Quantifying the trade-off between income stability and the number of members in a pooled annuity fund

Thomas Bernhardt[*] and Catherine Donnelly[†]

September 11, 2020

Abstract

The number of people who receive a stable income for life from a closed pooled annuity fund is studied. Income stability is defined as keeping the income within a specified tolerance of the initial income in a fixed proportion of future scenarios. The focus is on quantifying the effect of the number of members, which drives the level of idiosyncratic longevity risk in the fund, on the income stability. To do this, investment returns are held constant and systematic longevity risk is omitted. An analytical expression that closely approximates the number of fund members who receive a stable income is derived and is seen to be independent of the mortality model. An application of the result is to calculate the length of time for which the pooled annuity fund can provide the desired level of income stability.

Keywords: Longevity credit; decumulation; tontine; unsystematic risk; pooling.

1. Introduction

The lack of innovative products to convert a lump-sum into a life-long retirement income is increasingly seen as a problem in the United Kingdom (Work and Pensions Committee, 2018, Paragraph 12, page 10). Previously, pension savers typically either bought a life annuity from an insurance company or saved within a defined benefit pension scheme. With the former option rather unattractive for those with smaller pension pots due to low annuity rates (caused by low investment returns, uncertainty about future lifespans and solvency capital requirements) and the latter becoming less available due to the closing of defined benefit pension schemes, there is an opportunity for attractive, alternative ways to convert savings into an income for life.

Pooled annuity funds, a type of tontine, are an alternative which offer the opportunity of a lifelong, stable income to its participants. They pool the retirement wealth of their participants and pay it out to the survivors as a regular income. They should give a higher income than under

[*]T.Bernhardt@hw.ac.uk. Department of Mathematics, University of Michigan, Ann Arbor, Michigan MI 48109-1043, USA.

[†]C.Donnelly@hw.ac.uk (corresponding author). Risk Insight Lab, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, Scotland EH14 4AS.

self-insurance (also called income drawdown in UK terminology), for the same chance of exhausting funds. They can do this because they pool the participants' longevity risk, and hence subsidize the payments to the longer-lived members, and they do not pay a bequest to the participants' beneficiaries. It is the same principle underpinning life annuities. The critical difference is that the income from a conventional life annuity product is guaranteed by an insurance company whereas that from a pooled annuity fund is not guaranteed by anyone. The income from a pooled annuity fund will vary due to fluctuations in investment returns and longevity experience, unlike those of a conventional life annuity, which gives a constant income regardless of investment and longevity experience.

In this paper a mathematical expression is derived which allows the calculation of a lower bound on the number of participants who receive a stable, life-long income once the degree of income volatility is specified. A numerical study suggests that the lower bound is close in value to the true value. By keeping investment returns constant and not including systematic longevity risk (the risk that the wrong distribution of future lifetime has been chosen), the risk attributable to idiosyncratic longevity risk (the risk that the distribution of future lifetime is not observed perfectly among the annuitants) is studied in isolation. This is important since the *raison d'être* of joining the fund for individual participants - a stable income paid for life - is reliant on idiosyncratic longevity risk being sufficiently diversified. The analysis in this paper will allow the risk inherent in random investment returns and systematic longevity risk to be quantified and attributed separately in further research. Importantly, the derived results give us the insight that the proportion of people who receive a stable, life-long income is approximately independent of the mortality model, once the number of members initially in the fund is fixed.

Tontines have enjoyed recently some attention in the academic literature. Piggott et al. (2005) proposed Group Self-Annuitization, which is the structure studied here. In it, the income paid to the annuitants is calculated from their individual fund values and annuity factors. The income calculated for each annuitant should be payable for life as long as both the investment and mortality experience turn out as assumed in the annuity factors. Group Self-Annuitization is similar to the Pooled Annuity Fund introduced in Stamos (2008), although the latter does not specify how income is withdrawn from their structure, but lets income be withdrawn as the annuitant desires. These structures are designed to operate when all annuitants are independent and identical copies of each other. The same membership assumption is made in this paper; indeed, the pooled annuity fund studied here is that of Piggott et al. (2005) in the homogeneous case. An extension of Group Self-Annuitization to heterogeneous groups is given in Valdez et al. (2006).

Alternative ways of sharing out the funds in a tontine are proposed by Donnelly et al. (2014) and Sabin (2010), with the latter further discussed in Forman and Sabin (2015). Their tontines pay survivors an explicit longevity credit, which represents a share of the funds of the recently deceased, and the level of income withdrawn is not mandated. In contrast, tontines like those of Milevsky and Salisbury (2015) use the level of income withdrawn to implicitly distribute the funds of the deceased among the survivors. Chen et al. (2019) proposes and analyzes a product that is comprised of a tontine until a fixed age and a deferred life annuity that comes into payment at the fixed age. Donnelly and Young (2017) propose a tontine product with a minimum guaranteed payment, and the proposed product is developed and examined further in Chen and Rach (2019).

The question of actuarial fairness - that the expected value of the benefits received should equal the value of each annuitant's fund - can arise in a discussion of tontine structures. Some are actuarially fair only for homogeneous memberships, such as the Group Self-Annuitization and the Pooled Annuity Fund (as shown in Donnelly 2015). Others are actuarially fair for heterogeneous memberships, such as those in Donnelly et al. (2014) and Sabin (2010). Milevsky and Salisbury

(2016) argue instead for actuarial equitability rather than fairness - everyone gets the same expected value of benefit payments even if it is less than what they put in - and propose a structure to do this.

Qiao and Sherris (2013) analyze and study how to manage systematic longevity risk in Group Self-Annuitization structures. Noting that annuity payments decline at older ages if the annuity factors used to calculate the income withdrawn are based on up-to-date mortality experience, they suggest a way of sharing systematic longevity risk across cohorts, through an adjustment to the calculation of Piggott et al. (2005). They do a numerical study of this, calculating the quantiles of the income distribution at fixed time points.

However, our aim is to study idiosyncratic longevity risk. The mathematical results presented hold for any mortality model as long as systematic longevity risk is excluded. The evolution of each sample path of the income is analysed, which permits a precise quantification of the behaviour of the income streams. Piggott et al. (2005) and Sabin (2010) both consider a particular mortality model and plot the income amounts, observing that they are more volatile at older ages. Their conclusions are qualitative and not quantitative. None of these studies defines income stability and so cannot study it rigorously and generally, as is done here. While one definition of income stability is proposed in this paper, there are many other possible definitions which may be more suitable for different needs.

After the operation of the pooled annuity fund is detailed in Section 2, the results in Section 3.1 enable the calculation of the number of people who can receive a life-long, stable income from the pooled annuity fund. In effect, this is the number of people who die while the desired level of income is stable. Our results give a lower bound which can be used as an approximation (Section 3.2) to the true number of people. Importantly, the approximation is independent of the mortality model and thus the numbers calculated hold for any age, sex or other sub-population, once the fund parameters are fixed.

Determining for how long a pooled annuity fund can pay a stable income to its members from the theoretical results is studied in Section 4. Finally, an approximation to the lower bound is presented in Section 5, with the derivation relegated to the Appendix.

2. The operation of the pooled annuity fund

Consider a group of $N \geq 2$ individuals who constitute the entire membership of the pooled annuity fund at time 0. Each member is an independent and identical copy of the rest. No member joins after time 0 and no member can leave except through death.

2.1. Future lifetime random variables and survival probabilities

The members are each aged $x \geq 0$ at time 0. Represent the ith member's future lifetime from age x by the real-valued random variable $T_i > 0$, for $i = 1, \ldots, N$. The random variables $T_1, T_2, \ldots, T_N$ are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, are independent and identically distributed and have a continuous distribution. For each $\omega \in \Omega$, the order statistics $(T_{(i)}(\omega))_{i=1}^N$ are the increasingly ordered $(T_i(\omega))_{i=1}^N$, i.e. $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(N)}$.

The number of individuals observed to be alive at age $x + t$ is

$$L_{x+t} = \sum_{i=1}^N \mathbb{1}_{[T_i > t]}, \quad \text{for } t \geq 0,$$

in which $\mathbb{1}_A$ is the zero-one indicator function of the set $A \subset \Omega$.

The empirical survival probability to age $x + t + s$ conditional on being alive at age $x + t$, for $s, t \geq 0$, is

$$_s\hat{p}_{x+t} = L_{x+t+s}/L_{x+t}, \quad \text{if } L_{x+t} > 0,$$

and the assumed true survival probability from age $x + t$ to age $x + t + s$ is

$$_sp_{x+t} = \mathbb{P}\left[T_i > t + s \mid T_i > t\right].$$

In line with actuarial convention, $\hat{p}_{x+t} := {}_1\hat{p}_{x+t}$ and $p_{x+t} := {}_1p_{x+t}$. The calculation of longevity credits paid to survivors in the fund uses the empirical survival probabilities $({}_s\hat{p}_{x+t})_{s,t \geq 0}$, whereas the calculation of the income withdrawn from the members' account values relies on the assumed true survival probabilities $({}_sp_{x+t})_{s,t \geq 0}$.

2.2. The income calculation

Each member of a pooled annuity fund has a fund account with value $W(t) \geq 0$ at time t, with constant $W(0) > 0$ at time 0. The fund account value changes over time due to investment returns, income withdrawals and longevity credits, the latter coming from the re-allocation of the fund accounts of the members who died.

The account values are invested to get a constant effective rate of return $R > -1$ over each unit of time.

The income withdrawn from each account by each surviving member is the amount that a fair life annuity would pay if purchased with the current fund value. Thus the member withdraws an income of amount

$$C(t) = W(t)/\ddot{a}_{x+t} \quad \text{at } t = 0, 1, 2, \ldots, \tag{1}$$

in which

$$\ddot{a}_{x+t} = 1 + \sum_{j=1}^{\infty}(1 + R)^{-j} \, {}_jp_{x+t}.$$

2.3. The longevity credit calculation

The account value at time $t + 1$ of a member who dies over the time period $(t, t + 1]$ is $(W(t) - C(t))(1 + R)$. The number of deaths observed over $(t, t + 1]$ is $L_{x+t} - L_{x+t+1}$ and so an amount equal to $(W(t) - C(t))(1 + R)(L_{x+t} - L_{x+t+1})$ is distributed equally among the L_{x+t+1} survivors observed at time $t + 1$. Thus the longevity credit paid at time $t + 1$ to each member alive at time $t + 1$ is

$$M(t + 1) = (W(t) - C(t))(1 + R)(L_{x+t} - L_{x+t+1})/L_{x+t+1}, \quad \text{if } L_{x+t+1} > 0. \tag{2}$$

If $L_{x+t+1} = 0$ then there is no-one left alive at time $t + 1$, and the pooled annuity fund ceases to exist. With no-one left in the fund at time $t + 1$, the account values of those who died over $(t, t+1]$ would be paid to their estate at time $t + 1$.

Immediately after the payment of the longevity credit, the account value at time $t + 1$ of a member who is alive at that time is

$$W(t + 1) = (W(t) - C(t))(1 + R) + M(t + 1), \quad \text{for } t \geq 0. \tag{3}$$

In contrast, the account value of a member who is dead at time $t + 1$ is zero since the funds of deceased members are distributed among the survivors.

Using the identity $\ddot{a}_{x+t} - 1 = p_{x+t}\, \ddot{a}_{x+t+1}/(1+R)$ and substituting for $M(t+1)$ from equation (2) and for $W(t)$ and $W(t+1)$ from equation (1) into equation (3), shows that the income at time $t+1$ can be written as

$$C(t+1) = C(t)\, p_{x+t}/\hat{p}_{x+t}, \quad \text{if } L_{x+t+1} > 0. \tag{4}$$

Thus fluctuations in the income are caused only by fluctuations in the observed survival probability $\hat{p}_{x+t}$ against the true survival probability p_{x+t} and the level of investment returns do not affect the fluctuations. This is a consequence of setting the investment returns to be constant and known, which allows idiosyncratic mortality risk to be studied in isolation. The income in one future scenario is shown in Figure 1.
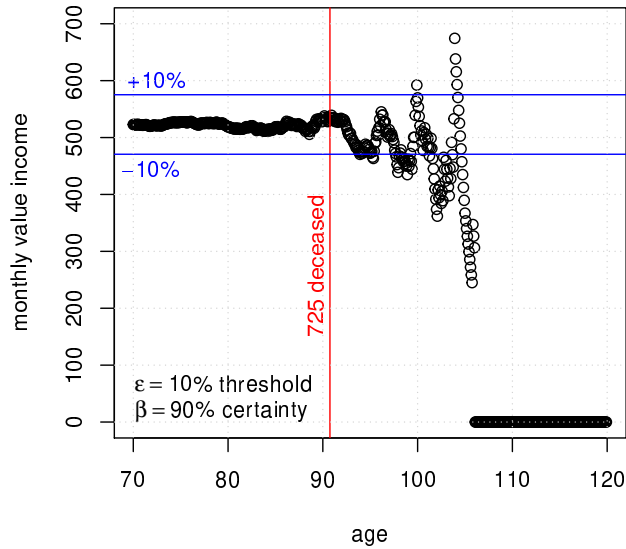

Figure 1: A sample of the income process, shown as black circles, for a fund with initially $N = 1000$ members, each aged $x = 70$ and bringing $W(0) = 100\,000$ units of wealth to the fund at time 0. Time is measured in months, the mortality model used is the UK-based life table S1PFL (Continuous Mortality Investigation, 2008) and a uniform distribution of deaths is assumed between integer ages. The blue horizontal lines indicate the lower and upper income thresholds ($\varepsilon_1 = \varepsilon_2 = 0.1$). The plot shows only one sample. However, for 90% of all samples, the income stays between the income thresholds until 725 members died. In the sample income process shown, the income happens to stay within the band for a few years after the time of the 725th death.

If idiosyncratic longevity risk is completely diversified in the pooled annuity fund then, for all $t \geq 0$, $_t\hat{p}_x = {}_tp_x$ and it would follow that $C(t) = C(0)$. However, the analysis in this paper focuses on less-than-perfect pooling of idiosyncratic longevity risk. The key investigation is to measure how well does the income stream $(C(t); t = 0, 1, 2, \ldots)$ stay close to the initial amount of income withdrawn, $C(0)$.

3. The problem

The purpose of the pooled annuity fund is to pay a regular, stable income to the fund's members for as long as they live. The level of the initial income withdrawn, $C(0)$, is important since the stability of the subsequent income withdrawn by the annuitants is measured by reference to it.

Fix a value $\varepsilon_1 \in (0, 1)$, called the lower threshold parameter, a value $\varepsilon_2 > 0$, called the upper threshold parameter and a value $\beta \in (0, 1)$, called the certainty. The income $C(\omega, t)$ at time $t > 0$ and in future scenario ω is deemed to be stable if $C(\omega, t) \in [(1 - \varepsilon_1)C(0), (1 + \varepsilon_2)C(0)]$. The values $(1 - \varepsilon_1)C(0)$ and $(1 + \varepsilon_2)C(0)$ are called the lower and upper income thresholds, respectively. The aim is to measure for how long is the income stable in at least $100\beta\%$ of the future scenarios. Denote the number of people who receive an income in the range $[(1 - \varepsilon_1)C(0), (1 + \varepsilon_2)C(0)]$ in at least $100\beta\%$ of the future scenarios for the whole of their lifetime by k_C. Then all participants receive a stable income with certainty β up to the time of the k_Cth death. But it is only the k_C annuitants who die first who receive a stable income with certainty β for the whole of their lifetime, rather than for the first part of their lifetime only.

The motivation for the definition of income stability is that it should be understandable to finance professionals. The closest definition to ours is the "4% rule", introduced in Bengen (2001) and further studied in Guyton (2004), Guyton and Klinger (2006) and Pfau and Kitces (2014). The idea is to calculate the initial amount of income that can be withdrawn for 30 years, with subsequent year's withdrawals adjusted for inflation, such that there is at least a 90% chance of this strategy being sustainable. In contrast to the probabilistic approach taken here, an alternative approach can be to maximize the expected value of a function of the income withdrawn. For example, He and Liang (2013) minimize the discounted expected value of the squared distance of the income from a target value. Others such as Bruhn and Steffensen (2013), Constantinides (1990), Munk (2008) and Van Bilsen et al. (2020) are based on habit formation. The latter paper and Curatola (2017) assume a loss averse investor.

The time when exactly k members have passed away is denoted by the random variable $T_{(k)}$, the kth order statistic of the future lifetime random variables. Suppose that the maximum integer $k := k_C$ to satisfy

$$\mathbb{P}\left[(1 + \varepsilon_2)C(0) \geq C(s) \geq (1 - \varepsilon_1)C(0) \text{ for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right] \geq \beta, \qquad (5)$$

in which $\lfloor T_{(k)} \rfloor$ is the integer part of $T_{(k)}$, is to be determined. Figure 1 illustrates a choice of the income thresholds, with a realisation of the income stream.

In this paper, a close, lower bound to k_C is found, which holds for any mortality model that excludes systematic longevity risk. This surprising independence result is a consequence of the inverse transform method, which is applied in Theorem 3.1 below. It means that the values of the lower bound calculated for a choice of ε, β and N remain the same regardless of what distribution is chosen for $T_1, T_2, \ldots, T_N$. In contrast, calculating the value of k_C directly from (5) does require the choice of a mortality model.

It would be ideal to find an explicit expression for the maximum integer that satisfies (5), rather than a lower bound on it. Indeed, the exact distribution of the order statistics of $T_1, T_2, \ldots, T_N$ are known (Birnbaum and Lientz, 1969; Csörgö, 1965). Unfortunately, they are not amenable to calculations since their distribution functions are polynomials of degree N, and in a pooled annuity fund the value of N represents the initial number of members. Thus the value of N is likely to be in the order of hundreds or more.

Going further, the preferred goal would be to find the maximal integer time $t \geq 1$ such that

$$\mathbb{P}\left[(1 + \varepsilon_2)C(0) \geq C(s) \geq (1 - \varepsilon_1)C(0) \text{ for all } s \in \{1, 2, \ldots, t\}\right] \geq \beta.$$

However, the maximal integer time would depend on a mortality model and would not yield a general result, like the one in this paper.

3.1. The main results

Theorem 3.1. *Let $U_{(1)}, U_{(2)}, \ldots, U_{(N)}$ be the order statistics of N independent and standard uniformly distributed random variables $U_1, U_2, \ldots, U_N$. Similarly, denote the order statistics of the independent and identically distributed future lifetime random variables $(T_i)_{i=1}^N$, which have a continuous distribution, by $T_{(1)}, T_{(2)}, \ldots, T_{(N)}$. Fix constants $\varepsilon_1 \in (0,1)$, $\varepsilon_2 > 0$ and $k \in \{1, 2, \ldots, N\}$. Then*

$$\mathbb{P}\left[(1+\varepsilon_2)C(0) \geq C(s) \geq (1-\varepsilon_1)C(0) \ \text{for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right]$$
$$\geq \mathbb{P}\left[(1-\varepsilon_1)\tfrac{i-1}{N} + \varepsilon_1 \geq U_{(i)} \geq (1+\varepsilon_2)\tfrac{\min\{i, N-1\}}{N} - \varepsilon_2 \ \text{for all } i \in \{1, 2, \ldots, k\}\right],$$

in which $C(s)$ is the income at time s that is calculated via equation (1).

Proof. Fix $t \in \mathbb{N}$ and assume that at least one person is alive at age $x + t + 1$, i.e. $L_{x+t+1} > 0$. Starting with equation (4), it follows by induction on $t \in \mathbb{N}$ that

$$C(t+1) = C(0) \prod_{j=0}^{t} \frac{p_{x+j}}{\hat{p}_{x+j}} = C(0)\frac{_{t+1}p_x}{_{t+1}\hat{p}_x}.$$

As the joint distribution of $T_1, T_2, \ldots, T_N$ is continuous, the set $[\lfloor T_{(k)} \rfloor < T_{(k)}$, for $k = 1, 2, \ldots, N]$ has measure one. Noting that $_sp_x > 0$ for $s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}$ on this set and working only on this set, it follows that

$$\left[(1+\varepsilon_2)C(0) \geq C(s) \geq (1-\varepsilon_1)C(0) \ \text{for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right]$$
$$= \left[1 + \varepsilon_2 \geq \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1 \ \text{for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right]$$
$$= \left[\inf_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1\right] \cap \left[\sup_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2\right].$$

Let both $T_{(N+1)} := \infty$ and $_sp_x/_s\hat{p}_x := 1$, if $_s\hat{p}_x = 0$. Since $\{1, 2, \ldots, \lfloor T_{(k)} \rfloor\} \subset [0, T_{(k)}) \subset [0, T_{(k+1)})$, it follows

$$\inf_{s \in [0, T_{(k)})} \frac{_sp_x}{_s\hat{p}_x} \quad \leq \quad \inf_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x},$$
$$\sup_{s \in [0, T_{(k+1)})} \frac{_sp_x}{_s\hat{p}_x} \quad \geq \quad \sup_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x},$$

and so

$$\inf_{s \in [0, T_{(k)})} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1 \quad \Rightarrow \quad \inf_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1,$$
$$\sup_{s \in [0, T_{(k+1)})} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2 \quad \Rightarrow \quad \sup_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2.$$

Therefore

$$\left[\inf_{s \in [0, T_{(k)})} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1\right] \cap \left[\sup_{s \in [0, T_{(k+1)})} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2\right]$$
$$\subset \left[\inf_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1\right] \cap \left[\sup_{s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2\right].$$

In summary,

$$\left[\inf_{s \in [0, T_{(k)})} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1\right] \cap \left[\sup_{s \in [0, T_{(k+1)})} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2\right]$$
$$\subset \left[(1+\varepsilon_2)C(0) \geq C(s) \geq (1-\varepsilon_1)C(0) \ \text{for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right]. \tag{6}$$

Now the goal is to write the left-hand side of (6) in terms of the order statistics of independent, standard uniformly distributed random variables. This is done by partitioning $[0, T_{(k)})$ into intervals $[T_{(i-1)}, T_{(i)})$ and considering the minimum value of $_sp_x/_s\hat{p}_x$ over each sub-interval $[T_{(i-1)}, T_{(i)})$.

The empirical distribution function of the proportion of the initial membership who have died up to time $t \geq 0$ is defined as

$$\hat{F}_N(t) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{T_i \leq t\}} = \frac{1}{N} \max\{i : T_{(i)} \leq t\}.$$

Denote by F the distribution function of the death times $T_1, T_2, \ldots, T_N$. Let $(1 - F(s))/(1 - \hat{F}_N(s)) := 1$ if $1 - \hat{F}_N(s) = 0$. It follows immediately from the definition of the (empirical) survival probability that $_s\hat{p}_x = 1 - \hat{F}_N(s)$ and $_sp_x = 1 - F(s)$. Hence

$$
\left[\inf_{s \in [0, T_{(k)})} \frac{_sp_x}{_s\hat{p}_x} \geq 1 - \varepsilon_1 \right] \cap \left[\sup_{s \in [0, T_{(k+1)})} \frac{_sp_x}{_s\hat{p}_x} \leq 1 + \varepsilon_2 \right]
$$
$$
= \left[\inf_{s \in [0, T_{(k)})} \frac{1 - F(s)}{1 - \hat{F}_N(s)} \geq 1 - \varepsilon_1 \right] \cap \left[\sup_{s \in [0, T_{(k+1)})} \frac{1 - F(s)}{1 - \hat{F}_N(s)} \leq 1 + \varepsilon_2 \right]. \tag{7}
$$

Let $T_{(0)} := 0$. As the joint distribution of $T_1, T_2, \ldots, T_N$ is continuous, the set $[T_{(i-1)} < T_{(i)}, \text{ for } i = 1, 2, \ldots, N+1]$ has measure one. In the following, we work on this set only.

Consider an arbitrary time interval, $[T_{(i-1)}, T_{(i)})$. The empirical distribution function $\hat{F}_N$ changes value only when a member dies, namely only at times $T_{(1)}, \ldots, T_{(N)}$. In particular, for all $s \in [T_{(i-1)}, T_{(i)})$, $\hat{F}_N(s) = \hat{F}_N(T_{(i-1)}) = \hat{F}_N(T_{(i)}-)$, the left limit of $\hat{F}_N$ at $T_{(i)}$.

As the distribution function F of the death time is an increasing function, it follows that $1 - F(s)$ is a decreasing function that approaches its infimum over $s \in [T_{(i-1)}, T_{(i)})$ at the end of the interval. Thus, as $\hat{F}_N(s)$ is constant over the same interval, the fraction $(1 - F(s))/(1 - \hat{F}_N(s))$ approaches its infimum over $s \in [T_{(i-1)}, T_{(i)})$ as s approaches $T_{(i)}$. By the continuity of F, $F(T_{(i)}-) = F(T_{(i)})$ so that

$$\inf_{s \in [T_{(i-1)}, T_{(i)})} \frac{1 - F(s)}{1 - \hat{F}_N(s)} = \frac{1 - F(T_{(i)}-)}{1 - \hat{F}_N(T_{(i)}-)} = \frac{1 - F(T_{(i)})}{1 - \hat{F}_N(T_{(i-1)})}.$$

On the other hand, as $1 - F(s)$ is a decreasing function, it attains its largest value over $s \in [T_{(i-1)}, T_{(i)})$ at the start of the interval. Thus the fraction $(1 - F(s))/(1 - \hat{F}_N(s))$ attains its largest value over $s \in [T_{(i-1)}, T_{(i)})$ at $s = T_{(i-1)}$, i.e.

$$\sup_{s \in [T_{(i-1)}, T_{(i)})} \frac{1 - F(s)}{1 - \hat{F}_N(s)} = \frac{1 - F(T_{(i-1)})}{1 - \hat{F}_N(T_{(i-1)})}.$$

The argument above shows that the infimum of $(1 - F(s))/(1 - \hat{F}_N(s))$ over $s \in [0, T_{(k)})$ is equal to the infimum of $(1 - F(T_{(i)}))/(1 - \hat{F}_N(T_{(i-1)}))$ over $i \in \{1, 2, \ldots, k\}$ and that the supremum of $(1 - F(s))/(1 - \hat{F}_N(s))$ over $s \in [0, T_{(k+1)})$ is equal to the supremum of $(1 - F(T_{(i-1)}))/(1 - \hat{F}_N(T_{(i-1)}))$ over $i \in \{1, 2, \ldots, k+1\}$. In summary,

$$\left[\inf_{s \in [0, T_{(k)})} \frac{1 - F(s)}{1 - \hat{F}_N(s)} \geq 1 - \varepsilon_1 \right] = \left[\inf_{i \in \{1, 2, \ldots, k\}} \frac{1 - F(T_{(i)})}{1 - \hat{F}_N(T_{(i-1)})} \geq 1 - \varepsilon_1 \right], \tag{8}$$

and, using $T_{(0)} := 0$,

$$\left[\sup_{s \in [0, T_{(k+1)})} \frac{1 - F(s)}{1 - \hat{F}_N(s)} \leq 1 + \varepsilon_2\right] = \left[\sup_{i \in \{1,2,\ldots,k+1\}} \frac{1 - F(T_{(i-1)})}{1 - \hat{F}_N(T_{(i-1)})} \leq 1 + \varepsilon_2\right]$$

$$= \left[\sup_{i \in \{2,3,\ldots,k+1\}} \frac{1 - F(T_{(i-1)})}{1 - \hat{F}_N(T_{(i-1)})} \leq 1 + \varepsilon_2\right] \tag{9}$$

$$= \left[\sup_{i \in \{1,2,\ldots,k\}} \frac{1 - F(T_{(i)})}{1 - \hat{F}_N(T_{(i)})} \leq 1 + \varepsilon_2\right].$$

As F is continuous, the random variables $U_i = F(T_i)$ for $i \in \{1, 2, \ldots, N\}$ are independent and standard uniformly distributed. Their order statistics, $(U_{(i)})_{i=1}^N$, are linked to the order statistics of the future lifetime random variables by the identity

$$U_{(i)} = F(T_{(i)}) \quad \text{for } i = 1, 2, \ldots, N.$$

Furthermore, by the definition of the empirical distribution function, $\hat{F}_N(T_{(i-1)}) = (i-1)/N$, for $i = 1, 2, \ldots, N$ with $\hat{F}_N(T_{(0)}) = \hat{F}_N(0) = 0$. Let $(1 - U_{(N)})/0 := 1$, then

$$\left[\inf_{i \in \{1,2,\ldots,k\}} \frac{1 - F(T_{(i)})}{1 - \hat{F}_N(T_{(i-1)})} \geq 1 - \varepsilon_1\right] \cap \left[\sup_{i \in \{1,2,\ldots,k\}} \frac{1 - F(T_{(i)})}{1 - \hat{F}_N(T_{(i)})} \leq 1 + \varepsilon_2\right]$$

$$= \left[\inf_{i \in \{1,2,\ldots,k\}} \frac{1 - U_{(i)}}{1 - (i-1)/N} \geq 1 - \varepsilon_1\right] \cap \left[\sup_{i \in \{1,2,\ldots,k\}} \frac{1 - U_{(i)}}{1 - i/N} \leq 1 + \varepsilon_2\right]$$

$$= \left[(1 - \varepsilon_1)\tfrac{i-1}{N} + \varepsilon_1 \geq U_{(i)} \text{ for all } i \in \{1, 2, \ldots, k\}\right]$$

$$\cap \left\{(1 + \varepsilon_2)\tfrac{i}{N} - \varepsilon_2 \leq U_{(i)} \text{ for all } i \in \{1, 2, \ldots, k\} \setminus \{N\}\right\}$$

$$= \left[(1 - \varepsilon_1)\tfrac{i-1}{N} + \varepsilon_1 \geq U_{(i)} \geq (1 + \varepsilon_2)\tfrac{\min\{i,N-1\}}{N} - \varepsilon_2 \text{ for all } i \in \{1, 2, \ldots, k\}\right].$$

Combining the last equation with equations (6)-(9) and taking the probability, the desired result is obtained. □

In the sequel, the focus is on either a symmetric income threshold – representing a desire to avoid both upside and downside income volatility – or a lower income threshold only – representing a desire to avoid downside income volatility.

Corollary 3.2. *Suppose that for $k \in \{1, 2, \ldots, N\}$ and $\varepsilon \in (0,1)$,*

$$\mathbb{P}\left[(1 - \varepsilon)\tfrac{i-1}{N} + \varepsilon \geq U_{(i)} \geq (1 + \varepsilon)\tfrac{\min\{i,N-1\}}{N} - \varepsilon \text{ for all } i \in \{1, 2, \ldots, k\}\right] \geq \beta. \tag{10}$$

Then

$$\mathbb{P}\left[(1 + \varepsilon)C(0) \geq C(s) \geq (1 - \varepsilon)C(0) \text{ for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right] \geq \beta. \tag{11}$$

Proof. Apply Theorem 3.1 with $\varepsilon = \varepsilon_1 = \varepsilon_2$. □

Corollary 3.3. *Suppose that, for $k \in \{1, 2, \ldots, N\}$ and $\varepsilon \in (0,1)$,*

$$\mathbb{P}\left[(1 - \varepsilon)\tfrac{i-1}{N} + \varepsilon \geq U_{(i)} \text{ for all } i \in \{1, 2, \ldots, k\}\right] \geq \beta. \tag{12}$$

Then

$$\mathbb{P}\left[C(s) \geq (1 - \varepsilon)C(0) \text{ for all } s \in \{1, 2, \ldots, \lfloor T_{(k)} \rfloor\}\right] \geq \beta. \tag{13}$$

Proof. Apply Theorem 3.1 with $\varepsilon = \varepsilon_1$, let $\varepsilon_2 \uparrow \infty$ and observe

$$(1 + \varepsilon_2)\tfrac{i}{N} - \varepsilon_2 = \tfrac{i}{N} - \varepsilon_2(1 - \tfrac{i}{N}) \longrightarrow -\infty \quad \text{for all } i \in \{1, 2, \ldots, N - 1\}.$$

<div align="right">□</div>

The motivation for proving the results in Section 3.1 is that, rather than calculating the maximal integer $k := k_C$ satisfying (11), the maximal integer $k := k_U$ satisfying (10) can instead be calculated. The same idea applies for (12) and (13). This has some significant benefits.

The value of k_U is independent of the distribution of $T_1, T_2, \ldots, T_N$ which, if k_U is close in value to k_C, allows general conclusions to be drawn as to the ability of the pooled annuity fund to diversify longevity risk. Additionally, the calculation of k_U, which requires only the sampling of random variables, is faster than the calculation of k_C, which requires the simulation of the income process in the pooled annuity fund. It is straightforward to show that the calculation of k_C requires the order of $(L + N) \times M$ operations whereas k_U takes the order of $N \times M$ operations, where L is the number of times that the income is calculated from time 0 to the maximum time of the last death in the fund, in each simulation. (For example, if the income was calculated monthly, the participants were age 70 at time 0 and lived until at most age 120 then, at most, $L = 600$ time-steps are needed in each simulation to calculate the income paid to the surviving participants.)

Therefore, a key question is how close is the lower bound k_U to the true value k_C? As is shown next for some chosen mortality models, k_U is close enough to k_C to use it to make conclusions on the pooled annuity fund's ability to diverse idiosyncratic longevity risk.

3.2. How close is k_U to k_C?

To examine how close k_U is to k_C, the maximal integers satisfying each of (10) to (13) individually are calculated; the method for each calculation is described in Sections 3.2.1 and 3.2.2, and all simulations were carried out in the statistical software package R.

Figure 2, which details the extent to which the relative difference between them improves as the number of members N initially in the fund increases, shows that k_C is quite close to k_U. For example, k_C is less than 3% above the value of k_U when $N = 2\,000$, and the percentage falls to 2% when $N = 4\,000$ and falls further to about 1% when $N = 8\,000$. These values apply for the two considered life tables and for the two different starting ages, 50 and 70 years old. The relative difference falls as N increases, due to k_U/N and k_C/N converging to one as N increases.

It is seen from Figure 2 that the younger group of initially 50-year-olds has a lower relative error than the older group of initially 70-year-olds. The reason is a technical one. Denote the distribution function of T_1 for the 50-year-olds by F_{50} and that for the 70-year-olds by F_{70}, with $F_{50} < F_{70}$. This inequality implies that for any $y \in (0, 1)$ and any given time grid $(t_i)_i$, there are more indexes i such that $F_{50}(t_i) < y$ than there are indexes i satisfying $F_{70}(t_i) < y$. Suppose that for a sequence of uniform order statistics $(U_{(k)})_{k=1}^{N}$, the goal is to check if condition (10) holds until the kth member. To do this, the condition is checked for all indexes i satisfying $F_{50}(t_i) < U_{(k)}$, and similarly for all indexes i satisfying $F_{70}(t_i) < U_{(k)}$. Since for the 70-year-olds there are fewer indices i to check than for 50-year-olds, the condition is more often fulfilled by the 70-year-olds, yielding a higher maximal k_U for the 70-year-olds. This is true even though the values $F_x(t_i)$ are different for $x = 50$ and $x = 70$, since the values are close to each other. The relative differences are magnified further when N is small, due to the smaller numbers involved.

A closer look at the results displayed in Figure 2a shows that the value of the threshold parameter ε is more important than the value of the certainty β in determining the goodness of the

approximation. The larger the value of ε, and therefore the wider are the thresholds, the better is the approximation. For example, when $\varepsilon = 0.1$ and $N = 2\,000$, k_C is at most 1% higher than k_U. However, the relative difference increases to 3% when $\varepsilon = 0.05$. The same observations apply to the other plots in Figure 2.

Turning to how these relative differences affect the likely time for which the fund can provide a stable, life-long income, Figure 3 indicates that the difference is at most 10 months. For funds with at least $2\,000$ members, using k_U instead of k_C understates the likely time for which a stable, life-long income is provided by between 2 to 4 months. An explanation of the likely time is given in Section 4.

The conclusion is that, since it is a close lower bound, k_U can be used as an approximation to k_C. The calculation methods of k_U and k_C are described next.

3.2.1. Calculation of the maximal number k_U

To find the maximal number $k := k_U$ that fulfills either (10) or (12), Monte Carlo simulation is used. The initial number of members in the fund, N, and the certainty β are fixed. As the income thresholds are symmetric about the initial income, call $\varepsilon := \varepsilon_1 = \varepsilon_2$ the threshold parameter. Suppose that $M \in \mathbb{N}$ sample vectors are generated of the uniform order statistics $(U_{(1)}, U_{(2)}, \ldots, U_{(N)})$.

Denote the mth sample of the uniform order statistics by $(u_{(1)}^{(m)}, u_{(2)}^{(m)}, \ldots, u_{(N)}^{(m)})$, for $m \in \{1, 2, \ldots, M\}$. For each $m \in \{1, 2, \ldots, M\}$, the first integer $i := i(m) \in \{1, 2, \ldots, N\}$ that fails

$$\varepsilon + (1 - \varepsilon)(i - 1)/N \geq u_{(i)}^{(m)} \geq (1 + \varepsilon)\min\{i, N - 1\}/N - \varepsilon$$

is determined and $k(m) = i(m) - 1$ is recorded. If there is no such $i(m)$, then $k(m) = N$ is recorded.

The values $(k(m))_{m=1}^M$ are considered as samples from a random variable K and are used to calculate the empirical distribution of K. Finally, the value of k_U is calculated as the β-quantile of this empirical distribution of K. The same method is used to find the maximal number that fulfills (12). The final calculated value, k_U, is the result of 10 million simulations.

Note that the procedure can be implemented efficiently since the sorting of uniform random variables can be avoided. According to Devroye (1986, f.207), uniform order statistics are ratios of sums of exponential random variables. More precisely, let $(E_i)_{i=1}^{N+1}$ be independent Exp(1)-distributed random variables and define $S_i = \sum_{j=1}^i E_j$ for integers $1 \leq j \leq N + 1$. Then the distributions of $(U_{(i)})_{i=1}^N$ and $(S_i/S_{N+1})_{i=1}^N$ are the same.

Table 1 lists the maximal number k_U as the number of initial members N increases, for a selection of values of ε and β, depending on whether either a lower income threshold or a symmetric lower and upper income threshold are applied. Plotting the maximal numbers (Figure 4), it can be seen that the value of k_U increases approximately linearly for $N \geq 2\,000$. The linear approximation improves as the income thresholds widen (i.e. as ε increases) and as the certainty β decreases.

3.2.2. Calculation of the maximal number k_C

Calculating k_C involves a straightforward simulation of $M = 10$ million sample paths of the income process, based on monthly time units and applying an appropriate life table. Along the mth sample path, the maximum value of $k(m)$ satisfying the event in inequality (11) or (13), as appropriate, is determined. Similar to the calculation above, the value of k_C is calculated as the β-quantile of the empirical distribution of a random variable K, for which $(k(m))_{m=1}^M$ are the observed samples.
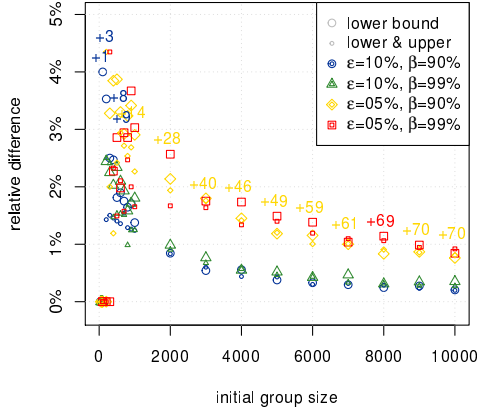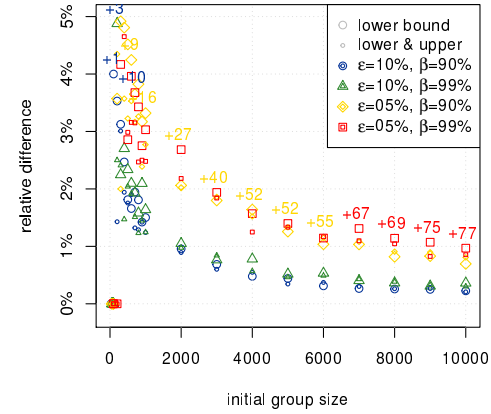
(a) Human Mortality Database life table, starting age 50.
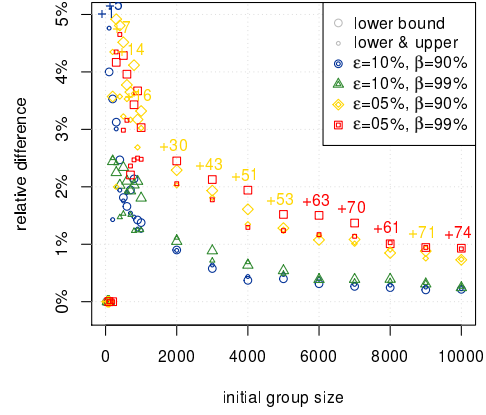
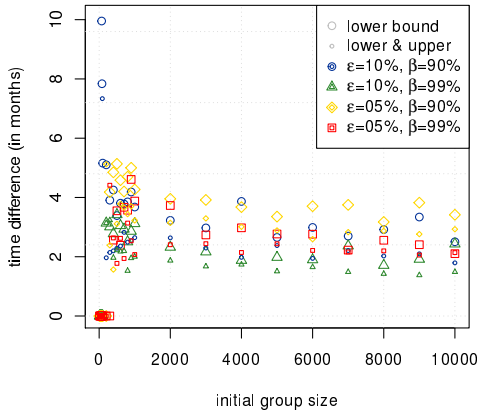(b) Human Mortality Database life table, starting age 70.
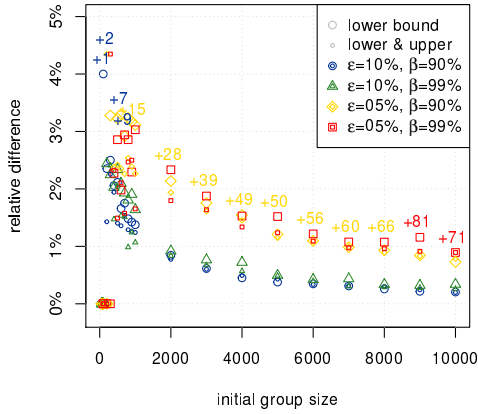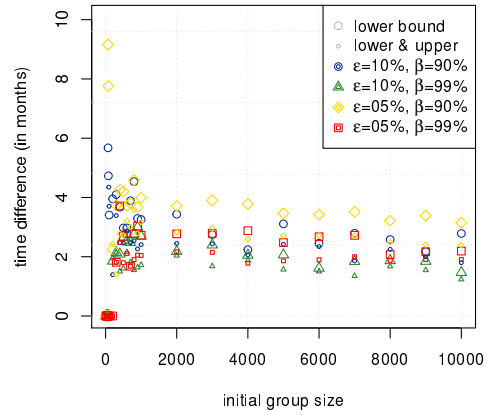

(c) Life table S1PFL, starting age 50.

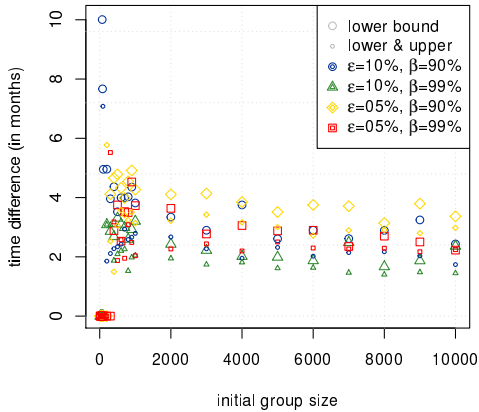(d) Life table S1PFL, starting age 70.

Figure 2: Relative distance of the maximal integer $k := k_C$ fulfilling (11) from the maximal integer $k := k_U$ fulfilling (10), i.e. $(k_C - k_U)/k_U$, as the initial group size increases. The smallest initial group size is 10. The relative distances are displayed for a selection of values of ε and β, and are indicated by the smaller symbols. The same calculation is shown for (13) and (12), respectively, and indicated by the larger symbols. The numbers in the plot state the maximum value of $k_C - k_U$, for each value of N, over the eight different combinations of the values of β and ε, and whether a lower income threshold only or both income thresholds are included. For the calculation of k_C from (11) and (13), it is assumed that income payments are paid monthly to survivors, the initial age is either 50 or 70 years old, and the mortality distribution is based on either the Human Mortality Database' life table for the UK for 2016 (Human Mortality Database, 2018) or the UK-based life table S1PFL (Continuous Mortality Investigation, 2008).
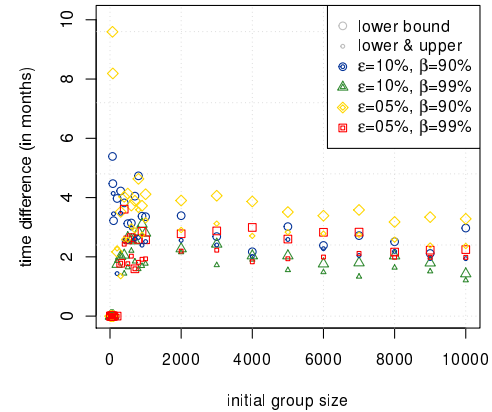
(a) Human Mortality Database life table, starting age 50.

(b) Human Mortality Database life table, starting age 70.

(c) Life table S1PFL, starting age 50.

(d) Life table S1PFL, starting age 70.

Figure 3: Additional likely time for which the fund can provide a stable, life-long income when using the maximal integer $k := k_C$ instead of the maximal integer $k := k_U$, as the initial group size increases. Thus using $k := k_U$ understates the time for which the fund can provide a stable, life-long income by at most 10 months, for the considered life tables and parameters. The assumptions are the same as in Figure 2. An explanation of the likely time is given in Section 4.

| | $\varepsilon = 10\%$ | | $\varepsilon = 10\%$ | | $\varepsilon = 5\%$ | | $\varepsilon = 5\%$ | |
| | $\beta = 90\%$ | | $\beta = 99\%$ | | $\beta = 90\%$ | | $\beta = 99\%$ | |
N	k_U^{above}	k_U^{both}	k_U^{above}	k_U^{both}	k_U^{above}	k_U^{both}	k_U^{above}	k_U^{both}
100	25	21	9	9	6	6	1	1
200	85	70	41	40	28	23	9	9
500	331	285	214	196	155	124	70	67
1000	799	725	610	562	483	397	264	242
2000	1778	1680	1524	1436	1310	1135	857	779
3000	2770	2662	2485	2377	2224	1988	1599	1466
4000	3766	3652	3463	3342	3171	2894	2421	2242
5000	4764	4645	4450	4320	4137	3829	3291	3072
6000	5762	5641	5440	5304	5113	4781	4192	3940
7000	6761	6638	6434	6292	6093	5744	5112	4831
8000	7760	7636	7427	7283	7079	6715	6049	5742
9000	8759	8634	8424	8276	8067	7692	6997	6670
10000	9758	9632	9420	9269	9059	8673	7952	7608

Table 1: The values k_U^{above} and k_U^{both} are the maximal integer k that satisfies (12) and (10), respectively.
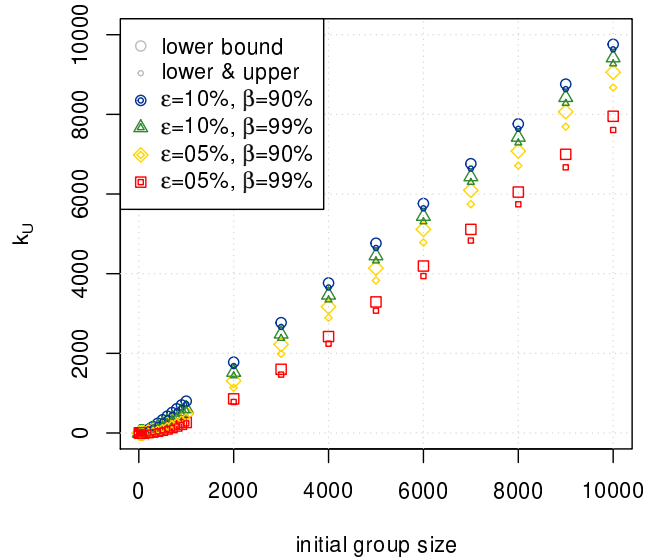

Figure 4: The maximal number of members k_U for whom the lifetime income is stable, plotted against the initial number of members N in the fund, for a selection of threshold parameters ε and certainties β. The number is calculated from either (12), represented by the larger symbols in the plot, or (10), represented by the smaller symbols.

3.3. Discussion of the main results

The purpose of the pooled annuity fund is to pay a stable life-long income to its participants. What universal features can be noted about its ability to do so, if volatility is due only to idiosyncratic longevity risk and k_U is used as a close approximation to k_C? In this section, stability is defined as the future income payments each being within $100\varepsilon\%$ of the initial income $C(0)$ (as in Corollary 3.2). This is equivalent to considering the values of i for which the condition

$$\varepsilon + (1 - \varepsilon)(i - 1)/N \geq U_{(i)} \geq (1 + \varepsilon)\min\{i, N - 1\}/N - \varepsilon \tag{14}$$

holds, in which the upper bound $\varepsilon + (1 - \varepsilon)(i - 1)/N$ corresponds to the lower income threshold $100(1 - \varepsilon)C(0)$. The lower bound $(1 + \varepsilon)\min\{i, N - 1\}/N - \varepsilon$ corresponds to the upper income threshold $100(1 + \varepsilon)C(0)$.

First, note that the distance between the lower and upper bounds narrows as i increases. Geometrically, considering the bounds as lines, the gradient to the lower bound is higher than the gradient of the upper bound. Thus as i increases, it may be more likely that a sample of $U_{(i)} \sim \text{Beta}(i, N - i + 1)$ (Shorack and Wellner, 2009, pp.97) lies outside of the bounds. To investigate this possibility further, the bounds and the 0.5% and 99.5% quantiles of the uniform order statistics are plotted in Figure 5 for $\varepsilon = 0.05$. Note that although the support of the distribution functions of the uniform order statistics is $[0, 1]$, the quantiles indicate the location of the vast majority (99%) of their possible values.

Figure 5 illustrates that the income received at the end of the life of the longest-lived members is less likely to lie between the upper and lower income thresholds than for the shortest-lived members. This is seen from Figure 5: the 0.5% and 99.5% quantile lines (the uppermost and lowermost black, dashed lines in the figure) lie inside the bounds (red, solid lines) for the lower order statistics (i.e. the smaller values of i/N, which denote the shorter-lived members). For larger values of i/N (denoting the longer-lived members), the quantile lines lie outside the bounds in the plot. When this occurs depends on the number of people initially in the fund. For $N = 200$, the quantiles are seen to fall quickly outside the bounds, as i/N increases (Figure 5a). Only about the first 6 members ($i/200 \approx 0.03$) are extremely likely to get a life-long, stable income. As N increases to $2\,000$ (Figure 5b), it is only for approximately the first 900 members (i.e. the first 800 members; $i/2000 \approx 0.45$) that the 0.5% to 99.5% quantiles of $U_{(i)}$ falls outside the bounds. Note that these values are not the same as the maximal number of members who get a stable, life-long income with a certainty level of 99%, since they look only at the quantiles at each time of death, rather than along each possible future scenario.
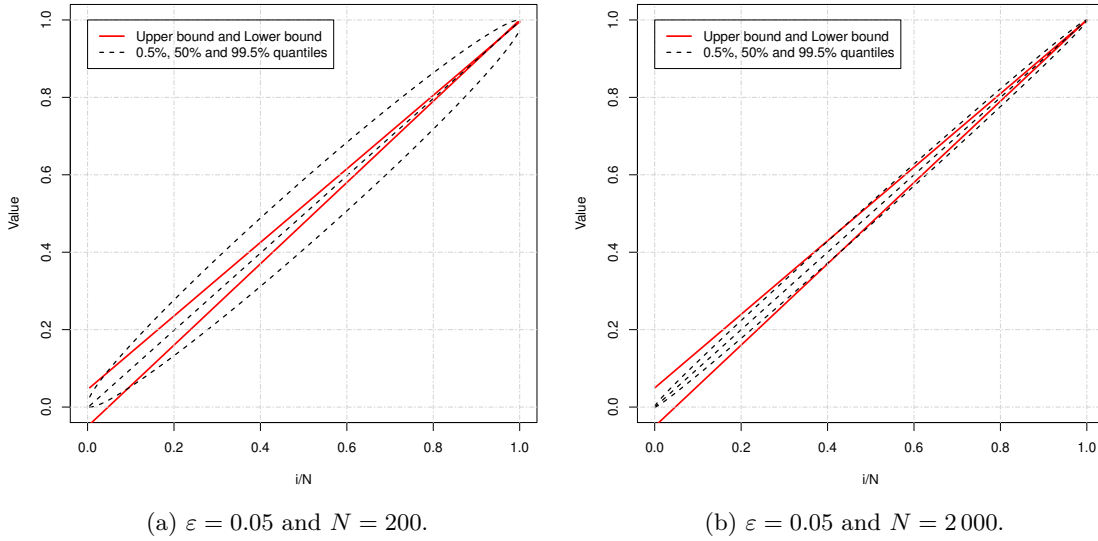
(a) $\varepsilon = 0.05$ and $N = 200$.   (b) $\varepsilon = 0.05$ and $N = 2\,000$.

Figure 5: The symmetric upper and lower bounds on the order statistics, for $\varepsilon = 0.05$ and either $N = 200$ or $N = 2\,000$, shown as solid red lines. The 0.5%, 50% and 99.5% quantiles of the ith uniform order statistic are displayed as dashed black lines. The x-axis values are i/N, rather than i, so that the domains of the plotted functions are all $[0, 1]$.

The second item to note on the condition (14) is that it fails to hold for the highest values of i, by virtue of the lower and upper bounds crossing each other. The point at which they cross is found by solving

$$\varepsilon + (1 - \varepsilon)(i - 1)/N = (1 + \varepsilon)\min\{i, N - 1\}/N - \varepsilon$$

for i. Setting $i = N$ gives a contradiction ($N = N - 1$) so assume that $i < N$ to obtain that the bounds have crossed for each i satisfying

$$i \geq N + \left(1 - \varepsilon^{-1}\right)/2. \tag{15}$$

Since the bounds are independent of the order statistics, in all future states of the world the following can be observed.

- In view of $\varepsilon \leq 1$, the right-hand side of the inequality (15) is less than or equal to N. Thus there is always at least one value of i (i.e. $i = N$) for which the bounds have crossed. Indeed, as $\varepsilon \to 1$, the right-hand side of the inequality (15) tends to N, which means that the two bounds cross only as the last member of the fund dies.

- As $\varepsilon \to 0$, the right-hand side of the inequality (15) tends to $-\infty$ and the two bounds cross before any fund member dies. In the limit, no fund member can receive a stable (i.e. constant) income in any future state of the world, meaning that the choice of β is irrelevant.

- For fixed ε, there are a fixed number of members who are still alive when the bounds have crossed each other. For example, for $\varepsilon = 0.05$ the longest-lived 10 members do not receive a stable income for the entirety of their future lifetime in any future state of the world, regardless of the initial number of members in the fund (since the bounds have crossed for $i \geq N + (1 - \varepsilon^{-1})/2 = N - 9.5$, i.e. $i \in \{N - 9, N - 8, \ldots, N\}$, a set of 10 elements.). The remaining $N - 10$ members may or may not receive a life-long income in any particular state of the world.

16

In summary, the longest-lived members are unlikely to receive a stable, life-long income when a symmetric income threshold is used to define income stability. This is by dint of the narrowing between the symmetric bounds for the longest-lived members and the distribution of the uniform order statistics relating to those longest-lived members not concentrating at the same rate. In fact, as long as $\varepsilon < 1$, the longest-lived members have no opportunity to get a stable income for the entirety of their future lifetime, in any future state of the world. Nonetheless, they may still receive a stable income for the majority of their future lifetime.

These observations support the increasing volatility in income for the longest-lived members (in the last cohort to enter if the fund is an open one), remarked upon in Piggott et al. (2005), Qiao and Sherris (2013) and Sabin (2010). The results in this paper enable the precise elucidation in our model of when "increasingly volatile" becomes "too volatile".

Next turn to the case when income stability is defined as the income lying above a lower income threshold only. This is equivalent to considering the values of i for which the condition

$$\varepsilon + (1 - \varepsilon)(i - 1)/N \geq U_{(i)}$$

holds. From Figure 5, it appears that the 50% quantile line lies below the upper bound (the upper solid red line in Figure 5; the lower solid red line can be ignored as it represents the redundant upper income threshold). Considering the mean of the ith order statistic, $i/(N + 1)$, instead of the median – since it has a simple formula that is amenable to algebra and is not dissimilar in value (Kerman, 2011) – then

$$\varepsilon + (1 - \varepsilon)(i - 1)/N \geq i/(N + 1) \qquad \Rightarrow \qquad \begin{cases} i \leq N + 1 & \text{if } \varepsilon > 1/(N + 1) \\ i \geq N + 1 & \text{if } \varepsilon \leq 1/(N + 1). \end{cases}$$

Thus the means of the N order statistics all lie above the upper bound $\varepsilon + (1 - \varepsilon)(i - 1)/N$, if the income threshold parameter $\varepsilon \leq 1/(N + 1)$. Approximating the medians of the N order statistics by their means, this means that there is a less than 50% chance of each order statistic lying below the upper bound.

In summary, the value of ε should not be too small, i.e. it should be at least above $1/(N + 1)$. Otherwise, life-long income stability will not be achieved for a substantial proportion of the fund membership. This comment applies in the presence of either a lower income threshold only or both income thresholds.

4. Application: determining for how long the fund can provide a stable income

The time at which the income of a pooled annuity fund becomes unstable – i.e. when the income no longer lies between the thresholds in at least $100\beta\%$ of future states of the world – is calculated in terms of the number of deaths that occur up to that time. However, it is useful to communicate this as a length of time which represents how long all members receive a stable income with a specified level of certainty. The members who die before that length of time receive a life-long stable income, whereas those who die after it receive a stable income for only part of their lifetime.

Since the time at which a specified number of deaths occurs varies in each future state of the world, an estimate of the time at which it occurs is calculated, which is called the likely time. This is an approximation of the true time, since the number of deaths used, k_U, is calculated by maximising the left-hand side of inequality (12) over k, rather than maximising that of inequality (13).

The calculation of the length of time requires the selection of a mortality law, and the mortality law matters. For example, the length of time over which the fund is likely to provide a stable income will increase as mortality lightens (i.e. people live longer).

For example, assume that $N = 2\,000$ members all join the pooled annuity fund at age 70. The annual income $C(m)$ which they withdraw at time $m \in \{0, 1, 2, \ldots\}$ is calculated by dividing their fund value at time m by the value of a single life annuity calculated at age $70 + m$, as described in Section 2.2. Suppose that the future income is considered stable if it is at least 95% of the initial income (i.e. lower threshold parameter of $\varepsilon = 0.05$ and no upper threshold parameter), in 90% of all future scenarios (i.e. certainty level $\beta = 0.9$). To determine for how long the fund is likely to provide a stable income, a two-step calculation is required.

- First, by Monte Carlo simulation (as described in Section 3.2.1) it is calculated that the highest value of k which satisfies inequality (12) is $k_U = 1\,310$.

  Then, by inequality (13), all members of the fund get an income of at least $0.95C(0)$ up to time $\lfloor T_{(1310)} \rfloor$, in at least 90% of future scenarios. As shown in Section 3.2, the maximal value $k := k_C$ that satisfies inequality (13) is about 2% higher than the maximal value $k := k_U = 1\,310$ that satisfies inequality (12) for $N = 2\,000$, for the considered mortality table. So in reality, $k_C \approx 1\,338$ and thus the stable income is provided for longer than suggested by k_U. Mitigating the small error caused by this under-estimation is that, in the second step described next, the higher time $T_{(1310)}$ at which $1\,310$ deaths have occurred is used, rather than the time $\lfloor T_{(1310)} \rfloor \leq T_{(1310)}$.

- Second, calculate at what time is it likely that exactly $k_U = 1\,310$ deaths have been observed in the initial population of $2\,000$ 70-year-olds. This is the problem of finding t that satisfies $_t q_{70} = 1\,310/2\,000$, in which the constant $_t q_{70}$ is the true probability that a 70-year old dies between age 70 and age $70 + t$. Here, t is called the likely time. It is a straightforward task if one has a life table at hand. For example, assuming that the mortality of each of the $2\,000$ members initially in the fund follows the UK-based life table S1PFL (Continuous Mortality Investigation, 2008) and assuming a uniform distribution of deaths between integer ages, gives that $t = 19.4$ years.

Note that only small variations of the time of death of the k_Uth member are expected. This is because as N becomes larger, the probability density of the kth uniform order statistic tends to a Dirac delta function. To examine how quickly this happens, the probability that the time of the k_Uth death occurs at least $d \geq 0$ time units before its likely time of occurrence t, i.e. $\mathbb{P}[T_{(k_U)} \leq t - d]$ is plotted against d in Figure 6, for different choices of N, ε and β and for one mortality law. Symmetric lower and upper income thresholds are imposed. The figure when only a lower income threshold is imposed is very similar and is omitted.

For an initial number of members $N = 100$ in the fund (the thickest lines in Figure 6), the time of the k_Uth death has a 99% probability of appearing between 2.5 to 3 years below t. For groups with initially $N = 1\,000$ members, there is a 99% probability that the k_Uth death appears at most one year below t. This falls to 0.5 years with $N = 10\,000$ members. The conclusion is that the likely time t calculated in the second step above is very close to the actual time of the k_Uth death when there are at least $1\,000$ members initially in the fund, being at most one year greater than the actual time.

Figure 7 shows the likely time t that the income is stable, for the same mortality law. The time increases with N, since more members mean that idiosyncratic longevity risk is more diversified and thus the income paid from each surviving member's account is stable for longer. Although the

rate of increase slows as N increases, the income is stable for quite a long time once the number of initial members N is at least $2\,000$. For example, when symmetric lower and upper income thresholds are imposed, the likely time for which a stable income can be provided with certainty β increases from 17 years to 25 years as N increases from $2\,000$ to $10\,000$, for $\varepsilon = 0.05$ and $\beta = 0.9$.

Now turn to the case when there is only a lower threshold imposed in the definition of income stability. Starting with $N = 2\,000$ members, the income is at least 95% of the initial income (i.e. $\varepsilon = 0.05$) for nearly 15 years in $\beta = 99\%$ of future states of the world. This lower income stability extends to almost 20 years if the certainty is dropped to $\beta = 90\%$. From Figure 7 it is seen that including both an upper and a lower threshold reduces by, at most, 2 years the length of time for which a fund is likely to provide an income above a lower threshold.

It is also seen from the figure that the threshold parameter ε has a larger effect on the likely times than the certainty β, for each value of N. For example, for a fund with initially $N = 2\,000$ members, decreasing the threshold parameter ε from 10% to 5% reduces the likely time by up to 7 years, everything else being the same. However, decreasing the certainty level β from 99% to 90% reduces the likely time by up to 5 years, everything else being the same. These reductions decline in magnitude as N increases.
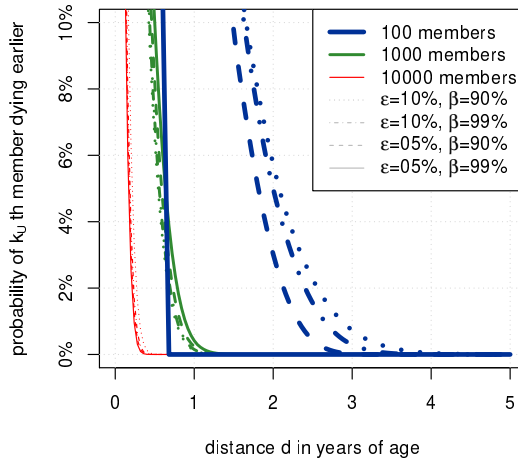

Figure 6: $\mathbb{P}[T_{(k_U)} \leq F^{-1}(k_U/N) - d]$ plotted against d, in which $F(s) = {}_s q_{70}$ is calculated using the UK-based life table S1PFL (Continuous Mortality Investigation, 2008). The plots are shown for a selection of values of the initial number N of 70-year-old members in the fund, certainties β and threshold parameter ε and when a symmetric lower and upper income threshold are imposed. The plot using the Human Mortality Database's life table for the UK for 2016 (Human Mortality Database, 2018) looks very similar and is omitted.
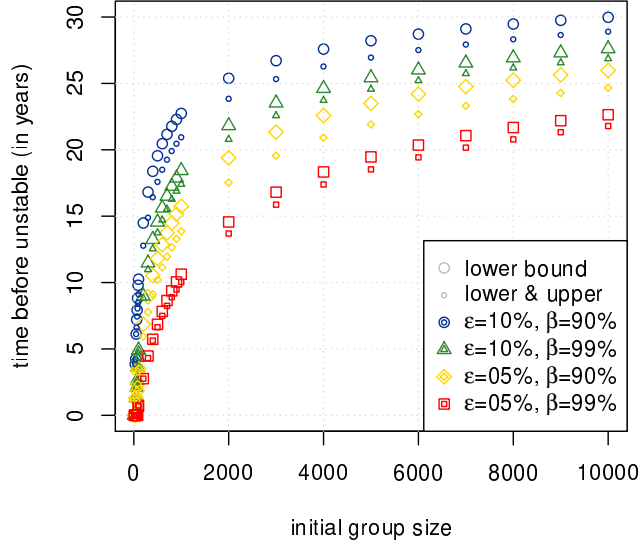
Figure 7: The likely length of time for which the fund can provide a stable income for a group of 70-year-olds, plotted as a function of the initial number of members in the fund N. The plot is based on the UK-based life table S1PFL (Continuous Mortality Investigation, 2008). The smaller symbols indicate a symmetric lower and upper threshold on the income, and the larger symbols indicate a lower threshold only on the income.

Our results also provide evidence to support broadly the statement in Qiao and Sherris (2013, page 967) that "Most of the significant pooling benefits are realized when the pool size reaches $1\,000$"; Figure 7 indicates a slowing up of the rate of increase of the likely time for which a stable income can be provided around $N = 2\,000$ (the study of Qiao and Sherris 2013 considered only $N \in \{1, 100, 1\,000, 10\,000\}$). It is important to note that when the number of members who receive a stable income is considered, there is still a steadily increasing proportion of members who receive a life-long, stable income, as can be observed from Figure 4.

5. Approximate formula

Here an approximate formula to determine the maximum integer $k := k_U$ that satisfies inequality (12) is presented. Fix the initial number of fund members N, certainty $\beta \in (0,1)$, threshold parameter $\varepsilon \in (0,1)$ and let Φ^{-1} be the inverse of the standard normal distribution function Φ. Then

$$k_U \approx k_U^{\mathrm{approx}} := N \left(1 - \left\lfloor \frac{1}{1-\varepsilon} \left(1 - \frac{1}{1 + \frac{1}{N} \left(\frac{1-\varepsilon}{\varepsilon} \right)^2 \left(\Phi^{-1} \left(\frac{1-\beta}{2} \right) \right)^2} \right) \right\rfloor_N \right), \qquad (16)$$

in which $\lfloor u \rfloor_N = \max\{i/N : i/N \leq u \text{ and } i \in \{0, 1, \ldots, N\}\}$ for $u \geq 0$. The derivation is shown in Appendix A.

Figure 8a shows that $1 - k_U^{\mathrm{approx}}/N$ is visually almost indistinguishable from $1 - k_U/N$ on a scale from 0% to 100%. However, the relative errors reveal small systematic discrepancies. Observe in Figure 8b that the relative errors change most noticeably with the threshold parameter ε, upon fixing the certainty β. On the other hand, the relative errors are similar for different certainties β, upon fixing the threshold parameter ε. This indicates that the treatment of ε is the major source

of error. In fact, in the derivation of the approximation (16), a vague argument that involves ε and works by compensating one effect with another is employed; see the expressions (17) and (18).
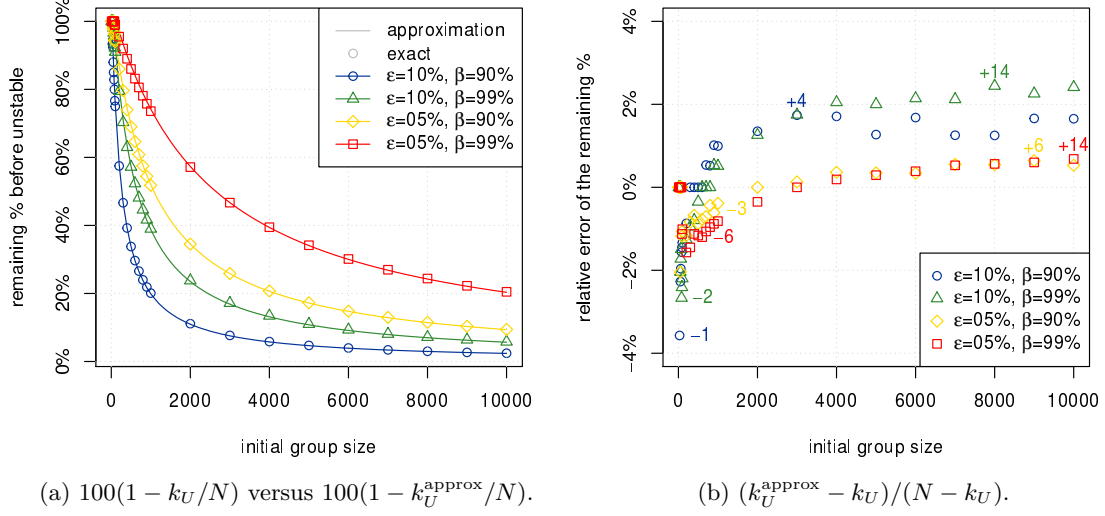

(a) $100(1 - k_U/N)$ versus $100(1 - k_U^{\mathrm{approx}}/N)$.

(b) $(k_U^{\mathrm{approx}} - k_U)/(N - k_U)$.

Figure 8: Representations of the error in the approximation k_U^{approx}. Figure 8a plots the percentage of members still alive after the k_Uth death, i.e. $100(1 - k_U/N)$ and displayed as discrete symbols, and its approximation $100(1 - k_U^{\mathrm{approx}}/N)$, plotted as solid lines, as the initial number of members N increases. The results are shown for various choices of the certainty β and threshold parameter ε. Figure 8b plots the relative error of these values, which reduces to calculating $(k_U^{\mathrm{approx}} - k_U)/(N - k_U)$. The displayed integers inside the graph show the maximal value of $k_U^{\mathrm{approx}} - k_U$ in terms of number of members.

More generally, it is possible to apply the same methodology to find an approximate formula to determine the maximum integer $k := k_U$ that satisfies

$$\mathbb{P}\left[(1 - \varepsilon_1)\frac{i-1}{N} + \varepsilon_1 \geq U_{(i)} \geq (1 + \varepsilon_2)\frac{\min\{i, N-1\}}{N} - \varepsilon_2 \text{ for all } i \in \{1, 2, \ldots, k\}\right] \geq \beta.$$

However, the result involves a special function that is as inexplicit as the target function, and we give it only for completeness:

$$k_U \approx k_U^{\mathrm{approx}} := N \lfloor \Psi_{\varepsilon_1, \varepsilon_2, N}^{-1}(\beta) \rfloor_N$$

with

$$\Psi_{\varepsilon_1, \varepsilon_2, N}(y) := \mathbb{P}\left[-\sqrt{N}\frac{\varepsilon_2}{1+\varepsilon_2} \leq \inf_{s \leq \frac{1}{(1+\varepsilon_2)(1-y)} - 1} W(s); \sup_{s \leq \frac{1}{(1-\varepsilon_1)(1-y)} - 1} W(s) \leq \sqrt{N}\frac{\varepsilon_1}{1-\varepsilon_1}\right]$$

for all $y \in (0, 1)$ and W a standard Brownian motion.

6. Conclusion

Here a pooled annuity fund has been analyzed to see how well it provides a stable income to its participants, with a specified degree of certainty. Income stability is defined path-wise. The main theoretical result is the derivation of a lower bound that can be used as a good approximation to the true number of members. The lower bound gives both insights and a greater understanding of the pooled annuity fund's ability to diversify idiosyncratic longevity risk:

- It is independent of the mortality distribution of the fund members. As the lower bound is close to the true value, it suggests that the true number of members is unlikely to change much if the mortality distribution is varied, all other parameters being unchanged.

- It can be calculated faster than the true number, since the former requires only the simulation of the order statistics of independent, standard uniform random variables rather than a time-dependent stochastic simulation.

- When a stable income is defined as the income lying between a symmetric upper and non-zero lower income threshold, the longest lived members do not receive a stable income in any future state of the world. Even if such members may receive a stable income for the majority of their lifetime, this means that there should be an alternative method of providing a stable income with a desired degree of certainty for the entirety of their lifetime, for example by using deferred life annuity contracts or the opportunity to exit the fund at a certain age or the modifications proposed in, for example, Chen et al. (2019), Chen and Rach (2019) and Donnelly and Young (2017). This is an insight into the failure of the diversification of idiosyncratic longevity risk for the longest lived members.

The main theorem quantifies the number of members required to give the desired degree of income stability. In other words, the number of members needed to diversify idiosyncratic longevity risk to a specific degree. At a high-level, the results suggest that the membership of the fund should number in the thousands if it is important to reduce the income instability derived from longevity risk pooling to low levels. Our finding complements the results of Qiao and Sherris (2013), who suggest the same order of magnitude based on examining the quantiles of the income. However, our results allow the calculation of the precise number of members required and, moreover, use a path-wise definition of income stability. This means that we know how the income streams have behaved up to the time that the calculated number of people die, rather than only how they behave at one point in time.

Applying the theoretical results, the length of time for which the pooled annuity fund could provide a stable income with a specified degree of certainty was determined. Again, the length of time depends on how income stability is specified. It also depends on a mortality law; the results for one law are presented. Based on the chosen mortality law, the fund can provide a stable income from 10 years to 30 years depending on the specification of income stability and the initial number of members in the fund. Finally, an approximation to the lower bound is derived, and is found to be a close approximation to it.

The limitations of the assumptions made in this paper motivate future research into this area. It is assumed that all annuitants join with the same amount of money and are the same age. While it would be possible to impose such restrictions in real life, it is overly prescriptive and makes the continued success of the pooled annuity fund less likely. All annuitants are assumed to have the same chance of dying at each age. How well this approximates reality will depend on the group of annuitants; it would be unrealistic if anyone was allowed to join but may be appropriate for a fund offered only to university professors, for example. Deaths are assumed to occur independently, but this may not be true if annuitants catch a deadly virus from each other. These restrictions on the characteristics of the fund membership could be relaxed in future research.

The historical record of actuaries under-estimating future lifetimes and the fast spread across the world of the coronavirus both suggest that the distribution of the annuitants' future lifetimes assumed in the model is unlikely to be observed in practice. How the inclusion of systematic longevity risk affects the results very broadly is discussed briefly here. Suppose that the day after

the fund started, the future lifetime distribution of the annuitants was found to understate for how long the annuitants would live. Assume for simplicity that the definition of stability includes only a lower income threshold.

While the income paid to the annuitants would be consequently revised downwards, the number k_U who would receive the stable version of this new income for life would be unchanged (assuming no further updates to the future lifetime distribution are needed). This is because k_U is independent of the choice of the future lifetime distribution, but it requires that this distribution is the one observed in practice. This would be the case if the distribution changed the day after the fund started, and never changed subsequently.

However, suppose that the understatement of future lifetime was not discovered while the fund's annuitants were alive. Then the income paid to the annuitants would be too high and would gradually decline, as fewer people died than expected meaning that too much money was paid out. Thus fewer than k_U people would receive a stable, lifelong income for some given level of certainty since the declining income would breach the lower income threshold earlier than anticipated.

If instead more people died than suggested by the future lifetime distribution used to calculate the annuity payments, then less income is paid out than expected. In that case, more than k_U people would receive a stable, lifelong income, because more people die earlier - thus getting a lifelong income over a shorter lifetime - and there is more money for the longer-lived.

As the focus of this paper is idiosyncratic longevity risk, the quantification of the effect of systematic longevity risk on the results requires a further study. This is important so that how the annuitants' income may vary over their lifetime is understood better, thus allowing the pooled annuity fund to meet the needs and expectations of annuitants.

Acknowledgments

References

Bengen, W. (2001). Conserving client portfolios during retirement, Part IV. *Journal of Financial Planning*, 5:110–119.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley, 3rd edition.

Birnbaum, Z. and Lientz, B. (1969). Exact distributions for some Rényi-type statistics. *Applicationes Mathematicae*, 10:179–192.

Bruhn, K. and Steffensen, M. (2013). Optimal smooth consumption and annuity design. *Journal of Banking and Finance*, 37:2693–2701.

Chen, A., Hieber, P., and Klein, J. K. (2019). Tonuity: A novel individual-oriented retirement plan. *ASTIN Bulletin*, 49(1):5–30.

Chen, A. and Rach, M. (2019). Options on tontines: An innovative way of combining tontines and annuities. *Insurance: Mathematics and Economics*, 89:182–192.

Constantinides, G. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy*, 98:519–543.

Continuous Mortality Investigation (2008). S1PFL. Technical report, Institute and Faculty of Actuaries, UK. Data downloaded at https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/cmi-mortality-and-morbidity-tables/s1-series-tables on 28 Sept 2019.

Csörgö, M. (1965). Exact probability distribution functions of some Rényi type statistics. *Proceedings of the American Mathematical Society*, 16(6):1158–1167.

Curatola, G. (2017). Optimal portfolio choice with loss aversion over consumption. *The Quarterly Review of Economics and Finance*, 66:345–358.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag New York.

Donnelly, C. (2015). Actuarial fairness and solidarity in pooled annuity funds. *ASTIN Bulletin*, 45(1):49–74.

Donnelly, C., Guillén, M., and Nielsen, J. (2014). Bringing cost transparency to the life annuity market. *Insurance: Mathematics and Economics*, 56:14–27.

Donnelly, C. and Young, J. (2017). Product options for enhanced retirement income. *British Actuarial Journal*, 22(3):636–656.

Donsker, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 23(2):277–281.

Forman, J. and Sabin, M. (2015). Tontine pensions. *University of Pennsylvania Law Review*, 163(3):755–831.

Guyton, J. (2004). Decision rules and portfolio management for retirees: is the 'safe' initial withdrawal rate too safe? *Journal of Financial Planning*, pages 54–62. October 2004.

Guyton, J. and Klinger, W. (2006). Decision rules and maximum initial withdrawal rates. *Journal of Financial Planning*, pages 50–58.

He, L. and Liang, Z. (2013). Optimal dynamic asset allocation strategy for ELA scheme of DC pension plan during the distribution phase. *Insurance: Mathematics and Economics*, 52(2):404–410.

Human Mortality Database (2018). United Kingdom, Life tables (period 1x1), Total (both sexes). Technical report, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Data downloaded at https://www.mortality.org on the 28/09/2019.

Karatzas, I. and Shreve, S. (1988). *Brownian motion and stochastic calculus*. Springer-Verlag.

Kerman, J. (2011). A closed-form approximation for the median of the Beta distribution. Available at arXiv. https://arxiv.org/pdf/1111.0433.

Milevsky, M. and Salisbury, T. (2015). Optimal retirement income tontines. *Insurance: Mathematics and Economics*, 64:91–105.

Milevsky, M. A. and Salisbury, T. S. (2016). Equitable retirement income tontines: Mixing cohorts without discriminating. *ASTIN Bulletin*, 46(3):571–604.

Munk, C. (2008). Portfolio and consumption choice with stochastic investment opportunities and habit formation in preference. *Journal of Economic Dynamics and Control*, 32(11):3560–3589.

Pfau, W. and Kitces, M. (2014). Reducing retirement risk with a rising equity glide path. *Journal of Financial Planning*, 27:38–45.

Piggott, P., Valdez, E., and Detzel, B. (2005). The simple analytics of a pooled annuity fund. *The Journal of Risk and Insurance*, 72(3):497–520.

Qiao, C. and Sherris, M. (2013). Managing systematic mortality risk with group self-pooling and annuitization schemes. *Journal of Risk and Insurance*, 80(4):949–974.

Sabin, M. (2010). Fair tontine annuity. Available at SSRN. http://dx.doi.org/10.2139/ssrn.1579932.

Shorack, G. and Wellner, J. (2009). *Empirical Processes with Applications to Statistics*. SIAM.

Stamos, M. Z. (2008). Optimal consumption and portfolio choice for pooled annuity funds. *Insurance: Mathematics and Economics*, 43:56–68.

Valdez, E. A., Piggott, J., and Wang, L. (2006). Demand and adverse selection in a pooled annuity fund. *Insurance: Mathematics and Economics*, 39(2):251–266.

Van Bilsen, S., Laeven, R., and Nijman, T. (2020). Consumption and portfolio choice under loss aversion and endogenous updating of the reference level. *Management Science*. In press.

Vrbik, J. (2018). Small-sample corrections to Kolmogorov-Smirnov test statistic. *Pioneer Journal of Theoretical and Applied Statistics*, 15(1-2):15–23.

Work and Pensions Committee (2018). Pension freedoms. Technical report, House of Commons (United Kingdom). Ninth Report of Session 2017-19, HC917.

Appendix

A. Derivation

Claim A.1. *For the order statistics $(U_{(i)})_{i=1}^N$ of N independent and standard uniformly distributed random variables, certainty $\beta \in [0,1]$, and threshold parameter $\varepsilon \in (0,1)$, let $k_U \leq N$ be the last integer that fulfills*

$$\mathbb{P}\left[(1-\varepsilon)\frac{i-1}{N} + \varepsilon \geq U_{(i)} \ \text{ for all } i \in \{1,2,\ldots,k\}\right] \geq \beta$$

For $u \geq 0$, let $\lfloor u \rfloor_N = \max\{i/N : i/N \leq u \ \text{ and } \ i \in \{0,1,\ldots,N]\}\}$, and let Φ^{-1} be the inverse of the standard normal distribution function Φ. Then k_U can be calculated approximately from

$$1 - \frac{k_U}{N} \approx \left\lfloor \frac{1}{1-\varepsilon}\left(1 - \frac{1}{1 + \frac{1}{N}\left(\frac{1-\varepsilon}{\varepsilon}\right)^2 \left(\Phi^{-1}\left(\frac{1-\beta}{2}\right)\right)^2}\right)\right\rfloor_N .$$

Derivation of claim. Assume that both the time of deaths of members and payment times to surviving members are reasonably dense in time so that from (6) in the proof of Theorem 3.1,

$$\mathbb{P}\left[\inf_{s \in [0, T_{(k_U)})} \frac{{}_s p_x}{{}_s \hat{p}_x} \geq 1 - \varepsilon\right] \geq \beta$$

is approximately equivalent to

$$\mathbb{P}\left[\inf_{s \in [0,t]} \frac{{}_s p_x}{{}_s \hat{p}_x} \geq 1 - \varepsilon\right] \geq \beta, \quad \text{with } t \approx T_{(k)}.$$

Using $\hat{F}_N(s) = 1 - {}_s \hat{p}_x$ and $F(s) = 1 - {}_s p_x$ from the proof of Theorem 3.1,

$$\beta = \mathbb{P}\left[\frac{{}_s p_x}{{}_s \hat{p}_x} \geq 1 - \varepsilon, \ \forall s \leq t\right] = \mathbb{P}\left[\frac{1 - F(s)}{1 - \hat{F}_N(s)} \geq 1 - \varepsilon, \ \forall s \leq t\right]$$

$$= \mathbb{P}\left[\frac{\varepsilon}{1-\varepsilon} \geq \frac{F(s) - \hat{F}_N(s)}{1 - F(s)}, \ \forall s \leq t\right].$$

It is more convenient to look at the complementary probability, i.e.

$$1 - \beta = \mathbb{P}\left[\exists s \leq t : \ \frac{\varepsilon}{1-\varepsilon} < \frac{F(s) - \hat{F}_N(s)}{1 - F(s)}\right] = \mathbb{P}\left[\frac{\varepsilon}{1-\varepsilon} < \sup_{s \leq t} \frac{F(s) - \hat{F}_N(s)}{1 - F(s)}\right].$$

Let F^{-1} be the generalized inverse of F. Re-writing the above expression in a form amenable to applying well-known results from the literature yields

$$1 - \beta = \mathbb{P}\left[\frac{\varepsilon}{1-\varepsilon} < \sup_{u \leq F(t)} \frac{u - \hat{F}_N(F^{-1}(u))}{1 - u}\right]. \tag{17}$$

One of the early versions of Donsker's theorem states that the process $u \mapsto \sqrt{N}(u - \hat{F}_N(F^{-1}(u)))$ converges in distribution for $N \to \infty$ to a Brownian bridge (Donsker, 1952). However, applying this theorem to (17) yields an approximation that noticeably underestimates the true value. The issue seems to be that the discontinuous jump-processes $(\hat{F}_N)_{N=1}^\infty$ are approximated by a continuous Brownian bridge. Heuristically, a continuous process should be a better fit for a continuous process than a discontinuous one. In fact, the classical Donsker theorem is based on continuous interpolations of random variables (Billingsley, 1999).

Let $\hat{F}_n^c$ be the process that continuously interpolates the jumps of $\hat{F}_n$ for given n. Then $\hat{F}_n^c \geq \hat{F}_n$ are functions on the real line. This leads to a smaller supremum and, in turn, to an under-estimation of (17) when $\hat{F}_n$ is replaced with its continuous version $\hat{F}_n^c$ in (17).

One way to compensate for this under-estimation when using the Brownian bridge is to enlarge the set $\{u : u \leq F(t)\}$. A similar argument has been used in Vrbik (2018) for small sample corrections in the context of the Kolmogorov-Smirnov test. Interestingly, there is an enlargement that leaves the value of (17) almost unchanged: let y be the value of the last jump of $\hat{F}_N(F^{-1}(u))$ in $\{u : u \leq F(t)\}$. Assume that $\varepsilon/(1-\varepsilon)$ hasn't been crossed so far (the other case doesn't change the outcome of the inequality in (17) for any enlargement). Then

$$\frac{\varepsilon}{1-\varepsilon} \geq \frac{u-y}{1-u} \quad \text{for } u \leq F(t).$$

Rearranging the inequality and using $y \approx F(t)$ gives

$$\varepsilon + (1-\varepsilon)y \geq u, \qquad \text{or} \qquad \varepsilon + (1-\varepsilon)F(t) \geq u.$$

Hence, $\{u : u \leq F(t)\}$ can be approximately enlarged to $\{u : u \leq \varepsilon + (1-\varepsilon)F(t)\}$ without changing the value in (17).

Overall, let B be a Brownian bridge in $[0,1]$, then

$$1 - \beta \approx \mathbb{P}\left[\frac{\varepsilon}{1-\varepsilon} < \sup_{u \leq \varepsilon + (1-\varepsilon)F(t)} \frac{B(u)}{\sqrt{N}(1-u)}\right]. \tag{18}$$

The process $u \mapsto B(u)/(1-u)$ is a time-changed Brownian motion, as it is a continuous, zero-mean, Gaussian process with covariance function $(u,v) \mapsto u/(1-u)$ for $0 \leq u \leq v \leq 1$ (Karatzas and Shreve, 1988, pp.103-104).

Then, letting W be a Brownian motion,

$$1 - \beta \approx \mathbb{P}\left[\frac{\varepsilon}{1-\varepsilon} < \sup_{u \leq \varepsilon + (1-\varepsilon)F(t)} \frac{W(\frac{u}{1-u})}{\sqrt{N}}\right] = 2\,\mathbb{P}\left[\sqrt{N}\frac{\varepsilon}{1-\varepsilon} < W(\tfrac{\varepsilon+(1-\varepsilon)F(t)}{1-(\varepsilon+(1-\varepsilon)F(t))})\right]$$

$$= 2\,\Phi\left(-\sqrt{\tfrac{1-(\varepsilon+(1-\varepsilon)F(t))}{\varepsilon+(1-\varepsilon)F(t)}}\sqrt{N}\frac{\varepsilon}{1-\varepsilon}\right)$$

$$= 2\,\Phi\left(-\sqrt{\tfrac{1}{1-(1-\varepsilon)_t p_x}-1}\sqrt{N}\frac{\varepsilon}{1-\varepsilon}\right).$$

Rearranging yields

$$_t p_x \approx \frac{1}{1-\varepsilon}\left(1 - \frac{1}{1 + \frac{1}{N}\left(\frac{1-\varepsilon}{\varepsilon}\right)^2\left(\Phi^{-1}\left(\frac{1-\beta}{2}\right)\right)^2}\right).$$

Observe that $_t p_x \approx {}_t \hat{p}_x \approx {}_{T_{(k_U)}}\hat{p}_x = L_{x+T_{(k_U)}}/N = (N-k_U)/N = 1 - k_U/N$. Hence, the claim follows from $_t p_x \approx 1 - k_U/N$ together with $k_U \leq N$ is an integer. $\qquad\square$