B.A.J. **9**, IV, 903-958 (2003)

THE USE OF UTILITY FUNCTIONS FOR INVESTMENT CHANNEL CHOICE IN DEFINED CONTRIBUTION RETIREMENT FUNDS

II: A PROPOSED SYSTEM

By R. J. Thomson

ABSTRACT

In this paper a system for recommending investment channel choices to members of defined contribution retirement funds is proposed. The system is interactive, using a member's answers to a series of questions to derive a utility function. The observed values are interpolated by means of appropriate formulae to produce a smooth utility function over the whole positive range of benefits at retirement. The resulting function, together with stochastic models of the returns on the available channels and of the annuity factor at exit, is then used to recommend an optimum apportionment of the member's investment. The proposed system is applied to the observed values of utility functions of post-retirement income elicited from members of retirement funds. Difficulties in the application are discussed and the results are analysed. The sensitivity of the recommendations to the parameters of the stochastic model is discussed.

KEYWORDS

Expected Utility Theory; Decision Theory; Optimisation; Investment Channel Choice; Defined Contribution Retirement Funds; Dynamic Programming

CONTACT ADDRESS

R. J. Thomson, B.Sc., F. I. A., School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, Private Bag 3, WITS 2050, South Africa. Tel: +27-(0)11-646-5332; Fax: +27-(0)11-717-6285; E-mail: rthomson@icon.co.za

Money is round and rolls away.	English proverb
Money is flat and meant to be piled up.	Scottish proverb

1. Introduction

1.1 In defined contribution retirement funds members may have the right to choose the investment channels in which their contributions are to be invested, and to apportion their fund investments between those channels. It is proposed in this paper that an interactive system can be used to help a member decide on such choices and apportionments.

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

1.2 The paper appears in two parts. In the first part (Thomson, 2003), the use of expected utility theory, for the purposes suggested in this part, was defended against arguments that have been levelled against it, and empirical evidence was discussed. In this, the second part, the method is discussed. The system uses a member's answers to a series of questions to derive a utility function. The observed values are interpolated by means of appropriate formulae to produce a smooth utility function over the whole positive range of benefits at retirement. The resulting function, together with stochastic models of the returns on the available channels and of the annuity factor at exit, is then used to recommend an optimum apportionment of the member's investment based on dynamic programming. Allowance is made for future service as well as past service. The proposed system is applied to the observed values of the utility functions of post-retirement income elicited from a sample of members of retirement funds. Difficulties in the application are discussed and the results are analysed. The sensitivity of the recommendations to the parameters of the stochastic model is discussed.

1.3 The member's utility function, and the probability density function of the amount of the benefit, must be determined as at the date on which the member intends to draw his or her benefit (the 'exit date'). The probability density function will depend on the circumstances of the member, the rules of the fund and the output of the stochastic model. Both the utility function and the probability density function are specified in real terms. The benefit under consideration is the amount payable as at the exit date in respect of service up to that date, converted into an annuity if applicable.

1.4 It may be noted from the above that no allowance is made for the different benefits that may become payable on death in service or on ill-health retirement. This is deliberate. The only modes of exit allowed for are resignation, normal retirement and voluntary early or late retirement. The reason for this is that it is assumed that appropriate risk benefits are provided for other contingencies out of contributions other than those payable towards retirement benefits.

1.5 In Section 2 relevant literature is surveyed. Section 3 sets out the fund model. Section 4 discusses the stochastic investment model and the distribution of the annuity factor at exit. Section 5 discusses the principles and explains the method to be used for the elicitation of a utility function. Section 6 discusses issues relating to the interpolation (and extrapolation) of the utility function over the whole positive range of benefits and explains the interpolation method proposed. Section 7 explains how the interpolated utility function and the stochastic model may be used to derive a recommended investment channel choice. Section 8 shows the results of applying the methods of Sections 3 to 7 to the utility functions elicited from a sample of subjects. Section 9 draws conclusions, gives some caveats regarding the implementation of the system, and suggests avenues for further research.

2. Literature

2.1 Modern portfolio theory developed from Markowitz's (1952) analysis in two-dimensional space defined by the mean and variance of the single-period return. In the actuarial literature, this was modified by Wise (1984a, 1984b, 1987a, 1987b) and Wilkie (1985) to define the space in terms of the mean and variance of the surplus at a specified time horizon. This adaptation was useful for the analysis of the problem for a financial institution (such as a life office or a defined benefit retirement fund) with specified liabilities. Although the time horizon might be some years hence, the problem was still treated as a single-period problem; no allowance was made in the optimisation process for the review of portfolio selections before the time horizon. In the financial economics literature, the theory has also been extended to allow for the liabilities of the investor (Sharpe & Tint, 1990).

2.2 In the mean time, in the economics literature two developments had taken place. First, it was observed (e.g. Ingersoll, 1987, pp95-97) that a sufficient condition for mean–variance analysis to apply was that the investor should have quadratic utility or that the distribution of returns should be multivariate normal (or at least elliptically symmetrical). This gave rise to the pursuit of more general theories based on the maximisation of expected utility, stimulated by recognition of the weaknesses of quadratic utility functions and by findings that returns on risky assets were not generally normally distributed. Secondly, Mossin (1968) showed how allowance could be made by dynamic programming for the review of portfolio selections at discrete time intervals up to the time horizon, so that, in the optimisation process, the decision-maker would not be assumed to be committed to the current portfolio selection until the time horizon. Mossin's work was extended by Samuelson (1969) and Merton (1969) to solve the dynamic consumption and portfolio selection problems simultaneously in discrete and continuous time respectively, including allowance for a bequest motive. This involved determining the optimum consumption from time to time, as well as the optimum exposure of savings to a risky asset and a risk-free asset, using an intertemporal utility function. Using a martingale technique instead of dynamic programming, Cox & Huang (1989) extended Merton's results to allow for non-negativity constraints on consumption and final wealth.

2.3 Sherris (1992) generalised Wise's and Wilkie's work to allow for more realistic utility functions and for a dynamic approach to portfolio selection in discrete time. Further work on the portfolio selection problem in defined benefit retirement funds followed, notably O'Brien (1986, 1987), Haberman & Sung (1994), Boulier, Trussant & Florens (1995), Cairns (1995, 1996, 1997), Boulier, Michel & Wisnier (1996), Cairns & Parker (1997), Ziemba & Mulvey (1998), Siegmann & Lucas (1999), Chang (1999) and Josa-Fombellida & Rincón-Zapatero (2001).

2.4 Booth & Ong (1994) applied expected utility maximisation to the investment decision of a personal investor with no liabilities and no future cash flow on a single-period basis with an exponential utility function.

2.5 Khorasanee (1995) derived (in closed form) the mean and variance of the terminal fund assuming normal distributions of forces of return. Expected utilities of outcomes are not considered, so no definite conclusions can be drawn with regard to recommendations to individual members.

2.6 Khorasanee & Smith (1997) considered the typical strategy whereby, as a member approaches retirement age, his or her fund balance is shifted from risky assets to risk-free or matched assets, generally referred to as a 'lifestyle' strategy. Because of the simplicity of the approach, it was possible to obtain expressions for the maximum expected utility of the terminal fund in closed form. However, there are a few problems with their approach:

- The method of eliciting the utility function is quite arbitrary, and does not conform to the axioms of expected utility theory.
- The shapes permitted for the utility function are quite restrictive; in the present study, subjects' utility functions were found to be more variable than those allowed for by Khorasanee & Smith.
- The channels available may not correspond to those assumed; in particular, the processes governing the unit prices of the respective channels may be more complex.
- The errors introduced by the assumption of a lognormal distribution of the terminal fund are not quantified.
- The errors introduced by the assumption of constant annuity rates are also not quantified.
- It is not generally optimal to be 100% exposed to a particular asset category.
- No allowance is made for the dynamic revision of portfolio choices.

Thus, while the paper provides some useful insights, it cannot constitute a basis for explicit recommendations to members with regard to investment channel apportionment.

2.7 Booth & Yakoubov (2000) considered various alternative gradual lifestyle strategies. They found no evidence to suggest that a lifestyle strategy is appropriate.

2.8 Working with a dynamic approach in continuous time, Gerber & Shiu (2000) assumed that the prices of risky assets follow a multidimensional geometric Wiener process (i.e. with constant drift and covariance). They found that the optimal investment strategy can be specified in closed form. While this is a very elegant result, which gives considerable insight into the nature of the asset allocation problem, there are a few problems with this approach:

- — The price of a bond cannot be represented by a geometric Wiener process.
- — The assumption of constant drift and variance in the processes driving prices is quite restrictive.
- — The argument of the utility function is the terminal fund, not the post-retirement income.
- — While the shape of the utility curve allowed for is quite adaptable, it does not cope with the variability found in the present study.
- — The concept of a risk-free asset begs the question whether it is a nominal short-term deposit or an inflation-protected long-term bond.

2.9 Vigna & Haberman (2001) adopted a dynamic discrete-time approach using a joint normal distribution for the real forces of return on a low-risk and a high-risk asset. The utility function was effectively quadratic, and it was applied at every year end by means of a summation of discounted utilities. The results suggest the appropriateness of the lifestyle strategy. The authors did not intend to justify their utility functions in terms of members' attitudes to risk; but, by the same token, their results cannot be used to make recommendations to members.

2.10 Working in continuous time, Boulier, Huang & Taillard (1999) developed a defined contribution fund model involving the Vasicek interest rate model, three assets, deterministic contribution rates, a guaranteed minimum benefit and terminal utility measured as a power function of surplus cash over the guarantee. Deelstra, Grasselli & Koehl (1999) developed a similar model involving the Cox-Ingersoll-Ross model of interest rates. Generalising (in some respects) the work of these authors, Cairns, Blake & Dowd (2000) developed an optimal asset allocation model for a defined contribution retirement fund in the presence of non-hedgeable salary risk in continuous time. The authors found that it is optimal to invest in a combination of three portfolios: (A) minimum risk relative to salary; (B) minimum risk relative to salary divided by annuity factor; and (C) risky portfolio, efficient with risk and return relative to those quantities. The problems with this approach are as follows:

- — The use of a power utility function as in the solutions given is too restrictive.
- — The process driving prices is still too restrictive.
- — In Cairns, Blake & Dowd (*op. cit.*), the argument of the utility function is the replacement ratio, not the post-retirement income.
- — In general the channels available to a member are not confined to, and may not include, the portfolios described.
- — The error due to the approximation of the annuity factor is not quantified.

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

2.11 In all the work relating to defined contribution funds described in ¶¶2.4 to 2.10, it is assumed that the contribution rate is constant (or at least predetermined). The theoretical framework for the relaxation of that assumption is provided by the work of Samuelson (*op. cit.*) and Merton (*op. cit.*); but the practicalities involved are non-trivial. This matter is further discussed in Section 5.1.1.

2.12 In the light of the discussion in this section, it is clear that, in order to facilitate recommendations to members of defined contribution retirement funds with regard to investment channel choice, further work is needed. It is the aim of this paper to meet that need.

3. The Fund Model

3.1 *Benefit Provisions and Members' Particulars*

3.1.1 For the purposes of this paper, it is assumed that the benefit at the exit date will be an amount equal to, or a pension based on, the accumulated contributions paid by, or in respect of, the member towards that benefit. For this purpose, the contributions paid are accumulated according to the returns earned on the investment channels chosen by the member from time to time. In this paper that amount is referred to as the 'benefit base'. If the benefit is a lump sum, it is equal to the benefit base. If it is a pension, it is equal to the benefit base divided by an annuity factor at a real rate of interest. The benefit base is expressed in real terms. In this paper, 'real' means before price inflation; real salary increases are discussed in Section 3.4.

3.1.2 The member must either specify the current benefit base or enter an identity code that will enable the system to obtain those particulars from the fund's database. The latter is preferable, as it avoids errors that might otherwise be made by the member, and serves as a member audit to identify errors that may have been made by the administrators of the fund.

3.1.3 In order to determine the amounts of future contributions payable from time to time, it will also be necessary to obtain the following information:

(1) in respect of the fund:

$c =$ the contribution rate towards retirement benefits; and

(2) in respect of the member:

$S(m) =$ the member's salary (in real terms) at certain future dates.

With regard to item (1), if there is a vesting scale in terms of which the benefit base will be reduced if the member retires on the specified exit date, the benefit base and the contribution rate must be correspondingly reduced. The future dates referred to in item (2) are those described in Section 3.2. Values of the benefit at exit will then be determined from the stochastic model as explained in Section 3.3.

3.1.4 It may be noted that the contribution rates are assumed to be constant, and that future real salaries are assumed to be deterministic. This is further discussed in Section 3.4.

3.2 *Subdivision of the Period to Exit*

3.2.1 In defining the problem, we must bear in mind that the member is able to review his or her choices from time to time, and that those choices can be based on revised probability distributions of the benefit at exit, allowing for the returns earned to the date of the review.

3.2.2 We therefore divide the term from the date of calculation to the date of exit into N years, each of which is assumed for simplicity to be an integral number of calendar years. Let $t(m)$ be the length (in integral years) of period m and let $T(m)$ denote the period from the initial date to the end of the mth period, so that:

$$T(m) = \sum_{w=1}^{m} t(w).$$

The lengths of $t(m)$ are further discussed in ¶3.3.4.

3.3 *The Fund Process*

3.3.1 For the sake of simplicity (as explained below), it is assumed that half the contributions payable during each period (in real terms) are paid at the beginning of that period and the other half at the end.

3.3.2 Let $r_k(m)$ be a random variable denoting the total real force of return on the kth channel during the mth period, so that:

$$r_k(m) = \sum_{u=T(m-1)+1}^{T(m)} y_{k,u}$$

where $y_{k,u}$ is the real average force of return on channel k during year u.

3.3.3 Let $B(m)$ denote the benefit base accumulated at the end of the mth period before the contribution then due. Let $C(m)$ denote the contributions due during the mth period and accruing towards the benefit at the exit date, and let $p_k(m)$ denote the proportion invested in channel k during the mth period. Then, for $m = 1, \ldots, N$:

$$\begin{aligned} B(m) &= [B(m-1) + \tfrac{1}{2}\{C(m-1) + C(m)\}] \exp\left\{\sum_{k=1}^{K} p_k(m) r_k(m)\right\} \text{for } m > 1 \\ &\quad [B(0) + \tfrac{1}{2}C(1)] \exp\left\{\sum_{k=1}^{K} p_k(1) r_k(1)\right\} \text{for } m = 1 \qquad (1) \end{aligned}$$

where $B(0)$ is as specified in ¶3.1.2, and:

$$C(m) = \begin{cases} t(m)S(m)c \text{ for } 1 \leq m \leq N \\ 0 \text{ for } m = 0. \end{cases} \tag{2}$$

It should be noted that equation (1) assumes continuous rebalancing. For this reason it is the total force of return for period m that is multiplied by $p_k(m)$, not the total return for that period. In practice, if a member has chosen a particular asset allocation, it is reasonable to assume that that allocation will be maintained until the next occasion on which a selection is made, so that this assumption is itself reasonable.

3.3.4 For the purposes of subdivision into periods as explained in ¶3.2.2, it should be noted that $t(m)$ should be sufficiently large so that, for all values of m, j and k (including $j = k$):

$$\text{Cov}\{r_k(m), r_j(m-1)\} \approx 0.$$

The purpose of this requirement, which is discussed further in ¶4.2.3, is explained in ¶7.1.10. On the other hand, subject to that requirement, it should not be made so great that the simplified assumption regarding the payment of contributions might result in excessive errors of approximation. The trade-off depends on the model used; it would be advisable to test the sensitivity of results to alternative values of $t(m)$.

3.4 *The Salary and Contributions Process*

3.4.1 It may be noted that, in the specification of the fund process, the member's salary is assumed to be deterministically specified (in real terms). In fact, prospective salaries are required at each prospective review date, and at the exit date. The system could calculate typical values based on the member's current salary, age and length of service, and allow the member to substitute alternative values.

3.4.2 The problem with the use of a stochastic salary process is that the parameters of that process would have to be specified by the member concerned. While it is reasonable for the idealised decision-maker to make assumptions about the investment channel process, the salary process is a different matter. Here it would be unreasonable to suppose that the idealised decision-maker is better placed to model the process than the member.

3.4.3 Furthermore, while it may be possible to devise a stochastic model of the salaries of a large sample of members for a given current salary, age and length of service, it would be necessary to allow for the additional variance of an individual member's future salary. The distinction is analogous to that between diversifiable and undiversifiable risk in the analysis of returns on capital assets.

3.4.4 It would be possible to elicit the parameters of an individual's salary process and to allow for the non-diversifiable portion of that process, but until further research has been completed that will not be practicable. In the mean time it would be better to present results on alternative salary prospects and allow the member to vary them in order to test the sensitivity of the recommended apportionment, and the certainty equivalent of the corresponding maximum expected utility, to such alternatives. (The 'certainty equivalent' of a risky prospect is the outcome whose utility is equal to the expected utility of the risky prospect. It therefore constitutes the normative risk-free value to the decision-maker of the risky prospect as at the date of its outcome.)

3.4.5 In the above formulation, contribution rates are assumed to be constant. It would be a simple matter to allow for contribution rates that vary deterministically by age. Additional voluntary contributions have not been dealt with in the model.

4. The Stochastic Model

With regard to the modelling of the investment returns on the investment channels available to members of the fund, and of inflation, two alternative approaches may be adopted. The first is to use a model that produces a multivariate normal distribution of the forces of return and inflation, and the second is to use a generalised simulation model. The advantage of the first approach is that solutions to part of the problem may be found analytically, thereby reducing the calculation time required. However, it may not be possible to model the variables required in this manner. The generalised case must therefore also be considered.

4.1 *A State-Space Model*

4.1.1 For the purposes of the first approach, it is assumed that a stochastic model of investment returns on the investment channels available to members of the fund, and of inflation, has been developed. The model is assumed to be a state-space model specified as:

$$\boldsymbol{y}_u = \mathbf{H}\boldsymbol{z}_u + \boldsymbol{d} \tag{3}$$

where:

$$\boldsymbol{z}_u = \mathbf{F}\boldsymbol{z}_{u-1} + \mathbf{G}\boldsymbol{e}_u. \tag{4}$$

In this model:

— $\boldsymbol{y}_u$ is a K-component vector comprising the real average forces of return on the respective channels during year u;

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

— $\boldsymbol{z}_u$ is a vector with at least K components representing the state space;
— $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ are parameter matrices of appropriate dimensions;
— $\boldsymbol{d}$ is a K-component vector representing the means of the real returns on the respective channels; and
— $\boldsymbol{e}_u$ is a vector of mutually and serially independent random variables with standard normal distributions.

The components of the state-space vector $\boldsymbol{z}_u$ may have explicit economic meaning, or they may merely be statistically explanatory. In essence, the vector defines a state that may be modelled as a linear Markov process and of which the required variables are linear functions. (As shown in the core reading for subject 109 (Faculty & Institute of Actuaries, 2001), the Wilkie model (Wilkie, 1986) can be reduced to a linear Markov model of a state-space vector, i.e. equation (4); but, because it models variables of which the forces are non-linear functions, it does not conform to the requirements of this approach; i.e. equation (3) does not apply.)

4.1.2 From equations (3) and (4), it may be seen that $\boldsymbol{y}_u$, being a linear function of $\boldsymbol{e}_u$, which is a multivariate normal, is itself a multivariate normal. This makes it possible to use an analytical method to solve the optimisation problem, as explained in Section 7.1. The mean of $\boldsymbol{y}_u$ may be shown to be:

$$\mathrm{E}(\boldsymbol{y}_u) = \mathbf{H}\mathbf{F}^u \boldsymbol{z}_0 + \boldsymbol{d} \tag{5}$$

and its covariance matrix:

$$\mathrm{Cov}(\mathbf{y}_u, \mathbf{y}_v) = \mathbf{H}\left(\sum_{w=v-u}^{v-1} \mathbf{F}^w \mathbf{G}\mathbf{G}' \mathbf{F}'^w\right)\mathbf{H}' \text{ for } u \le v \tag{6}$$

in particular:

$$\mathrm{Var}(\mathbf{y}_u) = \mathbf{H}\left(\sum_{w=0}^{u-1} \mathbf{F}^w \mathbf{G}\mathbf{G}' \mathbf{F}'^w\right)\mathbf{H}'.$$

4.1.3 From ¶¶3.3.2 and 4.1.2 it may be noted that, since $\boldsymbol{y}_u$ is a multivariate normal, $r_k(m)$ is normally distributed with mean:

$$\mu_k(m) = \sum_{u=T(m-1)+1}^{T(m)} \mu_{ku}$$

and covariance:

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

$$\sigma_{kj}(m) = \sum_{u,v=T(m-1)+1}^{T(m)} \text{Cov}(y_{k,u}, y_{j,v})$$

where μ_{ku} is the kth component of $\text{E}(\boldsymbol{y}_u)$ obtained from equation (5) and $\text{Cov}(y_{k,u}, y_{j,v})$ is obtained from equation (6).

4.1.4 The annuity factor at the member's exit age may be assumed to be fixed, or (as justified by Cairns, Blake & Dowd, 2000) it may be defined as a lognormally distributed variable.

4.2 *Simulation Model*

4.2.1 For the purposes of the second approach, it is assumed that $\boldsymbol{y}_u$ is a random variable, pseudo-random samples of which may be generated by means of a stochastic model. Suppose that J simulations are made of this variable. We denote by y_{jku} the jth pseudo-random sample value of the real average force of return on channel k during year u for $j = 1, \ldots, J$ and $k = 1, \ldots, K$.

4.2.2 We may then define $y_{jk}(m)$ to be the jth pseudo-random sample value of the real aggregate force of return on channel k during period m as:

$$y_{jk}(m) = \sum_{u=T(m-1)+1}^{T(m)} y_{jku}.$$

4.2.3 Under this model, if:

$$\text{Cov}\{r_k(m), r_j(m-1)\} \gg 0$$

for small values of $t(m)$, it would be possible to modify the fund model described in ¶3.3.3 by defining:

$$B(m) = B(m-1)\exp D + cS(m)\frac{t(m)(\exp D - 1)}{D}$$

where:

$$D = \sum_{k=1}^{K} p_k(m) r_k(m)$$

which would give a better approximation than equation (1).

4.2.4 The conversion into an annuity at exit may be dealt with by means of an additional component of the vector $\boldsymbol{y}_u$ to represent the real long-term interest rate, which can then be used at exit to determine an annuity factor, and hence, from the terminal benefit base, the sample real pension.

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

5. The Elicitation of the Utility Function

In this section consideration is first given to context and framing effects in the elicitation process, and then to the format of the questions posed to the member.

5.1 *The Argument of the Utility Function*

As discussed in Section 2, authors have used various definitions for the argument of the utility function. Some have used the terminal fund (i.e. the benefit base at exit), others have used the purchased pension, and yet others have used the replacement ratio. None has yet used a lifetime consumption function in the context of a defined contribution retirement fund. There is little justification in the literature for the use of the replacement ratio. In this section consideration is accordingly given to lifetime consumption, post-retirement income and the terminal fund.

5.1.1 *Lifetime consumption*

5.1.1.1 As discussed in ¶2.2, Samuelson (1969) and Merton (1969) developed a framework for the simultaneous solution of the consumption and portfolio selection problem, in discrete and continuous time respectively, including an allowance for a bequest on death. For practical application Samuelson's approach would generally be more convenient. It would be quite possible to use that method to determine, not only the optimal choice of investment channels, but also (and simultaneously) the optimal contribution rate subject to a minimum of the mandatory contribution rate if applicable (or a minimum of zero otherwise).

5.1.1.2 However, this would necessitate further questions in the elicitation of the utility function, so as to elicit, not only the utility of the benefit at retirement, but a two-dimensional utility function of consumption by year. In practice, the inter-temporal effect could be captured by means of a discounting of a one-dimensional utility function to allow for inter-temporal preferences, but the discount rate would not necessarily be constant, and the elicitation process would therefore be complex.

5.1.1.3 To simplify the problem, it may be assumed that the member's exit date and contribution rate are determined; but it would nevertheless be quite simple to extend the system to allow for assets and liabilities outside the retirement fund. Utility would then be defined in terms of post-retirement income (including conversion of assets outside the fund into an annuity at the exit date if applicable).

5.1.2 *Fund benefit*

5.1.2.1 If assets and savings outside the retirement fund can be ignored, it is justifiable to consider only the fund benefit. In these circumstances the question arises whether it is necessary to consider the utility of the pension

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

purchased by the terminal fund, or whether it would be adequate to consider that of the terminal fund itself. In fact the benefit may be payable not entirely in the form of a pension, but either wholly or partly in the form of a lump sum.

5.1.2.2 The arguments in favour of considering the utility of the terminal fund are as follows:

— The member may intend to invest the lump-sum benefit in unconventional assets (such as cattle or agricultural land) that might carry more utility to the member than the income that they produce. This may be particularly true in third-world countries.
— The member may have substantial capital requirements at retirement, such as the repayment of a mortgage bond.
— Members may tend to focus on the accumulation of the benefit base, rather than on the amount of income that it can produce. Where there are cash options, members may tend to elect them. Since utility is subjectively determined, such subjective preferences cannot be ignored.
— As pointed out by Booth (1995, p110), since real long-term interest rates are relatively stable, real annuity conversion factors are also stable. It may therefore not be inaccurate to assume that they are fixed. This being the case, the use of a utility function of the lump sum should give the same results as the use of a utility function of the pension. While this effect may be offset to some extent by the effect of the longer durations of index-linked bonds, it is not invalidated; it remains true for zero-coupon bonds, where the duration of an index-linked bond is equal to that of a nominal bond.
— The member may be strongly driven by the bequest motive.

5.1.2.3 The arguments in favour of considering the utility of a purchased annuity (either notionally or actually purchased) are as follows:

— The major need of the member after retirement is for the replacement of income.
— If the member is obliged to convert the terminal fund in full into a pension, the amount of the terminal fund is irrelevant. If a portion must be converted into a pension, it would be unnecessarily complicated to consider the utility of a combination of a lump-sum benefit and a pension.
— If a lifetime-consumption approach were to be adopted, it would be necessary to consider the income generated by the assets in which any lump-sum benefit would be invested.

5.2 *Framing*

One of the major sources of bias in assessment procedures identified by Hershey, Kunreuther & Schoemaker (1982) is that of context and framing effects. They state that: "context or framing differences strongly affect choice

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

in a non-normative manner". They argue, on the basis of their experiments, that: "people do not hold preferences free of context". They also cite a number of other texts to support this argument. It is therefore important to ensure that the hypothetical prospects offered to the member relate as closely as possible to the context of the actual prospects. Similarly, as discussed in Thomson (2003, ¶3.1.8), emotive framing effects should be avoided. Keeney & Raiffa (1976, pp189-90) suggest that, in preparation for assessment, a decision-maker should be reassured that:

(1) there are no right or wrong answers to any of the questions;
(2) all answers should reflect the decision-maker's own preferences; and
(3) the decision-maker can change his or her mind with regard to any of the answers if so desired.

These reassurances should be built into the system, and allowance should be made for the member to change answers before proceeding to the recommendation.

5.3 *Elicitation Method*

5.3.1 Farquhar (1984) categorises and analyses the various question formats used to elicit decision-makers' utility functions by means of hypothetical prospects. He uses the symbol R to denote a preference relationship, which may be $\succ$, $\sim$ or $\prec$ (i.e. strong preference of the preceding to the succeeding, indifference between them, or strong preference of the succeeding to the preceding), or $\succeq$ or $\preceq$ (i.e. weak preference of the preceding to the succeeding or vice versa), as defined in ¶2.4.5 of Thomson (2003). As in ¶2.5.3 of that paper, the possible outcomes of a particular hypothetical prospect and the associated probabilities are specified as $a_s p a_t$; where: $a_1 \succeq a_s \succeq a_t \succeq a_r$; and the probabilities of outcomes of a_s and a_t are p and $1 - p$ respectively. Using that notation, a comparison of hypothetical prospects presented to the decision-maker may be represented as $(a_1 p a_r) R (b_1 q b_r)$; or $(a_1 p a_r) R b$; the first expression representing a comparison of two uncertain prospects, and the second (as proposed in this paper) a comparison between an uncertain and a certain prospect. To represent the format of the question, the item to be specified by the decision-maker is underlined. Thus, for example, the format proposed in this paper may be represented as $(a_1 0.5 a_r) \sim \underline{b}$, since the decision-maker has to specify the certainty-equivalent of the uncertain prospect $a_1 0.5 a_r$.

5.3.2 It is considered that a member will find it easier to compare a certain prospect with an uncertain prospect, because there are less data to bear in mind. In view of the fact that the process is automated, it is important to avoid complexities as much as possible. We therefore confine our attention to comparisons of the format $(a_1 p a_r) R b$. In doing so, it must be acknowledged that we may be exposing the procedure to biases due to the certainty effect (Thomson, 2003, Section 3.9). It may be worthwhile to test empirically whether such biases are more serious than the confusion arising from comparison of two uncertain prospects. Pending such tests, the simpler approach is proposed.

5.3.3 Farquhar (*op. cit.*) analyses such comparisons as follows:

preference comparison:

$$(a_1 p a_r)\underline{R}b$$

probability equivalence:

$$(a_1 \underline{p} a_r) \sim b$$

value equivalence:

$$(\underline{a} p a_r) \sim b \text{ or } (a_1 p \underline{a}) \sim b$$

certainty equivalence:

$$(a_1 p a_r) \sim \underline{b}.$$

The comparison proposed in this paper is thus identified as certainty equivalence, with $p = 0.5$. This method is referred to by Anderson, Dillon & Hardaker (1977) as the 'equally likely certainty equivalent' (or 'ELCE') method.

5.3.4 The simplest of the comparisons is arguably preference comparison. However, in view of the problems of subjective probability weighting (Thomson, 2003, Section 3.10), any value of p other than 0.5 is likely to create a bias. Schoemaker (1980, pp14, 182) cites Karmarkar (1978) and Van Dam (1973) as evidence that utility functions constructed with 50–50 prospects differ systematically from those constructed with, say, 75–25 prospects. "Which is the more faithful one," he avers, "is open to debate." Apart from framing effects, it is difficult, in view of the symmetry of a 50–50 chance, to see in which direction such a chance is likely to be biased.

5.3.5 Preference comparison may be used repetitively with $p = 0.5$ and varying values of a_1 (or a_r) or b, in order to converge towards a critical value representing value or certainty equivalence respectively (Keeney & Raiffa, 1976, pp193-4). This method may be preferable if a member has difficulty in responding direct to value or certainty equivalence; but otherwise the member may find it tedious. Consideration could be given to offering members this approach as an alternative. In the mean time the more direct method is proposed.

5.3.6 Probability equivalence is similarly avoided because of subjective probability weighting. Hershey, Kunreuther & Schoemaker (*op. cit.*) found that certainty equivalence methods generally yield greater risk seeking than

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

probability equivalence methods. However, their tests were made only for prospects involving losses. Such prospects do tend to elicit risk seeking (Friedman & Savage, 1948); but in the context of this paper it is not clear what constitutes a loss, so that such tendencies are less likely to occur.

5.3.7 As between value and certainty equivalence there is relatively little to choose. Certainty equivalence has the advantage that it allows for the subdivision of an appropriate range of benefits on a predetermined basis.

5.3.8 Farquhar (*op. cit.*) subdivides certainty equivalence methods into fractile methods and chaining methods. Under the fractile method the ith comparison is $(a_1 p_i a_r) \sim \underline{b_i}$. This method has also been rejected because of subjective probability weighting.

5.3.9 If previously elicited values are used in subsequent comparisons, the responses are 'chained'. Chaining methods of certainty equivalence are further subdivided into the fractionation and midpoint methods. Under the fractionation method the ith comparison is $(a_i p a_r) \sim \underline{a_{i+1}}$. This method has the disadvantage that, if $p = 0.5$, then no utility values can be obtained for a in the range $a_s < a < a_1$, where $u(a_s) = 0.5$; otherwise subjective probability weighting again becomes problematic.

5.3.10 The midpoint chaining method involves comparisons of the form $(a_j 0.5 a_k) \sim \underline{a_i}$, where a_j and a_k are one of the closest pairs of values already found, including a_1 and a_r, closeness being defined with reference to difference in utility. This is the method adopted for the purposes of this paper (i.e. the ELCE method referred to above).

5.3.11 As Farquhar (*op. cit.*) points out, this method does have some drawbacks. Besides the problem of certainty effects discussed above, it suffers from serial dependence, range effects, distortions in risk behaviour and certain other biases. These effects are considered by Krzysztofowicz & Duckstein (1980) and Novick, Dekeyrel & Chuang (1981). In the present context it appears that, of these effects, only range effects need be further considered.

5.3.12 Range effects relate to the difficulty experienced by decision-makers in making comparisons involving uncertain prospects with widely differing outcomes. To overcome this problem, Krzysztofowicz & Duckstein (*op. cit.*) propose a 'variable range' method. This involves partitioning the range into two arbitrary subintervals and then applying the midpoint method separately to each subinterval. The utility function over the entire range is then obtained by offering a final comparison involving one elicited value from each subinterval. This method may constitute an improvement over the proposed method.

5.3.13 Quiggin (1993, p49) claims that the ELCE method is one of the most successful methods of eliciting utility functions.

5.4 *Changes in Utility Functions*

5.4.1 In this paper it is assumed that members will be prepared to state their preferences now for future benefits. It may at first sight appear that we are assuming that there will be no change in a member's preferences between the date of calculation and the date of exit.

5.4.2 In fact, what we are assuming is that a member will be prepared to anticipate such changes. If the member is able to do so, the recommended apportionment will fairly reflect the anticipated preferences. Otherwise it will only reflect the member's current preferences.

5.4.3 It would be possible to allow for assumed future preferences as a function of current preferences, but this would have to be based on empirical studies, which would not necessarily provide a fair reflection of the way in which a particular member's preferences will evolve over the period to exit.

5.4.4 As time passes, it may be expected that members' preferences will converge towards their preferences at exit. This convergence will be due, not only to changing attitudes, but also to changing circumstances.

5.5 *The Elicitation Process*

5.5.1 For the purposes of this paper the axioms of von Neumann & Morgenstern (1947) are assumed to hold. The justification of this assumption was dealt with in Thomson (2003). In order to determine a member's utility function at a particular time, it is necessary to obtain the member's answers to certain questions. The following questions, which have been adapted from sources such as Bowers *et al.* (1986), may be posed by means of an interactive program.

5.5.2 The first question is of the following form: "If you could choose your channels:

(1) so that there is a 50-50 chance of getting:
    (a) a benefit of X; or
    (b) a benefit of Y; or

(2) so that you will get Z with certainty;

which of (1) and (2) would you choose?
Enter 1 or 2, or, if you are indifferent, leave blank:"

In this question, X and Y could, for example, be the upper and lower 90% confidence limits of the benefit, and Z could be their geometric mean. For the purpose of determining the confidence limits, a high-variance investment channel apportionment may be assumed and the stochastic model of Section 4 used. If the member is indifferent between the alternatives (1) and (2), the use of the geometric mean suggests a logarithmic utility function.

5.5.3 The second question is asked if and only if the member has entered 1 or 2 in response to the first question. If the member has entered 1, the question is: "To what value would the amount in (2) have to be increased in order to change your answer?" If the member has entered 2, the question

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

is: "To what value would the amount in (2) have to be decreased in order to change your answer?''

5.5.4 We now provisionally fix three points on the member's utility curve. For X and Y we may set arbitrary values $u(X) = 0$ and $u(Y) = 1$. These values may be set arbitrarily because utility curves may be subjected to positive linear transformation without affecting the maximisation of expected utility. In other words, provided $A > 0$, the problem:

$$\text{maximise } \mathrm{E}\{Au(x) + B\}$$

gives the same result as:

$$\text{maximise } \mathrm{E}\{u(x)\}.$$

5.5.5 Because of the specification of the 50–50 chance, the value of the utility function for a benefit of Z (after any increase or decrease indicated by the member in the second question) is 0.5; i.e.:

$$u(Z) = \tfrac{1}{2}\{u(X) + u(Y)\} = 0.5.$$

5.5.6 The range (X, Y) may now be split into two ranges (X, Z) and (Z, Y). For each of these ranges we ask the same pair of questions, again with geometric means in alternative (2). The answers to these questions give us five points on the provisional utility curve, two of which are arbitrary and three of which provide meaningful information on the shape of the member's utility curve.

5.5.7 It would be theoretically possible to continue asking similar questions indefinitely, thus ostensibly obtaining more and more information on the shape of the member's utility curve. However, not only is the member's patience likely to wear thin, but we are not likely to obtain much more information than we could get by fitting a suitable function to the five observed values. Furthermore, because the responses are subjective, excessive accuracy is spurious. In any event, at some stage it will be necessary to find a method of determining values for benefit levels between those specified. This matter is further discussed in Section 6. For the purposes of this paper, however, it is assumed that we have five observed utility values.

6. The Interpolation of the Utility Function

6.1 *Interpolation vs. Regression*

6.1.1 In the early literature — before Mosteller & Nogee (1951) — a utility function was supposed to be deterministically and objectively quantifiable. Mosteller & Nogee found that subjects were not consistent in

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

their preferences, which suggested a shift from the deterministic to the stochastic and from the objective to the subjective. Savage (1954) confirmed the shift to the subjective by basing his axioms on subjective probabilities. Subsequent authors (e.g. Schoemaker, 1980, pp117-8) have emphasised that preferences are not only subjective, but stochastic; there is an error in a decision-maker's comparison of prospects that may be considered to have a probability distribution of its own.

6.1.2 At the same time the assessment methods only derive a finite number of values of the utility function. For many applications, such as envisaged by this paper, a continuous function is required.

6.1.3 These considerations suggest the possibility of maximum likelihood estimation of a parametric utility function, or at least — in the absence of information about the distribution of the error — a least squares method. Such methods are not extensively dealt with in the literature. Exceptions include French *et al.* (1992), where a function is fitted using least squares. The situation may be considered as analogous to the graduation of mortality data, and it may be possible, if information can be obtained regarding the distribution of errors in preference statements, to adopt a combination of parametric and non-parametric methods along the lines suggested by Thomson (1999).

6.1.4 Another possible interpretation of the assessment process, which is a modification of the deterministic approach, is suggested by French (unpublished): "Subjects do not possess a utility function internally which they need to estimate accurately. Rather during the analysis they construct a function and the construction process guides the evolution of their ... preferences.''

6.1.5 French's approach suggests that, during the construction process all errors are eliminated, or alternatively, that the concept of an 'error' is meaningless. In either case, it would seem to be inconsistent to fit a parametric function by means of least squares; all that would be needed is an interpolation formula to estimate values between those assessed. The only role of a parametric function would be to facilitate calculations; but that would constitute an approximation, not an inherent improvement in the assessment process.

6.1.6 French (1986, pp174, 199, 346) repeatedly stresses the need for sensitivity tests after the initial determination of utilities. For each such test a recommended strategy is determined by maximising expected utility. A lower bound of the expected utility of that strategy is found by testing the limits of the agent's uncertainty in response to the hypothetical comparisons. Similarly, an upper bound of the expected utility of every other strategy is found. If the former is greater than the latter, the agent can adopt the recommendation with confidence.

6.1.7 Unfortunately, while it may be argued that French's approach makes sense for a finite number of discrete alternative strategies, it breaks

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

down where, as in the present context, the range of alternative strategies is continuous.

6.1.8 On the other hand, the member's uncertainties could be used as a basis for determining the subjective probability distribution of the error in the responses given. There is room for research in this direction.

6.2 *The Interpolation Process*

6.2.1 Let us denote the five benefit amounts as $x_0, \ldots, x_4$ in increasing order of magnitude, and the corresponding observed utility values as:

$$u(x_i) = \frac{i}{4}. \tag{1}$$

If for any i, $x_i = x_{i+1}$, the member should be challenged to reconsider the question that gave rise to that result. It could be pointed out, for example, that this implies that the possibility of getting an extra benefit of $x_{i+2} - x_{i+1}$ is valueless to the member. If the member persists, he or she should be informed that the system is unable to make a recommendation.

6.2.2 For the purposes of this part of the paper, we use a utility function of the form:

$$u(x) = v_i(x) \text{ for } x \in S_i\ ,\ i = 1, \ldots, 4 \tag{7}$$

where:

$$S_i = (0, x_1] \text{ and } v_i(x) = u_1(x) \text{ for } i = 1 \tag{8}$$

$$S_i = [x_3, \infty) \text{ and } v_i(x) = u_3(x) \text{ for } i = 4 \tag{9}$$

$$S_i = [x_{i-1}, x_i) \text{ and } v_i(x) = \frac{(x_i - x)u_{i-1}(x) + (x - x_{i-1})u_i(x)}{x_i - x_{i-1}} \text{ otherwise} \tag{10}$$

$$u_i(x) = a_i \ln x + b_i \text{ if } n_i = 0 \tag{11}$$

$$a_i x^{n_i} + b_i \quad \text{otherwise} \tag{12}$$

and n_i, a_i and b_i are determined so that, for $i = 1, 2, 3$:

— $u_i(x_{i-1}) = u(x_{i-1})$; (13)

— $u_i(x_i) = u(x_i)$; and (14)

— $u_i(x_{i+1}) = u(x_{i+1})$. (15)

It may be noted from equations (7), (8) and (10) that $u(x_1) = u_1(x_1)$ whichever of the latter two formulae is used. Similarly, $u(x_2) = u_2(x_2)$ and $u(x_3) = u_3(x_3)$ whichever of the applicable formulae is used. Equation (10), in

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

fact, represents interpolated values over the respective ranges S_2 and S_3: between $u_1(x)$ and $u_2(x)$ over S_2; and between $u_2(x)$ and $u_3(x)$ over S_3.

6.2.3 The rationale behind the form of the utility function given in ¶6.2.2 is that $u_i(x)$ exhibits constant relative risk aversion (Pratt, 1964):

$$\begin{aligned} R_R &= -\frac{u''(x)}{u'(x)} \\ &= -(n-1). \end{aligned}$$

This means that, over the intervals S_1 and S_4, $u(x)$ also exhibits constant relative risk aversion. Furthermore, for all i, $v_i(x_i) = u_i(x_i)$ and $v_i'(x_i) = u_i'(x_i)$. The average value of $-v_i''(x)$ (the numerator of R_R) over the interval $[x_{i-1}, x_i)$ for $i = 2, 3$ is therefore:

$$-\frac{u_i'(x_i) - u_{i-1}'(x_{i-1})}{x_i - x_{i-1}}$$

and the average value of $v_i'(x)$ (the denominator of R_R) over that interval is:

$$\frac{u_i(x_i) - u_{i-1}(x_{i-1})}{x_i - x_{i-1}}.$$

Thus, the average relative risk aversion over that interval (weighted by its denominator) is:

$$-\frac{u_i'(x_i) - u_{i-1}'(x_{i-1})}{u_i(x_i) - u_{i-1}(x_{i-1})}$$

which is an unbiased approximation to its true value. Finally, for a subject whose utility function exhibits constant relative risk aversion over its entire domain, the interpolated utility function will preserve that constant.

6.2.4 Except where $n_i = 0$, we have, from equations (12) to (15):

$$a_i x_{i-1}^{n_i} + b_i = u(x_{i-1}) \tag{16}$$

$$a_i x_i^{n_i} + b_i = u(x_i) \tag{17}$$

and

$$a_i x_{i+1}^{n_i} + b_i = u(x_{i+1}) \tag{18}$$

which, after some algebra, gives:

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

$$w_i^{n_i} - 2 + w_{i-1}^{-n_i} = 0 \tag{19}$$

where:

$$w_i = \frac{x_{i+1}}{x_i}. \tag{20}$$

6.2.5 Equation (19) can be solved for n_i by a Newton-Raphson method. However, it does not need to be accurately determined. All that is required in order to ensure that equation (7) is satisfied for $x_0, x_1, \ldots, x_4$ is that $u_1(x_0)$, $u_1(x_1)$, $u_2(x_2)$, either $u_2(x_1)$ or $u_2(x_3)$, $u_3(x_3)$ and $u_3(x_4)$ satisfy equations of the form of (16) to (18). In other words, $u_1(x)$ must replicate the observed values of $u(x_0)$ and $u(x_1)$; $u_2(x)$ must replicate that of $u(x_2)$; and $u_3(x)$ those of $u(x_3)$ and $u(x_4)$. Equation (7) then ensures that $u(x)$ replicates all the observed values. While the fidelity of the interpolated values of $u(x)$ to the observed values is not compromised by approximations in the values of n_i, it should be recognised that interpolated values — and more particularly extrapolated values at the extreme ends of the range of x — will be affected. Nevertheless, in view of the subjective elements in the elicitation of the utility function, excessive accuracy is spurious.

6.2.6 Once n_i has been found, the values of a_i and b_i may be determined by means of the equations:

$$a_1 = \frac{1}{4\Delta x_0^{n_1}} \tag{21}$$

$$a_2 = \frac{1}{4\Delta x_1^{n_2}} \tag{22}$$

or

$$\frac{1}{4\Delta x_2^{n_2}} \tag{23}$$

$$a_3 = \frac{1}{4\Delta x_3^{n_3}}. \tag{24}$$

$$b_1 = u(x_1) - a_1 x_1^{n_1} \tag{25}$$

$$b_2 = u(x_2) - a_2 x_2^{n_2} \tag{26}$$

and

$$b_3 = u(x_3) - a_3 x_3^{n_3}. \tag{27}$$

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

As indicated above, it is immaterial whether equation (22) or (23) is used for a_2. To avoid indeterminacy, equation (22) is used below.

6.2.7 Now from equation (19) it is evident that:

$$\lim_{w_i \to w_{i-1}} n_i = 0$$

whence, from equations (21) to (27):

$$\lim_{w_i \to w_{i-1}} a_i = \pm\infty$$

and

$$\lim_{w_i \to w_{i-1}} b_i = \pm\infty.$$

If w_i is close to w_{i-1}, therefore the numerical solution of equations (13) to (15) becomes problematic. However, if, for example, equations (16) and (17) apply, then $\lim_{n_i \to \infty} a_i x^{n_i} + b_i$ is of the form $a \ln x + b$, and equation (11) may be used.

6.2.8 Thus, if $n_i \approx 0$, then:

$$u_i(x) = a_i \ln x + b_i.$$

The values of a_i and b_i are then:

$$a_1 = \frac{1}{4(\ln x_1 - \ln x_0)} \tag{28}$$

$$a_2 = \frac{1}{4(\ln x_2 - \ln x_1)} \tag{29}$$

$$a_3 = \frac{1}{4(\ln x_4 - \ln x_3)} \tag{30}$$

$$b_1 = u(x_1) - a_1 \ln x_1 \tag{31}$$

$$b_2 = u(x_2) - a_2 \ln x_2 \tag{32}$$

and

$$b_3 = u(x_3) - a_3 \ln x_3. \tag{33}$$

6.2.9 Interpolated values of $u(x)$ may therefore be determined by means

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

of formula (10). If it is found that the above method does not give sufficient accuracy, the number of questions asked may be increased, as contemplated in ¶5.5.7, in which case the above formulae would be modified accordingly.

6.2.10 A further problem arises from the fact that, although $u_{i-1}(x)$ and $u_i(x)$ are monotonically increasing, $v_i(x)$ may not be. This matter is further discussed in Section 8.7.

7. The Problem and its Solution

7.1 *Analytical Method*

7.1.1 We first deal with an analytical method using the state-space model described in Section 4.1. Given the fund model described in Section 3, the stochastic model described in Section 4.1 and the utility function elicited as in Section 5 and interpolated as in Section 6, the problem to be solved by the proposed system is to find the apportionment during the period from the date of calculation to the next review date that will maximise the expected utility of the member's benefit at exit. In doing so, allowance must be made for the member's ability to review the apportionment at every future review date.

7.1.2 In the following paragraphs the dynamic programming method (Mossin, 1968) is applied to the problem as stated above. This involves considering each period from the last to the first, allowing the member to make a decision at the start of each period depending on the circumstances at the time. As the circumstances at the time are random, so is the member's decision. The member's utility function of the benefit at exit, and thus the maximum value of its expected utility, can, however, be expressed as a function of the benefit base as at the end of the previous period. The determination of this objective function (and of its gradient vector and Hessian matrix, which are required for the purposes of optimisation) is set out in Section A.1. The 'indirect' or 'derived' utility function (*ibid.*), defined as that maximum, may then be used to find a corresponding function as at the start of the previous period, and so on down to the first period. For the first period the commencing value of the benefit base is known. At that stage, therefore, an absolute value of the maximum expected utility can be determined, and the apportionment that produces that maximum is the apportionment that may be recommended for the first period. For the purposes of calculation, a functional form of the indirect utility function is required; it is assumed, for the sake of simplicity, that it will have the same form as the true utility function (but with different parameters). The process is complicated by the necessity to allow for contributions during each period, but the complexity is minimised by the assumption (in ¶3.3.1) that half the contributions payable during each period are paid at the beginning of that period and the other half at the end.

7.1.3 First, we consider the effect of the final contribution of $\frac{1}{2}C(N)$ assumed to be made on the date of exit and the conversion of the terminal benefit base into a pension. The benefit at exit may be defined as:

$$P = \frac{B(N) + \frac{1}{2}C(N)}{\bar{a}}$$

where $\bar{a}$ is the annuity factor for conversion into a pension, which may be a constant (possibly equal to one in the case of a lump-sum benefit) or, as discussed in ¶4.1.4, a lognormally distributed variable. If $\bar{a}$ is a constant, then it follows from equation (7) that the utility function of P may be expressed as a function of the value of $B(N)$ as follows:

$$u(P|B(N) = x) = v_i\Big(\frac{x+c}{\bar{a}}\Big) \text{ for } \frac{x+c}{\bar{a}} \in S_i,\ i = 1, \ldots, 4$$

where $c = \frac{1}{2}C(N)$, and S_i and $v_i(\bullet)$ are defined as in equations (8) to (10).

7.1.4 Using equations (19) to (33) and the values of x_i elicited, we calculate, for $i = 1, 2, 3$, the parameters n_i, a_i and b_i for the purposes of equations (11) and (12).

7.1.5 Next, we calculate appropriate (say 90%) confidence limits of $B(N)$, which are denoted by x_{N0} and $x_{N,I+1}$. (The meaning of I, and the determination of an appropriate value of I, are discussed below.)

7.1.6 Now, for $\iota = 1, \ldots, I$, we calculate:

$$x_{N\iota} = x_{N0}^{1-\iota/(I+1)} x_{N,I+1}^{\iota/(I+1)}.$$

7.1.7 For $\iota = 0, \ldots, I+1$, if $\bar{a}$ is a constant, we now calculate:

$$u\Big(\frac{x_{N\iota} + c}{\bar{a}}\Big)$$

using equations (7) to (12) and the results of ¶7.1.4. Otherwise we calculate:

$$E\{u(P)|B(N) = x_{N\iota}\}$$

using definition (A6) in Appendix A, with:

$$\mu(m) = \ln(x_{N\iota} + c) - E(\ln \bar{a})$$

and

$$\sigma^2(m) = \text{Var}(\ln \bar{a}).$$

7.1.8 Now, for $m = N, \ldots, 1$ we proceed as follows. First, for $\iota = 1, \ldots, I$, from ¶A.1.1, we calculate the parameters $n_{m\iota}$, $a_{m\iota}$ and $b_{m\iota}$ for the purposes of interpolating a derived utility function (Mossin, *op. cit.*):

$$\tilde{u}_m(x) = u\left(\frac{x+c}{\bar{a}}\right) \qquad \text{for } m = N \text{ if } \bar{a} \text{ is a constant}$$

$$E\{u(P)|B(N) = x + c\} \qquad \text{for } m = N \text{ if } \bar{a} \text{ is lognormally distributed}$$

$$\max E\left\{u(P)|B(m) = x\right\} \qquad \text{for } m < N.$$

For this purpose we use the values calculated in ¶7.1.7 (for $m = N$) or ¶7.1.11 (for $m < N$) for $x = x_{m1}, \ldots, x_{m,I+1}$. It may now be observed that I is the number of points in the range of $B(m)$ at which derived utility functions of the form:

$$\tilde{u}_{m\iota}(x) = a_{m\iota} h_{m\iota}(x) + b_{m\iota} \qquad \text{(equation (A4))}$$

are determined. These points are referred to below as 'knots'. For $x < x_{m1}$:

$$\tilde{u}_m(x) = \tilde{u}_{m1}(x). \qquad \text{(equation (A1))}$$

For $x_{m,\iota-1} \leq x < x_{m\iota}$, $2 \leq \iota \leq I$, $\tilde{u}_m(x)$ is interpolated between $\tilde{u}_{m,\iota-1}(x)$ and $\tilde{u}_{m\iota}(x)$, as stated in equation (A2); and for $x \geq x_{mI}$:

$$\tilde{u}_m(x) = \tilde{u}_{mI}(x). \qquad \text{(equation (A3))}$$

7.1.9 Next, for $m > 1$, we calculate appropriate (say 90%) confidence limits of $B(m-1)$, which are denoted by $x_{m-1,0}$ and $x_{m-1,I+1}$, and for $\iota = 1, \ldots, I$, we calculate:

$$x_{m-1,\iota} = x_{m-1,0}^{1-\iota/(I+1)} x_{m-1,I+1}^{\iota/(I+1)}.$$

7.1.10 If $m > 1$ then, for $\iota = 0, \ldots, I+1$, we now need to calculate the maximum expected utility of P, given that $B(m-1) = x_{m-1,\iota}$. If $m = 1$ then we need to calculate the maximum expected utility of P given that $B(0)$ is as specified in ¶3.1.2. From ¶¶3.3.2, 3.3.3 and 4.1.3, the conditional probability density function of $Y = \ln B(m)$, given that $B(m-1) = x$, is:

$$f\left\{y\middle|B(m-1) = x\right\} = \frac{1}{\sqrt{2\pi}\sigma(m)} \exp\left\{-\tfrac{1}{2}\left(\frac{y-\mu(m)}{\sigma(m)}\right)^2\right\}$$

where, from equation (1):

$$\mu(m) = \ln(x + c_m) + \sum_{k=1}^{K} p_k(m)\mu_k(m) \tag{34}$$

$$\sigma^2(m) = \sum_{k,j=1}^{K} p_k(m)p_j(m)\sigma_{kj}(m) \tag{35}$$

and

$$c_m = \tfrac{1}{2}\{C(m-1) + C(m)\} \text{ for } m > 1$$

$$\tfrac{1}{2}C(1) \text{ for } m = 1.$$

Note that, because it is assumed (in ¶3.3.4) that $\mathrm{Cov}\{r_k(m), r_j(m-1)\} \approx 0$, $\sigma^2(m)$ may be taken, for purposes of calculation, to be independent of $B(m-1)$. Thus, the expected utility of P, given that $B(m-1) = x$, is:

$$\begin{aligned} E\{u(P)|B(m-1) = x\} &\approx \frac{1}{\sqrt{2\pi}\sigma(m)} \int_{-\infty}^{\infty} u(e^y)e^{-\frac{1}{2}\left(\frac{y-\mu(m)}{\sigma(m)}\right)^2} dy \\ &= \sum_{\iota=1}^{I+1} E_{m\iota} \end{aligned} \tag{36}$$

where $E_{m\iota}$ is as defined in ¶A.1.6. Note that, although we are now working with the mth period, the utility we are approximating is still that of the benefit at exit.

7.1.11 The maximisation of:

$$E\{u(P)|B(m-1) = x_{m-1,\iota}\} = \sum_{\iota=1}^{I+1} E_{m\iota}\Big|_{x=x_{m-1,\iota}}$$

subject to:

$$p_k(m) \geq 0 \text{ for all } k$$

and

$$\sum_{k=1}^{K} p_k(m) = 1$$

is a non-linear programming problem. (Note that the substitution $x = x_{m-1,\iota}$ occurs in the determination of $\mu(m)$ in terms of equation (34); the values of l_ι

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

and s_ι in equations (A8) and (A9) are determined with reference to $S_{m\iota}$, and thus $x_{m\iota}$, not with reference to $x_{m-1,\iota}$.) For that purpose, the gradient vector and the Hessian matrix may be used as specified in ¶¶A.1.9-A.1.10, and a Newton-Raphson ridge optimisation method may be used. Under this method a pure Newton step is used when the Hessian is positive definite, provided that the Newton step increases the value of the objective function. If the Hessian is not positive definite, a multiple of the identity matrix is added to it so as to make it positive definite (Eskow & Schnabel, 1991). If the Newton step does not increase the value of the objective function, a ridging method is used to compute successful steps.

7.1.12 The process described in ¶¶7.1.8 to 7.1.11 is continued until we reach the first period. For that period, $B(m-1) = B(0)$ is known. So, instead of finding the maximum value of formula (36) for a series of values of $B(0)$, we find the values of $p_k(1) \geq 0$ that maximise formula (36) for:

$$\mu(1) = \ln\{B(0) + \tfrac{1}{2}C(1)\} + \sum_{k=1}^{K} p_k(1)\mu_k(1)$$

and

$$\sigma^2(1) = \sum_{k,j=1}^{K} p_k(1)p_j(1)\sigma_{kj}(1).$$

From the results given by Mossin (*op. cit.*), it follows that those values maximise the expected utility of the member's benefit at the exit date. They may thus be presented to the member as the recommended apportionment.

7.1.13 It may be noted that formula (36) is expressed in closed form, thus obviating the need for approximate integration, which would necessitate considerably more computer time. It is for this reason that the simplifications referred to in ¶¶3.3.1 and 7.1.8 are introduced. The dynamic programming approach outlined above is more realistic than a method that assumes that the apportionment decision is made once and for all at the calculation date. Furthermore, the results are likely to be very different, particularly for a very risk-averse member; such a member is more likely to accept exposure to risky assets if there are opportunities to revisit that decision before retirement.

7.1.14 The accuracy of this method depends entirely on the lengths $t(m)$ of the periods specified, on the value of I (i.e. on the number of knots used for derived utility functions) and on the confidence limits chosen. Clearly, though, to avoid unnecessarily long response times, it is desirable to avoid large values of I. This matter is further discussed in Section 8.3.

7.2 *Simulation Method*

7.2.1 Under the generalised model of ¶4.2, the process is identical to that of the analytical method, except that, instead of equation (A7), we use equation (A10).

7.2.2 As in ¶7.1.11, the maximisation of:

$$E\{u(P)|B(m-1)=x_{m-1,\iota}\}=\frac{1}{J}\sum_{i=1}^{J}u_i$$

where u_i is as defined in ¶A.2.1, subject to:

$$p_k(m)\geq 0 \text{ for all } k$$

and

$$\sum_{k=1}^{K}p_k(m)=1$$

is a non-linear programming problem.

7.2.3 While the mathematics of the generalised model are much simpler, the solution of the non-linear programming problem, with an objective function requiring summations over a large number of simulations, necessitates considerably more computer time than the multivariate normal distribution. While the simulations themselves may be made in advance, the summations tend to lengthen response times, as shown in Section 8.5.

8. Application

This section shows the results of applying the system described in Sections 3 to 7. For this purpose, the utility functions of 49 members of South African pension schemes elicited in Thomson (2000), which follow the principles of Section 5, were used.

8.1 *Assumptions*

It was assumed that:

— each subject enters a fund at age 35 and remains in service until age 60;
— real gross income from employment remains constant during that period;
— the utility function remains the same throughout that period;
— the contributions to the fund are 15% of gross income;
— the annuity factor used to convert the benefit base to a pension on retirement at age 60 is 12; and

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

— two investment channels are available — a low-risk channel 1 and a high-risk channel 2 — whose annual real forces of return are bivariate normal with means 3% and 6% and standard deviations 3% and 12% respectively, and with zero autocorrelation.

8.2 *Method*

8.2.1 At quinquennial intervals for durations from zero to 20 years and at annual intervals for durations from 21 to 24 years, the recommended apportionment was calculated on the above assumptions. For this purpose the contributions were accumulated at the expected real forces of return up to the date of calculation, assuming that the recommendations made by the system at earlier calculation dates were accepted.

8.2.2 For the purpose of dynamic programming, it was assumed that $t(m) = 5$ for all m if the duration at the date of calculation is less than 20 years, or 1 for all m otherwise. In other words, for each date of calculation of a recommended apportionment as specified in the previous paragraph, the assumed review dates for the purposes of dynamic programming are the same as the subsequent dates of calculation of recommended apportionments.

8.2.3 The analytical method described in Section 7.1 was used.

8.2.4 The values of n_i and n_{mi} were taken to the nearest first decimal place, and their absolute values were made subject to a maximum of 50.

8.3 *Accuracy versus Time: Analytical Method*

8.3.1 As indicated in ¶7.1.14, the confidence limits and the number of knots used affect the accuracy of the system, and the number of knots used also affects its response times. In order to consider the trade-off between accuracy and response time, the confidence limits and the number of knots were varied. For each subject, the recommended apportionment at each calculation date up to duration 20 years was calculated, as well as the expected accumulated benefit base at that duration. For this purpose, six knots were used between 90% confidence limits. For each subject for whom the apportionment at that duration was non-zero, the sensitivity of that apportionment to the confidence limits and the number of knots used after that duration were tested.

8.3.2 As expected, it was found that, for given confidence limits, the apportionment at duration 20 years converged as the number of knots increased. For a given confidence limit, the apportionment at that duration showed convergence about the third decimal place as the number of knots approached 12. For 12 knots, the apportionment showed convergence at various levels of confidence limit, generally in the region of 99%. The best accuracy was obtained with 12 knots between 99% confidence limits.

8.3.3 However, the average response time for 12 knots between 99% confidence limits was 25 minutes 11 seconds at duration 20 years, which was considered inordinately slow. Response times at earlier durations would be even slower. While response times are highly dependent on the hardware and

Table 1. Accuracy versus time: analytical method

Basis	Provisional	A	B	C	D	E
Confidence limits	90%	95%	95%	99%	99%	99%
I = number of knots	6	6	10	6	10	12
RMSE	0.048	0.026	0.014	0.011	0.002	–
Average response time (min:sec)	3:08	5:20	9:37	2:50	9:12	25:11

software used, the response times found in this study may serve to indicate relative values. Table 1 shows, for selected combinations of confidence limits and numbers of knots:

— the root-mean-squared error (RMSE) of the recommended apportionment at duration 20 years relative to that for 99% confidence limits with 12 knots; and
— the average response time.

8.3.4 From Table 1, it may be noted that six knots between 99% confidence limits (basis C) gave an RMSE of only 1.1%, while the response time was one-tenth of that for 12 knots. This was considered adequate for the purposes of this paper, and that basis was accordingly adopted.

8.4 *Results: Analytical Method*

8.4.1 For 26 subjects (i.e. 57% of the total), the recommended apportionment to channel 1 (the low-risk channel) was zero throughout. Figure 1 shows the recommended apportionment (p_1) to channel 1 for the

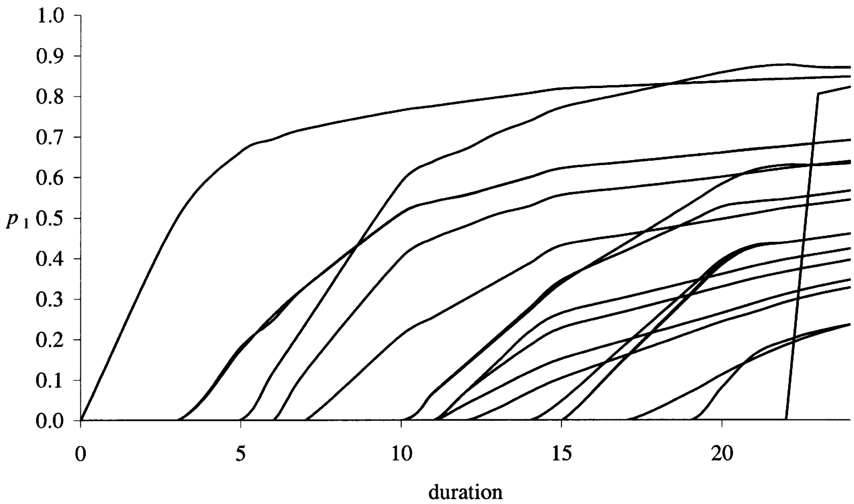

Figure 1. Recommended apportionments to low-risk channel

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

other subjects. The balance ($p_2 = 1 - p_1$) is the recommended apportionment to channel 2 (the high-risk channel).

8.4.2 In every case shown in Figure 1 the apportionment to channel 1 increases monotonically. It is clear that the pattern of movement from channel 2 to channel 1 is highly dependent on the utility function of the subject, with regard to both the date of commencement and the pace of movement. In no case does the apportionment to channel 1 reach 100%. It may be shown mathematically that, for the parameters assumed, the value of n would have to exceed 33 in order to justify full investment in the low-risk channel. No member showed such high risk seeking over the whole range. Also, in no case is there any apportionment to channel 1 at duration 0. This may be justified by considering future contributions as an extra asset. Even the most risk averse subjects will wish to diversify out of that asset with the marginal amount represented by the first contribution, at least until the next review date.

8.4.3 The steeply sloping curve on the right-hand side of the figure is in respect of subject 28. As explained in Section 8.7, the apportionment for that subject at dates close to retirement is virtually a matter of indifference, so that the substantial change is not indicative of any material change in preference. In order to avoid such spurious effects, subject 28 has been omitted from the analyses in ¶8.4.4 and Sections 8.5 and 8.6.

8.4.4 Figure 2 shows the mean apportionment to the low-risk channel at each duration, together with the means of the highest three apportionments

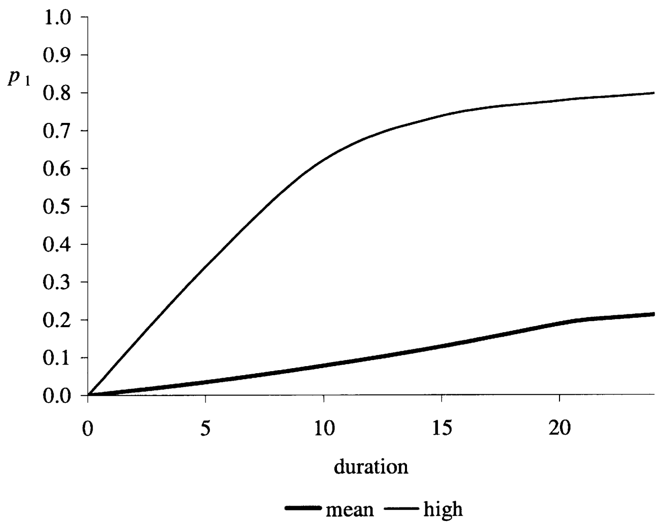

Figure 2. Mean and spread of apportionment to low-risk channel under standard assumptions

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

Table 2. Accuracy versus time: simulation method

Basis	C	F	G	H	I	J
J = number of simulations	Analytical	1,000	3,000	5,000	1,000	5,000
I = number of knots	6	6	6	8	10	10
RMSE	0.011	0.033	0.047	0.026	0.034	0.025
Average response time (min:sec)	2:50	1:51	5:07	16:10	3:48	36:48

bar one (denoted 'high' in the legend). The means of the lowest three apportionments bar one are zero at all durations. In the figure, the mean is shown in bold and the spread is from the horizontal axis to the fine curve. This gives a convenient indication of the level and dispersion of apportionments in respect of the observed sample for the purposes of the investigations described in the following sections.

8.5 *The Simulation Method*

8.5.1 The simulation method described in Section 7.2 was also applied, with the same parameters as those used for the analytical method. Here again the issue of time versus accuracy must be addressed; but in this case the response time is dependent, not only on the number of knots and the confidence limits, but also on the number of iterations J.

8.5.2 The RMSEs and average response times under the simulation method are shown in Table 2. As before, the RMSEs were calculated relative to the apportionments derived from the analytical method, using 12 knots between 99% confidence limits.

8.5.3 From Table 2, it may be seen that, in order to get an RMSE of less than 3%, the number of simulations required is of the order of 5,000.

8.6 *Passive Approach*

8.6.1 In order to establish the effect of the dynamic approach outlined above, it was compared with a passive approach, under which, for each subject at each calculation date, a single optimisation is performed. Under this approach, therefore, for the purpose of determining the recommended apportionment at a particular calculation date, no allowance is made for subsequent reapportionments; it is assumed that the recommended apportionment will remain effective until retirement.

8.6.2 Because of the effect of contributions payable after the calculation date, the analytical method under the passive approach would involve a distribution function that would not be normal over the optimisation period. This would result in an excessively unwieldy objective function. For the purposes of this approach the simulation method was therefore used.

8.6.3 Under this approach, instead of the procedure outlined in ¶¶7.1.4 to 7.1.12, after determining the parameters of the utility function (as in ¶7.1.4), we solve the problem:

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

$$\text{maximise } E\{u(B)\}$$

subject to:

$$p_k \geq 0 \text{ for all } k$$

and

$$\sum_{k=1}^{K} p_k = 1.$$

Note that $E\{u(B)\}$, which is defined in Section A.3, is no longer a conditional expectation, and p_k is no longer dependent on the time elapsed after the calculation date.

8.6.4 The results on the passive approach are compared with those on the dynamic approach in Figure 3. For this purpose, 5,000 simulations were made with eight knots between 99% confidence limits. As in Figure 2, the mean apportionment to the low-risk channel is shown as a bold curve, and the means of the highest three apportionments bar one as a fine curve. The means of the lowest three apportionments bar one are again zero at all durations for both approaches.

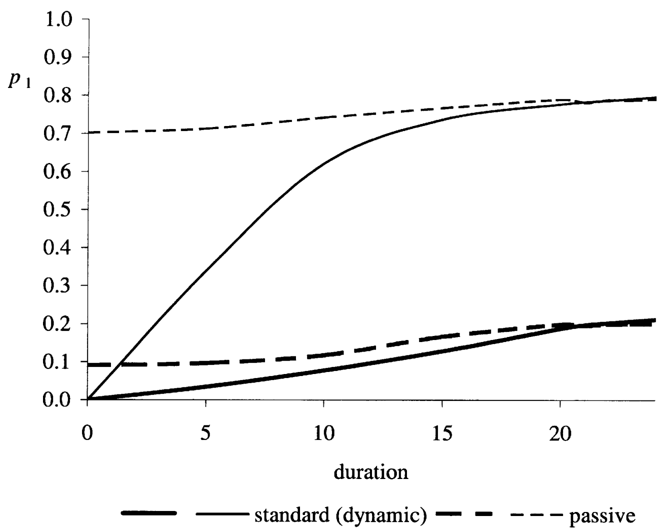

Figure 3. Mean and spread of apportionment to low-risk channel for active and passive approaches

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

8.6.5 As expected, the recommended allocations under the passive approach tended (with small errors due to simulation) to the same limits as those under the dynamic approach as the duration tended to the retirement date; but at shorter durations there were marked differences, particularly for subjects with high levels of risk aversion (such as subject 25 — cf. ¶8.8.1). On average, however, as might be expected, the recommended apportionment to channel 1 under the passive approach at each duration was commensurate with the average under the dynamic approach over the remaining durations; but, for subjects with high levels of risk aversion, the passive approach required much more substantial investment in that channel in earlier years.

8.7 *Downward Sloping Utility Functions*

8.7.1 As mentioned in ¶6.2.10, there may be circumstances in which, although $u_{i-1}(x)$ and $u_i(x)$ are monotonically increasing, $v_i(x)$ is not. This occurred in four cases. In each of those cases the problem arose from a discontinuity. In two of them the effect was so small that it did not result in negative values of $k_{m-1,t}$. In the other two, which are illustrated in Figures 4 and 5 (for subjects 28 and 48 respectively), the effect was more substantial.

8.7.2 It may be questioned, in a case such as that of subject 28, whether the interpolation of the elicited utility function is itself appropriate, giving, as it does, a quite noticeably downward sloping curve over quite a large portion of its domain. Where there is a short term to exit, it may happen that

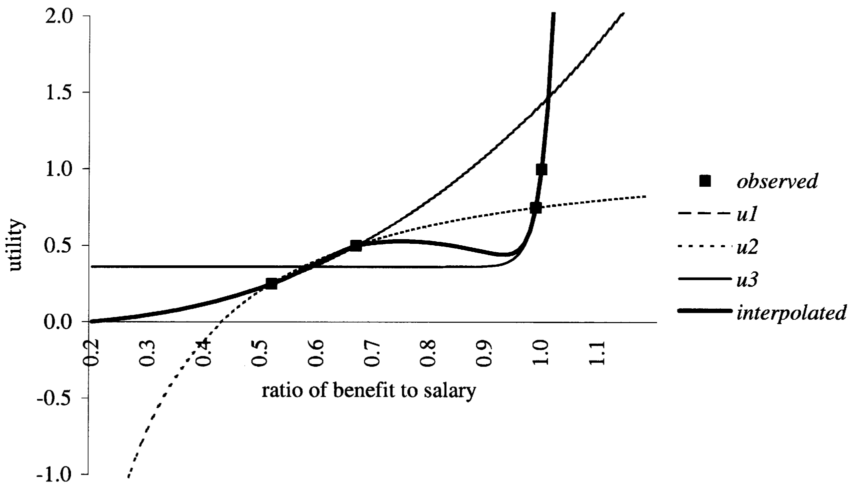

Figure 4. Interpolation of utility function: subject 28

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

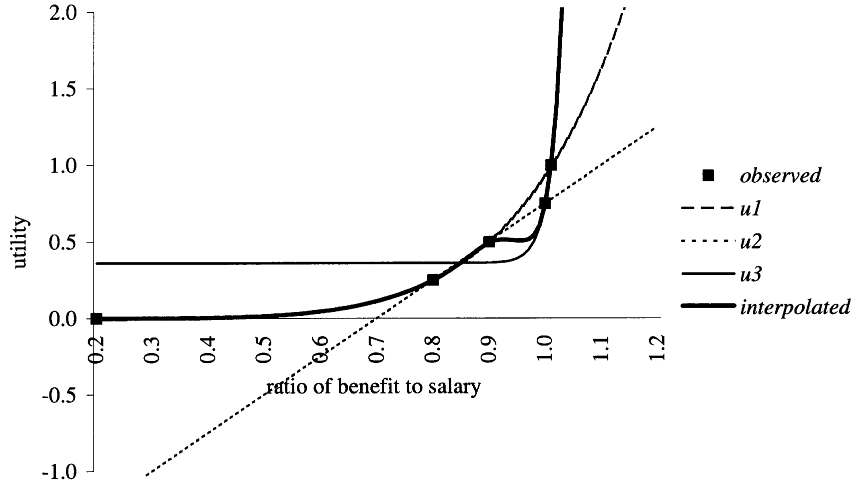

Figure 5. Interpolation of utility function: subject 48

the benefit will almost certainly fall within the range over which the utility function is downward sloping. This may result in widely fluctuating recommendations over the last few years. However, if the utility function was constant over the interval concerned, recommendations would be irrelevant anyway. So, in such cases widely fluctuating recommendations do not reflect widely fluctuating expected utilities. On the other hand, in the case of subject 48 it would be justifiable to regard the effect as immaterial. In practice, the resolution of the discontinuities as suggested in ¶6.2.1 might reduce the incidence of this effect. The discontinuities may also have arisen from framing effects, which could be avoided in the implementation of the system by avoiding reference to post-retirement income as a percentage of salary; but, even if it is not possible to avoid it, the fact remains that, as explained in ¶6.2.3, the average relative risk aversion over the interval (weighted by its denominator) is an unbiased approximation to its true value.

8.8 *Utility Functions Tending Strongly to 1*

8.8.1 There may be cases in which the elicited utility function tends very strongly to 1. An example (that of subject 25 in the sample) is shown in Figure 6.

8.8.2 If, as a result, the derived utility function is very close to 1 over the entire range for which it is calculated (say a 99% confidence limit), it may be difficult to solve the optimisation problem, as the gradient will be very close to zero. For such cases the apportionment is virtually a matter of

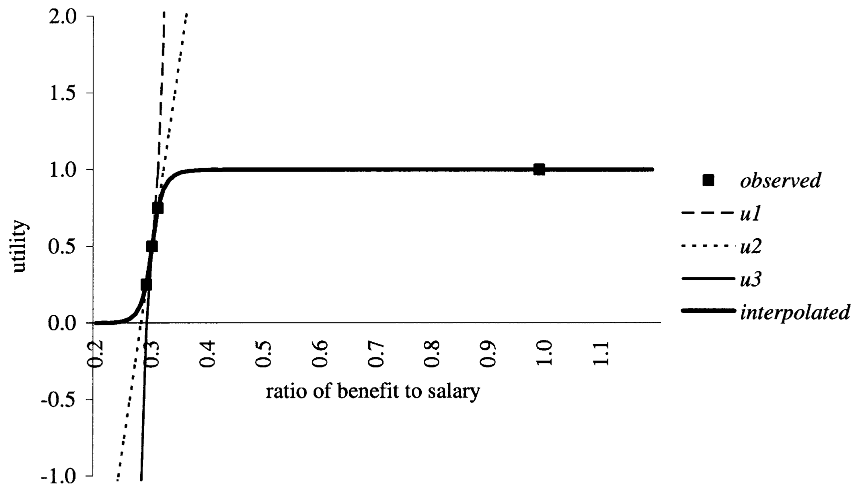

Figure 6. Interpolation of utility function: subject 25

indifference. However, if it is desired to make recommendations for such cases, it should be ensured that the termination condition for the optimisation problem is appropriately specified. Even a very low absolute gradient may give a recommended apportionment quite different from the optimum. In order to accommodate such cases, it is better to use a termination condition specified in terms of the absolute value of the gradient relative to the Hessian matrix, such as:

$$\left| \boldsymbol{g}' \boldsymbol{H}^{-1} \boldsymbol{g} \right|$$

where $\boldsymbol{g}$ is the gradient vector and $\boldsymbol{H}$ is the Hessian matrix, or to confine termination conditions to changes in the components of the apportionment vector $\boldsymbol{p}$.

8.9 *Sensitivity Tests*

8.9.1 In order to test the sensitivity of the recommended apportionments to the assumptions, results were obtained on various alternative sets of assumptions. One assumption was varied at a time, and the results are discussed below.

8.9.2 *Mean returns*

8.9.2.1 Figures 7 and 8 show the effects of changing the mean returns of the respective channels. In Figure 7 the mean return of channel 1 has been

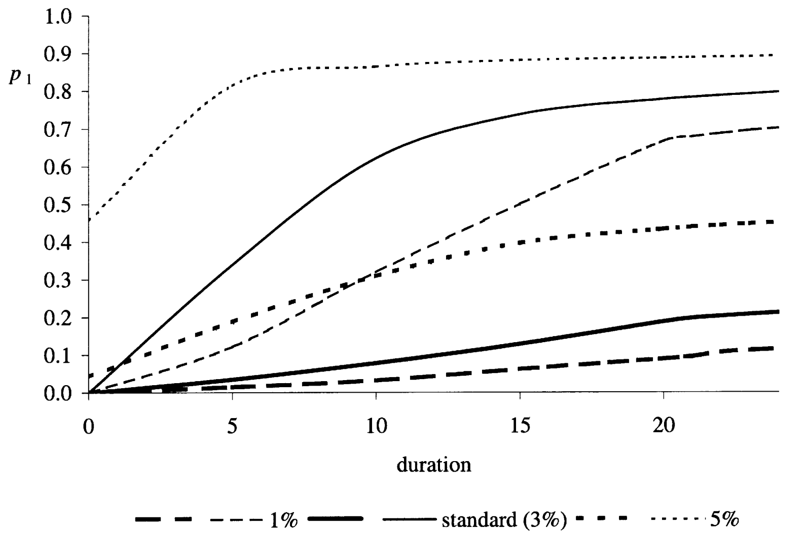

Figure 7. Mean and spread of apportionment to low-risk channel for various mean returns on channel 1

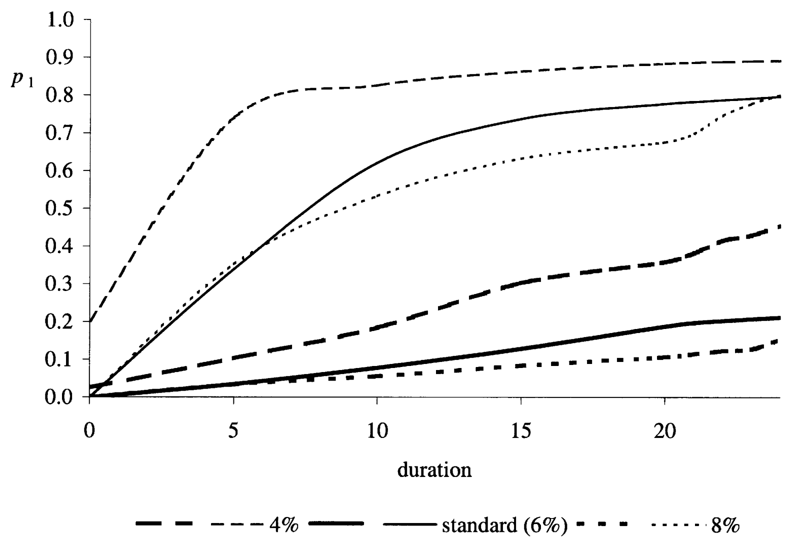

Figure 8. Mean and spread of apportionment to low-risk channel for various mean returns on high-risk channel

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

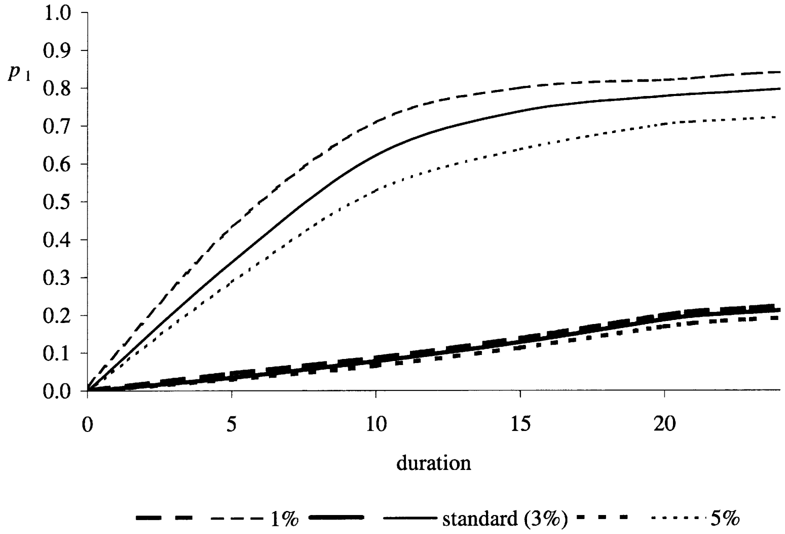

Figure 9. Mean and spread of apportionment to low-risk channel for various standard deviations on low-risk channel

varied, and in Figure 8 the mean return of channel 2. In these and the subsequent figures the same style of presentation is used as in Figures 2 and 3. The mean apportionment to the low-risk channel is shown as a bold curve, and the means of the highest three apportionments bar one as a fine curve. The means of the lowest three apportionments bar one are zero throughout.

8.9.2.2 The mean returns on the respective channels clearly have a marked effect on the apportionments. From Figures 7 and 8 it may be noted that, if the mean return on the low-risk channel is close to that on the high-risk channel, the advantages of diversification discussed in ¶8.4.2 may be insufficient to warrant the full investment of the initial contributions in the high-risk channel over the five-year period to the next review.

8.9.3 *Standard deviations of returns*

8.9.3.1 Figures 9 and 10 show the effects of changing the standard deviations of the returns of channels 1 and 2 respectively.

8.9.3.2 Figure 9 merely reflects the relatively small room for variation in the standard deviation of the low-risk channel. Figure 10 shows that the standard deviation of the high-risk channel can have just as important an effect on apportionments as its mean, though of course in the opposite direction.

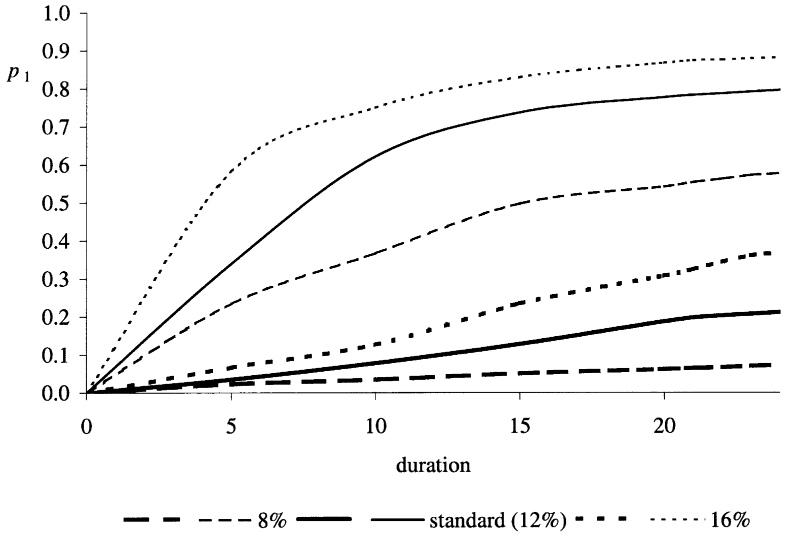

Figure 10. Mean and spread of apportionment to low-risk channel for various standard deviations on high-risk channel

8.9.4 *Correlation of returns*

8.9.4.1 Figure 11 shows the effect of changing the covariance of the returns. In that figure, the results for:

$$\boldsymbol{G} = \begin{pmatrix} 0.03 & 0.0125 \\ 0.0125 & 0.12 \end{pmatrix}$$

are compared with those on the standard basis, where:

$$\boldsymbol{G} = \begin{pmatrix} 0.03 & 0 \\ 0 & 0.12 \end{pmatrix}.$$

The covariance matrices are thus:

$$\boldsymbol{GG}' = \begin{pmatrix} 0.001056 & 0.001875 \\ 0.001875 & 0.014556 \end{pmatrix} \text{ and } \boldsymbol{GG}' = \begin{pmatrix} 0.0009 & 0 \\ 0 & 0.0144 \end{pmatrix}$$

respectively, giving correlation coefficients of 0.478 and zero respectively.

8.9.4.2 Figures 12 and 13 show the effects of increasing the autocorrelation of the returns of channels 1 and 2 respectively.

8.9.4.3 An increase in the covariance makes very little difference to the

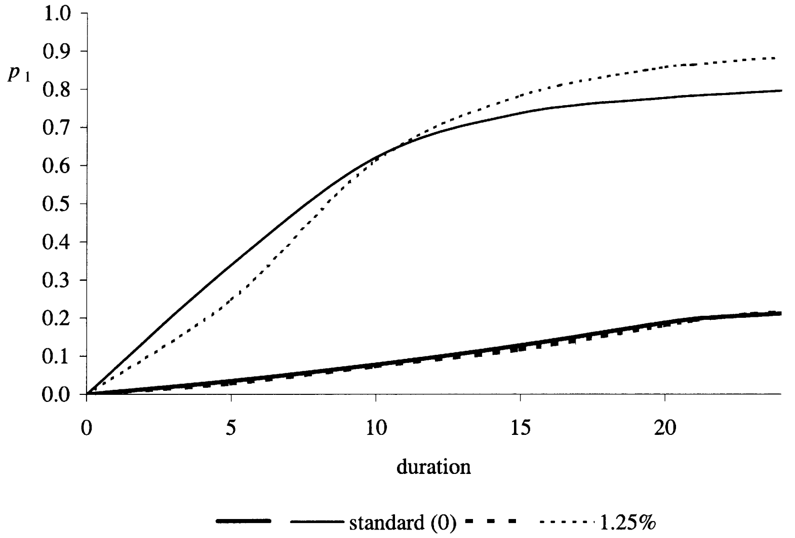

Figure 11. Mean and spread of apportionment to low-risk channel for alternative values of $\boldsymbol{G}_{12}$

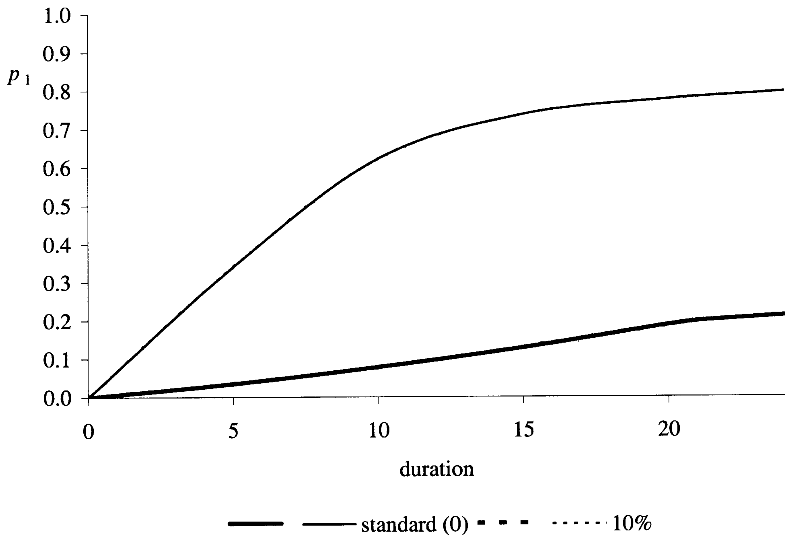

Figure 12. Mean and spread of apportionment to low-risk channel for autocorrelations on low-risk channel

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

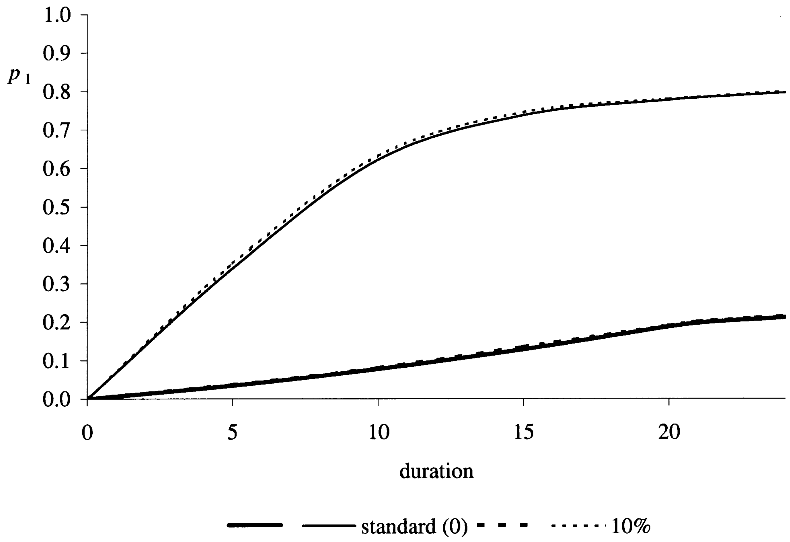

Figure 13. Mean and spread of apportionment to low-risk channel for autocorrelations on high-risk channel

mean apportionment; but for subjects with higher risk aversion the apportionment to the low-risk channel is reduced at early durations, reflecting the fact that the advantages of diversification are reduced relative to the effective asset represented by future contributions. At later durations the value of future contributions is reduced and the effect is reversed.

8.9.4.4 With regard to the effects of autocorrelation, the requirements of ¶3.3.4 should be noted. Those requirements imply that, the greater the autocorrelation of the returns, the greater should be the lengths of the periods between reviews of investment channel choices. For the sake of comparability, the lengths of the periods have not been changed for the purposes of Figures 12 and 13. However, for that reason the extent of the increase in autocorrelation has been restricted.

8.9.4.5 The effects of autocorrelations are virtually nil; only if the starting value of z_u (i.e. z_0) were non-zero would non-zero autocorrelations affect apportionments.

8.9.5 *Contribution rates and annuity factors*

8.9.5.1 Figure 14 shows the effect of changing the contribution rate. Because the annuity factor was assumed to be constant, the effect of changing that factor corresponds to that of an inversely proportionate change to the contribution rate.

8.9.5.2 The effect of reducing the contribution rate is to shift the

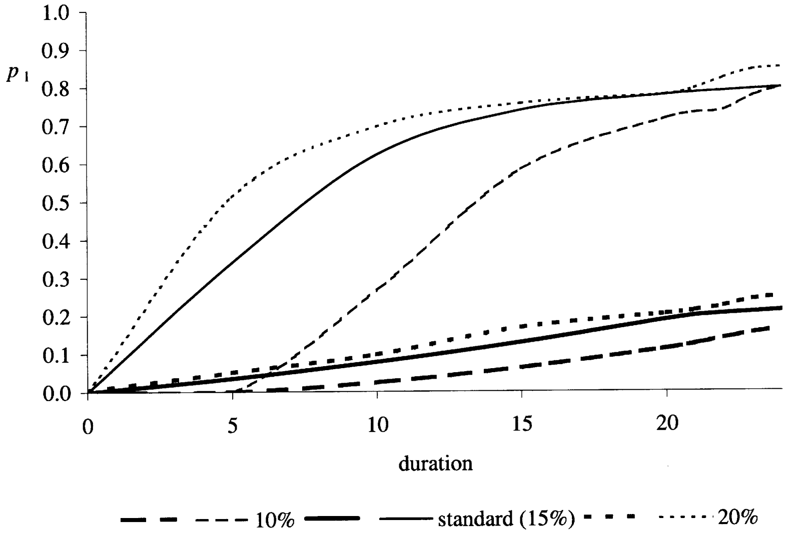

Figure 14. Mean and spread of apportionment to low-risk channel for various contribution rates

distribution of the benefit to correspondingly lower values. The result of this shift is to place greater emphasis on a subject's risk aversion over those values. For subjects that showed high risk aversion at higher levels of post-retirement income, the effect is a reduction in the apportionment to the low-risk channel. The converse applies to an increase in the contribution rate, though the effect is smaller.

9. Conclusion, Caveats and Further Research

9.1 *Conclusion*

9.1.1 As set out in Section 1, this paper has specified the design of an interactive system to recommend investment channel apportionments to members of defined contribution retirement funds so as to maximise their expected utility. Subject to constraints relating to the autocorrelation of returns in the investment model, the design allows for any degree of accuracy required. Higher degrees of accuracy are achieved at the expense of longer response times. The investment model may be a state-space model with normally distributed forces of return, or any other model that the user wishes to specify. In the former case an analytical method may be used for the determination of the objective function used in the optimisation process, resulting in a faster response time for any required degree of accuracy.

9.1.2 The application of the system was tested on a sample of subjects. Apart from subjects for which it was optimal to remain in the high-risk channel throughout, the recommended apportionment to the low-risk channel increased monotonically. The pattern of movement was highly dependent on the utility function of the subject, with regard to both the date of commencement and the pace of movement. In general, the commencing apportionment to the low-risk channel was zero, and in no case did it reach 100% at retirement.

9.1.3 The advantage of the dynamic approach used in the proposed system over a passive approach was demonstrated.

9.1.4 Sensitivity tests showed that, except for autocorrelation, the parameters of the model had marked effects on the apportionments, though not as marked as the variability between subjects.

9.2 *Caveats regarding the Implementation of the System*

9.2.1 The sample selected for this analysis was not a random sample, and there was some evidence of bias. Furthermore, there may have been framing effects in the elicitation of the utility functions of the subjects. Conclusions drawn from this sample should therefore be treated with circumspection. However, it can form the basis of the next step, which will be to introduce the system on a trial basis to selected retirement funds.

9.2.2 It is argued by Asher (1999) that the imposition of investment channel choice on the members of a retirement fund places an unnecessary burden on them, which should be borne by the trustees. In South Africa there is also considerable antagonism from trades unions against allowing investment channel choice to members of retirement funds. The omission from this article of discussion of this issue is not intended to imply that it can be ignored. Depending on the law of the country in which the system is implemented, consideration should be given to the issuance of caveats or disclaimers so as to avoid unnecessary liabilities to members.

9.2.3 In recent years, in addition to the concepts of 'descriptive' and 'normative' analyses (which seek respectively to explain and, as in this article, to recommend decisions), the concept of 'prescriptive' analyses has been propounded (e.g. Bell *et al*., 1988; French & Xie, 1994). Although, in the author's opinion, 'prescriptive' is an unfortunate epithet, the concept is useful. It envisages using decision theory as a process to assist the decision-maker to understand the analysis, and thereby to make a more informed decision. In discussions with Professor S. French, the author was strongly urged to ensure that, if a normative system was made available to the members of a retirement fund to make recommendations regarding investment channel choice, trained personnel should be available to assist in this process and to help members to understand the model being used and the basis on which the recommendation was made. The extent to which this will be possible may be limited by considerations of cost effectiveness.

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

9.3 *Further Research*

9.3.1 Once the system has been introduced on a trial basis, the way in which members of retirement funds use the system in making decisions can be analysed. Recommendations acted on can then form the basis of further research. In particular, it would be of interest to establish whether indicators of the shape of members' utility functions can be obtained from questions that may be easier for them to answer. Further research into alternative elicitation methods in the context of the proposed system may also be fruitful.

9.3.2 Other areas for further research include:

— methods of allowing for autocorrelation of returns in the investment model without excessive computing time;

— the development of a stochastic model of salaries for individual members of retirement funds; and

— the application of the consumption problem to the development of systems for the recommendation of investment channel choice, particularly for the purposes of optimising retirement dates and voluntary additional contribution rates.

Acknowledgements

The author acknowledges with thanks the help of Professors S. French, L.P. Fatti and A. Asher and Mr A.J. Maitland with regard to certain aspects of this article, the helpful criticism of the Journal's scrutineers, the work of the students who returned completed questionnaires and the comments of participants in seminars at the Universities of the Witwatersrand and Cape Town and in Harare.

References

Aitken, W.H. (1994). *A problem-solving approach to pension funding and valuation*. Actex, Winstead.

Anderson, J.R., Dillon, J.L. & Hardaker, B. (1977). *Agricultural decision analysis*. Iowa State University Press, Ames; cited in Farquhar (1984).

Asher, A. (1999). Loading retirement fund members with investment channel choice. *Journal of Pensions Management*, **5** (1), 76-84.

Bell, D.E., Raiffa, H. & Tversky, A. (Eds.) (1988). *Decision making: descriptive, normative and prescriptive interactions*. Cambridge. Cited in French & Xie (1994).

Bicksler, J.L. (1974). Theory of portfolio choice and capital market behavior: an introductory survey. In Bicksler & Samuelson (1974), 1-25.

Bicksler, J.L. & Samuelson, P.A. (Eds.) (1974). *Investment portfolio decision-making*. Lexington Books, Lexington, Mass.

Blake, D., Cairns, A.J.G. & Dowd, K. (2001). Pensionmetrics: stochastic pension plan design and value-at-risk during the accumulation phase. *Insurance: Mathematics and Economics*, **29**, 187-215.

Booth, P.M. (1995). The management of investment risk for defined contribution pension schemes. *Transactions of the International Congress of Actuaries*, **25** (3), 105-122.

Booth, P.M. & Ong, A.S.K. (1994). A simulation based approach to asset allocation. *Proceedings of the 4th AFIR International Colloquium*, **1**, 217-239.

Booth, P. & Yakoubov, Y. (2000). Investment policy for defined contribution pension scheme members close to retirement: an analysis of the 'lifestyle' concept. *North American Actuarial Journal*, **4** (2), 1-19.

Boulier, J.-F., Huang, S.-J. & Taillard, G. (1999). Optimal management under stochastic interest rates: the case of a protected pension fund. *Proceedings of the 3rd IME Conference, London*, **2**, 1-23.

Boulier, J.-F., Michel, S. & Wisnia, V. (1996). Optimizing investment and contributions policies of a defined benefit pension fund. *Proceedings of the 6th AFIR International Colloquium*, **1**, 593-607.

Boulier, J.-F., Trussant, E. & Florens, D. (1995). A dynamic model for pension funds management. *Proceedings of the 5th AFIR International Colloquium*, **1**, 361-384.

Bowers, N.L., Gerber, H.U., Hickman, J.J., Jones, D.A. & Nesbitt, C.J. (1986). *Actuarial mathematics.* Society of Actuaries, Itasca.

Cairns, A.J.G. (1995). Pension funding in a stochastic environment: the role of objectives in selecting an asset-allocation strategy. *Proceedings of the 5th AFIR International Colloquium*, **1**, 429-453.

Cairns, A.J.G. (1996). Continuous-time stochastic pension fund modelling. *Proceedings of the 6th International AFIR Colloquium*, **1**, 609-624.

Cairns, A.J.G. (1997). A comparison of optimal and dynamic control strategies for continuous-time pension fund models. *Proceedings of the 7th AFIR International Colloquium*.

Cairns, A.J.G., Blake, D. & Dowd, K. (2000). Optimal dynamic asset allocation for defined-contribution pension plans. *Proceedings of the 10th AFIR International Colloquium,* Tromsø, 131-154.

Cairns, A.J.G. & Parker, G. (1997). Stochastic pension fund modelling. *Insurance: Mathematics and Economics*, **21**, 43-79.

Chang, S.-C. (1999). Optimal pension funding through dynamic simulations: the case of Taiwan public employees retirement system. *Insurance: Mathematics and Economics*, **24**, 187-199.

Cox, J.C. & Huang, C.-F. (1989). Optimum consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, **49**, 33-83.

Deelstra, G., Grasselli, M. & Koehl, P.-F. (1999). Optimal investment strategies in a CIR framework. Cited as preprint in Cairns, Blake & Dowd (2000).

Eskow, E. & Schnabel, R.B. (1991). Algorithm 695: software for a new modified Cholesky factorization. *ACM Trans. Math. Software*, **17**, 306-312.

Exley, C.J., Mehta, S.J.B. & Smith, A.D. (1997). The financial theory of defined benefit pension schemes. *British Actuarial Journal*, **3**, 835-966.

Faculty & Institute of Actuaries (2001). Core reading: subject 109, financial economics. Faculty & Institute of Actuaries.

Farquhar, P.H. (1984). Utility assessment methods. *Management Science*, **30**, 1283-1294.

French, S. (1986). *Decision theory: an introduction to the mathematics of rationality*. Wiley, New York.

French, S. (unpublished). Correspondence with the author 5/7/1998.

French, S., Allatt, P., Slater, J.B., Vassiloglou, M. & Willmot, A.S. (1992). Implementation of a decision analytic aid to support examiners judgements in aggregating components. *British Journal of Mathematical and Statistical Psychology*, **45**, 75-91.

French, S. & Xie, Z. (1994). A perspective on recent developments in utility theory. In Rios (1994).

Friedman, M. & Savage, L.J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, **56**, 279-304.

Gerber, H.U. & Shiu, E.S.W. (2000). Investing for retirement: optimal capital growth and dynamic asset allocation. *North American Actuarial Journal,* **4** (2), 42-62.

Haberman, S. & Sung, J.-H. (1994). Dynamic approaches to pension funding. *Insurance: Mathematics and Economics,* **15**, 151-162.

Hakansson, N.H. (1969). Optimal investment and consumption strategies under risk, an uncertain lifetime, and insurance. *International Economic Review*, **10** (3), 443-466.

Hakansson, N.H. (1970). Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica*, **38**, 587-607.

Hershey, J.C., Kunreuther, H.C. & Schoemaker, P.J.H. (1982). Sources of bias in assessment procedures for utility functions. *Management Science*, **28**, 936-954.

Ingersoll, J.E. (1987). *Theory of financial decision making*. Rowman & Littlefield, Savage, Maryland.

Josa-Fombellida, R. & Rincón-Zapatero, J. (2001). Minimization of risks in pension funding by means of pension contributions and portfolio selection. *Insurance: Mathematics and Economics*, **29**, 35-45.

Karmarkar, U.S. (1978). Subjective weighted utility: a descriptive extension of the expected utility model. *Organizational Behavior and Human Performance*, **21**, 61-72.

Keeney, R.L. & Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value trade-offs*. Wiley, New York.

Khorasanee, M.Z. (1995). Simulation of investment returns for a money purchase fund. *Journal of Actuarial Pratice*, **3** (1), 93-115.

Khorasanee, M.Z. & Smith, D.A. (1997). A utility maximisation approach to individual investment choices in a money purchase pension scheme. *Proceedings of the 1997 Investment Conference*, **2**, 261-289.

Knox, D.M. (1993). A critique of defined contribution plans using a simulation approach, *Journal of Actuarial Pratice*, **1** (2), 49-66.

Krzysztofowicz, R. & Duckstein, L. (1980). Assessment errors in multiattribute utility functions. *Organizational Behvior and Human Performance*, **26**, 326-348.

Ludvik, P.M. (1994). Investment strategy for defined contribution plans. *Proceedings of the 4th AFIR International Colloquium,* **3**, 1389-1400.

Markowitz, H.M. (1952). Portfolio selection. *Journal of Finance*, **7**, 77-91.

Merton, R.C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics*, **51**, 247-257.

Merton, R.C. (1971). Optimal consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, **3**, 373-413.

Mossin, J. (1968). Optimal multiperiod portfolio policies. *Journal of Business*, **41**, 215-229.

Mosteller, F. & Nogee, P. (1951). An experimental measure of utility. *Journal of Political Economy*, **59**, 371-404.

Novick, M.R., Dekeyrel, D.F. & Chuang, D.T. (1981). Local and regional coherence utility assessment procedures. *Bayesian Statistics, Proceedings of the First International Meeting,* University Press, Valencia, 557-568.

O'Brien, T. (1986). A stochastic-dynamic approach to pension funding. *Insurance: Mathematics and Economics*, **5**, 141-146.

O'Brien, T. (1987). A two-parameter family of pension contribution functions and stochastic optimization. *Insurance: Mathematics and Economics*, **6**, 129-134.

Phelps, E.S. (1962). The accumulation of risky capital: a sequential utility analysis. *Econometrica*, **30**, 729-743.

Pratt, J.W. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122-136.

Quiggin, J. (1993). *Generalized expected utility theory*. Kluwer, Boston.

Richard, S.F. (1975). Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model. *Journal of Financial Economics*, **2**, 187-203.

RIOS, S. (ed.) (1994). *Decision theory and decision analysis: trends and challenges*. Kluwer, Dordrecht.
SAMUELSON, P.A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, **51**, 239-246.
SAVAGE, L.J. (1954). *The foundations of statistics*. Wiley; 2nd ed., Dover, New York (1972).
SCHLAIFER, R.O. (1969). *Analysis of decisions under uncertainty*. McGraw-Hill, New York.
SCHOEMAKER, P.J.H. (1980). *Experiments on decisions under risk: the expected utility hypothesis*. Nijhoff, Boston.
SHARPE, W.F. & TINT, L.G. (1990). Liabilities: a new approach. *Journal of Portfolio Management,* 5-10.
SHERRIS, M. (1992). Portfolio selection and matching: a synthesis. *Journal of the Institute of Actuaries*, **119**, 87-105.
SIEGMANN, A.H. & LUCAS, A. (1999). Continuous-time dynamic programming for ALM with risk-averse loss functions. *Proceedings of the 9th AFIR International Colloquium*, **2**, 183-193.
THOMSON, R.J. (1999). Non-parametric likelihood enhancements to parametric graduations. *British Actuarial Journal*, **5**, 197-236.
THOMSON, R.J. (2000). An analysis of the utility functions of members of retirement funds. *Proceedings of the 10th AFIR International Colloquium,* 615-630.
THOMSON, R.J. (2003). The use of utility functions for investment channel choice in defined contribution retirement funds: part I. *British Actuarial Journal*, **9**, 653-709.
VAN DAM, C. (1973). *Beslissen in onsekerheid.* Stenfert Kroese, Leiden.
VIGNA, E. & HABERMAN, S. (2001). Optimal investment strategy for defined contribution pension schemes. *Insurance: Mathematics and Economics*, **28**, 233-62.
VON NEUMANN, J. & MORGENSTERN, O. (1947) *Theory of games and economic behaviour*, 2nd ed. Princeton.
WILKIE, A.D. (1985). Portfolio selection in the presence of fixed liabilities: a comment on 'the matching of assets to liabilities'. *Journal of the Institute of Actuaries*, **112**, 229-278.
WILKIE, A.D. (1986). A stochastic investment model for actuarial use. *Transactions of the Faculty of Actuaries*, **39**, 341-403.
WINKLER, R.L. (1967). The quantification of judgment: some methodological suggestions. *Journal of the American Statistical Association,* **62**, 1105-20. Reprinted as Chap. 9 in Bicksler & Samuelson (1974).
WISE, A.J. (1984a). A theoretical analysis of the matching of assets to liabilities. *Journal of the Institute of Actuaries*, **111**, 375-402.
WISE, A.J. (1984b). The matching of assets to liabilities. *Journal of the Institute of Actuaries*, **111**, 445-502.
WISE, A.J. (1987a). Matching and portfolio selection: part 1. *Journal of the Institute of Actuaries*, **114**, 113-133.
WISE, A.J. (1987b). Matching and portfolio selection: part 2. *Journal of the Institute of Actuaries*, **114**, 551-568.
ZIEMBA, W.T. & MULVEY, J.M. (eds.) (1998). *Worldwide asset and liability modeling.* Cambridge Univ. Press, Cambridge.

APPENDIX A

THE OBJECTIVE FUNCTIONS

This appendix sets out the derivation of the objective functions, as well as the corresponding gradient vectors and Hessian matrices, that are used for the purpose of optimisation.

A.1 *Dynamic Approach: Analytical Method*

A.1.1 In the dynamic programming, as described in Section 7.1, it is necessary to interpolate the derived utility function $\tilde{u}_m(x)$ between (and extrapolate it beyond) a small number of determined values. The interpolation of the derived utility function is similar (with some exceptions noted in the following paragraphs) to that of the elicited utility function described in Section 6.2. The formula used for this interpolation is:

$$\tilde{u}_m(x) = v_{m\iota}(x) \text{ for } x \in S_{m\iota}\ ,\ \iota = 1, \ldots, I+1$$

where:

$$\begin{aligned}
S_{m\iota} &= (0, x_{m1}] && \text{for } \iota = 1 \\
&\quad [x_{m,\iota-1}, x_{m\iota}] && \text{for } \iota = 2, \ldots, I \\
&\quad [x_{mI}, \infty) && \text{for } \iota = I+1 \\
v_{m\iota}(x) &= \tilde{u}_{m1}(x) && \text{for } \iota = 1 && \text{(A1)} \\
&\quad \frac{(x_{m\iota} - x)\tilde{u}_{m,\iota-1}(x) + (x - x_{m,\iota-1})\tilde{u}_{m\iota}(x)}{x_{m\iota} - x_{m,\iota-1}} && \text{for } \iota = 2, \ldots, I && \text{(A2)} \\
&\quad \tilde{u}_{mI}(x) && \text{for } \iota = I+1 && \text{(A3)}
\end{aligned}$$

and, for $\iota = 1, \ldots, I$:

$$\begin{aligned}
\tilde{u}_{m\iota}(x) &= a_{m\iota}h_{m\iota}(x) + b_{m\iota} && && \text{(A4)} \\
h_{m\iota}(x) &= \ln x && \text{if } k^{\#}_{m\iota} = 1 \\
&\quad x^{n_{m\iota}} && \text{otherwise.} \\
n_{m\iota} &= \frac{\ln k_{m\iota}}{\ln w_m} \\
k^{\#}_{m\iota} &= 1 && \text{if } n_{m\iota} = 0 \text{ and } k^{*}_{m\iota} = 1 \\
&\quad 0 && \text{otherwise.}
\end{aligned}$$

$$
\begin{aligned}
k^*_{m\iota} &= 1 && \text{if } \tilde{u}_m(x_{m,\iota-1}) < \tilde{u}_m(x_{m\iota}) < \tilde{u}_m(x_{m,\iota+1})\\
&\quad 0 && \text{otherwise.}\\
k_{m\iota} &= \frac{\tilde{u}_m(x_{m,\iota+1}) - \tilde{u}_m(x_{m\iota})}{\tilde{u}_m(x_{m\iota}) - \tilde{u}_m(x_{m,\iota-1})} && \text{if } k^*_{m\iota} = 1\\
&\quad 1 && \text{otherwise.} \qquad \text{(A5)}\\
w_m &= \left(\frac{x_{m,I+1}}{x_{m,0}}\right)^{\frac{1}{I+1}}\\
a_{m\iota} &= \frac{\tilde{u}_m(x_{m\iota}) - \tilde{u}_m(x_{m,\iota-1})}{h_{m\iota}(x_{m\iota}) - h_{m\iota}(x_{m,\iota-1})} && \text{if } k_{m\iota} > 0,\ \iota < I\\
&\quad \frac{\tilde{u}_m(x_{m,I+1}) - \tilde{u}_m(x_{mI})}{h_{mI}(x_{m,I+1}) - h_{mI}(x_{m,I})} && \text{if } k_{mI} > 0,\ \iota = I\\
&\quad 0 && \text{otherwise.}\\
b_{m\iota} &= \tilde{u}_m(x_{m\iota}) - a_{m\iota}h_{m\iota}(x_{m\iota}) && \text{for } \iota = 1, \ldots, I.
\end{aligned}
$$

A.1.2 Note that, in the case of the derived utility function (unlike the case of the elicited utility function), the values of $x_{m\iota}$ are geometrically spaced. This means that w_m is constant for all ι, and $n_{m\iota}$ may be defined in closed form, so that the iterative method of ¶6.2.5 is unnecessary. The values of $x_{m\iota}$ are given in ¶7.1.9.

A.1.3 As mentioned in ¶6.2.10, the interpolated version of the elicited utility function may be downward sloping over part of its domain. This means that, for $m = N$ (and similarly for lower values of m), there may be values of ι for which:

$$\tilde{u}_m(x_{m,\iota+1}) < \tilde{u}_m(x_{m\iota}).$$

Furthermore, for $n_{m\iota} \ll 0$, $\tilde{u}_{m\iota}(x)$ may tend so strongly to a constant limit that, to the levels of accuracy of which the computer is capable, calculated values are indistinguishable from that value. In the first case, but for the distinction drawn for $k^*_{m\iota} \neq 1$ in equation (A5), $k_{m\iota}$ would be negative for certain values of ι and in the second it would be zero. In either case $n_{m\iota}$ would be indeterminate. In order to avoid this effect, such cases are distinguished by means of the Boolean variables $k^{\#}_{m\iota}$ (for the purposes of the definition of $h_{m\iota}(\bullet)$) and $k^*_{m\iota}$ (for the purposes of the definitions of $k_{m\iota}$, $a_{m\iota}$ and $b_{m\iota}$).

A.1.4 In the first case, the justification of this procedure is discussed in Section 9.3. In the second case, it is justified by the fact that, if for some ι:

$$\tilde{u}_m(x_{m,\iota+1}) = \tilde{u}_m(x_{m\iota})$$

then it is appropriate to assume that:

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

$a_{m\iota} = 0$;
$n_{m\iota} = 0$; and
$h_{m\iota} = x^{n_{m\iota}}$;

so that:

$$\tilde{u}_{m\iota}(x) = a_{m\iota}x^{n_{m\iota}} + b_{m\iota}$$
$$= b_{m\iota}$$

i.e. that $\tilde{u}_{m\iota}(x)$ is constant, which is the result achieved by this procedure.

A.1.5 From Section A.1.1 we have (ignoring the subscript m):

$$v_\iota(x) = (\alpha_{1\iota}x + \beta_{1\iota})h_{\iota-1}(x) + (\alpha_{2\iota}x + \beta_{2\iota})h_\iota(x) + (\gamma_\iota x + \delta_\iota) \text{ for } \iota = 1, \ldots, I+1$$

where:

$\alpha_{1\iota} = 0$	for $\iota = 1, I+1$
$-\dfrac{a_{\iota-1}}{x_\iota - x_{\iota-1}}$	otherwise.
$\alpha_{2\iota} = 0$	for $\iota = 1, I+1$
$\dfrac{a_\iota}{x_\iota - x_{\iota-1}}$	otherwise.
$\beta_{1\iota} = 0$	for $\iota = 1$
$\dfrac{a_{\iota-1}x_\iota}{x_\iota - x_{\iota-1}}$	for $\iota = 2, \ldots, I$
a_I	for $\iota = I+1$.
$\beta_{2\iota} = a_1$	for $\iota = 1$
$-\dfrac{a_\iota x_{\iota-1}}{x_\iota - x_{\iota-1}}$	for $\iota = 2, \ldots, I$
0	for $\iota = I+1$.
$h(x) = \ln x$	if $k^{\#}_{m\iota} = 1$
x^{n_ι}	otherwise.
$\gamma_\iota = 0$	for $\iota = 1, I+1$
$\dfrac{b_\iota - b_{\iota-1}}{x_\iota - x_{\iota-1}}$	otherwise.
$\delta_\iota = b_1$	for $\iota = 1$
$\dfrac{b_{\iota-1}x_\iota - b_\iota x_{\iota-1}}{\Delta x_{\iota-1}}$	for $\iota = 2, \ldots, I$
b_I	for $\iota = I+1$.

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

A.1.6 For the analytical method, as stated in ¶7.1.11, the objective function is:

$$E\{u(P)|B(m-1)=x\}=\sum_{\iota=1}^{I+1}E_{m\iota} \quad \text{(A6)}$$

where:

$$E_{m\iota}=\frac{1}{\sqrt{2\pi}\sigma(m)}\int_{l_\iota}^{s_\iota}\left\{\sum_{r=1}^{2}\left(\alpha_{r\iota}e^y+\beta_{r\iota}\right)h_{\iota+r-2}(e^y)+\left(\gamma_\iota e^y+\delta_\iota\right)\right\}e^{-\frac{1}{2}\left(\frac{y-\mu(m)}{\sigma(m)}\right)^2}dy \quad \text{(A7)}$$

$$l_\iota=\ln\inf S_{m\iota} \quad \text{(A8)}$$

$$s_\iota=\ln\sup S_{m\iota} \quad \text{(A9)}$$

and $\mu(m)$ and $\sigma^2(m)$ are as defined in equations (34) and (35), with:

$$x=x_{m-1,\iota}.$$

A.1.7 It may be shown that (A7) reduces to:

$$E_{m\iota}=\sum_{r=1}^{10}\eta_{r\iota}\Delta_{r\iota}$$

where:

$$\Delta_{r\iota}=\psi_r(s_\iota^*-\sigma_r)-\psi_r(l_\iota^*-\sigma_r)$$

$$l_\iota^*=\frac{l_\iota-\mu}{\sigma}$$

$$s_\iota^*=\frac{s_\iota-\mu}{\sigma}$$

and $\eta_{r\iota}$, $\psi_r(\bullet)$ and σ_r are as shown in Table A1. In that table:

$$n_t\equiv n_{\iota-1} \text{ for } r \text{ odd}$$
$$n_\iota \text{ for } r \text{ even.}$$

A.1.8 From Table A1, we may generalise $\eta_{r\iota}$ as follows:

$$\eta_{r\iota}=\rho_{r\iota}\chi_{r\iota}\exp(\tau_{r\iota})$$

where:

$\chi_{r\iota}=\theta_{r\iota}\mu+\lambda_{r\iota}\sigma+\xi_{r\iota}\sigma^2+\zeta_{r\iota}$;

$\tau_{r\iota}=\varepsilon_{r\iota}\mu+\frac{1}{2}\varepsilon_{r\iota}^2\sigma^2$; and

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

Table A1. Components of E_i

	$n_\iota = 0$			$n_\iota \neq 0$		
r	$\eta_{r\iota}$	$\psi_r(x)$	σ_r	$\eta_{r\iota}$	$\psi_r(x)$	σ_r
1	$\alpha_{1\iota}\sigma e^{\mu+\frac{1}{2}\sigma^2}$	$-\phi(x)$	σ	$\alpha_{1\iota}e^{(n_\iota+1)\mu+\frac{1}{2}(n_\iota+1)^2\sigma^2}$	$\Phi(x)$	$(n_\iota+1)\sigma$
2	$\alpha_{2\iota}\sigma e^{\mu+\frac{1}{2}\sigma^2}$	$-\phi(x)$	σ	$\alpha_{2\iota}e^{(n_\iota+1)\mu+\frac{1}{2}(n_\iota+1)^2\sigma^2}$	$\Phi(x)$	$(n_\iota+1)\sigma$
3	$\alpha_{1\iota}(\mu+\sigma^2)e^{\mu+\frac{1}{2}\sigma^2}$	$\Phi(x)$	σ	0	0	0
4	$\alpha_{2\iota}(\mu+\sigma^2)e^{\mu+\frac{1}{2}\sigma^2}$	$\Phi(x)$	σ	0	0	0
5	$\beta_{1\iota}\sigma$	$-\phi(x)$	0	$\beta_{1\iota}e^{n_\iota\mu+\frac{1}{2}n_\iota^2\sigma^2}$	$\Phi(x)$	$n_\iota\sigma$
6	$\beta_{2\iota}\sigma$	$-\phi(x)$	0	$\beta_{2\iota}e^{n_\iota\mu+\frac{1}{2}n_\iota^2\sigma^2}$	$\Phi(x)$	$n_\iota\sigma$
7	$\beta_{1\iota}\mu$	$\Phi(x)$	0	0	0	0
8	$\beta_{2\iota}\mu$	$\Phi(x)$	0	0	0	0
9	$\gamma_\iota e^{\mu+\frac{1}{2}\sigma^2}$	$\Phi(x)$	σ	$\gamma_\iota e^{\mu+\frac{1}{2}\sigma^2}$	$\Phi(x)$	σ
10	δ_ι	$\Phi(x)$	0	δ_ι	$\Phi(x)$	0

$\rho_{r\iota}$, $\theta_{r\iota}$, $\lambda_{r\iota}$, $\xi_{r\iota}$, $\zeta_{r\iota}$ and $\varepsilon_{r\iota}$ are defined with reference to the definition of $\eta_{r\iota}$ in that table.

A.1.9 The jth component of the gradient vector is:

$$\sum_{\iota=1}^{I+1} D_j E_{m\iota}$$

where D_j is the partial derivative operator with respect to p_j, and

$$D_j E_{m\iota} = \sum_{r=1}^{10} D_j \eta_{r\iota} \Delta_{r\iota} + \eta_{r\iota} D_j \Delta_{r\iota}$$

$$D_j \Delta_{r\iota} = \psi_r'(s_\iota^* - \sigma_r)(D_j s_\iota^* - D_j \sigma_r) - \psi_r'(l_\iota^* - \sigma_r)(D_j l_\iota^* - D_j \sigma_r)$$

$$D_j \eta_{r\iota} = \rho_{r\iota}(D_j \chi_{r\iota} + \chi_{r\iota} D_j \tau_{r\iota}) \exp(\tau_{r\iota})$$

$$D_j \chi_{r\iota} = \theta_{r\iota} D_j \mu + \lambda_{r\iota} D_j \sigma + \xi_{r\iota} D_j \sigma^2$$

$$D_j \tau_{r\iota} = \varepsilon_{r\iota} D_j \mu + \tfrac{1}{2}\varepsilon_{r\iota}^2 D_j \sigma^2$$

$$D_j \sigma_r = \varepsilon_{r\iota} D_j \sigma$$

$$D_j l_\iota^* = -\frac{1}{\sigma^2}\left\{\sigma D_j \mu + (l_\iota - \mu) D_j \sigma\right\}$$

$$D_j s_\iota^* = -\frac{1}{\sigma^2}\left\{\sigma D_j \mu + (s_\iota - \mu) D_j \sigma\right\}$$

$$D_j \mu = \mu_j$$

$$D_j \sigma = \frac{1}{2\sigma} D_j \sigma^2$$

$$D_j \sigma^2 = 2\sum_{k=1}^{K} p_k \sigma_{jk}.$$

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

A.1.10 Element (j, k) of the Hessian matrix is:

$$\sum_{\iota=1}^{I+1} D_{jk} E_{m\iota}$$

where D_{jk} is the second-order partial derivative operator with respect to p_j and p_k, and:

$$D_{jk}E_{m\iota} = \sum_{r=1}^{10} D_{jk}\eta_{r\iota}\Delta_{r\iota} + D_j\eta_{r\iota}D_k\Delta_{r\iota} + D_k\eta_{r\iota}D_j\Delta_{r\iota} + \eta_{r\iota}D_{jk}\Delta_{r\iota}$$

$$D_{jk}\Delta_{r\iota} = \psi_r''(s_\iota^* - \sigma_r)(D_j s_\iota^* - D_j\sigma_r)(D_k s_\iota^* - D_k\sigma_r) + \psi_r'(s_\iota^* - \sigma_r)(D_{jk}s_\iota^* - D_{jk}\sigma_r)$$
$$\quad - \psi_r''(l_\iota^* - \sigma_r)(D_j l_\iota^* - D_j\sigma_r)(D_k l_\iota^* - D_k\sigma_r) - \psi_r'(l_\iota^* - \sigma_r)(D_{jk}l_\iota^* - D_{jk}\sigma_r)$$

$$D_{jk}\eta_{r\iota} = \rho_{r\iota}(D_{jk}\chi_{r\iota} + \chi_{r\iota}D_{jk}\tau_{r\iota} + D_j\chi_{r\iota}D_k\tau_{r\iota} + D_k\chi_{r\iota}D_j\tau_{r\iota} + \chi_{r\iota}D_j\tau_{r\iota}D_k\tau_{r\iota})\exp(\tau_{r\iota})$$

$$D_{jk}\chi_{r\iota} = \lambda_{r\iota}D_{jk}\sigma + \xi_{r\iota}D_{jk}\sigma^2$$

$$D_{jk}\tau_{r\iota} = \tfrac{1}{2}\varepsilon_{r\iota}^2 D_{jk}\sigma^2$$

$$D_{jk}\sigma_r = \varepsilon_{r\iota}D_{jk}\sigma$$

$$D_{jk}l_\iota^* = \frac{1}{\sigma^3}\left[\sigma\left\{D_j\mu D_k\sigma + D_k\mu D_j\sigma - (l_\iota - \mu)D_{jk}\sigma\right\} + 2(l_\iota - \mu)D_j\sigma D_k\sigma\right]$$

$$D_{jk}s_\iota^* = \frac{1}{\sigma^3}\left[\sigma\left\{D_j\mu D_k\sigma + D_k\mu D_j\sigma - (s_\iota - \mu)D_{jk}\sigma\right\} + 2(s_\iota - \mu)D_j\sigma D_k\sigma\right]$$

$$D_{jk}\sigma = \frac{1}{2\sigma^2}\left(\sigma D_{jk}\sigma^2 - \tfrac{1}{2}D_j\sigma^2 D_k\sigma^2\right)$$

$$D_{jk}\sigma^2 = 2\sigma_{jk}.$$

A.2 *Dynamic Approach: Simulation Method*

A.2.1 For the simulation method the objective function is

$$E\{u(P)|B(m-1) = x\} = \frac{1}{J}\sum_{i=1}^{J} u_i \tag{A10}$$

where:

$$u_i = (\alpha_{1\iota}z_i + \beta_{1\iota})h_{\iota-1} + (\alpha_{2\iota}z_i + \beta_{2\iota})h_\iota + (\gamma_\iota z_i + \delta_\iota)$$

$$h_\iota = h_\iota(z_i)$$

$$z_i = (x_{m-1,\iota} + c_m)\exp\sum_{l=1}^{K} p_l(m)y_{il}(m)$$

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

$$\iota = \begin{cases} 1 & \text{if } z_i < x_{m1} \\ v & \text{if } x_{m,v-1} \le z_i < x_{mv} \text{ for } v = 2, \ldots, I \\ I+1 & \text{if } z_i > x_{mI}. \end{cases}$$

A.2.2 Under this method the jth component of the gradient vector is:

$$D_j E = \frac{1}{J} \sum_{i=1}^{J} D_j u_i$$

where:

$$D_j u_i = \{\alpha_{1\iota} z_i h_{\iota-1} + (\alpha_{1\iota} z_i + \beta_{1\iota}) h_{\iota-1}^{(1)} + \alpha_{2\iota} z_i h_\iota + (\alpha_{2\iota} z_i + \beta_{2\iota}) h_\iota^{(1)} + \gamma_\iota z_i\} y_{ij}(m)$$

$$h_\iota^{(1)} = \begin{cases} 1 & \text{if } n_i = 0 \\ n_i z_i^{n_i} & \text{otherwise.} \end{cases}$$

A.2.3 Element (j, k) of the Hessian matrix is:

$$D_{jk} E = \frac{1}{J} \sum_{i=1}^{J} D_{jk} u_i$$

where:

$$D_{jk} u_i = \{\alpha_{1\iota} z_i h_{\iota-1} + 2\alpha_{1\iota} z_i h_{\iota-1}^{(1)} + (\alpha_{1\iota} z_i + \beta_{1\iota}) h_{\iota-1}^{(2)} + \alpha_{2\iota} z_i h_\iota + 2\alpha_{2\iota} z_i h_\iota^{(1)} + (\alpha_{2\iota} z_i + \beta_{2\iota}) h_\iota^{(2)} + \gamma_\iota z_i\} y_{ij}(m) y_{ik}(m)$$

$$h_\iota^{(2)} = n_i^2 z_i^{n_i}.$$

A.3 *Passive Approach: Simulation Method*

A.3.1 Under the passive approach described in Section 9.8, the objective function is:

$$E\{u(B)\} = \frac{1}{J} \sum_{i=1}^{J} u_i$$

where u_i is as defined in ¶A.2.1, but with:

$$z_i = \sum_{M=1}^{N} b_M \exp \sum_{l=1}^{K} p_l Y_{il}(M) + \tfrac{1}{2} C(N)$$

$$\iota = \begin{cases} 1 & \text{if } z_i < x_1 \\ v & \text{if } x_{v-1} \le z_i < x_v \quad \text{for } v = 2, 3 \\ 4 & \text{if } z_i > x_3 \end{cases}$$

$$Y_{il}(M) = \sum_{m=M}^{N} y_{il}(m)$$

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press

and

$$b_M = B(0) + c_0 \qquad \text{for } M = 0$$
$$\qquad c_M \qquad \text{for } M > 0.$$

A.3.2 Here the gradient vector is:

$$D_j E = \frac{1}{J}\sum_{i=1}^{J} D_j u_i$$

where:

$$D_j u_i = \{\alpha_{1\iota} h_{\iota-1} + (\alpha_{1\iota} z_i + \beta_{1\iota}) h_{\iota-1}^{(1)} + \alpha_{2\iota} h_\iota + (\alpha_{2\iota} z_i + \beta_{2\iota}) h_\iota^{(1)} + \gamma_\iota\} D_j z_i$$

$$D_j z_i = \sum_{M=1}^{N} b_M Y_{ij}(M) \exp \sum_{l=1}^{K} p_l Y_{il}(M)$$

$$h_\iota^{(1)} = \frac{1}{z_i} \qquad \text{if } n_i = 0$$
$$\qquad n_i z_i^{n_i - 1} \qquad \text{otherwise.}$$

A.3.3 The Hessian matrix is:

$$D_{jk} E = \frac{1}{J}\sum_{i=1}^{J} D_{jk} u_i$$

where:

$$D_{jk} u_i = \{\alpha_{1\iota} h_{\iota-1} + (\alpha_{1\iota} z_i + \beta_{1\iota}) h_{\iota-1}^{(1)} + \alpha_{2\iota} h_\iota + (\alpha_{2\iota} z_i + \beta_{2\iota}) h_\iota^{(1)} + \gamma_\iota\} D_{jk} z_i$$
$$+ \{2\alpha_{1\iota} h_{\iota-1}^{(1)} + (\alpha_{1\iota} z_i + \beta_{1\iota}) h_{\iota-1}^{(2)} + 2\alpha_{2\iota} h_\iota^{(1)} + (\alpha_{2\iota} z_i + \beta_{2\iota}) h_\iota^{(2)}\} D_j z_i D_k z_i$$

$$D_{jk} z_i = \sum_{M=1}^{N} b_M Y_{ij}(M) Y_{ik}(M) \sum_{l=1}^{K} p_l Y_{il}(M)$$

$$h_\iota^{(2)} = -\frac{1}{z_i^2} \qquad \text{for } n_i = 0$$
$$\qquad n_i(n_i - 1) z_i^{n_i - 2} \qquad \text{otherwise.}$$

https://doi.org/10.1017/S1357321700004402 Published online by Cambridge University Press