Insurance: Mathematics and Economics 111 (2023) 214–229

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Actuarial fairness and social welfare in mixed-cohort tontines

An Chen [a], Manuel Rach [b,*]

[a] Institute of Insurance Science, Ulm University, Helmholtzstr. 20, 89069 Ulm, Germany
[b] Institute of Insurance Economics, University of St. Gallen, Tannenstrasse 19, 9000 St. Gallen, Switzerland

ARTICLE INFO

Article history:
Received September 2022
Received in revised form May 2023
Accepted 8 May 2023
Available online 15 May 2023

JEL classification:
G22
G52
H31
J32

Keywords:
Tontine
Actuarial fairness
Participation rates
Age heterogeneity
Risk aversion

ABSTRACT

We study actuarial fairness in tontines with heterogeneous cohorts. For a given tontine, we show that both collective and individual actuarial fairness can be achieved. While it is impossible to design a tontine scheme with mixed cohorts that is optimal (utility-maximizing) for each single cohort (Milevsky and Salisbury (2016), Chen et al. (2021d)), we design a socially optimal tontine that maximizes the collective's expected utility which is characterized through a weighted sum of individual utility functions. In particular, we compare the resulting collectively optimal tontine to existing schemes in the literature, and identify similarities and differences as well as potential (dis-)advantages.

1. Introduction

Societal challenges with ageing have generated a growing stream of literature on innovative retirement products providing random, mortality-linked benefits to pensioners. Such products are known as pooled annuity funds, group self-annuitization schemes and tontines (cf. Piggott et al. (2005), Sabin (2010), Donnelly et al. (2014), Milevsky and Salisbury (2015)). Although they carry different names, all these products have the similarity that a pool of pensioners shares mortality risks. Given a sufficiently large pool, the idiosyncratic mortality risk can be diversified, while the pool shares the systematic mortality risk. Among the main challenges of these products lies the fact that pensioners with different age and wealth typically cannot be simply joint in one and the same scheme without discriminating at least some pensioners. While the majority of the literature in this field assumes a homogeneous pool for analytical convenience (e.g. Milevsky and Salisbury (2015), Chen et al. (2019, 2020)), there is also some literature dealing with heterogeneous cohorts (e.g. Sabin (2010), Donnelly et al. (2014), Milevsky and Salisbury (2016), Denuit (2019) and Chen et al. (2021d)). Milevsky and Salisbury (2016) come up with the concept *equitability*, a weaker concept than fairness, which states that each individual loses an identical fraction of their wealth if participating in the tontine with mixed cohorts. Chen et al. (2021d) determine the optimal payment design for a given policyholder in a tontine with mixed cohorts, and do not find a unique withdrawal rate which satisfies the fairness condition simultaneously for all the individuals. In contrast to the equitable result obtained in Milevsky and Salisbury (2016), we show in this article that both collective and individual actuarial fairness are achievable in a closed scheme applying the heterogeneous cohorts of Milevsky and Salisbury (2016). Due to the inadequate documentation of actuarial fairness for tontines with mixed cohorts, there is hardly literature addressing optimal tontines with mixed cohorts in terms of utility maximization. In the present paper, we come up with an optimal tontine scheme designed by a social planner, which takes account of various risk aversion levels of all the participants. We determine this socially optimal tontine in such a way that it ensures at least collective fairness and compare it to existing schemes in the literature.

* Corresponding author.
E-mail addresses: an.chen@uni-ulm.de (A. Chen), manuel.rach@unisg.ch (M. Rach).

https://doi.org/10.1016/j.insmatheco.2023.05.002

Actuarially fair annuities are typically considered to be the optimal source of retirement income (Yaari (1965)). However, the inclusion of safety loadings in annuity premiums opens a market for mortality risk-carrying products like tontines, which require no safety loadings because providers do not promise guaranteed benefits (cf. Piggott et al. (2005), Stamos (2008), Sabin (2010), Hanewald et al. (2013), Donnelly et al. (2014) and Milevsky and Salisbury (2015)). Thus, if tontines can be designed fairly for each participant, they could be offered at actuarially fair prices, which represents a huge advantage for any retiree who is not fond of full annuitization.[1] We therefore believe that it is important to extend the article Milevsky and Salisbury (2016) for mixed cohorts to actuarially fair tontines for heterogeneous groups.

In this article, we distinguish between collective and individual fairness. In the latter one, we postulate that, for each policyholder, the expected present value of future benefits shall be equal to the single up-front premium paid by this policyholder. Collective fairness is a weaker requirement as it only requires the present value of future benefits that the pool receives to be equal to the sum of the initial up-front premiums. Hence, it is clear that individual fairness implies collective fairness. We extend the model setup in Milevsky and Salisbury (2016) by including systematic mortality risk and prove that the design they propose is not collectively fair, which prohibits it from being individually fair for all the individuals. Consequently, we propose an alternative way to design the withdrawal rate which ensures collective fairness. Furthermore, we then derive conditions under which individual fairness can be achieved.

Although it is possible to design a tontine for heterogeneous groups which even ensures individual fairness for each group, this tontine cannot maximize the expected discounted utility of each group simultaneously (see also Milevsky and Salisbury (2016) and Chen et al. (2021d). In the second part of this paper, we design a tontine for heterogeneous groups that is optimal for the collective of policyholders. For this, we take a social planner's viewpoint and maximize the weighted sum of the individual utility functions to achieve a withdrawal rate which is at least optimal for the collective of heterogeneous policyholders.[2] The resource constraint of the social planner is the collective fairness criterion. In our setting, we can choose, in the second step, appropriate participation rates for each group to ensure individual fairness. If all the participants use logarithmic utility to describe their risk preferences (relative risk aversion all equal to 1), we are able to determine the optimal tontine withdrawal rate analytically and can determine individually fair participation rates for the given withdrawal rate numerically. In contrast, we need to rely on numerical procedures to determine the optimal tontine withdrawal rate along with individually fair participation rates when policyholders from various cohorts own different levels of relative risk aversion.

We compare the socially optimal withdrawal rate of tontines with mixed cohorts to those with homogeneous cohorts, and to the proportional and natural tontines introduced by Milevsky and Salisbury (2016). Both the proportional and natural tontine are generalizations of the natural single-cohort tontine introduced in Milevsky and Salisbury (2015) which delivers constant retirement benefits if mortality evolves as expected. We find that all these tontine models yield approximately identical certainty equivalents under logarithmic utility. However, under heterogeneous power utility preferences, we find that the social planner-designed tontine and the single-cohort tontine outperform the proportional and natural tontines. This is because the natural and proportional tontines are unable to account for the degree of relative risk aversion of policyholders. However, our numerical results also suggest that the proportional and natural tontines perform better than the social planner's tontine when the ages of the cohorts differ significantly. So, among the parameters considered, the natural and proportional tontines perform better for mixed-cohort tontines with large age differences, while the social planner's tontine performs better when age differences are closer and utility preferences differ greatly from log utility. In contrast to Milevsky and Salisbury (2016), we find that the benefits of pooling heterogeneous cohorts do not significantly exceed those of individually utility-maximizing withdrawal rates in single-cohort tontines. In conclusion, we find that fair participation rates for reasonable tontine designs are roughly identical and can be approximated by the ratio of annuity factors. The reason is that participation rates measure differences between age groups for which annuity prices are at least a good approximation (see also Milevsky and Salisbury (2016)).

The remainder of this article is structured as follows: In Section 2, we consider a tontine with a given withdrawal rate and derive conditions under which such a tontine is collectively and individually fair. In Section 3, we then solve the collective optimization problem of the social planner and analyze the expected discounted lifetime utility of the heterogeneous policyholders. In Section 4, we conclude. Some proofs and technical details as well as a pseudo code are collected in the appendix.

2. Achieving fairness in given tontines

In this section, it is our goal to derive conditions under which fairness in mixed-cohort tontines can be achieved, if withdrawal rates are assumed to be given. In particular, it is not the goal of this section to derive individually utility-maximizing withdrawal rates as, for instance, done in Milevsky and Salisbury (2015) and Chen et al. (2019, 2021d).

2.1. Model setup

We consider L cohorts that differ in initial wealth and age. We denote the initial size of cohort $j \in \{1, \dots, L\}$ by n_j, the age of the members in cohort j by x_j and the initial wealth of a member of cohort j by w_j. We assume that the members in each cohort are identical copies of each other. The total initial pool size is then $n = n_1 + \cdots + n_L$.

The remaining lifetime of policyholder i will be denoted by T_i for $i = 1, \dots, n$. The (possibly stochastic) force of mortality of each member is μ_{x+t}, i.e. all the members are subject to the same mortality law (but still differ in their ages). We use $\mathcal{F}_t := \sigma(\{\mu_{x+s}\}_{s \le t})$ to denote the sigma-algebra containing the information regarding the systematic mortality risk up to time t. Furthermore, we introduce the notation

$$S_x(t) := \mathbb{E}\left[\mathbb{1}_{\{T>t\}} \mid \mathcal{F}_t\right] = e^{-\int_0^t \mu_{x+s} \mathrm{d}s}$$

[1] Naturally, there will still be some fees (e.g. for the management of investments, distribution of dividends, the organization of a tontine, and providing regular information to policyholders), but these will be small compared to annuities (cf. Chen et al. (2021a)).

[2] This type of collective utility function is frequently considered in the finance literature, see e.g. Wilson (1968), Dumas (1989), Weinbaum (2009) and Jensen and Nielsen (2016).

for the random survival probabilities conditional on the systematic mortality outcome. The (deterministic) survival probabilities are then given by

$$s_x(t) := \mathbb{E}\left[\mathbb{1}_{\{T>t\}}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{T>t\}} \mid \mathcal{F}_t\right]\right] = \mathbb{E}\left[e^{-\int_0^t \mu_{x+s}\mathrm{d}s}\right].$$

The number of policyholders alive at time t is then given by

$$N(t) = \sum_{j=1}^{n} \mathbb{1}_{\{T_j>t\}},$$

where $(\mathbb{1}_{\{T_j>t\}} \mid \mathcal{F}_t) \sim \mathrm{Ber}(S_{x_j}(t))$. We use $N_i(t)$ to denote the number of living members in cohort i, such that $N(t) = \sum_{i=1}^{L} N_i(t)$. We assume that the remaining lifetimes of individuals are conditionally (on $\mathcal{F}_t$) independent. Conditional on $\mathcal{F}_t$, the overall number of living policyholders $N(t)$ then follows a Poisson Binomial distribution.

Following Milevsky and Salisbury (2016), we define the payoff to an individual policyholder by

$$b^{(i)}(t) := wd(t) \cdot \frac{\pi_i w_i}{\sum_{j=1}^{n} \pi_j w_j \mathbb{1}_{\{T_j>t\}}} \mathbb{1}_{\{T_i>t\}} \tag{1}$$

where $d(t)$ is a deterministic withdrawal rate specified at the beginning of the contract and $w = \sum_{i=1}^{n} w_i$ is the total initial wealth from all the cohorts. π_i is the so-called participation rate, or, in other words, $1/\pi_i$ can be interpreted as the share price for an individual i to participate in the tontine product. The main purpose of this parameter is to arrive at a higher level of fairness among heterogeneous policyholders, particularly those with different ages. Thinking of policyholders with different ages, with all the other parameters being identical, roughly speaking, the older shall be entitled to higher tontine payments, as the entire period during which they obtain payoffs is expected to be shorter. In this sense, the participation level for older policyholders shall be higher, or equivalently, the share price for older policyholders shall be lower. The quantity $\pi_i w_i$ can be considered as the number of shares of individual i.

In the remainder of this section, we will assume that $d(t)$ is a given withdrawal rate and analyze under which conditions fair participation rates π_i exist and whether they are unique.

2.2. Collective and individual fairness

In this article, we disregard financial market risk and focus exclusively on mortality risk. Moreover, we follow Stamos (2008), Hanewald et al. (2013) and Milevsky and Salisbury (2015) and neglect safety margins in tontines, as they can be neglected for sufficiently large pools even in the presence of systematic mortality risk (cf. Chen et al. (2019)).[3] Let r be the risk-free interest rate. The actuarially fair premium for an individual j in a given cohort can then be computed as the expected present value of the benefits:

$$\begin{aligned} P_0^j &= \mathbb{E}\left[\int_0^\infty e^{-rt} wd(t) \frac{\pi_j w_j}{\sum_{i=1}^{n} \pi_i w_i \mathbb{1}_{\{T_i>t\}}} \mathbb{1}_{\{T_j>t\}} \mathrm{d}t\right] \\ &= \int_0^\infty e^{-rt} \mathbb{E}\left[wd(t) \frac{\pi_j w_j}{\sum_{i=1}^{n} \pi_i w_i \mathbb{1}_{\{T_i>t\}}} \mathbb{1}_{\{T_j>t\}}\right] \mathrm{d}t \\ &= \int_0^\infty e^{-rt} wd(t) \mathbb{E}\left[S_{x_j}(t) \mathbb{E}\left[\frac{\pi_j w_j}{\sum_{i=1}^{n} \pi_i w_i \mathbb{1}_{\{T_i>t\}}} \;\middle|\; \mathcal{F}_t, T_j > t\right]\right] \mathrm{d}t. \end{aligned} \tag{2}$$

Using (2), we can now specify the term "fairness".

Definition 2.1. We distinguish between collective and individual fairness:

- Individual fairness is defined for a member in cohort $j \in \{1, \ldots, L\}$ if $w_j = P_0^j$.
- Collective fairness is defined as

$$w = \sum_{i=1}^{n} P_0^i.$$

If individual fairness holds for all the individuals in the mixed cohorts, it implies collective fairness. However, the reverse statement is not valid. Milevsky and Salisbury (2016) have shown that their design is not collectively fair (and thus cannot be individually fair for all policyholders). For the sake of completeness, we want to briefly show that their design is not collectively fair in our slightly generalized setting and, particularly, discuss the main assumption responsible for this result.

[3] Note that, in addition to the findings in Chen et al. (2019), intuitively there is no need for a specific safety margin with respect to systematic mortality risk in a tontine, since (almost) exclusively policyholders, not providers, bear this risk.

Proposition 2.2. *Let $\pi_i > 0$ and $w_i > 0$ for all $i = 1, \dots, n$. Following Milevsky and Salisbury (2016), i.e. particularly applying the assumption used in their paper*

$$\int_0^\infty e^{-rt} d(t) \mathrm{d}t = 1,$$

the mixed-cohort payoff design in (1) is not collectively fair.

Proof. First, note that the premium charged from the collective, or in other words, the sum of the individual premiums, is given by

$$\begin{aligned}\sum_{j=1}^n P_0^j &= \sum_{j=1}^n \int_0^\infty e^{-rt} \mathbb{E}\left[wd(t) \frac{\pi_j w_j}{\sum_{i=1}^n \pi_i w_i \mathbb{1}_{\{T_i > t\}}} \mathbb{1}_{\{T_j > t\}} \right] \mathrm{d}t \\ &= \int_0^\infty e^{-rt} wd(t) \sum_{j=1}^n \mathbb{E}\left[\frac{\pi_j w_j}{\sum_{i=1}^n \pi_i w_i \mathbb{1}_{\{T_i > t\}}} \mathbb{1}_{\{T_j > t\}} \right] \mathrm{d}t.\end{aligned}$$

Here, it holds

$$\begin{aligned}&\sum_{j=1}^n \mathbb{E}\left[\frac{\pi_j w_j}{\sum_{i=1}^n \pi_i w_i \mathbb{1}_{\{T_i > t\}}} \mathbb{1}_{\{T_j > t\}} \right] = \mathbb{E}\left[\sum_{j=1}^n \frac{\pi_j w_j}{\sum_{i=1}^n \pi_i w_i \mathbb{1}_{\{T_i > t\}}} \mathbb{1}_{\{T_j > t\}} \right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^n \frac{\pi_j w_j}{\sum_{i=1}^n \pi_i w_i \mathbb{1}_{\{T_i > t\}}} \mathbb{1}_{\{T_j > t\}} \,\middle|\, \mathcal{F}_t \right] \right] \\ &= \mathbb{E}\left[\mathbb{P}(N(t) > 0 \mid \mathcal{F}_t) \mathbb{E}\left[\sum_{j=1}^n \frac{\pi_j w_j}{\sum_{i=1}^n \pi_i w_i \mathbb{1}_{\{T_i > t\}}} \mathbb{1}_{\{T_j > t\}} \,\middle|\, \mathcal{F}_t, N(t) > 0 \right] \right] + \mathbb{E}\left[\mathbb{P}(N(t) = 0 \mid \mathcal{F}_t) \cdot 0 \right] \\ &= \mathbb{E}\left[\mathbb{P}(N(t) > 0 \mid \mathcal{F}_t) \right] \\ &= \mathbb{E}\left[1 - \prod_{j=1}^n (1 - S_{x_j}(t)) \right].\end{aligned}$$

It has then the consequence

$$\begin{aligned}\sum_{j=1}^n P_0^j &= \int_0^\infty e^{-rt} wd(t) \mathbb{E}\left[1 - \prod_{j=1}^n (1 - S_{x_j}(t)) \right] \mathrm{d}t \\ &< \int_0^\infty e^{-rt} wd(t) \mathrm{d}t = w. \end{aligned} \tag{3}$$

The inequality in (3) shows that the fairness does not hold in the mixed-cohort tontine on the collective level. □

The assumption discussed in Proposition 2.2 implies that the withdrawal rate $d(t)$ is chosen in such a way that the tontine provider makes payments up to an infinite time horizon, independent of whether there are living policyholders left. Or in other words, it can be understood that Milevsky and Salisbury (2016) assume that the pool contains at least one policyholder who lives forever. In the following, we will slightly adjust this assumption and discuss under what conditions collective fairness can be achieved. We slightly adjust the constraint on the withdrawal rate $d(t)$ to account for the pool size eventually dropping to zero. For this, we introduce the notation

$$\begin{aligned}P_0 &:= \mathbb{E}\left[\int_0^\infty e^{-rt} d(t) \mathbb{1}_{\{N(t) > 0\}} \mathrm{d}t \right] \\ &= \int_0^\infty e^{-rt} d(t) \mathbb{E}\left[1 - \prod_{j=1}^n (1 - S_{x_j}(t)) \right] \mathrm{d}t,\end{aligned}$$

where

$$\mathbb{P}\left(N(t) > 0\right) = \mathbb{E}\left[1 - \prod_{j=1}^n (1 - S_{x_j}(t)) \right],$$

i.e. we can directly observe from the expression P_0 that payments are made to policyholders as long as at least one policyholder is alive. In particular,

$$wP_0 = \int_0^\infty e^{-rt} wd(t)\mathbb{P}\left(N(t) > 0\right) \mathrm{d}t$$

denotes the expected present value of payments that the tontine provider pays out to the policyholders. In expectation, the premiums collected will perfectly finance the retirement benefits of the pool. However, if the actual mortality deviates from the expectation, the tontine provider can make a gain or suffer a loss. In particular, if individuals survive longer than expected, the tontine provider needs to fulfill the promise of providing $d(t)$ to the pool as long as at least one policyholder is alive. In this sense, the tontine provider carries the mortality risk of the last survivor of the pool.

Proposition 2.3. *Let $\pi_i > 0$ and $w_i > 0$ for all $i = 1, \ldots, n$. Further, let the withdrawal rate $d(t)$ be chosen in such a way that $P_0 = 1$. Then, the mixed-cohort payoff design in (1) is collectively fair.*

Proof. Assuming that $d(t)$ is chosen in such a way that $P_0 = 1$, it is a straightforward calculation (see the proof of Proposition 2.2) to show that

$$\sum_{j=1}^n P_0^j = w \cdot P_0.$$

Hence, choosing $d(t)$ such that $P_0 = 1$ directly leads to the collective fairness. □

Given that the collective fairness is fulfilled, it seems natural to examine whether there are some choices of $\pi_1, \ldots, \pi_n$ such that the individual fairness is additionally fulfilled. For the case with two cohorts $L = 2$, we can theoretically show that individual fairness for all individuals can be ensured under mild assumptions.

Proposition 2.4. *Let $L = 2$ with $n_1 > 0$, $n_2 > 0$. Furthermore, assume without loss of generality that the first member of the pool belongs to the first cohort, while the second member belongs to the second cohort. Additionally, assume that the collective fairness is fulfilled, i.e.*

$$P_0 = 1. \tag{4}$$

Assume further that

$$n_1 w_1 < n_2 w_2 \left(\frac{\int_0^\infty e^{-rt} d(t)\mathbb{E}\left[1 - \left(1 - S_{x_1}(t)\right)^{n_1}\right] \mathrm{d}t}{1 - \int_0^\infty e^{-rt} d(t)\mathbb{E}\left[1 - \left(1 - S_{x_i}(t)\right)^{n_1}\right] \mathrm{d}t} \right). \tag{5}$$

Assuming that $\pi_1 = 1$, there exists a unique value π_2 such that the individual fairness is satisfied for both cohorts.

Proof. We want to have

$$w_i = P_0^i, \;\; i = 1, 2. \tag{6}$$

We set $\pi_1 := 1$ without loss of generality. Under this specification, we get:

$$P_0^1 = \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{w_1}{w_1 N_1(t) + w_2\pi_2 N_2(t)}\mathbb{1}_{\{T_1 > t\}}\right] \mathrm{d}t,$$

$$P_0^2 = \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{\pi_2 w_2}{w_1 N_1(t) + w_2\pi_2 N_2(t)}\mathbb{1}_{\{T_2 > t\}}\right] \mathrm{d}t.$$

Note that for (6) to be fulfilled, it suffices to find π_2 such that $w_1 = P_0^1$. Once this is fulfilled, the second condition $w_2 = P_0^2$ follows directly from the collective fairness condition (4), because it follows directly that

$$w = n_1 w_1 + n_2 P_0^2$$

which is equivalent to $w_2 = P_0^2$, because $w - n_1 w_1 = n_2 w_2$. It is clear that $P_0^1 : [0, \infty) \to \left(0, \int_0^\infty e^{-rt}\mathbb{E}\left[\frac{wd(t)}{N_1(t)}\mathbb{1}_{\{T_1 > t\}}\right] \mathrm{d}t\right]$, $\pi_2 \mapsto P_0^1(\pi_2)$ is a strictly decreasing and thus bijective function in π_2. Therefore, if

$$w_1 < \int_0^\infty e^{-rt}\mathbb{E}\left[\frac{wd(t)}{N_1(t)}\mathbb{1}_{\{T_1 > t\}}\right] \mathrm{d}t, \tag{7}$$

it follows that there exists a unique positive number π_2^* such that $P_0^1(\pi_2^*) = w_1$. □

Inequality (5) (or, equivalently, (7)) means that the initial wealth invested by a member of the first cohort shall be smaller than the initial value of the tontine benefits if they were paid only to a member of the first cohort. Note that for larger values of n_1, the denominator in (5) will be close to zero, because $\int_0^\infty e^{-rt} d(t)\mathbb{E}\left[1-\left(1-S_{x_1}(t)\right)^{n_1}\right]dt$ is close to 1. Thus, in realistic situations with appropriate pool sizes and two cohorts, individual fairness can be achieved. Inequality (5) can therefore be seen as a technical requirement rather than an economic restriction on the setup of the tontine.

Next, we prove that for any cohort size, given the existence of a set of fair participation rates, this set is unique up to multiplicative constant. For this, we directly follow a proof provided in Milevsky and Salisbury (2016).

Proposition 2.5. *Assume without loss of generality that the i-th individual belongs to the i-th cohort. Let $L \geq 2$ and $d(t)$ be given with $n_i > 0$ and $w_i > 0$ and assume that a set of fair participation rates $\pi = (\pi_1, \dots, \pi_L)$ with $\pi_j \in (0, \infty)$ exists. Then, this set is unique up to a multiplicative constant.*

Proof. The proof follows similar steps as the proof of Theorem 4(a) in Milevsky and Salisbury (2016) and is provided in Appendix A.1 for the sake of completeness. □

Proposition 2.4 shows the existence of fair participation rates for two cohorts. Proposition 2.5 shows the uniqueness of the fair participation rates (given existence) for any cohort size. Consequently, the question is now whether a set of fair participation rates exists for larger numbers of cohorts $L \geq 3$. In the following, assume without loss of generality that the i-th individual belongs to the i-th cohort. For such a number of cohorts, we get:

$$w_1 = \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{w_1}{w_1N_1(t)+\sum_{j=2}^L w_j\pi_jN_j(t)}\mathbb{1}_{\{T_1>t\}}\right]dt,$$

$$w_i = \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{\pi_i w_i}{w_1N_1(t)+\sum_{j=2}^L w_j\pi_jN_j(t)}\mathbb{1}_{\{T_i>t\}}\right]dt,\quad i=2,\dots,L-1.$$

Note that we can again omit the last equation ($i = L$) due to the collective fairness. Therefore, we have $L-1$ nonlinear equations and $L-1$ unknowns $(\pi_2, \dots, \pi_L)$. In the following, we consider the expected present values of future benefits P_0^i as functions of the participation rates. Note that these functions

$$\begin{aligned}
P_0^1 &: (0,\infty) \to \left(0, \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{w_1}{w_1N_1(t)+\sum_{j=3}^L w_j\pi_jN_j(t)}\mathbb{1}_{\{T_i>t\}}\right]dt\right),\\
&\quad \pi_2 \mapsto P_0^1(\pi_2),\\
P_0^i &: (0,\infty) \to \left(0, \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{\pi_i w_i}{w_1N_1(t)+\sum_{j=2,j\neq i+1}^L w_j\pi_jN_j(t)}\mathbb{1}_{\{T_i>t\}}\right]dt\right),\\
&\quad \pi_{i+1} \mapsto P_0^i(\pi_{i+1}),\quad i=2,\dots,L-1
\end{aligned}$$

are strictly decreasing in π_2 and π_{i+1}, treating $\pi_j,\ j = 3, \dots, L$ and $\pi_j,\ j \neq i+1$ as given constants, respectively. In other words, for any values of $\pi_j,\ j = 3, \dots, L$, fulfilling

$$w_1 < \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{w_1}{w_1N_1(t)+\sum_{j=3}^L w_j\pi_jN_j(t)}\mathbb{1}_{\{T_i>t\}}\right]dt, \tag{8}$$

we can find a unique value π_2^* such that $P_0^1(\pi_2^*) = w_1$. Similarly, for any values $\pi_j,\ j \neq i+1$ satisfying

$$w_i < \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{\pi_i w_i}{w_1N_1(t)+\sum_{j=2,j\neq i+1}^L w_j\pi_jN_j(t)}\mathbb{1}_{\{T_i>t\}}\right]dt, \tag{9}$$

we can find a unique value π_{i+1}^* such that $P_0^i(\pi_{i+1}^*) = w_i$. Thus, we can successively determine all the values $\pi_2^*, \dots, \pi_L^*$ until all the fairness conditions are met. This can, for instance, be done using the numerical procedure provided in Appendix B. This numerical method is one possible approach to finding fair participation rates and has some weaknesses, depending on how precisely the parameters are defined. Most importantly, in some situations the search may amount to trial and error and may require several attempts to find a solution. For a more detailed discussion along with the technical details, we refer interested readers to Appendix B.

2.3. Numerical example

Throughout the numerical analyses, we rely on a shocked Gompertz law. The deterministic Gompertz law (see Gompertz (1825)) is frequently used in actuarial science, particularly for retirement planning (see e.g. Milevsky (2020)). In this paper, we follow e.g. Lin and

Table 1
Base case parameter setup. We assume without loss of generality that individuals $i = 1, 2, 3$ belong to cohorts $i = 1, 2, 3$, respectively.

Cohorts $L = 3$	Cohort sizes $n_i = 500$	Initial wealth levels $w_1 = 100,\ w_2 = 200,\ w_3 = 300$
Risk-free rate $r = 0.01$	Participation rate $\pi_1 = 1$	Initial ages $x_1 = 65,\ x_2 = 70,\ x_3 = 75$
Modal age $m = 88.721$	Dispersion $b = 10$	Longevity shock $\epsilon \sim \mathcal{N}_{(-\infty,1)}\left(-0.0035, 0.0814^2\right)$

Table 2
Fair participation rates π_2 and π_3 for the flat tontine. We rely on the parameters introduced in Table 1.

$w_i = 100\,i$		
	$n_i = 100$	$n_i = 500$
$x = 70$	$\pi_2 = 5.49$	$\pi_2 = 6.57$
$x = 75$	$\pi_3 = 11.76$	$\pi_3 = 15.16$
$w_i = 100$		
	$n_i = 100$	$n_i = 500$
$x = 70$	$\pi_2 = 2.52$	$\pi_2 = 2.93$
$x = 75$	$\pi_3 = 4.42$	$\pi_3 = 5.36$

Cox (2005) and apply a stochastic shock to this deterministic mortality law to take account of the systematic mortality risk. Such a model can be motivated by the fact that regulators in many countries require insurers to test their balance sheet against various stress scenarios. Dealing with retirement plans, we are particularly interested in longevity shocks. In total, the force of mortality is, for any x and $t \geq 0$, given by

$$\mu_{x+t} = (1-\epsilon)\frac{1}{b}e^{\frac{x+t-m}{b}},$$

where $m > 0$ denotes the modal age at death, $b > 0$ is the dispersion coefficient and ϵ is a random shock taking values in $(-\infty, 1)$. Table 1 provides the base case parameters used in the subsequent numerical analyses.

The parameters are chosen due to the following reasons:

- For the overall pool size, we follow Qiao and Sherris (2013) who recommend a pool size of at least 1000 for modern tontines.
- Based on the ongoing low interest rate environment across most countries, we set the constant risk-free interest rate close to zero. For example, the German average risk-free rate of investment in 2019 was 1.1% (see Statista (2019)).
- The ages of the retirees 65, 70 and 75 are typical retirement ages. The initial wealth levels of the retirees increase in the age, since individuals who decide to postpone their retirement have more years to earn income than individuals who retire at an earlier age.
- For the values of m and b, we follow Milevsky and Salisbury (2015).
- Concerning the longevity shock ϵ, we follow Chen et al. (2019) and assume that it follows a truncated normal distribution on the interval $(-\infty, 1)$. The parameters used for this distribution are also taken from Chen et al. (2019), in which the parameters are calibrated in accordance with the longevity shock scenario of the Solvency II standard formula. This amounts to a sudden and permanent decrease of annual death probabilities by 20%.

As a simple example, we consider a flat tontine as defined in Milevsky and Salisbury (2015), i.e. we consider a constant withdrawal rate

$$d(t) = d = \frac{1}{\int_0^\infty e^{-rt}\mathbb{E}\left[1 - \prod_{j=1}^n (1 - S_{x_j}(t))\right]\mathrm{d}t}.$$

In Table 2, we present the resulting fair participation rates, which ensure both collective and individual fairness.

Naturally, the participation rates increase with age, as older individuals are entitled to higher retirement benefits. Furthermore, we find that the participation rates rise more sharply with age if initial wealth levels also increase with age. This is another natural outcome, as individuals who contribute at higher levels are entitled to even higher benefits due to individual fairness. Finally, we find that an increase in pool size leads to an increase in participation rates, but has no effect on the relationship between participation rates.

3. Optimal tontine under actuarial fairness

For a given tontine, or more specifically, a withdrawal rate, we have shown both theoretically and numerically how collective and individual fairness can be achieved among various cohorts. If fairness is the utmost key component that matters, our results above show that, at least for two cohorts, it is possible to ensure individual fairness for all. However, if other optimality criteria such as maximizing expected utility are also relevant for decision makers, the considered tontines in the above sections may not be optimal.

3.1. Social planner's problem

It is clear that the withdrawal rate $d(t)$ of the mixed-cohort tontine considered in this article cannot be determined in such a way that the expected discounted lifetime utility level of each cohort is maximized, as it is frequently done for homogeneous-cohort tontines (cf. Milevsky and Salisbury (2015) and Chen et al. (2019, 2020)). In the following, we therefore choose a withdrawal rate $d(t)$ which maximizes the weighted sum of the individual utility functions. In this sense, the resulting withdrawal rate $d(t)$ is optimal for the collective, but not optimal for a single cohort. In particular, this approach is particularly different from the approach taken in Milevsky and Salisbury (2016), where the natural tontine design introduced in Milevsky and Salisbury (2015) is generalized to heterogeneous cohorts.

Let U_i be the utility function and ρ_i be the subjective discount factor of individual i. Furthermore, let $\beta_1, \ldots, \beta_n$ be nonnegative numbers adding up to 1. Then, the optimization problem of the social planner is given by

$$\begin{aligned}
&\max_{d(t)} \mathbb{E}\left[\int_0^\infty \sum_{i=1}^n e^{-\rho_i t}\beta_i U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right] \\
&\text{subject to } \sum_{i=1}^n P_0^i = \int_0^\infty e^{-rt} wd(t)\mathbb{E}\left[1-\prod_{i=1}^n (1-S_{x_i}(t))\right]\mathrm{d}t = w,
\end{aligned} \tag{10}$$

where w is the total initial wealth. In the construction of this optimization problem, we follow the literature on collective decisions under uncertainty, see e.g. Wilson (1968), Amershi and Stoeckenius (1983), Huang and Litzenberger (1985), Dumas (1989), Weinbaum (2009), Pazdera et al. (2016), Jensen and Nielsen (2016), Schumacher (2018) and Chen et al. (2021b,c). The constraint of the optimization problem is the collective fairness condition, i.e. the initial value to the tontine scheme coincides with the total initial wealth.

If all the individual utility preferences are identical and given by a log-utility function, we can find an explicit solution to this optimization problem.

Theorem 3.1. *Assume that $U_i(\cdot) = \ln(\cdot)$ for all $i = 1, \ldots, n$. Then, the optimization problem (10) has the following explicit solution:*

$$d^*(t) = \frac{\sum_{i=1}^n e^{-\rho_i t}\beta_i \mathbb{E}\left[\mathbb{1}_{\{T_i>t\}}\right]}{\lambda e^{-rt}\mathbb{E}\left[1-\prod_{j=1}^n (1-S_{x_j}(t))\right]}, \tag{11}$$

where the Lagrangian multiplier is given by

$$\lambda = \sum_{i=1}^n \beta_i \int_0^\infty e^{-\rho_i t}\mathbb{E}\left[\mathbb{1}_{\{T_i>t\}}\right]\mathrm{d}t.$$

Proof. See Appendix A.2. □

Note that, with logarithmic utility, the optimal withdrawal rate (11) depends on individual discount factors, utility weights and future lifetimes, but not on the individual participation rates. This implies that in the numerical analysis, we can therefore treat this withdrawal rate as given and proceed as described in Section 2 to determine the individually fair participation rates.

If the utility functions are not identical, the expected discounted lifetime utility can be rewritten in the following way:

$$\begin{aligned}
&\mathbb{E}\left[\int_0^\infty \sum_{i=1}^n e^{-\rho_i t}\beta_i U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right] \\
&= \int_0^\infty \sum_{i=1}^n \mathbb{E}\left[e^{-\rho_i t}\beta_i U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\right]\mathrm{d}t \\
&= \int_0^\infty \sum_{i=1}^n \mathbb{E}\left[S_{x_i}(t)\mathbb{E}\left[e^{-\rho_i t}\beta_i U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\,\middle|\,\mathcal{F}_t, T_i>t\right]\right]\mathrm{d}t.
\end{aligned}$$

The Lagrangian is given by

$$\begin{aligned}
\mathcal{L} = &\int_0^\infty \sum_{i=1}^n \mathbb{E}\left[S_{x_i}(t)\mathbb{E}\left[e^{-\rho_i t}\beta_i U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\,\middle|\,\mathcal{F}_t, T_i>t\right]\right]\mathrm{d}t \\
&+\lambda\left(w-\int_0^\infty e^{-rt} wd(t)\mathbb{E}\left[1-\prod_{j=1}^n (1-S_{x_j}(t))\right]\mathrm{d}t\right).
\end{aligned}$$

The first-order condition is then given by

$$\sum_{i=1}^{n}\mathbb{E}\left[S_{x_i}(t)\mathbb{E}\left[e^{-\rho_i t}\beta_i U_i'\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^{n}\pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)w\cdot\frac{\pi_i w_i}{\sum_{j=1}^{n}\pi_j w_j \mathbb{1}_{\{T_j>t\}}}\,\middle|\,\mathcal{F}_t, T_i>t\right]\right]$$
$$=\lambda e^{-rt}w\mathbb{E}\left[1-\prod_{j=1}^{n}(1-S_{x_j}(t))\right].$$

In this more general case with different risk aversion levels for various cohorts, the withdrawal rate $d(t)$ and the participation rates π_i are interconnected. To solve this equation, we need to rely on numerical procedures. In particular, we need to simultaneously determine the optimal withdrawal rate $d(t)$ and the individually fair participation rates π_i. The interconnection between these parameters can make the numerical computation of these parameters complex and time-consuming.

3.2. Certainty equivalents and additional designs

We compare the social planner's tontine to tontines with homogeneous cohorts set up for each of the individual cohorts. Note that for such homogeneous-cohort-tontines, an explicit solution to the optimal withdrawal rate exists under CRRA utility preferences (see e.g. Chen et al. (2019)). In addition, we modify two tontine designs from Milevsky and Salisbury (2016), the proportional and natural mixed-cohort tontine. The modification is done to ensure the collective fairness condition for these tontines, such that reasonable comparisons between these designs and the social planner's optimal tontine can be conducted. Both the proportional and natural mixed-cohort tontine are extensions of the single-cohort natural tontine introduced in Milevsky and Salisbury (2015), which is optimal for log-utility-maximizers.

First, the *proportional* tontine is specified by

$$d(t)=\sum_{j=1}^{n}\frac{w_j}{w}\cdot\frac{s_{x_j}(t)}{\bar{a}_{x_j}},$$

where $\bar{a}_{x_j}=\int_0^\infty e^{-rt}s_{x_j}(t)\mathrm{d}t$ is the money's worth of an annuity paying out 1 continuously until death. Note that this design is not collectively fair, since it provides payments up to an infinite time horizon, even with no living policyholders left:

$$\int_0^\infty e^{-rt}\mathbb{E}\left[1-\prod_{j=1}^{n}(1-S_{x_j}(t))\right]d(t)\mathrm{d}t<\int_0^\infty e^{-rt}d(t)\mathrm{d}t=1.$$

If we want to use this tontine structure in our setting, we can easily modify it to the following collectively fair scheme:

$$d(t)=\frac{1}{\mathbb{E}\left[1-\prod_{j=1}^{n}(1-S_{x_j}(t))\right]}\sum_{j=1}^{n}\frac{w_j}{w}\cdot\frac{s_{x_j}(t)}{\bar{a}_{x_j}}.$$

In a final step, we can then choose the participation rates π_j in such a way that individual fairness is achieved.

Furthermore, Milevsky and Salisbury (2016) consider the *natural* mixed-cohort tontine, specified by

$$d(t)=\sum_{j=1}^{n}\frac{\pi_j w_j}{\sum_{i=1}^{n}\bar{a}_{x_i}\pi_i w_i}\cdot s_{x_j}(t),$$

where the weights π_j can be chosen arbitrarily. This approach is also not collectively fair, as it provides payments up to an infinite time horizon, even if there are no living policyholders left:

$$\int_0^\infty e^{-rt}\mathbb{E}\left[1-\prod_{j=1}^{n}(1-S_{x_j}(t))\right]d(t)\mathrm{d}t<\int_0^\infty e^{-rt}d(t)\mathrm{d}t=1.$$

Again, we can easily modify this withdrawal rate into the following collectively fair scheme:

$$d(t)=\frac{1}{\mathbb{E}\left[1-\prod_{j=1}^{n}(1-S_{x_j}(t))\right]}\sum_{j=1}^{n}\frac{\pi_j w_j}{\sum_{i=1}^{n}\bar{a}_{x_i}\pi_i w_i}\cdot s_{x_j}(t),$$

where we can again choose the participation rates to achieve individual fairness.

For our numerical analyses, we assume that the utility functions are of the constant relative risk aversion (CRRA) type, i.e., for $\gamma_i>0$, the utility function U_i is defined by

$$U_i(y)=\begin{cases}\frac{y^{1-\gamma_i}}{1-\gamma_i}, & \gamma_i\neq 1\\ \ln(y), & \gamma_i=1.\end{cases}$$

In the utility function $\sum_{i=1}^{n}\beta_i U_i(\cdot)$, we add different types of CRRA utility functions U_i. To make sure that the units in the collective utility function can be added, we need to ensure that the weights β_i are chosen in such a way that $\beta_i U_i$ owns the same unit for all i. This is e.g. ensured by the choice of weights

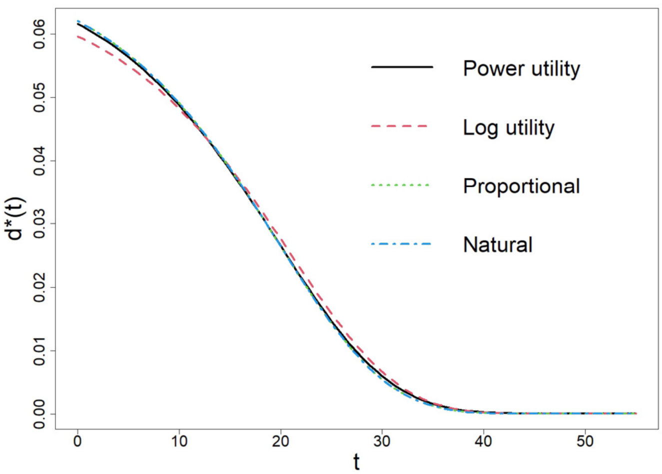

Fig. 1. Withdrawal rates depending on time t for different tontine designs, where we use $\gamma_1 = 6$ and $\gamma_2 = 8$ for the case with power utility. The parameters are chosen as in Table 1. Furthermore, we assume $\rho_i = r$.

$$\beta_i = \frac{\bar{w}^{\gamma_i}}{\sum_{j=1}^{n} \bar{w}^{\gamma_j}}, \quad \bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i, \tag{12}$$

which we have taken from Chen et al. (2021c).

In this section, we fix the first two cohorts introduced in Table 1 as base case parameters. Throughout the numerical analyses, we assume $\rho_i = r$ to ensure a fair comparison between the tontine of the social planner and Milevsky and Salisbury (2016)'s designs. For the CRRA utility, we assume that For the CRRA utility, we assume that $\gamma_1 = 6$ and $\gamma_2 = 8$, i.e. risk aversion increases with age, as found in the literature (at least from age 65 on, see e.g. Riley and Chow (1992) and DaSilva et al. (2019)).

In Fig. 1, we show the optimal withdrawal rate $d(t)$ obtained under log utility and power utility along with the withdrawal rates of the proportional and natural tontine. Note that the optimal withdrawal rate under log utility is independent of the participation rates π_i. For the power utility, we set $\pi_1 = 1$ and $\pi_2 = 1.222$ to achieve individual fairness. The withdrawal rate of the proportional tontine is again independent of the participation rates and for the natural tontine, we set $\pi_2 = 1.211$ to achieve individual fairness. We observe that the withdrawal rates are decreasing and therefore coincide roughly with optimal withdrawal rates derived in the literature (cf. Milevsky and Salisbury (2015)). In total, there are no major differences in the structure of the different withdrawal rates.

We analyze the utility benefit/loss generated by the collective problem by comparing it to optimal solutions when they are treated as separated cohorts. To compare the benefits resulting from the different tontines, we consider the certainty equivalent CE as the level of the deterministic annuity payoff that yields the same expected utility as a given mixed-cohort tontine. That is, the certainty equivalent CE_i of individual i is determined by

$$\int_0^\infty e^{-\rho_i t} s_{x_i}(t) U_i(\mathrm{CE}_i)\mathrm{d}t = \mathbb{E}\left[\int_0^\infty e^{-\rho_i t} U_i\left(wd(t)\cdot \frac{\pi_i w_i}{\sum_{j=1}^{n}\pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right].$$

Under CRRA utility functions, the certainty equivalent can be determined as

$$\mathrm{CE}_i = \begin{cases} \left(\frac{(1-\gamma_i)}{\int_0^\infty e^{-\rho_i t} s_{x_i}(t)\,\mathrm{d}t}\mathbb{E}\left[\int_0^\infty e^{-\rho_i t} U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^{n}\pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right]\right)^{\frac{1}{1-\gamma_i}}, & \gamma_i \neq 1 \\ e^{\mathbb{E}\left[\int_0^\infty e^{-\rho_i t} U_i\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^{n}\pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right]\left(\int_0^\infty e^{-\rho_i t} s_{x_i}(t)\,\mathrm{d}t\right)^{-1}}, & \gamma_i = 1, \end{cases}$$

depending on whether power or log utility is used.

3.3. Numerical results

3.3.1. Base case

In Table 3, we provide the certainty equivalents of the mixed-cohort and the single-cohort tontines for log-utility. We observe that the proportional and natural tontine slightly outperform the social planner's tontine for both pool sizes considered. However, they are unable to outperform the homogeneous tontines. In particular, we observe that the younger cohort prefers the homogeneous tontine to the mixed-cohort tontines. This result is reversed for the older cohort. All in all, the differences between the various tontine designs are rather small. Furthermore, we observe that an increase in the pool size increases all certainty equivalents, a natural result which is consistent with the literature.

In Table 4, we provide the certainty equivalents of the mixed-cohort and the single-cohort tontines for CRRA utility. We observe that the social planner's optimal tontine and the homogeneous tontines now outperform the natural and proportional tontine. The reason for this result is that both these tontine designs are generalizations of the natural tontine introduced in Milevsky and Salisbury (2015), which is optimal for a log-utility maximizer. Furthermore, we again find that the younger cohort prefers the homogeneous tontine to the mixed-cohort tontine set up by the social planner. This result reverses for the older cohort. Thus, it is not clear which of the two approaches of

Table 3
Certainty equivalents obtained under log-utility. We rely on the parameters introduced in Table 1 and the participation rates making the heterogeneous tontine individually fair. Furthermore, we assume $\rho_i = r$.

	$x = 65$	$x = 70$	π_2
$n_1 = n_2 = 100$			
Homogeneous tontine	5.423	13.09	-
Social planner	5.40	13.12	1.236
Proportional tontine	5.417	13.12	1.216
Natural tontine	5.416	13.13	1.216
$n_1 = n_2 = 500$			
Homogeneous tontine	5.44	13.127	-
Social planner	5.42	13.130	1.230
Proportional tontine	5.43	13.14	1.211
Natural tontine	5.43	13.14	1.211

Table 4
Certainty equivalents obtained for $\gamma_1 = 6$ and $\gamma_2 = 8$. We rely on the parameters introduced in Table 1 and the participation rates making the heterogeneous tontine individually fair. Furthermore, we assume $\rho_i = r$.

	$x = 65$	$x = 70$	π_2
$n_1 = n_2 = 100$			
Homogeneous tontine	5.31	12.64	-
Social planner	5.27	12.89	1.236
Proportional tontine	5.16	12.00	1.216
Natural tontine	5.18	12.15	1.216
$n_1 = n_2 = 500$			
Homogeneous tontine	5.38	12.91	-
Social planner	5.36	13.02	1.222
Proportional tontine	5.28	12.68	1.211
Natural tontine	5.29	12.75	1.211

Table 5
Certainty equivalents obtained under log-utility. We rely on the parameters introduced in Table 1 (except for the cohort sizes) and the participation rates making the heterogeneous tontine individually fair. Furthermore, we assume $\rho_i = r$.

	$x = 65$	$x = 70$	π_2
$n_1 = 100$, $n_2 = 25$			
Homogeneous tontine	5.423	12.97	-
Social planner	5.412	13.11	1.230
Proportional tontine	5.418	13.12	1.218
Natural tontine	5.417	13.11	1.218
$n_1 = 500$, $n_2 = 125$			
Homogeneous tontine	5.438	13.09	-
Social planner	5.430	13.14	1.222
Proportional tontine	5.436	13.15	1.211
Natural tontine	5.435	13.14	1.211

is preferable. One could just as easily argue that the decision to establish a single tontine with heterogeneous cohorts or separate tontines with homogeneous cohorts could be left with the plan provider.

3.3.2. Unequal pool sizes

Next, we want to study the impact of unequal pool sizes on the certainty equivalents and participation rates. We assume that the second cohort (retirement age 70) makes out only 25% of the first cohort (retirement age 65). In Table 5 we provide the certainty equivalents of the mixed-cohort and the single-cohort tontines for log-utility. Comparing the results of Table 5 to those in Table 3, we observe that the certainty equivalent of the second cohort decreases significantly under the homogeneous tontine, whereas this effect is reduced in the heterogeneous tontines.

In Table 6, we provide the corresponding certainty equivalents of the mixed-cohort and the single-cohort tontines for CRRA utility. Comparing the results in Table 6 to those in Table 4, we observe that members of the second cohort now suffer a larger loss in utility from choosing the single-cohort tontine. This is a natural result, given that their pool size now only makes out 25% of the younger cohort. In particular, any mixed-cohort tontine performs better for this cohort than the single-cohort tontine. Nevertheless, we still observe that the social planner's tontine is the most preferable mixed-cohort tontine for both cohorts. The reasons are, again, the consideration of the different degrees of risk aversion (which differ from log utility).

Table 6
Certainty equivalents obtained for $\gamma_1 = 6$ and $\gamma_2 = 8$. We rely on the parameters introduced in Table 1 and the participation rates making the heterogeneous tontine individually fair. Furthermore, we assume $\rho_i = r$.

	$x = 65$	$x = 70$	π_2
$n_1 = 100$, $n_2 = 25$			
Homogeneous tontine	5.308	11.98	-
Social planner	5.281	12.86	1.234
Proportional tontine	5.180	12.10	1.218
Natural tontine	5.180	12.11	1.218
$n_1 = 500$, $n_2 = 125$			
Homogeneous tontine	5.383	12.70	-
Social planner	5.366	13.05	1.222
Proportional tontine	5.325	12.87	1.211
Natural tontine	5.317	12.82	1.211

Table 7
Certainty equivalents obtained under log-utility. We rely on the parameters introduced in Table 1 and the participation rates making the heterogeneous tontine individually fair. Furthermore, we assume $\rho_i = r$.

	$x = 65$	$x = 80$	π_2
$n_1 = n_2 = 100$			
Homogeneous tontine	5.423	20.77	-
Social planner	5.063	20.57	2.495
Proportional tontine	5.405	20.90	1.945
Natural tontine	5.404	20.88	1.942
$n_1 = n_2 = 500$			
Homogeneous tontine	5.438	20.86	-
Social planner	5.103	20.62	2.462
Proportional tontine	5.428	20.92	1.931
Natural tontine	5.428	20.92	1.930

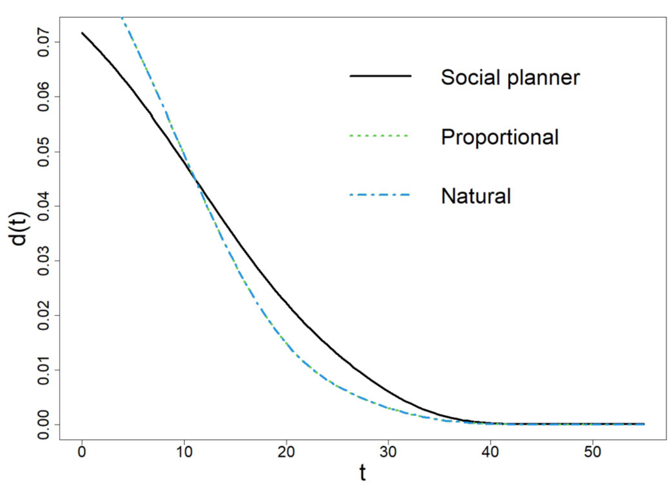

Fig. 2. Withdrawal rate $d(t)$ depending on time t for the proportional, natural and social planner's tontine (under log utility). The parameters are chosen as in Table 1 except for $x_2 = 80$. Furthermore, we assume $\rho_i = r$.

3.3.3. Increasing the age difference

As a final analysis, we want to increase the heterogeneity of the cohorts by increasing the age of members in the second cohort to 80. In Table 7, we provide the corresponding certainty equivalents of the mixed-cohort and the single-cohort tontines for log-utility. Now we observe that the social planner's tontine delivers the lowest certainty equivalents for both cohorts and has a significantly higher participation rate for the second cohort than the proportional and natural tontine. To analyze these findings in more detail, in Fig. 2, we show the corresponding withdrawal rates of the different tontine designs. We observe that the social planner's tontine delivers lower (higher) benefits at younger (older) retirement ages compared to the other two designs, which nearly coincide. This slower decline of the withdrawal rate results in a higher participation rate for the older cohort, because these have lower probabilities of receiving retirement benefits at older ages, i.e. they need higher participation rates for their remaining lifetime in order to fulfill the individual fairness constraint. The flatter withdrawal rate results because, in this example, apparently the social planner takes the younger cohort stronger into consideration. The payoff could be adjusted by modifying the utility weights β_i in (12) to account, for example, for the ages of the different cohorts as well.

Table 8
Certainty equivalents obtained for $\gamma_1 = 6$ and $\gamma_2 = 8$. We rely on the parameters introduced in Table 1 and the participation rates making the heterogeneous tontine individually fair. Furthermore, we assume $\rho_i = r$.

	$x = 65$	$x = 80$	π_2
$n_1 = n_2 = 100$			
Homogeneous tontine	5.308	19.78	-
Social planner	5.080	20.43	2.11
Proportional tontine	5.117	20.45	1.95
Natural tontine	5.111	20.43	1.94
$n_1 = n_2 = 500$			
Homogeneous tontine	5.383	20.35	-
Social planner	5.227	20.62	2.06
Proportional tontine	5.287	20.73	1.93
Natural tontine	5.281	20.72	1.93

The more steeply declining proportional and natural withdrawal rates, however, lead to lower participation rates for the second cohort as well as higher certainty equivalents for both cohorts. Hence, while the natural and proportional tontine turn out not to be optimal for higher degrees of risk aversion, in this example, they can better account for large differences in age groups than the tontine constructed by the social planner.

Next, in Table 8, we provide the certainty equivalents of the mixed-cohort and the single-cohort tontines for CRRA utility. Comparing the results in Table 8 to those in Table 7, we observe that the certainty equivalents of the collectively optimal and the natural and proportional tontine have become closer. The reason for this is that the more adequate reflection of risk aversions offsets the utility loss caused by the age difference observed in Table 7. Similar to Table 7, we observe that the participation rates of the social planner are higher than those of the natural and proportional tontine.

Thus, among the parameters considered, we find that the natural and proportional tontines are better suited to mixed cohort tontines with large age differences, while the social planner's tontine works better when ages are closer together and utility preferences differ strongly from log utility.

Throughout the respective numerical analyses, we additionally observe that the fair participation rates π_2 are roughly identical and approximately given by

$$\pi_2 \approx \frac{\bar{a}_{x_1}}{\bar{a}_{x_2}} = \begin{cases} 18.35290/15.19852 = 1.207545, & x_2 = 70, \\ 18.35290/9.552432 = 1.921280, & x_2 = 80. \end{cases} \tag{13}$$

The reason for this is that the participation rates measure the differences between age groups, for which annuity prices are at least a good approximation (see also Milevsky and Salisbury (2016)). However, for the collectively optimal tontine, the approximation is less accurate when the certainty equivalents are low (see in particular the cases $x_1 = 65$ and $x_2 = 80$). In particular, our numerical analyses suggest that for any well-functioning tontine design (yielding acceptable certainty equivalents), the above approximation (13) can be used for the participation rates. Particularly, this approximation of the fair participation rates as the ratio of the present values of the annuities can be used to find more accurate estimates of the fair participation rates. For the details, we refer the interested reader to Appendix B.

Finally, we observe that the numerically found participation rates coincide with sensible participation rates from the literature. In particular, we confirm the findings by Donnelly (2015) in the context of equitable retirement products. Moreover, our numerical results suggest that risk aversion heterogeneity might be a second-degree effect in mortality risk sharing dominated by the first-degree effect of age heterogeneity.

4. Conclusion

This paper studies actuarial fairness and social welfare in tontines with heterogeneous cohorts. To study actuarial fairness, the tontine design we consider is a given one and the one introduced in Milevsky and Salisbury (2016). We distinguish between collective and individual fairness, show how collective fairness can be achieved and demonstrate that, under certain conditions, even individual fairness can be additionally ensured. In the second part of this paper, we consider a utilitarian social planner setting up a tontine on behalf of the collective of heterogeneous policyholders. Comparing the collectively optimal tontine to separate, cohort-optimized tontines as well as the proportional and natural tontine presented in Milevsky and Salisbury (2016), we find that approximately identical certainty equivalents are achieved under homogeneous log-utility preferences. However, in our numerical analysis, under higher and heterogeneous degrees of relative risk aversion, the social planner and the homogeneous-cohort tontines outperform the proportional and natural tontine, whereas the natural and proportional tontines perform better for mixed cohort tontines with large age differences.

As a possible extension to this article, we might assume that policyholders are not only heterogeneous in their ages and wealth levels, but, for example, also in their health status (cf. care-dependent tontines as in Chen et al. (2022) and Hieber and Lucas (2022)). We might for example assume that each cohort can be split into two groups, where one group has higher mortality rates (i.e. a lower health status) than the other group. We might then also assume that the health status is correlated with the wealth, i.e. less healthy individuals are also less wealthy. Such an additional degree of heterogeneity would increase the diversity in the pool and might change the relation between the social planner and the single-cohort tontines.

Declaration of competing interest

There are no competing interests.

Data availability

No data was used for the research described in the article.

Appendix A. Proofs

A.1. Proof of Proposition 2.5

Note that individual fairness is equivalent to

$$\frac{P_0^j}{w_j} = 1 \text{ for all } j = 1\dots, L.$$

The proof now follows similar steps as the proof of Theorem 4(a) in Milevsky and Salisbury (2016). Assume that π and $\tilde{\pi}$ both fulfill individual fairness, are unequal and are not multiples of each other. We use the notation $P_0^j(\pi)$ throughout this proof to emphasize the dependence of P_0^j on the different sets of participation rates. Define $\pi(s) := s\pi + (1-s)\tilde{\pi}$. Then, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}P_0^j(\pi(s)) &= \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{\mathrm{d}}{\mathrm{d}s}\frac{\pi_j(s)}{\sum_{i=1}^L \pi_i(s)w_iN_i(t)}\mathbb{1}_{\{T_j>t\}}\right]\mathrm{d}t \\ &= \int_0^\infty e^{-rt}\mathbb{E}\left[wd(t)\frac{\sum_{i=1}^L w_iN_i(t)\left(\pi_i(s)\pi_j'(s) - \pi_j(s)\pi_i'(s)\right)}{\left(\sum_{i=1}^L \pi_i(s)w_iN_i(t)\right)^2}\mathbb{1}_{\{T_j>t\}}\right]\mathrm{d}t.\end{aligned}$$

Here, it holds

$$\begin{aligned}\pi_i(s)\pi_j'(s) - \pi_j(s)\pi_i'(s) &= \left(\tilde{\pi}_i + s\left(\pi_i - \tilde{\pi}_i\right)\right)(\pi_j - \tilde{\pi}_j) - (\pi_i - \tilde{\pi}_i)\left(\tilde{\pi}_j + s\left(\pi_j - \tilde{\pi}_j\right)\right) \\ &= \tilde{\pi}_i(\pi_j - \tilde{\pi}_j) - (\pi_i - \tilde{\pi}_i)\tilde{\pi}_j = \tilde{\pi}_i\pi_j - \pi_i\tilde{\pi}_j = \pi_j\pi_i\left(\frac{\tilde{\pi}_i}{\pi_i} - \frac{\tilde{\pi}_j}{\pi_j}\right).\end{aligned} \tag{14}$$

Now fix j such that $\tilde{\pi}_j/\pi_j$ is minimal. Then, it follows that (14) is ≥ 0 for all i and > 0 for at least one i (because $\tilde{\pi}$ is not a multiple of π). Hence, $\frac{\mathrm{d}}{\mathrm{d}s}P_0^j$ is positive for this j which implies $1 = P_0^j(\pi) < P_0^j(\tilde{\pi}) = 1$, clearly a contradiction. □

A.2. Proof of Theorem 3.1

Note that we can rewrite the expected lifetime utility of the collective as

$$\begin{aligned}&\mathbb{E}\left[\int_0^\infty \sum_{i=1}^n e^{-\rho_i t}\beta_i \ln\left(wd(t)\cdot\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right] \\ &= \mathbb{E}\left[\int_0^\infty \sum_{i=1}^n e^{-\rho_i t}\beta_i\left(\ln(w) + \ln(d(t)) + \ln\left(\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\right)\mathbb{1}_{\{T_i>t\}}\mathrm{d}t\right] \\ &= \int_0^\infty \sum_{i=1}^n e^{-\rho_i t}\beta_i\mathbb{E}\left[\left(\ln(w) + \ln(d(t)) + \ln\left(\frac{\pi_i w_i}{\sum_{j=1}^n \pi_j w_j \mathbb{1}_{\{T_j>t\}}}\right)\right)\mathbb{1}_{\{T_i>t\}}\right]\mathrm{d}t.\end{aligned}$$

The first-order condition is therefore:

$$\sum_{i=1}^n e^{-\rho_i t}\beta_i\frac{1}{d(t)}\mathbb{E}\left[\mathbb{1}_{\{T_i>t\}}\right] = \lambda e^{-rt}\mathbb{E}\left[1 - \prod_{j=1}^n(1 - S_{x_j}(t))\right],$$

which delivers

$$d^*(t) = \frac{\sum_{i=1}^n e^{-\rho_i t}\beta_i\mathbb{E}\left[\mathbb{1}_{\{T_i>t\}}\right]}{\lambda e^{-rt}\mathbb{E}\left[1 - \prod_{j=1}^n(1 - S_{x_j}(t))\right]}.$$

The Lagrangian multiplier λ is obtained from the budget constraint as follows:

$$1 = \int_0^\infty e^{-rt}\frac{\sum_{i=1}^n e^{-\rho_i t}\beta_i\mathbb{E}\left[\mathbb{1}_{\{T_i>t\}}\right]}{\lambda e^{-rt}\mathbb{E}\left[1 - \prod_{j=1}^n(1 - S_{x_j}(t))\right]}\mathbb{E}\left[1 - \prod_{j=1}^n(1 - S_{x_j}(t))\right]\mathrm{d}t$$

$$= \int_0^\infty \frac{\sum_{i=1}^n e^{-\rho_i t} \beta_i \mathbb{E}\left[\mathbb{1}_{\{T_i > t\}}\right]}{\lambda} \mathrm{d}t$$

$$\Leftrightarrow \quad \lambda = \int_0^\infty \sum_{i=1}^n e^{-\rho_i t} \beta_i \mathbb{E}\left[\mathbb{1}_{\{T_i > t\}}\right] \mathrm{d}t. \quad \square$$

Appendix B. Pseudo code for determining the fair participation rates

Let us again assume that individual i is a member of cohort i ($i = 1, \dots, L$). For a given withdrawal rate, the following pseudo code finds the fair participation rates $\pi_1, \dots, \pi_L$.

1. Fix a tolerance level *tol* and initialize all the necessary parameters.
2. Specify upper and lower bounds $\pi_2^u, \dots, \pi_L^u$ and $\pi_2^l, \dots, \pi_L^l$ for $\pi_2, \dots, \pi_L$ and set $\pi_i = \frac{1}{2}(\pi_i^u + \pi_i^l)$ for $i = 2, \dots, L$ such that (8) and (9) are fulfilled.
3. Compute P_0^1.
4. While $|P_0^1 - w_1| > tol$
   (a) Specify upper and lower bounds $\pi_3^u, \dots, \pi_L^u$ and $\pi_3^l, \dots, \pi_L^l$ for $\pi_3, \dots, \pi_L$ and set $\pi_i = \frac{1}{2}(\pi_i^u + \pi_i^l)$ for $i = 3, \dots, L$ such that (8) and (9) are fulfilled.
   (b) While $|P_0^2 - w_2| > tol$
      i. ... (Continue as in (a) and (b) for all $i = 3, \dots, L-2$.)
      ii. Specify upper and lower bounds π_L^u and π_L^l for π_L and set $\pi_L = \frac{1}{2}(\pi_L^u + \pi_L^l)$ such that (8) and (9) are fulfilled.
      iii. While $|P_0^{L-1} - w_{L-1}| > tol$
         A. If $P_0^{L-1} - w_{L-1} > 0$, set $\pi_L = \pi_L^l$.
         B. If $P_0^{L-1} - w_{L-1} < 0$, set $\pi_L = \pi_L^u$.
         C. Compute P_0^{L-1}.
      iv. ... (Set all the π_i values as in B and C for the corresponding $i = L-2, \dots, 3$.)
      v. If $P_0^2 - w_2 > 0$, set $\pi_3 = \pi_3^l$.
      vi. If $P_0^2 - w_2 < 0$, set $\pi_3 = \pi_3^u$.
      vii. Compute P_0^2.
   (c) If $P_0^1 - w_1 > 0$, set $\pi_2 = \pi_2^l$.
   (d) If $P_0^1 - w_1 < 0$, set $\pi_2 = \pi_2^u$.
   (e) Compute P_0^1.

We want to point out that choosing the upper and lower bounds for the participation rates may, in fact, not be straightforward. For the lower bound, there is some information we can use:

- For identical wealth levels and heterogeneous ages, we can assume that the first cohort is the youngest cohort and fix $\pi_1 = 1$. It is then clear that all the other cohorts will have $\pi_j > 1$ for $j = 2, \dots, L$.
- Assuming identical ages and heterogeneous wealth levels, we can do the same thing by fixing $\pi_1 = 1$, where this time the first cohort ($j = 1$) is the least wealthy. It is then again clear that all the other cohorts will have $\pi_j > 1$ for $j = 2, \dots, L$.
- In the numerical analyses in this paper, we assume that older cohorts are also the more wealthy ones, allowing us to rely on the previous two observations. From our point of view, this assumption is not unrealistic because people who retire at older ages have more time to accumulate savings for their retirement.
- The last case would then be decreasing wealth levels combined with increasing ages. In this case, choosing a lower bound would not be so clear and we would have to work out the more dominant factor (age vs. wealth) first.

Choosing an appropriate upper bound, however, is not straightforward if we want to allow for general withdrawal rates $d(t)$.

- For example, for the withdrawal rates in Section 3, we could use the approximation $\pi_j \approx \bar{a}_{x_1}/\bar{a}_{x_j}$. For example, we can fix the upper bound as $c \cdot \bar{a}_{x_1}/\bar{a}_{x_j}$, where c is some constant greater than 1 (suggested by our numerical findings, as $\pi_j > \bar{a}_{x_1}/\bar{a}_{x_j}$), for example, $c = 2$.
- However, for general withdrawal rates (e.g. the flat tontine), we indeed have to first take a guess and see whether the algorithm can find a solution within the defined interval. If it cannot do that (it will then simply deliver the upper bound as a solution), we need to restart the algorithm with a larger upper bound. Finding appropriate upper bounds can therefore amount to trial and error in some cases.

References

Amershi, A.H., Stoeckenius, J.H., 1983. The theory of syndicates and linear sharing rules. Econometrica. Journal of the Econometric Society 51 (5), 1407–1416.

Chen, A., Chen, Y., Xu, X., 2022. Care-dependent tontines. Insurance. Mathematics & Economics 106, 69–89.

Chen, A., Guillen, M., Rach, M., 2021a. Fees in tontines. Insurance. Mathematics & Economics 100, 89–106.

Chen, A., Hieber, P., Klein, J.K., 2019. Tonuity: a novel individual-oriented retirement plan. ASTIN Bulletin. The Journal of the IAA 49 (1), 5–30.

Chen, A., Nguyen, T., Rach, M., 2021b. A collective investment problem in a stochastic volatility environment: the impact of sharing rules. Annals of Operations Research 302 (1), 85–109.

Chen, A., Nguyen, T., Rach, M., 2021c. Optimal collective investment: the impact of sharing rules, management fees and guarantees. Journal of Banking & Finance 123, 106012.
Chen, A., Qian, L., Yang, Z., 2021d. Tontines with mixed cohorts. Scandinavian Actuarial Journal 2021 (5), 437–455.
Chen, A., Rach, M., Sehner, T., 2020. On the optimal combination of annuities and tontines. ASTIN Bulletin. The Journal of the IAA 50 (1), 95–129.
DaSilva, A., Farka, M., Giannikos, C., 2019. Age-dependent increasing risk aversion and the equity premium puzzle. Financial Review 54 (2), 377–412.
Denuit, M., 2019. Size-biased transform and conditional mean risk sharing, with application to p2p insurance and tontines. ASTIN Bulletin. The Journal of the IAA 49 (3), 591–617.
Donnelly, C., 2015. Actuarial fairness and solidarity in pooled annuity funds. ASTIN Bulletin. The Journal of the IAA 45 (1), 49–74.
Donnelly, C., Guillén, M., Nielsen, J.P., 2014. Bringing cost transparency to the life annuity market. Insurance. Mathematics & Economics 56, 14–27.
Dumas, B., 1989. Two-person dynamic equilibrium in the capital market. The Review of Financial Studies 2 (2), 157–188.
Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London 115, 513–583.
Hanewald, K., Piggott, J., Sherris, M., 2013. Individual post-retirement longevity risk management under systematic mortality risk. Insurance. Mathematics & Economics 52 (1), 87–97.
Hieber, P., Lucas, N., 2022. Modern life-care tontines. ASTIN Bulletin. The Journal of the IAA 52 (2), 563–589.
Huang, C.-F., Litzenberger, R., 1985. On the necessary condition for linear sharing and separation: a note. Journal of Financial and Quantitative Analysis 20 (3), 381–384.
Jensen, B.A., Nielsen, J.A., 2016. How suboptimal are linear sharing rules? Annals of Finance 12 (2), 221–243.
Lin, Y., Cox, S.H., 2005. Securitization of mortality risks in life annuities. The Journal of Risk and Insurance 72 (2), 227–252.
Milevsky, M.A., 2020. Calibrating Gompertz in reverse: what is your longevity-risk-adjusted global age? Insurance. Mathematics & Economics 92, 147–161.
Milevsky, M.A., Salisbury, T.S., 2015. Optimal retirement income tontines. Insurance. Mathematics & Economics 64, 91–105.
Milevsky, M.A., Salisbury, T.S., 2016. Equitable retirement income tontines: mixing cohorts without discriminating. ASTIN Bulletin. The Journal of the IAA 46 (3), 571–604.
Pazdera, J., Schumacher, J.M., Werker, B.J., 2016. Cooperative investment in incomplete markets under financial fairness. Insurance. Mathematics & Economics 71, 394–406.
Piggott, J., Valdez, E.A., Detzel, B., 2005. The simple analytics of a pooled annuity fund. The Journal of Risk and Insurance 72 (3), 497–520.
Qiao, C., Sherris, M., 2013. Managing systematic mortality risk with group self-pooling and annuitization schemes. The Journal of Risk and Insurance 80 (4), 949–974.
Riley Jr, W.B., Chow, K.V., 1992. Asset allocation and individual risk aversion. Financial Analysts Journal 48 (6), 32–37.
Sabin, M.J., 2010. Fair tontine annuity. Available at SSRN https://ssrn.com/abstract=1579932.
Schumacher, J.M., 2018. Linear versus nonlinear allocation rules in risk sharing under financial fairness. ASTIN Bulletin. The Journal of the IAA 48 (3), 995–1024.
Stamos, M.Z., 2008. Optimal consumption and portfolio choice for pooled annuity funds. Insurance. Mathematics & Economics 43 (1), 56–68.
Statista, 2019. Average risk-free investment rate in Germany 2015-2019. Website. Available online on https://www.statista.com/statistics/885774/average-risk-free-rate-germany/. (Accessed 23 October 2019).
Weinbaum, D., 2009. Investor heterogeneity, asset pricing and volatility dynamics. Journal of Economic Dynamics and Control 33 (7), 1379–1397.
Wilson, R., 1968. The theory of syndicates. Econometrica. Journal of the Econometric Society 36 (1), 119–132.
Yaari, M.E., 1965. Uncertain lifetime, life insurance, and the theory of the consumer. The Review of Economic Studies 32 (2), 137–150.