Fairness vs. efficiency of pension schemes

Esben Masotti Kryger

April 21, 2010

Abstract

The benefits that members of with-profits pension schemes obtain are determined by the scheme design and the controlled funding level at the time of entry. This paper examines efficiency and intergenerational fairness of with-profits pension schemes.

1 Introduction

The price of a traded security reacts promptly to changes in the fundamental determinants of its value. As opposed to this, in spite of fair value accounting standards, the price of entering a with-profits pension scheme is typically fixed, regardless of changes to the financial outlooks for participation.

The manifestation of this paradox is that members contribute equally to the collective bonus reserve – even when their prospects for enjoying it are vastly different. In particular, the value of the implicit, compound bonus option that comes with membership depends substantially on the (random, yet controlled) funding ratio at the time of entry. This difference is a source for *systematic* intergenerational redistribution, which may be seen as unfair.

The aim of this study is to discuss and quantify the loss of efficiency associated with imposing bounds on the pension fund's design in order to achieve a certain degree of intergenerational fairness. Or reversely put: to analyse the loss of fairness stemming from restrictions imposed for the sake of reaching a specific level of efficiency. This trade-off should be of utmost importance to any regulator or altruistic board. In order to discuss the problem we consider a with-profits pension scheme that does take intergenerational redistribution into account, thereby constraining scheme design.

In a pension context intergenerational redistribution has – to our knowledge – been addressed mainly by Døskeland and Nordahl (2008). Their model, however, is so vastly different from the one presented below that comparison is futile. They conclude that it is unfavourable to take part in the accumulation phase of a pension fund, and vice versa. One *particular* distinction between Døskeland and Nordahl (2008) and the model of the present paper is that they consider overlapping generations explicitly whereas we deal with disjoint generations. Overlapping or contemporary generations can easily be studied within this paper's framework, however. Hansen and Miltersen (2002) also briefly discuss redistribution between different generations in the presence of a collective bonus reserve.
There is a rich literature – initiated by Briys and de Varenne (1994) – on the related problem of constructing contracts that are fair between owners and policyholders as a whole. That setup *could* be interpreted as imposing intergenerational fairness, albeit in a rather different way from what we have in mind. Also, none of those papers distinguish between the set of fair contracts (because they value under a unique equivalent martingale measure this would not make sense).

1.1 Outline

Section 2 introduces the underlying mathematical model. The measurement of fairness and efficiency is discussed in Section 3, where some optimisation criteria are subsequently suggested. These criteria are illustrated through Monte Carlo simulation in Section 4, while Section 5 considers an extension of the model, which reduces redistribution markedly. Finally, Section 6 provides a discussion of the preceding modelling and results, and gives concluding remarks.

2 Model

We consider a pension fund, which is owned by its present members. The board, which designs the scheme, represents future entrants as well, although these have no formal stakes in the scheme yet. Thus, the board can be seen as a device for solving the coordination problem that arises in any intergenerational enterprise. Such fairness motives are non-standard in the literature, but highly relevant from a practical perspective. Recently, the concept has regained popularity through the book by Akerlof and Shiller (2009).

Rather than starting from scratch the framework of Kryger (2010) is used, but as opposed to that paper the concern is with the finite time properties of the system. The model is summarised below, and Section 5 introduces various extensions that were not dealt with in previous work.

The market values of the scheme's assets and liabilities at time $t \geq 0$ are denoted A_t respectively L_t, while the *funding ratio* is derived as $F \triangleq A/L$. Between reporting periods, indexed by $0, 1, 2, \ldots$, the asset value follows the controlled process

$$A_i > 0, \quad dA_{i+t} = A_{i+t} \left((r + \pi_{i+t} \Lambda \sigma) \, dt + \pi_{i+t} \sigma dB_{i+t} \right), \; (i \in \mathbb{N}_0, \; t \in [0, 1)),$$

where r is the constant risk free interest rate, $\Lambda > 0$ the constant market price of risk, $\sigma > 0$ the constant market volatility, and π the time-varying, controlled proportion of assets allocated to risky assets. B is a one-dimensional standard Brownian motion on a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ driving the financial market, which is frictionless and complete as seen from the scheme's point of view. Individuals are assumed, however, to have limited access to the financial markets. In particular, an individual's guaranteed future benefits cannot be sold or pawned.

In order to avoid insolvency it is required that funding is strictly above one at all times. Hence, we assume that $F_0 > 1$, and that the investment strategy is Constant Proportion Portfolio Insurance (CPPI), that is

$$\pi_t = \alpha \frac{A_t - L_t}{A_t}, \; (t \geq 0),$$

where the so-called *multiplier*, $\alpha > 0$, is chosen by the board. Liabilities develop deterministically between reporting times:

$$L_i > 0, \quad L_{i+t} = L_i e^{rt} \; (i \in \mathbb{N}_0, \; t \in [0, 1)) \, .$$

Consequently, the funding ratio process follows the discrete time controlled Markov process

$$F_{0^+} > 1, \quad F_{i^-} = \left(F_{(i-1)^+} - 1 \right) \exp(Z_i) + 1, \; (i \in \mathbb{N}), \tag{1}$$

where $(Z_i)_{i\in\mathbb{N}}$ is an i.i.d. sequence with $Z_1 \sim N\left(s\Lambda - s^2/2, s^2\right)$, and $s \triangleq \alpha\sigma$ is denoted "risk" as it measures the volatility of the bonus reserve, $A - L$.
At the end of each accounting period, between times i^- and i^+, benefits fall due, new contributions are paid in, and members are awarded bonus. This brings about jumps in asset and liability values, and consequently in the funding ratio. These three types of updates are as follows:
$\Pi_i \in [0,1)$ denotes the proportion of existing liabilities paid out as benefits (e.g. expiring policies), and similarly, $\Gamma_i \geq 0$ is the amount of new contributions (e.g. new underwritings) relative to existing liabilities. These contributions are converted into liabilities $g_i\Gamma_i L_{i^-}$ for some $g_i > 0$, which is the proportion of the contribution that buys guaranteed benefits. Hence, $1-g_i$ is the share that – implicitly - buys a compound bonus option. Finally, existing liabilities, L_{i^-}, are increased by a bonus factor $\exp(b_i) \geq 1$, which is determined by using all assets in excess of κL_{i^-}, for a *bonus barrier*, $\kappa > 1$, to enhance guarantees. This barrier is also determined by the board.
With the prescribed approach one arrives at the *post bonus* funding ratio

$$\begin{aligned} F_{i^+} &= \frac{A_{i^-} + L_{i^-}(\Gamma_i - \Pi_i)}{L_{i^-}(1 + g_i\Gamma_i - \Pi_i)} \wedge \kappa \\ &= \frac{F_{i^-} + \Gamma_i - \Pi_i}{1 + g_i\Gamma_i - \Pi_i} \wedge \kappa, \ \ (i \in \mathbb{N}). \end{aligned} \tag{2}$$

The bonus, b_i, that is in fact allotted such that

$$L_{i^+} = L_{i^-}\left((1 - \Pi_i)\exp(b_i) + g_i\Gamma_i\right), \quad A_{i^+} = A_{i^-} + L_{i^-}\left(\Gamma_i - \Pi_i \exp(b_i)\right),$$

and (2) is satisfied, is

$$b_i = \left(\log \frac{\left[F_{i^-} - \Gamma_i\left(\kappa g_i - 1\right)\right]^+}{\kappa - \Pi_i(\kappa - 1)}\right)^+, \ \ (i \in \mathbb{N}). \tag{3}$$

Note that the new contributions do not earn bonus immediately, whereas existing contracts *are* credited. As bonus is, partly, intended to pay for disposable capital, this is only natural.
In this paper we consider contracts, in which members contribute the nominal amount $\xi(t) = e^{(\eta+r)t}$ at time t, for some "net contribution inflation", η. This is converted into a guaranteed benefit at *horizon* time $n \geq t$ of $g_t\xi(t)e^{r(n-t)}$ with present value $g_t\xi(t)$ – plus a compound bonus option. The object of

interest is the (to individuals) non-tradeable, discounted terminal benefit

$$
\begin{aligned}
X &\triangleq e^{-nr}\sum_{j=0}^{n} g_j \xi(j) e^{r(n-j)} e^{\sum_{k=j+1}^{n} b_k} \\
&= \sum_{j=0}^{n} g_j e^{\eta j} e^{\sum_{k=j+1}^{n} b_k}.
\end{aligned}
\tag{4}
$$

If necessary, we will equip X with arguments (s, κ, F_{0+}) representing the "risk", the bonus barrier, and the initial funding ratio respectively.

Contributions are compulsory, and there is no free policy option nor any surrender option. This leaves no scope for speculation (via timing of contributions or lapses) against the scheme, i.e. the other members.

Actual life insurance contracts give rise to interest rate risk, which is hedgeable in competitive markets, and mortality risk. Also, benefits are typically not received as a lump sum. While none of those factors are considered X can be seen as a proxy for the value of a whole life annuity bought at market terms at time n.

In order to consider intergenerational redistribution we use the rule $g_i = 1$ for all i. Section 5 explores the consequences of applying other rules. Also, we assume that inflow and outflow match exactly, that is $\Gamma_i = \Pi_i$ for all i, except in Section 4.2.1.

Administrative costs, transaction costs, taxes, etc. are disregarded throughout.

From (4) we observe that it is the release of bonus that will govern the outcome. Therefore, the properties of bonus are discussed next.

2.1 Properties of bonus

In order to analyse the scheme consider the time until next bonus, as seen from an arbitrary time $i \in \mathbb{N}$,

$$
\tau(\theta; s) \triangleq \min\{j \geq 1 : \; b_{i+j} > 0 | \, F_{i+} = (\kappa - 1)\theta + 1\}, \; (\theta \in (0,1], \; s > 0),
$$

where θ measures how far the funding ratio is from the bonus barrier. With this specification the choice of κ will not determine when bonus is awarded, c.f. (1) and (2). The continuous version of τ is the stopping time

$$
\tilde{\tau}(\theta, s) \triangleq \inf\left\{t > 0 : \; B_t \geq -\frac{\log\theta}{s} + t\,(s/2 - \Lambda)\right\}, \; (\theta \in (0,1), \; s > 0).
$$

Due to discrete time sampling of the funding ratio

$$\tau \geq \lceil \tilde{\tau} \rceil \geq \tilde{\tau},$$

but the approximation error is fairly small, when the barrier is "distant", or the investment strategy is cautious, i.e. θ or s is low (or if time is measured in "small" units).

Proposition 2.1. *The distribution function of $\tilde{\tau}(\theta, s)$ is*

$$\Phi\left(\frac{\log\theta}{s\sqrt{t}} + \sqrt{t}(\Lambda - s/2)\right) + \theta^{\frac{s-2\Lambda}{s}}\Phi\left(\frac{\log\theta}{s\sqrt{t}} - \sqrt{t}(\Lambda - s/2)\right), \ (t > 0).$$

For $s \leq 2\Lambda$ this is an inverse Gaussian distribution, but otherwise $\tilde{\tau}$ is defective.

A mere focus on the time until the first bonus allotment certainly has its shortcomings, but it is a nice way of illustrating that cautious strategies (corresponding to low values of s) are dominated on short horizons (in particular if initial funding is low), essentially because of the near-absence of downside risk. On longer horizons cautious strategies are more attractive, precisely because of downside risk. To realise this one must study $\tau(1, s)$, the time between bonus allotments, which exhibits negative first order stochastic dominance with respect to s, i.e. smaller values of s are preferable. Its distribution can be calculated exactly via the method in Jarner and Kryger (2009), and is shown in Figure 1 for four different investment strategies.

A natural supplement to the properties of bonus allotment is the (one-step) *conditional bonus* distribution:

Proposition 2.2.

$$\begin{aligned}\mathbb{P}\left(b_{i+1} \leq y \,|F_{i^+} = (\kappa - 1)\theta + 1, \ b_{i+1} > 0\right) \\ = 1 - \frac{\Phi\left(\Lambda - \frac{s}{2} + s^{-1}\log\theta - s^{-1}\log\frac{\kappa y + \kappa - 1}{\kappa - 1}\right)}{\Phi\left(\Lambda - \frac{s}{2} + s^{-1}\log\theta\right)}, \ (y > 0)\,. \quad (5)\end{aligned}$$

The conditional distribution in (5) exhibits first order stochastic dominance in s (and κ and θ), so that higher values are preferable.

Altogether, cautious investment strategies do not give rise to much bonus on short horizons, especially if initial funding is low, whereas on longer horizons the matter is more ambiguous – but with both very cautious and very aggressive strategies inducing only little bonus. As for the barrier – when initial funding is $(\kappa - 1)\theta + 1$, higher barriers are always preferable. The board could, however, encounter a fixed initial funding, and be asked to set a barrier subsequently, which would complicate matters. This is because, for short horizons, more bonus would be given with low barriers, in particular if initial funding is low. But as the horizon increases higher barriers again become more attractive.

For both design parameters one should also have in mind, that in case of several contributions ($\eta > -\infty$) the final wealth distribution depends more on later bonuses than on early ones, c.f (4). Therefore, the long-run properties are more important than this discussion perhaps suggests.

3 Fairness and efficiency

In this section the measurement of fairness and efficiency will be discussed. Subsequently, some tangible measures of these two vague notions are introduced.

3.1 Preliminary reflections

Since we consider a compulsory scheme the design must satisfy (this approach is inspired by David McCarthy) that

> almost every outcome is acceptable for almost every one.

Hence, it is required that – on the vast majority of paths and for almost all types of members – the degree of redistribution between different generations is low, while the outcome must at the same time be satisfactory for almost everyone. The precise materialisation of these notions will be made clear below.

To further approach a decision rule we rely on the *original position* of moral philosophy (see Rawls (1971)), which loosely states that any design agreed upon by agents whose identity is unknown to themselves during the bargaining is fair.

It is clear that not all generations can get the same outcome. And as hinted above it is probably not desirable to follow a very cautious investment strategy with the aim of approximating such equality. The question is then, how the board should evaluate the inherent redistribution against a possible advantage from investing more aggressively. Although such trade-offs are *acknowledged* in most economic analysis, they are typically disregarded.

In order to pick fairness and efficiency measures we consider a hypothetical bargaining between two members entering the scheme at funding ratios κ and $f \in (1, \kappa]$ respectively, but with the caveat that they must design the system without knowing who enters at which funding. f may depend on the design parameters (s, κ), as will be explained below, but the notation f is used as a shorthand nevertheless, since the meaning will always be clear from the context. The generations represented by the two members are taken to experience identical institutional conditions during their membership periods, but their financial markets are assumed to be governed by independent versions of B, and f is taken to be independent of those. This means that a proper, although not imperative, interpretation of the setup is that the member with funding f enters first, and then after at least n years the other member enters at a time with full funding.

As for the actual measurement, the initial *dogma* of this section and the idea of the original position guides us. Efficiency of the outcome should be associated with the distribution of the terminal benefit, which must overcome some minimum target with a high probability. Fairness ought to be related to the ratio of the terminal benefits, which we want to be close to one with a high probability. Non-overlapping generations are compared, and the ratio of the benefits of two such disjoint generations has a wider distribution than in the case of overlapping generations. Thus, in this respect the discussion in this paper is "worst case" in terms of intergenerational subsidisation.

The two most widely applied measures for evaluating a monetary random variable is measuring its arbitrage-free value or its expected utility. We discard the former approach because of the assumed non-tradeability of the guarantee, and we discard the latter because we prefer to ensure attractive outcomes with a high probability.

3.2 Two formulations of the problem

Regarding the choice of initial fundings to compare, it is uncontroversial to use κ as the higher level. As for the lower value one may choose to consider a constant. In this case it is meaningful to compare different barriers, for – as discussed above – there is a trade-off between high and low barriers.
Alternatively, we could let $f = 1 + (\kappa - 1)\theta$ for some $\theta \in (0, 1)$ as in Section 2.1, in which case higher barriers are more attractive for both parties, so that it is not possible to optimise over κ. Still, to properly differentiate between candidates for the optimal investment strategy we let θ depend on s, since a high distance from the barrier is more likely with more aggressive strategies. To this end, fix an $\epsilon \in (0, 1)$, and choose $\theta(s)$ such that $\mathbb{P}\left(1 + (\kappa - 1)\theta(s)\right) = \epsilon$ in stationarity. From Preisel *et al.* (2010) we then get

$$\theta(s) = \left(\frac{\epsilon}{1 - s\rho/\sqrt{2}}\right)^{\rho^{-1}}, \ (0 < s < 2\Lambda),$$

where ρ is the unique non-zero solution to

$$1 - \rho^2 s^2/2 = \exp\left(-\rho s(\Lambda - s/2)\right), \ (0 < s < 2\Lambda).$$

For $s \geq 2\Lambda$ no stationary distribution exists, therefore we truncate θ at the value corresponding to the somewhat arbitrary $s = 1.99\Lambda$.
The former setting corresponds to an existing scheme encountering a (low) funding ratio, and wishing to design a fair and efficient scheme going forward. On the other hand, the latter formulation covers the case of a new scheme with all the good intentions at the outset, but with an exogenously fixed κ. We refer to the two settings as "case A" and "case B" respectively.

3.3 Measuring fairness and efficiency

Sections 3.3.1 and 3.3.2 present our choices for measuring fairness and efficiency respectively. They are combined to form two different constrained optimisation problems in Section 3.3.3.

3.3.1 Fairness

To measure intergenerational fairness we focus on a threshold for the ratio of the respective terminal benefits:

$$\mathbb{P}\left(\frac{X(s,\kappa,f)}{X(s,\kappa,\kappa)} > 1-\delta\right), \tag{6}$$

where $0 \leq \delta < 1$ measures the maximum permitted redistribution (up to some probability). If the generations were contemporary the ratio in question would be bounded by one. Although studying disjoint generations we use the measure nevertheless, and thus disregard the extent to which the ratio exceeds one.

3.3.2 Efficiency

The quantification of efficiency follows similar lines as above. As previously hinted, a target is needed to calculate efficiency. To this end expected power utility is used as a measure of satisfaction. Since the target will only be used as an auxiliary we prefer this simple approach, because it is easy to communicate, and requires one parameter only. The certainty equivalent of a positive random variable, Y, is then

$$CE(\gamma;Y) \triangleq \begin{cases} \mathbb{E}\left\{Y^{1-\gamma}\right\}^{\frac{1}{1-\gamma}}, & \gamma \in [0,\infty)\backslash\{1\} \\ \exp\left(\mathbb{E}\left\{\log Y\right\}\right), \ \gamma = 1 \end{cases}, \tag{7}$$

where $\gamma \geq 0$ denotes the coefficient of relative risk aversion. The suggested efficiency measure for a generation with funding $f < \kappa$ is

$$\mathbb{P}\left(\frac{X(s,\kappa,f)}{\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}} \ CE\left(\gamma; X(\bar{s},\bar{\kappa},f)\right)} > 1-\beta\right). \tag{8}$$

For the two generations efficiency is measured by merely adding their respective terminal benefits to a single random variable. Due to linearity of the certainty equivalent, $0 \leq \beta < 1$, can be interpreted as the maximum permitted relative *cost* of obtaining fairness (up to some probability). $\mathcal{S} \times \mathcal{K} \subseteq (0,\infty) \times (1,\infty)$ is a range of considered design variables.

3.3.3 Constrained optimisation

Combining the criteria (6) and (8) produces two different constrained optimisation problems. First, if we maximise efficiency subject to a fairness side condition we get:

$$\max_{s,\kappa\in\mathcal{S}\times\mathcal{K}} \mathbb{P}\left(\frac{X\left(s,\kappa,\kappa\right)+X\left(s,\kappa,f\right)}{\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}}\ CE\left(\gamma;X\left(\bar{s},\bar{\kappa},\bar{\kappa}\right)+X\left(\bar{s},\bar{\kappa},f\right)\right)}>1-\beta\right) \tag{9a}$$

subject to

$$\mathbb{P}\left(\frac{X\left(s,\kappa,f\right)}{X\left(s,\kappa,\kappa\right)}>1-\delta\right)\geq p, \tag{9b}$$

and subject to Pareto optimality:

$\forall(\tilde{s},\tilde{\kappa})\in(\mathcal{S}\times\mathcal{K})\backslash\{(s,\kappa)\}$: (9c) or (9d) satisfied.

$$\frac{\mathbb{P}\left(X\left(s,\kappa,f\right)>(1-\beta)\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}}\ CE\left(\gamma;X\left(\bar{s},\bar{\kappa},f\right)\right)\right)}{\mathbb{P}\left(X\left(\tilde{s},\tilde{\kappa},f\right)>(1-\beta)\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}}\ CE\left(\gamma;X\left(\bar{s},\bar{\kappa},f\right)\right)\right)}\geq 1 \tag{9c}$$

$$\frac{\mathbb{P}\left(X\left(s,\kappa,\kappa\right)>(1-\beta)\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}}\ CE\left(\gamma;X\left(\bar{s},\bar{\kappa},\bar{\kappa}\right)\right)\right)}{\mathbb{P}\left(X\left(\tilde{s},\tilde{\kappa},\kappa\right)>(1-\beta)\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}}\ CE\left(\gamma;X\left(\bar{s},\bar{\kappa},\bar{\kappa}\right)\right)\right)}\geq 1. \tag{9d}$$

A natural consequence of having individual measures of efficiency is to consider only the Pareto optimal candidates.

The reverse constrained optimisation is the one where a fairness criterion is maximised subject to an efficiency threshold condition:

$$\max_{s,\kappa\in\mathcal{S}\times\mathcal{K}} \mathbb{P}\left(\frac{X\left(s,\kappa,f\right)}{X\left(s,\kappa,\kappa\right)}>1-\delta\right). \tag{10a}$$

subject to

$$\mathbb{P}\left(\frac{X\left(s,\kappa,\kappa\right)+X\left(s,\kappa,f\right)}{\max_{\bar{s},\bar{\kappa}\in\mathcal{S}\times\mathcal{K}}\ CE\left(\gamma;X\left(\bar{s},\bar{\kappa},\bar{\kappa}\right)+X\left(\bar{s},\bar{\kappa},f\right)\right)}>1-\beta\right)\geq p. \tag{10b}$$

In (10) the inclusion of (9c)-(9d) is not imperative, nor meaningful. In (9b) and (10b) $p\in[0,1]$ is a minimum acceptance probability decided along with β and δ. Before proceeding we warn that for some parameterisations one might end up with probabilities of zero or one, in which case the clever approach is to re-parameterise, unless it was intentional.

Next, we illustrate the suggested criteria through simulations.

4 Simulation-based illustrations

In this section we mainly analyse a "mature" fund with equal in- and outflow of $\Pi=\Gamma=0.02$, net contribution inflation $\eta=0.02$, a horizon of $n=50$, and market price of risk, $\Lambda=0.25$. Following the analysis of the base case each of the main parameters (Π, Γ, η, n, and Λ) are changed, and the derived consequences are briefly discussed. Also, the auxiliary parameters are fixed at $\beta = 0.05$, $\delta = 0.05$, $\gamma = 0.5$, $f = 1.02$ (case A), $\epsilon = 0.05$ (case B), and $\kappa = 1.3$ (case B), but a sensitivity analysis is conducted in Section 4.3. Finally, Section 4.4 reviews the simulation details.

4.1 Base case

Figures 2 and 3 show the trade-off between fairness and efficiency. In the former graph the bonus barrier and the investment strategy are both to be optimised over (case A), whereas the latter considers a pre-specified bonus barrier (case B). The results are qualitatively in line with the predictions of Section 2.1.

In case A, higher barriers yield less fairness (because generations are more different), but more efficiency. Also, the optimal strategies associated with higher barrier are more cautious. If one uses the maximum-efficiency criterion (9), a very narrow range of (for this parametrisation) modestly aggressive strategies are non-dominated. The most cautious as well as the most aggressive investment strategies are excluded by Pareto inoptimality, while others are merely dominated. When the maximum-fairness criterion (10) is imposed instead, all investment strategies *above* some threshold are candidates for optimum, because redistribution is less for more aggressive strategies (bonus becomes rare).

Conversely, in case B, the heritage to future generations is implicitly considered (through the long-run funding level), which leads to less aggressive strategies being favourable. With a main focus on efficiency, through criterion (9), a range of rather cautious investment strategies are non-dominated, as in case A, though these values of s are generally much lower in case B. If fairness is emphasised instead, all strategies *below* a certain threshold are potentially optimal, which highlights the difference between the two cases. In case B, even modestly aggressive investment strategies are penalised severely by a higher initial distance from the barrier – in turn making them poor candidates for yielding sufficient efficiency.

Altogether, the results show that it is hard to obtain outcomes that are adequately efficient and fair to a high degree at the same time. One way of overcoming this is to alter the way in which contributions are transformed into benefits, which is the topic for Section 5.

4.2 Alternative environments

4.2.1 Demography

When the net outflow is positive ($\Pi > \Gamma$) bonuses are higher and more frequent than otherwise. Oppositely, of course, in a fund that is in the process of building up its balance. The aggregate effect on fairness and efficiency is not clear at the outset. The situation with $\Pi = 0.1$, and $\Gamma = 0.02$ is shown in Figure 4, which demonstrates that intergenerational subsidisation is slightly less in such a non-accumulating scheme, without harming efficiency. In general, the trade-off governing the design decision is very similar to the base case in Figures 2 and 3.

4.2.2 Economy

As net contribution inflation, η, is increased, a higher degree of fairness is obtained, because – as previously mentioned – later bonus matters more, and later bonus is more likely to be the same for both generations. Also, all strategies become more efficient because the terminal distribution is narrower, i.e. less dependent on bonus.

When the market price, Λ, increases, bonuses are larger and more frequent. More aggressive investment strategies are, of course, preferable.

The qualitative *and quantitative* conclusions from the base case are surprisingly insensible to changes to η and Λ. The only noteworthy effect is that higher contribution inflation implies slightly more cautious investment in case A because of the amplified importance of later bonuses.

4.2.3 Horizon

Extending the horizon, n, makes the system fairer, due to the longer period with identically distributed bonuses, $(n - \tau)^+$. The shapes of the curves in Figures 2 and 3 are essentially unaltered, however. Only, in case A, slightly more cautious investment strategies are preferred on longer horizons, as the long-run properties become more important. This point emphasises that when sampling from a *fixed* (not affected by the controls) initial distribution (in this case a Dirac distribution), and the terminal conditions do not matter there is a need for someone, be it the board or the regulator, to require long-run stability, for otherwise it is tempting to gradually – or swiftly – exhaust the bonus reserve to the disadvantage of future generations.

4.3 Sensitivity

The choice of β, γ, δ, in case A: f, and in case B: ϵ and κ matters little as far as the qualitative conclusion goes.

As for the former two, the reason that the parameters play minor roles is that the benefit distribution is quite narrow (on most trajectories bonus is small compared to contributions), and bounded (far) away from zero. Of course, the higher the values of β and γ, the higher the efficiency. Therefore it is instructive to use a rather low γ-value, since this ensures that the efficiency probabilities are not too high for any agent.

The choice of δ affects the *level* of fairness profoundly – but the *shapes* of the curves in Figures 2 and 3 are unaltered.

In case A higher f implies more fairness and a shift towards slightly more cautious investment.

The tail probability, ϵ, was introduced to allow a design-dependent distinction between the two generations, and the extent to which this differentiation is carried out does not affect the results much as long as ϵ is kept reasonably small.

Also, by construction, the choice of κ, in case B, only influences the conditional bonus, and only so to an extent which hardly affects the optimal choice of s.

4.4 Simulation details

The simulations were done in the freeware statistical computing package `R`. In all cases 100,000 paths were simulated. The seed was set manually to allow all results to be reproduced, and reused across experiments. It is particularly important to ensure that disjoint generations experience independent market innovations. Oppositely, if contemporary generations were considered, it would be equally important that they sampled the same financial market. Distribution functions are estimated by their sample counterparts.

5 Extensions

The previous section showed that although some designs are superior to others it is not possible to obtain high levels of fairness and efficiency at the same time, when the entire contribution is transformed into guaranteed future benefits, i.e. $g = 1$. It is straightforward, however, to design rules that take the differing conditions into account, and hence – partly – overcome systematic intergenerational redistribution. Below we present two such rules.

5.1 A solidary rule

When a policy expires its accrued guarantees (including bonus) are paid out, but the free reserve stays in the fund. Therefore, ceteris paribus, the funding ratio increases as a result of an expiry. This gain is split between existing members and new contributions according to some rule. In the standard case, $g = 1$, new money always benefits from entering in the sense that the value of their guarantee can be approximated by $g_i F_{i+} 1$, which is greater than one, but highly dependent on the random timing of entry

Instead, the way in which contributions are transformed into guaranteed future benefits *may* be based on the solidary point of view that $g_i F_{i+}$ should be the same for all generations, that is

$$g_i \left(\frac{F_{i-} + \Gamma_i - \Pi_i}{1 + g_i \Gamma_i - \Pi_i} \wedge \kappa \right) = C_i, \tag{11}$$

for some positive $C_i < \frac{1+\Gamma_i - \Pi_i}{\Gamma_i}$, which - for the natural choice $C_i = 1$ - is solved by the *solidary rule*

$$g_i = \frac{1 - \Pi_i}{F_{i-} - \Pi_i} \vee \kappa^{-1}, \ (i \geq 1) \tag{12a}$$

$$g_0 = \frac{1}{F_{0+}} \tag{12b}$$

In general C_i could depend on e.g. demographics forecasts and time to expiry.
Because $g < 1$ the funding ratio gets a boost upwards, so that bonuses are larger and more frequent (in return for lower guarantees). As in the case of positive net inflow the combined effect on fairness is ambiguous, depending on the horizon, among others. For the base case it turns out that fairness is enhanced slightly. The fairness measure (6) is ill-suited, however, since it focuses exclusively on one-sided deviations. As a matter of fact the generations can be made approximately equally well off when using the solidary rule (in the sense that the density of the ratio between the benefits is much more balanced around 1), which is a major advance over the results obtained with $g = 1$.

5.2 An indemnifying rule

As another example we present the *indemnifying rule*

$$\frac{F_{i^-} + \Gamma_i - \Pi_i}{1 + g_i\Gamma_i - \Pi_i} \wedge \kappa = F_{i^-} \wedge \kappa,$$

which gives staying members the same funding ratio regardless of the amount of new entrants and exits. This rule yields

$$g_i\Gamma_i = \Pi_i + \frac{\Gamma_i - \Pi_i}{F_{i^-}}, \ (i \geq 1) \tag{13a}$$

$$g_0\Gamma_0 = \Pi_0 + \frac{\Gamma_0 - \Pi_0}{F_{0^-}}, \tag{13b}$$

the latter assuming F_{0^-} is known. The indemnifying rule enhances fairness still more the higher the ratio of inflow to outflow, precisely because an accumulating scheme releases relatively little bonus reserve, and thus the higher the pre-bonus reserve the less the new entrants will receive (in terms of g). The *neutralising* effect of the solidary rule is only achieved if there is no outflow, however – in which case the two rules are almost identical.

6 Discussion and conclusion

This section discusses the insights gained from the paper's model. Also, limitations and possible alternatives approaches are described. Policy implications are touched upon, and finally, concluding remarks are given.

One of the main lessons that can be derived from the paper stems from the vast difference between cases A and B. In the former situation both investment strategy and bonus barrier are optimisation variables, and the trade off is that higher barriers induce more efficient systems that are less fair. Conversely, in case B higher barriers are always preferable. If efficiency is the maximisation object and fairness the side condition there is a narrow range of non-dominated strategies in either case (though those ranges are substantially different the cases between). But if fairness is maximised subject to an efficiency constraint the two cases differ markedly. In case A cautious strategies are dominated, but in case B aggressive strategies are dominated because they imply low initial funding and thus low fairness. These qualitative conclusions are stable under different parameterisations, but are likely sensitive to different *formulations* of the objective.

Another important insight comes from realising how fairness can be improved substantially upon by introducing new ways of transforming contributions into guaranteed future benefits.

A third outcome is the ability to exclude certain investment strategies based on dominance arguments.

6.1 Limitations and alternatives

Obviously the evaluation of intergenerational fairness is much broader than what can be covered here. For instance, one could argue that by facing identical *rules* generations are treated fairly in that the economic conditions they face are not *explicitly* controlled by other generations. Also, even within the present paradigm, redistribution and efficiency could be measured quite differently.

The presented rules do not aim at converting each contribution into guarantees in a fair manner. That is, the *value* of the compound bonus option does not equal its implicit *price*, as the former depends on time to maturity and forecasts of demographics etc. Instead, it is assumed that all members have identical contribution plans, so there is no heterogeneity nor any free policy or surrender options.

Default is precluded by construction in the present setting. To overcome this weakness one could allow for default by fixing the portfolio only at the beginning of each period. This would mimic real life investment behaviour more closely than the often employed constant allocation to stocks, while still allowing for bankruptcy. Within the realm of no-default one could change the asset allocation in some non-linear way, while maintaining $\lim_{F_t \to 1^+} \pi(F_t) = 0^+$, e.g. through an Option Based Portfolio Insurance (OBPI) strategy.

Another way of introducing default is to allow for non-marketed shocks to the value of liabilities – interpreted as unanticipated changes in mortality, statute, or the like. Such jumps could occur periodically or at random points in time.

Instead of distributing all excess funding as bonus, some authors argue in favour of smoothing bonus allotting over time precisely with the aim of reducing the effect of random entry time funding levels. This would reduce subsidisation slightly. Another widespread alternative consists of basing the bonus on the past year's financial performance exclusively, which reduces intergenerational redistribution, but enhances solvency problems (if the members do not participate in the downside).

Finally, as previously mentioned one could consider perfectly contemporary generations sampling the same market. Then the interpretation of fairness would be somewhat different, namely related to joining schemes with different initial funding, but which are operated identically. This appears somewhat less interesting from a designer's point of view, but can be very useful in other settings.

6.2 Policy implications

A regulator overseeing scheme design, or an altruistic designer should discuss the weighing between short and long term objectives as well as the trade-off between fairness and efficiency. In order to use the analysis in the framework laid out here to make an informed decision they must also choose whether case A or B is more appropriate for their purpose.

The most important recommendation stems from noting how fairness can be enhanced greatly at no cost by following the suggestions in Section 5.

6.3 Concluding remarks

This paper discusses fair and efficient design of with-profits pension schemes. More specifically, strategies for investment and bonus allotment are treated. As in many other problems in social science an important, but often neglected, feature of this problem is the crucial choice of measure for the intangible quantities fairness and efficiency. We have suggested a set of criteria and sketched the characteristics of an optimal design in two situations. First, one where only the present generation is considered, and second the case where the long-run properties (i.e. the heritage to future generations) are implicitly taken into account. It turns out that the optima are very different – quite precisely representing the different approaches. Finally, as a consequence of the somewhat dismal results of that analysis, different ways of converting contributions into guarantees are suggested and shown to yield a substantial improvement.

A Figures

Figure 1: The distribution of the non-delayed regeneration time $\tau\left(1;s\right)$. Fixed parameters: $g=1$, $\Gamma-\Pi=0$, and $\Lambda=0.25$.

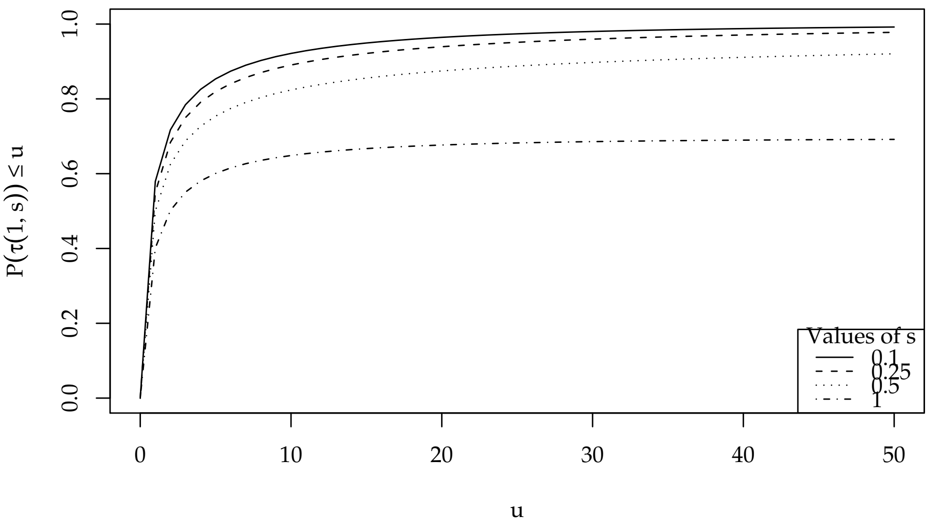


Figure 2: Case A: Trade-off between efficiency (8) and fairness (6) at different values of s indicated in the diagram. The left curve corresponds to $\kappa = 1.5$, and the right curve represents $\kappa = 1.1$. The dashed parts of the graphs correspond to strategies that were discarded due to (9c)-(9d). Fixed parameters: $g = 1$, $\Gamma = \Pi = 0.02$, $\eta = 0.02$, $n = 50$, $\Lambda = 0.25$, $\gamma = 0.5$, $\beta = \delta = 0.05$, $\mathcal{S} \subseteq (0, 1]$, $\mathcal{K} = \{1.1, 1.5\}$, and $f = 1.02$.

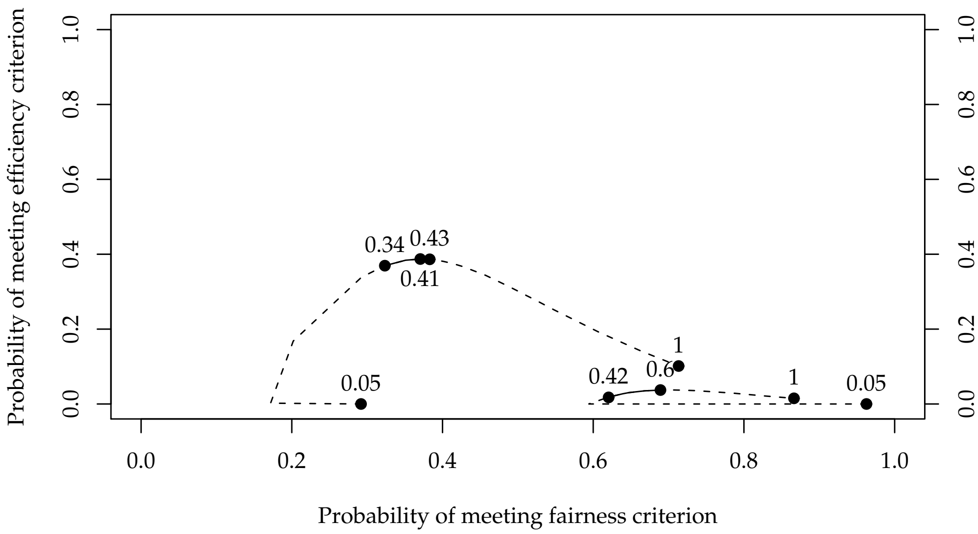


Figure 3: Case B: Trade-off between efficiency (8) and fairness (6) at different values of s indicated in the diagram. The dashed parts of the graph corresponds to strategies that were discarded due to (9c)-(9d). Fixed parameters: $\kappa = 1.3$, $g = 1$, $\Gamma = \Pi = 0.02$, $\eta = 0.02$, $n = 50$, $\Lambda = 0.25$, $\gamma = 0.5$, $\beta = \delta = 0.05$, $\mathcal{S} \subseteq (0, 1]$, $\mathcal{K} = \{\kappa\}$, and $\epsilon = 0.05$.

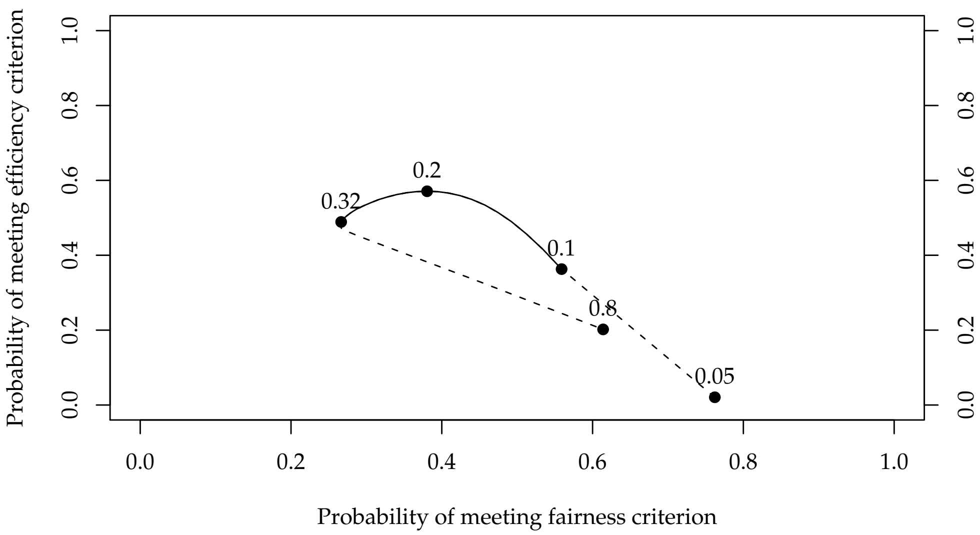
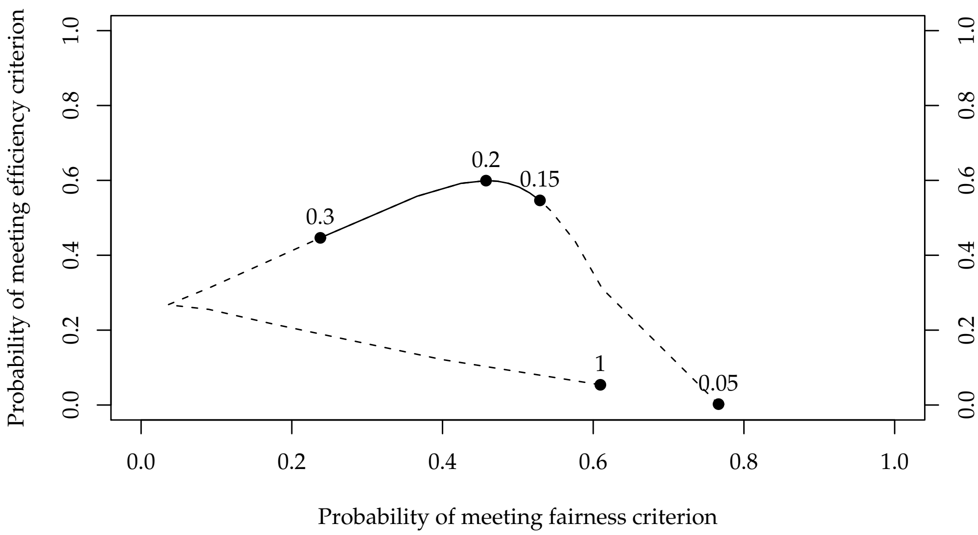


Figure 4: Case B: Trade-off between efficiency (8) and fairness (6) at different values of s indicated in the diagram. The dashed parts of the graph corresponds to strategies that were discarded due to (9c)-(9d). Fixed parameters: $\kappa = 1.3$, $g = 1$, $\Gamma = 0.02$, $\Pi = 0.1$, $\eta = 0.02$, $n = 50$, $\Lambda = 0.25$, $\gamma = 0.5$, $\beta = \delta = 0.05$, $\mathcal{S} \subseteq (0, 1]$, $\mathcal{K} = \{\kappa\}$, and $\epsilon = 0.05$.



References

Akerlof, G. A. and Shiller, R. J. (2009). *Animal Spirits.* Princeton University Press, New Jersey.

Briys, E. and de Varenne, F. (1994). Life Insurance in a Contingent Claim Framework: Pricing and Regulatory Implicatons. *The Geneva Papers on Risk and Insurance Theory* **19**, 53–72.

Døskeland, T. M. and Nordahl, H. A. (2008). Intergenerational Effects of Guaranteed Pension Contracts. *The Geneva Risk and Insurance Review* **33**, 19–46.

Hansen, M. and Miltersen, K. R. (2002). Minimum Rate of Return Guarantees: The Danish Case. *Scandinavian Actuarial Journal*, 280–318.

Jarner, S. F. and Kryger, E. M. (2009). Computation of exact hitting time probabilities of random walks. Non-published manuscript available on request.

Kryger, E. M. (2010). Optimal Pension Fund Design Under Long-Term Fairness Constraints. *The Geneva Risk and Insurance Review, forthcoming.*

Preisel, M., Jarner, S. F., and Eliasen, R. (2010). Investing for Retirement Through a With-Profits Pension Scheme: A Client's Perspective. *Scandinavian Actuarial Journal*, 15–35.

Rawls, J. (1971). *A Theory of Justice.* Cambridge, Massachusetts.