Utilitarian versus neutralitarian design of endowment fund policies

Schumacher, J.M.

DOI
10.1080/03461238.2022.2025892

Publication date
2022

Document Version
Final published version

Published in
Scandinavian Actuarial Journal

Citation for published version (APA):
Schumacher, J. M. (2022). Utilitarian versus neutralitarian design of endowment fund policies. *Scandinavian Actuarial Journal*, *2022*(8), 718-748.
https://doi.org/10.1080/03461238.2022.2025892

Scandinavian Actuarial Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/sact20

Utilitarian versus neutralitarian design of endowment fund policies

Johannes M. Schumacher

To cite this article: Johannes M. Schumacher (2022) Utilitarian versus neutralitarian design of endowment fund policies, Scandinavian Actuarial Journal, 2022:8, 718-748, DOI: 10.1080/03461238.2022.2025892

To link to this article: https://doi.org/10.1080/03461238.2022.2025892

Published online: 30 Jan 2022.

SCANDINAVIAN ACTUARIAL JOURNAL
2022, VOL. 2022, NO. 8, 718–748
https://doi.org/10.1080/03461238.2022.2025892

Utilitarian versus neutralitarian design of endowment fund policies

Johannes M. Schumacher

Amsterdam School of Economics, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT

This paper addresses investment and spending policies of endowment funds aiming to generate a stable income stream in perpetuity. The standard academic approach to the design of such policies is based on optimization of utility aggregated over time. However, the explicit purpose of many funds to serve current and future generations 'in equal measure' suggests incorporation of a suitable notion of neutrality. The utilitarian and neutralitarian approaches are compared in two settings: one in which the preferences of individual generations are described by a standard CRRA utility function, and one in which these utility functions are modified by the introduction of a saturation level. Results are expressed in terms of the implied assumed interest rate (AIR), which reflects the apportionment of initially available capital to the time-0 values of individual future benefits. Under CRRA preferences, the neutralitarian point of view can be seen as a way of determining the discount factor that is used in the utilitarian method. When a saturation level is added, the neutralitarian and utilitarian policies are essentially different. The introduction of saturation generally induces a shift of value from earlier to later generations.

ARTICLE HISTORY
Received 9 August 2021
Accepted 3 January 2022

KEYWORDS
Endowment funds; saturation; fair division; variable perpetuity; assumed interest rate

1. Introduction

Variable annuities are financial products that provide the holder with an income stream that is linked to an equity index, until termination occurs due to a preset final date or to death of the holder. These instruments are used in particular for retirement income provision and are often equipped with riders such as the Guaranteed Minimum Withdrawal Benefit. An extensive introduction to variable annuities is provided by Dellinger (2006); recent research, focused on pricing, hedging, and welfare comparisons, includes Bacinello et al. (2011), Balter & Werker (2019), Delong (2014), Horneff et al. (2015), Koijen et al. (2011), Mahayni & Schneider (2012) and Trottier et al. (2018). It should be noted that variable annuities do not only arise as commercial products sold to individual consumers by insurance companies. Collective pension funds that provide benefits to retirees depending on financial market indices, for instance, via 'conditional indexation' (Kleinow & Schumacher 2017), effectively deliver a variable life-contingent annuity to participants.

In this paper, attention is focused on *perpetual* non-life-contingent variable annuities. Such products can be referred to as 'variable perpetuities'. A consumer market for variable perpetuities does not seem to exist; one of the reasons may be that few insurance companies can credibly claim to remain in existence forever. This does not mean, however, that the design of variable perpetuities is not an important issue. Billions of dollars worldwide are managed by endowment funds such as the Bill and

CONTACT Johannes Schumacher j.m.schumacher@uva.nl

Melinda Gates Foundation, the Getty Foundation, the Bertelsmann Stiftung, and the Nobel Foundation, to name just a few examples. Many of these institutions invest at least partly in risky assets and aim to generate a non-ending income stream while allowing for some fluctuations in annual payouts in response to variations in the return on the investment portfolio and possibly other factors relating to the specific purposes of the fund. In this way, endowment funds act, in effect, as providers of variable perpetuities, in the same way as collective pension funds that pay conditionally indexed benefits are effectively providers of variable annuities.

The policy design problem of endowment funds can be viewed as a problem of intertemporal social choice. A standard approach to this problem in the academic literature is to form a social welfare function, in which utilities of individual benefits[1] are aggregated by means of chosen weights. This 'utilitarian' point of view leads, in the case of endowment funds, to a problem formulation that is similar to usual formulations of investment problems with intermediate consumption for a single individual. An alternative approach that avoids the problem of selecting the weights is to apply a notion of 'equity' or 'fairness'. This can be called the 'neutralitarian' approach.[2] In a deterministic and atemporal setting, the principle of neutrality simply states that all agents should receive the same amount. In an intertemporal and stochastic context, one may still formulate a similar principle, in the sense that a certain functional of the random variable representing the uncertain future benefit should be the same for all agents. Which functional should be chosen for this purpose is open for debate. The objective of the present paper is to compare notions of aggregate utility and neutrality in the context of endowment fund policy. Attention is paid in particular to the balance between generations as expressed by the time-0 values of the stochastic future benefits that are implied by the chosen policy and the economic model. These time-0 values are represented in terms of the implied *assumed interest rate* (AIR), which is in general horizon-dependent.

The focus in this paper lies on endowment funds that aim for a benefit stream that stays more or less constant at a sufficiently high level in the course of the years, rather than funds that may spend much more in one year than in another, such as disaster relief funds. The notion of fairness used here is different from the concept of financial fairness as used in the study of a multiperiod risk-sharing problem by Bao et al. (2017). In the present paper, it is assumed that the capital provider(s) and the beneficiaries are different individuals, so that the notion of financial fairness does not apply.

In most of the literature on endowment funds (see, for instance, Cejnek et al. 2014 for a survey), there is no formalization of the notion of intergenerational fairness. An exception is Gilbert & Hrdlicka (2011), where fairness is defined via an aggregation of period utilities that deviates from the usual exponentially weighted sum. A similar approach is followed in Balter & Schweizer (2021) (in an atemporal context) using certainty equivalents rather than expected utilities. These approaches are still based on optimization, whereas, in the present paper, concepts of neutrality are proposed and compared to the frequently applied method of weighted-sum optimization.

The historical paper by Ramsey (1928) presents an early mathematical analysis of the intertemporal wealth allocation problem within a deterministic framework. The approach taken by Ramsey is based on optimization, but he does use considerations of fairness to argue that the discount rate in the intertemporal optimization problem should be taken equal to zero. Later authors, starting with Merton (1969), have mainly addressed the problem in the context of life cycle planning for individuals; this work is often based on aggregation of period utilities, typically using exponential weights with a nonzero discount rate. In the context of endowment funds, Tobin (1974, p. 427) expressed the principle that 'the trustees are supposed to have a zero subjective rate of time preference'. This appears to be close to Ramsey's point of view. The operational form that Tobin chooses for the principle, within a

[1] The term 'benefit' will generally be used in this paper to refer to the monies that are made available by an endowment fund for consumption in whatever form. Equivalent terms that may be found in the literature, and that will occasionally also be used in this paper, include 'spendings', 'disbursements', 'payouts', and 'withdrawals'. When benefits are viewed as contingent claims, they may be referred to as 'payoffs', in line with common terminology in mathematical finance.

[2] In the theory of distributive justice (see, for instance, Roemer 1998), the point of view of fairness is often referred to as the 'Rawlsian' perspective (Rawls 1971).

deterministic environment, is that spendings should be constant in inflation-corrected terms. Below, analogous principles are applied in a stochastic context.

Several authors have carried out policy evaluation studies for endowment funds, assuming an investment and benefit policy of a given type (for instance: a fixed-mix investment portfolio combined with benefits defined as a moving-window average of realized returns). These authors look at the effect of adjusting policy parameters on quantities of interest, such as the average level of benefits, the variability of benefits, and the probability of ruin. Examples of such studies include Pye (2017), Lindset & Matsen (2018), and Brown & Scholz (2019). Similar evaluation studies can also be found in the related literature on the management of collective pension funds, using either analytical or simulation methods; see, for instance, Dufresne (1990), Milevsky & Robinson (2000), and Blake et al. (2003). Closer to the present paper is the work of Balter & Werker (2019), who determine the AIR under exponentially weighted aggregate CRRA utility as well as under the neutralitarian policy of equal expectations, under the assumption that the investment policy is of the fixed-mix type. They also find the AIR under equal expectations for a smoothing investment policy. In the present paper, the investment policy is not preset but is rather determined as a consequence of a chosen design principle. A more technical difference is that, because perpetuities are considered rather than fixed-term annuities as in Balter & Werker (2019), it becomes possible to discuss asymptotic properties of the AIR.

Investment problems that include the design of a sharing rule between heterogeneous agents have been studied by several authors; see, for instance, Kryger & Steffensen (2010), Jensen & Nielsen (2016), Pazdera et al. (2016) and Chen et al. (2021). In these papers, it is allowed that agents are different in terms of preferences, but it is assumed that payoff takes place for all agents at the same time. The present paper studies a collective investment problem in which the heterogeneity of the collective is due to agents being placed differently in time, rather than being equipped with different preferences.

The paper is organized as follows. Section 2 introduces assumptions on the economic environment and presents the different design principles that will be considered in this paper. Section 3 contains a general result on the relation between, on the one hand, the asymptotic AIR, and on the other hand, the asymptotic probability that benefits will equal or exceed a given level. Section 4 covers the differences between utilitarian schemes on the one hand and neutralitarian schemes on the other hand for the case of utilities of standard isoelastic/CRRA type. Subsequently, the consequences of introducing a saturation level are investigated in Section 5. Section 6 presents conclusions and perspectives for further research. There is an appendix which contains proofs and details of calculations that are needed in the main text.

2. Preliminaries

2.1. The AIR

The benefit flow from a variable annuity is often described in terms of the assumed interest rate (AIR). It is sometimes said that the AIR represents the growth rate of the annuity, but that is an oversimplified and misleading explanation. A typical example of the use of the AIR occurs when the benefit b_j paid at time T_j is determined in terms of the benefit paid at time T_{j-1} by

$$b_j = \frac{1+z_j}{1+\rho}\, b_{j-1}, \tag{1}$$

where z_j is the return earned on a reference portfolio in the period from T_{j-1} to period T_j, and ρ is the AIR. The formula implies that the benefits will be constant if the returns on the reference portfolio are equal to the AIR. The benefit b_0 at time $T_0 = 0$, which also serves as the initial value for recursion (1), is determined from the budget constraint that is obtained as follows. Note that the payment at time

T_j can be written as

$$b_j = b_0(1+\rho)^{-j}\prod_{i=1}^{j}(1+z_i). \tag{2}$$

Payment of $\prod_{i=1}^{j}(1+z_i)$ units in period j can be realized, in all return scenarios, by investing one unit in the reference portfolio at time 0. Therefore, the benefit stream as a whole can be financed if b_0 is chosen that

$$V_0 = b_0\sum_{j=0}^{T}(1+\rho)^{-j},$$

where V_0 is the initially available capital and T is the terminal time of the annuity. In the case of a perpetuity, we have $T=\infty$, and consequently, the initial benefit b_0 is given by[3]

$$b_0 = \frac{\rho}{1+\rho}\,V_0. \tag{3}$$

For a given initial capital, increasing the AIR leads to higher initial benefits, but also to a higher probability that later benefits will suffer from disappointing returns on the reference portfolio.

Actual annuity products are usually more complicated than the simple scheme (1). However, the concept of AIR can be made applicable to *any* annuity product if the AIR is allowed to be horizon-dependent, rather than constant across all maturities, and if it can be assumed that a time-0 value can be assigned to all of the future benefits as specified by the product definition. Since the benefits in a variable annuity scheme are typically uncertain, this presumes the availability of a valuation model; in this paper, the standard Black–Scholes (BS) model will be used as such. In terms of time-0 values, the continuously compounded AIR for time T_j is defined by Balter & Werker (2019, Def. 2.2)[4]

$$\rho_j = -\frac{1}{T_j}\log\frac{v_j}{v_0}, \tag{4}$$

where v_j is the time-0 value of the (stochastic) benefit paid at time T_j.[5] To see that this agrees with usage in (2), note that the time-0 value of the payment defined by (2) is $v_j=(1+\rho)^{-j}b_0$. The *asymptotic AIR* is defined by

$$\rho_\infty = \lim_{j\to\infty}\rho_j \tag{5}$$

under the assumption that the limit exists.[6] By the budget constraint $\sum_{j=0}^{\infty}v_j \le V_0$, we must have $v_j \le v_0$ for all but finitely many values of j, so that the asymptotic AIR is nonnegative for all schemes. Generally speaking, the asymptotic AIR of a benefit scheme is an indication of the extent to which the interests of generations far into the future are taken into account. Higher values of the asymptotic AIR correspond to lower levels of generosity towards recipients in the distant future.

A simple policy that might be adopted by an endowment fund is to let the benefit paid each year be determined as a fixed fraction r_s of available capital (the 'spending rate'). In this case, the time-0 value of the benefit paid to generation j is equal to $r_s(1-r_s)^jV_0$. The corresponding AIR according to

[3] It is assumed that the first benefit is paid at the *start* of the first period. If it is paid at the *end* of the first period, then b_0 still appears as a parameter in (2), and its value is given by $b_0=\rho V_0$ instead of (3).

[4] The definition as stated applies for $j\ge 1$. If time-0 values are derived from an expression $v_j=v(T_j)$ where $v(T)$ is a differentiable function defined for all $T\ge 0$, then a natural definition of the AIR for $j=0$ is $\rho_0=-v'(0)/v(0)$).

[5] In this paper, only schemes are considered with $P(b_j<0)=0$ and $P(b_j>0)>0$ for all j. As a consequence, $v_j>0$ for all j.

[6] Many of the results in this paper can be generalized by using the limes superior/inferior (lim sup/inf) instead of the limit. The generalization appears to have little practical relevance, however; for simplicity, the existence of the limit will always be assumed. In particular, the limit does exist (and is computed explicitly) in all schemes that are analyzed in this paper.

definition (4) is $\rho = -\log(1 - r_s)$ so that the AIR for this scheme is constant and, for small values of r_s, approximately equal to r_s. The AIR as defined in (4) can, therefore, also be viewed as a generalized (possibly horizon-dependent) version of the spending rate.

2.2. Economic model

The Black–Scholes model will be used in this paper to model the investment opportunities. It will be convenient to assume that all prices are expressed in inflation-corrected units, so that the rate of inflation in the model is zero by definition. As a consequence, all statements related to growth are to be understood in real terms; 'expected return' is to be interpreted as 'expected real return', 'interest rate' is 'real interest rate', and so on. The adjective 'real' will be added for emphasis in a few cases, but will usually be suppressed in the interest of brevity. It is *not* assumed in this paper that the real interest rate is positive.

In the BS model, there are two assets available for trading, namely a risky asset with value S_t and a riskless asset with value B_t at time t. The evolution of the values of these assets is described by the differential equations

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t, \tag{6}$$

$$\mathrm{d}B_t = rB_t \,\mathrm{d}t, \tag{7}$$

where the drift parameter μ, the volatility $\sigma > 0$ and the real interest rate r are constants, and where W_t is a standard Brownian motion. The parameter λ defined by

$$\lambda = \frac{\mu - r}{\sigma} \tag{8}$$

is called the *Sharpe ratio* or the *market price of risk*.

By forming a fixed-mix portfolio that keeps a fraction a of total portfolio value in the risky asset and a fraction $1-a$ in the riskless asset, one obtains a new asset whose value V_t satisfies the differential equation

$$\mathrm{d}V_t = ((1 - a)r + a\mu)V_t \,\mathrm{d}t + a\sigma V_t \,\mathrm{d}W_t = (r + \lambda\sigma_V)V_t \,\mathrm{d}t + \sigma_V V_t \,\mathrm{d}W_t, \tag{9}$$

where $\sigma_V := a\sigma$. Since the portfolio volatility σ_V can be chosen arbitrarily by varying the parameter a, the formula above shows that the relevant parameters in the BS model, as an idealized model for describing investment opportunities, can be taken to be only r and λ. The market price of risk λ is the number of percentage points of expected return that is obtained for every additional percentage point of volatility that is accepted. It will be assumed throughout the paper that λ is positive.

The *pricing kernel process* ξ_t in the BS model is given by the stochastic differential equation

$$\mathrm{d}\xi_t = -r\xi_t \,\mathrm{d}t - \lambda\xi_t \,\mathrm{d}W_t, \quad \xi_0 = 1. \tag{10}$$

The time-0 price of a time-T stochastic payoff X_T in the BS model is given by $E[\xi_T X_T]$. In the BS model, both the pricing kernel ξ_T at time T and the value of the risky asset S_T at time T are monotonically related to the value taken by the driving Brownian motion W_T at time T. Therefore, any payout expressed in terms of the pricing kernel ξ_T (see for instance (12)) can be re-expressed in terms of the risky asset value S_T. The payout then appears as a contingent benefit, with S_T as the value of the reference portfolio that determines the payment that will be made.

2.3. Optimality and fairness

As argued in the introduction, endowment funds can be viewed as the typical providers of variable perpetuities. Such funds are not all similar; for instance, the spendings of funds aimed at providing relief in case of calamities may vary greatly from year to year, while other funds are geared towards generating a stable stream of disbursements. The point of view taken in this paper will be that of an endowment fund of the latter type. It will be assumed that an initial capital is made available at the time of origination of the fund and that the fund receives no donations later on. As an example of the type of endowment fund addressed in this paper, one might think of the Nobel Foundation.

Benefits are paid at discrete and equispaced time instants T_j ($j = 0, 1, 2, \ldots$) with $T_0 = 0$; the length of the interval between payment times is denoted by ΔT, so that $T_j = j\Delta T$. In examples, ΔT is taken to be one year. Benefits at different times are treated as being paid to different individuals. These individuals are referred to as 'recipients' or 'beneficiaries'; the terms 'agents' and 'generations' are also used. The benefit paid at time T_j is denoted by b_j.

If the benefit policy is fully specified in terms of the variables appearing within the BS model, then the payout b_j is a well-defined stochastic variable within the model with time-0 value given by $v_j = E[\xi_{T_j} b_j]$. The overall budget constraint is that the sum of the values v_j must be equal to the capital available at time 0:

$$\sum_{j=0}^{\infty} v_j = \sum_{j=0}^{\infty} E[\xi_{T_j} b_j] = V_0, \tag{11}$$

where V_0 is the initial capital. The constraint is stated as an equality, rather than an inequality, so that a requirement of efficiency is expressed as well (the initial capital should be fully used).

The focus of the present paper is on the intertemporal allocation problem of apportioning the initial capital V_0 to the individual budgets v_j.[7] For the application of the various design principles that will be studied below, it is necessary to specify the benefits b_j as random variables depending on the assigned budgets v_j. Such a specification can be given, for instance, by defining b_j as the amount at time T_j that results from following a prescribed investment policy that starts with capital v_j at time 0. Alternatively, one can specify a 'period utility function' $u_j(x)$ for time T_j and define b_j as the stochastic benefit that optimizes the criterion $E[u_j(b_j)]$ subject to the constraint $E[\xi_{T_j} b_j] = v_j$. Under the complete market assumption, the replication theorem of mathematical finance (see, for instance, Björk 1998, Thm. 7.3) ensures that the payoff b_j can indeed be realized and also provides the corresponding investment strategy. In line with tradition in academic work, the approach based on period utility functions will be followed in this paper. Given such a function $u_j(b_j)$ of standard type (strictly increasing, strictly concave, twice continuously differentiable), optimization of $E[u_j(b_j)]$ subject to a budget constraint leads (see, for instance, Föllmer & Schied 2008, § 3.3) to the solution

$$b_j^* = (u_j')^{-1}(y_j \xi_{T_j}), \quad y_j \in \mathbb{R}_+, \tag{12}$$

where ξ_T is the stochastic pricing kernel at time T, and where y_j is a parameter whose value is determined by the budget constraint for time T_j. When the period utility is of CRRA type (constant relative risk aversion), then, under the BS assumptions, the corresponding investment strategy is to keep a fixed percentage of capital in risky assets ('fixed-mix' strategy) (Merton 1969). Therefore, at least within the BS model, to say that the stochastic benefit b_j is defined by optimizing a CRRA period utility function is the same as to say that the benefit b_j is obtained from the initial capital v_j by applying a fixed-mix investment policy.

In the utilitarian approach, a common choice is to aggregate period utilities by means of weighted summation with exponential weights. This leads to an objective function of the form

$$\mathcal{J} = \sum_{j=0}^{\infty} e^{-\delta T_j} E[u_j(b_j)], \tag{13}$$

[7] In the context of variable annuities, the time-0 budgets v_j are called 'pension buckets' by Balter & Werker (2019).

where δ is a discount rate.[8] When beneficiaries at different times are different individuals, using an expression such as the one above entails an assumption that it is indeed possible to aggregate utilities of different individuals. The alternative viewpoint of 'neutrality' or 'fairness' is in line with often stated objectives of endowment funds to serve current and future generations 'in equal measure'. In a deterministic setting, a natural interpretation of fairness is that everyone should receive the same. Under the assumption that the continuously compounded real interest rate r is constant, this is realized by defining

$$b_j = (1 - \mathrm{e}^{-r\Delta T})V_0 \quad (j = 0, 1, 2, \ldots) \tag{14}$$

in combination with investing all of the capital in riskless assets.[9] In this scheme, the time-0 value of the time-T_j payout b_j is given by $v_j = \mathrm{e}^{-rT_j} b_j$, so that the continuously compounded AIR for time T_j according to definition (4) is simply given by $\rho_j = r$ for all j. However, many endowment funds choose to invest at least partly in risky assets and accept that, as a consequence, benefits become stochastic.[10] In this context, the meaning of 'fairness' is less obvious; even if one agrees that it means that some functional of the stochastic variable that represents the benefit should be the same for all beneficiaries, it still does not seem evident which functional should be chosen for this purpose. Perhaps the first candidate should be the time-0 expectation $E[b_j]$, as used, for instance, in Trautmann (2009) and in Balter & Werker (2019). When the form of the benefit is selected on the basis of an explicit period utility function $u_j(x)$, then another functional that comes to mind as a yardstick for fairness is the certainty equivalent $u_j^{-1}(E[u_j(b_j)])$, i.e. the hypothetical deterministic benefit that leads to the same utility level as the actual stochastic benefit, as seen from time 0.

In summary, each of the following three principles might be followed, given a period utility function:

(i) optimization of aggregated utility, with exponential weights;
(ii) equalization of expected benefits;
(iii) equalization of certainty equivalents of benefits.

Below, these three approaches will be compared to each other in terms of their associated AIRs. The three approaches will be referred to as 'utilitarian', 'expectation neutral', and 'CE neutral', respectively. As a notational device, the subscripts U, E, and C will be used for quantities related to utilitarian, expectation neutral, and CE neutral schemes, respectively. Since period utility functions are determined by risk aversion only up to positive affine transformations, the approach (i) presupposes that some kind of normalization has been applied so that a weighted combination of period utilities becomes meaningful. The problem is avoided in the two other approaches since neither the criterion in (ii) nor the one in (iii) is sensitive to positive affine transformations of the period utility functions.

To prepare for the application of the utilitarian scheme, it is noted here how the parameters y_j appearing in (12) are related to the discount rate δ. The optimal time-j utility $E[u(b_j^*)]$ can be thought of as a function $U_j(\theta_j)$ of the fraction of initial capital reserved for generation j. The problem of optimizing criterion (13) subject to budget constraint (11) can then be written as the optimization

[8] Like the real interest rate r, the discount rate δ is allowed to be zero or negative in this paper. See Herdegen et al. (2021) for arguments that support considering the possibility of negative discount rates.

[9] It appears that this is the scheme that Alfred Nobel had in mind when writing his will:

> All of my remaining realisable assets are to be disbursed as follows: the capital, converted to *safe securities* by my executors, is to constitute a fund, the *interest* on which is to be distributed annually as prizes to those who, during the preceding year, have conferred the greatest benefit to humankind. (English translation taken from www.nobelprize.org; emphasis added)

Nobel may have underestimated the effect of inflation.

[10] In 1953, in view of the substantially reduced value of the Nobel Prize in real terms, the directors of the Nobel Foundation sought and obtained permission from the Swedish government to invest part of the capital in equities and real estate. Constraints on investing have been relaxed further in later years (Rose & Nilsson 1999).

problem

$$\sum_{j=0}^{\infty} U_j(\theta_j) \to \max \quad \text{s.t.} \quad \theta_j \geq 0, \sum_{j=0}^{\infty} \theta_j = 1. \tag{15}$$

We have

$$U_j'(\theta_j) = E\left[u_j'(b_j^*)\,\frac{\partial b_j^*}{\partial \theta_j}(\theta_j)\right] = y_j\,\frac{\partial}{\partial \theta_j}E[\xi_{T_j} b_j^*(\theta_j)] = y_j V_0 \tag{16}$$

since $u_j'(b_j^*) = y_j \xi_{T_j}$ by (12). Given the interpretation of y_j as a Lagrange multiplier, the statement above just represents the well-known fact that, in an optimization problem subject to a single equality constraint, the derivative of the optimal value with respect to the constraint level is given by the multiplier. A necessary condition for optimality[11] is that $\mathrm{e}^{-\delta T_j} U_j'(\theta_j) = U_0'(\theta_0)$ for all $j \geq 1$, since otherwise an improvement is possible by increasing θ_0 at the expense of θ_j or vice versa. In view of (16), it follows that there exists a constant y such that

$$y_j = \mathrm{e}^{\delta T_j} y \quad \text{for all } j \geq 0. \tag{17}$$

The value of y is determined by the overall budget constraint. The corresponding benefits are given by

$$b_j^* = (u')^{-1}(y\,\mathrm{e}^{\delta T_j} \xi_{T_j}). \tag{18}$$

This solution could also have been obtained directly by optimizing the objective $\sum_{j=0}^{\infty} \mathrm{e}^{-\delta T_j} E[u(b_j)]$ subject to the single equality constraint $\sum_{j=0}^{\infty} E[\xi_{T_j} b_j^*] = V_0$; in fact, that is the more standard route, in particular when benefits are paid continuously rather than discretely. For the purposes of the present paper, however, it is useful to split the problem into a part that relates to the allocation of initial capital across generations, and another part that relates to optimization of the payoff to an individual generation, given its assigned budget. Similar approaches, with continuously paid benefits, are found, for instance, in Wachter (2002) and Steffensen (2011).

3. Guarantee properties

This brief section is devoted to certain asymptotic properties of the AIR that relate to guarantee properties of the associated benefit stream. The following terminology will be used.

Definition 3.1: Given an initial capital V_0 and a sequence of payment dates T_j, an *admissible benefit scheme* is any sequence $(b_j)_{j=0,1,2,\ldots}$ of T_j-measurable nonnegative random variables such that $\sum_{j=0}^{\infty} E[\xi_{T_j} b_j] \leq V_0$.

Definition 3.2: A benefit scheme with stochastic payouts $(b_0, b_1, b_2, \ldots)$ is said to satisfy

(i) the *strict guarantee property* if there exists $B > 0$ such that

$$P(b_j \geq B) = 1 \quad \text{for all } j \geq 0; \tag{19}$$

(ii) the *strong asymptotic guarantee property* if there exists $B > 0$ such that

$$\lim_{j \to \infty} P(b_j \geq B) = 1; \tag{20}$$

[11] It is assumed here that the utility functions u_j are such that, in the optimal solution, the fractions θ_j are all positive; in other words, each generation has a positive probability of receiving a positive benefit. Well-posedness of the optimization problem is assumed as well.

(iii) the *weak asymptotic guarantee property* if there exists $B > 0$ such that

$$\lim_{j\to\infty} P(b_j \geq B) > 0. \tag{21}$$

The weak asymptotic guarantee property is a weak property indeed; if it is not satisfied, then the sequence $(b_j)_{j=0,1,2,\ldots}$ converges to 0 in probability. Admissible schemes that satisfy the strong (weak) asymptotic guarantee property will for brevity be referred to as ASAG (AWAG) schemes; admissible schemes with the strict guarantee property are named ASG schemes.

The following proposition gives an upper bound on the asymptotic AIR for schemes that satisfy either of the three properties above. The statement under (i) in the proposition below is well known and is included for completeness and comparison. The proof of the proposition is provided in Appendix 1.

Proposition 3.3: *Let ξ_T denote the pricing kernel at time T in the BS model with interest rate r and price of risk λ. Let $0 = T_0 < T_1 < \cdots$ be a sequence of equispaced points on $[0, \infty)$, and let $V_0 > 0$ be given.*

(i) *Admissible schemes that satisfy the strict guarantee property exist if and only if $r > 0$. The asymptotic AIR of any such scheme satisfies*

$$\rho_\infty \leq r. \tag{22}$$

This bound is sharp; i.e. if $r > 0$, ASG schemes can be constructed whose asymptotic AIR is equal to r.

(ii) *Admissible schemes that satisfy the strong asymptotic guarantee property exist if and only if $r + \frac{1}{2}\lambda^2 > 0$. The asymptotic AIR of any such scheme satisfies*

$$\rho_\infty \leq r + \tfrac{1}{2}\lambda^2. \tag{23}$$

Moreover, this bound is sharp.

(iii) *Admissible schemes that satisfy the weak asymptotic guarantee property exist if and only if $r + \frac{1}{2}\lambda^2 \geq 0$. The asymptotic AIR of any such scheme satisfies* (23). *Moreover, this bound is sharp.*

It is remarkable that, in cases where $r + \frac{1}{2}\lambda^2 > 0$, the largest value attainable for the asymptotic AIR is the same for AWAG schemes and for ASAG schemes.

4. CRRA utility

This section provides a review of the AIR for schemes that are constructed on the basis of the standard CRRA utility function given by

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma} \quad (\gamma > 1). \tag{24}$$

Calculations for the log utility case and the case $0 < \gamma < 1$ in this and the following section are analogous to those for the case $\gamma > 1$, but partly different to the extent that they would have to be presented separately. In the interest of brevity, the treatment in this paper will be restricted to the case $\gamma > 1$, which empirically is the most relevant case; when conclusions deviate for $0 < \gamma < 1$ and/or in the logarithmic case, this will be mentioned in footnotes, without proof. The expressions for the AIR that are obtained in this section are well known or easily derived from standard results; they serve mainly as a benchmark for comparison with the AIR for schemes with saturated power utility, which will be discussed in the next section. The results below relating to situations with a possibly negative real interest rate appear to be new, however.

First, consider the scheme based on optimization of aggregate utility. This is the classical 'Merton problem' (Merton 1969). Maximization of period utility in the BS model under power utility is a standard problem; the optimal policy is to invest the budget in a fixed-mix portfolio with volatility λ/γ, leading to the payout

$$b_j^* = \theta_j V_0 \exp[(r + \lambda^2/\gamma - \tfrac{1}{2}\lambda^2/\gamma^2)T_j + (\lambda/\gamma)W_{T_j}] \tag{25}$$

with expected utility

$$E[u(b_j^*)] = \frac{1}{1-\gamma}\theta_j^{1-\gamma} V_0^{1-\gamma} \exp\left[(1-\gamma)\left(r + \frac{1}{2}\lambda^2/\gamma\right)T_j\right].$$

Using this, one can determine the optimal split of initial capital, when utility is aggregated by means of exponential weights. As seen from the above, this is a problem of the form

$$\sum_{j=0}^{\infty} \mathrm{e}^{-\alpha T_j}\theta_j^{1-\gamma} \to \min \quad \text{subject to } \theta_j > 0 \ (j = 0, 1, 2, \ldots),\ \textstyle\sum_{j=0}^{\infty}\theta_j = 1, \tag{26}$$

where the parameter α is given by

$$\alpha = \delta + (\gamma - 1)(r + \tfrac{1}{2}\lambda^2/\gamma). \tag{27}$$

If $\alpha > 0$, the optimal solution is given by

$$\theta_j \propto \exp(-(\alpha/\gamma)T_j) \tag{28}$$

where the proportionality constant is determined by the constraint $\sum_{j=0}^{\infty}\theta_j = 1$. It follows that the AIR for the scheme based on optimization of aggregate CRRA utility, with discount rate δ, is given by

$$\rho_{\mathrm{U}} = \alpha/\gamma = \frac{1}{\gamma}\delta + \left(1 - \frac{1}{\gamma}\right)\left(r + \frac{1}{2}\lambda^2/\gamma\right). \tag{29}$$

This is the same result as in the finite-horizon case (Balter & Werker 2019, eqn. (2.27)).

The assumption $\alpha > 0$ is not always satisfied. Exceptions occur in particular when the real interest rate is negative, and when moreover a small value is chosen for the discount rate δ and a high value for the risk aversion parameter γ.[12] The proposition below shows that these are exactly the situations in which the problem of optimizing aggregate utility is not well-posed, in the sense that there are no admissible schemes that lead to a criterion value larger than $-\infty$.

Proposition 4.1: *Let δ, r, and $\gamma > 1$ be given real numbers. Let $0 = T_0 < T_1 < T_2 < \cdots$ be a sequence of equispaced points on the real line, and let ξ_{T_j} denote the value of the pricing kernel in the Black–Scholes model at time T_j. There exists a sequence $(b_j)_{j=0,1,2,\ldots}$ of T_j-measurable positive random variables such that*

$$\sum_{j=0}^{\infty} E[\xi_{T_j} b_j] < \infty \quad \text{and} \quad \sum_{j=0}^{\infty} \mathrm{e}^{-\delta T_j} E\left[\frac{b_j^{1-\gamma}}{1-\gamma}\right] > -\infty \tag{30}$$

if and only if

$$\frac{1}{\gamma}\delta + \left(1 - \frac{1}{\gamma}\right)\left(r + \frac{1}{2}\lambda^2/\gamma\right) > 0. \tag{31}$$

[12] In the case $0 < \gamma < 1$, the expression for the optimal policy is of the same form, and the situation $\alpha \leq 0$ may arise even when the discount rate δ and the real interest rate r are both positive.

Proof: First, assume that a sequence of positive random variables b_j as described in the proposition exists.[13] From the equality

$$e^{-(\delta/\gamma)T_j}\xi_{T_j}^{(\gamma-1)/\gamma} = (e^{-\delta T_j}b_j^{1-\gamma})^{\frac{1}{\gamma}}(\xi_{T_j}b_j)^{1-\frac{1}{\gamma}},$$

it follows, by Hölder's inequality (see, for instance, Rudin 1974, Thm. 3.5), that

$$E[e^{-(\delta/\gamma)T_j}\xi_{T_j}^{(\gamma-1)/\gamma}] \le (E[e^{-\delta T_j}b_j^{1-\gamma}])^{\frac{1}{\gamma}}\,(E[\xi_{T_j}b_j])^{1-\frac{1}{\gamma}}. \tag{32}$$

By the assumption $\gamma > 1$, it follows from the second inequality in (30) that

$$\sum_{j=0}^{\infty} E[e^{-\delta T_j}b_j^{1-\gamma}] < \infty$$

which implies that $\lim_{j\to\infty} E[e^{-\delta T_j}b_j^{1-\gamma}] = 0$. By the first inequality in (30), we also have $\lim_{j\to\infty} E[\xi_{T_j}b_j] = 0$. It follows that the quantity at the left-hand side of (32) tends to 0 as j tends to infinity. Calculation shows that

$$E[e^{-(\delta/\gamma)T_j}\xi_{T_j}^{(\gamma-1)/\gamma}] = \exp\left[-\left(\frac{1}{\gamma}\delta + \left(1-\frac{1}{\gamma}\right)\left(r+\frac{1}{2}\lambda^2/\gamma\right)\right)T_j\right]$$

so that (31) follows. The converse is shown by noting that the optimal benefits (25) satisfy the conditions in (30). ■

Next, consider the neutralitarian scheme based on equalization of expected benefits. The benefits are still given by (25), but the fractions θ_j are determined in a different way, namely by requiring that $E[b_j^*]$ is the same for all j. From (25), it follows that

$$E[b_j^*] = \theta_j V_0 \exp[(r+\lambda^2/\gamma)T].$$

If the exponent $r+\lambda^2/\gamma$ is positive, then expectations are equalized by taking $\theta_j = \theta_j^{\mathrm{E}}$ with

$$\theta_j^{\mathrm{E}} \propto \exp[-(r+\lambda^2/\gamma)T] \tag{33}$$

where the proportionality constant is determined by the constraint $\sum_{j=0}^{\infty}\theta_j^{\mathrm{E}} = 1$. Consequently, the AIR that equalizes expectations is given by

$$\rho_{\mathrm{E}} = r+\lambda^2/\gamma. \tag{34}$$

Again, this is the same as in the finite-horizon case (Balter & Werker 2019, eqns. (2.6) and (2.14)). If the exponent $r+\lambda^2/\gamma$ is zero or negative, then the approach via equalization of expectations fails. Indeed, in this case, 'fairness' in the sense of equal expectations can only be achieved by letting all benefits be equal to 0; this is hardly a satisfactory solution, though.

[13] Here and occasionally also below, the notational device is used of referring to a sequence by means of its typical element.

The third principle that was proposed is to equalize certainty equivalents.[14] The certainty equivalent that corresponds to the benefit b_j^* is given by

$$\mathrm{CE}(b_j^*) = u^{-1}(E[u(b_j^*)]) = \theta_j V_0 \exp[(r + \tfrac{1}{2}\lambda^2/\gamma)T].$$

Therefore, the AIR that corresponds to the principle of equalizing certainty equivalents is

$$\rho_{\mathrm{C}} = r + \tfrac{1}{2}\lambda^2/\gamma. \tag{35}$$

This is notably less than the AIR obtained by equalizing expectations, although the difference does tend to zero when the risk aversion parameter γ tends to infinity. As in the case of equalizing expectations, the approach via equalization of certainty equivalents is feasible if and only if the associated AIR is positive.

Under CRRA utility within the BS environment, it is found that all three design principles outlined in Section 2.3 lead to a horizon-independent AIR. It follows that both neutralitarian schemes can be obtained as special cases of the utilitarian scheme if the discount rate within the utilitarian scheme is chosen such that the utilitarian AIR becomes equal to the AIR of either of the neutralitarian schemes. Comparing expressions (29), (34), and (35), one finds that the utilitarian scheme coincides with the expectation neutral scheme when $\delta = \delta_E$ and with the CE neutral scheme when $\delta = \delta_C$, where

$$\delta_E = r + \tfrac{1}{2}\lambda^2/\gamma + \tfrac{1}{2}\lambda^2, \quad \delta_C = r + \tfrac{1}{2}\lambda^2/\gamma. \tag{36}$$

At typical values of the Sharpe ratio for a broadly diversified portfolio (for instance, $\lambda = 0.3$), the difference between these two discount rates is substantial. Both δ_E and δ_C can be thought of as an answer to the question that is implicitly posed by Ramsey (1928), namely: What is a fair discount rate? The answers in (36) differ from the reply $\delta = 0$ that is suggested in Ramsey's paper, even though they are still based on principles of fairness. The deviation from the choice $\delta = 0$ is motivated by expected growth. The principle of expectation neutrality leads to a higher value for the discount rate because it does not take risk into account, in contrast to the principle of CE neutrality.

It follows from (25) and (28) that the utilitarian benefit scheme with discount rate δ satisfies the weak asymptotic guarantee property if and only if

$$\frac{\alpha}{\gamma} \le r + \lambda^2/\gamma - \tfrac{1}{2}\lambda^2/\gamma^2 \tag{37}$$

or equivalently

$$\delta \le r + \tfrac{1}{2}\lambda^2. \tag{38}$$

In other words, for values of the discount rate that exceed $r + \frac{1}{2}\lambda^2$, the optimal benefits tend to zero in probability as the time of payment tends to infinity. The conditions for the strong asymptotic guarantee property to hold are obtained by replacing the non-strict inequalities in the above by strict inequalities. Expression (29) for the utilitarian AIR ρ_{U} may be rewritten as

$$\rho_{\mathrm{U}} = r + \frac{1}{2}\lambda^2 - \frac{1}{2}\lambda^2\left(1 - \frac{1}{\gamma}\right)^2 - \frac{r + \frac{1}{2}\lambda^2 - \delta}{\gamma}.$$

This shows that, under condition (38), we have $\rho_{\mathrm{U}} \le r + \frac{1}{2}\lambda^2$, as should be the case according to Proposition 3.3. For high values of γ, the AIR can, at most, only slightly exceed the upper bound r that holds for schemes satisfying the strict guarantee property.

[14] In a case such as considered here, where the same utility function is used for all periods, equalization of certainty equivalents comes down to the same as equalization of expected utilities. Generally speaking, however, the use of certainty equivalents is essential to remove the effects of different scaling of utility functions at different time periods.

Conditions for the weak asymptotic guarantee property to hold in the case of expectation neutral and CE neutral schemes are analogous to (37), with substitution of the left-hand side by the corresponding AIRs. From this, it is readily verified that, for all values $\gamma > 1$ of the risk aversion parameter, CE neutral schemes are ASAG schemes.[15]

On the other hand, as can be seen by comparing (38) with (36), not even the weak asymptotic guarantee property is satisfied by any expectation neutral scheme under CRRA period utility. This fact may seem surprising, since one may feel that 'fairness' is not compatible with benefits for later generations tending to zero in probability. The reason why such a phenomenon may occur under equalization of expectations is that the expected value of the benefit at long horizons is influenced heavily by extremely high payouts that occur in very few scenarios. This is a consequence of using CRRA period utility (i.e. fixed-mix investment strategies) in combination with the lognormal BS model. Many endowment funds may prefer to have stable payouts at a reasonable level in as many scenarios as possible, rather than having a small probability of being able to pay very large sums at some point in the future. The notion of a 'reasonable level' of payouts is incorporated in the model that is proposed in the next section. It will be seen that, under saturated utility, equalization of expectations leads to schemes that do satisfy the weak asymptotic guarantee property.

5. The impact of saturation

The period utility functions that have been used so far are of the strictly increasing type that is most commonly found in the literature. However, such utility functions may not accurately reflect the preferences of trustees of endowment funds. The purposes of many funds can be well achieved by an annual amount that is approximately fixed in real terms; say, paying for a professor's salary, covering the costs of an annual music festival, or providing prize money for an annual award.[16] To reflect the notion that spending beyond a given amount is not really necessary, it is natural to introduce a saturation level.[17] This will be done in this paper by a straightforward amendment of the classical CRRA utility function. Specifically, the following modification of the utility function $u(x) = x^{1-\gamma}/(1-\gamma)$ is proposed:

$$\bar{u}(x) = \min(u(x), u(C)) = \frac{(\min(x, C))^{1-\gamma}}{1-\gamma}. \tag{39}$$

Here, the constant $C > 0$ represents the saturation level; in many cases, it can be seen as a target level of spending. The utility function above might be described as 'capped power utility'.[18]

Under the conditions $r > 0$ and

$$C/V_0 \leq 1 - \mathrm{e}^{-r\Delta T} \tag{40}$$

where ΔT is the length of the interval between payment times, one can make sure that the target is always achieved by keeping all of the capital in riskless assets. For brevity, the ratio C/V_0 of saturation level to initial capital will be referred to as the *target ratio* below. Its reciprocal V_0/C is the number of years of benefits at the saturation level that could be paid from the initial capital if there would be no investment returns.

Many endowment funds set their targets in such a way that the inequality (40) is not met, counting on returns from risky investments to make it possible to pay the full target amount in many scenarios.

[15] The assumption $\gamma > 1$ is crucial here. In the case of logarithmic period utility, CE neutral schemes are AWAG but not ASAG. When $0 < \gamma < 1$, CE neutral schemes do not have the weak asymptotic guarantee property.

[16] A similar argument may be made for funds that promise a certain benefit to recipients, such as for instance collective pension funds. The fund will have fully achieved its goal when the promise can be kept.

[17] One may argue that additional money is always welcome. However, one can also argue (as illustrated above) that the utility of payoffs beyond a certain level is overstated under CRRA preferences. The model with saturation can be viewed as the simplest way of introducing a 'satisfaction level'.

[18] An alternative form that might be used is $\tilde{u}(x) = \min(u(x) - u(C), 0)$. The difference between the two formulations has no effect on single-period optimization, but does affect the finiteness of the aggregate objective (13) in cases in which the discount rate is negative or zero.

Clearly, for higher values of the target ratio C/V_0, there will be fewer scenarios in which the goal is achieved. The relation between the target ratio and the probability of achieving the target depends on the choice of the coefficient or risk aversion γ. High values of γ lead to a fairly slow reduction of disbursements in situations in which the full target is not paid, while low values lead to a sharper dropoff. This does mean that the probability of achieving the target goes down when γ is raised while the target ratio C/V_0 remains the same. Since the allocation problem under capped power utility has a trivial solution when (40) holds, it will be a standing assumption in this paper that the condition (40) is *not* satisfied.

Under the specification (39) of the utility function, marginal utility equals $x^{-\gamma}$ for $c < C$, and jumps from $C^{-\gamma}$ to 0 at the saturation point C. Expression (12) for efficient benefits can in such cases be interpreted in a generalized sense (Basak 1995, Berkelaar et al. 2004, Carassus & Pham 2009, Bernard et al. 2015, Bian & Zheng 2015). The inverse marginal utility $(\bar{u}')^{-1}(z)$ is viewed as the multivalued function that is defined as follows:

$$(\bar{u}')^{-1}(z) = \begin{cases} [C, \infty) & \text{for } z = 0 \\ C & \text{for } 0 < z \le C^{-\gamma} \\ z^{-1/\gamma} & \text{for } z \ge C^{-\gamma}. \end{cases} \tag{41}$$

This will be used freely below.

5.1. Efficient payoffs

It follows from (12) and (41) that, under saturated CRRA utility, the one-parameter family of efficient payoffs at time T_j is of the form

$$\bar{b}_j^* = \min((y_j \xi_{T_j})^{-1/\gamma}, 1)C \tag{42}$$

where the parameter y_j is adjusted to meet the budget constraint for time T_j. In particular, the full target benefit will be paid when the pricing kernel ξ_{T_j} at time T_j satisfies $\xi_{T_j} < y_j^{-1}$. It will be convenient below to use, for $j \ge 1$, the following reparametrization in terms of parameters d_j rather than y_j:

$$\bar{b}_j^* = \exp\left(\min\left(\frac{\lambda}{\gamma}\sqrt{T_j}(Z_j + d_j), 0\right)\right) C, \quad Z_j := \frac{1}{\sqrt{T_j}} W_{T_j}. \tag{43}$$

The relation between d_j and y_j is given by

$$d_j = \frac{r + \frac{1}{2}\lambda^2}{\lambda}\sqrt{T_j} - \frac{\log y_j}{\lambda\sqrt{T_j}}. \tag{44}$$

The random variable Z_j follows a standard normal distribution. As a consequence, the probability that the target level C will be achieved at time T_j is given by

$$P(\bar{b}_j^* = C) = P(Z_j + d_j \ge 0) = \Phi(d_j).$$

The parameter d_j, therefore, has a direct interpretation as the standard normal quantile that corresponds to the probability of paying the target benefit at time T_j for $j \ge 1$.

The probability that $\bar{b}_j^*$ as given by (43) achieves at least the value ηC, with $0 < \eta \le 1$, is given by

$$P(\bar{b}_j^* \ge \eta C) = \Phi\left(d_j - \frac{\gamma \log \eta}{\lambda\sqrt{T_j}}\right).$$

This shows that a scheme with benefits given by (43) has the strong asymptotic guarantee property if the sequence d_j tends to infinity as j tends to infinity, and that the weak asymptotic guarantee property is satisfied if the sequence d_j is bounded below. Therefore, in both cases, if there is a (weak) asymptotic guarantee level, then the saturation level serves as such.

To determine the policies implied by the three different design principles that have been discussed above, expressions are needed for the expected value of $\bar{b}_j^*$ defined in (43), the utility level associated to $\bar{b}_j^*$, and the time-0 value of $\bar{b}_j^*$. Due to the relatively simple specification of $\bar{b}_j^*$, all of these can be given in analytic form. By direct computation, it can be shown that

$$E[\exp(a(Z+x)\mathbb{1}_{Z+x\leq 0})] = F(x,a) \quad (Z \sim N(0,1)), \tag{45}$$

where the function $F(x,a)$ is defined by

$$F(x,a) = \mathrm{e}^{ax+\frac{1}{2}a^2}\Phi(-(x+a)) + \Phi(x). \tag{46}$$

Using this, one finds for $j \geq 1$:

$$E[\bar{b}_j^*] = F\left(d_j, \frac{\lambda}{\gamma}\sqrt{T_j}\right)C, \tag{47}$$

$$E[\bar{u}(\bar{b}_j^*)] = F\left(d_j, \frac{1-\gamma}{\gamma}\lambda\sqrt{T_j}\right)\frac{C^{1-\gamma}}{1-\gamma}, \tag{48}$$

$$E[\xi_{T_j}\bar{b}_j^*] = \mathrm{e}^{-rT_j}F\left(d_j - \lambda\sqrt{T_j},\ \frac{\lambda}{\gamma}\sqrt{T_j}\right)C. \tag{49}$$

Expressions (47) and (48) follow directly from relation (45). Claim (49) can be proved by using the relation

$$E[\mathrm{e}^{-\frac{1}{2}\alpha^2-\alpha Z}g(Z)] = E[g(Z-\alpha)]$$

which holds for $Z \sim N(0,1)$, $\alpha \in \mathbb{R}$, and any function g such that the expectation is defined. A plot of the function F is shown in Figure A1 in the appendix.

5.2. Utilitarian scheme

The allocation under the utilitarian principle is obtained from (17) and (42); one finds

$$\bar{b}_j^{\mathrm{U}} = \min((y\,\mathrm{e}^{\delta T_j}\xi_{T_j})^{-1/\gamma}, 1)C, \tag{50}$$

where the parameter y is adapted to meet the budget constraint associated with the initial capital V_0. For purposes of comparison with other schemes, it is useful to have the representation in terms of the parameters d_j in (43) as well. It follows from (44) that the parameters d_j are of the form $d_j = d_{\mathrm{U}}(T_j)$ with

$$d_{\mathrm{U}}(T) = \frac{r+\frac{1}{2}\lambda^2-\delta}{\lambda}\sqrt{T} + \frac{\kappa_{\mathrm{U}}}{\sqrt{T}}, \tag{51}$$

where $\kappa_{\mathrm{U}} \in \mathbb{R}$ is a constant. The expression above determines the benefits at times T_j for $j \geq 1$ through (43). It follows from (42), (44), and (51) that the time-0 benefit can be expressed in terms of κ_{U} by

$$\bar{b}_0^{\mathrm{U}} = \min(\exp((\lambda/\gamma)\kappa_{\mathrm{U}}), 1)C. \tag{52}$$

The constant κ_{U} is chosen to meet the budget constraint; in this way, it is monotonically related to the target ratio C/V_0. Higher values of κ_{U} correspond to lower values of the target ratio. Irrespective of the choice of the target ratio, however, it is seen from (51) that the *asymptotic* probability of reaching the target is 1, $\frac{1}{2}$, or 0 according to whether the discount rate δ is less than, equal to, or higher than

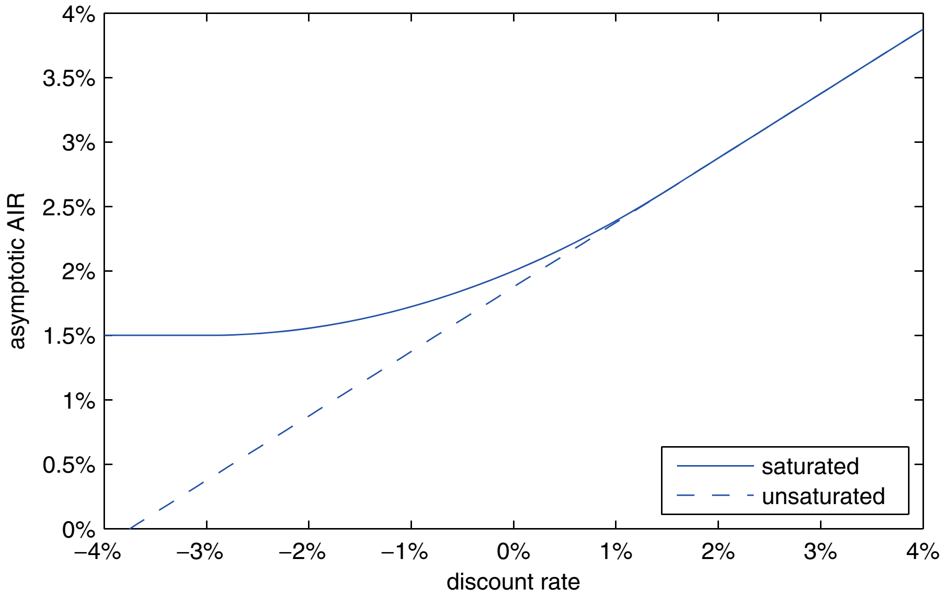

Figure 1. The asymptotic AIR under the utilitarian principle, as a function of the discount rate. The period utility function is CRRA with or without saturation (drawn/dashed curve). Parameter values: interest rate $r = 1.5\%$, market price of risk $\lambda = 0.3$, coefficient of risk aversion $\gamma = 2$.

$r + \frac{1}{2}\lambda^2$. The utilitarian scheme under saturated utility is, therefore, an ASAG scheme if and only if $\delta < r + \frac{1}{2}\lambda^2$, and an AWAG scheme if and only if $\delta \leq r + \frac{1}{2}\lambda^2$. These are the same conditions as the ones that were found in the classical (unsaturated) case. While the probability of reaching the saturation level in the utilitarian scheme does not depend on the target ratio in the limit as the horizon length tends to infinity, the ratio does impact that probability at finite horizons.

The time-0 values of the payoffs under optimization of aggregate saturated utility are given by (49) with $d_j = d_{\mathrm{U}}(T_j)$. Let the asymptotic AIR be denoted by $\bar{\rho}_\infty^{\mathrm{U}}$. It can be shown (see Appendix 3) that we have

$$\bar{\rho}_\infty^{\mathrm{U}} = \begin{cases} r & \text{if } \delta \leq r - \frac{1}{2}\lambda^2 \\ r + \dfrac{1}{2}\left(r - \delta - \dfrac{1}{2}\lambda^2\right)^2 / \lambda^2 & \text{if } r - \frac{1}{2}\lambda^2 \leq \delta \leq r - \frac{1}{2}\lambda^2 + \lambda^2/\gamma \\ \dfrac{1}{\gamma}\delta + \left(1 - \dfrac{1}{\gamma}\right)\left(r + \dfrac{1}{2}\lambda^2/\gamma\right) & \text{if } \delta \geq r - \frac{1}{2}\lambda^2 + \lambda^2/\gamma \end{cases} \tag{53}$$

provided the quantity defined in this way is positive.[19] The asymptotic AIR $\bar{\rho}_\infty^{\mathrm{U}}$ is defined by the above as a continuously differentiable nondecreasing function of the discount rate δ; a numerical example is shown in Figure 1. Due to the saturation, the asymptotic AIR is always at least equal to the riskless interest rate, even if future generations are strongly favored by the choice of a negative discount rate. On the other hand, if the discount rate is sufficiently high, then saturation has no effect on the asymptotic AIR. The upper bound for the asymptotic AIR that follows from (53) is the same as in the unsaturated case, since the critical value $r + \frac{1}{2}\lambda^2$ for δ always lies in the region where the expression for $\bar{\rho}_\infty^{\mathrm{U}}$ as given in (53) is the same as the expression for ρ_{U} as given in (29).

[19] If positivity does not hold, then the utilitarian policy defined by (51) is not implementable under the condition of the finiteness of the initial capital; compare the discussion of the utilitarian scheme in Section 4. It can be shown that, if positivity does hold, then the policy defined by (43) and (51) defines a finite objective value either under (39) or under the alternative formulation mentioned in note 18.

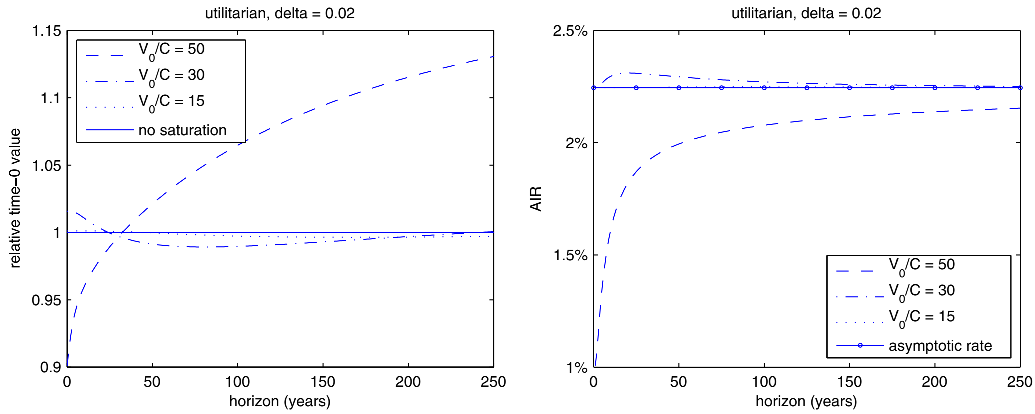

Figure 2. The left panel shows the time-0 values of benefits at different horizons for the utilitarian scheme with discount rate $\delta = 0.02$ under saturated utility, relative to the same values when there is no saturation. Results are shown under different assumptions on the ratio V_0/C of initial capital to target spending. The right panel shows the corresponding implied AIRs. The following parameter values are assumed: market price of risk $\lambda = 0.3$, real interest rate $r = 1\%$, coefficient of relative risk aversion $\gamma = 10/3$. The value of γ corresponds to a 60/40 asset mix when the risky asset in the BS model has 15% volatility. Under these parameter values, the asymptotic AIR in the scheme under saturation is equal to the AIR of the scheme without saturation.

A graph of the time-0 values of the utilitarian scheme under saturation, relative to the time-0 values in the scheme without saturation, is shown for specific parameter values in the left panel of Figure 2. The right panel shows the corresponding horizon-dependent AIRs. It is seen from the figure that, under the assumed parameter values, the effect of introducing a saturation level (i.e. placing a cap on spending) for a utilitarian social planner is strong in particular in cases in which the initial capital is high relative to the cap. In these cases, in which the fund is initially 'rich' relative to the cap, the planner would spend more in early years if the cap would not be present. Therefore, the introduction of the cap is disadvantageous for early generations (up to about 30 years in the case $V_0/C = 50$) and advantageous for later generations. In cases in which the fund is initially 'poor', the presence of the cap has little effect, since in most scenarios the benefits will be below saturation level anyway. There are intermediate cases in which the introduction of the saturation level is slightly advantageous for early generations and disadvantageous for later generations. In these cases, the initial capital is not high enough to pay benefits at the saturation level, but the fund has a good chance of reaching that position at a later stage. In favorable scenarios, the level of benefits will be lower when the cap is present than when the cap is not present. Consequently, the time-0 values corresponding to benefits for later generations are reduced. As is seen from the graph, the effect does not reach beyond the magnitude of one or two percentage points in the example calculation.

5.3. Expectation neutral scheme

From expression (47) for $E[b_j^*]$, it follows that in the expectation neutral scheme the fractions of capital allocated to successive generations are chosen in such a way that $d_j = d_{\mathrm{E}}(T_j)$, where $d_{\mathrm{E}}(T)$ satisfies the equation

$$F(d_{\mathrm{E}}(T), (\lambda/\gamma)\sqrt{T}) = \kappa_{\mathrm{E}} \tag{54}$$

in which $\kappa_{\mathrm{E}} \in (0, 1)$ is a constant. The time-0 benefit under expectation neutrality is given by $\bar{b}_0^{\mathrm{E}} = \kappa_{\mathrm{E}} C$. Analogously to the case of the utilitarian scheme, the parameter κ_{E} is monotonically related to the target ratio.

In contrast to the utilitarian scheme, the full target benefit is paid at time 0 under the expectation neutral scheme only if the target ratio satisfies condition (40), under which it becomes possible (within the BS model) to pay the target benefit to all generations with full certainty. Indeed, it is not 'fair' (in the sense of expectations, and in other senses as well) to pay the target benefit in full to the current generation, when later generations do run a risk of experiencing a shortfall.

In the case of a classical CRRA period utility function, without saturation, it was found that expectation neutrality always gives rise to schemes that do not satisfy the weak asymptotic guarantee property. The situation is different when saturation is added. It is shown in Appendix 4 that the function $d_{\mathrm{E}}(T)$ which is defined in (54) satisfies $\lim_{T\to\infty} d_{\mathrm{E}}(T) = \Phi^{-1}(\kappa_{\mathrm{E}})$. In other words, κ_{E} is the asymptotic probability of achieving the target as the horizon T_j tends to infinity. Therefore, under the saturated period utility function, expectation neutral schemes are AWAG schemes. On the other hand, they do not satisfy the strong asymptotic guarantee property.

In view of Proposition 3.3, the fact that expectation neutral schemes under saturated utility satisfy the weak asymptotic guarantee property implies that the asymptotic AIR for these schemes must be reduced with respect to the unsaturated case. Indeed, it is proved in Appendix 4 that the fractions $\bar{\theta}_j^{\mathrm{E}}$ of initial capital allocated to different generations under the expectation neutral scheme are such that

$$\bar{\theta}_j^{\mathrm{E}} \propto \chi_{\mathrm{E}}(T_j)\exp(-\bar{\rho}_\infty^{\mathrm{E}} T_j), \quad \bar{\rho}_\infty^{\mathrm{E}} = r + \lambda^2/\gamma - \tfrac{1}{2}\lambda^2/\gamma^2, \tag{55}$$

where $\chi_{\mathrm{E}}(T)$ is a function that satisfies

$$\chi_{\mathrm{E}}(0) = \kappa_{\mathrm{E}}, \quad \lim_{T\to\infty} \chi_{\mathrm{E}}(T) = \infty, \quad \lim_{T\to\infty} \mathrm{e}^{-\varepsilon T}\chi_{\mathrm{E}}(T) = 0 \quad \text{for all } \varepsilon > 0. \tag{56}$$

Therefore, while the function $\chi_{\mathrm{E}}(T)$ tends to ∞ as T tends to ∞, it does so at a subexponential rate; the asymptotic AIR is hence given by $\bar{\rho}_\infty^{\mathrm{E}}$ as defined in (55). It is seen that the introduction of a saturation level results in a reduction of the asymptotic AIR from $\rho_{\mathrm{E}} = r + \lambda^2/\gamma$ to $\bar{\rho}_\infty^{\mathrm{E}} = r + \lambda^2/\gamma - \frac{1}{2}\lambda^2/\gamma^2$. This indicates a shift of value from earlier to later generations. In fact, it can be shown (see also Appendix 4) that the ratio $\bar{\theta}_j^{\mathrm{E}}/\theta_j^{\mathrm{E}}$ of fractions of capital assigned to generation j with or without saturation is increasing in j. Although the later generations are, therefore, better off under expectation neutrality when saturation is introduced, it can be noted that the asymptotic AIR of the expectation neutral scheme with saturation is still larger than the AIR of the CE neutral scheme without saturation, as is seen from comparing (55) to (35).[20]

The left panel of Figure 3 shows a plot of time-0 values in the scheme with saturation, relative to time-0 values in the scheme without saturation. The corresponding AIRs are shown in the right panel. The figure can be compared to Figure 2 which shows the same quantities in the case of a utilitarian planner. The curves shown in the left panel are all increasing, contrary to the case of the utilitarian planner. Similarly to that case, the introduction of a saturation level has a stronger effect when the ratio of initial capital to the spending cap is higher ('rich' funds). For low values of this ratio ('poor' funds), later generations are still exponentially better off under the scheme with saturation than in the scheme without saturation, but the effect becomes only strong at horizons in the order of several human lifetimes. The crossover point between generations that profit from the introduction of a saturation level and those who are worse off does not depend much on the ratio of initial capital to saturation level, and lies at approximately the 30-year horizon. For rich funds, the asymptotic value of the implied AIR in the scheme with saturation is somewhat misleading, since the AIR still lies considerably below the asymptotic value even at quite long horizons.

5.4. CE neutral scheme

It is seen from (48) that, to equalize certainty equivalents, the parameters d_j ($j = 1, 2, \ldots$) must be chosen as $d_j = d_{\mathrm{C}}(T_j)$, where the function $d_{\mathrm{C}}(T)$ is defined by

$$F\left(d_{\mathrm{C}}(T), \frac{1-\gamma}{\gamma}\lambda\sqrt{T}\right) = \kappa_{\mathrm{C}} \tag{57}$$

[20] In case $0 < \gamma \le 1$, it follows in the same way as above that expectation neutral schemes under saturation have the weak asymptotic guarantee property. Moreover, it can be shown that their asymptotic AIR, in this case, equals $r + \frac{1}{2}\lambda^2$.

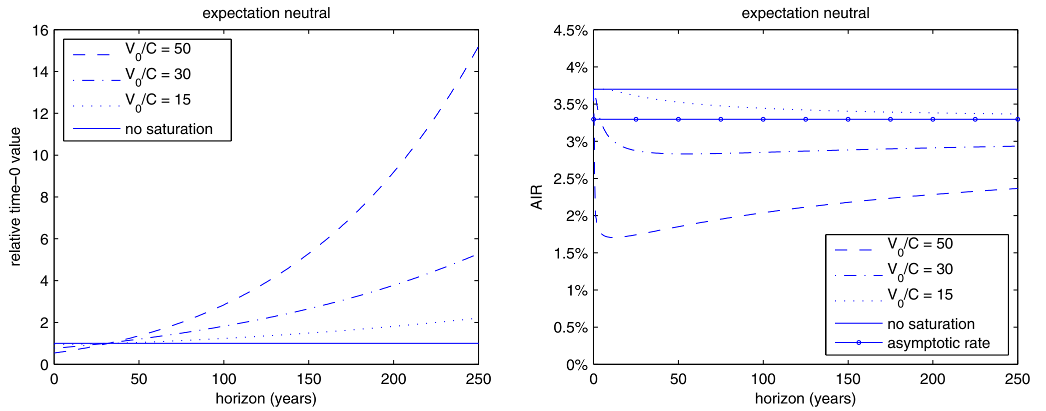

Figure 3. The left panel shows the time-0 values of benefits at different horizons for the expectation neutral scheme under saturated utility, relative to the same values when there is no saturation. Results are shown under different assumptions on the ratio V_0/C of initial capital to target spending. The right panel shows the corresponding implied AIRs. Parameter values are as in Figure 2.

and $\kappa_C > 1$ is a fixed constant. The payoff at time 0 that leads to the same level of utility is given by

$$\bar{b}_0^E = \kappa_C^{1/(1-\gamma)} C. \tag{58}$$

The constant κ_C is chosen such that the budget constraint is met. As in the case of the expectation neutral scheme, and contrary to the case of the utilitarian scheme, the initial payment always falls short of the target benefit C unless condition (40) holds.

The fractions of initial capital that are allocated to generations under the CE neutral scheme are determined by (see Appendix 5)

$$\bar{\theta}_j^C \propto \chi_C(T) \exp(-\bar{\rho}_\infty^C T_j), \quad \bar{\rho}_\infty^C = r + \tfrac{1}{2}\lambda^2/\gamma, \tag{59}$$

where $\chi_C(T)$ is an increasing function with limit values given by

$$\lim_{T\downarrow 0} \chi_C(T) = \kappa_C^{1/(1-\gamma)}, \quad \lim_{T\to\infty} \chi_C(T) = (\kappa_C - 1)^{1/(1-\gamma)}. \tag{60}$$

The asymptotic AIR for the CE neutral scheme is, therefore, equal to $\bar{\rho}_\infty^C = r + \frac{1}{2}\lambda^2/\gamma$, which is the same as in the unsaturated case.[21]

It follows from results in the Appendix (Section 2) that the function $d_C(T)$ defined in (57) satisfies $\lim_{T\to\infty} d_C(T) = \infty$. Consequently, the CE neutral schemes under saturated utility are always ASAG schemes, just as in the unsaturated case (with $\gamma > 1$).[22]

The plots in Figure 4 provide a comparison between the schemes with and without saturation that would be chosen by a CE neutral social planner, analogously to Figures 2 and 3 for the cases of a utilitarian planner and an expectation neutral planner, respectively. In the CE neutral case, there is no difference in the asymptotic AIR between schemes with and without saturation. Consequently, differences between these schemes are not as dramatic as in the case of an expectation neutral planner. Nevertheless, the fact that the function $\chi_C(T)$ appearing in (59) is increasing means that there is still a shift in time-0 value from earlier generations to later generations. This shift may be explained as follows. The quantity $\xi_T^{-1/\gamma}$ that appears in (42) has a larger variance for longer horizons. The upward

[21] In case $0 < \gamma \leq 1$, it can be shown that the asymptotic AIR of CE neutral schemes under saturation is equal to $r + \frac{1}{2}\lambda^2$. For $0 < \gamma < 1$, this is different from the AIR in the case without saturation.

[22] This conclusion remains valid in the logarithmic case. In case $0 < \gamma < 1$, CE neutral schemes under saturation are AWAG but not ASAG.

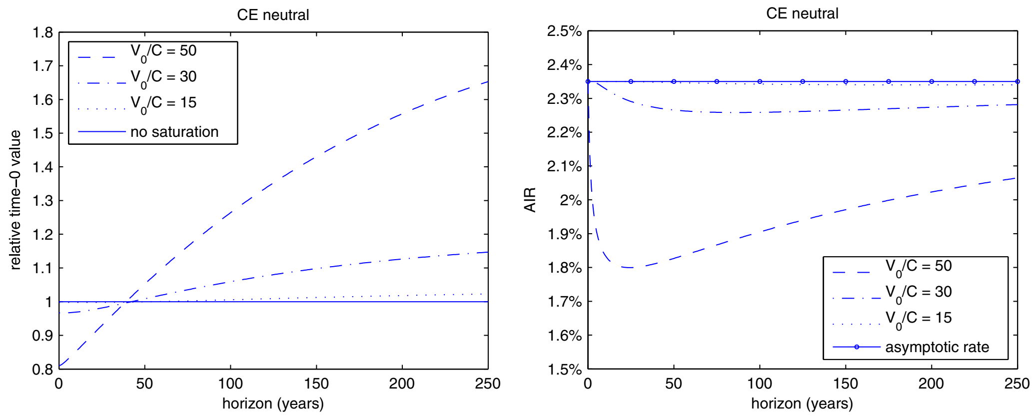

Figure 4. The left panel shows the time-0 values of benefits at different horizons for the CE neutral scheme under saturated CRRA utility, relative to the same values under standard CRRA utility. Results are shown under different assumptions on the ratio V_0/C of initial capital to target spending. The right panel shows the corresponding implied AIRs. Parameter values are as in Figure 2. The asymptotic rate is the same as the AIR in the CE neutral scheme without saturation.

Table 1. The probability of achieving the target for two different neutralitarian schemes at different horizons.

	Expectation neutral			CE neutral		
	50	30	15	50	30	15
20 years	88.3%	43.9%	3.7%	82.1%	32.4%	1.6%
50 years	91.5%	55.6%	11.4%	90.3%	59.0%	18.1%
100 years	93.2%	62.7%	18.7%	94.8%	77.1%	45.6%
∞	98.0%	84.9%	52.1%	100%	100%	100%

Notes: The numbers 50/30/15 indicate values of the ratio V_0/C of initial capital to target annual spending. Parameter values are as in Figure 2.

potential created by this variance is eliminated by the truncation of payoffs at the saturation level C, but the downward risk still remains. In particular, later generations can be hit hard in scenarios in which there is a long string of low returns to capital. To compensate for this, these generations receive a larger portion of initial capital, which allows a more favorable probability distribution of their benefits.

A comparison between the expectation neutral and the CE neutral schemes under saturation can also be made in terms of the probabilities of reaching the target level of spending at different horizons. Results for a particular case are shown in Table 1; the assumed parameter values are the same as the ones in Figures 2–4. To place the numbers shown in the table in perspective, it may be noted that, at the assumed interest rate of 1%, to achieve the target level with certainty at all horizons would require the ratio of capital to target annual spending to be equal to 100.5. It is seen that, at half of the capital required for certainty, one still obtains fairly high probabilities of achieving the target. Naturally, these probabilities decline if the amount of initial capital is still lower. As the horizon lengthens, the probability of reaching the target tends to a limit less than 1 in the expectation neutral schemes, but tends to 1 in the CE neutral schemes, even for funds that are initially poor. As shown in the table, this means that the probability of achieving the target is lower under CE neutrality than under expectation neutrality for early generations, but is higher for later generations. The point of overtaking comes earlier when the fund is initially poorer.

6. Conclusions and outlook

The endowment fund model that has been discussed in this paper combines elements of social choice theory with notions from mathematical finance. It belongs to an area that might be called 'social finance', in analogy with 'personal finance', and that is concerned with the design of financial policy on behalf of multiple (typically heterogeneous) agents. Collective investment is an example. The endowment fund model used in this paper is intended as a relatively simple model in which the aspect of intertemporal choice can be addressed.

In the classical BS/CRRA context, there is no structural difference between policies derived from either of the three principles that were discussed in this paper. The two neutralitarian principles can in fact both be viewed as providing a way of deriving the discount rate in the utilitarian principle from the economic parameters r and λ and the parameter γ which describes the attitude towards risk. The rate $r + \lambda^2/\gamma$, which is recommended under expectation neutrality, is the optimal expected return[23] of a CRRA investor with coefficient γ of relative risk aversion. In that sense, it can be said that the principle of expectation neutrality leads to the recommendation to 'discount by the expected return'. Along the same lines, the principle of CE neutrality can be said to recommend discounting by the average of the expected return and the riskless rate.

When the CRRA utility function is modified by introducing a saturation level, the policies derived from the three principles discussed in this paper become essentially different. It seems likely that this will also happen when the CRRA utility is modified in a different way, for instance, by introducing a guarantee level. The introduction of a saturation level is usually beneficial for later generations, with an exception in the utilitarian scheme in the case of funds with a fairly low ratio of initial capital to the saturation level of spending. The quantitative differences between expectation neutral schemes and CE neutral schemes are substantial under the standard CRRA period utility function. Under saturated CRRA, these differences are mitigated, without becoming negligible.

The utilitarian schemes discussed in this paper are all time-consistent: re-evaluation at any time in the future does not lead to a change in policy. On the other hand, the neutralitarian strategies must in general be viewed as precommitment strategies. The reason is that fairness is a time-indexed notion; what is fair today may no longer be fair tomorrow, when new information has come in. This fact need not be a source of concern if one takes a 'contract' point of view. One can imagine that, at time 0, a contract is proposed to all generations, which specifies for each of them a promise for a benefit to be paid at the due date T_j. The promise is expressed as a formula that takes relevant economic variables that will be known at time T_j as inputs and that produces the actual amount to be paid as an output. Under the assumptions of the complete market model, the generations are certain that the promise will be kept under all circumstances. When the proposed contract is considered fair at time 0 by all generations, it is agreed between them, and benefit payments will take place as stated in the contract, for better or for worse. The idea that re-evaluation will never be necessary rests strongly on model assumptions, however, and seems hard to maintain in practice. The question then arises whether it is possible to develop policies that are both time-consistent and fair.[24] The design of time-consistent policies with respect to various non-classical optimization criteria has been extensively researched in recent years; see, for instance, Ekeland & Pirvu (2008), Vieille & Weibull (2009), Björk & Murgoci (2014), Basak & Chabakauri (2010), Ekeland & Lazrak (2010), Asheim & Ekeland (2016), Björk et al. (2017), Vigna (2020), Kryger et al. (2020), Balter et al. (2021), and Desmettre & Steffensen (2021). This literature may become applicable to fairness criteria as discussed in this paper by viewing these criteria in a maxmin framework, i.e. as criteria for optimality with respect to the minimum of expected benefits or certainty equivalents across recipients.

The capped utility functions that have been used in this paper were chosen as an example of non-CRRA utilities because they still allow an almost completely analytical treatment; moreover, the saturation effect may be fairly realistic for some funds. It would be of interest to investigate other non-CRRA utility functions as well. For instance, one might include a minimal subsistence level that must

[23] The term 'expected return' is used here to refer to $(1/T)\log(E[V_T]/V_0)$, where V_t indicates portfolio value at time t, rather than to $(1/T)E[\log(V_T/V_0)]$.

[24] A suitable form of efficiency is to be added as a third requirement. For instance, the policy of paying nothing to all generations is both time-consistent and fair (indeed, all generations are treated the same), but inefficient in the sense that not all of the available budget is used.

be achieved with certainty or with very high probability, as for instance in Benartzi & Thaler (1995), Berkelaar et al. (2004) and Chen et al. (2021). It would be of interest to see how this impacts the distribution of time-0 values across generations. Another question concerns the corresponding investment policies. Under capped utility, investment in risky assets is reduced in favorable scenarios, when the available capital approaches the amount that is sufficient to keep the benefit stream at the target level forever. When a subsistence level is added, it can be expected that derisking will also take place in strongly adverse scenarios.

It is often seen in practice that trustees of endowment funds aim to suppress the year-to-year variation of benefits, by applying some form of smoothing. For instance, a typical benefit policy is to disburse a fixed percentage of the *average* value of assets during the last three years, rather than a fixed percentage of the *current* value of assets. For a similar effect, Balter & Werker (2019) propose to reduce the exposure to risky assets of the portfolio earmarked for the benefit at time T_j according to a fixed schedule in the period leading up to the payment time. Alternatively, one may attempt to capture preference for smoothness in an objective function that can subsequently be optimized. For instance, to express aversion against reduction of benefits with respect to the previous period, one could let the period utility function for time T_j ($j \geq 1$) be a mixture of the saturated CRRA utility function (39) with a similar utility function that has payment b_{j-1} as a saturation level, instead of the constant C:

$$\bar{u}_j(x) = \beta \frac{(\min(x, C))^{1-\gamma}}{1-\gamma} + (1-\beta)\,\frac{(\min(x, b_{j-1}))^{1-\gamma}}{1-\gamma}, \tag{61}$$

where $\beta \in (0,1)$ is a chosen weight. The efficient payoff at time T_j is still given by (12), where now b_{j-1} appears as a parameter in the utility function. The payoff has two flat regions, one corresponding to the saturation level C, and another (for a range of lower values of the reference portfolio) corresponding to the time-T_{j-1} payoff b_{j-1}. The location of the latter region becomes only determined at time T_{j-1}. By decreasing the weight β, one can diminish the probability of scenarios in which a reduction of benefits takes place, at the cost of lower expected growth. The form of the utility function in (61) is quite simple; other ways of expressing preference for smoothness might be investigated as well.

The endowment fund model as used in this paper assumes that the fund aims to pay inflation-indexed benefits at regular intervals, such as would be the case for instance for funds supporting an endowed professorship, or for an institution such as the Nobel Foundation. However, there are also endowment funds that pay benefits at irregular intervals, in response to unpredictable opportunities or emergencies that may arise. For instance, a fund supporting an art museum may want to spend a substantial amount at times when an interesting piece appears on the market. In such cases, one still needs to find a balance between spending now and saving for later; the models supporting such decisions would need to be different from the one used here.

The Black–Scholes market used in this paper is of course a simple model that incorporates only the most basic features of financial markets. In particular, interest rates are assumed constant. As a result, the analysis of the present paper cannot shed light on the question of how endowment funds should respond to changes in interest rates. To extend the scope of the analysis, one might, for instance, combine the BS model for equity with an affine term structure model. It is to be expected that it would then still be possible to carry out a large part of the analysis on the basis of explicit formulas.

Acknowledgments

For helpful suggestions and comments on earlier versions of this paper, I would like to thank Jennifer Alonso García, Servaas van Bilsen, An Chen, Frank de Jong, Sander Muns, Theo Nijman, Manuel Rach, Peter Schotman, Bas Werker, and an anonymous reviewer. The usual disclaimer applies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Johannes M. Schumacher http://orcid.org/0000-0001-5753-3412

References

Asheim, G. B. & Ekeland, I. (2016). Resource conservation across generations in a Ramsey–Chichilnisky model. Economic Theory **61**, 611–639.
Bacinello, A. R., Millossovich, P., Olivieri, A. & Pitacco, E. (2011). Variable annuities: a unifying valuation approach. Insurance: Mathematics and Economics **49**, 285–297.
Bajeux-Besnainou, I. & Portait, R. (1997). The numeraire portfolio: a new perspective on financial theory. European Journal of Finance **3**, 291–309.
Balter, A. G., Mahayni, A. & Schweizer, N. (2021). Time-consistency of optimal investment under smooth ambiguity. European Journal of Operational Research **293**, 643–657.
Balter, A. G. & Schweizer, N. (2021). Robust decisions for heterogeneous agents via certainty equivalents. Available at: ssrn.com/abstract = 3873267
Balter, A. G. & Werker, B. J. M. (2019). The effect of the assumed interest rate and smoothing on variable annuities. ASTIN Bulletin **50**, 131–154.
Bao, H., Ponds, E. H. M. & Schumacher, J. M. (2017). Multi-period risk sharing under financial fairness. Insurance: Mathematics and Economics **72**, 49–66.
Basak, S. (1995). A general equilibrium model of portfolio insurance. Review of Financial Studies **8**, 1059–1090.
Basak, S. & Chabakauri, G. (2010). Dynamic mean-variance asset allocation. Review of Financial Studies **23**, 2970–3016.
Benartzi, S. & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. Quarterly Journal of Economics **110**, 73–92.
Berkelaar, A. B., Kouwenberg, R. & Post, T. (2004). Optimal portfolio choice under loss aversion. Review of Economics and Statistics **86**, 973–987.
Bernard, C., Chen, J. S. & Vanduffel, S. (2015). Rationalizing investors' choices. Journal of Mathematical Economics **59**, 10–23.
Bian, B. & Zheng, H. (2015). Turnpike property and convergence rate for an investment model with general utility functions. Journal of Economic Dynamics and Control **51**, 28–49.
Björk, T. (1998). Arbitrage theory in continuous time. Oxford: Oxford University Press.
Björk, T., Khapko, M. & Murgoci, A. (2017). On time-inconsistent stochastic control in continuous time. Finance and Stochastics **21**, 331–360.
Björk, T. & Murgoci, A. (2014). A theory of Markovian time-inconsistent stochastic control in discrete time. Finance and Stochastics **18**, 545–592.
Blake, D., Cairns, A. J. G. & Dowd, K. (2003). Pensionmetrics 2: stochastic pension plan design during the distribution phase. Insurance: Mathematics and Economics **33**, 29–47.
Brown, D. & Scholz, D. (2019). Discount rates and asset returns: implications for endowment strategies. Available at: ssrn.com/abstract = 3426197
Carassus, L. & Pham, H. (2009). Portfolio optimization for piecewise concave criteria functions. In S. Ogawa, editor, *8th Workshop on Stochastic Numerics*, RIMS Kôkyuroku series, Vol. 1620. Kyoto: Kyoto University. p. 81–111.
Cejnek, G., Franz, R., Randl, O. & Stoughton, N. (2014). A survey of university endowment management research. Journal of Investment Management **12**(3), 90–117.
Chen, A., Nguyen, T. & Rach, M. (2021). Optimal collective investment: the impact of sharing rules, management fees and guarantees. Journal of Banking & Finance **123**, 106012.
Dellinger, J. K. (2006). The handbook of variable income annuities. Hoboken, NJ: Wiley.
Delong, Ł. (2014). Pricing and hedging of variable annuities with state-dependent fees. Insurance: Mathematics and Economics **58**, 24–33.
Desmettre, S. & Steffensen, M. (2021). Optimal investment with uncertain risk aversion. Available at: https://ssrn.com/abstract = 3805069
Dufresne, D. (1990). The distribution of a perpetuity, with applications to risk theory and pension funding. Scandinavian Actuarial Journal **1990**, 39–79.
Ekeland, I. & Lazrak, A. (2010). The golden rule when preferences are time inconsistent. Mathematics and Financial Economics **4**, 29–55.
Ekeland, I. & Pirvu, T. A. (2008). Investment and consumption without commitment. Mathematics and Financial Economics **2**, 57–86.
Feller, W. (1968). An introduction to probability theory and its applications, Vol. **I**. New York: Wiley.
Föllmer, H. & Schied, A. (2008). Stochastic finance: an introduction in discrete time. Berlin: Walter de Gruyter.
Gilbert, T. & Hrdlicka, C. M. (2011). Are universities fair? Risking the endowment for future generations. Available at: ssrn.com/abstract = 1919764

Herdegen, M., Hobson, D. & Jerome, J. (2021). An elementary approach to the Merton problem. Mathematical Finance **31**, 1218–1239.
Horneff, V., Maurer, R., Mitchell, O. S. & Rogalla, R. (2015). Optimal life cycle portfolio choice with variable annuities offering liquidity and investment downside protection. Insurance: Mathematics and Economics **63**, 91–107.
Jensen, B. A. & Nielsen, J. A. (2016). How suboptimal are linear sharing rules? Annals of Finance **12**, 221–243.
Kleinow, T. & Schumacher, J. M. (2017). Financial fairness and conditional indexation. Scandinavian Actuarial Journal **2017**, 651–669.
Koijen, R. S. J., Nijman, T. E. & Werker, B. J. M. (2011). Optimal annuity risk management. Review of Finance **15**, 799–833.
Kryger, E. M., Nordfang, M.-B. & Steffensen, M. (2020). Optimal control of an objective functional with non-linearity between the conditional expectations: solutions to a class of time-inconsistent portfolio problems. Mathematical Methods of Operations Research **91**, 405–438.
Kryger, E. M. & Steffensen, M. (2010). Some solvable portfolio problems with quadratic and collective objectives. Available at: ssrn.com/abstract = 1577265
Lindset, S. & Matsen, E. (2018). Institutional spending policies: implications for future asset values and spending. Financial Markets and Portfolio Management **32**(1), 53–76.
Long, Jr, J. B. (1990). The numeraire portfolio. Journal of Financial Economics **26**, 29–69.
Mahayni, A. & Schneider, J. C. (2012). Variable annuities and the option to seek risk: why should you diversify? Journal of Banking & Finance **36**, 2417–2428.
Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. Review of Economics and Statistics **51**, 247–257.
Milevsky, M. A. & Robinson, C. (2000). Self-annuitization and ruin in retirement. North American Actuarial Journal **4**(4), 112–124.
Pazdera, J., Schumacher, J. M. & Werker, B. J. M. (2016). Cooperative investment in incomplete markets under financial fairness. Insurance: Mathematics and Economics **71**, 394–406.
Platen, E. & Heath, D. (2006). A benchmark approach to quantitative finance. Berlin: Springer.
Pye, G. B. (2017). Disbursements. Journal of Portfolio Management **43**(3), 136–151.
Ramsey, F. P. (1928). A mathematical theory of saving. Economic Journal **38**, 543–559.
Rawls, J. (1971). A theory of justice. Cambridge, MA: Belknap Press.
Roemer, J. E. (1998). Theories of distributive justice. Cambridge, MA: Harvard University Press.
Rose, J. & Nilsson, A. (1999). Nobel foundation seeks looser financial reins. Science **285**(5426), 323–325.
Rudin, W. (1974). Real and complex analysis, 2nd ed. New Delhi: Tata McGraw-Hill.
Steffensen, M. (2011). Optimal consumption and investment under time-varying relative risk aversion. Journal of Economic Dynamics and Control **35**, 659–667.
Tobin, J. (1974). What is permanent endowment income? American Economic Review **64**, 427–432.
Trautmann, S. T. (2009). A tractable model of process fairness under risk. Journal of Economic Psychology **30**, 803–813.
Trottier, D.-A., Godin, F. & Hamel, E. (2018). Local hedging of variable annuities in the presence of basis risk. ASTIN Bulletin **48**, 611–646.
Vieille, N. & Weibull, J. W. (2009). Multiple solutions under quasi-exponential discounting. Economic Theory **39**, 513–526.
Vigna, E. (2020). On time consistency for mean-variance portfolio selection. International Journal of Theoretical and Applied Finance **23**, 2050042.
Wachter, J. A. (2002). Portfolio and consumption decisions under mean-reverting returns: an exact solution for complete markets. Journal of Financial and Quantitative Analysis **37**, 63–91.

Appendices

Appendix 1. Proof of Proposition 3.3

Consider first the claim in (i). Since $E[\xi_T] = \mathrm{e}^{-rT}$, the relation $b_j \geq B$ implies that the time-0 value v_j of the benefit paid at time T_j satisfies $v_j \geq \mathrm{e}^{-rT_j} B$. Therefore, we can only have $\sum_{j=0}^{\infty} v_j < \infty$ if $r > 0$. Moreover, it follows from

$$\log \frac{v_j}{v_0} \geq -rT_j + \log \frac{B}{v_0}$$

that $\rho_\infty \leq r$. Finally, an example of an ASG scheme that achieves the upper bound is the one defined in (14).

Since every ASAG scheme is also an AWAG scheme, the statements under (ii) and (iii) can be rearranged into the following claims:

(a) if $r + \frac{1}{2}\lambda^2 \leq 0$, then no ASAG scheme exists;
(b) for any AWAG scheme, we have $\rho_\infty \leq r + \frac{1}{2}\lambda^2$;

(c) if $r + \frac{1}{2}\lambda^2 > 0$, then there exists an ASAG scheme with asymptotic AIR equal to $r + \frac{1}{2}\lambda^2$;
(d) if $r + \frac{1}{2}\lambda^2 = 0$, then there exists an AWAG scheme with asymptotic AIR equal to 0.

To prove these claims, first, an inequality is shown that generalizes the inequality $v_j \geq \mathrm{e}^{-rT_j}B$ on which the proof for the strict guarantee case is based. Let a scheme $(b_j)_{j=0,1,2,\ldots}$ and a level $B > 0$ be given. For $j = 0, 1, 2, \ldots$, define $p_j = P(b_j \geq B)$, and let $d_j = \Phi^{-1}(p_j)$ denote the standard normal quantile that corresponds to the probability p_j. Since the pricing kernel ξ_T in the BS model is a deterministic and monotonically decreasing function of the driving Brownian motion W_T, the smallest time-0 value of a benefit at time T that is at least equal to B with probability p is given by $E[\xi_T \mathbb{1}_{W_T \geq \eta}]B$, where the constant η is chosen such that $P(W_T \geq \eta) = p$. A standard computation shows that

$$E[\xi_T \mathbb{1}_{W_T \geq \eta}] = \mathrm{e}^{-rT}\Phi(d - \lambda\sqrt{T}), \tag{A1}$$

where $d = \Phi^{-1}(p)$. It follows that the time-0 values v_j of the benefits b_j satisfy

$$v_j \geq B\,\mathrm{e}^{-rT_j}\Phi(d_j - \lambda\sqrt{T_j}) \quad (j = 0, 1, 2, \ldots). \tag{A2}$$

If the sequence $d_j - \lambda\sqrt{T_j}$ contains a subsequence that is bounded below, then it follows that $r > 0$ just as in the case of the strict guarantee, and the conclusion under part (a) is seen to hold. To prove part (a), it is therefore sufficient to consider ASAG schemes that satisfy

$$\lim_{j\to\infty} d_j - \lambda\sqrt{T_j} = -\infty. \tag{A3}$$

Due to the fact that

$$\lim_{x\to\infty} x\Phi(-x)/\varphi(x) = 1 \tag{A4}$$

(see, for instance, Feller 1968, Lemma VII.1.2), one can write, for $0 < a < 1$ and for d and T such that $\lambda\sqrt{T} - d$ is sufficiently large,

$$\Phi(d - \lambda\sqrt{T}) \geq \frac{a}{\sqrt{2\pi}}\,\frac{\exp(-\frac{1}{2}d^2 + d\lambda\sqrt{T} - \frac{1}{2}\lambda^2 T)}{\lambda\sqrt{T} - d}. \tag{A5}$$

From the asymptotic guarantee property, it follows that the sequence d_j tends to ∞. Inequalities (A2) and (A3) imply the existence of a positive constant a_1 such that

$$v_j \geq a_1 \exp\left(-\left(r + \frac{1}{2}\lambda^2\right)T_j\right)\frac{\exp(-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j})}{\lambda\sqrt{T_j} - d_j}$$

for all sufficiently large values of j. We can write

$$-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j} = \frac{1}{2}d_j(\lambda\sqrt{T_j} - d_j) + \frac{1}{2}d_j\lambda\sqrt{T_j}. \tag{A6}$$

Since the sequences d_j and $\lambda\sqrt{T_j} - d_j$ both tend to ∞, it follows that there exists a positive constant a_2 such that

$$v_j \geq a_2 \exp\left(-\left(r + \frac{1}{2}\lambda^2\right)T_j\right)\frac{\exp\left(\frac{1}{2}d_j\lambda\sqrt{T_j}\right)}{\lambda\sqrt{T_j} - d_j}$$

for all sufficiently large j. Let a_3 be an arbitrary positive constant. Since $d_j \geq a_3$ for all sufficiently large j, we can write

$$v_j \geq a_2 \exp\left(-\left(r + \frac{1}{2}\lambda^2\right)T_j\right)\frac{\exp\left(\frac{1}{2}a_3\lambda\sqrt{T_j}\right)}{\lambda\sqrt{T_j}} \tag{A7}$$

for all sufficiently large j. In case $r + \frac{1}{2}\lambda^2 \leq 0$, the right-hand side of (A7) tends to infinity, and we have a contradiction. This proves part (a).

For part (b), suppose that $(b_j)_{j=0,1,2,\ldots}$ is an AWAG scheme with weak asymptotic guarantee level B. From inequality (A2), it follows that the asymptotic AIR ρ_∞ of the scheme satisfies

$$\rho_\infty \leq r - \liminf_{j\to\infty}\frac{1}{T_j}\log\Phi(d_j - \lambda\sqrt{T_j}). \tag{A8}$$

From the weak asymptotic guarantee property, it follows that the sequence d_j is bounded below. Therefore, the claim in (b) will be verified if it can be shown that

$$\liminf_{j\to\infty}\frac{1}{T_j}\log\Phi(d_j - \lambda\sqrt{T_j}) \geq -\frac{1}{2}\lambda^2 \tag{A9}$$

for any such sequence d_j. For a proof by contradiction, assume that there is a sequence d_j which is bounded below and which is such that

$$\lim_{j\to\infty} \frac{1}{T_j} \log \Phi(d_j - \lambda\sqrt{T_j}) < -\frac{1}{2}\lambda^2.$$

Without loss of generality, it can be assumed that $d_j - \lambda\sqrt{T_j}$ tends to $-\infty$. For any subsequence of d_j that remains bounded, we have $\lim_{j\to\infty}(-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j})/T_j = 0$. From (A6), it is seen that, for subsequences of d_j that tend to infinity, the expression $-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j}$ is positive for all sufficiently large values of j. It follows that

$$\liminf_{j\to\infty} \frac{1}{T_j}\left(-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j}\right) \geq 0.$$

We now have a contradiction as a result of (A5). This completes the proof of part (b).

In part (c), it is given that $r + \frac{1}{2}\lambda^2 > 0$. A benefit scheme with the required properties can be constructed in the 'digital' form

$$b_j = \mathbb{1}_{W_{T_j} \geq \eta_j}.$$

The sequence η_j is chosen as follows. First, take a sequence d_j such that $d_j \to \infty$ and $d_j/\sqrt{T_j} \to 0$ as j tends to infinity; for instance, one can take $d_j = T_j^\alpha$ with $0 < \alpha < \frac{1}{2}$. Now, define η_j by $\eta_j = -d_j\sqrt{T_j}$. We have $P(W_{T_j} \geq \eta_j) = \Phi(d_j) \to 1$ as j tends to infinity. The time-0 values of the payoffs b_j are given by (A1). From the property $d_j/\sqrt{T_j} \to 0$, it follows that $\lim_{j\to\infty} \lambda\sqrt{T_j} - d_j = \infty$. Making use of (A4), one finds that the time-0 values of the digital payoffs satisfy

$$v_j \leq \frac{a}{\sqrt{2\pi}} \exp\left(-\left(r + \frac{1}{2}\lambda^2\right)T_j\right) \frac{\exp\left(-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j}\right)}{\lambda\sqrt{T_j} - d_j} \tag{A10}$$

for all sufficiently large j, where a is a constant larger than 1. From the property $\lim_{j\to\infty} d_j\sqrt{T_j} = 0$, it also follows that $\lim_{j\to\infty}(-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j})/T_j = 0$. Consequently, we have

$$-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j} \leq \tfrac{1}{2}\left(r + \frac{1}{2}\lambda^2\right)T_j$$

for all sufficiently large j. Using this, it is seen from (A10) that the sum of the time-0 values of the digital payoffs b_j is finite. To conclude, again using (A4), one can write the asymptotic AIR of the scheme as

$$\rho_\infty = r + \frac{1}{2}\lambda^2 - \lim_{j\to\infty} \frac{1}{T_j}\left(-\frac{1}{2}d_j^2 + d_j\lambda\sqrt{T_j}\right) = r + \frac{1}{2}\lambda^2. \tag{A11}$$

Finally, in part (d), it is assumed that $r + \frac{1}{2}\lambda^2 = 0$. Consider again a digital benefit scheme, constructed, as in part (c), from a sequence d_j. This time choose the sequence d_j such that it converges to a finite negative limit; for instance, take $d_j = -1$ for all j. The weak asymptotic guarantee property is then satisfied since the sequence d_j is bounded below. The property $\lim_{j\to\infty} \lambda\sqrt{T_j} - d_j = \infty$ obviously holds, so that inequality (A10) applies. Finiteness of the sum of time-0 values then follows from the fact that the series $\sum_{j=0}^\infty e^{-\alpha\sqrt{n}}/\sqrt{n}$ is convergent for positive values of α, since the corresponding integral is convergent. Relation (A11) can be used and shows in this case that $\rho_\infty = 0$. This completes the proof.

The scheme that is used in the proof of part (c) allows benefits to take the value 0 with nonzero probability. An alternative scheme with benefits that are positive with probability 1 is given by

$$b_j = \min(G_{T_j}/K_j, 1), \tag{A12}$$

where $G_t := \xi_t^{-1}$ is the time-t value of the so-called growth optimal portfolio,[25] and the constants K_j are chosen such that $P(G_{T_j} \geq K_j) = \Phi(d_j)$ where d_j is defined implicitly by

$$\varphi(d_j) - d_j\Phi(-d_j) = \frac{\kappa}{\lambda\sqrt{T_j}}, \quad \kappa > 0. \tag{A13}$$

By somewhat laborious computations that are not shown here, it can be demonstrated that (i) the sum of the time-0 values of this scheme is finite and can be adjusted to match any given positive value of initial capital by choice of the parameter κ, (ii) the probability of the event $b_j = 1$ tends to 1 as j tends to infinity, and (iii) the asymptotic AIR

[25] The growth optimal portfolio (Long 1990, Bajeux-Besnainou & Portait 1997, Platen & Heath 2006) can be constructed as a fixed-mix portfolio where the weights are chosen such that the portfolio volatility is equal to the Sharpe ratio. This may require a short position in bonds.

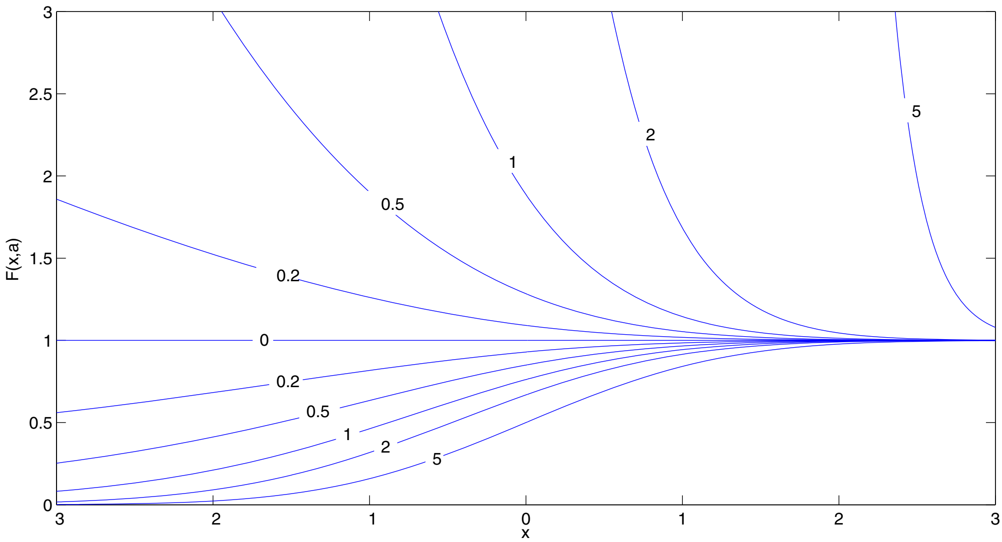

Figure A1. $F(x, a)$ as a function of x, for several values of a as indicated in the plot.

of the scheme equals $r + \frac{1}{2}\lambda^2$. The scheme defined by (A12) and (A13) is not as contrived as it may seem; it is in fact the scheme that would be chosen on the basis of equalization of certainty equivalents when the utility function is log utility with saturation at level 1.

Appendix 2. Auxiliaries

The function $F(x, a)$ was introduced in (46); it plays a major role in describing key quantities related to the efficient benefits under capped CRRA utility, as seen in (47)–(49). A plot F as a function of x, for several fixed values of the parameter a, is shown in Figure A1. It is sometimes convenient to express the function $F(x, a)$ in terms of the Mills ratio of the standard normal distribution, i.e. the function $m(x)$ given by

$$m(x) := \Phi(-x)/\varphi(x).$$

We have

$$F(x, a) = \varphi(x)m(x + a) + \Phi(x) = \varphi(x)(m(x + a) + m(-x)). \tag{A14}$$

For future use, it is noted that (see, for instance, Feller 1968, Lemma VII.1.2)

$$\frac{1}{x} - \frac{1}{x^3} < m(x) < \frac{1}{x} \tag{A15}$$

for all $x > 0$. Consequently,

$$0 < 1 - xm(x) < \frac{1}{x^2} \quad (x > 0). \tag{A16}$$

The inequality $1 - xm(x) > 0$ in fact holds for all $x \in \mathbb{R}$. The derivative of $m(x)$ is given by $m'(x) = xm(x) - 1$. It follows that the Mills ratio $m(x)$ is a monotonically decreasing function.

The partial derivatives of $F(x, a)$ with respect to x and a are given, respectively, by

$$\frac{\partial F}{\partial x}(x, a) = a\varphi(x)m(x + a), \tag{A17}$$

$$\frac{\partial F}{\partial a}(x, a) = \varphi(x)((x + a)m(x + a) - 1). \tag{A18}$$

It is seen from these formulas (as can also be concluded from stochastic representation (45)) that $F(x, a)$ as a function of x is increasing if $a > 0$ and decreasing if $a < 0$, and that $F(x, a)$ is decreasing in a for every fixed value of x. Moreover, we have the limit relationships

$$\lim_{x\to-\infty} F(x,a) = \begin{cases} 0 & (a>0), \\ \infty & (a<0), \end{cases} \qquad \lim_{x\to\infty} F(x,a) = 1 \quad (a\in\mathbb{R}). \tag{A19}$$

From these properties, it can be concluded that the equation

$$F(x(a),a) = \kappa \tag{A20}$$

defines, for $0<\kappa<1$, a function $x(a)$ of a for $a>0$, and in case $\kappa>1$, defines a function $x(a)$ of a for $a<0$. Moreover, the functions that are defined in this way for $a>0$ are increasing, while the functions defined for $a<0$ are decreasing. The asymptotic behavior of both types of functions is of interest in the analysis in the main body of the paper. The inverse function of $x(a)$ will be written as $a(x)$ and is defined by $F(x,a(x))=\kappa$.

Properties of the function $x(a)$ defined by (A20) for $a>0$ and $\kappa\in(0,1)$ can be obtained as follows. Introduce a new variable y by $y=x+a$, and write the equation $F(x,a)=\kappa$ in the form

$$\frac{\Phi(-y)}{\varphi(y)} = \frac{\kappa-\Phi(x)}{\varphi(x)}. \tag{A21}$$

From this, it is seen that the inequality $\Phi(x)<\kappa$ must hold. Also, it can be concluded from (A15) that $\lim_{x\uparrow\Phi^{-1}(\kappa)} y(x)=\infty$. This implies that $\lim_{x\uparrow\Phi^{-1}(\kappa)} a(x)=\infty$ as well, so that we have

$$\lim_{a\to\infty} x(a) = \Phi^{-1}(\kappa). \tag{A22}$$

Concerning the behavior of $x(a)$ for small values of the argument, from the fact that $x(a)$ is increasing in a and the fact that $F(x,0)=1$ for all x it follows that $x(a)$ tends to $-\infty$ when a tends to zero. For a more precise description, rewrite (A20) in the form

$$\frac{1}{2}a^2 + ax(a) = \log\frac{\kappa-\Phi(x(a))}{\Phi(-(x+a))}. \tag{A23}$$

Taking the limits of both sides as a tends to zero, one finds

$$\lim_{a\downarrow 0} ax(a) = \log\kappa. \tag{A24}$$

This is sufficient information for the applications below.

Next, consider the function $x(a)$ defined by (A20) for $a<0$, with $\kappa>1$. Since $F(x,a)$ tends to infinity when a tends to $-\infty$ while x remains bounded, the function $x(a)$ must satisfy $\lim_{a\to-\infty} x(a)=\infty$. Consider again (A21). Since the right-hand side of the equation tends to ∞ as x tends to ∞, it follows that $a(x)+x\to-\infty$ as $x\to\infty$. This implies that $a+x(a)\to-\infty$ as $a\to-\infty$. Since the right-hand side in (A23) (where now $\kappa>1$ is assumed) remains bounded as $a\to-\infty$ (in fact tends to $\log(\kappa-1)$), it follows that $x(a)$ behaves asymptotically for $a\to-\infty$ as $-\frac{1}{2}a$, and consequently, $x(a)+a$ behaves asymptotically as $\frac{1}{2}a$. This in turn implies that the right-hand side of (A23) tends to $\log(\kappa-1)$ at a faster than polynomial rate in a as a tends to $-\infty$. Therefore, we can write

$$x(a) = -\frac{1}{2}a + \frac{\log(\kappa-1)}{a} + o(a^{-N}) \quad (a\to-\infty) \tag{A25}$$

for any natural number N. Considering now the behavior of $x(a)$ as a tends to 0 from below, the fact that $\lim_{a\uparrow 0} F(x,a)=1$ for all fixed x implies that we must have $x(a)\to-\infty$ as $a\uparrow 0$. This of course also implies that $x(a)+a\to-\infty$ as $a\uparrow 0$. Equation (A23) then shows that $\lim_{a\uparrow 0} ax(a)=\log\kappa$, which implies that the right-hand side of (A23) converges to $\log\kappa$ at a faster than polynomial rate in $1/a$ as a tends to 0 from below. Consequently, we can write

$$x(a) = \frac{\log\kappa}{a} - \frac{1}{2}a + o(a^N) \quad (a\uparrow 0) \tag{A26}$$

for any natural number N.

Appendix 3. Claims in Section 5.2

The expressions in (53) are obtained as follows. Define a function $f(T)$ by

$$f(T) = F\left(\zeta\sqrt{T} + \frac{\kappa}{\sqrt{T}}, \sigma\sqrt{T}\right),$$

where ζ and σ are defined by

$$\zeta = \frac{r-\delta-\frac{1}{2}\lambda^2}{\lambda}, \qquad \sigma = \frac{\lambda}{\gamma}$$

and where κ is the constant appearing in (51). We have $\lim_{T\downarrow 0} f(T)=1$. If $\zeta>0$, then $\lim_{T\to\infty} f(T)=1$, so that the asymptotic AIR as given by (49) is equal to the riskless interest rate r. The same rate holds for $\zeta=0$, since in that case

$\lim_{T\to\infty} f(T) = \frac{1}{2}$. Assume now that $\zeta < 0$. By l'Hôpital's rule, the asymptotic rate of decrease of the function $f(T)$ can be found from

$$\lim_{T\to\infty} -\frac{1}{T}\log f(T) = -\lim_{T\to\infty} \frac{f'(T)}{f(T)}$$

if the limit on the right-hand side exists. Computation on the basis of (A17) and (A18) shows that

$$\frac{f'(T)}{f(T)} = \frac{(\zeta\sigma + \frac{1}{2}\sigma^2)m((\zeta+\sigma)\sqrt{T} + \kappa/\sqrt{T}) - \frac{1}{2}\sigma/\sqrt{T}}{m((\zeta+\sigma)\sqrt{T} + \kappa/\sqrt{T}) + m(-\zeta\sqrt{T} - \kappa/\sqrt{T})}. \tag{A27}$$

First consider the case in which $\zeta + \sigma > 0$. Making use of the property $\lim_{x\to\infty} xm(x) = 1$ of the Mills ratio, one finds that, in this case,

$$\lim_{T\to\infty} \sqrt{T}\, m((\zeta+\sigma)\sqrt{T} + \kappa/\sqrt{T}) = \frac{1}{\zeta+\sigma}, \quad \lim_{T\to\infty} \sqrt{T} m(-\zeta\sqrt{T} - \kappa/\sqrt{T}) = -\frac{1}{\zeta}.$$

Multiplying both the numerator and the denominator of the right-hand side of (A27) by $\sqrt{T}$, one finds that

$$\lim_{T\to\infty} \frac{f'(T)}{f(T)} = \frac{(\zeta\sigma + \frac{1}{2}\sigma^2)/(\zeta+\sigma) - \frac{1}{2}\sigma}{1/(\zeta+\sigma) - 1/\zeta} = -\frac{1}{2}\zeta^2.$$

Lastly, assume that $\zeta + \sigma \leq 0$. In this case, the term $m((\zeta+\sigma)\sqrt{T} + \kappa/\sqrt{T})$ dominates, and it follows that $\lim_{T\to\infty} f'(T)/f(T) = \zeta\sigma + \frac{1}{2}\sigma^2$.

Appendix 4. Claims in Section 5.3

It follows from (47) and (49) that the fractions $\bar{\theta}_j^{\mathrm{E}}$ $(j \geq 1)$ of initial capital reserved for successive generations under the expectation neutral scheme are such that $\bar{\theta}_j^{\mathrm{E}} \propto \bar{\theta}_{\mathrm{E}}(T_j)$, where

$$\bar{\theta}_{\mathrm{E}}(T) = \mathrm{e}^{-rT} F\left(d_{\mathrm{E}}(T) - \lambda\sqrt{T}, \frac{\lambda}{\gamma}\sqrt{T}\right) \quad \text{with} \quad F\left(d_{\mathrm{E}}(T), \frac{\lambda}{\gamma}\sqrt{T}\right) = \kappa_{\mathrm{E}}.$$

As shown in Appendix 2, the relation $F(d_{\mathrm{E}}(T), (\lambda/\gamma)\sqrt{T}) = \kappa_{\mathrm{E}}$ indeed defines $d_{\mathrm{E}}(T)$ uniquely as a function of $T > 0$ for any given value of $\kappa_{\mathrm{E}} \in (0, 1)$. Define the function $\chi_{\mathrm{E}}(T)$ by

$$\chi_{\mathrm{E}}(T) = \exp(\bar{\rho}_{\infty}^{\mathrm{E}} T)\bar{\theta}_{\mathrm{E}}(T) = \exp\left(\left(\lambda^2/\gamma - \frac{1}{2}\lambda^2/\gamma^2\right) T\right) F\left(d_{\mathrm{E}}(T) - \lambda\sqrt{T}, \frac{\lambda}{\gamma}\sqrt{T}\right).$$

Introduce a new time parameter τ by $\tau = (\lambda/\gamma)\sqrt{T}$. We can then write $d_{\mathrm{E}}(T) = x(\tau)$, where $x(\tau)$ satisfies the relation $F(x(\tau), \tau) = \kappa_{\mathrm{E}}$. Moreover, we have $\chi_{\mathrm{E}}(T) = \chi(\tau)$, where

$$\chi(\tau) = \exp\left(\left(\gamma - \frac{1}{2}\right)\tau^2\right)\alpha(\tau) \tag{A28}$$

and the function $\alpha(\tau)$ is defined by

$$\alpha(\tau) = F(x(\tau) - \gamma\tau, \tau) \quad \text{with } F(x(\tau), \tau) = \kappa_E. \tag{A29}$$

To verify the claims in (55)–(56), it is sufficient to prove the following:

$$\lim_{\tau\to\infty} x(\tau) = \Phi^{-1}(\kappa_{\mathrm{E}}), \quad \lim_{\tau\downarrow 0} \chi(\tau) = \kappa_{\mathrm{E}}, \quad \lim_{\tau\to\infty} \chi(\tau) = \infty, \quad \lim_{\tau\to\infty} \mathrm{e}^{-\varepsilon\tau^2}\chi(\tau) = 0 \quad \text{for all} \varepsilon > 0. \tag{A30}$$

The first property follows from (A22). Note that

$$\chi(\tau) = \mathrm{e}^{(\gamma-\frac{1}{2})\tau^2} F(x(\tau) - \gamma\tau, \tau) = \mathrm{e}^{\tau x(\tau)}\Phi((\gamma-1)\tau - x(\tau)) + \mathrm{e}^{(\gamma-\frac{1}{2})\tau^2}\Phi(x(\tau) - \gamma\tau). \tag{A31}$$

Since $\lim_{\tau\downarrow 0} x(\tau) = -\infty$ and $\lim_{\tau\downarrow 0} \tau x(\tau) = \log\kappa_{\mathrm{E}}$ by (A24), the second claim in (A30) follows. Consider now the limit as τ tends to infinity. Since $x(\tau)$ remains bounded for $\tau \to \infty$, the first term on the right in (A31) tends to ∞ at a rate that is exponential in τ. The second term can be rewritten as follows:

$$\exp\left(\left(\gamma - \frac{1}{2}\right)\tau^2\right)\Phi(x(\tau) - \gamma\tau) = \exp\left(-\frac{1}{2}(\gamma-1)^2\tau^2 - \frac{1}{2}x^2(\tau) + \gamma\tau x(\tau)\right)\frac{m(\gamma\tau - x(\tau))}{\sqrt{2\pi}}.$$

This shows that the second term tends to 0 as τ tends to ∞. Hence, the final two claims in (A30) are verified.

It remains to verify the claim that the ratio $\bar{\theta}_j^{\mathrm{E}}/\theta_j^{\mathrm{E}}$ is increasing in j. Using the same transformation of time as above, this claim will be proved if it can be shown that the function $\hat{\chi}(\tau) := \exp(\frac{1}{2}\tau^2)\chi(\tau) = \exp(\gamma\tau^2)\alpha(\tau)$ is increasing in

τ. This can be done by showing that positivity of its derivative, which will follow if it can be proved that $2\gamma\tau\alpha(\tau) + \alpha'(\tau) > 0$ for $\tau > 0$. We can write

$$\begin{aligned}\frac{2\gamma\tau\alpha(\tau)+\alpha'(\tau)}{\varphi(x(\tau)-\gamma\tau)} &= 2\gamma\tau(m(x(\tau)+(1-\gamma)\tau)+m(-x(\tau)+\gamma\tau)) \\ &\quad + \tau m(x(\tau)+(1-\gamma)\tau)\left(-\frac{x(\tau)m(x(\tau)+\tau)}{\tau m(x(\tau)+\tau)} - 1 - \gamma\right) \\ &\quad + (x(\tau)+(1-\gamma)\tau)m(x(\tau)+(1-\gamma)\tau) - 1 \\ &= 2\gamma\tau m(-x(\tau)+\gamma\tau) + \frac{m(x(\tau)+(1-\gamma)\tau)}{m(x(\tau)+\tau)} - 1.\end{aligned}$$

Since the function $m(x)$ is decreasing, the second term in the final expression above is larger than 1, and the desired conclusion follows.

Appendix 5. Claims in Section 5.4

It follows from (48) and (49) that the fractions $\bar{\theta}_j^{C}$ ($j \geq 1$) of initial capital reserved for successive generations under the scheme that equalizes certainty equivalents are such that $\bar{\theta}_j^{C} \propto \bar{\theta}_{C}(T_j)$, where

$$\bar{\theta}_{C}(T) = e^{-rT} F\left(d_{C}(T) - \lambda\sqrt{T}, \frac{\lambda}{\gamma}\sqrt{T}\right) \quad \text{with } F\left(d_{C}(T), \frac{1-\gamma}{\gamma}\lambda\sqrt{T}\right) = \kappa_{C}.$$

As shown in Appendix 2, the relation $F(d_{C}(T), (1-\gamma)(\lambda/\gamma)\sqrt{T}) = \kappa_{C}$ indeed defines $d_{C}(T)$ uniquely as a function of $T > 0$ for any given value of $\kappa_{C} > 1$. Define the function $\chi_{C}(T)$ by

$$\chi_{C}(T) = \exp(\bar{\rho}_{\infty}^{C} T)\bar{\theta}_{C}(T) = \exp\left(\left(\frac{1}{2}\lambda^2/\gamma\right) T\right) F\left(d_{C}(T) - \lambda\sqrt{T}, \frac{\lambda}{\gamma}\sqrt{T}\right).$$

Introduce, as in the previous section, a new time parameter τ by $\tau = (\lambda/\gamma)\sqrt{T}$. We can then write $d_{C}(T) = x(\tau)$, where $x(\tau)$ satisfies the relation $F(x(\tau), (1-\gamma)\tau) = \kappa_{C}$. Moreover, we have $\chi_{C}(T) = \chi(\tau)$, where

$$\chi(\tau) = \exp\left(\frac{1}{2}\gamma\tau^2\right)\alpha(\tau) \tag{A32}$$

and the function $\alpha(\tau)$ is defined by

$$\alpha(\tau) = F(x(\tau) - \gamma\tau, \tau) \quad \text{with } F(x(\tau), (1-\gamma)\tau) = \kappa_{C}. \tag{A33}$$

To verify the claims in Section 5.3, it is sufficient to prove the following:

$$\chi(\tau) \text{is increasing}, \quad \lim_{\tau\downarrow 0} \chi(\tau) = \kappa_{C}^{1/(1-\gamma)}, \quad \lim_{\tau\to\infty} \chi(\tau) = (\kappa_{C}-1)^{1/(1-\gamma)}. \tag{A34}$$

Note that

$$\chi(\tau) = e^{\frac{1}{2}\gamma\tau^2}\alpha(\tau) = e^{\frac{1}{2}(1-\gamma)\tau^2 + x(\tau)\tau}\Phi((\gamma-1)\tau - x(\tau)) + e^{\frac{1}{2}\gamma\tau^2}\Phi(x(\tau)-\gamma\tau). \tag{A35}$$

From (A33) and (A25), we have

$$x(\tau) = \frac{1}{2}(\gamma-1)\tau - \frac{\log(\kappa_{C}-1)}{(\gamma-1)\tau} + o(\tau^{-N})(\tau\to\infty), \tag{A36}$$

where N can be any natural number. This implies

$$\lim_{\tau\to\infty} x(\tau) = \infty, \quad \lim_{\tau\to\infty} (\gamma-1)\tau - x(\tau) = \infty, \quad \lim_{\tau\to\infty} x(\tau) - \gamma\tau = -\infty,$$

and moreover,

$$\frac{1}{2}(1-\gamma)\tau^2 + x(\tau)\tau = -\frac{\log(\kappa_{C}-1)}{\gamma-1} + o(\tau^{-1}).$$

Concerning the second term on the right-hand side of (A35), we can write

$$e^{\frac{1}{2}\gamma\tau^2}\Phi(x(\tau)-\gamma\tau) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2(\tau) + \gamma\tau x(\tau) - \frac{1}{2}\gamma^2\tau^2 + \frac{1}{2}\gamma\tau^2\right) m(\gamma\tau - x(\tau)),$$

where, by (A36), $\gamma\tau x(\tau) - \frac{1}{2}\gamma^2\tau^2 + \frac{1}{2}\gamma\tau^2 = O(1)$ as $\tau\to\infty$. It follows that

$$\lim_{\tau\to\infty} \chi(\tau) = (\kappa_{C}-1)^{1/(1-\gamma)}. \tag{A37}$$

In a similar way, one finds from (A33) and (A26):

$$\lim_{\tau \downarrow 0} \chi(\tau) = \kappa_{\mathrm{C}}^{1/(1-\gamma)}. \tag{A38}$$

It remains to prove that the function $\chi(\tau)$ is increasing. As in the previous section, this can be done by showing that the derivative is positive. Define functions

$$F_x^1(\tau) = (\partial F/\partial x)(x(\tau) - \gamma\tau, \tau), \quad F_x^2(\tau) = (\partial F/\partial x)(x(\tau), (1-\gamma)\tau),$$
$$F_a^1(\tau) = (\partial F/\partial a)(x(\tau) - \gamma\tau, \tau), \quad F_a^2(\tau) = (\partial F/\partial a)(x(\tau), (1-\gamma)\tau)$$

and note from (A17) and (A18) that

$$\frac{F_x^1(\tau)}{F_x^2(\tau)} = \frac{1}{1-\gamma}\,\frac{\varphi(x(\tau)-\gamma\tau)}{\varphi(x(\tau))}, \quad \frac{F_a^1(\tau)}{F_a^2(\tau)} = \frac{\varphi(x(\tau)-\gamma\tau)}{\varphi(x(\tau))}.$$

From the definition of $x(\tau)$ in (A33), it follows that $F_x^2(\tau)x'(\tau) + (1-\gamma)F_a^2(\tau) = 0$. We then can write

$$\alpha'(\tau) = F_x^1(\tau)\left(-\frac{(1-\gamma)F_a^2(\tau)}{F_x^2(\tau)} - \gamma\right) + F_a^1(\tau) = -\gamma F_x^1(\tau) = -\gamma\tau[\alpha(\tau) - \Phi(x(\tau)-\gamma\tau)]. \tag{A39}$$

It is now straightforward to compute the derivative of $\chi(\tau)$:

$$\chi'(\tau) = \mathrm{e}^{\frac{1}{2}\gamma\tau^2}(\gamma\tau\alpha(\tau) + \alpha'(\tau)) = \gamma\tau\,\mathrm{e}^{\frac{1}{2}\gamma\tau^2}\Phi(x(\tau)-\gamma\tau). \tag{A40}$$

This is indeed seen to be positive.