

The quantification of type-2 prudence in asset allocation by the trustees of a retirement fund

Robert J. Thomson*

School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg 2195, South Africa

Taryn L. Reddy

School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg 2195, South Africa

Abstract

In this paper, consideration is given to the normative use of expected-utility theory for the purposes of asset allocation by the trustees of retirement funds. A distinction is drawn between “type-1 prudence”, which relates to deliberate conservatism on the part of actuaries in the setting of assumptions and the determination of model parameters, and “type-2 prudence”, which relates to the risk aversion of the trustees. The intention of the research was to quantify type-2 prudence for the purposes of asset allocation, both for defined-contribution (DC) and defined-benefit (DB) funds. The authors propose new definitions of the objective variables used as the argument of the utility function: one for DC funds and another for DB funds. A new class of utility functions, referred to as the “weighted average relative risk aversion” class is proposed. Practicalities of implementation are discussed. Illustrative results of the application of the method are presented, and it is shown that the proposed approach resolves the paradox of counter-intuitive results found in the literature regarding the sensitivity of the optimal asset allocation to the funding level of a DB fund.

Keywords

Expected utility; Weighted average relative risk aversion; WARRA class; Retirement funds; Type-2 prudence

1. Introduction

The purpose of this paper is to explore the meaning of “prudence” as it relates to decision-making by the trustees of retirement funds.

The legal provisions relating to the control of retirement funds vary between jurisdictions, but the members of the controlling body of a retirement fund (i.e. an occupational pension or provident fund) are in general fiduciaries of the moneys entrusted to them for the prospective beneficiaries of the fund. In this paper, they are referred to as “trustees”. As such, though they may in some countries be referred to collectively as a “board”, they are jointly and severally accountable not as representatives of constituencies of stakeholders but as fiduciaries for the moneys entrusted to them. This requirement raises implicit questions as to the criteria to be used in order to establish whether a particular decision meets it. For the purposes of such decisions a normative approach to decision-making is required – in particular a normative theory that quantifies the requirements not only of risk aversion but also of prudence.

*Correspondence to: Robert J. Thomson, School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg 2195, South Africa. Tel: +27 11 646 5332. E-mail: rthomson@icon.co.za

In this paper, consideration is given to the normative use of expected-utility theory for the purposes of asset allocation by the trustees of retirement funds. Normative applications relate to the use of expected-utility theory in situations in which decision-makers have obligations, or in which they wish to apply reasonable axioms to their decision-making. Descriptive applications relate to the use of that theory in describing how decision-makers make decisions.

“Prudence” may refer to the assumptions made by actuaries in the valuation of pension schemes (i.e. to deliberately conservative or “somewhat pessimistic” assumptions – Haberman *et al.*, 2003), or it may refer to the risk appetite of the trustees as decision-makers. More attention is given in the actuarial literature to the former than to the latter. For example, Thornton & Wilson (1992: 275) define “prudence” as a probability of 60% that events will prove more favourable in the long term. This compares with a “best estimate” and a “cautious” basis, for which there are probabilities of 50% and 70% that events will prove more favourable in the long term. These suggestions are criticised by Cowling *et al.* (2005: 72–73).

In this paper, a distinction is therefore drawn between “type-1 prudence”, which relates to deliberate conservatism in the setting of assumptions and the determination of model parameters as contemplated in the preceding paragraph, and “type-2 prudence”, which relates to the risk aversion of the trustees. In terms of expected-utility theory, type-1 prudence relates to the distribution of the outcome (i.e. the argument of the utility function) and type-2 prudence to the specification of the utility function in terms of that outcome.

To clarify this distinction we may consider the expression

$$E = \int_0^{\infty} u(x)f(x)dx \quad (1)$$

where $u(x)$ is the decision-maker’s utility of wealth x at a time horizon and $f(\bullet)$ the probability density function of her/his wealth at that time horizon.

Under expected-utility theory, the decision-maker’s objective is to maximise E . Here type-1 prudence focuses on $f(\bullet)$. For example, it would involve setting a deliberately low value to the expected value of the decision-maker’s wealth at the time horizon or a deliberately high value to its standard deviation. Type-2 prudence, on the other hand, focusses on the utility function $u(x)$. For example, it may involve adopting a utility function with high risk aversion.

The problem with type-1 prudence is that it fails to quantify the level of prudence required. If every assumption on which the distribution of the outcome is based is set at a deliberately prudent level, the result is likely to be excessively prudent. It may also be argued that it is more rational to apply prudence to that part of the decision function which is explicitly designed to account for the level of prudence required than to make the distribution arbitrarily unrealistic. Provided it is possible to establish a utility function that on the one hand is not arbitrary and on the other reflects the required level of prudence, the proper place for the application of prudence is therefore in the specification of the utility function. The intention of the authors’ research was to quantify type-2 prudence for the purposes of asset allocation by the trustees of a retirement fund. It is argued that the requirement of type-2 prudence implies that certain criteria should be satisfied by a utility function for normative use by the trustees of a retirement fund. Whilst the elicitation of utility functions inevitably involves subjective choices, it is also demonstrated in this paper how an arbitrary choice of utility function may be avoided.

The requirement of type-2 prudence is understood to imply risk aversion in the sense that the trustees require extra expected returns for extra risks and the quantification of further requirements of prudence is discussed in section 3.

Reference is made in this paper to “retirement funds”. Following usage in certain countries, the intention behind the use of this expression is to include both occupational pension funds (which pay pensions after retirement) and occupational provident funds (which pay lump sums on retirement). This includes both defined-contribution (DC) and defined-benefit (DB) retirement funds.

In section 2, the objective variables used as the argument of the utility function are considered, both for DC funds and for DB funds. In section 3, problems regarding the specification and parameterisation of utility functions for use by the trustees of a retirement fund (whether DC or DB) are considered. In the light of that discussion, criteria for an appropriate utility function are proposed and a new class of utility functions is proposed that satisfies those criteria. This is referred to as the “weighted average relative risk aversion” (WARRA) class of utility functions. In section 4, two problems relating to the application of a WARRA-class utility function to the asset-allocation decision are addressed: the determination of the parameters of utility functions of that class and the use of dynamic programming for the purposes of optimising the asset-allocation decision. Again the discussion is equally relevant to DC and DB funds.

Counter-intuitive results found by other authors paradoxically suggest that the trustees of a DB retirement fund should invest more in risky assets when the fund is in shortfall and less when it is in surplus. In section 5, illustrative results from the application of the method proposed in this paper are presented and the sensitivity of asset allocations to the assumptions made is analysed. Those results resolve that paradox. They also illustrate the application of expected-utility theory to asset allocation using discrete intervals, with future liability cashflows and a model of risky assets defined by the user.

The findings are summarised in section 6, together with suggestions for further research.

2. The Argument of the Utility Function

In this section, the objective variable used as the argument of the utility function is considered. In section 2.1, the objective variable for DC funds is considered and in section 2.2, the objective variable for DB funds is considered.

2.1. DC funds

In the economics literature (e.g. Von Neumann & Morgenstern, 1947; Savage, 1954; Samuelson, 1969) expected-utility theory is developed as a basis for decision-making by individual agents. The argument of the utility function is the consumption of the decision-maker, discounted at a rate equal to the subjective discount rate (Merton, 1969), which reflects the decision-maker’s liquidity preference.

Some actuarial authors – particularly in addressing the asset-allocation problem of a member of a DC retirement fund during the accumulation phase (e.g. Merton, 1993; Gerber & Shiu, 2000 – cf. Cairns *et al.*, 2000; Nielsen, 2006; Deelstra *et al.*, 1999) – define utility functions in terms of terminal wealth, i.e. the lump-sum benefit on retirement. The problem with this definition is that it does not take cognisance of the post-retirement income that can be purchased from the benefit on retirement. (Here a “member” includes a contributing member, a deferred pensioner and a pensioner.)

Cairns *et al.* (2000) use the pension purchased at retirement. The problem with this approach is that it treats members with large current balances as having greater expected utilities at retirement. Whilst this definition may be useful for a member, it does not solve the trustees' problem of designing a default asset allocation (i.e. of determining the asset allocation in the default investment channel) for all members of a certain attained age.

Vigna & Haberman (2001), Levitan & Thomson (2009) and Levitan (2006) use the net replacement ratio as the argument of the utility function for the asset-allocation problem facing a member of a DC retirement fund. The use of a net replacement ratio may be justified for the purposes of investment-channel choice by an individual member. However, it still does not address the trustees' problem of designing a default asset allocation for all members of a certain age. For a group of members of a particular attained age it would be possible to use an aggregate net replacement ratio for the whole group, but this would again place greater weight on those with large current balances than on those with small balances.

The above critique suggests that the argument of the utility function should be a ratio of the post-retirement income to a reference income. As we are using expected-utility theory to determine the optimal exposure to risky assets, we should use as our reference income a risk-free post-retirement income based on a member's current balance. Furthermore, in order to ensure that we can conveniently and fairly assign members to groups so that the same asset allocation will apply to each member of a group, our argument should satisfy the requirement that, for each member of a particular group, the argument is the same.

For a member of a certain group, the argument of the utility function may be taken as the ratio of her/his post-retirement income to an inflation-protected deferred annuity notionally purchased from her/his current balance, expressed as

$$z_T = \frac{\left(\frac{A_T}{a_T}\right)}{\left(\frac{A_0}{{}_T a_0}\right)} \quad (2)$$

where A_T is the member's accumulated balance at time T , being the outstanding term to her/his retirement date; a_T the value per unit at time T of an inflation-protected immediate annuity; A_0 the member's current balance at time 0; and ${}_T a_0$ the value per unit of an inflation-protected deferred annuity payable with effect from retirement date, notionally purchased at time 0.

Equation (2) thus expresses the argument of the utility function as the ratio of an outcome to a benchmark outcome, both of which relate to the objectives of a DC fund.

As z_T is the same for all members of a particular attained age, with a particular retirement age and with particular annuity terms, the same ratio – and therefore the same asset allocation – may be used for all such members. Within each such group of members it also avoids placing greater emphasis on members with large current balances than on members with small balances. It therefore solves the trustees' problem of designing a default asset allocation: the same asset allocation may be used for every member in the group. This ratio is referred to below as the "DC benefit ratio".

The above definition presupposes that the benefit on retirement will be payable as a lump sum. In effect, however, it is assumed either that the fund will purchase an annuity for a retiring member or that she/he will do so. Longevity risks will therefore not be borne by the fund. Because preferences

must ultimately be expressed in terms of goods and services rather than money, z_T is defined in terms of inflation-protected annuities. However, if and to the extent that the benefit on retirement is payable in the form of an annuity, the terms and conditions of that annuity (i.e. its frequency, its provision for a term certain or a survivor's annuity and whether it is inflation-protected) should be applied in the determination both of ${}_T a_0$ and of a_T .

Both A_T and a_T are stochastic. Allowance will have to be made for correlation between a_T and the returns on the various investment channels as these correlations will affect the variance of the distribution of z_T .

If the fund buys annuities from a life office, the value of a_T will have to be modelled accordingly. If a member has options between alternative annuities, or the option to defer annuitisation – or not to annuitise – then the trustees will need to decide on a default annuity option. This option may be applied before retirement, unless and until the member makes an option, for the purposes of simulating equation (2) during the accumulation phase. It may be appropriate to allow for a combination of alternative annuities, optimising between them by maximising the expected utility of real post-retirement income. For this purpose, the bequest motive may be disregarded, because the trustees' fiduciary duties are to the members and stated dependants, not to other legatees. As shown by Samuelson (1969), though, the consumption–saving problem will not necessarily be separable, except for constant relative risk aversion. In general, the simultaneous maximisation of expected utility for annuitisation (and for asset allocation during deferment of annuitisation) and for post-retirement consumption year by year will therefore be complicated. In practice, it would be preferable to assume no savings or dissavings – i.e. to assume that the accumulated savings will be immediately annuitised on attainment of the retirement age and that all post-retirement income will be required for consumption year by year. For this purpose, a level inflation-protected annuity would need to be assumed for the default option. Even if most members select an annuity that is constant in nominal terms, the requirement of type-2 prudence suggests that the default should be expressed in real terms. There may be some room for compromise if allowance is made for life-cycle expenses varying by age, but this would also need to be expressed in real terms. The features of the default option – including frequency of payment and the provision of annuities to surviving spouses and dependants – would need to be reflected in the determination of a_T . There may be grounds for arguing that some capital may be required on retirement, either for paying off debts or for investment in income-generating activities. The counter-argument to this is that such arrangements are preferable only if and to the extent that they improve post-retirement income. The latter should therefore be used as the default argument.

As mentioned above, the value of ${}_T a_0$ is intended to be notional; it need not be based on the prices of deferred annuities available in the market. As markets for such annuities are generally illiquid, it would be better to determine the value internally. It should, however, be consistent with the basis used for determining a_T so that

$$\lim_{T \rightarrow 0} ({}_T a_0) = a_0.$$

In principle, A_t should include not only the retirement savings in the fund, but also those outside of the fund. If it is assumed that the retirement savings outside of the fund will be invested (and taxed) similarly to the asset allocation in the fund, the DC benefit ratio is unaffected. However, the trustees have no responsibility for members' retirement savings outside of the fund.

As shown in equation (2), the time horizon for a particular member is the date of attainment of the retirement age. However, as explained above, it should be recognised that it will be possible to

re-determine the default asset allocation at regular intervals during the accumulation phase. This necessitates dynamic asset allocation, as explained in section 4.2 below.

For the purposes of dynamic asset allocation, some authors (e.g. Cairns, 2000) use continuous time. As explained by Cairns (2000), this may be useful for the purposes of analysis of the sensitivity of asset allocations to various parameters. However, for practical applications the utility function and the models of risky assets may be too complex to permit solutions in closed form. In principle, as argued elsewhere by one of the authors (Thomson, 2011), the time interval to be used for dynamic asset allocation should be the same as the decision interval. Thus, if the trustees review the mandates of investment managers annually, if they reconsider their model of risky assets and the parameters of their utility function annually, and if members are grouped into annual age intervals for the purposes of determining default allocations, it would be appropriate to use an annual interval for dynamic modelling. Dynamic modelling is further considered in section 4.2 below.

2.2. DB funds

Some authors (e.g. Sherris, 1993; Boulier *et al.*, 1995; Siegmann & Lucas, 1999) treat the sponsoring employer as the beneficiary of the pension fund. Others (e.g. Sharpe & Tint, 1990; Cairns, 2000; Cairns *et al.*, 2000; Haberman *et al.*, 2003; Cardinale *et al.*, 2006; McCarthy & Miles, 2013) attempt to reflect a compromise between the interests of the sponsoring employer and the obligations of the trustees to members.

Whilst for an individual the argument of the utility function may be defined in terms of her/his post-retirement income, it must be recognised that this is not necessarily appropriate for the trustees of a retirement fund. In the case of a DB fund, the argument of the utility function must take cognisance not only of the assets of the fund but also of its liabilities. An individual has choices with regard to the liabilities that she/he may incur from time to time, but with the possible exceptions of increases to pensions, the trustees have no such choices. The risks faced by the trustees of a retirement fund cannot be defined without reference to its liabilities.

For a DB fund, it would be natural to take “wealth” to mean the excess of the assets over the liabilities; i.e. the surplus at a time horizon (as does Sherris (1992) in the single-period case). However, this would mean that the trustees would prefer a large fund to a small fund with the same funding ratio. That would clearly be wrong; the size of a fund does not determine its financial soundness, and it is the latter, not the former, for which the trustees are responsible. For example, the amalgamation of a well-funded scheme with a scheme with a relatively low funding ratio should not be accepted by the trustees of the former without arrangements to safeguard the interests of its members and its sponsoring employer. This suggests that the funding ratio itself would be more appropriate. The funding ratio also has more meaning with regard to the security of the prospective beneficiaries. The numerator of the funding ratio can also be adjusted to allow for the sponsor’s solvency and share of surpluses and shortfalls (e.g. McCarthy & Miles, 2013).

A problem with the use of the funding ratio is that, if the argument of the utility function is required at the date of expiry of the liabilities, the denominator of the ratio is 0. Also – and more importantly – whilst in the absence of a sponsoring employer’s guarantee and insurance against its failure a high funding ratio gives greater security to members, it does not relate to the benefits actually payable to members.

The above critique suggests that the argument of the utility function should be expressed in terms of the benefits payable up to the time horizon, together with the assets available to pay subsequent benefits. (The inclusion of the sponsor covenant is dealt with below.) As for the DC benefit ratio, the argument should be relative to a standard value. In contrast to the DC case, however, a DB fund creates rights to benefits that are unrelated to the contributions paid by and in respect of members. The standard value should therefore be expressed in terms of benefits payable up to the time horizon, together with liabilities for subsequent benefits.

The argument may be taken as

$$z_t = \frac{A_t + P_t}{L_t + P_t} \quad (3)$$

where A_t is the value of the assets of the fund available at time t for subsequent benefits; L_t the value of the liabilities of the fund at time t for subsequent benefits in respect of service to time 0; and P_t the value of benefits payable during the period $[0, t]$, accumulated to time t with interest at the beneficiary's risk-free rate from time to time.

In particular

$$z_0 = \frac{A_0}{L_0} \quad \text{and} \quad (4)$$

$$z_T = \frac{A_T + P_T}{P_T} \quad (5)$$

where T is the date of expiry of the liabilities.

The numerator of equation (3) represents the value of benefits payable to members up to and beyond time t , together with surplus assets available. (P_t represents the benefits payable up to time t , whilst A_t represents the assets required to cover subsequent benefits, and any surplus assets available for possible improvements.) The denominator represents the value of benefits payable up to and beyond time t . We refer to the ratio as the "DB benefit ratio". Equation (3) thus expresses the argument of the utility function as the ratio of an outcome to a benchmark outcome, both of which relate to the objectives of a DB fund.

It may be questioned why P_t should be accumulated with interest at the beneficiary's risk-free rate. The reason for this is that the benefits cannot be assumed to have been invested in risky assets after they have been paid; it is the trustees who are accountable for investment in risky assets. In effect, the beneficiary's risk-free rate is assumed to represent the liquidity preference of the beneficiary.

It may also be questioned whether the choice of T as the date of expiry of the liabilities is prejudicial to pensioners as at time 0, whose pensions will cease long before T . For a DC fund the definition of z_T is specific to a member whose retirement date is T , so there is no issue; the optimal exposure to risky assets applies to that member. But for a DB fund the question is material: if there was only one member in the fund the optimal exposure to risky assets would generally be very different from the optimal exposure for a diverse range of members. From the point of view of a particular member, the proposed choice of T is arbitrary. This matter is dealt with in section 4.2 below.

As noted above, consideration also needs to be given to the sponsor covenant. In principle, the argument of the utility function should take cognisance of the objectives both of the members and of

the sponsoring employer. Whilst the trustees' prime concern is not the minimisation of employers' contributions, they do have an indirect responsibility to avoid precipitating the default of the sponsoring employer, particularly while the fund is in shortfall. In fact, the suggested DB benefit ratio in equation (3) addresses not only the objectives of the members but also those of the sponsoring employer. The pension fund is a mechanism for assisting the negotiation between sponsor and beneficiaries with regard to future benefits. By focussing on the fund's capacity to pay benefits, such negotiations become integrative rather than distributive bargaining.

It would, however, be possible to analyse the difference between the interests of the members and those of the sponsoring employer. This could be done by analysing the assets A_t as those available for members less the shortfall (if any) payable by the employer. Thus

$$A_t = A_t^M + A_t^E \tag{6}$$

where $A_t^M = L_t + jS_t$ constitutes the funds available to members; $A_t^E = (1 - j)S_t$ constitutes the funds available to the employer; and $j \in [0, 1]$ is the proportion of the surplus available to members. The value of j will depend on the rules of the fund and the regulatory requirements regarding the funding of shortfalls and the transfer of surplus to employers, as well as the probability of default by the sponsoring employer and the arrangements regarding insurance against that contingency. It will typically depend on whether the surplus is positive and it may itself depend on the funding ratio.

Now suppose that we wish to apply a weighting $l \in [0, 1]$ to the employer relative to members, so that

$$A'_t = A_t^M + lA_t^E$$

i.e.

$$\begin{aligned} A'_t &= \{L_t + jS_t\} + l\{(1 - j)S_t\} \\ &= \{L_t + S_t\} - (1 - j)S_t + l\{(1 - j)S_t\} \\ &= A_t - \{(1 - l)(1 - j)S_t\} \end{aligned}$$

Equation (3) can then be adjusted as

$$z_t^M = \frac{A_t^M - kS_t^E + P_t}{L_t + P_t} \tag{7}$$

where

$$k = (1 - j)(1 - l).$$

When $k = 0$ (either because $j = 1$ or $l = 1$), equation (7) reduces to equation (3), which gives neutral weighting. When $k = 1$ (because $j = 0$ and $l = 0$), the numerator of equation (7) ignores the surplus. The optimal asset allocation can then be found for various values of k and the decision can be made by distributive bargaining.

It may be argued that members' concerns are confined to shortfalls, whilst employers have an interest both in surpluses and in shortfalls. This can be accommodated by adjusting k accordingly. However, it is questionable whether members' interests are in fact confined to the downside; the greater the surplus at time t the lower the probability of a shortfall at time $t + 1$.

The above discussion relates only to past-service benefits, i.e. to benefits for service up to time 0. The optimal portfolio based on the maximisation of the expected utility of that argument relates to the assets

of the fund at time 0. The argument for service currently accruing may be similarly determined. The optimal portfolio based on the maximisation of the expected utility of the latter argument relates to the investment of contributions to the fund during the period $[0, 1)$ or the annual rate of contributions at time 0. (Although dynamic asset allocation is used to determine future optimal asset allocations as described in section 4.2 below, the object of that exercise is to find the optimum asset allocation for current contributions. The whole exercise will be repeated at the start of the next year, which will then be defined as time 0.) This approach would correspond to the use of the projected-unit-credit method of valuation, which would be appropriate for an open fund. The optimal asset allocation of contributions can be determined accordingly. For a closed fund it would be appropriate to allow for all future service, which would correspond to an aggregate-funding method.

As for DC funds, it should be recognised that it will be possible to re-determine the asset allocation at regular intervals. This necessitates dynamic asset allocation. Again, the time interval to be used for dynamic asset allocation should be the same as the decision interval. Thus, if the trustees review the mandates of investment managers annually, if they reconsider their model of risky assets and the parameters of their utility function annually, and if the fund is valued annually, it would be appropriate to use an annual interval for dynamic modelling.

3. The Utility Function

In this section, the specification of the utility function for use by the trustees of a retirement fund is considered. In section 3.1, utility functions proposed in the literature are reviewed and problems with those functions are identified. Section 3.2 addresses the parameterisation of utility functions – in particular, the determination of parameters relating to risk aversion and prudence. In the light of sections 3.1 and 3.2, section 3.3 proposes five criteria that need to be satisfied by the utility function in order to meet the requirements of type-2 prudence. In section 3.4, the use of decreasing relative risk aversion is considered and a new class of utility functions is proposed that satisfies the criteria established in section 3.3.

3.1. The functional form of the utility function: a review

Numerous empirical tests of the descriptive validity of various utility functions have been made. For example, Levy (2005) finds, in an experiment involving an actual monetary gain or loss indexed to the participant's investment performance, that “[decreasing absolute risk aversion] is strongly supported, but [increasing relative risk aversion] is rejected”. Friend (1973) and Cohn *et al.* (1975) provide empirical evidence of decreasing relative risk aversion. Some evidence of decreasing relative risk aversion is also found by Bossons (1973) and Projector & Weiss (1966). In principle, the functional form of the utility function for a DC fund should be the same as for a DB fund; it is only the argument of the functions that differ.

The shapes of the utility functions of individual decision-makers may be quite variable. Levitan & Thomson (2009) show those elicited from a sample of members of a South African retirement fund. In that sample, there are examples of individuals who show aspiration levels – i.e. points below which risk aversion is negative and above which it is positive. The preferences implied by such utility functions would not be appropriate for fiduciaries as they constitute aspirational behaviour (Diecidue & Van de Ven, 2008). Prudence requires, *inter alia*, that trustees should not take higher risks when the fund is in shortfall than when it is in surplus. Indeed, any utility function that exhibits negative risk aversion over any part of its range would constitute aspirational behaviour. Thus, the value

function defined in prospect theory would typically be inappropriate. Furthermore, it may be argued that any utility function that exhibits increasing absolute or relative risk aversion (depending on which of these measures is adopted) over any part of its range would constitute aspirational behaviour; it would suggest that lower risk aversion may be acceptable at lower values of the benefit ratio than at higher values. Also, the elicitation of a utility function may be affected by the problems associated with the descriptive application of expected-utility theory. These matters are considered further below.

For most purposes, the hyperbolic absolute risk aversion (HARA) class of utility functions has been considered to be sufficiently flexible and its mathematical tractability has established its popularity in the literature. As the name suggests, this class satisfies the requirement (Mossin, 1968) that, for a utility function $u(\bullet)$, the absolute risk aversion is

$$r(x) = -\frac{u''(x)}{u'(x)} = \frac{1}{ax+b}. \tag{8}$$

For $a > 0, b > 0$, this establishes decreasing absolute risk aversion. Defining “absolute risk tolerance” as

$$t(x) = \frac{1}{r(x)} \tag{9}$$

we obtain the result that absolute risk tolerance is linear. As shown by Merton (1971), the utility function for this class is

$$u(x) = \begin{cases} \frac{\gamma}{1-\gamma} \left(\frac{ax}{\gamma} + \beta\right)^{1-\gamma} & \text{for } a \neq 0, \gamma \neq 1 \\ \ln(x + \beta) & \text{for } a \neq 0, \gamma = 1 \\ -\exp(ax) & \text{for } a = 0 \end{cases} \tag{10}$$

These forms are referred to as “generalised power utility”, “generalised logarithmic utility” and “exponential utility”, respectively. This class includes

- quadratic utility, with $\gamma = -1$;
- constant absolute risk aversion, with $a = 0$; and
- constant relative risk aversion with $a \neq 0, \beta = 0$.

HARA utility functions show decreasing relative risk aversion only if $\beta < 0$. This is generally problematic for small values of x . The problems with quadratic utility functions are well documented (e.g. Pratt, 1964; Booth, 1995) and are not further discussed here.

The utility function

$$u(x) = \begin{cases} \frac{x^{1-\gamma}-1}{1-\gamma} & \text{for } \gamma \neq 1 \\ \ln(x) & \text{for } \gamma = 1 \end{cases} \tag{11}$$

is referred to by Samuelson (1969) as the “isoelastic” utility function. The use of this expression is explained by Gerber & Shiu (2000). Isoelastic utility implies (and is implied by) constant relative risk aversion:

$$\begin{aligned} \gamma(x) &= -x \frac{u''(x)}{u'(x)} \\ &= \gamma \end{aligned} \tag{12}$$

The use of a discontinuous utility function such as those used by Booth (1995) and Khorasane & Smith (1997) are inappropriate as they imply infinite relative risk aversion (generally negative below

the point of discontinuity and positive above it). Similar problems arise with kinked utility functions (i.e. with utility functions exhibiting discontinuities in marginal utility).

For the purposes of determining the form of the utility function, it would be possible to elicit the utility functions of at least a sample of members. The functional form so elicited may be parametric or non-parametric. Whilst such utility functions might be informative, they could not be used in any mechanical way to construct a utility function for use by the trustees. In the first place, the trustees are required to be prudent, and second, if such a method had been contemplated, it would not have been necessary to appoint them. The way in which the utility functions of a sample of members may be used to inform the development of a utility function for use by the trustees is illustrated in sections 3.2 and 4.1.

We first consider the HARA class. From equations (8) and (12) it follows that the relative risk aversion for HARA utility is

$$\gamma(z) = -z \frac{u''(z)}{u'(z)} = \frac{z}{az+b}. \quad (13)$$

The derivative of this is

$$\frac{d}{dz}(\gamma(z)) = \frac{b}{(az+b)^2} \quad (14)$$

For our purposes we require $a \geq 0$; otherwise risk aversion becomes negative for large z . We also require that $b \geq 0$ or that the range of z must be restricted to $z > \frac{-b}{a}$; otherwise there is a discontinuity at $z > \frac{-b}{a}$. The range of z cannot be so restricted. This means that, unless $b = 0$ (in which case relative risk aversion is constant), relative risk aversion is increasing. This applies both to generalised power utility and to generalised logarithmic utility. The latter has been advanced by Rubinstein (1976), but neither of these forms is possible with an argument $z \in (0, \infty)$.

The findings of Projector & Weiss (1966), Bossons (1973), Friend (1973), Cohn *et al.* (1975) and Levy (2005) that, even amongst individual decision-makers, non-increasing relative risk aversion is supported suggests that constant relative risk aversion is not generally appropriate for trustees. Instead, non-increasing relative risk aversion should be used. It would be aspirational for trustees to take more risk when funding ratios are low than when they are high. As relative risk aversion is the measure to be used, this means that increasing relative risk aversion (and therefore constant absolute risk aversion) would be aspirational.

Another advantage of the use of relative risk aversion over that of absolute risk aversion is that the former is dimensionless. The dimension of absolute risk tolerance is that of the argument of the utility function. This means that the dimension of absolute risk aversion is the inverse of that of the argument of the utility function. The advantage of a dimensionless measure is that it applies regardless of the size of the fund. If, as in equation (3), the argument is itself dimensionless, it appears at first sight that the problem falls away. However, it must be borne in mind that, for given values of future benefits, the proposed argument is effectively expressed in terms of currency. The use of relative risk aversion ensures that the problem is truly dimensionless even for given values of the denominator of equation (3).

In the light of the above arguments, apart from the case of constant relative risk aversion, the HARA class is inappropriate.

The fitting of utility functions as contemplated by Thomson (2003b) and Levitan & Thomson (2009) allows for degrees of risk tolerance that, whilst they might be acceptable for individual members, would not be appropriate for the purposes of decision-making by trustees as contemplated in this paper.

Because of the requirements of type-2 prudence, the relative risk aversion will invariably be greater than 1; this is discussed further in section 3.2 below. The case of constant relative risk aversion may be acceptable to the trustees, and there is no intrinsic reason why it should not be. However, if the trustees require decreasing relative risk aversion, or if they wish to effect a compromise between different levels of relative risk aversion amongst themselves, then another class of utility functions is necessary. That is considered in sections 3.2 and 3.3 below.

3.2. Levels of risk aversion and the requirements of prudence

The legislative and regulatory requirements of many jurisdictions explicitly or implicitly require fiduciary duties including prudence, and common law in this regard is well established. We now turn to the relationship between risk appetite and type-2 prudence. In terms of expected-utility theory, risk appetite is measured in terms of risk aversion. As we have accepted expected-utility theory as normatively valid, we shall consider the relationship between risk aversion and type-2 prudence.

Biasca (2009) uses illustrative levels of constant relative risk aversion of 10–30. These levels are selected without justification. Cardinale *et al.* (2006) use illustrative levels of constant relative risk aversion of 0–7. McCarthy & Miles (2013) use constant relative risk aversion of 5. Examples of widely differing levels of constant absolute risk aversion are given by Pézier & Scheller (2011).

If the trustees are to make a reasonable decision about the level of relative risk aversion they should adopt, it would arguably be appropriate for them to be informed by the levels of relative risk aversion exhibited by members of the fund. In doing so, they need to recognise that not all members will be prudent. Nevertheless, such information could at least be used as a check on their own utility functions, so as to establish how they as trustees compare with members. For this purpose, the trustees may elicit the utility functions of a sample of members.

This approach may be illustrated with reference to the sample of members of a retirement fund that was used in Levitan & Thomson (2009). More details of that sample, and of the research conducted on it, are contained in Levitan (2006). In those works, for each respondent m , an average relative risk aversion was found as

$$\bar{\gamma}_m = \frac{1}{x_w - x_\alpha} \int_{x_\alpha}^{x_w} \gamma_m(x) dx \quad (15)$$

where x is the member's net replacement ratio at retirement and $\gamma_m(x)$ the relative risk aversion of the member as a function of the net replacement ratio x . Whilst the net replacement ratio differs from the benefit ratios used in this paper, the elicitation could be framed in terms of benefit ratios. The range of integration $[x_\alpha, x_w]$ was the range of net replacement ratios over which the member's utility function was elicited. For the purpose of calculating $\gamma_m(x)$, the member's utility function was elicited using the equally likely certainty equivalent (ELCE) method (Anderson *et al.*, 1977; Farquhar, 1984; Thomson, 2003b) at discrete values of x over the range. The values of the utility function for intermediate values of x were found by interpolation, using an interpolation method suggested in Thomson (2003b).

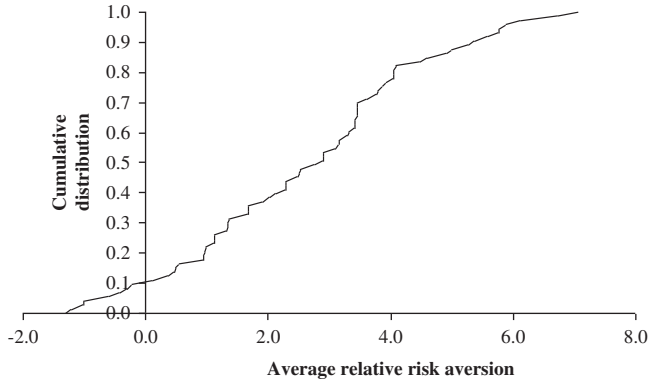


Figure 1. Cumulative sample distribution of average relative risk aversion.

The cumulative distribution of the average relative risk aversion of members in this sample is shown in Figure 1. As discussed in Levitan & Thomson (2009), two outliers have been ignored. The median was 2.90 and the upper extreme was 7.06.

Because the trustees need to be more prudent than most members, the relative risk aversion to be adopted by the trustees should be at least the median. Clearly, though, one would not expect them to adopt the upper extreme. The relative risk aversion must therefore be between 2.90 and 7.06.

However, the requirement of type-2 prudence does not only relate to risk aversion; it also relates to the levels of risk aversion exhibited over the range of possible outcomes. Analogously to the Pratt–Arrow definitions of absolute and relative risk aversion, Kimball (1990) defines a “coefficient of absolute prudence”

$$p(x) = -\frac{u'''(x)}{u''(x)} \quad (16)$$

and a “coefficient of relative prudence”

$$\pi(x) = -x\frac{u'''(x)}{u''(x)} \quad (17)$$

He motivates these coefficients by a definition of “prudence” as “the sensitivity of the optimal choice of a decision variable” (e.g. the amount to be invested in the fund’s risk-free asset) to risk. He suggests that, if $p(x) > 0$ or $\pi(x) > 0$ then the decision-maker exhibits absolute or relative prudence, respectively.

In the context of the obligations of trustees, it seems that Kimball’s (1990) definition is insufficient. If, for example, an isoelastic utility function is adopted by the trustees, then Kimball’s coefficient of relative prudence will be

$$\pi(z) = -z\frac{u'''(z)}{u''(z)} = \gamma(z) + 1 \quad (18)$$

This will be positive (implying “prudence”) even for $-1 < \gamma < 0$. In that range a decision-maker would not even be risk-averse (and in fact not even unsatiated), let alone “prudent” in the sense intended in the context of trustees’ obligations.

In our context, it is clearly intended that type-2 prudence relates to risk aversion. First, it requires greater risk aversion than that of the median member, and certainly greater risk aversion than 1. Second, the requirement of type-2 prudence would be incompatible with increasing relative risk aversion, let alone increasing absolute risk aversion. As discussed in section 3.1 above, it would not be prudent for a trustee to adopt lower risk aversion when the fund is in shortfall than when it is in surplus. (Some literature shows that, on certain assumptions, the adoption of this principle produces results suggesting the contrary. That paradox is considered in section 5 below.) By the same token, as also indicated in that section, there should not be a discontinuity in marginal utility. Third, for any given level of the argument of the utility function (the DC benefit ratio in the case of a DC fund or the DB benefit ratio in the case of a DB fund) the trustees' level of prudence should be greater than or equal to the level of prudence that would be exhibited by an agent with constant relative risk aversion equal to the trustees' relative risk aversion at that level of the argument.

3.3. Criteria required for type-2 prudence

In the light of the discussion in sections 3.1 and 3.2 it follows that, in order to satisfy the requirements of type-2 prudence, the trustees' utility function $u(\bullet)$ should conform to the following criteria:

1. The range criterion: it should map the open interval $(0, +\infty)$ into the open interval $(-\infty, +\infty)$ or into a subset of that interval.
2. The continuity criterion: it should be thrice differentiable.
3. The unsatiation criterion: the trustees should be unsatiated, so that, for all z , $u'(z) > 0$.
4. The relative risk-aversion criterion: for all z there exists a γ^* such that $\gamma(z) \geq \gamma^* > 1$.
5. The non-increasing relative risk aversion criterion: for any z^* and for all $z < z^*$, $\gamma(z) \geq \gamma(z^*)$.

In condition 4, following from the preceding discussion, the value of γ^* needs to be set high enough to establish the trustees' level of risk aversion as adequately prudent. On the other hand, the value of γ^* should not be unreasonably high.

We refer to a utility function satisfying these criteria as a "prudent utility function".

3.4. The WARRA class

If, as observed above, the trustees require decreasing relative risk aversion, or if they wish to effect a compromise between different levels of relative risk aversion amongst themselves, then another class of utility functions is necessary. In this section, such a class is defined. As before, the same principles apply to DC funds as to DB funds.

We define the WARRA class by a generalisation of the isoelastic utility function as follows:

$$u(z) = \frac{u_0(z) + cu_\infty(z)}{1 + c} \tag{19}$$

where

$$u_0(z) = \frac{z^{1-\gamma_0} - 1}{1 - \gamma_0} \tag{20}$$

$$u_{\infty}(z) = \frac{z^{1-\gamma_{\infty}} - 1}{1 - \gamma_{\infty}}; \tag{21}$$

$$c > 0 \quad \text{and}$$

$$\gamma_0 \geq \gamma_{\infty} > 1.$$

The rationale behind the subscripts becomes clear below.

The relative risk aversion of this utility function is

$$\begin{aligned} \gamma(z) &= -z \frac{u''(z)}{u'(z)} \\ &= \frac{\gamma_0 z^{-\gamma_0} + c\gamma_{\infty} z^{-\gamma_{\infty}}}{z^{-\gamma_0} + cz^{-\gamma_{\infty}}} \\ &= \frac{\gamma_0 + c\gamma_{\infty} z^{\lambda}}{1 + cz^{\lambda}} \end{aligned} \tag{22}$$

where

$$\lambda = \gamma_0 - \gamma_{\infty}$$

The relative risk aversion of the WARRA class is thus a weighted average of those of a more risk-averse decision-maker and a less risk-averse decision-maker, the weighting increasing to the latter as the argument increases.

Figure 2 shows the relative risk aversion of a WARRA-class utility function as a function of the argument z . In that figure the following parameters have been used

$$\gamma_0 = 5$$

$$\gamma_{\infty} = 3 \quad \text{and}$$

$$c = 1$$

From equation (22) it may be shown that

$$\lim_{z \rightarrow 0} (\gamma(z)) = \gamma_0 \quad \text{and} \tag{23}$$

$$\lim_{z \rightarrow \infty} (\gamma(z)) = \gamma_{\infty}. \tag{24}$$

This means that, if WARRA is adopted as a compromise between the most risk-averse trustee A, with constant relative risk aversion of γ_0 , and the least risk-averse trustee B, with relative risk aversion of γ_{∞} , then, as shown in Figure 2, the relative risk aversion of trustee A is accommodated when benefit ratios are very low, and that of trustee B is accommodated when benefit ratios are very high. This provides a suitable basis for compromise, as trustee A is likely to be more concerned about the risk at low levels, whilst trustee B is likely to be more concerned about excessive risk aversion at higher levels.

It may also be shown that

$$\frac{d}{dz} (\gamma(z)) \leq 0 \tag{25}$$

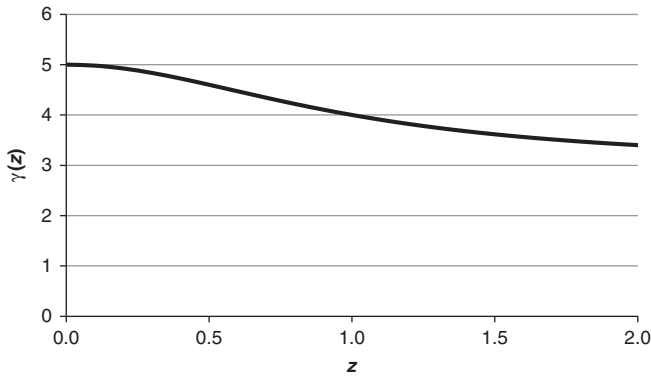


Figure 2. Relative risk aversion of a WARRA-class utility function.

thus establishing non-increasing risk aversion over the whole range, as shown in Figure 2. This shows that WARRA may be adopted not only for the purpose of effecting a compromise between different trustees, but also for the purpose of deliberately achieving decreasing relative risk aversion.

For suitable values of the parameters, the WARRA class satisfies the requirements of type-2 prudence as set out in section 3.2. This class does not uniquely satisfy those requirements, but it provides a simple formulation of a utility function that does satisfy them, with sufficient flexibility for most purposes.

Finally, it follows from equation (22) that

$$\gamma(1) = \frac{\gamma_0 + c\gamma_\infty}{1 + c} \tag{26}$$

so that, as shown in Figure 2, when the benefits are equal to their benchmarks, the weighting given to trustee B’s risk aversion relative to trustee A’s is equal to the weighting given to trustee B’s utility function relative to trustee A’s.

4. The Application of a WARRA-Class Utility Function to the Asset-Allocation Decision

In this section, two problems relating to the application of a WARRA-class utility function to the asset-allocation decision are addressed. In the first place, the parameters of utility functions of that class need to be determined. This problem is addressed in section 4.1. In order to apply the WARRA class of utility functions to retirement-fund asset-allocation decision-making, dynamic asset allocation is required. A method of implementing dynamic programming for this purpose is outlined in section 4.2.

4.1. The determination of the parameters

If the WARRA class is adopted for the purpose of deliberately achieving decreasing relative risk aversion (as opposed to a compromise between less risk-averse and more risk-averse trustees) then the trustees must reach consensus as to the relative risk aversion at extremely low funding levels (γ_0),

at extremely high funding levels (γ_∞) and when the assets are equal to the liabilities, so that, from equation (26)

$$c = \frac{\gamma_0 - \gamma(1)}{\gamma(1) - \gamma_\infty} \tag{27}$$

For this purpose, the trustees' utility functions at these levels need to be elicited. This may be done using the ELCE method (Anderson *et al.*, 1977; Farquhar, 1984; Thomson, 2003*b*). This method addresses some of the problems raised above. If the trustees find it difficult to respond to the elicitation process with reference to extreme levels, a more reasonable range of levels may be used. Suppose the levels elicited are

$$z_i \text{ for } i = 1, \dots, 5$$

and that the corresponding values of the utility function are

$$u_i \text{ for } i = 1, \dots, 5.$$

Without loss of generality we may assume that the utility function is shifted and scaled by the factors θ and ϕ , so that

$$u_i = \theta + \phi v_i \tag{28}$$

where v_i is the utility function before shifting and scaling, i.e.

$$v_i = \frac{1}{1+c} \left(\frac{z_i^{1-\gamma_0} - 1}{1-\gamma_0} + c \frac{z_i^{1-\gamma_\infty} - 1}{1-\gamma_\infty} \right) \tag{29}$$

Let

$$w_i = \theta + \phi v_i - u_i \tag{30}$$

Then we require

$$w_i = 0.$$

We may then express the relationships between the five given values of z_i and u_i in terms of the simultaneous equations

$$w_i = \theta + \frac{\phi}{1+c} \left(\frac{z_i^{1-\gamma_0} - 1}{1-\gamma_0} + c \frac{z_i^{1-\gamma_\infty} - 1}{1-\gamma_\infty} \right) - u_i \text{ for } i = 1, \dots, 5. \tag{31}$$

In order to solve these equations we may apply Newton's method as follows. Using the subscript n to denote the n th estimate of the respective parameters we have

$$w_n = \begin{pmatrix} w_{n1} \\ \vdots \\ w_{n5} \end{pmatrix} = \begin{pmatrix} \theta_n + \frac{\phi_n}{1+c_n} \left(\frac{z_1^{1-\gamma_{0n}} - 1}{1-\gamma_{0n}} + c_n \frac{z_1^{1-\gamma_{\infty n}} - 1}{1-\gamma_{\infty n}} \right) - u_1 \\ \vdots \\ \theta_n + \frac{\phi_n}{1+c_n} \left(\frac{z_5^{1-\gamma_{0n}} - 1}{1-\gamma_{0n}} + c_n \frac{z_5^{1-\gamma_{\infty n}} - 1}{1-\gamma_{\infty n}} \right) - u_5 \end{pmatrix} \tag{32}$$

The Jacobian matrix of this equation is

$$J_n = \begin{pmatrix} \frac{\partial w_{n1}}{\partial \theta_n} & \frac{\partial w_{n1}}{\partial \phi_n} & \frac{\partial w_{n1}}{\partial \gamma_{0n}} & \frac{\partial w_{n1}}{\partial \gamma_{\infty n}} & \frac{\partial w_{n1}}{\partial c_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial w_{n5}}{\partial \theta_n} & \frac{\partial w_{n5}}{\partial \phi_n} & \frac{\partial w_{n5}}{\partial \gamma_{0n}} & \frac{\partial w_{n5}}{\partial \gamma_{\infty n}} & \frac{\partial w_{n5}}{\partial c_n} \end{pmatrix} \tag{33}$$

where

$$\frac{\partial w_{ni}}{\partial \theta_n} = 1 \tag{34}$$

$$\frac{\partial w_{ni}}{\partial \phi_n} = \frac{1}{1 + c_n} \left(\frac{z_i^{1-\gamma_{0n}} - 1}{1 - \gamma_{0n}} + c_n \frac{z_i^{1-\gamma_{\infty n}} - 1}{1 - \gamma_{\infty n}} \right) \tag{35}$$

$$\frac{\partial w_{ni}}{\partial \gamma_{0n}} = \frac{\phi_n \left[z_i^{1-\gamma_{0n}} \{ 1 - (1 - \gamma_{0n}) \ln z_i \} - 1 \right]}{(1 + c_n)(1 - \gamma_{0n})^2} \tag{36}$$

$$\frac{\partial w_{ni}}{\partial \gamma_{\infty n}} = \frac{\phi_n c_n \left[z_i^{1-\gamma_{\infty n}} \{ 1 - (1 - \gamma_{\infty n}) \ln z_i \} - 1 \right]}{(1 + c_n)(1 - \gamma_{\infty n})^2} \text{ and} \tag{37}$$

$$\frac{\partial w_{ni}}{\partial c_n} = \frac{\phi_n}{(1 + c_n)^2} \left(\frac{z_i^{1-\gamma_{\infty n}} - 1}{1 - \gamma_{\infty n}} - \frac{z_i^{1-\gamma_{0n}} - 1}{1 - \gamma_{0n}} \right). \tag{38}$$

Let

$$\chi_n = \begin{pmatrix} \theta_n \\ \phi_n \\ \gamma_{0n} \\ \gamma_{\infty n} \\ c_n \end{pmatrix}. \tag{39}$$

Applying Newton’s method we may, in principle, solve the equation iteratively using the updating equation

$$\chi_{n+1} = \chi_n - J_n^{-1} w_n. \tag{40}$$

A problem arises with this approach: however, in general the Jacobian matrix is almost singular, so that Newton’s method fails to converge. If we assume that the parameter c is known, then the above method may be applied, omitting that parameter. For that purpose, we use only four given values, we omit the last column and the bottom row of the Jacobian and the last component of w_n and χ_n . This procedure may be accessed by another function optimising c so as to approach a fifth observation as closely as possible. If c is allowed to become large, the higher risk aversion at lower levels of γ_0 would be undermined. In optimising c , it would therefore be appropriate to apply a maximum.

Algorithms and R code for the determination of the parameters of the utility function are available from the authors free of charge.

Multiple solutions are possible. For this reason it may be necessary to use a range of starting values $\tilde{\gamma}_1$. It is also possible that no solution exists. If this is the case, it would be necessary to question the appropriateness of the trustees’ attitudes to risk. If constant relative risk aversion is required the determination of the parameters of the utility function may be simplified accordingly.

The approach suggested in this section presupposes two postulates: first, that the trustees accept the axioms of expected-utility theory as normative, and second, that the elicitation process does not introduce any bias into the measurement of trustees' risk aversion. As argued in Thomson (2003a), whilst the use of expected-utility theory is not uniquely rational, it is rational. There is therefore no reason why it should not be accepted for normative purposes.

The second postulate, whilst problematic in principle, need not necessarily be regarded as problematic in practice. In the first place, the method of elicitation avoids – or at least reduces – the problems of framing, continuity, the common ratio condition, probability weighting, biases in subjective probabilities and the reference effect. Second, problems such as the Allais problem, preference reversal, inconsistent orderings, ambiguity and violations of the independence axiom may be reduced by education. Details of these problems have been described in Thomson (2003a). Other problems raised in the literature have also been addressed in that paper.

It should, nevertheless, be recognised that there may be some remaining bias, and the risk aversions determined should not be regarded as accurate; they should rather be interpreted as reasonably indicative.

4.2. Dynamic asset allocation

As explained by Jarvis *et al.* (2009), the use of a long-term time horizon without allowance for dynamic asset allocation at shorter intervals is generally sub-optimal. This applies both to DC funds and to DB funds. An exception may occur if it is assumed that the trustees adopt constant relative risk aversion. However, as discussed in section 3 above, that assumption is too restrictive and cannot be made in this paper. Because the trustees of a retirement fund may revisit their asset-allocation decision regularly, dynamic modelling is required. In this section, the use of dynamic asset allocation is discussed.

As discussed in sections 2.1 and 2.2, the time interval to be used for dynamic programming will depend on the decision interval, both for DC funds and for DB funds. Without loss of generality it is assumed in this section that the decision interval is a year.

As shown by Mossin (1968), we may allow for annual optimisation by dynamic programming as follows. Let X_{it} denote the i th simulation of a vector representing the state of the world at time t for $i = 1, \dots, I$ and $t = 1, \dots, T$. The components of that vector comprise the variables needed to determine the values of z_T from X_{iT} at time T (namely A_T and a_T in equation (2) for a DC fund or A_T and P_T in equation (5) for a DB fund) as well as any other variables on which the value of $X_{i,t+1}$ will depend (e.g. L_t in equation (3) yield curves at time t and cumulative improvements in mortality to that time). We refer to these simulations as “primary simulations”. From each primary simulation i we select, for $t = 1, \dots, T-1$, the 10th, 30th, 50th, 70th and 90th percentiles of z_{it} , and the corresponding values of its components, which we refer to as \tilde{z}_{it} and \tilde{X}_{it} , respectively, for $i = 1, \dots, 5$.

For selected primary simulation i at time $t = T-1$ we may optimise for the final year before retirement by determining the maximum expected utility at that time, namely

$$u_{T-1}^*(\tilde{z}_{iT}) = \max_{\mathbf{g}_{i,T-1}} \left[E \left\{ u(\tilde{z}_{iT}) \mid \tilde{X}_{i,T-1}, \mathbf{g}_{i,T-1} \right\} \right] \quad (41)$$

where

$\mathbf{g}_{i,T-1} = \begin{pmatrix} g_{1i,T-1} \\ \vdots \\ g_{Ni,T-1} \end{pmatrix}$ is a vector representing the portfolio – i.e. the proportion invested in each investment channel available – at time $T-1$, such that

$$\sum_{n=1}^N g_{ni,T-1} = 1. \tag{42}$$

The function $u_{T-1}^*(\bullet)$ represents what Mossin (1968) refers to as an “indirect” or “derived” utility function at time $T-1$. For each selected primary simulation i the estimation of the value of that function requires a set of “secondary simulations” $j = 1, \dots, J$ of \tilde{X}_{iT} (say \tilde{X}_{ijt}), so that $\mathbf{g}_{i,T-1}$ may be calculated as the portfolio that maximises

$$E_{i,T-1}^* = \hat{E}_{T-1} \left\{ u(\tilde{z}_{iT}) \mid \tilde{X}_{i,T-1}, \mathbf{g}_{i,T-1} \right\} = \frac{1}{J} \sum_{j=1}^J u(\tilde{z}_{ijt}) \tag{43}$$

where \tilde{z}_{ijt} is calculated by updating $\tilde{X}_{i,T-1}$ to $\tilde{X}_{i,t}$ for trial values of $\mathbf{g}_{i,T-1}$ until convergence to a maximum is obtained. That maximum then constitutes an estimate of $u_{i,T-1}^*(\tilde{z}_{iT})$ for the purposes of equation (41).

At time $t = T-2, T-3, \dots, 0$, as explained by Mossin (1968) we may obtain the indirect utility function as

$$u_{t-1}^*(\tilde{z}_{it}) = \max_{\mathbf{g}_{i,t-1}} \left[E \left\{ u_t^*(\tilde{z}_{it}) \mid \tilde{X}_{i,t-1}, \mathbf{g}_{i,t-1} \right\} \right] \tag{44}$$

As before, for that purpose we require secondary simulations. Here we need to calculate

$$E_{i,t-1}^* = \hat{E} \left\{ u_t^*(\tilde{z}_{it}) \mid \tilde{X}_{i,t-1}, \mathbf{g}_{i,t-1} \right\} = \frac{1}{J} \sum_{j=1}^J u_t^*(\tilde{z}_{ijt}) \tag{45}$$

For this purpose, we need values of the function $u_t^*(\mathbf{X})$ for all $\mathbf{X} \in (\mathbf{X}_{i1t}, \dots, \mathbf{X}_{ijt})$. But we only have those values that have been simulated during the primary simulation, i.e. $\mathbf{X} \in (\mathbf{X}_{1t}, \dots, \mathbf{X}_{5t})$. The required values must be determined by fitting an indirect utility function to the selected values $\mathbf{X} \in (\mathbf{X}_{1t}, \dots, \mathbf{X}_{5t})$.

For this purpose, we need to use prudent utility functions for the maximisation of indirect expected utilities at the beginning of each year. This is because type-2 prudence must be applied not only to the asset allocation decision at the time horizon, but also to those at the beginning of each prior year. We therefore obtain a prudent utility function from the selected simulations. This is done in the same manner as for the elicitation of such a utility function from trustees or members, as explained in section 4.1 above, using J 1-year-ahead secondary simulations for each of the five selected values \tilde{z}_{ijt} and the corresponding values of E_{it}^* to determine the parameters of the indirect utility function $u_t^*(\bullet)$ and hence, for each of the five selected simulations, the optimal asset allocation $\mathbf{g}_{i,t-1}$ and $E_{i,t-1}^*$ as in equation (45).

Although the utility function is expressed in terms of the outcome at the time horizon, the indirect utility function is expressed from year to year in terms of the outcome over the following year; it reflects what the trustees’ 1-year-ahead utility functions will be. This means that the shape – and

therefore the parameters – of the latter may differ from that of the former. Nevertheless, it must still conform to the requirements of prudence.

From the maximisation for $t = 0$ we obtain an estimate of the value of \mathbf{g} that represents the optimal asset allocation during the period $[0, 1]$. The method is the same as for $t = T - 2, \dots, 1$, except that $\tilde{X}_{i,t-1} = X_0$ is known. Here we need to determine the portfolio \mathbf{g}_0 that maximises

$$E_0^* = \hat{E}\{u_1^*(z_1) | X_0, \mathbf{g}_0\} = \frac{1}{J} \sum_{j=1}^J u_1^*(z_{j1}). \quad (46)$$

As noted in section 2.2, it may be questioned whether the choice of T as the date of expiry of the liabilities is prejudicial to those members and pensioners as at time 0 whose pensions will cease long before T . However, the use of dynamic asset allocation addresses this issue by defining utility functions along the whole path from time 0 to time T ; all that is required at time 0 is that the trustees find an optimal allocation during the period $[0, 1]$. The utility function used for that purpose is an indirect utility function, so that for a fund with young members it will differ from that for a fund with pensioners only. Nevertheless, minimum and maximum constraints may be applied to the parameters of the indirect utility function, so as to avoid unrealistic indirect utility functions. It may also be questioned whether the accumulation of the assets to time T is preferable to discounting to time 0. In fact this is immaterial; if the numerator and the denominator were both discounted from time T to time 0 then the value of the ratio would be unaffected.

The values of I and J must be determined so that the process converges to a satisfactorily stable result.

5. Illustrative Results for a DB Fund: Resolving the Paradox of Counter-Intuitive Results

In this section, illustrative results from the application to a DB fund of the method proposed in this paper are presented and the sensitivity of asset allocations to the assumptions made is analysed. In section 5.1, the matter of counter-intuitive results is discussed. In section 5.2, the model used in the illustrative approach is formulated. In section 5.3, we specify the parameters and in section 5.4 the results are presented and discussed.

5.1. Counter-intuitive results

As argued in section 3.1, it would not be prudent for a trustee to adopt lower risk aversion when a DB fund is in shortfall than when it is in surplus. If a person takes more risk when his/her assets fall short of his requirements than when they exceed his requirements, he may well be acting rationally, but in terms of neo-classical economics, his preferences would generally be described as aspirational (Diecidue & Van de Ven, 2008). For example, trustees may prefer to lock in surpluses by investing more conservatively when the fund is in surplus. In itself such a policy would not be imprudent. But if, when the fund is in shortfall, the trustees go for broke by adopting a deliberately riskier policy than they would when the fund is in surplus, then they may justifiably be accused of attempting to save their own reputations in conflict with the requirements of prudence. As shown in the behavioural-finance literature, members may well exhibit aspirational preferences. However, trustees should not. Though aspirational behaviour may be intuitively appropriate to a member, it cannot be treated as intuitively appropriate to a trustee. We therefore refer to findings that support aspirational behaviour by trustees as “counter-intuitive”.

One would expect that, if the trustees' utility function shows higher relative risk aversion when the fund is in shortfall, then their exposure to the risky asset should be lower. In general, the literature finds that, in a DB fund, for certain arguments and forms of the utility function, the asset allocation is counter-intuitive: at lower levels of funding, optimal investment allocations require greater exposure to risky assets when the fund is in shortfall than when it is in surplus (Boulier *et al.*, 1995; Siegmann & Lucas, 1999; Cairns, 2000; Haberman *et al.*, 2003; McCarthy & Miles, 2013). In this section, the implications of these findings for the application of type-2 prudence are explored.

A problem with some of the approaches in the literature is that they treat the minimisation of contributions as the objective. But that is not a concern to members. In fact high contributions from the sponsoring employer are advantageous to the security of members' benefits. Whilst it is appropriate to take cognisance of the interests of the sponsoring employer, it is not appropriate to disregard those of members. The DB benefit ratio such as that proposed in section 2.2 effectively embraces both the interests of members and those of the employer and the proposed adjustments allow the trustees to take the interests of both parties into account. The emergence of asset allocations that are counter-intuitive in terms of members' interests is therefore not surprising.

More importantly, though, the objective functions used in the literature do not generally conform to the requirement of type-2 prudence.

Because of the problems outlined above, it cannot be concluded that the dynamic use of a prudent utility function of the DB benefit ratio would necessarily produce counter-intuitive investment allocations. However, the circumstances (if any) in which type-2 prudence would produce such results must be considered.

5.2. Formulation of the model

Suppose for the purposes of illustration that there are two assets available: an asset matching the fund's liabilities for members' benefits and a risky asset. From the point of view of the trustees the former may be regarded as risk-free: if the assets comprised just enough of the matching asset to match the liabilities, then the fund itself would be risk-free. The value of those assets and the payments from those assets will at all times be equal to the corresponding value and payments in respect of the liabilities. (In practice, such an asset will not generally be available; practical applications will require the modelling of all the assets available. But such a model here would have complicated the analysis unnecessarily.) Let δ_t denote the expected force of growth of the value of the liabilities (and therefore the expected force of return on the matching asset) during year t (where $t = 1$ denotes the forthcoming year). Let p_t denote the amount of the benefits payable during year t , allowing for expected future increases in salaries and pensions and for future mortality on the basis of specified assumptions. Let ξ_t denote the actual increase in the benefits payable during that year in excess of the expected increase. That excess may be positive or negative. We assume that the expected value of ξ_t is 0 and that these increases follow a Markov process. Let h_t be the force of return on the risky asset in year t in excess of that on the matching asset. This means that the force of return on the fund's risk-free asset is $\delta_t + \xi_t$ and the force of return on the risky asset is $\delta_t + h_t$.

Then the amount of benefits payable during year t , before allowance for increases during that year, is

$$\tilde{p}_t = p_t S_t \quad (47)$$

where

$$S_t = \begin{cases} 1 & \text{for } t = 1 \\ \exp\left(\sum_{s=1}^{t-1} \xi_s\right) & \text{for } t > 1. \end{cases} \quad (48)$$

Suppose for convenience that the benefits paid during that year comprise the amount of $\frac{1}{2}\tilde{p}_t$ at the beginning of the year and the amount of $\frac{1}{2}\tilde{p}_t \exp(\xi_t)$ at the end.

For simplicity of illustration, suppose that $j = 1$ (so that all the surplus is available to members) and that $l = 0$ (so that no weighting is applied to the sponsoring employer). This means that $k = 0$ and we may use equation (3) instead of equation (7).

From equation (3), the value of the DB benefit ratio at the end of year t will be

$$z_t = \frac{A_t + P_t}{L_t + P_t} \quad (49)$$

where

$$\begin{aligned} A_t &= \left(A_{t-1} - \frac{1}{2}\tilde{p}_t \right) \exp\{\delta_t + (1 - \alpha_t)\xi_t + \alpha_t h_t\} - \frac{1}{2}\tilde{p}_t \exp(\xi_t) \text{ for } t > 0 \\ &= A_0 \text{ for } t = 0 \end{aligned} \quad (50)$$

is the value of the assets at the end of year t ;

$$\begin{aligned} P_t &= \left\{ P_{t-1} \exp(\delta_t) + \frac{1}{2}\tilde{p}_t (1 + \exp(\delta_t)) \right\} \exp(\xi_t) \text{ for } t > 0 \\ &= 0 \text{ for } t = 0 \end{aligned} \quad (51)$$

the value at the end of year t of the benefits paid from year 1 up to and including year t ;

$$\begin{aligned} L_t &= \left\{ L_{t-1} \exp(\delta_t) - \frac{1}{2}\tilde{p}_t (1 + \exp(\delta_t)) \right\} \exp(\xi_t) \text{ for } t > 0 \\ &= L_0 \text{ for } t = 0 \end{aligned} \quad (52)$$

the value at the end of year t of the liabilities in respect of benefits payable after that year; α_t the proportion of the fund's assets invested in the risky asset; A_0 the value of the assets at the end of year 0; L_0 the value of the liabilities at the end of year 0, i.e. the price of the matching asset; and T the year of expiry of the liabilities; i.e. the year for which $p_t = 0$ for $t > T$.

In the definition of P_t it is assumed for convenience that the beneficiary's risk-free rate during year t is equal to the fund's expected risk-free rate during that year.

Thus

$$\begin{aligned}
 z_t &= \frac{A_t + P_t}{L_t + P_t} \\
 &= \frac{(A_{t-1} - \frac{1}{2}\tilde{p}_t)\exp(\delta_t + (1 - \alpha_t)\xi_t + \alpha_t h_t) + (P_{t-1} + \frac{1}{2}\tilde{p}_t)\exp(\delta_t + \xi_t)}{(L_{t-1} - \frac{1}{2}\tilde{p}_t)\exp(\delta_t) + (P_{t-1} + \frac{1}{2}\tilde{p}_t)\exp(\delta_t)} \\
 &= \frac{(A_{t-1} - \frac{1}{2}\tilde{p}_t)\exp((1 - \alpha_t)\xi_t + \alpha_t h_t) + (P_{t-1} + \frac{1}{2}\tilde{p}_t)\exp(\xi_t)}{L_{t-1} + P_{t-1}} \text{ for } t > 0 \\
 &= \frac{A_0}{L_0} \text{ for } t = 0.
 \end{aligned} \tag{53}$$

In particular

$$z_T = 1 + \frac{A_T}{P_T}. \tag{54}$$

The accumulation in the numerator of equations (3) and (49) is the same as in the denominator, so, as noted in section 4.2, the ratio is independent of δ_t . This means that, for the purposes of optimisation, we may redefine A_t , P_t and L_t so that $\delta_t = \delta$ is a constant. From equation (52) it follows that δ is such that

$$L_{t-1} = \{L_t + \frac{1}{2}\tilde{p}_t \exp(\xi_t)\} \exp(-\delta - \xi_t) + \frac{1}{2}\tilde{p}_t$$

i.e.

$$L_0 = \sum_{t=1}^T \frac{1}{2}\tilde{p}_t \{ \exp(\delta + \xi_t) + \exp(\xi_t) \} \exp\left\{ -\sum_{r=1}^t (\delta + \xi_r) \right\} \tag{55}$$

where L_0 is the value of the liabilities (i.e. the price of the matching asset) at time 0. Alternatively, δ may be defined by the user and L_0 calculated from equation (55). This feature constitutes a further advantage of the proposed DB benefit ratio: because at the beginning of a year the expected return on the fund’s risk-free asset is known, it should not affect the optimum asset allocation for that year. The proposed benefit ratio satisfies that criterion.

Equations (50), (51) and (52) may then be used to determine A_t , P_t and L_t for $t = 1, \dots, T$. By definition we shall have $L_T = 0$.

The value of α_1 that gives the greatest indirect utility at the beginning of year 1 may be found by dynamic programming as explained in section 4.2. In this application, for a given value of z_{T-1} we may determine the value of α_T (say α_T^*) that maximises the expected utility at the end of year T . The expected utility for that value (say E_{T-1}^*) then constitutes the indirect utility for that value of z_{T-1} . We may determine such values for a range of values of A_{T-1} to give indirect utility function values $E_T^*(z)$. These values may then be used to solve for a WARRA-class utility function. This procedure may be continued for t from T to 1. For $t = 0$ the known value A_0 of the assets may be used. It may happen that the resulting parameters will give lower relative risk aversion than the utility function used for year T . Whilst it may be acceptable to reduce the risk-aversion parameters of the indirect utility functions below those of the terminal values γ_{0T} and $\gamma_{\infty T}$, minimum values should be applied. The risk-aversion parameters are therefore subject to minima of $\gamma_{0\min}$ and $\gamma_{\infty\min}$, respectively. Also, the weighting parameter c is subject to a minimum of 0 and a selected maximum c_{\max} . An algorithm for the determination of α_1^* and R code are available from the authors free of charge.

5.3. Parameters used

Suppose that the variables ξ_t and h_t are normally distributed with means μ_ξ and μ_h and standard deviations σ_ξ and σ_h , respectively. As p_t is the expected value of the benefits during year t before increases in salaries and pensions during that year and ξ_t is the excess of the benefit increase during year t over the expected increase, μ_ξ must be taken to be 0. For illustrative purposes, the algorithm was used to determine the parameters of the indirect utility function and the optimum values α_0^* for the following parameters

$$\begin{aligned}
 \gamma_{0T} &= 5 \\
 \gamma_{\infty T} &= 3 \\
 c_T &= 1 \\
 \gamma_{0\min} &= 4 \\
 \gamma_{\infty\min} &= 2 \\
 \mu_\xi &= 0 \\
 \sigma_\xi &= 0.05 \\
 \mu_h &= 0.035 \\
 \sigma_h &= 0.14 \\
 \delta &= 0.02 \\
 T &= 20 \\
 z_0 &= 1 \text{ and} \\
 \mathbf{p} &= \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{20} \end{pmatrix} = \begin{pmatrix} \frac{40}{20 \times 21} \\ \frac{38}{20 \times 21} \\ \vdots \\ \frac{2}{20 \times 21} \end{pmatrix}
 \end{aligned} \tag{56}$$

so that

$$P_{20} = \sum_{t=1}^{20} p_t = 1 \tag{57}$$

and p_t decreases linearly.

This set of assumptions is referred to below as the “standard” set.

A_0 is thus equal to the value of the liabilities for future payments, i.e. the value of those payments before allowance for growth, discounted at the expected risk-free rate δ or, alternatively, the value of those payments including allowance for growth, discounted at the fund’s risk-free rate $\delta + \xi_t$ from time to time. The values of γ_{0T} , $\gamma_{\infty T}$, $\gamma_{0\min}$ and $\gamma_{\infty\min}$ are informed by the values found in section 3.2.

Stopping conditions were used that generally produced reasonable convergence. It was found, however, that better convergence could be obtained by rerunning the program numerous times and averaging the results. For positive values of σ_ξ – i.e. where future increases in benefits are uncertain – it was found that convergence was relatively slow. This is because the dimensionality of the problem

is effectively increased: at each node of the primary simulations the state of the world depends not only on the excess returns earned on the risky asset but also on the increases in salaries and pensions.

5.4. Results

The parameters of the indirect utility functions are shown in Figure 3. Some smoothing was applied so as to avoid spurious fluctuation due to random effects. From that figure it may be noted that, as t decreases from T to 1 (i.e. following the curve backwards), the values initially decrease towards their minima, after which they increase to or beyond the values at time T . The values at time 1 represent the values that will implicitly be applied by the trustees for deciding on the asset allocation during the 1st year.

The resulting initial exposure to the risky asset was 0.52. In considering this result it should be borne in mind that the assumption that $A_0 = L_0$ implies 100% funding on a risk-free basis, which is quite a strong financial condition.

This means that, for the model of investment returns assumed, and for a strong fund with future liability payments decreasing linearly in real terms, even under the rigorous requirements of type-2 prudence, the optimal exposure to the risky asset is over 50%.

The sensitivity of this result to the assumptions was also tested. Figure 4 shows the sensitivity of the optimal exposure to the risky asset to the ultimate risk-aversion parameters γ_{0T} and $\gamma_{\infty T}$, (“gamma.0.T” and “gamma.inf.T”, respectively). For each parameter, the result on the standard set of assumptions is shown by a large square marker; the lines show results for alternative assumptions for that parameter, assuming that the other parameters are unchanged. Because of the limits of the various parameters relative to each other, the ranges of the various parameters are restricted; the intention is merely to give an indication of the effects of changes to the respective parameters one by one.

It may be seen from Figure 4 that increasing values tend to result in lower optimal exposures to the risky asset, which is intuitive.

The sensitivity of the optimal exposure to the minimum values γ_{0min} and $\gamma_{\infty min}$ was found to be very low.

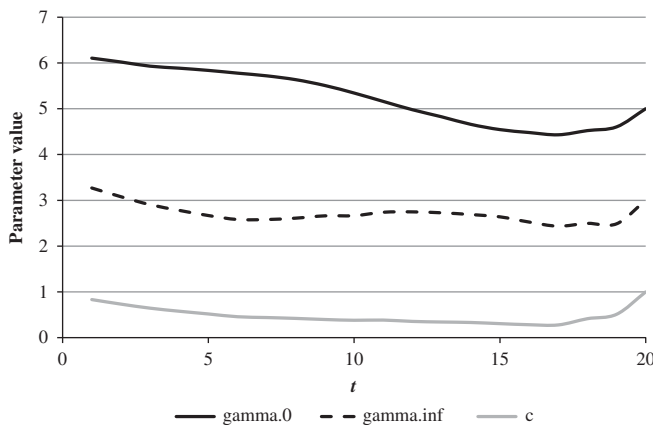


Figure 3. Parameters of the indirect utility functions.

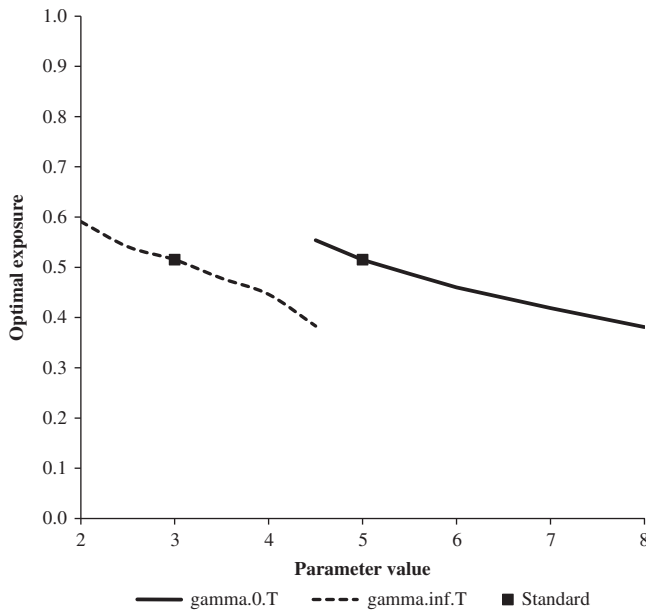


Figure 4. Sensitivity of the optimal exposure to the risky asset to the risk-aversion parameters of the utility function.

Figure 5 shows the sensitivity of the optimal exposure to the risky asset to the weighting parameters c_T and c_{\max} (“c.T” and “c.max”, respectively).

Increases in c_T result in slight increases in the optimal allocation to the risky asset. This is also intuitive, higher values of c_T place greater emphasis on the lower risk aversion represented by $\gamma_{\infty T}$. This sensitivity is low, because the values of c_t in the implied utility functions in earlier years are affected more by the optimal values of $\gamma_{\infty t}$ relative to those of γ_{0t} than by the ultimate value c_T . In effect, because of the long period to time T , variations in the value of c_T are neutralised. As shown in Figure 3, c_t may reduce quite considerably for shorter periods to expiry, so that the ultimate value c_T is not very relevant. The optimal allocation is not sensitive to c_{\max} ; where the limit applies it tends to be offset by lower levels of $\gamma_{\infty t}$.

Figure 6 shows the sensitivity of the optimal exposure to the risky asset to the parameters μ_b (“mu.h”) and σ_b (“sigma.h”) of the distribution of the excess return on that asset, and its sensitivity to the parameter σ_{ξ} (“sigma.xi”) of the distribution of increases in benefits.

These results for the parameters of the risky asset are intuitive; as expected returns on risky assets increase, the exposure to them increases and as their volatility increases, exposure to them decreases. In fact, following the findings of Jones *et al.* (2013) regarding likely future investment returns under resource constraints, the lower levels of μ_b are more likely in the long term. If those findings are followed, the level selected will depend on the scenario adopted by the trustees for decision-making. Furthermore, under certain scenarios, Jones *et al.* (2013) contemplate considerably increased standard deviations of returns on investments. This will result in even lower optimal allocations to the risky asset.

The sensitivity of the optimal exposure to the risky asset to the parameter σ_{ξ} of the distribution of increases in benefits is much lower. The trustees cannot offset variability from this source by reducing

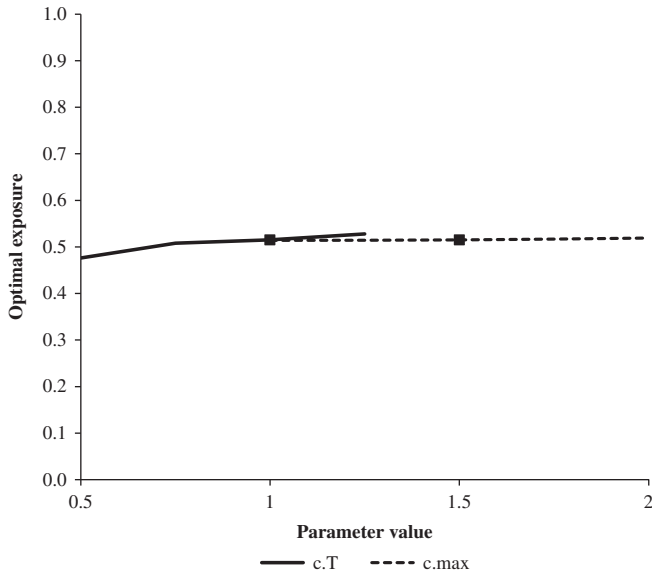


Figure 5. Sensitivity of the optimal exposure to the risky asset to the weighting parameters of the utility function.

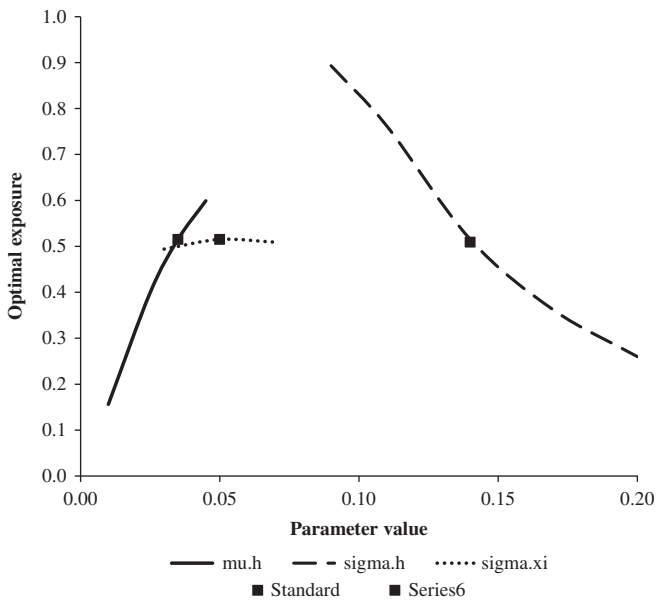


Figure 6. Sensitivity of the optimal exposure to the matched asset to the parameters of the distribution of the excess return on that asset.

their exposure to the risky asset. This is because there is no correlation between unexpected increases (and decreases) to pensions due to increases in salaries, pensions in payment and mortality and returns on the risky asset. If allowance were made for such correlation the sensitivity of the results to this parameter could be expected to be greater.

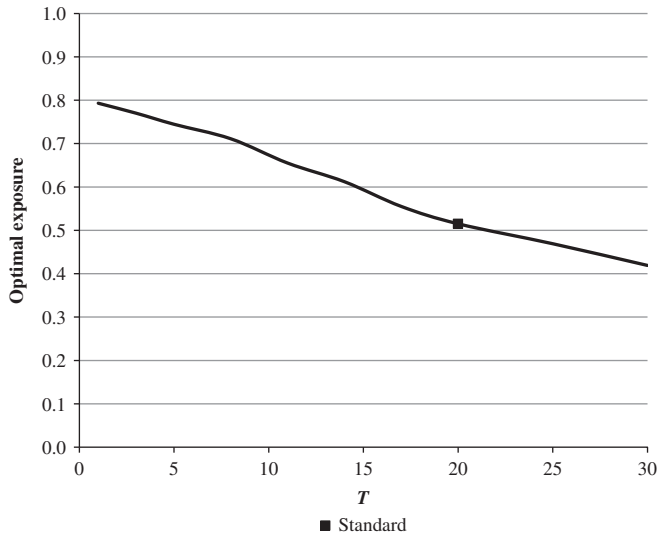


Figure 7. Sensitivity of the optimal exposure to the risky asset to the term to expiry of the liabilities.

Figure 7 shows the sensitivity of the optimal exposure to the risky asset to the term to expiry of the liabilities (T). Some smoothing has been applied.

The downward trend is counter-intuitive; according to conventional wisdom one would expect that the expected advantages of investing in risky assets would emerge over the long term. The reason for this effect is that, as shown in Figure 3, for longer terms to expiry (i.e. when t is much < 20), the relative risk-aversion factors in the indirect utility functions at time 0 are generally greater than for shorter terms (i.e. when t is closer to 20). The virtually linear relationship is remarkable.

Figure 8 shows the sensitivity of the optimal exposure to the initial funding ratio z_0 , i.e. the asset value at time 0 per unit of the value of the liabilities at that time. The “high prudence” curve shows results with amended risk-aversion parameters: γ_{0T} has been increased and $\gamma_{\infty T}$ reduced so as to increase the downward slope in relative risk aversion and therefore the level of prudence. The parameters used are shown in Table 1. Parameters other than those shown are unaltered from the standard assumptions.

For initial funding ratios below 1, the optimal exposure on the standard assumptions increases as the initial funding ratio increases. This is intuitive. But for initial funding ratios above 1 it remains almost constant. The latter effect, like the effects found in the literature as observed above, is counter-intuitive; in terms of neo-classical economics one would expect that, as the funding ratio increases, the exposure to risky assets would increase. However, unlike the findings in the literature, the exposure to risky assets does not decrease; it merely stays constant. This is because the counter-intuitive effect is offset by the reduction in relative risk aversion as the funding ratio increases.

However, when higher levels of prudence are considered the picture changes. For the “high prudence” curve in Figure 8, the risk-aversion parameters for low funding levels have been increased and those for high funding levels have been decreased. This means that relative risk aversion decreases at a higher rate as the initial funding ratio increases. The result is intuitive: it suggests that higher the

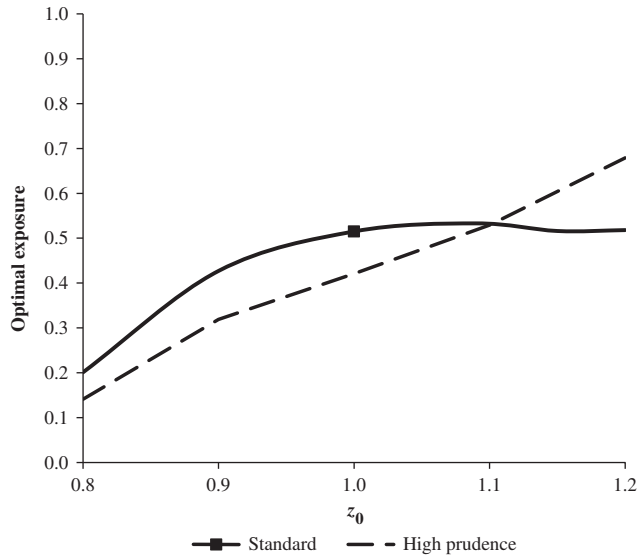


Figure 8. Sensitivity of the optimal exposure to the risky asset to initial funding ratio.

Table 1. Parameters used in Figure 8.

Parameter	Standard	High prudence
γ_{0T}	5	10
$\gamma_{\infty T}$	3	2
γ_{0min}	4	6
$\gamma_{\infty min}$	2	1.5

funding level, the higher the optimal exposure to risky asset becomes. This effect becomes stronger than the counter-intuitive effect discussed above. Ultimately, at initial funding ratios above 1.1, the level of relative risk aversion under the “high-prudence” assumptions in Table 1 becomes so low that the exposure to risky assets exceeds that under the standard assumptions.

This resolves the paradox in the findings of the articles discussed above. Whether the level of prudence reflected in Table 1 – or some level intermediate between the standard value and the high-prudence value – is justified is a matter of judgement, which should be informed with reference to members’ levels of risk aversion as discussed in section 3.2 above.

6. Summary and Suggestions for Further Research

6.1. Summary

In this paper, a distinction is drawn between “type-1 prudence” and “type-2 prudence”: the former relates to the adoption of deliberately conservative assumptions for the valuation of a retirement fund, whereas the latter, on which the paper focuses, relates to the risk aversion of the fund’s trustees. It is argued that, to satisfy the requirements of type-2 prudence, expected-utility theory may be used as a normative theory for the purposes of asset allocation by the trustees of retirement funds.

For that purpose, definitions of the arguments of the utility function for DC funds and for DB funds are proposed. These are referred to as the “DC benefit ratio” and the “DB benefit ratio”, respectively.

In order to place greater or less weight on the sponsoring employer’s interests relative to those of members, a method of modifying the DB benefit ratio is proposed.

It is found that the utility functions generally used for retirement-fund decision-making are problematic. Furthermore, there is little consensus in the literature about the levels of risk aversion that are appropriate for the trustees of a retirement fund. In order to address these problems, and in particular to satisfy the requirements of type-2 prudence, it is proposed that the trustees’ utility function $u(\bullet)$ should conform to five criteria. A class of utility functions (the “WARRA” class) that meets these requirements is introduced and its properties are discussed and shown to be satisfactory in terms of the above criteria. This class includes constant relative risk aversion as a special case.

Methods of determining the risk-aversion parameters of a WARRA-class utility function are described with reference to values elicited from the trustees of a retirement fund. In particular, it is shown how, for a particular fund, they may be determined with reference to the utility functions elicited from a sample of members of the fund.

In practice, spurious accuracy in the determination of the parameters of the trustees’ utility function should be avoided. Sensitivity tests may be made so as to establish the extent to which marginal changes in those parameters will affect the decisions to be made. This may reduce the difficulties that trustees might experience in reaching consensus or a compromise with regard to the adoption of a utility function.

In order to apply the WARRA class of utility functions to retirement-fund asset-allocation decision-making, dynamic asset allocation is required. A method of implementing dynamic programming for this purpose is outlined.

Results were obtained from an illustrative formulation of the model for a DB fund with illustrative parameter values. The sensitivity of those results to changes in the parameter values were analysed.

Counter-intuitive results found by other authors suggest that the trustees of a DB retirement fund should invest more in risky assets when the fund is in shortfall and less when it is in surplus.

For initial funding ratios below 1, the optimal exposure on the standard assumptions increases as the initial funding ratio increases. This is intuitive. But for initial funding ratios above 1 it remains almost constant. The latter effect, like the effects found in the literature, is counter-intuitive. However, unlike the findings in the literature, the exposure to risky assets does not decrease; it merely stays constant. This is because the counter-intuitive effect is offset by the reduction in relative risk aversion as the funding ratio increases.

However, when higher levels of prudence are considered, relative risk aversion decreases at a higher rate as the initial funding ratio increases. The result is intuitive: it suggests that the higher the funding level, the higher the optimal exposure to risky asset becomes. This effect becomes stronger than the counter-intuitive effect in the standard case. Ultimately, at initial funding ratios above 1.1, the level of relative risk aversion under the high-prudence assumptions becomes so low that the exposure to

risky assets exceeds that under the standard assumptions. This resolves the paradox in the findings of the articles referred to above.

The findings obtained from the illustrative examples in this paper cannot be regarded as generalised results. Whilst they serve to illustrate the effects of variations in the parameters on the asset-allocation decision in a DB fund, they do not show the effects of combinations of variations. Practitioners are therefore advised to apply their own sensitivity tests. For that purpose the algorithms and R code available from the authors may be useful. The algorithms and R code for a DC fund are similar.

6.2. Suggestions for further research

Further research on levels of relative risk aversion both amongst members of retirement funds and amongst trustees would be helpful, particularly research on explanatory variables that might be used to inform or predict the parameters of WARRA-class utility functions for use by trustees.

Further analysis of the implications of the proposed method for asset allocation by the trustees of retirement funds is also a matter for further research.

The application of WARRA-class utility functions to the pricing of the unhedgeable components of the liabilities of a DB retirement fund in an incomplete market could be explored. For this purpose, certainty equivalents could be used, which could be based on the trustees' utility function.

In this research, a simple model of returns on risky assets was used. In practice, more complex models may be used. The modelling of contributions by the sponsoring employer to fund shortfalls from time to time would also result in greater complexity, especially the allowance for correlation between the financial condition of the sponsoring employer and that of the fund. Such models require further research.

Whilst the focus of this research was on the form of the utility function rather than on the model of returns on risky assets, it has been noted that the findings of Jones *et al.* (2013) with regard to the effects of resource constraints on future returns on risky assets would have a substantial effect both on the parameters of such models and on the optimal exposures to risky assets. At present there are no models of such effects. The development of such models is a matter of urgency.

Acknowledgements

The authors acknowledge the advice of Megan Butler on certain aspects of this paper. The financial assistance of the Actuarial Society of South Africa is acknowledged. Opinions expressed and conclusions drawn are those of the authors and are not to be attributed to the Society. Valuable criticism by anonymous referees is also acknowledged.

References

- Anderson, J.R., Dillon, J.L. & Hardaker, B. (1977). *Agricultural Decision Analysis*. Iowa State University Press, Ames cited in Farquhar (1984).
- Biasca, C. (2009). Strategic asset allocation for pension plans, unpublished dissertation for M.Sc., University of Neuchâtel, Neuchâtel.
- Booth, P.M. (1995). The management of investment risk for defined contribution pension schemes. *Transactions of the International Congress of Actuaries*, 25, 105–120.

- Bossons, J. (1973). The distributions of assets among individuals of different age and wealth. In R. Goldsmith (Ed.), *Institutional Investors and Corporate Stock* (pp. 394–428). National Bureau of Economic Research, New York.
- Boulier, J.-F., Trussant, E. & Florens, D. (1995). A dynamic model for pensions funds management. Proceedings of the 5th AFIR International Colloquium, September, 1995, Leuven, Belgium, 1, 361–384.
- Cairns, A.J.G. (2000). Some notes on the dynamics and optimal control of stochastic pension fund models in continuous time. *Astin Bulletin*, 30(1), 19–55.
- 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, 20th to 23rd June, 2000, also working paper, Heriot-Watt University Edinburgh.
- Cardinale, M., Katz, G., Kumar, J. & Orszag, J. M. (2006). Background risk and pensions. *British Actuarial Journal*, 12(1), 79–134.
- Cohn, R., Lewellen, W., Lease, R. & Schlarbaum, G. (1975). Individual investor risk aversion and investment portfolio composition. *Journal of Finance*, 10(2), 605–620.
- Cowling, C.A., Gordon, T.J. & Speed, C.A. (2005). Funding defined benefit pension schemes. *British Actuarial Journal*, 11(1), 63–131.
- Deelstra, G., Grasselli, M. & Koehl, P.F. (1999). Optimal investment strategies in a CIR framework. Preprint.
- Diecidue, E. & Van de Ven, J. (2008). Aspiration level, probability of success and failure, and expected utility. *International Economic Review*, 49(2), 683–700.
- Farquhar, P.H. (1984). Utility assessment methods. *Management Science*, 30, 1283–1294.
- Friend, I. (1973). Methodology in finance. *Journal of Finance*, 28(2), 257–272.
- Gerber, H.U. & Shiu, E.S.W. (2000). Investing for retirement: optimal growth and dynamic asset allocation. *North American Actuarial Journal*, 4(2), 42–62.
- Haberman, S., Day, C., Fogarty, D., Khorasane, M.Z., McWhirter, M., Nash, N., Ngwira, B., Wright, I.D. & Yakoubov, Y. (2003). A stochastic approach to risk management and decision making in defined benefit pension schemes. *British Actuarial Journal*, 9(3), 493–618.
- Jarvis, S., Lawrence, A. & Miao, S. (2009). Dynamic asset allocation techniques. *British Actuarial Journal*, 15(3), 573–677.
- Jones, A., Allen, I., Silver, N., Cameron, C., Howarth, C. & Caldecott, B. (2013). Resource constraints: sharing a finite world: implications of limits to growth for the actuarial profession: the evidence and scenarios for the future. Presented by the Institute and Faculty of Actuaries, 17 January. Available online at the address <https://www.actuaries.org.uk/documents/resource-constraints-sharing-finite-world-evidence-and-scenarios-future>, accessed 29/6/2016.
- Khorasane, M.Z. & Smith, D.A. (1997). A utility maximisation approach to individual investment choices in a money purchase pension scheme. Proceedings of the Investment Conference, 2, 261–289.
- Kimball, M.S. (1990). Precautionary saving in the small and in the large. *Econometrica*, 58(1), 53–73.
- Levitan, S. (2006). A practical examination of Thomson's expected utility framework and its application for investment channel choice in defined contribution funds, unpublished dissertation for M.Sc., University of the Witwatersrand, Johannesburg.
- Levitan, S. & Thomson, R.J. (2009). The application of expected-utility theory to the choice of investment channels in a defined-contribution retirement fund. *ASTIN Bulletin*, 39(2), 615–647.
- Levy, H. (2005). Absolute and relative risk aversion: an experimental study. *Journal of Risk and Uncertainty*, 8(3), 289–307.
- McCarthy, D. & Miles, D. (2013). Optimal portfolio allocation for corporate pension funds. *European Financial Management*, 19(3), 599–629.

- Merton, R.C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics*, 51(3), 247–257.
- Merton, R.C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3, 373–413.
- Merton, R.C. (1993). *Continuous-time Finance*, revised edition. Blackwell, Cambridge MA.
- Mossin, J. (1968). Optimal multiperiod portfolio policies. *Journal of Business*, 41, 215–229.
- Nielsen, P.H. (2006). Optimal investment strategies: a short survey of classical and recent results. *Proceedings of the 28th International Congress of Actuaries*, 1B.
- Pézier, J. & Scheller, J. (2011). A comprehensive evaluation of portfolio insurance strategies. Available online at the address <http://dx.doi.org/10.2139/ssrn.1864533>.
- Pratt, J. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1), 122–136.
- Projector, D.S. & Weiss, G.S. (1966). *Survey of Financial Characteristics of Consumers*. Federal Reserve System, Washington, DC.
- Rubinstein, M. (1976). The strong case for the generalized logarithmic utility model as the premier model of financial markets. *Journal of Finance*, 31(2), 551–571.
- 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*, 2nd edition. Wiley, Dover, NY.
- Sharpe, W.F. & Tint, L.G. (1990). Liabilities: a new approach. *Journal of Portfolio Management*, winter, 5–10.
- Sherris, M. (1992). Portfolio selection and matching: a synthesis. *Journal of the Institute of Actuaries*, 119(1), 87–106.
- Sherris, M. (1993). Portfolio selection models for life insurance and pension funds. 3rd AFIR International Colloquium, 2, 915–930.
- Siegmann, A.H. & Lucas, A. (1999). Continuous-time dynamic programming for ALM with risk averse loss functions. *Proceedings of the 9th AFIR International Colloquium*, Joint day proceedings, 183–193.
- Thomson, R.J. (2003a). The use of utility functions for investment channel choice in defined contribution retirement funds: I: defence. *British Actuarial Journal*, 9(3), 653–709.
- Thomson, R.J. (2003b). The use of utility functions for investment channel choice in defined-contribution retirement funds: II: a proposed system. *British Actuarial Journal*, 9(4), 903–958.
- Thomson, R.J. (2011). The arbitrage-free equilibrium pricing of liabilities in an incomplete market: application to a South African Retirement Fund. *South African Actuarial Journal*, 11, 1–41.
- Thornton, P.N. & Wilson, A.F. (1992). A realistic approach to pension funding. *Journal of the Institute of Actuaries*, 119(2), 229–312.
- Vigna, E. & Haberman, S. (2001). Optimal investment strategy for defined contribution pension schemes. *Insurance: Mathematics and Economics*, 28, 233–262.
- Von Neumann, J. & Morgenstern, O. (1947). *Theory of Games and Economic Behavior*, 2nd edition. Princeton University Press, Princeton, NJ.