text{-}\theta)\cdot k_t(^{T}H_t)^2) \cdot (^{T}FR_t - 1)^2 + {}^{T}I^{*}(^{T}FR_{t+1}\text{-}1, t+1) \},$$

for each t ∈ Regime1; and similarly

$$^{T}I^{*}(^{T}FR_{t}\text{-}1, t) = \underset{k(^{T}H_t)}{\text{Min}} \{ e^{-\eta t} \cdot (\theta + (1\text{-}\theta)\cdot k(^{T}H_t)^2) \cdot (^{T}FR_t - 1)^2 + {}^{T}I^{*}(^{T}FR_{t+1}\text{-}1, t+1) \},$$

for each t ∈ Regime2.

Or alternatively, letting $^{T}J^{*}(^{T}FR_{t}\text{-}1, t) = e^{\eta t} \cdot {}^{T}I^{*}(^{T}FR_{t}\text{-}1, t)$ (i.e. converting the minimal future cost discounted at time 0 to the minimal future cost discounted at time t, as in sections 5.4.5 and 5.5.4), then these equations can be classified into two distinct groups, say time-varying and time-invariant: that is,

<u>G1. Time-varying group</u>: for every t ∈ Regime1

$$^{T}J^{*}(^{T}FR_{t}\text{-}1, t) = \underset{k_t(^{T}H_t)}{\text{Min}} \{(\theta + (1\text{-}\theta) \cdot k_t(^{T}H_t)^2) \cdot (^{T}FR_t\text{-}1)^2 + e^{-\eta} \cdot {}^{T}J^{*}(^{T}FR_{t+1}\text{-}1, t+1)\}, \quad \text{---} (5.51)$$

which represents the time-varying group of t^* equations and expresses, in particular, the switching structure from time-varying to time-invariant at time $t=t^*-1$, i.e.

$$^{T}J^{*}(^{T}FR_{t^{*}-1}-1, t^{*}-1) = \underset{k_{t^{*}-1}(^{T}H_{t^{*}-1})}{\text{Min}} \{(\theta + (1-\theta) \cdot k_{t^{*}-1}(^{T}H_{t^{*}-1})^2) \cdot (^{T}FR_{t^{*}-1}-1)^2 + e^{-\eta} \cdot {}^{T}J^{*}(^{T}FR_{t^{*}}-1)\};$$ and

G2. Time-invariant group: for every $t \in$ Regime2,

$$^{T}J^{*}(^{T}FR_{t}-1) = \underset{k(^{T}H_t)}{\text{Min}} \{ (\theta + (1-\theta) \cdot k(^{T}H_t)^2) \cdot (^{T}FR_t - 1)^2 + e^{-\eta} \cdot {}^{T}J^{*}(^{T}FR_{t+1}-1) \}, \quad \text{--- (5.52)}$$

which presents explicitly an infinite number of optimal functional equations but due to the time-invariability of function $^{T}J^{*}(.)$, it is sufficient to solve the optimal functional equation at time t^{*} in order to specify the optimal stationary spread policy $k^{*}(.)$, as considered in the quasi-stationary control problem of section 5.4.5 (we note that as assumed in the threshold system equation (5.50), the starting state $^{T}FR_{t^{*}}-1$ in Regime2 is set to be non-zero).

Here, the above equations (5.51) and (5.52) shall be called the threshold optimal functional equation, which is distinguishable from the time-invariant optimal functional equation (5.28) as well as the time-varying optimal functional equation (5.44) especially by the equation at $t=t^{*}-1$.

(ii) Solution of the threshold optimal functional equation:

As a preliminary to solving the threshold optimal functional equation, it is worth recalling that (a) the time-invariant part (5.52) is soluble simply by adjusting the main results (5.28)~(5.32), derived by the FDP approach in section 5.4.5, with replacing q_1 and q_2 involved in (5.28)~(5.32) with $^{T}q_1$ and $^{T}q_2$, respectively, and accordingly, we can obtain $^{T}J^{*}(^{T}FR_{t^{*}}-1)$; and (b) by using $^{T}J^{*}(^{T}FR_{t^{*}}-1)$ as a boundary condition of the time-varying part (5.51), it is soluble by the BDP approach in a similar manner to solving the finite-horizon deterministic control problem (4.5) examined in section 4.3.2: in other words, we may think that the optimal function value at the threshold time value t^{*}, i.e. $J^{*}(^{T}FR_{t^{*}}-1)$, provides the conversion from the framework of the infinite-horizon threshold control problem to that of the finite-horizon threshold control problem.

Now, we shall clarify mathematically the above descriptions (a) and (b) in turn.

Firstly, the solution of time-invariant part (5.52) is a specific case of the main results (5.28)~(5.32) derived in section 5.4.5: for any non-zero starting state ${}^{T}FR_{t*}$-1 in Regime2,

${}^{T}J^{*}({}^{T}FR_{t*}-1) = {}^{T}Q_1 \cdot ({}^{T}FR_{t*} - 1)^2 + {}^{T}Q_2 \cdot ({}^{T}FR_{t*} - 1) + {}^{T}Q_3$ (which plays the role of the boundary condition of the time-varying part (5.51)); and

$$k^{*}({}^{T}H_{t*}) = k^{*}({}^{T}FR_{t*}-1) = [(e^{-\eta} \cdot {}^{T}q_1^{\,2} \cdot {}^{T}Q_1) \cdot ({}^{T}FR_{t*} - 1) + e^{-\eta} \cdot {}^{T}q_1 \cdot ({}^{T}q_2 \cdot {}^{T}Q_1 + {}^{T}Q_2/2)]$$

$$/ \, [((1-\theta) + e^{-\eta} \cdot {}^{T}q_1^{\,2} \cdot {}^{T}Q_1) \cdot ({}^{T}FR_{t*} - 1)] \qquad \text{--- (5.53)}$$

where

$${}^{T}Q_1 = [\theta \cdot (1-\theta) + e^{-\eta} \cdot {}^{T}q_1^{\,2} \cdot {}^{T}Q_1] \, / \, [(1-\theta) + e^{-\eta} \cdot {}^{T}q_1^{\,2} \cdot {}^{T}Q_1];$$

$${}^{T}Q_2 = [2(1-\theta) \cdot e^{-\eta} \cdot {}^{T}q_1 \cdot ({}^{T}q_2 \cdot {}^{T}Q_1 + {}^{T}Q_2 / 2)] \, / \, [(1-\theta) + e^{-\eta} \cdot {}^{T}q_1^{\,2} \cdot {}^{T}Q_1]; \text{ and}$$

$${}^{T}Q_3 = [(1-\theta) \cdot e^{-\eta} \cdot ({}^{T}q_2^{\,2} \cdot {}^{T}Q_1 + {}^{T}q_2 \cdot {}^{T}Q_2 + {}^{T}Q_3) + e^{-2\eta} \cdot {}^{T}q_1^{\,2} \cdot ({}^{T}Q_1 \cdot {}^{T}Q_3 - {}^{T}Q_2^{\,2}/4)]$$

$$/ \, [(1-\theta) + e^{-\eta} \cdot {}^{T}q_1^{\,2} \cdot {}^{T}Q_1].$$

Further, the stationary optimal spread control policy $k^{*}(.)$ is not well defined for 100% funding level, in which we apply the possible spread rule for the uniqueness of the optimal spread control sequence $\{k^{*}({}^{T}H_t); t \in \text{Regime2}\}$ [see subsection (ii) of section 5.4.3.2].

Secondly, the non-stationary optimal spread policy $\{k^{*}_t(.); t \in \text{Regime1}\}$ can be found by solving the time-varying group of t^{*} equations with the boundary condition (5.53), that is, for each $t \in$ Regime1,

$${}^{T}J^{*}({}^{T}FR_t-1, t) = \underset{k_t({}^{T}H_t)}{\text{Min}} \{(\theta + (1-\theta) \cdot k_t({}^{T}H_t)^2) \cdot ({}^{T}FR_t-1)^2 + e^{-\eta} \cdot {}^{T}J^{*}({}^{T}FR_{t+1}-1, t+1)\} \text{ with}$$

${}^{T}J^{*}({}^{T}FR_{t*}-1)$ given by equation (5.53). --- (5.54)

The solution of the above equation (5.54) is uniquely determined by the following suggested form of quadratic function, subject to some condition (derived later in (5.56)): that is, for each $t \in$ Regime1,

$^{T}J^{*}(^{T}FR_t\text{-}1, t) = {}^{T}Q_1(t) \cdot (^{T}FR_t - 1)^2 + {}^{T}Q_2(t) \cdot (^{T}FR_t - 1) + {}^{T}Q_3(t)$ with the boundary condition $^{T}Q_1(t^*) = {}^{T}Q_1$, $^{T}Q_2(t^*) = {}^{T}Q_2$ and $^{T}Q_3(t^*) = {}^{T}Q_3$, --- (5.55)

which can be verified using an mathematical induction argument, as illustrated below.

This form holds clearly for $t = t^*$ and then proceeding by mathematical induction, we have for each $t \in$ Regime1,

$^{T}J^{*}(^{T}FR_{t+1}\text{-}1, t+1) = {}^{T}Q_1(t+1) \cdot (^{T}FR_{t+1} - 1)^2 + {}^{T}Q_2(t+1) \cdot (^{T}FR_{t+1} - 1) + {}^{T}Q_3(t+1)$ with the boundary condition $^{T}Q_1(t^*) = {}^{T}Q_1$, $^{T}Q_2(t^*) = {}^{T}Q_2$ and $^{T}Q_3(t^*) = {}^{T}Q_3$.

Introducing the above suggested solution into equation (5.54), we obtain the following form:

$$^{T}J^{*}(^{T}FR_t\text{-}1, t) = \underset{k_t(^{T}H_t)}{\text{Min}} \{ {}^{T}G(k_t(^{T}H_t), t) \} \text{ for all } t \in \text{Regime1},$$

where

$$\begin{aligned}^{T}G(k_t(^{T}H_t), t) = {} & \{ [(1\text{-}\theta) + e^{-\eta} \cdot {}^{T}q_1(t)^2 \cdot {}^{T}Q_1(t+1)] \cdot (^{T}FR_t\text{-}1)^2 \} \cdot k_t(^{T}H_t)^2 + \\ & \{ -2\, e^{-\eta} \cdot [\, {}^{T}q_1(t)^2 \cdot {}^{T}Q_1(t+1) \cdot (^{T}FR_t\text{-}1)^2 + (^{T}q_1(t) \cdot {}^{T}q_2(t) \cdot {}^{T}Q_1(t+1) + {}^{T}q_1(t) \cdot \\ & {}^{T}Q_2(t+1)/2) \cdot (^{T}FR_t\text{-}1)] \} \cdot k_t(^{T}H_t) + \{ \theta \cdot (^{T}FR_t\text{-}1)^2 + e^{-\eta} \cdot [(^{T}q_1(t) \cdot (^{T}FR_t\text{-}1) + \\ & {}^{T}q_2(t))^2 \cdot {}^{T}Q_1(t+1) + (^{T}q_1(t) \cdot (^{T}FR_t\text{-}1) + {}^{T}q_2(t)) \cdot {}^{T}Q_2(t+1) + {}^{T}Q_3(t+1)] \}.\end{aligned}$$

So, $^{T}G(k(^{T}H_t), t)$ is a strictly convex function of $k_t(^{T}H_t)$, subject to the condition given by

$$[(1\text{-}\theta) + e^{-\eta} \cdot {}^{T}q_1(t)^2 \cdot {}^{T}Q_1(t+1)] \cdot (^{T}FR_t\text{-}1)^2 > 0 \text{ for all } t \in \text{Regime1}. \qquad \text{--- (5.56)}$$

Hence, differentiating with respect to $k_t({}^TH_t)$ under the condition (5.56) and setting the resulting derivative equal to zero, we obtain the optimal value of $k_t({}^TH_t)$ (denoted by $k_t^*({}^TH_t)$) specified by

$$k_t^*({}^TH_t) = \{e^{-\eta} \cdot [{}^Tq_1(t)^2 \cdot {}^TQ_1(t+1) \cdot ({}^TFR_t-1) + ({}^Tq_1(t) \cdot {}^Tq_2(t) \cdot {}^TQ_1(t+1) + {}^Tq_1(t) \cdot {}^TQ_2(t+1)/2)]\} / \{[(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)] \cdot ({}^TFR_t-1)\} \quad \text{--- (5.57)}$$

; hence, the condition (5.56) corresponds to the uniqueness requirement for the optimal spread control sequence $\{k_t^*({}^TH_t);\ t \in \text{Regime1}\}$.

For completion, substituting $k^*({}^TH_t)$ into Min ${}^TG(k({}^TH_t))$ yields

$$\begin{aligned}{}^TJ^*({}^TFR_t-1, t) &= {}^TQ_1(t) \cdot ({}^TFR_t - 1)^2 + {}^TQ_2(t) \cdot ({}^TFR_t - 1) + {}^TQ_3(t) \\ &= \{[\theta \cdot (1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)] / [(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)]\} \cdot ({}^TFR_t-1)^2 + \\ &\quad \{e^{-\eta} \cdot (1-\theta) \cdot [2 \cdot {}^Tq_1(t) \cdot {}^Tq_2(t) \cdot {}^TQ_1(t+1) + {}^Tq_1(t) \cdot {}^TQ_2(t+1)] / \\ &\quad [(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)]\} \cdot ({}^TFR_t-1) + \{e^{-\eta} \cdot (1-\theta) \cdot [{}^Tq_2(t)^2 \cdot {}^TQ_1(t+1) + \\ &\quad {}^Tq_2(t) \cdot {}^TQ_2(t+1) + {}^TQ_3(t+1)] + e^{-2\eta} \cdot [{}^Tq_1(t)^2 \cdot {}^TQ_1(t+1) \cdot {}^TQ_3(t+1) - \\ &\quad {}^Tq_1(t)^2 \cdot {}^TQ_2(t+1)^2/4]\} / \{(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)\}.\end{aligned}$$

Thus, we complete the mathematical induction argument and find finally that the optimal functional equation (5.54) has a solution of the quadratic form (5.55) with ${}^TQ_1(t)$, ${}^TQ_2(t)$ and ${}^TQ_3(t)$ satisfying the following recurrence relation: for each $t \in$ Regime1,

$$\begin{aligned}{}^TQ_1(t) &= [\theta \cdot (1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)] / [(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)]; \\ {}^TQ_2(t) &= e^{-\eta} \cdot (1-\theta) \cdot [2 \cdot {}^Tq_1(t) \cdot {}^Tq_2(t) \cdot {}^TQ_1(t+1) + {}^Tq_1(t) \cdot {}^TQ_2(t+1)] / [(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)];\ \text{and} \\ {}^TQ_3(t) &= \{e^{-\eta} \cdot (1-\theta) \cdot [{}^Tq_2(t)^2 \cdot {}^TQ_1(t+1) + {}^Tq_2(t) \cdot {}^TQ_2(t+1) + {}^TQ_3(t+1)] + e^{-2\eta} \cdot [{}^Tq_1(t)^2 \cdot \\ &\quad {}^TQ_1(t+1) \cdot {}^TQ_3(t+1) - {}^Tq_1(t)^2 \cdot {}^TQ_2(t+1)^2/4]\} / \{(1-\theta) + e^{-\eta} \cdot {}^Tq_1(t)^2 \cdot {}^TQ_1(t+1)\}, \quad \text{--- (5.58)}\end{aligned}$$

starting with their respective boundary conditions ${}^TQ_1(t^*) = {}^TQ_1$, ${}^TQ_2(t^*) = {}^TQ_2$ and ${}^TQ_3(t^*) = {}^TQ_3$.

Since $k_t^*({}^TH_t)$ involves only the parametric functions ${}^TQ_1(t+1)$ and ${}^TQ_2(t+1)$, we need only to solve the backward recursive equations associated with ${}^TQ_1(t)$ and ${}^TQ_2(t)$. Further, we have proved in Proposition 5.2 in section 5.4.5.1 that ${}^TQ_1$ is positive and hence, the sequence $\{{}^TQ_1(0), {}^TQ_1(1), \ldots, {}^TQ_1(t^*)\}$ is also positive because $1 < \theta < 0$. Thus, the uniqueness requirement (5.56) for $\{k_t^*({}^TH_t); t\in\text{Regime1}\}$ leads to a simplified form that ${}^TFR_t\text{-}1 \neq 0$ for all $t\in\text{Regime1}$.

As a result, the optimal spread control at time t, i.e. $k^*_t({}^TH_t)$, is expressed as a function of the current state variable ${}^TFR_t\text{-}1$ and the resulting optimal function value, i.e. ${}^TJ^*({}^TFR_t\text{-}1, t)$, turns out to be quadratic, subject to ${}^TFR_t\text{-}1 \neq 0$. Further, the unique optimal spread control sequence $\{k^*_t({}^TH_t); t\in\text{Regime1}\}$ is provided under the assumption that ${}^TFR_t\text{-}1 \neq 0$ for all $t\in\text{Regime1}$.

<u>Remark 5.3</u>: (a) The case of ${}^TFR_j\text{-}1 = 0$ for some $j\in\text{Regime1}$ violates the uniqueness requirement (5.56) because we can determine $k^*_j(H_j)$ arbitrarily. For attaining a unique optimal spread control sequence $\{k^*_t({}^TH_t); t\in\text{Regime1}\}$, we will apply our possible spread rule of setting $k^*_j(H_j) = 0$ [see subsection (ii) of section 5.4.3.2];

(b) However, the parametric recursions (5.58) are no longer applicable, since the optimal functional equation at time j is given by ${}^TJ^*({}^TFR_j\text{-}1, j) = e^{-\eta}\cdot {}^TJ^*({}^TFR_{j+1}\text{-}1, j+1)$, which leads to, due to ${}^TFR_{j+1}\text{-}1 = {}^Tq_2(j)$ (for convenience, assumed here to be non-zero) for ${}^TFR_j - 1 = 0$ from the threshold system equation (5.48),

$$ {}^TJ^*({}^TFR_j\text{-}1, j) = e^{-\eta}\cdot[{}^Tq_2(j)^2\cdot {}^TQ_1(j+1) + {}^Tq_2(j)\cdot {}^TQ_2(j+1) + {}^TQ_3(j+1)]; \qquad \text{---}\ (5.59) $$

(c) Hence, letting Regime1' = {0, 1, 2, ..., j-1}, then the remaining optimal spread policy $\{k^*_t(.); t\in\text{Regime1'}\}$ can be found by solving the remaining time-varying group of j equations with the new boundary condition (5.59): that is, for each $t\in\text{Regime1'}$

$^TJ^*(^TFR_t\text{-}1, t) = \underset{k_t(^TH_t)}{\text{Min}} \{(\theta+(1-\theta)\cdot k_t(^TH_t)^2)\cdot(^TFR_t\text{-}1)^2 + e^{-\eta}\cdot{}^TJ^*(^TFR_{t+1}\text{-}1, t+1)\}$ with

$^TJ^*(^TFR_j\text{-}1)$ given by equation (5.59); and --- (5.60)

(d) Finally, we can then solve the remaining optimal functional equations (5.60) with a suggested solution of the following form in a similar manner to the previous optimisation procedures (5.55)~(5.58) for t=j+1, j+2, ..., t^*-1: in short, for each t∈Regime1'

$^TJ^*(^TFR_t\text{-}1, t) = {}^TQ'_1(t)\cdot(^TFR_t - 1)^2 + {}^TQ'_2(t)\cdot(^TFR_t - 1) + {}^TQ'_3(t)$ with the boundary condition $^TQ'_1(j) = 0$, $^TQ'_2(j) = 0$ and $^TQ'_3(j) = e^{-\eta}\cdot[^Tq_2(j)^2\cdot{}^TQ_1(j+1) + {}^Tq_2(j)\cdot{}^TQ_2(j+1) + {}^TQ_3(j+1)]$ in which $^TQ_1(j+1)$, $^TQ_2(j+1)$ and $^TQ_3(j+1)$ each are proceeded from the parametric recursions (5.60); and

(e) As a summary, whenever 100% funding level is encountered during Regime1, we carry out the same optimisation procedure for a unique optimal spread control sequence, as described above in (a)~(d). This kind of complementary optimisation procedure is required because of the weakness in the mathematical formula of the Spread method discussed earlier in subsection (ii) of section 5.4.3.2.

(iii) <u>Conclusion</u>:

The control optimisation procedure with respect to $k_t(^TH_t)\cdot I_{Regime1} + k(^TH_t)\cdot I_{Regime2}$ is classified into two parts together with our possible spread rule - designing optimally the non-stationary optimal spread control policy $k_t^*(.)$ over Regime1 and the stationary optimal spread control policy $k^*(.)$ over Regime2. These two spread control policies are commonly well defined in terms of the unfunded ratio (i.e. $^TFR_t\text{-}1 \in R^1\text{-}\{0\}$ for all t ∈[0, ∞)) as seen in formulae (5.53) and (5.57). Here, we are concerned about formulae (5.53) and (5.57).

Then, substituting $k^*_t(^TH_t)$ and $k^*(^TH_t)$ defined in (5.53) and (5.57) into equation (5.47) yields the optimal threshold spread funding formula given, for each t∈[0, ∞), by

$$^{T}CR^{f*}_{t} = {}^{T}\mu^{*}({}^{T}FR_{t}\text{-}1;\ k^{*}_{t}({}^{T}H_{t})\cdot l_{Regime1} + k^{*}({}^{T}H_{t})\cdot l_{Regime2}),$$

$$= {}^{T}NR^{f}_{t} - [k^{*}_{t}({}^{T}H_{t}) \cdot l_{Regime1} + k^{*}({}^{T}H_{t}) \cdot l_{Regime2}] \cdot [{}^{T}FR_{t} - 1] \qquad \text{--- (5.61)}$$

$$= {}^{T}NR^{f}_{t} - \{ [{}^{T}\varphi(t;\theta) \cdot ({}^{T}FR_{t}\text{-}1) - {}^{T}\xi(t;\theta)] \cdot l_{Regime1} + [{}^{T}\varphi(\theta) \cdot ({}^{T}FR_{t}\text{-}1) - {}^{T}\xi(\theta)] \cdot l_{Regime2} \}$$

where

${}^{T}\varphi(t;\theta) = [e^{-\eta} \cdot {}^{T}q_{1}(t)^{2} \cdot {}^{T}Q_{1}(t+1)] / [(1\text{-}\theta) + e^{-\eta} \cdot {}^{T}q_{1}(t)^{2} \cdot {}^{T}Q_{1}(t+1)],\ 0 < {}^{T}\varphi(t;\theta) < 1;$

${}^{T}\xi(t;\theta) = - [e^{-\eta} \cdot {}^{T}q_{1}(t) \cdot ({}^{T}q_{2}(t) \cdot {}^{T}Q_{1}(t+1) + Q_{2}(t+1)/2)] / [(1\text{-}\theta) + e^{-\eta} \cdot {}^{T}q_{1}(t)^{2} \cdot {}^{T}Q_{1}(t+1)];$

${}^{T}\varphi(\theta) = [e^{-\eta} \cdot {}^{T}q_{1}^{2} \cdot {}^{T}Q_{1}] / [(1\text{-}\theta) + e^{-\eta} \cdot {}^{T}q_{1}^{2} \cdot {}^{T}Q_{1}],\ 0 < {}^{Q}\varphi(\theta) < 1;$

${}^{T}\xi(\theta) = - [e^{-\eta} \cdot {}^{T}q_{1} \cdot ({}^{T}q_{2} \cdot {}^{T}Q_{1} + {}^{T}Q_{2}/2)] / [(1\text{-}\theta) + e^{-\eta} \cdot {}^{T}q_{1}^{2} \cdot {}^{T}Q_{1}];$ and

${}^{T}\mu^{*}(.\ ;\ k^{*}_{t}({}^{T}H_{t})\cdot l_{Regime1} + k^{*}({}^{T}H_{t})\cdot l_{Regime2})$ denotes the optimal threshold spread funding plan (and hence our dynamic pension funding plan) as an optimal control law for the threshold controlled object.

Hence, the above formula (5.61) takes mathematically a similar form to the funding formula (4.18) for the unconstrained form of CR_t derived in section 4.3.3.2. Similarly, ${}^{T}\varphi(t;\theta)\cdot l_{Regime1}$ + ${}^{T}\varphi(\theta)\cdot l_{Regime2}$ can be interpreted as the proportional state-feedback controlling parameter as in $\varphi(t;\theta)$ in (4.18), while ${}^{T}\xi(t;\theta)\cdot l_{Regime1}$ + ${}^{T}\xi(\theta)\cdot l_{Regime2}$ can also be interpreted as the additive controlling parameter as in $\xi(t;\theta)$ in (4.18).

Moreover, the optimal threshold control response ${}^{T}FR^{*}_{t+1}$ corresponding to ${}^{T}CR^{f*}_{t}$ is recursively generated with time t by the optimal threshold system equation: for each $t\in[0, \infty)$ and given ${}^{T}FR_{0}\text{-}1 = {}^{T}FR^{*}_{0}\text{-}1$,

$$^{T}FR^{*}_{t+1}\text{-}1 = \{ [{}^{T}q_{1}(t)\cdot(1 - {}^{T}\varphi(t;\theta)] \cdot [{}^{T}FR^{*}_{t} - 1] + [{}^{T}q_{1}(t)\cdot{}^{T}\xi(t;\theta) + q_{2}(t)] \} \cdot l_{Regime1} +$$
$$\{ [{}^{T}q_{1}\cdot(1 - {}^{T}\varphi(\theta))] \cdot [{}^{T}FR^{*}_{t} - 1] + [{}^{T}q_{1}\cdot{}^{T}\xi(\theta) + q_{2}] \} \cdot l_{Regime2} \qquad \text{--- (5.62)}$$

We note finally that our optimal spread control sequence, $\{k^{*}_{0}({}^{T}H_{0}), k^{*}_{1}({}^{T}H_{1}), \ldots, k^{*}_{t*-1}({}^{T}H_{t*-1}), k^{*}({}^{T}H_{t*}), k^{*}({}^{T}H_{t*+1}), \ldots\}$, will be uniquely generated with time t by switching policies from

$k^*_t(.)$ to $k^*(.)$ at time $t = t^*$, together with the possible spread rule. Further, we may think that in the light of optimal control theory, the possible spread rule represents the weakness in the mathematical formulation specified by the Spread method and hence this is different from optimising the form of the contribution rate examined in section 4.3.

5.5.5.5 Summary

Our control mechanism for the threshold controlled object can be summarised by the following block diagram (which will also give a useful comparison with Figure 5.2 of section 5.4.5.3).

[Threshold controlled object] $\longrightarrow$ $\{{}^{T}FR_t\text{-}1\}$ $\longrightarrow$

$\uparrow$ $\downarrow$

$\{\ k_t({}^{T}H_t)\cdot 1_{Regime1} + k({}^{T}H_t)\cdot 1_{Regime2} = k^*_t({}^{T}H_t)\cdot 1_{Regime1} + k^*({}^{T}H_t)\cdot 1_{Regime2})\ \}$ [Feedback]

$\uparrow$ $\downarrow$

$\{{}^{T}NR^f_t\}$ $\rightarrow$ [Optimal control law ${}^{T}\mu^*(.; k^*_t({}^{T}H_t)\cdot 1_{Regime1} + k^*({}^{T}H_t)\cdot 1_{Regime2})$] $\leftarrow$ $\lrcorner$

Figure 5.3 Optimal threshold spread funding control system.

The above Figure 5.3 shows that there is a control switching from the non-stationary optimal spread control policy $k^*_t(.)$ to the stationary optimal spread control policy $k^*(.)$ when time t has crossed Regime1.

5.5.5.6 Numerical illustrations

The section is to show numerically some illustrative relationships between the optimal value of the threshold spread parameter (i.e. $k^*_t({}^{T}H_t)\cdot 1_{Regime1} + k^*({}^{T}H_t)\cdot 1_{Regime2}$) and the resulting values of the control action error (i.e. ${}^{T}CR^{f*}_t\text{-}{}^{T}NR^f_t$) and the control error (i.e. ${}^{T}FR^*_{t+1}\text{-}1$), subject to the following assumptions given in subsection (i). All the observations fall into two groups - observations in Regime1 and observations in Regime2. Since the latter case has been illustrated in section 5.4.5.4, we here concentrate on the observations in Regime1. Due to the constraints of time and space, we are concerned only with the case of $\theta = 10\%$, in accordance with the comment on a suitable value of θ (described in subsection (i) of section 5.5.5.4) (- the other suitable values of θ would provide similar results as in subsection (iii) of section 5.5.5.4). All the numerical illustrations are given in graphical form in Appendix 5B.5.

(i) Assumptions:

(A1) All actuarial assumptions are the same as given in section 5.4.5.4 (in short, $\exp(\alpha) = 1.03$, $\exp(\beta) = 1.02$ and $\exp(\eta) = 1.06$);

(A2) Projection assumptions:

- infinite control horizon: $t \in [0, \infty)$;
- determined value of $\theta = 10\%$;
- given initial value $FR_0 = 50\%$;
- $\exp(\delta_{t+1})$ is governed by equation (5.44) where $i_e = 0.06$, $T_b = 12$, $\varnothing = 8\pi$; and
- the threshold time value: $t^* = 200$ (and hence Regime1 = {t: 0, 1, 2, ..., 199} and Regime2 = {t: 200, 201, 202, ...}).

(iii) Illustrative numerical results:

The projections of optimal threshold spread funding plan governed by formula (5.61) are presented in Graph 5.3 (composed of Graphs 5.3.1~5.3.4) of Appendix 5B.5.

The important observations are described below:

(a) The damped harmonic motion of real rates of return, illustrated in Graph 5.3.1, dominates the controlled overall motions as illustrated in Graphs 5.3.2~5.3.4: broadly,

- the pattern of $\{^{T}FR^{*}_{t}-1;\ t \in \text{Regime1}\}$ has a similar but proportional cycle to that of real rates of return, while the pattern of $\{^{T}CR^{f*}_{t}-{}^{T}NR^{f}_{t};\ t \in \text{Regime1}\}$ has a similar but adverse cycle to that of the real rates of return. This implies that there is a trade-off between the control errors and control action errors over Regime1, as we know from comparing Graphs 5.3.3 and 5.3.4;

- each peak and trough in Graph 5.3.2 corresponds to a point crossing the 100% funding target in Graph 5.3.3, except for the first peak point, (which seems to be not clearly illustrated in Graphs 5.3.2 and 5.3.3 but is clearly presented in our numerical experiments);

- the non-stationary spread control pattern of $\{k^{*}_{t}(^{T}H_{t});\ t \in \text{Regime1}\}$ can then be thought of as a series of control patterns similar to that in the turmoil stage of the optimal quasi-stationary spread control mechanism (described in section 5.4.5.4), except for the initial time periods [0, 5] prior to improving the initial given (unfunded) ratio 50% up to about 95% which corresponds to the introduction stage of the optimal quasi-stationary spread control mechanism; and

- as a result, under cyclical economic circumstances, it would be necessary to respond somewhat abruptly to the movements in the investment market in order to maintain both funding levels and contribution rates around their respective targets, so it is necessary to extend to some degree the typical spread parameter space $\{k_t: 0 \le k_t \le 1 \text{ for all } t\}$; and

(b) In addition, we note the following results associated with our model of (damped harmonic) business cycle without giving any numerical illustrations (- these are based on our intuition in relation to the numerical results of the quasi-stationary control problem given in section 5.4.5.4):

- increasing the initial phase shift $\varnothing$ reduces both the solvency and contribution rate risks because the larger $\varnothing$ yields the more stationary investment performance: in other words, although Mature and Young pension schemes faces commonly a series of peaks and troughs in the investment market, a Mature pension scheme is likely to control investment risk in a more secure direction than a Young pension scheme due to having relatively more capacity for maintaining reasonably well-diversified portfolios as well as improving the investment risk assessment (- in our model, this kind of a investment risk control capacity would be represented by a larger $\varnothing$, see the end part of subsection (ii) in section 5.5.2); and

- increasing the business cycle period T_b reduces both the solvency and contribution risks, which is quite understandable in view that a stationary economy can be thought of as a case of $T_b \to \infty$.

5.6 Conclusion

Even though dealing with an infinite-horizon control problem seems to be academic and impractical, we believe that our three different optimisation procedures for stationary, quasi-stationary and non-stationary (including threshold) provide a fundamental insight into the finite-horizon control problem.

According to the mathematical and numerical results derived in this chapter, we can make some general conclusions about how to control the value of the unknown spread parameter on a long-term basis for reducing both the solvency and contribution rate risks:

(a) Under a stationary economic and demographic circumstances, we may recommend both the algebraic and dynamic programming (DP) approaches. If their respective requirements are met (i.e. Requirements A1 ~ A3 described in section 5.3.4.2 and Requirements B1 ~ B3 described in section 5.4.5.2), the algebraic approach provides a solution within the typical spread parameter space, while the DP approach provides a distinct control pattern under an investment performance which is better than expected (i.e. $\delta > \eta$). The control pattern is characterised by four consecutive stages, introduction, turmoil, stabilising and maturity stages; and

(b) Under a non-stationary economic and demographic circumstances, the algebraic approach is inapplicable but the dynamic programming approach is suitable from the viewpoint that it provides a time-related response to the dynamic situation and then enables us to maintain the funding level and contribution rate around their respective control targets.

Finally, we shall finish this chapter with our suggestion about the typical spread parameter space $\{k_t: 0 \le k_t \le 1 \text{ for all } t\}$. As illustrated numerically in sections 5.4.5.4 and 5.5.5.6, we need to modify/extend the typical constraint of k_t. The formula $CR_t - NR_t = -k_t \cdot (FR_t - 1)$ with $0 \le k_t \le 1$ is designed to focus on the currently processed unfunded ratio (i.e. $FR_t \neq 100\%$) due to the constraint $0 \le k_t \le 1$. But it does not have a sufficient capacity to reflect (a priori) the more or less predictable future valuation outcomes for the future stability of both the funding levels and contribution rates and hence it is likely to lead to a potential risk of instability under a dynamic economic and demographic situation. In this respect, we would like to suggest a modification of the spread parameter space, e.g. $-1 \le k_t \le 2$ (here, the negative values take an optimistic perspective for the future on the current under-funded state, i.e. $FR_t < 100\%$, and a pessimistic perspective for the future on the current over-funded state, i.e. $FR_t > 100\%$: in particular, if the actual value of FR_t is approximately fully funded, i.e. $FR_t \cong 100\%$, we may allow a larger space as suggested in subsection (iii) of section 5.4.5.4.

Appendix 5A Equilibrium state and Zero-input stability

A.1 Introduction

The objective of this Appendix is to present some of the fundamental concepts of equilibrium state and its stability as discussed in Chapter 5. So, we concentrate on the so-called zero-input stability for a dynamic system governed by the zero-input linear first-order difference equation formulated in the following section A.2. In the application to pension funding, these concepts provide certain conditions for the actuarial (economic and demographic) parameters suitable for designing a stable pension scheme.

As a result, the stability of pension schemes will be a feature attractive to the trustees/members, advising actuary and sponsoring employer from the viewpoint of a long-term pension fund valuation.

First of all, we need to take into account the equilibrium state so as to study the zero-input stability concepts (- these are defined and commented on in section A.2). The concept of zero-input stability concerns the stability of the equilibrium state obtained from a dynamic system equation without input or equivalently from a dynamic system equation with a given input (e.g. a given form of controlling variable specified by the spread funding formula, see the dynamic system equation (3.25) in section 3.4.4), since there is no distinction between them in the light of mathematical modelling. The stability concepts and definitions that has been developed on a continuous-time domain are primarily due to Lyapunov (1892) and are concisely presented on a discrete-time domain in section A.3.

The concepts and definitions given below in sections A.2 and A.3 are based on Willems (1968; Ch.7) and Callier & Desoer (1991; Ch.7d).

A.2 Equilibrium state

Let us start by considering the zero-input linear first-order difference equation given by: for each $t = t_0, t_0+1, t_0+2, \ldots$, and $t_0 \in R^1$,

$X_{t+1} = a(t) \cdot X_t + b(t)$ with the prescribed initial condition $X(t_0)=x_0$ --- (A-1)

where

$X(t)$ = state variable representing the state of the dynamic system at time t;

a(t) and b(t) denotes the system parameters to apply between time t and t+1.

<u>Definition 1</u> (Equilibrium state): Consider the dynamic system governed by equation (A-1). A constant value xe is called an equilibrium state for this dynamic system if xe satisfies the equilibrium equation $xe = a(t) \cdot xe + b(t)$.

<u>Comment 1</u>: the equilibrium state has the following properties (P1)~(P4):

(P1; zero equilibrium state): In the case that either a(t) or b(t) is a time-varying parameter (i.e. in the case of a non-autonomous dynamic system) or a(t)=1 for all t, there exists no equilibrium state because $xe = b(t) / (1-a(t))$ is not constant for all t and is not defined for a(t)=1. However, if b(t)=0 for all t, there exists only a zero equilibrium state, i.e. xe = 0;

(P2; equilibrium solution): In the case that both $a(t)=a\neq1$ and b(t)=b are constant for all t (i.e. in a case of an autonomous dynamic system), there exists a unique equilibrium state $xe = b / (1-a)$ and another point to be noted is that the constant solution $X_t = xe$ is a particular solution to equation (A-1) irrespective of the initial condition. In this case, the definite solution to equation (A-1) can be represented as $X_t = (x_0 - xe) \cdot a^{t - t_0} + xe$ in which we notice that if $x_0 = xe$, then $X_{t_0+1} = X_{t_0+2} = \ldots = xe$ and the constant solution $X_t = xe$ with $X_{t_0} = xe$ is often referred to as the equilibrium solution (which means that a solution that passes through xe at some time (say, $t_{xe} \geq t_0$) remains at this value of xe for all $t \geq t_{xe}$);

(P3; moving equilibrium state): In the case that a(t) = 1 and b(t) = b for all t, the definite solution to equation (A-1) is $X_t = x_0 + b\cdot(t-t_0)$, which shows that with nonzero x_0, there will be a constant deviation x_0 from $b\cdot(t-t_0)$ for each t. From this point of view, the motion is invariant for a translation of time, i.e. $b\cdot(t-t_0)$, which can be thought of as a moving equilibrium state (not equilibrium state). In this case, the time-varying solution $X_t = b\cdot(t-t_0)$ is also a particular solution to equation (A-1) irrespective of the initial condition; and

(P4; the economic meaning of equilibrium): The economic meaning of equilibrium would be referred to Machlup (1958) expressed that an equilibrium is "a constellation of selected interrelated variables so adjusted to one another that no inherent tendency to change prevails in the model which they constitute."

A.3 Zero-input stability

Three most useful definitions for the zero-input stability will be discussed in this section: stability in the sense of Lyapunov, asymptotic stability and geometric stability (which will be described in turn). Let us restrict our attention to the zero-input linear dynamic system governed by equation (A-1). All of these definitions commonly specify whether or not the state-related behaviour of this dynamic system relative to its particular solution (including its equilibrium state and moving equilibrium state) is stable over the interval $[t_0, \infty)$.

To begin with, we should note the fact that the stability of any particular solution to a non-homogeneous linear system can always be reduced to the stability of the zero equilibrium state of the corresponding homogeneous linear system. In our application, letting PS_t be a given particular solution to equation (A-1), i.e. $PS_{t+1} = a(t) \cdot PS_t + b(t)$ with the initial condition $PS_{t_0} = s_0$, and then using the change of variable technique such as the one-to-one transformation from X_t to a new variable $Y_t \equiv X_t - PS_t$ for any t, then we have the following equation: for all t, substituting $X_t = Y_t + PS_t$ into equation (A-1) yields $Y_{t+1} + PS_{t+1} = a(t) \cdot [Y_t + PS_t] + b(t)$ with the initial condition $Y_{t_0} + PS_{t_0} = x_0$, which leads to

$$Y_{t+1} = a(t) \cdot Y_t \text{ with the initial condition } Y_{t_0} = y_0 \ (= x_0 - s_0). \qquad \text{- - - (A-2)}$$

So, the equilibrium state of the dynamic system governed by equation (A-2) (denoted by ye) is zero, i.e. ye=0. Here, we concentrate on the stability of this zero equilibrium state instead of the stability of a particular solution to equation (A-1) for the reasons that

firstly, dealing with the stability problem of zero equilibrium state is more convenient than dealing with that of any other particular solution because there are, in general, infinitely many particular solutions of equation (A-1), while the equilibrium state ye is the value of a particular solution of equation (A-2) and is uniquely given by a linear system [see properties P1 and P2 in A5.5.2];

secondly, the deviation of the definite solution to equation (A-1) about its particular solution PS_t, $X_t - PS_t$, is identical to the deviation of the definite solution of equation (A-2) about its zero equilibrium state, i.e. Y_t - ye; and then

lastly, it is valid that the stability of ye is equivalent to the stability of any particular solution to equation (A-1) (and hence speaking of the stability of equation (A-2) is equivalent to speaking of the stability of equation (A-1)).

Definition 2 (Stability in the sense of Lyapunov): The zero equilibrium state ye is said to be stable (or attracting) in the sense of Lyapunov iff, for any given t_0 and any positive number ε, there exists a positive number $\delta(\varepsilon, t_0)$ which may depend on ε and possibly the initial time t_0 as well, such that $|y_0| < \delta(\varepsilon, t_0)$ implies $|Y_t| < \varepsilon$ for all t. Otherwise, ye is said to be unstable (or repelling).

Comment 2: the above Definition 2 means that if a dynamic system is initially perturbed within an δ-open interval with centre at ye, then all subsequent state-related behaviour $\{Y_t;\ t=t_0+1, t_0+2, ...\}$ stays within an ε-open interval with centre at ye. However, this does not say that $\{Y_t;\ t = t_0+1, t_0+2, ...\}$ tends to ye after a small perturbation from ye and further there is no restriction on the size of the ε-open interval (and hence ε can be taken to be considerably larger than δ). From these points of view, this definition provides a very weak concept of zero-input stability.

Definition 3 (asymptotic stability): the zero equilibrium state ye is said to be asymptotically stable iff (a) ye is stable in the sense of Lyapunov and (b) ye is convergent (i.e. there exists a positive number $\beta(t_0)$ for all t_0, whose value may depend on t_0, such that $|y_0| < \beta(t_0)$ implies $\lim_{t\to\infty} Y_t = 0$).

Comment 3: Besides the stability in the sense of Lyapunov, asymptotic stability has the additional property that all subsequent state-related behaviour $\{Y_t:\ t=t_0+1, t_0+2, ...\}$ converges eventually to the zero equilibrium state ye after an initial small perturbation $\beta(t_0)$. Further, another point to be noted is that the condition (b) dominates the condition (a) because any subsequent state-related behaviour on any finite period, say $\{Y_t:\ t=t_0+m, t_0+m+1, \ldots, t_0+n$, with m and n positive integers and $m < n\}$, is finite whenever the initial perturbation is finite.

Definition 4 (geometric stability): The zero equilibrium state ye is geometrically stable iff there exist $\zeta \in (0, 1]$ and $h > 0$ such that for all $(t_0, y_0) \in R^1 \times R^1$, $|Y_t| \leq h \cdot (1-\zeta)^{t - t_0}$ for all $t \geq t_0$.

Comment 4: This definition states that the subsequent state-related behaviour $\{Y_t:\ t=t_0+1, t_0+2, \ldots\}$ should be characterised by decreasing geometric progression with a constant rate ζ, i.e.

ζ^{t-t_0} depending only on the elapsed time $t-t_0$: in other words, a uniform boundedness with the bound h and a uniform convergence to ye, no matter what the initial time t_0. This fact illustrates not only the aptness of the name 'geometric stability' but also a stronger form of zero-input stability than the other two described above. The constant rate ζ is sometimes referred to as geometric damping rate. Another point to be noted is that $(1-\zeta) \in [0, 1)$ can be transformed into an exponential form: in other words, there exists a positive rate ρ such that $(1-\zeta) = \exp(-\rho)$, in which ρ is sometimes referred to as the exponential damping rate. For this reason, geometric stability is normally called exponential stability but in our study, we shall use the term 'geometric' rather than the term 'exponential' on the grounds that the former term is the discrete-time analogue of the latter term (which is associated with a continuous-time domain) and we emphasise that we are dealing with the stability problem on a discrete-time domain.

A.4 Summary

We have explored the concepts of the equilibrium state and the zero-input stability, associated with the state-related motion of a zero-input first-order linear difference equation (A-1) (- this system equation corresponds to a general form of the zero-input, 100%-target solvency level growth equation (3.25) derived by using the given spread funding formula (3.22) in section 3.4.4). As noted in A5.5.1, the equilibrium state xe is the value of a constant solution (as a particular solution) to equation (A-1) irrespective of a given initial value x_0 at an initial time t_0: in particular, if $xe = x_0$, then the constant solution is called the equilibrium solution, i.e. $X_t = xe$ constant for all t.

If continuous improvement in the stability of a pension scheme is its management goal, the preference ranking among the stability concepts described in A5.5.2 is clearly 'stability in the sense of Lyapunov' $\prec$ 'asymptotic stability' $\prec$ 'geometric stability'

where the preference notation $\prec$ implies that the right-hand side is preferred to the left-hand side.

Therefore, we may classify a pension scheme into four different groups in accordance with the above stability preference relation:

- the ideal pension scheme would be a pension scheme under control characterised by the qualitative properties of geometric stability;
- the quasi-ideal pension scheme would be a pension scheme under control characterised by the qualitative properties of asymptotic stability;
- the threshold pension scheme would be a pension scheme under control characterised by the qualitative properties of stability in the sense of Lyapunov; and
- the chaotic pension scheme would be a pension scheme out of control (that is, an unstable pension scheme).

Lastly, we note that in the application of the spread funding plan (e.g. a funding plan governed by the spread funding formula (3.22) in section 3.4.4), the stability of a pension scheme is an essential consideration of a long-term pension fund valuation, together with the scheme's security. This is in view of the fact that designing the spread funding plan in the direction of reducing the contribution rate risk has much to do with the state-related stability of the pension scheme because the spread funding formula can be expressed as a linear function of the state variable [see paragraph (b) of section 5.1].

Appendix 5B.1: Solutions of Biquadratic Equation (5.13) [see, Upensky (1948, Ch. 5)]

Applying Ferrari's method, the biquadratic equation (5.13) has the following roots $f_j(k)$, j=1, 2, 3 and 4 (note that any real-valued $f_j(k)$'s are applied to Procedures 2 or 3 established in section 5.3.4.1):

$$\{f_j(k); j=1, 2, 3, 4\} = \{[h_1(y) \pm \surd(h_1(y)^2 + 8h_2(y))] / 4, \ [h_3(y) \pm \surd(h_3(y)^2 + 8h_4(y))] / 4\}$$

where

$h_1(y) = \surd(c_1^2 - 4c_2 + 4y) - c_1$, $h_2(y) = \surd(y^2 - 4c_4) - y$, $h_3(y) = -\surd(c_1^2 - 4c_2 + 4y) - c_1$ and $h_4(y) = -\surd(y^2 - 4c_4) - y$, in which $h_j(y)$ is a function of any root y satisfying the following cubic resolvent of the biquadratic equation (5.12), that is,

$$y^3 - c_2y^2 + (c_1c_3 - 4c_4)y + (4c_2c_4 - c_1^2c_4 - c_3^2) = 0.$$

This cubic equation produces roots, i.e. y_1, y_2 and y_3, by Cardan's method, that is,

$y_1 = c_2/3 + u_1 + u_2$, y_2 and $y_3 = c_2/3 - (u_1 + u_2)/2 \pm \frac{\sqrt{3}}{2} i (u_1 - u_2)$, in which

$i = \sqrt{-1}$, u_1 and $u_2 = [c_2^3/27 + (c_1^2c_4 + c_3^2)/2 - (c_1c_2c_3 + 8c_2c_4)/6 \pm \frac{1}{6\sqrt{3}}\sqrt{w}\ ]^{1/3}$ and

$$w = (72c_2c_4 + 9c_1c_2c_3 - 27c_1^2c_4 - 27c_3^2 - 2c_2^3)/27 + \frac{4}{27}(3c_1c_3 - 12c_4 - c_2^2)^3.$$

We note finally that w will be positive, zero or negative: if $w \geq 0$, then y_1 is real-valued but y_2 and y_3 are complex conjugates, so y_1 is chosen for y and then the roots $f_j(k)$ each are simply computable; on the contrary, if $w < 0$, then y_1, y_2 and y_3 are all real-valued, so y_1, y_2 and y_3 each can be chosen for y and then the roots $f_j(k)$ are to be calculated for y_1, y_2 and y_3 each.

Appendix 5B.2: Numerical illustrations for stationary LQP optimisation problem (5.8)

A1. (Comparison with respect to δ, θ and FR_0)

Table 5.1

Optimal value of k (denoted by k^*)

θ	$\exp(\delta) = \exp(\eta) = 1.06$			$\exp(\delta) = 1.08$, $\exp(\eta) = 1.06$		
	$FR_0 = 1.0$	$FR_0 = 0.5$	$FR_0 = 0.0$	$FR_0 = 1.0$	$FR_0 = 0.5$	$FR_0 = 0.0$
$100\%^-$	1.0^-	1.0^-	1.0^-	1.0^-	1.0^-	1.0^-
90%	1.0	0.956498	0.956477	1.0	0.961861	0.961662
80%	1.0	0.932877	0.932806	1.0	0.940839	0.940506
70%	1.0	0.920467	0.920332	1.0	0.929514	0.929088
60%	1.0	0.914384	0.914181	1.0	0.923643	0.923156
50%	1.0	0.911622	0.911356	1.0	0.920554	0.920032
40%	1.0	0.910190	0.909864	1.0	0.918482	0.917944
30%	1.0	0.908880	0.908496	1.0	0.916453	0.915903
20%	1.0	0.907226	0.906781	1.0	0.914183	0.913611
10%	1.0	0.905243	0.904733	1.0	0.911741	0.911134
$0\%^+$	0.056603^+	0.056603^+	0.056603^+	0.056603^+	0.056603^+	0.056603^+

Note that x^- and x^+ for a real number x are approximately equal to x but are smaller and larger than x, respectively; here, $0.056603^+ \cong 1 - \exp(-\eta)$.

Appendix 5B.3: Diagrammatical illustrations for computational infeasiblity of quasi-stationary LQP optimisation problem (5.22) subject to [Constraints set I]

Decision tree

(Basic structure for the FDP approach)

Stage 0: ${}^{Q}J_0({}^{Q}FR_t-1, t) = 0$ (given starting condition)

↓

Stage 1: ${}^{Q}J_1({}^{Q}FR_t-1, t)$

↓

$k_1({}^{Q}FR_t-1)$

↓

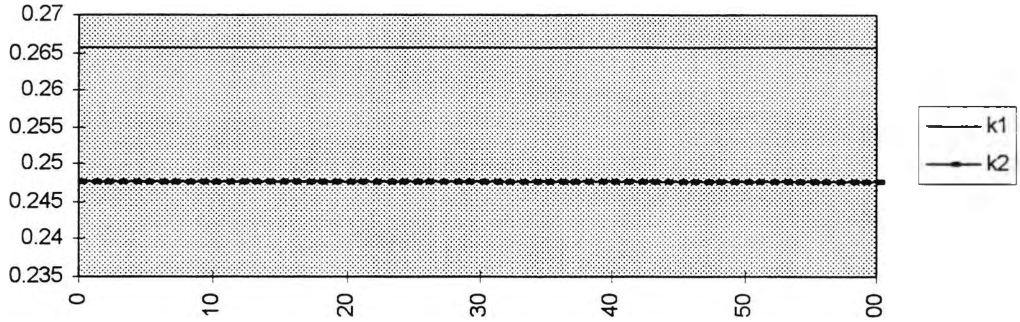
	Event1	Event2	Event3
	↓	↓	↓
Stage 2:	${}^{Q}J_2({}^{Q}FR_t-1, t;$ Event1)	${}^{Q}J_2({}^{Q}FR_t-1, t;$ Event2)	${}^{Q}J_2({}^{Q}FR_t-1, t;$ Event3)
	↓	↓	↓
	$k_2({}^{Q}FR_t-1;$ Event1)	$k_2({}^{Q}FR_t-1;$ Event2)	$k_2({}^{Q}FR_t-1;$ Event3)
	↓	↓	↓
	Event1 Event2 Event3	Event1 Event2 Event3	Event1 Event2 Event3
	↓ ↓ ↓	↓ ↓ ↓	↓ ↓ ↓

Note that three disjoint and exhaustive events each are given by Event1 = $\{kk: 0 \leq kk \leq 1\}$, Event2 = $\{kk: kk < 0\}$ and Event3 = $\{kk: kk > 1\}$, and further, the above decision tree demonstrates that the number of possible decisions at stage n, i.e. $k_n({}^{Q}FR_t-1;$ Event j) where j = 1, 2 or 3 and n = 1, 2, 3, ..., is up to 3^{n-1} (and hence leading to a significantly intractable problem of computational dimensionality as $n \rightarrow \infty$).

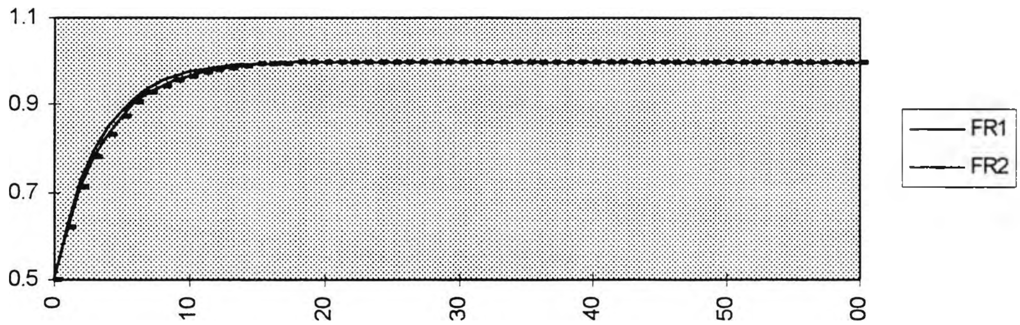
Appendix 5B.4: Numerical illustrations for quasi-stationary LQP optimisation problem (5.22) subject to [Constraints set II]

Graph 5.1

Projections of optimal quasi-stationary spread funding plan governed by formula (5.33) under the situation of the most likely expected investment performance (i.e. $\delta = \eta$)

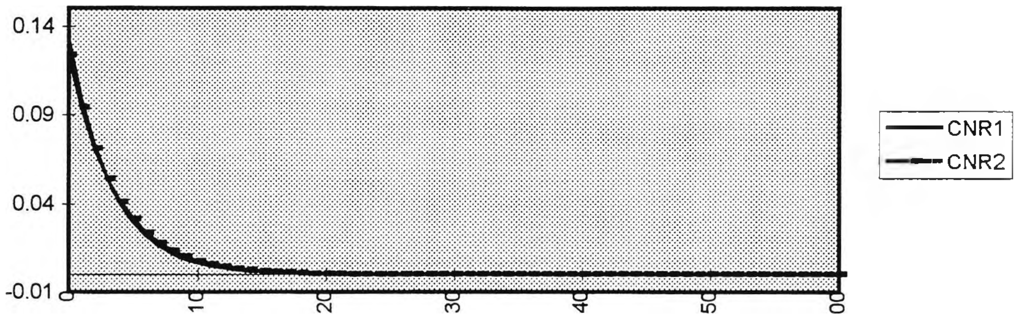


Graph 5.1.1. The time path of optimal spread controls: k1 = $\{k^*(^QH_t) = 0.265734$; t = 0, 1, 2, ... under θ = 10%} and k2 = $\{k^*(^QH_t) = 0.247557$; t = 0, 1, 2, ... under θ = 1%}.



Graph 5.1.2. The time path of optimal funding levels: FR1 = $\{FR^*_t$; t = 0, 1, 2, ... under θ = 10%} and FR2 = $\{FR^*_t$; t = 0, 1, 2, ... under θ = 1%}.

Notice that $FR^*_{t\geq 47} = 1^-$ for θ = 10%; and $FR^*_{t\geq 51} = 1^-$ for θ = 1%, where 1^- denotes to be approximately equal to 1 but smaller than 1.

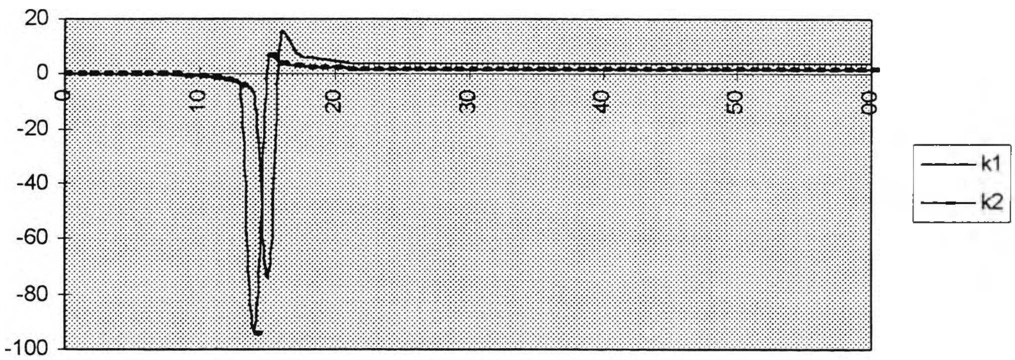


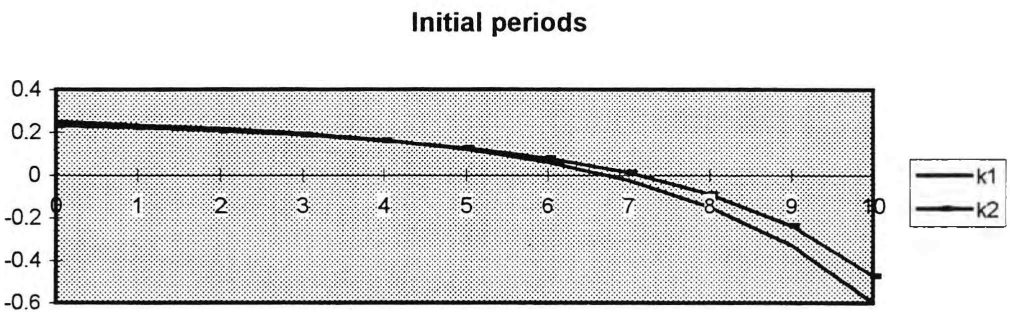
Graph 5.1.3. The time path of optimal control action errors: CNR1 = $\{{}^{Q}CR^{f*}_{t} - {}^{Q}NR^{f}_{t};\ t = 0, 1, 2, \ldots$ under $\theta = 10\%\}$ and CNR2 = $\{{}^{Q}CR^{f*}_{t} - {}^{Q}NR^{f}_{t};\ t = 0, 1, 2, \ldots$ under $\theta = 1\%\}$.

Notice that ${}^{Q}CR^{f*}_{t\geq 47t} - {}^{Q}NR^{f}_{t\geq 47} = 0^{+}$ for $\theta = 10\%$; and ${}^{Q}CR^{f*}_{t\geq 51} - {}^{Q}NR^{f}_{t\geq 51} = 0^{+}$ for $\theta = 1\%$, where 0^{+} denotes to be approximately equal to 0 but larger than 0.

Graph 5.2

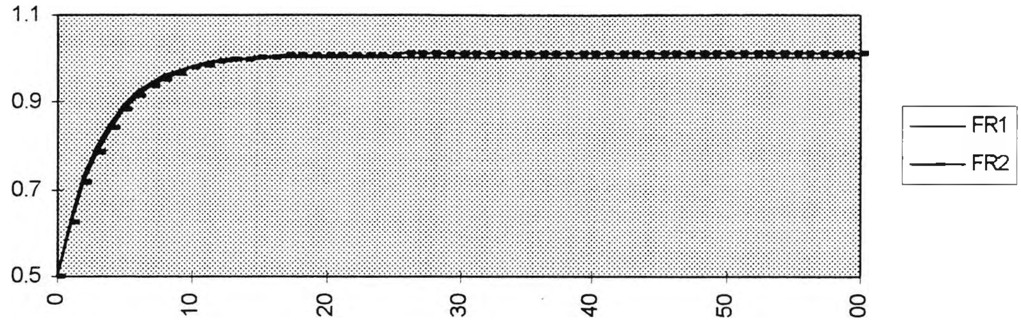
Projections of optimal quasi-stationary spread funding plan (5.33) under the better investment performance than most likely expected (i.e. $\delta > \eta$)





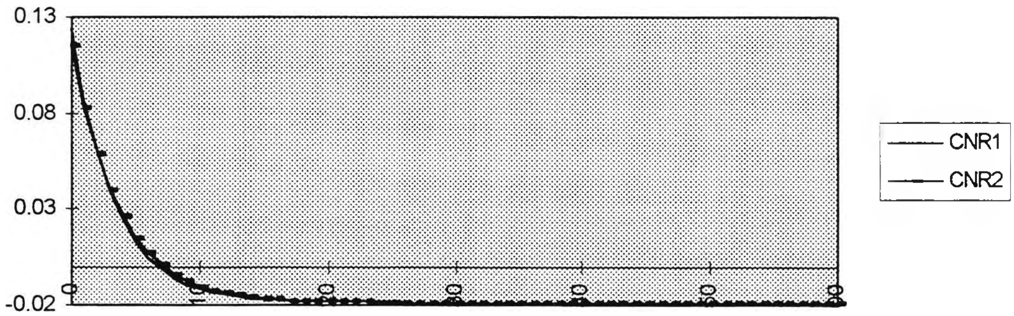
Graph 5.2.1. The time path of optimal spread controls: k1 = $\{k^*({}^QH_t)$; $t = 0, 1, 2, \ldots$ under $\theta = 10\%\}$ and k2 = $\{k^*({}^QH_t)$; $t = 0, 1, 2, \ldots$ under $\theta = 10\%\}$.

Notice that for $\theta = 10\%$, $k^*({}^QH_{t=15}) = -74.685113$, $k^*({}^QH_{t=16}) = 14.659166$ and $k^*({}^QH_{t\geq 69}) = 3.567606$; and for $\theta = 1\%$, $k^*({}^QH_{t=14}) = -94.099515$, $k^*({}^QH_{t=15}) = 6.885535$ and $k^*({}^QH_{t\geq 66}) = 1.781531$.



Graph 5.2.2. The time path of optimal funding levels: FR1 = $\{FR^*_t; t = 0, 1, 2, \ldots$ under $\theta = 10\%\}$ and FR2 = $\{FR^*_t; t = 0, 1, 2, \ldots$ under $\theta = 1\%\}$.

Notice that for $\theta = 10\%$, $FR^*_{t=15} = 0.999773$, $FR^*_{t=16} = 1.001185$ and $FR^*_{t\geq46} = 1.005184$; and for $\theta = 1\%$, $FR^*_{t=14} = 0.999832$, $FR^*_{t=15} = 1.002399$ and $^{Q}SR^*_{t\geq48} = 1.010462$.



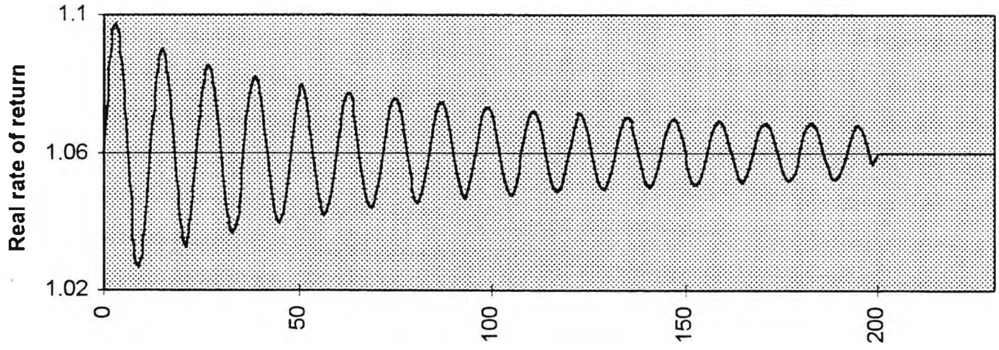
Graph 5.2.3. The time path of optimal control action errors: CNR1 = $\{^{Q}CR^{f*}_t - {}^{Q}NR^f_t; t = 0, 1, 2, \ldots$ under $\theta = 10\%\}$ and CNR2 = $\{^{Q}CR^{f*}_t - {}^{Q}NR^f_t; t = 0, 1, 2, \ldots$ under $\theta = 1\%\}$.

Notice that for $\theta = 10\%$, $^{Q}CR^{f*}_{t=6} - {}^{Q}NR^f_{t=6} = 0.004639$, $^{Q}CR^{f*}_{t=7} - {}^{Q}NR^f_{t=7} = -0.001399$ and $^{Q}CR^{f*}_{t\geq32} - {}^{Q}NR^f_{t\geq32} = -0.018487$; and for $\theta = 1\%$, $^{Q}CR^{f*}_{t=7} - {}^{Q}NR^f_{t=7} = 0.000679$, $^{Q}CR^{f*}_{t=8} - {}^{Q}NR^f_{t=8} = -0.003988$ and $^{Q}CR^{f*}_{t\geq38} - {}^{Q}NR^f_{t\geq38} = -0.018635$.

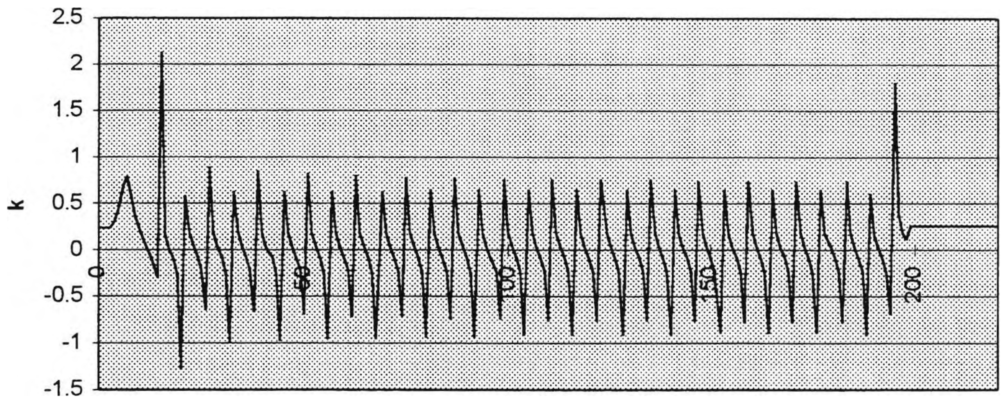
Appendix 5B.5: Numerical illustrations for threshold LQP optimisation problem (5.50)

Graph 5.3

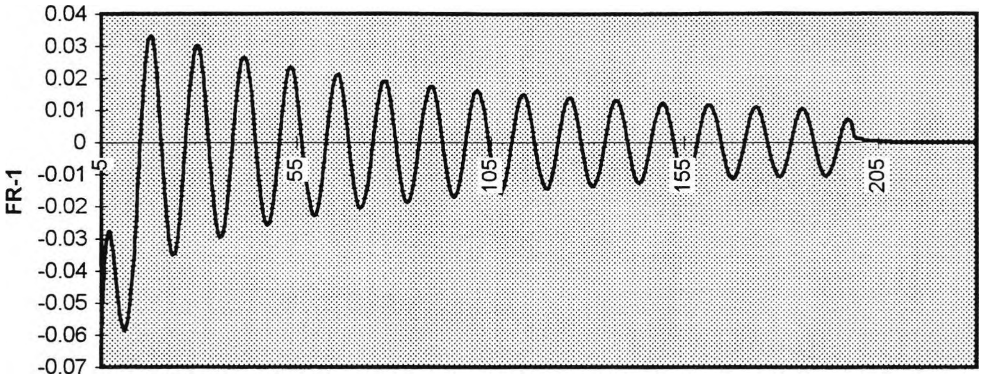
Projections of optimal threshold spread funding plan (5.61)



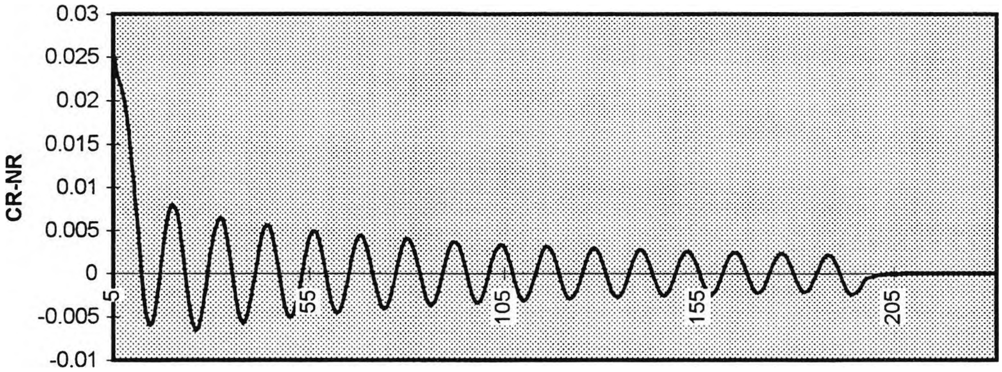
Graph 5.3.1. The time path of projected real rates of return



Graph 5.3.2. The time path of optimal spread controls



Graph 5.3.3. The time path of optimal control errors: ($^{Q}FR^{*}_{t}$-1; t=0, 1, ..., 5) = (-0.5, -0.367567, -0.260195, -0.173442, -0.106452, -0.060117).



Graph 5.3.4. The time path of optimal control action errors: ($^{Q}CR^{f*}_{t}$-$^{Q}NR^{f}_{t}$; t=0, 1, ..., 5) = (0.124376, 0.087490, 0.060336, 0.042133, 0.031298, 0.025711).

Chapter 6 Conclusions

In the thesis, we have applied the optimisation instruments/tools of optimal control theory, especially the method of dynamic programming (DP), in order to find a sensible pension funding plan that can be implemented for defined benefit pension schemes. The distinctiveness of the thesis arises from its demonstration of the usefulness and applicability of the control-theoretical approach to this problem by way of matching conceptually and mathematically a general framework for optimal control theory with a general mechanism for the funding of defined benefit pension schemes. We shall briefly review the main concepts and/or results of each of the previous Chapters 1~5 together with some suggestions for future research as an extension to the thesis.

In Chapter 1, we have asserted that the sponsoring employer and trustees of a defined benefit pension scheme have conflicting viewpoints on its pension funding plan due to their respective best interests, i.e. stability and security.

Chapter 2 has discussed how to define mathematically the solvency risk (as a measure for the security concept) and contribution rate risk (as a measure for the stability concept) and how to bring the concepts of optimal control theory into the field of the funding of defined benefit pension schemes. According to these discussions, we introduce and define the dynamic pension funding plan as a feedback funding mechanism characterised by controlling sequentially and optimally the solvency and contribution rate risks at the same time and adjusting these two risks, as time progresses, by means of processing the available information, updated with time.

In Chapter 3, we have constructed several linear dynamic system models representing a reduced structural model of the real financial structure of defined benefit pension schemes; in particular, the modified solvency-level growth equation and zero-input, 100%-target solvency-level growth equation These equations are treated as a dynamic system equation governing a controlled object.

In Chapter 4, we have formulated three distinct finite-horizon control optimisation problems, designed for a short-term, winding-up valuation with contribution rates unconstrained by any pre-determined funding plans - deterministic, stochastic with complete state information and stochastic with incomplete state information (note that the incomplete state information case allows for a one-unit time delay in the availability of current state information, reflecting, for example, the time needed for the background accounting and auditing work). Their respective solutions (derived by using the DP approach) are defined as our dynamic pension funding plan. A large portion of this chapter is devoted to the mathematical comparison between the dynamic pension funding plan and the spread funding plan and to suggestions for reducing the risk of insolvency. Thus, we reach the following conclusions:

(a) We believe that the following funding formula (specifying the dynamic pension funding plan) is suitable for balancing the conflicting interests of the employer and trustees:

$$DC_t = NC_t + pc_t \cdot UV_t + ac_t$$

where

DC_t = dynamic contribution rate applying between time t and t+1 (i.e. recommended contribution rate provided by the application of optimal control theory to pension funding);

NC_t = normal cost applying between time t and t+1 (i.e. a regular cost provided by a chosen primary funding method, such as Projected Unit or Entry Age methods);

UV_t = undesirable valuation outcome at time t (e.g. SL_t-F_t for a solvency valuation, i.e. short-term, winding-up valuation, or AL_t-F_t for a classical actuarial valuation, i.e. long-term and going-concern valuation);

pc_t = proportional controlling parameter applying between t and t+1 $\in$ (0, 1); and

ac_t = additive controlling parameter applying between t and t+1.

Here, we note that the proportional and additive terms are specified by the optimal funding control procedure and the additive term is distinct from the spread funding formula and hence this formula can be thought of as a dynamic version of the spread funding formula.

(b) In order to strengthen the confidence in the financial soundness of defined benefit pension schemes, the actuary should employ one or a combination of the three distinct options - placing more emphasis on the solvency risk than on the contribution rate risk, increasing the solvency or funding target and using a more conservative valuation basis than the best estimate valuation basis.

(c) The performance comparison measure, based on the concept of mean-squared error, seems to be a useful measure for comparing the incomplete state information control problem with the corresponding complete state information control problem.

Chapter 5 is mainly devoted to investigating how the spread funding plan is optimally designed with respect to the spread parameter in the application of optimal control theory. Different to Chapter 4 (considering the situation of a short-term, winding-up valuation), this chapter considers the situation of a long-term, going-concern valuation from the viewpoint that the spread funding plan is designed and adapted primarily for dealing with such a valuation. In this respect, we have formulated four distinct infinite-horizon deterministic control optimisation problems, designed for a long-term, going-concern valuation under contribution rates constrained by the spread funding plan - stationary, quasi-stationary, non-stationary and threshold. We note that the formulation of the infinite-horizon control problems may appear unrealistic but has been made on grounds that dealing with infinite-horizon control problems could provide an analytically convenient approximation for optimising the value of the spread parameter under finite but long-term control problems. The main concepts and results are as follows:

(a) The stationary, quasi-stationary, non-stationary and threshold control problems each are characterised by how to treat the spread parameter k_t. Thus, denoting the currently available information vector by H_t, then k_t is considered to be a constant function in the stationary control problem (i.e. $k_t = k$), a time-invariant function of H_t in the quasi-stationary problem (i.e. $k_t = k(H_t)$), a time-varying function of H_t in the non-stationary control problem (i.e. $k_t = k_t(H_t)$) and a function of H_t switching from time-varying over an initial finite period Regime1, to time-invariant over the last infinite period Regime2, in the threshold control problem (i.e. $k_t = k_t(H_t) \cdot 1_{Regime1} + k(H_t) \cdot 1_{Regime2}$.

(b) The stationary control problem is soluble by means of an algebraic approach, while the others are soluble by means of the DP approach, in particular the forward DP approach.

(c) For a systematic and unique optimisation procedure with respect to the spread parameters (except for the stationary control problem), it is necessary to set up a rule suitable for dealing with the case of encountering 100% funding levels over the control horizon - this rule is termed the possible spread rule: that is, if the funding level is 100% at some time in a process of sequential optimisation, then we set the optimal value of the corresponding spread parameter to be zero. This is needed because the spread funding formula has a weakness in its mathematical formulation for the case of a 100% funding level in the light of optimal control theory.

(d) The larger the number of strong boundary constraints on the spread parameter and the longer the control horizon, the higher is the potential risk that the sequential optimisation procedure can not be carried on (owing to the problem of computational dimensionality).

(e) Due to the computational insolubility of the non-stationary control problem; the threshold control problem is suggested as a best approximate alternative.

(f) Finally, a possible extension of the typical spread parameter space $\{k_t: 0 \le k_t \le 1\}$ to, say, $\{k_t: -1 \le k_t \le 2\}$ is suggested in order to reflect somewhat the unpredictable (optimistic or pessimistic) nature of the future financial status of the scheme. This is believed to lead to an improvement in the stability and security of the spread funding plan, especially in a more realistic environment allowing for dynamic economic and demographic changes.

I would like to finish the thesis by restating some interesting and useful subjects for future research as possible extensions to the thesis, which have been mentioned in the thesis, except for (e):

(a) We have not sufficiently/satisfactorily discussed how to design (not necessary optimally) the supplementary performance criterion as an additional means for improving the performance of the primary performance criterion in the light of the pace of funding and/or the progress of solvency (or funding) levels in relation to their respective target values. This subject would involve a variety of numerical and/or mathematical comparisons between all suitable types of supplementary performance criteria (as noted in section 4.2.3.2).

(b) We have suggested three distinct ways of reducing the potential risk of insolvency but there will be some other available options. So, we need to analyse each of these options in tandem with all possible combinations in order to set a best (not necessary optimal) policy for this problem. Following the result (observed in Chapter 4) that an improved level of protection against insolvency requires an additional funding burden on the employer, the analysis should take into account the magnitude/volatility of the extra financial burden on the employer with respect to the improvement in the solvency of the scheme.

(c) All dynamic system models in the thesis are constructed on a simplified financial structure of the defined benefit pension scheme. In the real world, the financial structure of the scheme may be characterised by non-linearity and complexity. In this respect, the stochastic models for investment returns and benefit outgoes represented by IID normal random variables need to be modified by using, say, ARIMA type introduced by Box and Jenkins (1976) (as in Haberman (1994)) and then some rigorous approximation procedure is required for solving our stochastic control problem due to its insolubility for a general ARIMA model (as mentioned in 4.4.1).

(d) It would be possible to generalise our incomplete state information control problem with one-unit time delay by way of introducing a b-unit time delay ($b > 1$) (as in Zimbidis & Haberman (1993)).

(e) Further, it would be possible to generalise the frequency of valuations with the inter-valuation period being n time units, n being an positive integer (as in Haberman (1993)) - note that in the thesis we have assumed that valuations are performed every one time unit.

(f) Finally, it would be possible to allow specifically for MFR type minimum rules with liabilities calculated on a prescribed basis different from that used for a going-concern valuation basis (as discussed in Greenwood & Keogh (1997)).

References

Anderson, A.W. (1992). "Pension Mathematics for Actuaries", 2nd Edition, ACTEX Publications, Inc., Connecticut.

Astrom, K.J. (1970). "Introduction to stochastic control theory", Academic Press, New York.

Bellman, R. (1957). "Dynamic Programming", Princeton University Press, Princeton.

Bellman, R. (1961). "Adaptive control processes", Princeton University Press, Princeton.

Benjamin, S. (1984). "An actuarial layman looks at control theory", Transactions of 22nd International Congress of Actuaries, Sydney 1984, vol.4, 295-310.

Benjamin, S. (1989), "Driving the pension fund", Journal of the Institute of Actuaries 116, 717-735.

Bertsekas, D.P. (1976). "Dynamic Programming and Stochastic Control", Academic Press, New York.

Bereskas, D.P. (1987). "Dynamic programming: Deterministic and Stochastic Models", Prentice-Hall, Inc., Englewood Cliffs.

Berger, J.O. (1985). "Statistical Decision Theory and Bayesian Analysis", 2nd Edition, Springer-Verlag, New York.

Blake, D. (1992a). "Issues in Pension Funding", Routledge, London.

Blake, D. (1992b). "Modelling Pension Fund Investment Behaviour", Routledge, London.

Bowers, N.L., Gerber, H.U., Hickman, J.C. & Nesbitt, C.J. (1986). "Actuarial Mathematics", The Society of Actuaries, Illinois.

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.

Box, G.E.P. & Jenkins, G.M. (1976). "Time Series Analysis", 2nd Edition, Holden-Day, San Francisco.

Buhlmann, H. (1992). "Stochastic discounting", Insurance: Mathematics and Economics 11, 113-127.

Bulmer, R. (1994). "Balancing risk and reward", The Actuary, November, 1994, 14-16.

Callier, F.M. & Desoer, C.A. (1991). "Linear System Theory", Springer-Verlag, New York.

Casey, B. (1993). "Employers' Choice of Pension Schemes: Report of a qualitative study", Department of Social Security Research Report No.17, HMSO, London.

Clifford Chance (1993). "Pensions in Europe", Clifford Chance, London.

Collins, A. (1992). "Funding and Solvency Levels for Pension Schemes", Paper presented to the Staple Inn Actuarial Society, London.

Dardis, A. & Huynh, V.L. (1995). "Bridging the Gap Between Financial Economists and Actuaries: A New Asset-Liability Management Model", Paper presented to the Staple Inn Actuarial Society, London.

Daykin, C.D., Bernstein, G.D., Coutts, S.D., Devitt, E.R.F., Hey, G.B., Reynolds, D.I.W. & Smith, P.D. (1987). "Assessing the solvency and financial strength of a general insurance company", Journal of the Institute of Actuaries 114, 227-310.

Daykin, C.D., Pentikainen, T. & Pesonen, M. (1994). "Practical Risk Theory for Actuaries", Chapman & Hall, London.

Dilnot, A., Disney, R., Johnson, P. & Whitehouse, E. (1994). "Pensions Policy in the UK: An Economic Analysis", The Institute for Fiscal Studies, London

Dorf, R.C. (1992). "Modern Control Systems", 6th Edition, Addison-Wesley, Reading.

Dufresne, D. (1986). "The Dynamics of Pension Funding", Ph.D Thesis, The City University, London

Dufresne, D. (1988). "Movements of pension fund contributions and fund levels when rates of return are random", Journal of 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. (1992). "Distributions of discounted values", Proceedings of the 26th Annual Actuarial Research Conference, 11-25, Illinois.

Fujiki, M.H. (1994). "Pension Fund Valuation", Ph. D. Thesis, The City University, London.

Government Actuary's National Insurance Fund Long Term Estimates (1990), HMSO, London.

Greenwood, P.M. & Keogh, T.W. (1997). "Pension funding and expensing in the minimum funding requirement environment", British Actuarial Journal (forthcoming)

Grimmett, G.R. & Stirzaker, D.R. (1992). "Probability and Random Processes", 2nd Edition, Clarendon Press, Oxford.

Haberman, S. (1990). "A stochastic approach to pension funding methods", Proceedings of 1st AFIR Colloquium Vol.4, 92-112, Paris.

Haberman, S. (1992). "Pension funding with time delays - A stochastic approach", Insurance: Mathematics & Economics 11, 179-189.

Haberman, S. (1993). "Pension funding: The effect of changing the frequency of valuations", Insurance: Mathematics & Economics 13, 263-270.

Haberman, S. (1994). "Pension funding modelling and stochastic investment returns", Paper presented to Joint Meeting of the Staple Inn Actuarial Society and the Royal Statistical Society, General Applications Section, London.

Haberman, S. (1997). "Stochastic investment returns and contribution rate risk in a defined benefit pension scheme", Insurance: Mathematics & Economics. To appear.

Haberman, S. & Sung, J.H. (1994). "Dynamic approaches to pension funding", Insurance: Mathematics and Economics 15, 151-162.

Harrison, D. (1995). "Pension Power", John Wiley & Sons, Chichester

Harvey, A. (1990). "The Econometric Analysis of Time Series", 2nd Edition, Philip Allan, New York.

Harvey, J. (1991). "Intermediate Economics", 5th Edition, Macmillan, London.

Hiller, F.S. & Lieberman, G.J. (1980). "Introduction to Operations Research", 3rd Edition, Holden-Day, San Francisco.

Houston, D.B. (1964). "Risk, insurance, and sampling", Journal of Risk and Insurance, Vol.31, No.4, 511-538.

Hull, J.C. (1993). "Options, Futures, and Other Derivative Securities", Prentice-Hall, Inc., London

Humphrys, R. (1984). "Standards and solvency requirements under Canadian insurance legislation", Transactions of the 22nd International Congress of Actuaries Vol.3, 489-497, Sydney.

Institute and Faculty of Actuaries (1984). "Terminology of Pension Funding Methods (TPFM)", Report of A Working Party of the Pension Standards Joint Committee, The Institute and Faculty of Actuaries, London.

Institute and Faculty of Actuaries (1988). "Pension Fund Terminology: Specimen Descriptions of Commonly Used Valuation Methods", The Institute and Faculty of Actuaries, London.

Institute and Faculty of Actuaries (1994). "Appraisal of Capital Projections", The Institute and Faculty of Actuaries, London.

Institute and Faculty of Actuaries Guidance Note (1996). "Pension Fund Terminology (GN26)", The Institute and Faculty of Actuaries, London.

Jacobs, O.L.R. (1993). "Introduction to Control Theory", 2nd Edition, Oxford University Press, Oxford

Janssen, J. (1994). "A dynamic stochastic ALM model for insurance companies", Proceedings of ASTIN Colloquium, 90-114, Cannes.

Joshi, H. (ed.) (1989). "The Changing Population of Britain", Basil Blackwell, Oxford.

Joseph, P.D. & J.T. Tou (1961). "On linear control theory", AIEE Transactions 80, 193-196.

Kemp, M. (1996). "Asset/Liability modelling for pension funds", Paper presented to the Staple Inn Actuarial Society, London.

Khorasanee, M.Z. (1993). "A comparison of pension funding strategies by means of simulations for a model scheme", Actuarial Research Paper No.49, The City University, London.

Lee, E.M. (1986). "An Introduction to Pension Schemes", The Institute and Faculty of Actuaries, London.

Lewin, C.G., Carne, S.A., De Rivaz, N.F.C., Hall, R.E.G., McKelvey, K.J. & Wilkie, A.D. (1995). "Capital Projections", British Actuarial Journal 1, 155-330.

Loades, D. H. (1992). "Instability in pension funding", Transactions of 24th International Congress of Actuaries Vol.2, 137-154.

London, D. (1985). "Graduation - the revision of estimates", ACTEX Publications, Inc., Connecticut.

Lumby, S. (1988). "Investment Appraisal and Financing Decisions", 3rd Edition, VNR, London.

Lyapunov, M.A. (1892). "Probleme general de la stabilite du mouvement", (Originally published in Russian, Kharkov, 1892: Reprinted in Annuals of Mathematical Studies 17, 1947, Princeton University Press, Princeton).

Lynch, J.M. (1979). "Identifying actuarial gains and losses in pension plans", Proceedings of the Conference of Actuaries in Public Practice 26, 582-594.

Machlup, F. (1958). "Equilibrium and Disequilibrium: Misplaced con", The Economic Journal, March, 1958, 1-24.

Marshall, D.R. & Reeve, J.G. (1993). "Defined Benefit Pension Schemes Funding for Ongoing Security", Paper presented at the Staple Inn Actuarial Society Meeting, London

McGillem, C.D. & Cooper, G.R. (1991). "Continuous and Discrete Signal and System Analysis", 3rd Edition, Saunders College Publishing, Philadelphia.

Mcleish, D.J.D. & Stewart, C.M.(1987). "Objectives and Methods of Funding Defined Benefit Pension Schemes", Journal of Institutes of Actuaries 114, 155-199.

Mehta, S.J.B. (1992). "Allowing for asset, liability and business risk in the valuation of a life office", Journal of the Institute of Actuaries 119, 385-455.

Mortensen, J. (ed.) (1992). "The Future of Pensions in the European Community", Brassey's, London.

Norberg, R. (1995). "Differential equations for moments of present values in life insurance", Insurance: Mathematics and Economics 17, 171-180.

O'Brien, T. (1984). "Projection valuation vs. Aggregate funding - Simulation", ARCH, 65-104.

O'Brien, T. (1987). "A two-parameter family of pension contribution functions and stochastic optimization", Insurance: Mathematics and Economics 6, 129-134.

Owadally, M.I. & Haberman, S. (1995). "Stochastic investment modelling and optimal pension funding strategies", Actuarial Research Paper No.76, The City University, London

Panjer, H. H. & D.R. Bellhouse, (1980). "Stochastic modelling of interest rates with applications to life contingencies", Journal of Risk and Insurance 47, 91-110.

Pension Law Review Committee (PLRC) (1993). "Pension Law Reform", Vol.I, HMSO, London.

Phillips and Drew Fund Management (1991). "Pension Fund Indicators - A Long-Term Perspective on Pension Fund Investment 1991", Phillips and Drew Fund Management, London.

PMI/PRAG (1992). "Pensions Terminology", 4th Edition, PMI/PRAG Publication, London.

Sandefur, J.T. (1990). "Discrete Dynamical System: Theory and Applications", Clarendon Press, Oxford

Secretary of State for Social security (1994). "Security, Equality, Choice: The Future for Pensions", Vol.I, HMSO, London.

Sienkiewicz, T. & Mascarenhas, A. (1993). "Accounts and Audit of Pension Schemes", Touche Ross & Co., London

Simon, H.A. (1956), "Dynamic programming under uncertainty with a quadratic criterion function", Econometrica 24, 74-81.

Smith, A.D. (1996). "How actuaries can use financial economics", British Actuarial Journal 2, 1057-1193.

Thornton, P.N. & Wilson, A.F. (1992). "A realistic approach to pension funding", Journal of Institute of Actuaries 119, 229-286.

Trowbridge, C.L. (1952). "Fundamentals of pension funding ", Transactions of Society of Actuaries Vol.4, 17-43.

Tignol, J.P. (1988). "Galois' theory of algebraic equations", Longman, London.

Tong, H. & Lim, K.S. (1980). "Threshold autoregression, limit cycles and cyclical data", Journal of Royal Statistical Society, Series B42, 245-292.

Vanderbroek, M. (1990). "Pension funding and optimal control", Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, 313-325.

Whittle, P. (1982). "Optimization over Time: Dynamic Programming and Stochastic Control", Vol.I, John Wiley & Sons, Chichester.

Whittle, P. (1983). "Optimization over Time: Dynamic Programming and Stochastic Control", Vol.II, John Wiley & Sons, Chichester.

Wilkie, A.D. (1986). "The myth of overfunding", The Financial Times Conference: Pensions in 1986, London, March, 1986.

Wilkie, A.D. (1995). "More on a stochastic asset model for actuarial use". British Actuarial Journal 1, 777-964.

Willems, J.L. (1970). "Stability Theory of Dynamical Systems", Thomas Nelson and Sons Ltd., London.

William, C.A. & Heins, R.M. (1971). "Risk Management & Insurance", McGraw-Hill Book Company, New York.

Zimbidis, A. & Haberman, S. (1993). "Delay, feedback and variability of pension contributions and fund levels". Insurance: Mathematics and Economics 13, 271-285.