

Pension Fund Design under Long-term Fairness Constraints

Esben Masotti Kryger

Laboratory of Actuarial Mathematics, Department of Mathematical Sciences, University of
Copenhagen, Universitetsparken 5, København Ø, DK-2100, Denmark.

We consider optimal portfolio insurance for a mutually owned with-profits pension fund. First, intergenerational fairness is imposed by requiring that the pension fund is driven towards a steady state. Subject to this condition the optimal asset allocation is identified among the class of constant proportion portfolio insurance strategies by maximising expected power utility of the benefit. For a simple contract approximate analytical results are available and accurate, whereas for a more involved contract Monte Carlo methods must be applied to pick out the best design. The main insights are (i) aggressive investment strategies are disastrous for the clients; (ii) in most cases it is optimal to gear the bonus reserve; and (iii) the results are far less sensitive to the agent's risk aversion than in the classical case of Merton (1969), and as opposed to Merton horizon matters even with constant investment opportunities (because of the serial dependence between bonuses).

The Geneva Risk and Insurance Review (2010) 35, 130–159. doi:10.1057/grir.2009.10;
published online 23 March 2010

Keywords: intergenerational fairness; bonus option; CPPI; stationarity; with-profits; Life and Pension Insurance

Introduction

Design of fair pension contracts has received some attention in the academic literature over the last 10 years. Traditionally the notion of fairness concerns the relationship between disjoint stakeholders, namely (equity) owners and a group of clients. In this line of work the main reference is Grosen and Jørgensen (2002) (extending Briys and de Varenne (1997)). Their model, however, is defined on a finite horizon and has no intergenerational considerations. Thus, in essence it is a single life model. We let the fund live on an infinite horizon, thus allowing transferring wealth between generations through the bonus reserve. This will lead to some degree of unfairness between different cohorts of clients. Therefore, we impose stationarity (of the funding ratio) at the outset, and consider the company only in its (distant future)

invariant condition. This implies that, as seen from today, (distant) future clients are all treated the same. Given this restriction, the “best” such distribution is identified by maximising expected power utility of the resulting benefit. In turn this yields an optimal strategy for portfolio insurance (to obtain stationarity it is necessary to impose an investment rule that guarantees absence of liquidation). Our model company has no group of equity holders disjoint from the clients, but nevertheless the framework of Grosen and Jørgensen (2002) is useful for designing fair contracts. For a number of reasons, however, we choose to adopt a different approach.

In designing optimal strategies for managing (distributing and investing) bonus reserves for individual contracts Steffensen (2004) uses the framework of Hindi and Huang (1993) to find the optimal distribution rule, which turns out to give rise to so-called local time payments (loosely speaking the optimal distribution policy consists of giving an infinitesimal amount of bonus whenever a certain barrier is hit, which happens infinitely often), and an optimal asset allocation strategy which, at least in a special case, turns out to be a modified version of the mutual fund separation theorem. Inspired by the results we impose a discrete time version of his bonus distribution rule.

The paper by Grosen and Jørgensen (2000) partly remedies the concerns over single life models and our approach is very much in the spirit of their work – even if we differ in vital aspects. Also, Preisel *et al.* (forthcoming) abandon the single life approach and their paper serves as the blueprint for this paper. Accordingly, it will be discussed thoroughly below and its framework will be used, albeit slightly modified, as a basis for the problems considered in this paper. Another important contribution taking the clients’ point of view and integrating the overall pension fund dynamics is the recent paper by Døskeland and Nordahl (2008b).

As implied in the preceding paragraphs, we believe that the main shortcoming in the existing literature on pension fund design is that it lacks disentanglement of the individual contracts from the overall financial status of the company. Such separation is necessary to fully understand the complex dynamics of the entire entity. We offer a new approach to optimising with-profits pension fund operation that can supplement the existing literature on this topic.

As pointed out by Døskeland and Nordahl (2008a) different, even non-overlapping, generations in a with-profits pension fund may systematically subsidise each other because the bonus reserve does not belong to a specific subset of the collective, but to the collective as a diffuse whole. One way of partially overcoming this is to price each contract *correctly* by adjusting the terms to reflect the fund’s financial and demographic condition at the time of underwriting. In practice, however, the stipulations are not adapted to

fit the economic situation of the fund, hence it is unattractive to enter funds that are poor (and possibly also funds that are increasing in size because bonus reserves may be crowded out by new entrants).

We take the view of an altruistic board of a mutually owned with-profits pension fund seeking to treat future generations equally. This property could be obtained by valuing the bonus option at the time of underwriting, and charging for it. But that approach is not desirable since it will compel the fund to put the bonus option on the balance sheet. Hence, in the short run, our board accepts that the clients are not treated equally. One reason for giving the board such influence over future generations could be that it represents some external party, say a union or governmental institution representing the common good rather than the present owners – or possibly subsidising the fund at its foundation. Alternatively the demand for intergenerational fairness may come from some authority. The point is that in the presence of a positive bonus reserve it may be in the interest of the present clients as a whole to dissolve the company rather than leave anything to generations to come. To counter that we let some external party design the fund.

In Section “Model” the pension fund, its clients, rules and environment are introduced along with the general contract that is analysed in the paper. Our notion of fairness is introduced in Section “Defining fairness”. Section “Analytical results” contains “approximate analytical” results for the optimal operation of the fund, considering a simple contract, whereas another type of approximation is used to get simulated results in Section “Numerical results”. The latter results are partly intended to assess the quality of the analytical approximations and partly intended to derive optimal rules for more complex contracts. Further, the speed at which the fund drives towards stationarity is analysed. Finally, Section “Concluding remarks” discusses and summarises the results of the preceding sections, while extensions and limitations are also touched upon.

Model

The environment in which the pension fund operates is as follows:

Let $(B_t)_{t \geq 0}$ be a Brownian motion on a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ generating the filtration $(\mathcal{F}_t)_{t \geq 0}$ with $\mathcal{F}_t \triangleq \sigma(B_s, 0 \leq s \leq t) \cup \mathcal{N}$, $(t \geq 0)$ – that is augmented by the null sets.

To the pension fund, but not necessarily to its individual clients, the financial market is frictionless regarding taxes, divisibility, transaction costs, liquidity and portfolio restrictions. This market consists of a bank account with interest intensity r , and a single risky asset (think of a well-diversified portfolio of risky

assets) with volatility σ and market price of risk Λ . Hence, the joint value process is

$$dS_t = \text{diag} S_t \left[(r_t, r_t + \Lambda_t \sigma_t)^\top dt + (0, \sigma_t)^\top dB_t \right], \quad (t \geq 0, S_0 > (0, 0)).$$

We will consider $r_t = r$, $\Lambda_t = \Lambda > 0$, and $\sigma_t = \sigma > 0$ for all $t \geq 0$, that is constant investment opportunities.

The fund divides its assets, A , between the risky asset and the bank account with a fraction, π , invested in the former. Liabilities, L , representing the progressive reserve, changes by the interest rate – hence between accounting periods indexed by 1, 2, ...

$$\begin{aligned} dA_t &= A_t [(r + \pi_t \Lambda \sigma) dt + \pi_t \sigma dB_t], \quad (t \in i + (0, 1), i \in \mathbb{N}, A_0 > 0). \\ dL_t &= L_t r dt, \quad (t \in i + (0, 1), i \in \mathbb{N}, L_0 > 0). \end{aligned}$$

One could allow for additional, non-marketed noise in the processes A and L , but we will impose the simplification that the only source of uncertainty is B , which is hedgeable. Hence, in line with recent studies, for example Døskeland and Nordahl (2008a), we focus on the savings part of the contract, in particular we disregard mortality.

The pension fund we consider is a mutual with-profits fund, that is one that is owned by its clients. Also, entry is not voluntary, but governed by, say, legislation and contributions are fixed. These assumptions imitate real life with-profits pension funds well. Yet, this ownership structure is non-standard in the literature and hence our model cannot be directly compared to those of Grosen and Jørgensen (2000, 2002); Døskeland and Nordahl (2008a, b). One could consider “old” and “new” clients as disjoint stakeholders; owners respectively customers, but that approach does not fit our purpose; nor does it reflect actual with-profits pensions systems. The compulsory membership may imply that some clients will enter on unacceptable terms, since it is likely preferable to enter a wealthy fund.

The *funding ratio* is defined as

$$F_t \triangleq \frac{A_t}{L_t}, \quad (t \geq 0).$$

Since avoiding insolvency is an integral part of intergenerational fairness we shall require throughout that the fund is always sufficiently liquid – in the special sense that $F > 1 + c$ for some $c > -1$ representing a minimum acceptable funding ratio (from the point of view of the fund’s board, but possibly laid down by some monitoring authority). Hence, we shall denote the surplus assets $A - L(1 + c)$ the *bonus reserve*. We let $c=0$ – corresponding to liquidation upon insolvency –

throughout (except in Subsection “Sensitivity analysis”). Any initial bonus reserve ($F_0 > 1 + c$) could have come from anywhere, for example as an inheritance from previous generations or as a subsidy from somewhere else.

In Preisel *et al.* (forthcoming), the asset process is controlled by choosing the fraction of wealth invested in the risky asset by optimising one-period expected power utility of the end-period funding ratio minus one. This criterion gives rise to a constant proportion portfolio insurance (CPPI) strategy (to be introduced in Subsection “Investment strategy”) parameterised by the *manager’s* coefficient of relative risk aversion, $\gamma > 0$. We, on the other hand, impose a parameterised investment strategy at the outset and take the clients’ point of view as a basis for optimisation. This disparity is a natural consequence of our objective being completely different from theirs. For where their aim is to point out the potential conflict between short- and long-viewed stakeholders our ambition is the study of *optimal* design as seen from an altruistic standpoint.

The fund has a rule of distributing bonus periodically,¹ but only when its funding ratio at the turn of the period exceeds some fixed threshold, $\kappa > 1 + c^+$, and in that case all funds above the threshold are distributed to the clients. We shall refer to κ as the *bonus barrier*.

At the turn of an accounting period assets and liabilities are A_{i-} and L_{i-} . Then new contracts are underwritten with value $\Gamma_i L_{i-}$, ($\Gamma_i \geq 0$) and converted into liabilities $g_i \Gamma_i L_{i-}$. The parameter $g_i > 0$ measures the proportion of contributions which is converted into liabilities at time i .² Due to the presence of a bonus reserve this parameter may be less than one. Traditionally, this contribution to the (collective) bonus reserve is not explicit. At the same time contracts mature with market value $\Pi_i L_{i-}$, ($\Pi_i \in [0, 1]$). This gives rise to the end year *post bonus* funding ratio

$$\begin{aligned}
 F_{i^+} &= \frac{A_{i-} + L_{i-}(\Gamma_i - \Pi_i)}{L_{i-}(1 + g_i \Gamma_i - \Pi_i)} \wedge \kappa \\
 &= \frac{F_{i-} + \Gamma_i - \Pi_i}{1 + g_i \Gamma_i - \Pi_i} \wedge \kappa, \quad (i \in \mathbb{N}).
 \end{aligned}
 \tag{1}$$

The bonus, b_i , that is, in fact, allotted such that $L_{i^+} = L_{i-}(e^{b_i} + g_i \Gamma_i - \Pi_i)$, $A_{i^+} = A_{i-} + L_{i-}(\Gamma_i - \Pi_i)$, and (1) is satisfied is

$$b_i \triangleq \left(\log \frac{F_{i-} + \Pi_i(\kappa - 1) - \Gamma_i(\kappa g_i - 1)}{\kappa} \right)^+, \quad (i \in \mathbb{N}).$$

¹ As an alternative to increasing future benefits the company could pay out a cash dividend. For an argument in favor of the former approach, see Nielsen (2006).

² The proportion $1 - g_i$ is intended to pay for the bonus option.

We let $g=1$ throughout such that if new contributions balance expiring claims, $\Gamma_i=\Pi_i$, leaving clients subsidise new clients by their “share” of the bonus reserve. If new net contributions are positive, $\Gamma_i>\Pi_i$, existing clients also subsidise new ones – and if they are negative existing clients also gain from terminating contracts. We shall assume that $\Gamma_i=\Pi_i=0$, ($i\in\mathbb{N}$), that is in some sense the fund is mature. It is indeed relevant to study growing or shrinking funds, but for our purpose it makes more sense to study the case of balanced cash flow. Consequently,

$$F_{i^+} = F_{i^-} \wedge \kappa, \text{ and } b_i = \left(\log \frac{F_{i^-}}{\kappa} \right)^+, \quad (i \in \mathbb{N}).$$

The described bonus rule is not widespread in the academic world nor in practice, where there are tactical, strategical, distributional (intergenerational) and political reasons for smoothing bonus distribution. Rather it is chosen for technical reasons and because of the results of Steffensen (2004) – and as an approximation to what is, in fact, practiced.

After settling on a short-term optimisation criterion Preisel *et al.* (forthcoming) investigate the properties of the implied stationary distribution of F . Their aim is to point out the divergence between long- and short-viewed stakeholders. Concerning the objectives of the present study their model has a few shortcomings, however. First, their optimisation criterion is inappropriate for our purpose. Second, they do not discuss the choice of bonus barrier. And third, their paper does not study the rate at which the system converges towards stationarity. The aim of this paper is to remedy these shortcomings. We shall address the first of these reservations by introducing a different, altruistic optimisation criterion in Sections “Defining fairness” and “Analytical results”. The second and third points of criticism turn out to be partially interrelated, and we shall discuss those topics in Section “Defining fairness” and Subsection “Speed towards stationarity”.

Investment strategy

An investor with assets, A_0 , and a, possibly random, “floor” on wealth $L_0 < A_0$ is said to follow a CPPI strategy (introduced by Black and Perold (1992)) with multiplier $\alpha > 0$, if his portfolio is self-financing and his absolute allocation to risky assets at time $t \geq 0$ is $\alpha(A_t - L_t)$. The strategy thus reduces exposure when the cushion, $A - L$, decreases and vice versa. In particular, as the cushion approaches zero, the allocation to risky assets approaches zero. Therefore, if paths are continuous, the strategy ensures that the cushion is always positive.

As mentioned above, the optimisation criterion of Preisel *et al.* (forthcoming) gives rise to a particular CPPI strategy, namely one with multiplier $\lambda\sigma^{-1}/\gamma$. We shall take a different approach and impose a general CPPI strategy at the outset. In particular, we use the family of parameterised investment strategies

$$\pi_t \triangleq \alpha \frac{A_t - L_t(1 + c)}{A_t}, \quad (t \geq 0, \alpha > 0). \tag{2}$$

The motivation for choosing this strategy at the outset is that it implies a zero probability of default in the present model framework.³

Under condition (2) the discrete time funding ratio dynamics between updates has i.i.d. lognormal innovations:

Theorem 1

$$\log \left(\frac{F_{i-} - (1 + c)}{F_{(i-1)^+} - (1 + c)} \right) = Z_i, \quad (i \geq 1).$$

Here $(Z_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence with $Z_1 \sim N(m, s^2)$, $(i \in \mathbb{N})$, $s \triangleq \alpha\sigma$ and $m \triangleq s(A - s/2)$.

The main result of Preisel *et al.* (forthcoming) is their Theorem 4.1, which states that $(F_{i+})_{i \in \mathbb{N}}$ admits a stationary distribution if and only if Z_1 has a strictly positive mean. In our case this translates into the condition $s \in (0, 2A)$. Note that this requirement is independent of the choice of barrier, κ . s is the volatility of the bonus reserve, $A - L(1 + c)$, and we will refer to s as “risk”.

Below, we come up with an analytical approximation for the stationary distribution of the funding ratio.

The contract

To fulfil the purpose of the paper, we consider a contract spanning a period of length $n \in \mathbb{N}$, which can be taken to represent the typical savings period for a

³ It is not clear by any means that it is optimal for the clients as a whole to impose zero probability of default. Having fairness as our primary concern it does *seem* reasonable, however, to put down this restriction already at the modelling level. An alternative with that property is to implement an Option-based Portfolio Insurance strategy, which – like the CPPI – secures a pre-specified lower boundary on portfolio value at the chosen horizon via a put option on the asset portfolio. For a wholly different approach we may decide to use a “free” strategy not protecting the excess assets, instead allowing for bankruptcy.

pension fund client, and with the first contribution made at time $M \in \mathbb{N}$ (with “today” being time 0). The contract consists of a set of contributions, $(\zeta(t))_{t=M}^{M+n}$, which are determined at time 0 even for a forward starting contract ($M > 0$), and a terminal benefit $W(n)$. To ease calculations we assume without losing much realism that there are no expenses (administrative costs etc.) associated with the contract. The benefit is

$$W(t) \triangleq \sum_{j=0}^{\lfloor t \rfloor} g(M+j)\zeta(M+j)e^{r(n-j)}e^{\sum_{k=j+1}^{\lfloor t \rfloor} b_{M+k}}, \quad (0 \leq t \leq n). \quad (3)$$

That is, by the terms of the contract, the contribution at time l is transformed into a claim at expiry of $g(l)\zeta(l)e^{r(M+n-l)}$ – and bonus may be added. Hence, contribution $\zeta(l)$ is not awarded bonus until time $l+1$. Since bonus is – in part – intended to reflect the return on the contribution it is natural to refrain from crediting bonus immediately. Typically, either $\zeta(t)=1_{(t=M)}$, or $\zeta(t)=1_{(t \in \{M, M+1, \dots, M+n\})}$, but it is also possible to have, say, an increasing contribution plan reflecting inflation. The contract can obviously be thought of as representing a capital pension. The presence of a bonus reserve, however, is usually linked to an insurance product. Surrender and free policy options are not considered.

Since the guaranteed interest rate equals the market rate, $g=1$, and there is no default risk the contract is an arbitrage. As argued by Døskeland and Nordahl (2008a) this does not really pose a problem. The fact that there is certain return to be earned beyond the guaranteed return of r is a consequence of the intergenerational subsidisation that built up a bonus reserve in the past. In other words, the return on the (random) amount $A_{M-}-L_{M-}(1+c)$ that previous generations, or some external party, made available, is handed over to the generation entering the fund at time M .

For a provocative setting, essentially suggesting no guarantees whatsoever, see Sørensen and Jensen (2001). It seems obvious, however, that this can be a gateway for managerial malpractice. Also, they are partially countered by Døskeland and Nordahl (2008b).

Defining fairness

Fairness between the company and its collective of clients is imposed by construction, since they form a closed circuit. Our concept of fairness, however, is an intergenerational one. It is based on the wish that future clients who, in fact, join the fund at times when it is funded differently will get a benefit with the same distribution – as seen from today, regardless of the conditions at the time of underwriting.

Definition 1 A with-profits pension fund is *long-term fair* if $W(n)$ is independent of F_0 for some forward-starting contract and for each n .

As a way of obtaining fairness we shall require that an invariant distribution for the funding ratio exists, and further we shall assume that stationarity occurs in finite time.

Assumption 1 Assume that F is invariantly distributed at time M , or more formally

$$\forall F_{0+} > 1 + c \exists M \in \mathbb{N} : \forall f \in \mathbb{R} : \mathbb{P}(F_{M-} \leq f) = \lim_{i \rightarrow \infty} \mathbb{P}(F_{i-} \leq f)_{0 < s < 2A}.$$

If no stationary distribution exists the probability of obtaining bonus over any fixed horizon will tend to zero. It is not completely obvious that this is, in fact, unsatisfactory to the clients as a whole, but it is clear that it will imply that intergenerational fairness cannot be obtained.

Theorem 2 Under Assumption 1 the pension fund is long-term fair if and only if $0 < s < 2A$.

Given the fairness side condition, $0 < s < 2A$, we wish to find the stationary distribution of bonus that will satisfy future clients better. To this end we shall maximise expected power utility of the benefit, $W(n)$. The reason for choosing a utility criterion over a financial one is that pension contracts are likely non-tradeable.

The only control we have at our disposal is the “risk”, s , since by stochastic dominance it does not make any sense to optimise over κ ; for the higher is κ the better off is any client joining at time M or later (for fixed s), c.f. Theorem 3 below. The real trade-off here is between waiting for an attractive funding ratio distribution (high κ) on one side and getting bonus underway and getting to stationarity quickly (low κ) on the other side, c.f. Subsection “Speed towards stationarity”. Once stationarity is attained no such prioritisation has to be made.

Analytical results

The stationary distribution of F is not explicitly known; and hence, we shall study a different Markov chain for which the invariant distribution can be identified. Following Preisel *et al.* (forthcoming), we impose

Assumption 2 The sequence $(Z_i)_{i \in \mathbb{N}}$ is i.i.d. with Z_1 Laplace distributed with location m and scale $\lambda^{-1} \triangleq s/\sqrt{2}$ (picked to match the variance of the true Z_1).

Under Assumption 2 the density of Z_I is

$$\frac{\lambda}{2} \exp(-\lambda|x - m|), \quad (x \in \mathbb{R}). \quad (4)$$

We are not familiar with any continuous time stochastic processes for the financial market bringing about this dynamics, but shall use the approximation nevertheless. The assumption may be justified by referring to the fact that this Laplace distribution is also symmetric about its mean and is constructed to have the same variance as the true one. Considering the tail behaviour, however, some differences occur because of the fatter tails in (4). The 4th central moment is twice that of the true distribution (corresponding to excess kurtosis of 3), and at higher order, even moments will, of course, differ more. Notice, however, that such fatter tails comply with some of the criticism of using normally distributed returns.

To study the quantitative properties of bonus we need the funding ratio *prior* to bonus distribution, which has the following stationary distribution function.

Theorem 3 Under Assumptions 1 (stationarity) and 2 (Laplace)

$$G^{(-)}(x) \triangleq \mathbb{P}(F_{M^-} \leq x) = \begin{cases} \frac{\lambda - \rho}{\lambda} \left(\frac{x - (1+c)}{\kappa - (1+c)} \right)^\rho & 1 + c < x \leq \kappa e^m - (1+c)(e^m - 1) \\ 1 - \frac{\rho e^{\lambda m}}{\lambda + \rho} \left(\frac{x - (1+c)}{\kappa - (1+c)} \right)^{-\lambda} & x > \kappa e^m - (1+c)(e^m - 1). \end{cases} \quad (5)$$

The parameter ρ is the non-zero solution to the non-linear equation $1 - (\rho/\lambda)^2 = \exp(-\rho m)$. This implies that $\rho \in (0, \lambda)$.

Expression (5) differs from that of Preisel *et al.* (forthcoming), who mix Laplace and Gaussian distributed innovations to derive an approximation to the stationary pre-bonus funding ratio distribution.

Because ρ and λ both decrease in s , the distribution is spread out more the higher is s (and the higher is κ). As the stationary marginal probability of obtaining bonus is unaffected by κ the bonus increases in κ . On the other hand that bonus probability decreases with s , so the *conditional* bonus increases with s . An example of the stationary funding ratio distribution can be seen in Figure 2.

When we consider the Markov chain obtained via the Laplace assumption (Assumption 2) the stationary moments of bonus can be derived. They turn out to be expressed in terms of hypergeometric functions (see Weisstein, 2008), which can be evaluated precisely and quickly. The second moment can only be derived when $c=0$. Therefore, we shall only calculate the moments of b_M explicitly in that case.

Theorem 4 Under Assumptions 1 (stationarity) and 2 (Laplace) all moments of b_M exist, and for $c=0$

$$\begin{aligned} \mathbb{E}\{b_M\} &= \frac{\rho(\lambda - \rho)}{\lambda} (\kappa - 1)^{-\rho} [H_1(\rho, (\kappa - 1)e^m + 1) - H_1(\rho, \kappa)] \\ &\quad - \frac{\lambda \rho e^{\lambda m}}{\lambda + \rho} (\kappa - 1)^\lambda H_1(-\lambda, (\kappa - 1)e^m + 1) - \frac{\rho}{\lambda} \log \kappa. \\ \mathbb{E}\{b_M^2\} &= \frac{\rho(\lambda - \rho)}{\lambda} (\kappa - 1)^{-\rho} [H_2(\rho, (\kappa - 1)e^m + 1) - H_2(\rho, \kappa)] \\ &\quad - \frac{\lambda \rho e^{\lambda m}}{\lambda + \rho} (\kappa - 1)^\lambda H_2(-\lambda, (\kappa - 1)e^m + 1) \\ &\quad - \frac{\rho}{\lambda} (\log \kappa)^2 - 2 \log \kappa \mathbb{E}\{b_M\}. \end{aligned}$$

The functions H_1 and H_2 can be found in the Appendix.

A simple contract

Analytically we shall, for technical reasons, only consider the contract consisting of a single unit of contribution initially (at time M) transformed into a benefit at time $M + n$ of

$$W(n) = e^{rn} \exp\left(\sum_{k=1}^n b_{M+k}\right),$$

that is a contract with $\zeta(t)=1_{(t=M)}$ and $g(M)=1$. Due to the presence of serially correlated bonuses, $W(n)$ has a complicated distribution that we cannot explicitly derive. Instead we shall impose two additional assumptions to facilitate a semi analytical solution. First, we approximate the serial correlation of $(b_{M+i})_{i \geq 0}$ by analysing the unrestricted underlying random walk with positive drift $(Z_1, Z_1 + Z_2, \dots)$:

Theorem 5 Given Assumption 1 (stationarity), and assuming that $\forall i \geq 0: (b_{M+i} | b_M b_{M+i} > 0)$ are identically distributed.

$$\begin{aligned} \mathbb{E}\{b_M b_{M+1}\} &= \mathbb{E}\{b_M\}^2 P_1 / \mathbb{P}(b_M > 0). \\ \mathbb{E}\{b_M b_{M+2}\} &= \mathbb{E}\{b_M\}^2 \left(\frac{P_2}{2} + \frac{P_1^2}{2!}\right) / \mathbb{P}(b_M > 0). \\ \mathbb{E}\{b_M b_{M+3}\} &= \mathbb{E}\{b_M\}^2 \left(\frac{P_3}{3} + \frac{P_1^3}{3!} + \frac{P_1 P_2}{2}\right) / \mathbb{P}(b_M > 0), \end{aligned}$$

where $P_j \triangleq \mathbb{P}(Z_1 + \dots + Z_j < 0)$, ($j \geq 1$).

Remark 1 To reduce the number of factors in our expressions below we choose to stop at three moments in Theorem 5. If desired one could include further moments to improve accuracy.

In order to proceed, we need to estimate the higher order serial correlations. To this end assume the following decay:

Assumption 3

$$\begin{aligned} \text{Corr}\{b_{M+i}; b_{M+j}\} &= \rho_{0,1}q^{(j-i)-1}, \quad (j > i \geq 0). \\ \rho_{i,j} &\triangleq \text{Corr}\{b_{M+i}; b_{M+j}\}, \quad (i, j \in \mathbb{N}). \\ q &\triangleq \frac{\rho_{0,3}}{\rho_{0,2}}. \end{aligned}$$

Remark 2 For the tail behaviour we used q – the relationship between step-3 and step-2 covariances. This implies that neither step-2 nor step-3 covariance contribute with our best estimates for them. This simplification does not matter much, though, and can be easily remedied.

Theorem 6 Under Assumptions 1 (stationarity) and 3 (correlation decay pattern)

$$\begin{aligned} \hat{V}(n) &\triangleq \mathbb{V}\left\{\sum_{i=1}^n b_{M+i}\right\} / n \\ &= \mathbb{V}\{b_M\} \left(1 + \frac{2\rho_{0,1}}{1-q} \left(1 - \frac{1-q^n}{n(1-q)}\right)\right), \quad (n \geq 1). \end{aligned}$$

Notice that $n \rightarrow \hat{V}(n)$ is increasing because $q \in (0, 1)$. To get an analytical expression for the optimisation problem we will also need a distributional assumption for the aggregated bonus. The computationally convenient choice is the normal distribution, which is also asymptotically correct by the central limit theorem. For finite horizons, however, it will be flawed by the artificial, negative value space. Nevertheless, we settle on that assumption:

Assumption 4

$$\sum_{i=1}^n b_{M+i} \sim N(n\mathbb{E}\{b_M\}, n\hat{V}(n)), \quad (n \geq 1).$$

To measure the merits of a particular contract we employ expected discounted power utility using a *deterministic*, integrable consumption discounting process $(v_t)_{t \geq 0}$, that is the maximisation object is

$$U_\gamma(W(n)) \triangleq \begin{cases} \mathbb{E} \left\{ e^{-\int_0^{M+n} v_t dt} \frac{W(n)^{1-\gamma}}{1-\gamma} \right\} & \gamma \in [0, \infty) \setminus \{1\}. \\ \mathbb{E} \left\{ e^{-\int_0^{M+n} v_t dt} \log W(n) \right\} & \gamma = 1. \end{cases} \tag{6}$$

Since v and r are deterministic, maximising (6) is equivalent to maximising expected power utility of

$$X \triangleq \exp \left(\sum_{k=1}^n b_{M+k} \right),$$

that is we can disregard v , r and the passing of time – and assume that the aggregated bonus, X , is received immediately. This is a special feature of the particular utility functions used here. We formalise this notion as

Theorem 7 Given Assumptions 1 (stationarity), 3 (correlation decay pattern) and 4 (normality), maximising expected discounted power utility (with relative risk aversion $\gamma \in [0, \infty)$) of the benefit $W(n)$ is equivalent to maximising the certainty equivalent bonus

$$b_{CE} \triangleq \mathbb{E}\{b_M\} + \frac{1-\gamma}{2} \hat{V}(n).$$

Remark 3 If $\gamma=1$ or $n=1$ only Assumptions 1 and 2 are required.

Remark 4 The maximisation object we have chosen is standard, but notice that it implicitly considers the contribution as sunk cost (else $X-1$ is the appropriate object).

Optimisation

We will now apply Theorem 7 to find the optimal s in various cases and analyse these. Throughout this section we will keep $\lambda=1/4$ (having periods of 1 year in mind). We consider variations in the length of the contract (n), the bonus barrier (κ), and the coefficient of relative risk aversion (γ).

Power utility

The first observation from Table 1 presenting the optimal risk is that it does not vary very much with κ nor with n (except at high levels of risk aversion).

Table 1 Optimal risk(s) based on analytical approximation

κ	n	γ					
		0	0.5	1	2	5	10
1.1	1	0.277	0.275	0.273	0.268	0.254	0.231
1.1	10	0.281	0.277	0.273	0.264	0.236	0.194
1.1	30	0.283	0.278	0.273	0.261	0.226	0.178
1.1	50	0.284	0.278	0.273	0.261	0.224	0.175
1.2	1	0.276	0.273	0.269	0.262	0.239	0.201
1.2	10	0.283	0.276	0.269	0.254	0.207	0.150
1.2	30	0.287	0.278	0.269	0.250	0.193	0.133
1.2	50	0.287	0.279	0.269	0.249	0.190	0.130
1.3	1	0.276	0.272	0.267	0.258	0.227	0.181
1.3	10	0.284	0.276	0.267	0.247	0.187	0.125
1.3	30	0.289	0.279	0.267	0.241	0.171	0.110
1.3	50	0.290	0.279	0.267	0.240	0.168	0.107

When $\gamma < 1$ the increased $\hat{V}(n)$ that results from a longer contract implies higher optimal risk – and oppositely if $\gamma > 1$. As can be seen from Theorem 7 this is a concavity effect. Similarly, at the lowest values of γ an increase in κ will induce higher optimal risk – and oppositely if γ is higher. In this case, the reason is that higher levels of κ are associated with relatively more variable outcomes than are lower barriers. As expected the optimal s is decreasing in γ (because $\hat{V}(n)$ increases with s , at least over the relevant range).

It is quite remarkable that even risk neutral clients prefer investment strategies, which are only modestly aggressive, far less than the upper boundary of 0.5. Apparently, the fat right tail of the stationary bonus distribution associated with aggressive investment strategies does not sufficiently compensate the lower marginal probability of obtaining bonus.

The classical Merton (1969) analogy to s is A/γ , which would be ∞ , 0.5, 0.25, 0.125, 0.05 and 0.025, respectively, in the rightmost columns of Table 1. Also, in the case of Merton (1969), horizon does not matter when investment opportunities are constant, but in our case n is clearly important (especially at high levels of risk aversion) because bonus is positively serially correlated, so that $\hat{V}(n)$ increases with n . Altogether, except at $\gamma=1$ the difference is enormous. This should come as no surprise since the problems are very unlike.

Seemingly, the choice of barrier, κ , is not so important for low and moderate levels of risk aversion, since the optimal s varies only little with this parameter. As regards the implied investment strategy this is true, of course, but the certainty equivalent bonus (not shown) differs very markedly across κ , that is in stationarity it is preferable to have a higher barrier because conditional

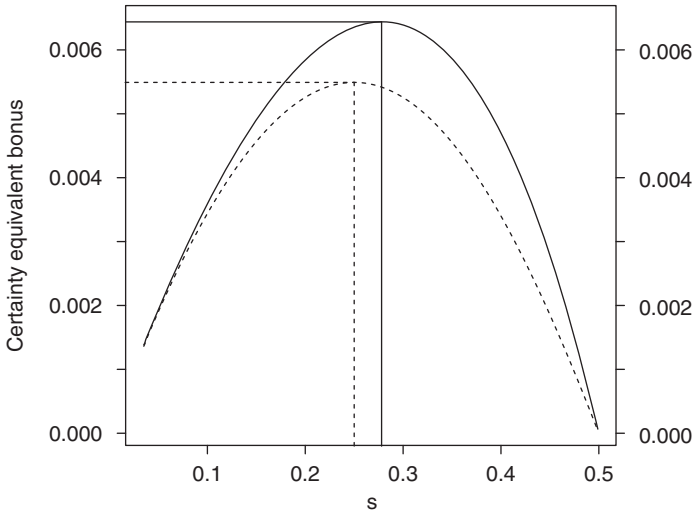


Figure 1. Certainty equivalent bonus as a function of s for $\gamma=0.5$ (full line) and $\gamma=2$ (dotted line). Fixed parameters: $\kappa=1.2$, $n=30$.
Note: Based on analytical approximation.

bonus increases with κ , whereas the marginal probability of obtaining bonus is not affected by this parameter.

Most optima are in the range $(0.2, 0.3)$; implying that for a risky asset volatility, σ , of less than 20 per cent, the bonus reserve should be geared since then $\alpha=s/\sigma > 1$.

Figure 1 shows the mapping $s \rightarrow b_{CE}$ for a certain parametrisation, but its appearance is quite representative across a broad range of configurations. The main insight is that it is not very steep around its maximum, implying that it is not overly important to evaluate γ correctly. In the example from Figure 1, when $\gamma=2$, the loss of terminal benefit from choosing $s=0.3$, rather than the optimal $s=0.25$, is far less than 1 per cent over 30 years. As discussed in the Introduction, entry to the pension fund provides an arbitrage, and therefore the certainty equivalent bonus is strictly positive.⁴

Mean–variance utility

It may be hard to be very specific as to your choice of utility function. To this end, as a pedagogical tool, we show in Figure 3 the mean–variance diagram,

⁴ But when using the normal approximation it needs not be so, of course.

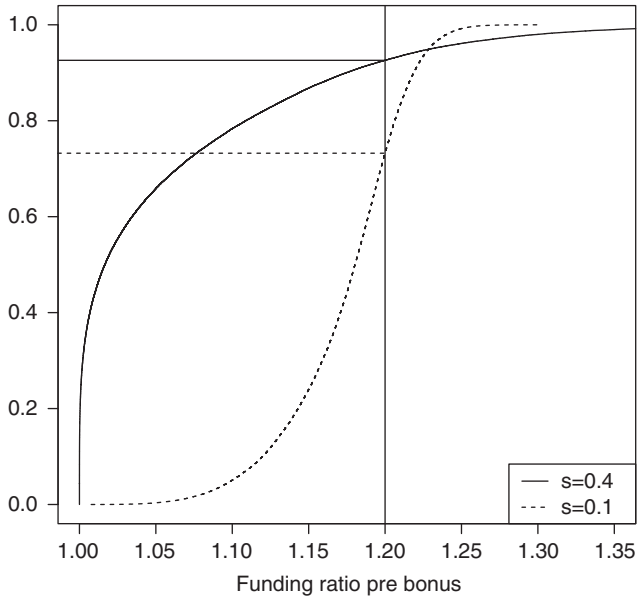


Figure 2. Stationary funding ratio distribution. Fixed parameters: $\kappa=1.2$.
Note: Horizontal lines indicate $\mathbb{P}(F_{M-} \leq \kappa) = \mathbb{P}(b_M = 0)$. Based on analytical approximation.

which will provide the set of optimal strategies for any agent with increasing utility of the mean and decreasing utility of the variance of aggregated bonus, and preferences over these two quantities only. One very useful insight conveyed by Figure 3 is the existence of investment strategies inducing stationarity, but which are mean–variance-inefficient. For as can be seen, as s is increased above a certain limit (depending on the parameters) the outcome worsens drastically. This phenomenon is also evident with power utility, c.f. Figure 1. The reason is that the previously mentioned trade-off between frequent and large bonuses, which was illustrated in Figure 2, tips in favour of more cautious strategies from a certain level of s because the bonus probability decreases too fast, or equivalently, the conditional bonus improves too slowly. In turn these characteristics, combined with the serial correlation, which also increases with s , translate into fatter tails of the benefit for more aggressive investment strategies. Figure 4 demonstrates this point, and thus implies how very adverse (as well as very favourable) outcomes are much more likely as risk increases. For a lengthy discussion of this important concept of being trapped at low funding levels, see Preisel *et al.* (forthcoming). Finally, notice that the normality assumption was not used to produce Figures 3 and 4.

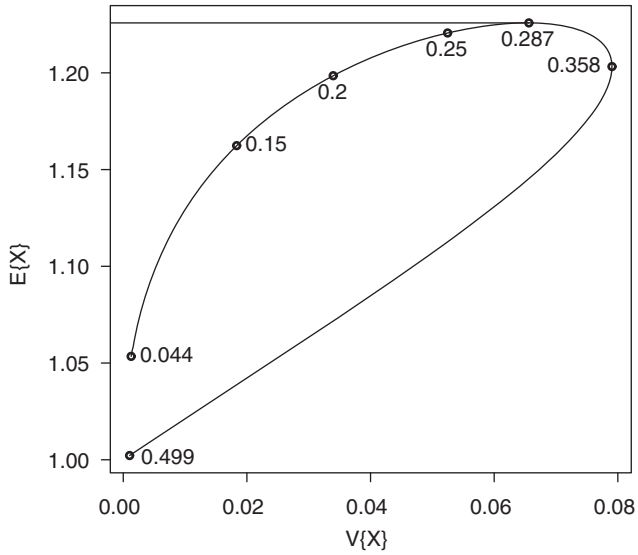


Figure 3. Mean–variance diagram for X . Fixed parameters: $\kappa=1.2$, $n=30$.
Note: The labelled points indicate various values of s . Based on analytical approximation.

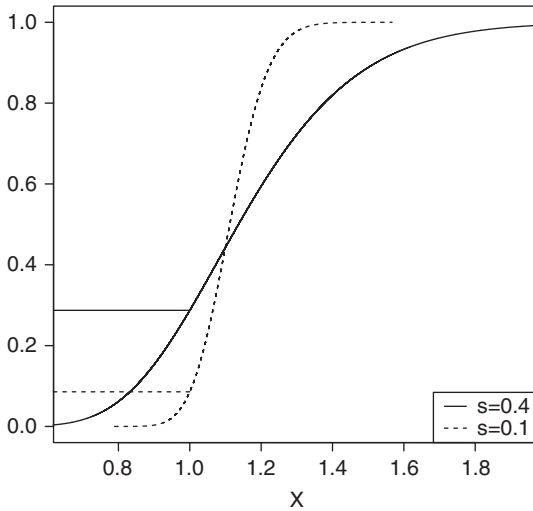


Figure 4. Distribution of X . Fixed parameters: $\kappa=1.2$, $n=30$.
Note: Horizontal lines indicate $\mathbb{P}(X \leq 1)$. Based on analytical approximation.

Limitations and accuracy

In Section “Numerical results”, we will provide support for the conclusions above by performing Monte Carlo simulation of the true dynamics. Applying this technique, we shall also find the optimal risk for a contract with several contributions. Further, the convergence of the funding ratio is analysed.

Numerical results

The simple contract considered above has a single contribution only, and is thus quite dissimilar to real life contracts. In this section, we shall meet this shortcoming by numerically maximising the expected power utility of a contract with several contributions. Also, we will provide support for the analytical approximation from the previous section. We start out by doing the latter in Subsection “Comparison to analytical results”. Then, in Subsection “A complex contract” we will identify the optimal design of contracts with several contributions. Subsection “Sensitivity analysis” briefly touches upon the dependence of the results towards the choice of model constants c and λ , whereas Subsection “Optimal strategies” discusses the optima. The rate at which the system converges is discussed in Subsection “Speed towards stationarity”.

Throughout we impose no approximations, except for using a finite state space rather than all of Ω . Our preferred tool in this section is Monte Carlo simulation, which is described briefly in Technical description of the Monte Carlo simulation section.

Comparison to analytical results

We shall now assess the overall quality of the analytical approximation by simulating the true dynamics using Theorem 1. First, the optima are compared, and afterwards, to explain the differences encountered, we look at the distribution of X .

Power utility optima

First, in Table 2 we provide optima comparable to those in Table 1. The comparison is quite uplifting; the difference in optimal allocations is less than a few percentage points *except* at high risk aversion, where the previous optima were too low. The main cause for this discrepancy is the fact the value space for X was extended below 1 by matching two moments only (i.e. probability mass was moved quadratically). Such transformation will make aggressive investment strategies appear artificially unattractive when the utility function is very concave.

Consequently, the true optima differ far less across γ than did the approximate ones which makes it easier to embrace individuals with different appetites for risk in a common investment policy.

Table 2 Optimal risk(s) based on simulation of true process as a percentage of the optima in Table 1

κ	n	γ					
		0	0.5	1	2	5	10
1.1	1	99	99	99	101	102	106
1.1	10	98	98	99	99	104	115
1.1	30	99	98	98	98	102	114
1.1	50	100	99	98	98	100	110
1.2	1	99	99	100	99	103	117
1.2	10	99	99	99	101	111	132
1.2	30	100	99	99	99	109	130
1.2	50	102	100	99	98	106	125
1.3	1	99	100	98	100	107	122
1.3	10	100	99	99	101	116	146
1.3	30	102	100	99	100	115	143
1.3	50	104	101	99	98	111	136

The optima can be backed out using Table 1.

At the outset it is not obvious if short or long contracts are optimised more precisely when applying the analytical approximation. The longer the contract in question the better the normality assumption (Assumption 4) works. Oppositely, the correlation assumption (Assumption 3) is worse for long contracts because the correlations decay slower than at rate q . From Table 2, we conclude that longer contracts are estimated more precisely, that is the normality assumption matters more (at least for the levels of risk aversion where the errors are more serious).⁵

The qualitative conclusions from the previous section regarding n , κ and γ hold true.

The quality of the analytical approximation

In terms of optima, we are pleased with the quality of the analytical approximation. The assumptions were made at a more primitive level, however, and we shall go to that to explain the deviations encountered. To this end consider Figure 5 comparing the true (simulated) distribution of X to that obtained via the analytical approximation. Clearly, the normality assumption (Assumption 4) brings about a rather substantial difference

⁵ In making this conclusion, we disregard the case $n=1$ where the variance of X is only approximated via Assumption 2 (Laplace).

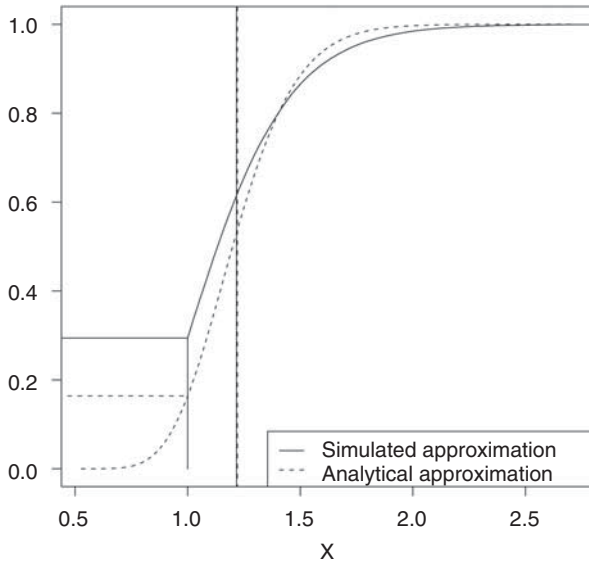


Figure 5. Distribution of X . Fixed parameters: $s=0.25$, $\kappa=1.2$, $n=30$.
Note: Horizontal lines indicate $\mathbb{P}(X \leq 1)$; vertical lines show $\mathbb{E}\{X\}$.

between the two by artificially extending the support of X below one. In fact, informal experiments indicate that almost the entire difference between the two distributions in Figure 5 stems from that assumption, whereas for reasonable values of horizon, n , the effect of the approximation regarding serial correlation (Assumption 3) is only of second order. This holds in spite of the latter assumption drastically reducing the true variance of X and thus contributing to lighter tails than the actual ones. Finally, the Laplace assumption (Assumption 2) is rather innocuous – it brings about slightly fatter tails, thus partially offsetting the aforementioned error.

A complex contract

In this section, we will find the optimal risk for a contract with contributions in every period. This construction implies that later bonuses are more important than earlier ones because they “act” on a larger amount. We assume that the contribution vector is

$$\xi \triangleq (1, \exp(i), \dots, \exp(in))$$

Table 3 Optimal risk(*s*) with net contribution inflation $x=0.1$ as a percentage of the optima with $x= -\infty$ (the simple contract, c.f. Table 2)

κ	n	γ					
		0	0.5	1	2	5	10
1.1	10	99	100	101	102	106	110
1.1	30	99	100	101	103	107	112
1.1	50	98	100	101	103	107	113
1.2	10	99	100	102	104	112	121
1.2	30	98	100	102	106	116	127
1.2	50	97	100	103	107	118	129
1.3	10	98	100	103	107	118	131
1.3	30	96	100	103	110	125	141
1.3	50	95	100	104	112	128	145

The optima can be backed out using Tables 1 and 2. Based on simulation of true process.

for some fixed contribution inflation, $i \in \mathbb{R}$. Then the terminal benefit is

$$\begin{aligned}
 W(n) &= \sum_{j=0}^n \exp\left(ij + (n-j)r + \sum_{k=j+1}^n b_{M+k} \right) \\
 &\propto \sum_{j=0}^n \exp\left(xj + \sum_{k=j+1}^n b_{M+k} \right),
 \end{aligned}$$

with $x \triangleq i-r$ denoting contribution inflation net of interest. We perform the same simulations as above and calculate expected power utility of $W(n)$. For illustrative purposes, we use the admittedly high net contribution inflation $x=0.1$, but the qualitative conclusions hold for $x=0$ as well.

The optima are shown, indirectly, in Table 3. It turns out that the consequence of increasing net contribution inflation, x , is similar to the effect of reducing n : At low risk aversions (say $\gamma < 1/2$), the optimal s decreases as x increases and oppositely at $\gamma \geq 1$. The simple explanation for this is that as x increases more emphasis is put on the last bonus allotted, and thus on the marginal properties of F , whereas the serial dynamics matters less.

Finally, notice three further points about the optima. First, at low levels of risk aversion the differences between $x= -\infty$ (the simple contract) and $x=0.1$ are small. This is because for such agents it is almost exclusively the mean bonus that determines expected utility. Consequently, the analytical approximation can be applied with high accuracy to this, somewhat different, problem as well. Second, and for essentially the same reason, the more risk-averse clients' optima increase substantially – in turn making it even more “feasible” to pool individuals with different attitudes towards risk in a common

investment policy. Third, the implied loss of certainty equivalent from choosing a slightly suboptimal s is almost zero because the distribution of X is relatively much narrower when there are several contributions.

Sensitivity analysis

The optimisations above were all performed with a fixed market price of risk, $A=1/4$. In practice, it is not widely agreed what the magnitude of this quantity might be. Therefore, we shall briefly investigate how the results depend on that vital parameter. Also, pension funds may exist in regimes differing w.r.t. liquidation rules. Hence, we also examine how our results are affected by choosing a non-zero constant for c . As c can probably be observed for every entity this part of the sensitivity analysis does not relate to the insecurity in applying the results; rather to the optima's robustness towards different environments. We have fixed $n=30$, $\kappa=1.2$ in this subsection. For simplicity, we perform the sensitivity analysis for the simple contract only.

As expected the magnitude of the optimal s is very sensitive towards the return/risk-relationship of the financial market. We will not provide the full results, but with $\gamma=0$ one gets the optima 0.223, 0.287, 0.356 at $A=0.2, 0.25, 0.3$, respectively. Similarly, at $\gamma=10$ the optimal s varies almost as radically, being 0.148, 0.174, 0.199, respectively, for the market prices of risk $A=0.2, 0.25, 0.3$. Clearly, as opposed to the classical rule of Merton (1969), optimal allocation to risky assets is not exactly linear (for fixed γ) in the market price of risk, A .

Increasing c corresponds to operating closer to the boundary – ceteris paribus. Therefore, the effect of increasing c is similar to that of lowering κ . Notice, however, that both $\kappa-(1+c)$ and $(\kappa-(1+c))/\kappa$ matter, so no direct translation can be made. The numbers confirm this conjecture and are available upon request.

We have not analysed the quality of the approximation in its own respect for different parameterisations, (c,A) , but we have no reason to believe that this choice should change the characteristics of the deviations in any substantial manner. Regarding the parameter c , it is not even possible to perform our analytical approximation unless $c=0$; hence no comparison can be made.

Optimal strategies

Following the discussion initiated in Subsection “Optimisation”, it is interesting to see which strategies, if any, are unlikely to ever be optimal. Casual studies show that across extreme parameterisations; $\gamma=0$, $c=-0.2$, $\kappa=1.5$, $n=100$ and $A \leq 0.4$ one gets an optimal s less than $0.85 \cdot 2A$.

Also, for reasonable values of γ the optimal s is far above zero. But by taking γ sufficiently high, one can, of course, get an optimum as close to zero as desired.

Hence, we can at least conclude that within this model $s > 0.85 \cdot 2A$ are highly unlikely to be optimal for any client. This modest upper limit is astonishing, as it *easily* induces stationarity. In most reasonable cases we are even further from the upper limit. The reason is, of course, that the implied stationary distribution becomes unattractive as s approaches $2A$ – as discussed above.

Speed towards stationarity

Assuming that it is desirable to obtain fairness via stationarity it is conceivably also attractive to approach such invariance as quickly as possible. For if stationarity is approached too slowly, today's clients will not even approximately sample the same distribution as will future clients. And in that case, stationarity is, more or less, in vain.

Casual experiments suggest that for low initial fundings stationarity is approached quicker with high values of s (even though the implied limiting distribution is more spread out, c.f. Theorem 3), whereas for high values of F_0 choosing s low results in the faster move towards the invariant distribution. The reason for this is partly the fact that the mean and variance of the innovation to the funding ratio process from Theorem 1 are both proportional to s . And in part the characteristics of the invariant distribution of Theorem 3, which is spread out as s increases.

Naturally, there is no connection between swift approach towards a particular stationary distribution and the desirability of that outcome.

The simulations also confirm the presupposition that the stationary distribution is approximately realised quicker when $(F_{0+} - (1 + c)) < (\kappa - (1 + c))$ is not too low, nor too high. This ratio, however, does not affect the limiting convergence *rate*, which is determined by s solely.

As an example, see Figure 6 showing – at a certain parametrisation – how the stationary distribution of the funding ratio is approached rather quickly.

The rate at which a stationary Markov chain moves towards its invariant distribution can, in fact, be bounded *analytically*. To this end see for example Baxendale (2005).

Technical description of the Monte Carlo simulation

The simulations were done in the freeware statistical computing package *R*. To evaluate the hypergeometric functions we have used the package *hypergeo*. The approximated stationary distribution of F (from Theorem 3) was stratified

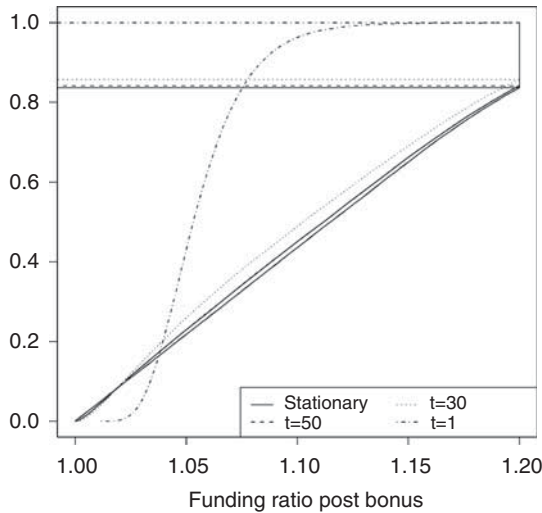


Figure 6. Funding ratio distribution at $t \in \{1, 30, 50, \infty\}$. Fixed parameters: $s=0.25$, $\kappa=1.2$, $F_{0+}=1.05$.

Notes: Horizontal lines show $\mathbb{P}(F_{t+} < \kappa) = \mathbb{P}(b_t=0)$. Based on simulation of true process.

using a trapezoidal method. To get a proxy for the true stationary distribution we did 200 Gaussian steps ahead from that stratified distribution.⁶ Then, we simulate the funding ratio process a further n periods ahead to estimate X .

Setting the seed manually allows all results to be reproduced. Throughout we did 100,000 trials. The value space for s was approximated by the set of equidistant points $\{0.001, 0.002, \dots, 2A-0.001\}$.

Concluding remarks

The bonus option

An important issue is whether or not to include the bonus option on the liability side of the balance sheet. From a strictly legal point of view one can absolutely argue in favour of excluding the option, since pension funds are

⁶ For virtually any choice of (s, κ) 200 steps seems to be more than enough – based on the distance between consecutive (pre-bonus) funding ratio distributions – to get an invariant distribution. In fact, the improvement after 20 steps is negligible for most parameter sets and 50 steps seems to be enough for almost all reasonable parameterisations. Starting from a fixed funding ratio, on the other hand, a comparable result requires far more time steps.

rarely strictly obliged by law to follow a particular bonus policy, even if it has been made public. Grosen and Jørgensen (2000) call this a *counter option* (held by the company). In addition, rules are subject to change for legal, political or strategic reasons. However, if the bonus option is disregarded, the principle of equivalence states that we must have $g=1$ for the set of pure savings contracts defined by (3). Alternatively, one could take a more pragmatic approach to accounting to justify the choice of some $g<1$ without explicitly regarding the bonus option. It would imply that new entrants contribute explicitly to the bonus reserves, even though they hold no strict statutory claim on it.

Policy implications

To apply the results one must first choose κ . This choice is a trade-off between obtaining a desirable stationary distribution (high κ) and attaining fairness quickly (low κ). After picking representative values for contract length (n), relative risk aversion (γ) and possibly net contribution inflation (x) one only needs to estimate the volatility (σ) and market price of risk (λ) of a suitable portfolio of risky assets.

The implied optima from cases with very unlike values of γ (and/or n) are not very different. Further, the implied difference in utility between quite different values of s is relatively modest which makes it possible to embrace such different attitudes towards risk (and possibly different horizons) in a common investment policy. For such an enterprise $s=\lambda$ seems to work well as a crude approximation. Alternatively, one could set up separate funds for individuals with varying appetite for risk.

References

- Baxendale, P.H. (2005) 'Renewal theory and computable convergence rates for geometrically Ergodic Markov chains', *Annals of Applied Probability* **15**(1A): 700–738.
- Black, F. and Perold, A.F. (1992) 'Theory of constant proportion portfolio insurance', *Journal of Economic Dynamics and Control* **16**: 403–426.
- Briys, E. and de Varenne, F. (1997) 'On the risk of insurance liabilities: Debunking some common pitfalls', *Journal of Risk and Insurance* **64**: 673–695.
- Døskeland, T.M. and Nordahl, H.A. (2008a) 'Intergenerational effects of guaranteed pension contracts', *The Geneva Risk and Insurance Review* **33**: 19–46.
- Døskeland, T.M. and Nordahl, H.A. (2008b) 'Optimal pension insurance design', *Journal of Banking & Finance* **32**: 382–392.
- Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn, New York: Wiley.
- Grosen, A. and Jørgensen, P.L. (2000) 'Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies', *Insurance: Mathematics and Economics* **26**: 37–57.

- Grosen, A. and Jørgensen, P.L. (2002) ‘Life insurance liabilities at market value: An analysis of insolvency risk, bonus policy, and regulatory intervention rules in a barrier option framework’, *Journal of Risk & Insurance* **69**(1): 63–91.
- Hindi, A. and Huang, C. (1993) ‘Optimal consumption and portfolio rules with durability and local substitution’, *Econometrica* **6**(1): 85–121.
- Merton, R.C. (1969) ‘Lifetime portfolio selection under uncertainty: The continuous-time case’, *Review of Economics and Statistics* **51**: 247–257.
- Nielsen, P.H. (2006) ‘Utility maximization and risk minimization in life and pension insurance’, *Finance and Stochastics* **10**: 75–97.
- Preisel, M., Jarner, S.F. and Eliassen, R. (forthcoming) ‘Investing for retirement through a with-profits pension scheme: A client’s perspective’, *Scandinavian Actuarial Journal*.
- Steffensen, M. (2004) ‘On Merton’s problem for life insurers’, *ASTIN Bulletin* **34**(1): 5–25.
- Sorensen, C. and Jensen, B.A. (2001) ‘Paying for minimum interest rate guarantees: Who should compensate who?’ *European Financial Management* **7**(2): 183–211.
- Weisstein, E.W. (2008) ‘Hypergeometric Function’, from MathWorld – A Wolfram Web Resource (<http://mathworld.wolfram.com/HypergeometricFunction.html>).

About the Author

Esben Masotti Kryger is an industrial Ph.D. student (actuarial science) at ATP (the Danish Labour Market Supplementary Pension Scheme) and the University of Copenhagen. His areas of research are pension scheme design, portfolio selection, and mortality modelling. Kryger’s previous research has been published in the *Scandinavian Actuarial Journal*.

Appendix

Proofs

Proof of Theorems 1, 2 and 3. Available upon request.

Proof of Theorem 4 To see that any moment of b_M exists let $n \geq 1$, $\varepsilon \in (0, (1 \wedge \lambda)/n)$. Then $y^\varepsilon/\varepsilon > \log y$, ($y > 0$). Hence,

$$\begin{aligned} \lim_{x \rightarrow \infty} \int (x - (1 + c))^{-\lambda-1} (\log x)^n dx &\leq \varepsilon^{-n} \lim_{x \rightarrow \infty} \int (x - (1 + c))^{-\lambda-1} x^{\varepsilon n} dx \\ &\leq \varepsilon^{-n} \lim_{x \rightarrow \infty} \int \left[(x - (1 + c))^{-\lambda-1} \right. \\ &\quad \left. ((x - (1 + c))^{\varepsilon n} + (1 + c)^{\varepsilon n}) \right] dx \\ &< \infty \text{ since } \varepsilon n - \lambda < 0. \end{aligned}$$

To derive the moments define for $q \neq 0, j \in N$ and $x > 1$

$$H_j(q, x) \triangleq \int (x - 1)^{q-1} (\log x)^j dx.$$

Then,

$$H_1(q, x) = \frac{(x - 1)^q}{q} \log x - \frac{x^q}{q^2} F_{2,1}(-q, -q, 1 - q, 1/x).$$

$$\rightarrow_{x \rightarrow \infty} 0 \text{ for } q \in (-\infty, 0).$$

$$H_2(q, x) = \frac{2x^q}{q^3} [F_{3,2}((-q, -q, -q), (1 - q, 1 - q), 1/x)$$

$$- qF_{2,1}(-q, -q, 1 - q, 1/x) \log x] + \frac{(x - 1)^q}{q} (\log x)^2$$

$$\rightarrow_{x \rightarrow \infty} 0 \text{ for } q \in (-\infty, 0),$$

where $F_{i,j}$ denotes the generalised hypergeometric function, c.f. Weisstein (2008).

$$\begin{aligned} \mathbb{E}\{b_M\} &= \int_{\kappa}^{\infty} \log(x/\kappa) dG^{(-)}(x) \\ &= \frac{\rho(\lambda - \rho)}{\lambda} (\kappa - 1)^{-\rho} \int_{\kappa}^{(\kappa-1)e^m+1} (x - 1)^{\rho-1} \log x dx \\ &\quad + \frac{\lambda \rho e^{\lambda m}}{\lambda + \rho} (\kappa - 1)^{\lambda} \int_{(\kappa-1)e^m+1}^{\infty} (x - 1)^{-\lambda-1} \log x dx \\ &\quad - \log \kappa \left(1 - G^{(-)}(\kappa)\right) \\ &= \frac{\rho(\lambda - \rho)}{\lambda} (\kappa - 1)^{-\rho} [H_1(\rho, (\kappa - 1)e^m + 1) - H_1(\rho, \kappa)] \\ &\quad - \frac{\lambda \rho e^{\lambda m}}{\lambda + \rho} (\kappa - 1)^{\lambda} H_1(-\lambda, (\kappa - 1)e^m + 1) - \frac{\rho}{\lambda} \log \kappa. \end{aligned}$$

Similarly,

$$\begin{aligned}
 \mathbb{E}\{b_M^2\} &= \int_{\kappa}^{\infty} (\log(x/\kappa))^2 dG^{(-)}(x) \\
 &= \int_{\kappa}^{\infty} \left[(\log x)^2 + (\log \kappa)^2 - 2 \log \kappa \log x \right] dG^{(-)}(x) \\
 &= \int_{\kappa}^{\infty} (\log x)^2 dG^{(-)}(x) + \frac{\rho}{\lambda} (\log \kappa)^2 \\
 &\quad - 2 \log \kappa \left[E\{b_M\} + \frac{\rho}{\lambda} \log \kappa \right] \\
 &= \frac{\rho(\lambda - \rho)}{\lambda} (\kappa - 1)^{-\rho} [H_2(\rho, (\kappa - 1)e^m + 1) - H_2(\rho, \kappa)] \\
 &\quad - \frac{\lambda \rho e^{\lambda m}}{\lambda + \rho} (\kappa - 1)^{\lambda} H_2(-\lambda, (\kappa - 1)e^m + 1) - \frac{\rho}{\lambda} (\log \kappa)^2 - 2 \log \kappa E\{b_M\}. \quad \square
 \end{aligned}$$

Proof of Theorem 5 Let (Z_i) be an i.i.d. sequence. For $i \in N$ introduce $Y_i \triangleq \log\left(\frac{\kappa-1}{F_i^+ - 1}\right)$ satisfying the recurrence $Y_{i+1} = (Y_i - Z_{i+1})^+$. Define

$$\begin{aligned}
 S_j &\triangleq - \sum_{i=1}^j Z_i, \quad (j \geq 1). \\
 P_j &\triangleq \mathbb{P}(S_j > 0), \quad (j \geq 1). \\
 \tau_j &\triangleq \mathbb{P}(S_1 \geq 0, \dots, S_{j-1} \geq 0, S_j < 0), \quad (j \geq 1). \\
 \tau(s) &\triangleq \sum_{j=1}^{\infty} \tau_j s^j, \quad (0 \leq s \leq 1).
 \end{aligned}$$

P_j gives the probability that the underlying, *unrestricted* random walk, S , is positive j periods ahead. τ_j is the probability that the *unrestricted* random walk stays positive before time j and goes negative (for the first time) at time j . Given we start at full funding, $F_{0+} = \kappa$ and thus $Y_0 = 0$, it is also the probability that bonus is awarded at time j , but not at times $1, \dots, j-1$. $\tau(\cdot)$ is the probability generating function for the (non-delayed) regeneration time of Y with density $(\tau_j)_j$. By differentiating $\tau(\cdot)$ and evaluating at zero we obtain the well-known relation

$$\tau_j j! = \tau^{(j)}(0), \quad (j \in \mathbb{N}).$$

Further, by Theorem 1 of Feller (1971, p. 413)

$$\tau(s) \triangleq 1 - \exp\left(-\sum_{j=1}^{\infty} \frac{s^j}{j} P_j\right), \quad (0 \leq s \leq 1).$$

Differentiation of this expression yields

$$\begin{aligned} \tau^{(1)}(s) &= (1 - \tau(s)) \sum_{j=1}^{\infty} s^{j-1} P_j. \\ \tau^{(2)}(s) &= -\tau^{(1)}(s) \sum_{j=1}^{\infty} s^{j-1} P_j + (1 - \tau(s)) \sum_{j=2}^{\infty} (j-1) s^{j-2} P_j. \\ \tau^{(3)}(s) &= -\tau^{(2)}(s) \sum_{j=1}^{\infty} s^{j-1} P_j - 2\tau^{(1)}(s) \sum_{j=2}^{\infty} (j-1) s^{j-2} P_j \\ &\quad + (1 - \tau(s)) \sum_{j=3}^{\infty} (j-1)(j-2) s^{j-3} P_j. \end{aligned}$$

And evaluating at 0 yields

$$\begin{aligned} \tau_1 &= P_1. \\ \tau_2 &= \frac{-P_1^2 + P_2}{2}. \\ \tau_3 &= \frac{2P_3 + P_1^3 - 3P_1P_2}{6}. \end{aligned}$$

Now consider the convolutions

$$F^{*j}(k) \triangleq \mathbb{P}(T_1 + \dots + T_j = k), \quad (k \geq j \geq 1),$$

where $\mathcal{T}_i: \Omega \rightarrow \mathbb{N}$ is the i th *non-delayed* regeneration time for Y , ($i \in \mathbb{N}_+$). The \mathcal{T}_i are i.i.d. according to $(\tau_j)_{j \geq 1}$. Hence, $F^{*j}(k)$ is the probability that the j th regeneration (and thus the j th bonus) occurs at time k . Writing the convolutions in terms of the τ_j 's gives

$$\begin{aligned} F^{*1}(k) &= \tau_k, \quad (k \geq 1). \\ F^{*j}(k) &= \sum_{i=1}^k F^{*(j-1)}(i) \tau_{k-i}, \quad (k \geq j > 1). \end{aligned}$$

In particular,

$$\begin{aligned} \sum_{j=1}^1 F^{*j}(1) &= \tau_1 = P_1. \\ \sum_{j=1}^2 F^{*j}(2) &= \tau_2 + \tau_1^2 + 0 = \frac{P_2}{2} + \frac{P_1^2}{2!}. \\ \sum_{j=1}^3 F^{*j}(3) &= \tau_3 + 2\tau_1\tau_2 + F^{*3}(3) \\ &= \tau_3 + 2\tau_1\tau_2 + \tau_1^3 \\ &= \frac{P_3}{3} + \frac{P_1^3}{3!} + \frac{P_1P_2}{2} \end{aligned}$$

Writing the “joint bonus probability” in terms of the convolutions yields

$$\begin{aligned} \mathbb{P}(b_M b_{M+i} > 0) / \mathbb{P}(b_M > 0) &= \mathbb{P}(F_{M^+} = F_{(M+i)^+} = \kappa) / \mathbb{P}(F_{M^+} = \kappa) \\ &= \mathbb{P}(F_{(M+i)^+} = \kappa \mid F_{M^+} = \kappa) \\ &= \mathbb{P}(T_1 = i) + \dots + \mathbb{P}(T_1 + \dots + T_i = i) \\ &= \sum_{j=1}^i F^{*j}(i), \quad (i \geq 1). \end{aligned}$$

Finally, use the assumption that $\forall i \geq 0: (b_{M+i} | b_M b_{M+i} > 0)$ are identically distributed. For then, since $(b_M | b_M b_{M+i})$ is independent of $(b_{M+i} | b_M b_{M+i} > 0)$ (due to regeneration at $b_M > 0$), we may calculate

$$\begin{aligned} \mathbb{E}\{b_M b_{M+i}\} &= \mathbb{P}(b_M b_{M+i} > 0) \mathbb{E}\{b_M b_{M+i} | b_M b_{M+i} > 0\} \\ &= \mathbb{P}(b_M b_{M+i} > 0) \mathbb{E}\{b_M | b_M b_{M+i} > 0\} \\ &\quad \mathbb{E}\{b_{M+i} | b_M b_{M+i} > 0\} \\ &= \mathbb{P}(b_M b_{M+i} > 0) \left(\frac{\mathbb{E}\{b_M\}}{\mathbb{P}(b_M > 0)} \right)^2 \\ &= \sum_{j=1}^i F^{*j}(i) \mathbb{E}\{b_M\}^2 \mathbb{P}(b_M > 0)^{-1}, \quad (i \geq 1). \quad \square \end{aligned}$$

Proof of Theorems 6 and 7. Available upon request.