ON THE OPTIMAL COMBINATION OF ANNUITIES AND TONTINES

BY

AN CHEN, MANUEL RACH AND
THORSTEN SEHNER

ABSTRACT

Tontines, retirement products constructed in such a way that the longevity risk is shared in a pool of policyholders, have recently gained vast attention from researchers and practitioners. Typically, these products are cheaper than annuities, but do not provide stable payments to policyholders. This raises the question whether, from the policyholders' viewpoint, the advantages of annuities and tontines can be combined to form a retirement plan which is cheaper than an annuity, but provides a less volatile retirement income than a tontine. In this article, we analyze and compare three approaches of combining annuities and tontines in an expected utility framework: the previously introduced "tonuity", a product very similar to the tonuity which we call "antine" and a portfolio consisting of an annuity and a tontine. We show that the payoffs of a tonuity and an antine can be replicated by a portfolio consisting of an annuity and a tontine. Consequently, policyholders achieve higher expected utility levels when choosing the portfolio over the novel retirement products tonuity and antine. Further, we derive conditions on the premium loadings of annuities and tontines indicating when the optimal portfolio is investing a positive amount in both annuity and tontine, and when the optimal portfolio turns out to be a pure annuity or a pure tontine.

KEYWORDS

Optimal retirement products, annuity, tontine, tonuity, antine.

JEL codes: G22, J32.

1. INTRODUCTION

Annuities provide (life-)long payment streams to the policyholder and thus, constitute a possible way to build protection against the increasing threat of the individual's incapability to keep her living standards at older ages. They have been considered as very desirable retirement products from a policyholder's perspective (see, e.g., Yaari, 1965; Mitchell, 2002; Davidoff *et al.*, 2005; or Peijnenburg *et al.*, 2016). However, in practice, annuitization rates have remained rather low (see, for instance, Hu and Scott 2007 and Inkmann *et al.* 2010). This phenomenon is already well known as the "annuity puzzle" in the academic world, and there exists a variety of literature exploring the main drivers responsible for this puzzle. Literature reviews can be found, for example, in Brown (2007) or Milevsky (2013). An overview of existing puzzles in life insurance can be found in Gottlieb (2012). Recent attempts to tackle the annuity puzzle include but are not limited to Poppe-Yanez (2017), Caliendo *et al.* (2017), Chen *et al.* (2018) and O'Dea and Sturrock (2019). Due to the tightening solvency regulation and the low interest rate environment, it yet seems unlikely that retirees are going to annuitize more of their wealth in the near future. Consequently, insurers and customers are searching for new, more attractive retirement products. In this context, tontine products, which were a popular source of retirement income back in the 17th, 18th and 19th centuries (see Milevsky and Salisbury, 2015), have attracted vast attention from academics and practitioners. For details about tontines, we refer the interested reader, for instance, to Sabin (2010), Milevsky (2015), Milevsky and Salisbury (2015, 2016), or Li and Rothschild (2019).[1] One of the main properties of tontines, in contrast to annuities, is that a pool of policyholders shares the longevity risk. In this sense, tontines and annuities are two extreme types of retirement products constructed in such a way that the longevity risk is, in the case of tontines, (almost) fully borne by the policyholders or, in the case of annuities, fully by the insurer.

Naturally, the question arises whether the advantages of annuities and tontines can be combined to form a product which is cheaper than an annuity and shifts the longevity risk not completely, but only partially toward the policyholder. Possible ways of combining annuities and tontines are already examined in Weinert and Gründl (2017) and Chen *et al.* (2019). Chen *et al.* (2019) present a new retirement product called "tonuity" which is a tontine at early retirement ages, but switches to an annuity at a predetermined switching time. Weinert and Gründl (2017) focus on how the policyholder can optimally invest fractions of her wealth in tontines and annuities in a cumulative prospect theory framework, where the tontine design is taken from Sabin (2010). In this article, we compare various combinations of annuities and tontines in a classical expected utility framework to find the "best" product from the policyholder's viewpoint, where we focus on the tontine design from Milevsky and Salisbury (2015). For this, we include not only the tonuity and a portfolio

consisting of an annuity and a tontine, but also a new product which we call "antine". The antine works similarly as the tonuity: it provides annuity-like payments at early retirement ages and, after a prespecified switching time, switches to tontine-like payments at older ages. All these three combinations of tontines and annuities contain the original products (annuity and tontine) as special cases.

In contrast to Chen *et al.* (2019), we extend their study by additionally considering and analyzing the portfolio and the novel concept of the antine, and comparing them with the tonuity. Further, we analytically investigate the impact of premium loadings on the decision about the optimal retirement product. Our resulting findings are hence all-new and importantly contribute to the discussion on optimal retirement products. Our article can also be considered as a straightforward extension to Milevsky and Salisbury (2015), where we take more retirement products into consideration. Compared to Sabin (2010) who deals with different ages, genders and initial contributions, we consider a simplified case with homogeneous policyholders. However, while Sabin (2010) focuses on how a fair tontine between members of different groups can be designed, we go beyond this and study utility-maximizing payoffs of various products.

According to Milevsky and Salisbury (2015), in an actuarially fair pricing framework, annuities yield a higher level of expected utility than tontines. However, more realistically, by adding appropriate safety loadings to the prices of these products, it is possible that tontines outperform annuities (see Milevsky and Salisbury 2015 or Chen *et al.* 2019). In the present article, we set ourselves in this more realistic setting and determine the utility-maximizing payoffs of the tonuity, the antine and the portfolio of an annuity and a tontine for a risk-averse policyholder with no bequest motive. While for the tonuity and the antine an explicit solution is available, the case with the portfolio requires us to rely on numerical procedures to determine the optimal annuity and tontine payoffs. The optimal payoff of the tonuity can be considered as a direct generalization of the optimal tonuity payoff in Chen *et al.* (2019) who derive the utility-maximizing payoff without incorporating safety loadings. The antine payoff can be determined analogously to the payoff of the tonuity. While, in the *portfolio*, the optimal payoff of the tontine coincides roughly with optimal tontine designs discussed in the literature (cf. Milevsky and Salisbury, 2015; Chen *et al.*, 2019), the corresponding annuity payoff structure deviates substantially from this literature as it first increases and then decreases rather strongly, leading to a bell-shaped curve. The reason for this structure is that the annuity provides secure payments at times when the tontine provides the most volatile payments. At rather advanced retirement ages, the tontine payments are relatively high due to the few surviving policyholders, which leads to a decrease in the annuity payoff. Based on these optimal income streams, we can implicitly determine the fractions of wealth initially invested in the annuity and the tontine, respectively.

Our main theoretical result shows that, from the policyholder's point of view, a portfolio consisting of an annuity and a tontine can outperform any tonuity and antine. The reason for this is that, under the considered structure of the (loaded) premiums, any tonuity and antine payoff can be replicated by such a portfolio, given the initial premiums of the three retirement plans are identical. Throughout this article, we incorporate safety loadings in the premiums using the expected value principle. Nevertheless, this main theoretical result remains valid when other premium calculation principles, like the variance or the standard deviation principle, are applied. Moreover, we derive conditions for the loadings of the tontine and the annuity, under which a pure annuity, a pure tontine, or an investment in both of them is utility-maximizing: if the annuity loading is smaller than or equal to the tontine loading, it is optimal to invest all initial wealth in the annuity. If the annuity loading drastically exceeds the tontine loading, a pure investment in the tontine is optimal. Under realistic loadings, that is, the annuity loading is reasonably larger than the tontine loading, an investment in both annuity and tontine yields the maximal utility.

In our numerical analysis, the expected lifetime utility of the optimal tonuity does get very close to that of the optimal portfolio. Given that the optimal payoffs of the tontine and the annuity in the portfolio are rather complex, this finding indicates that a single switch from tontine to annuity might be a more useful and simpler way for practice, although the optimal combination of annuities and tontines is, in fact, not the tonuity. Further, the newly proposed antine seems not to be a desirable product from the policyholder's perspective and is frequently outperformed by the tonuity. This is probably due to the design of the antine which leaves policyholders with volatile payments in the advanced retirement ages and is still rather expensive compared to tontines.

The remainder of the article is structured as follows: Section 2 describes the basic model setup, where, in particular, the assumptions regarding the mortality model and the design of the considered retirement products are discussed. In Section 3, we derive the optimal payoffs of the different retirement products and the optimal level of expected utility of each retirement plan. In Section 4, we theoretically and numerically compare the attractiveness of the different combinations of annuities and tontines from a policyholder's perspective. Section 5 concludes the article and is followed by appendices, where supplementary proofs and a pseudocode for the numerical determination of the optimal annuity and tontine payoffs in the portfolio are provided.

2. MODEL SETUP

In this section, we describe the basic model setup used throughout the remainder of our article. We start by describing our mortality model and continue by introducing the designs of the retirement plans under consideration.

2.1. Mortality model

We consider a simple mortality model which allows us to distinguish between two types of mortality risk: unsystematic or idiosyncratic mortality risk stems from the fact that the lifetime of a person is unknown, but still follows a certain mortality law, and can thus be diversified. Systematic or aggregate mortality risk stems from the fact that the true underlying mortality law cannot be determined with certainty. This type of mortality risk cannot be diversified and affects all the policies of an insurer in the same direction. Further explanations on these two different aspects of mortality risk are given, for example, in Piggott *et al.* (2005). We use the usual actuarial notation $_tp_x$ for the best-estimate survival curve of an x-year-old policyholder over time $t \geq 0$. These best-estimates can be computed from continuous-time mortality laws which are usually obtained from publicly available life tables. We follow Lin and Cox (2005) to incorporate the systematic mortality risk in the mortality law by applying a random shock ϵ to the best-estimates. The shocked survival curve is then given by $_tp_x^{1-\epsilon}$. The shock ϵ is a continuous random variable, whose density is denoted by $f_\epsilon(\cdot)$ and which takes values in $(-\infty, 1)$. Note that by restricting the shock ϵ to the interval $(-\infty, 1)$, the shocked survival probabilities $_tp_x^{1-\epsilon}$ still possess all the important properties we require from survival probabilities: first of all, they are still probabilities as they lie between zero and one. Furthermore, they fulfill the property $_tp_x^{1-\epsilon} = {}_sp_x^{1-\epsilon} \cdot {}_{t-s}p_{x+s}^{1-\epsilon}$ for all $0 \leq s \leq t$. As the shock affects all the policyholders in the same direction, it cannot be diversified by choosing the initial pool size large enough and is thus an important component in our model to capture the systematic mortality risk. The special case with no longevity shock is obtained by setting $\epsilon = 0$.

2.2. Retirement products

We consider an individual endowed with an initial wealth amounting to $v > 0$ who can buy one of the following five retirement plans. The first two are the annuity and the tontine. The remaining three are then combinations of the annuity and the tontine and contain the annuity and the tontine as special cases. In this section, we introduce the payoffs of the retirement products and determine their gross premiums obtained using the expected value principle.

2.2.1. *Annuity and tontine*

Let us first consider an *annuity* contract. Following Yaari (1965), we assume that by buying such a contract, the policyholder continuously receives the deterministic payment $c(t)$ which starts immediately and continues until her death. To denote the random remaining future lifetime of the policyholder, we use T_ϵ that takes account of the random longevity shock ϵ introduced above. Then, the payoff of the annuity to the policyholder at any time t can

be expressed as

$$b_A(t) = \mathbb{1}_{\{T_\epsilon > t\}} c(t). \tag{2.1}$$

Here, $\mathbb{1}_B$ is the indicator function that is equal to one if event B occurs and zero otherwise. By deploying the expected value principle, we can write the initially charged gross premium for the annuity as follows:

$$\widetilde{P}_0^A = (1 + C_A) P_0^A, \tag{2.2}$$

where $C_A \geq 0$ describes the proportional risk loading applied in the context of the annuity. The corresponding net premium P_0^A can be obtained, by noting that $\left(\mathbb{1}_{\{T_\epsilon > t\}} \big| \epsilon \right) \sim \text{Bernoulli} \left({}_t p_x^{1-\epsilon} \right)$, as (see also Equation (2.4) in Chen *et al.*, 2019 for a detailed derivation)

$$P_0^A = \mathbb{E}\left[\int_0^\infty e^{-rt} b_A(t) \, dt \right] = \int_0^\infty e^{-rt} \, {}_t p_x \, m_\epsilon(-\ln \, {}_t p_x) \, c(t) \, dt, \tag{2.3}$$

where r is the risk-free interest rate and $m_\epsilon(s) = \mathbb{E}\left[e^{s\epsilon} \right]$ for $s \in \mathbb{R}$ is the moment-generating function of ϵ.

Next, let us consider a *tontine* contract. We use $n \in \mathbb{N}$ to denote the initial number of homogeneous policyholders holding the same tontine contract. The policyholders can be considered as identical copies of each other. Note that, as we focus on the comparison between the different combinations of annuities and tontines, we keep the tontine modeling rather simple. Nevertheless, dealing with heterogeneity between the individuals in our context surely opens up an interesting perspective for future research (cf. Milevsky and Salisbury, 2016). By choosing the pool size n large enough, it is possible for the insurer to diversify the unsystematic mortality risk. However, it is not possible for the insurer to diversify the systematic mortality risk for this risk influences all the members in the pool in the same way. At older ages, when the pool size has decreased, the remaining policyholders are left with both systematic and unsystematic risk. We use $N_\epsilon(t)$ to denote the random number of policyholders in the pool who are still alive at time t. Following Milevsky and Salisbury (2015), the continuous payoff at any time t to a single policyholder in the pool, who holds a tontine contract, is then given by

$$b_{OT}(t) = \mathbb{1}_{\{T_\epsilon > t\}} \frac{n}{N_\epsilon(t)} d(t), \tag{2.4}$$

where $d(t)$ is a deterministic payoff function specified at contract initiation. When $N_\epsilon(t)$ equals zero, the tontine payoff is set equal to zero. While an annuity provides a deterministic payoff to a living policyholder, the future tontine payment to a living policyholder is a random variable depending on the number of pool members alive. Note that the payoff in (2.4) is paid out to a *single* policyholder. From the insurer's perspective, the payment $nd(t)$ is made at each time t. This payment is made until the last policyholder has died. The insurer carries the longevity risk of the last living policyholder in the pool. That is,

while in an annuity, the insurer promises a guaranteed payment to each single policyholder, in the tontine it is only promised to the pool, which leaves the policyholders with most of the mortality risk.

Conditional on the considered policyholder still being alive, given the longevity shock ϵ, and assuming the lifetimes of the policyholders to be independent, the number of surviving individuals follows a binomial distribution, that is,

$$(N_\epsilon(t) - 1 | T_\epsilon > t, \epsilon) \sim \text{Bin}\left(n - 1, \, {}_t p_x^{1-\epsilon}\right). \tag{2.5}$$

Similar to the case of the annuity described above, the gross premium for the tontine initially charged is specified through

$$\widetilde{P}_0^{OT} = (1 + C_{OT}) P_0^{OT}, \tag{2.6}$$

where $C_{OT} \geq 0$ describes the proportional risk loading applied in the context of the tontine. Using the property in (2.5) and the binomial theorem, the net premium P_0^{OT} of the tontine is computed as (see also Equation (2.5) in Chen *et al.*, 2019 for a detailed derivation)

$$P_0^{OT} = \mathbb{E}\left[\int_0^\infty e^{-rt} b_{OT}(t)\, dt\right] = \int_0^\infty e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\, d(t)\, dt. \tag{2.7}$$

In general, note that, for the magnitudes of the proportional risk loadings, it makes sense to assume $C_A > C_{OT}$. This is due to the fact that within an annuity product, the insurer carries the entire longevity risk, while the policyholders carry most of the longevity risk within a tontine product. Additionally, it is reasonable to assume that the loading for the tontine decreases in the pool size as less unsystematic risk is then involved in the tontine product. Note that, typically, the risk loadings should carefully reflect the risks associated with a retirement product. The setting we consider allows for very general risk loadings and still allows us to determine explicit solutions for the optimal payoffs of at least some of the retirement products, but is, due to its simplicity, not as accurate as the one in Chen *et al.* (2019). We solve this issue by carefully setting the loadings according to the values provided in Chen *et al.* (2019) in the numerical section.

2.2.2. *Portfolio*

Assume now that the policyholder can combine an annuity and a tontine by initially investing in both products. The resulting payoff of this portfolio at any time t is given by

$$b_{AT}(t) = \mathbb{1}_{\{T_\epsilon > t\}} \left(c_{AT}(t) + \frac{n}{N_\epsilon(t)} d_{AT}(t)\right), \tag{2.8}$$

where $c_{AT}(t)$ and $d_{AT}(t)$ are the payoff functions of the annuity and the tontine constituting the portfolio, respectively. In case $N_\epsilon(t)$ is equal to zero, the payoff in (2.8) is defined to be zero, similarly as the tontine payoff. The more wealth is invested in the annuity, the closer the payoff of the portfolio gets to an annuity-like payoff, and analogously for the tontine. The initial single gross premium for the portfolio is plainly given by

$$\widetilde{P}_0^{AT} = \widetilde{P}_0^{A,AT} + \widetilde{P}_0^{OT,AT} = (1 + C_A)P_0^{A,AT} + (1 + C_{OT})P_0^{OT,AT}, \qquad (2.9)$$

where $P_0^{A,AT}$ and $P_0^{OT,AT}$ are defined similarly as in Equations (2.3) and (2.7) with the payoffs $c(t)$ and $d(t)$ replaced by $c_{AT}(t)$ and $d_{AT}(t)$, respectively. That is, the gross premium of the portfolio corresponds to the gross premium of the contained annuity plus the gross premium of the contained tontine. Hence, both proportional loading factors C_A and C_{OT} appear in the above formula.

2.2.3. Tonuity

As the second way of combining the tontine and the annuity, we consider the *tonuity* with a prespecified switching time $\tau \geq 0$, originally introduced in Chen *et al.* (2019). Until time τ, the payoff to the policyholder coincides with that of the tontine. From time τ on, the payoff switches to the payoff of the annuity. Note that τ is not a random variable, but a constant fixed at contract initiation. This product provides the policyholder a secure payoff at more advanced retirement ages. At any time t, the payoff of a tonuity to a policyholder having a residual lifetime T_ϵ is given by

$$b_{[\tau]}(t) = \mathbb{1}_{\{0 \leq t < \min\{\tau, T_\epsilon\}\}} \frac{n}{N_\epsilon(t)} d_{[\tau]}(t) + \mathbb{1}_{\{\tau \leq t < T_\epsilon\}} c_{[\tau]}(t), \qquad (2.10)$$

where $d_{[\tau]}(t)$ and $c_{[\tau]}(t)$ are the payoff functions of the tontine and the annuity constituting the tonuity, respectively. Recall that $N_\epsilon(t)$ is the number of participants still alive at t and n is the initial number of participants. When choosing $\tau = \infty$, we obtain the payoff of a tontine and when choosing $\tau = 0$, we deal with an annuity. To determine the gross premium of a tonuity, we assume that the insurer again applies the expected value principle with proportional loading C_{OT} to the part of the payoff which corresponds to the tontine and the expected value principle with loading C_A to the part of the payoff which corresponds to the annuity. Consequently, the total premium for a tonuity is given by

$$\widetilde{P}_0^{[\tau]} = (1 + C_{OT}) P_0^{OT,\tau} + (1 + C_A) P_0^{A,\tau}, \qquad (2.11)$$

where the tonuity-specific premium parts $P_0^{OT,\tau}$ and $P_0^{A,\tau}$ are implicitly defined via the corresponding overall net premium which can be taken from Equation (4.2) in Chen *et al.* (2019):

$$P_0^{[\tau]} = \mathbb{E}\left[\int_0^\infty e^{-rt} b_{[\tau]}(t)\, dt\right]$$

$$= \int_0^\tau e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\, d_{[\tau]}(t)\, dt$$

$$+ \int_\tau^\infty e^{-rt} {}_t p_x \, m_\epsilon(- \ln {}_t p_x) \, c_{[\tau]}(t) \, dt \tag{2.12}$$
$$= P_0^{OT,\tau} + P_0^{A,\tau}.$$

Note that using the two safety loading factors C_{OT} and C_A also for the tonuity allows us to make reasonable comparisons between the different combinations of annuities and tontines, as is done in Section 4.

2.2.4. *Antine*

Inspired by tonuities, the third way of combining the tontine and the annuity is to start with annuity-like payments until a prespecified switching time $\sigma \geq 0$, after which tontine-like payments are made. Due to its structure, we name this new contract *antine*. At any time t, the payoff of an antine to a policyholder having a residual lifetime T_ϵ is given by

$$b_{[\sigma]}(t) = \mathbb{1}_{\{0 \leq t < \min\{\sigma, T_\epsilon\}\}} c_{[\sigma]}(t) + \mathbb{1}_{\{\sigma \leq t < T_\epsilon\}} \frac{n}{N_\epsilon(t)} d_{[\sigma]}(t), \tag{2.13}$$

where $c_{[\sigma]}(t)$ and $d_{[\sigma]}(t)$ are the payoff functions of the annuity and the tontine constituting the antine, respectively. When choosing $\sigma = \infty$, we obtain the payoff of an annuity and when choosing $\sigma = 0$, we deal with a tontine. The gross premium of an antine is determined similarly as for the tonuity. We assume that the insurer again applies the expected value principle with proportional loading C_A to the part of the payoff which corresponds to the annuity and the expected value principle with loading C_{OT} to the part of the payoff which corresponds to the tontine. Consequently, the total premium for the antine is given by

$$\widetilde{P}_0^{[\sigma]} = (1 + C_A) \, P_0^{A,\sigma} + (1 + C_{OT}) \, P_0^{OT,\sigma}, \tag{2.14}$$

where the antine-specific premium parts $P_0^{A,\sigma}$ and $P_0^{OT,\sigma}$ are implicitly defined via the corresponding overall net premium which can be computed analogously as for the tonuity:

$$
\begin{aligned}
P_0^{[\sigma]} &= \mathbb{E}\left[\int_0^\infty e^{-rt} b_{[\sigma]}(t) \, dt \right] \\
&= \int_0^\sigma e^{-rt} {}_t p_x \, m_\epsilon(- \ln {}_t p_x) \, c_{[\sigma]}(t) \, dt \\
&\quad + \int_\sigma^\infty e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi} \right)^n \right) f_\epsilon(\varphi) \, d\varphi \, d_{[\sigma]}(t) \, dt \\
&= P_0^{A,\sigma} + P_0^{OT,\sigma}.
\end{aligned}
\tag{2.15}
$$

Similarly as for the tonuity, we, for reasons of comparison, use the two distinct loading factors C_A and C_{OT} also for the antine.

3. OPTIMAL PAYOFF AND EXPECTED UTILITY

In this section, we derive the optimal payoff and the corresponding optimal level of expected utility for each of the retirement plans introduced in the previous section. To avoid redundancy, we only focus on the tonuity, the antine and the portfolio consisting of an annuity and a tontine. Note that the pure annuity and the pure tontine are contained in each of these three combined products, which is why there is no need to study them separately.

Before we start with the detailed consideration, let us first, in the style, for example, of Yaari (1965), generally introduce the policyholder's expected discounted lifetime utility as

$$U\left(\{\chi(t)\}_{t\geq 0}\right) = \mathbb{E}\left[\int_0^\infty e^{-\rho t} u\left(\chi(t)\right) \mathbb{1}_{\{T_\epsilon > t\}} \, dt\right], \tag{3.16}$$

where

$$\chi(t) = \begin{cases} c_{AT}(t) + \frac{n}{N_\epsilon(t)} d_{AT}(t), & \text{for portfolio,} \\ \mathbb{1}_{\{0 \leq t < \tau\}} \frac{n}{N_\epsilon(t)} d_{[\tau]}(t) + \mathbb{1}_{\{\tau \leq t\}} c_{[\tau]}(t), & \text{for tonuity,} \\ \mathbb{1}_{\{0 \leq t < \sigma\}} c_{[\sigma]}(t) + \mathbb{1}_{\{\sigma \leq t\}} \frac{n}{N_\epsilon(t)} d_{[\sigma]}(t), & \text{for antine,} \end{cases} \tag{3.17}$$

denotes the payoff of each retirement plan to a living policyholder. Note that we do not include a bequest motive in the utility of the policyholder. Instead, we assume that the policyholder has already set aside money to take care of the bequest motive beforehand, for example, by buying an insurance contract. The initial wealth v in our setting is therefore not the entire wealth she holds. This stipulation also allows us to compare our results with, for instance, Milevsky and Salisbury (2015), where the bequest motive is neglected as well. Further, we assume that $u(z) = \frac{z^{1-\gamma}}{1-\gamma}$ for $z > 0$ is a constant relative risk aversion (CRRA) utility function with a risk aversion parameter $\gamma > 0$ adhering to $\gamma \neq 1$ and ρ is the subjective discount rate of the policyholder. The policyholder chooses the deterministic payoff functions $c.(t)$ and $d.(t)$ in $\chi(t)$, so that (3.16) is maximized under the following budget constraint: her initial wealth v is fully used to purchase the corresponding retirement product. The purchase prices of the three products coincide with the different gross premiums and are thus given in (2.9), (2.11) and (2.14). Consequently, for $j = AT, [\tau], [\sigma]$, the budget constraint is generally given by

$$v = \widetilde{P}_0^j. \tag{3.18}$$

As the optimal payoffs of the tonuity and the antine can be determined explicitly, in contrast to the optimal payoff of the portfolio, we first discuss the optimization problems of the tonuity and antine before we deal with the portfolio. Let us start with the tonuity.

3.1. Tonuity

Translating the outlined policyholder's goal into the framework of the tonuity leads to the following optimization problem:

$$\max_{\left(c_{[\tau]}(t),\, d_{[\tau]}(t)\right)_{t\in[0,\infty)}} \mathbb{E}\left[\int_0^\infty e^{-\rho t} \left(\mathbb{1}_{\{0\leq t<\min\{\tau, T_\epsilon\}\}} u\left(\frac{n}{N_\epsilon(t)} d_{[\tau]}(t)\right) \right. \right.$$

$$\left. \left. + \mathbb{1}_{\{\tau\leq t<T_\epsilon\}} u\left(c_{[\tau]}(t)\right) \right) dt \right] \tag{3.19}$$

$$\text{subject to} \quad v = \widetilde{P}_0^{[\tau]} = (1+C_{OT})\, P_0^{OT,\tau} + (1+C_A)\, P_0^{A,\tau}.$$

Theorem 3.1 provides the solution to optimization problem (3.19).

Theorem 3.1. *For a tonuity with a switching time τ, the optimal payoff functions are given by*

$$d_{[\tau]}^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}} \left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma}}{\lambda_{[\tau]}^{1/\gamma}\, (1+C_{OT})^{1/\gamma} \left(\int_{-\infty}^1 \left(1-\left(1-{}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\right)^{1/\gamma}} \quad \text{for all } t \in [0,\tau) \tag{3.20}$$

and

$$c_{[\tau]}^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}}{\lambda_{[\tau]}^{1/\gamma}\, (1+C_A)^{1/\gamma}} \quad \text{for all } t \in [\tau, \infty), \tag{3.21}$$

where the optimal Lagrangian multiplier $\lambda_{[\tau]}$ is given by

$$\lambda_{[\tau]} = \left(\frac{1}{v} \left(\int_0^\tau (1+C_{OT})^{1-\frac{1}{\gamma}}\, e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^1 \left(1-\left(1-{}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\right)^{1/\gamma-1}}\, dt \right. \right.$$

$$\left. \left. + \int_\tau^\infty (1+C_A)^{1-\frac{1}{\gamma}}\, e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}\, {}_t p_x\, m_\epsilon(-\ln {}_t p_x)\, dt \right) \right)^\gamma \tag{3.22}$$

and $\kappa_{n,\gamma,\epsilon}({}_t p_x)$ by

$$\kappa_{n,\gamma,\epsilon}({}_t p_x) = \sum_{k=1}^n \binom{n}{k} \left(\frac{k}{n}\right)^\gamma \int_{-\infty}^1 \left({}_t p_x^{1-\varphi}\right)^k \left(1-{}_t p_x^{1-\varphi}\right)^{n-k} f_\epsilon(\varphi)\, d\varphi. \tag{3.23}$$

TABLE 1

BASE CASE PARAMETER SETUP.

Initial wealth $v = 300$ thousand euros	Pool size $n = 1000$	Risk aversion $\gamma = 6$
Risk-free rate $r = 0.01$	Subjective discount rate $\rho = r$	Risk loadings $C_A = 4\%$, $C_{OT} = 0.01\%$
Initial age $x = 65$	Gompertz law $m = 88.721$, $\beta = 10$	Longevity shock $\epsilon \sim \mathcal{N}_{(-\infty,1)}\left(-0.0035, 0.0814^2\right)$

The expected discounted lifetime utility is then given by

$$
\begin{aligned}
U_{[\tau]} = \mathbb{E}\Bigg[\int_0^\infty e^{-\rho t} \Bigg(&\mathbb{1}_{\{0 \leq t < \min\{\tau, T_\epsilon\}\}} u\left(\frac{n}{N_\epsilon(t)} d_{[\tau]}^*(t)\right) \\
&+ \mathbb{1}_{\{\tau \leq t < T_\epsilon\}} u\left(c_{[\tau]}^*(t)\right) \Bigg) dt \Bigg] = \frac{\lambda_{[\tau]}}{1 - \gamma} v.
\end{aligned}
\tag{3.24}
$$

Proof. See Appendix A.1. □

Note that the special cases $\tau = \infty$ and $\tau = 0$ lead to the tontine and the annuity, respectively. A further important observation resulting from Theorem 3.1 is that the optimal annuity payoff (3.21) is constant for all switching times if $\rho = r$, which is in line with Yaari (1965). It is also shown, for example, in Yagi and Nishigaki (1993) that constant annuities are suboptimal for individuals, whose subjective discount rate differs from the risk-free interest rate. If the subjective discount rate exceeds (falls below) the risk-free interest rate, that is, $\rho > r$ ($\rho < r$), the annuity payoff (3.21) is decreasing (increasing) over time.

As Theorem 3.1 holds for any τ, it is also possible for a specific policyholder to numerically find the optimal switching time τ^* for the tonuity such that the highest lifetime utility is achieved for this policyholder. More detailed explanations on the optimal switching time of a tonuity can also be found in Chen *et al.* (2019). We denote the optimal payoff functions resulting from τ^* by $d_{[\tau^*]}^*(t)$ and $c_{[\tau^*]}^*(t)$. In order to obtain τ^*, we can compute the expected utility levels for sufficiently many values of τ increasing from 0 to, for example, 55, and then choose the switching time τ^* which yields the highest expected lifetime utility. More details on the behavior of the optimal switching time τ^* can be found in Section 4.2.

In order to illustratively show how the optimal payoff of the tonuity can look like, we fix the parameter values summarized in Table 1 as our base case parameter setup.

Note the following remarks about our choice of parameters:

- To determine the value of the initial wealth v, we follow the estimation of Royal London (2018). They state that an average (British) employee needs

to invest around 260 thousand pounds sterling, which approximately corresponds to 300 thousand euros, in the private pension provision to keep her standard of living in the retirement phase beginning at the age of 65 years.

- In their simulation study of group self-annuitization schemes, Qiao and Sherris (2013) frequently apply a pool size of 1000 which we adopt for our analyses.
- For the risk-free rate, we choose a fairly low value to conform with the current situation in many European countries. As an example, consider Germany, where the average risk-free rate of investment in 2019 equals only 1.1% (see Statista, 2019).
- The values of the risk loadings are guided by the results for the risk capital charge in Chen *et al.* (2019). In this way, the reasonable assumption that $C_A > C_{OT}$ discussed in Section 2.2.1 remains in force.
- The best-estimates $_t p_x$ of the survival probability are assumed to follow the well-known Gompertz law (see Gompertz, 1825) as used, for example, in Gumbel (1958) or Milevsky and Salisbury (2015). In other words, we assume that

$$_t p_x = e^{e^{\frac{x-m}{\beta}}\left(1 - e^{\frac{t}{\beta}}\right)}, \tag{3.25}$$

with $\beta > 0$ being the dispersion coefficient and $m > 0$ being the modal age at death. The chosen values for β and m stem from Milevsky and Salisbury (2015).

- Regarding the chosen probability distribution for the shock ϵ, we comply with Chen *et al.* (2019) and assume that it follows a truncated normal distribution on $(-\infty, 1)$, that is, $\mathcal{N}_{(-\infty, 1)}\left(\mu, \nu^2\right)$. In accordance with the European Solvency II Directive, the parameters μ and ν are determined in such a way that the expected survival probabilities $\mathbb{E}\left[_t p_x^{1-\epsilon}\right]$ from our simple internal model are close to the best-estimate survival probabilities $_t p_x$.

For the base case, the optimal switching time of the tonuity is given by $\tau^* = 27$ as the maximal utility is attained at this time when considering $\{0, 1, \ldots, 54, 55\}$ as the possible choices for τ. Figure 1 shows the mean and the range bordered by the 0.01- and the 0.99-quantiles of the appropriate optimal tonuity payoff to the policyholder with respect to her age. The determination of all depicted quantities is done numerically and is based on the assumption that the individual is always alive, so that, at any time t, the applied optimal tonuity payoff is here given by

$$\mathbb{1}_{\{0 \leq t < \tau^*\}} \frac{n}{N_\epsilon(t)} d^*_{[\tau^*]}(t) + \mathbb{1}_{\{\tau^* \leq t\}} c^*_{[\tau^*]}(t). \tag{3.26}$$

As the only randomness in the optimal payoff stems from the uncertain future number of living policyholders in the pool, it is clear that, after the switch to the annuity at time $\tau^* = 27$, that is, when the policyholder turns 92 years, the two examined quantiles coincide and equal the constant annuity payment. As long as the tontine defines the tonuity, that is, while the individual
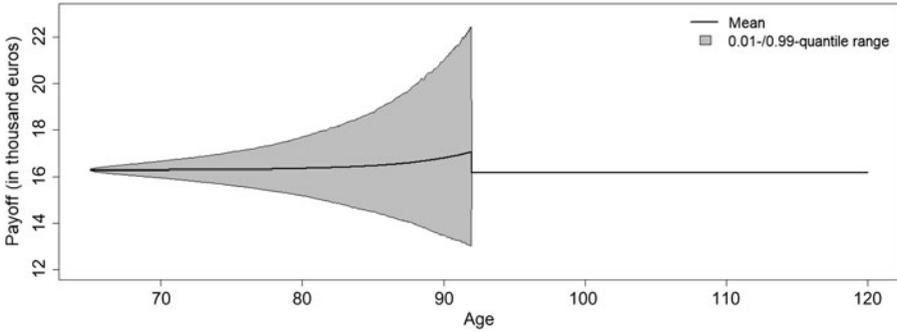
FIGURE 1: Mean and 0.01-/0.99-quantile range of the optimal payoff of the tonuity with $\tau^* = 27$ depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05. The plot is based on the assumption that the considered policyholder is always alive.

is between 65 and 92 years old, the payoff uncertainty overall increases as time goes by, where the possible upward movement, especially if the age gets closer to 92 years, intensifies considerably faster. This growing volatility trend is accompanied by a slight increase in the average payoff to the policyholder which however drops weakly afterward to also match the constant annuity payment for the remaining time.

3.2. Antine

The optimization problem for antines can be presented very similarly as for tonuities. Here, $c_{[\sigma]}(t)$ and $d_{[\sigma]}(t)$ are chosen in such a way that the corresponding expected discounted lifetime utility is maximized and that the appropriate budget constraint is met:

$$
\max_{\left(c_{[\sigma]}(t),\, d_{[\sigma]}(t)\right)_{t\in[0,\infty)}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\left(\mathbb{1}_{\{0\le t<\min\{\sigma,\, T_\epsilon\}\}} u\left(c_{[\sigma]}(t)\right)\right.\right.
$$

$$
\left.\left.+\mathbb{1}_{\{\sigma\le t<T_\epsilon\}} u\left(\frac{n}{N_\epsilon(t)} d_{[\sigma]}(t)\right)\right) dt\right] \tag{3.27}
$$

$$
\text{subject to} \quad v = \widetilde{P}_0^{[\sigma]} = (1+C_A)\, P_0^{A,\sigma} + (1+C_{OT})\, P_0^{OT,\sigma}.
$$

Theorem 3.2 provides the optimal payoff and expected discounted lifetime utility for the antine by analogy with Theorem 3.1.

Theorem 3.2. *For an antine with a switching time σ, the optimal payoff functions are given by*

$$
c_{[\sigma]}^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}}{\lambda_{[\sigma]}^{1/\gamma}\, (1+C_A)^{1/\gamma}} \quad \text{for all } t \in [0,\sigma) \tag{3.28}
$$

and

$$d^*_{[\sigma]}(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}} \left(\kappa_{n,\gamma,\epsilon}(_t p_x)\right)^{1/\gamma}}{\lambda_{[\sigma]}^{1/\gamma} \left(1 + C_{OT}\right)^{1/\gamma} \left(\int_{-\infty}^{1} \left(1 - \left(1 - _t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\right)^{1/\gamma}} \quad \text{for all } t \in [\sigma, \infty),$$

(3.29)

where the optimal Lagrangian multiplier $\lambda_{[\sigma]}$ is given by

$$
\begin{aligned}
\lambda_{[\sigma]} = \Bigg(\frac{1}{v} \Bigg(&\int_0^{\sigma} \left(1 + C_A\right)^{1-\frac{1}{\gamma}} e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} {_t p_x}\, m_\epsilon(-\ln {_t p_x})\, dt \\
&+ \int_{\sigma}^{\infty} \left(1 + C_{OT}\right)^{1-\frac{1}{\gamma}} e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} \\
&\qquad \cdot \frac{\left(\kappa_{n,\gamma,\epsilon}(_t p_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^{1} \left(1 - \left(1 - _t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\right)^{1/\gamma - 1}} dt \Bigg) \Bigg)^{\gamma}
\end{aligned}
$$

(3.30)

and $\kappa_{n,\gamma,\epsilon}(_t p_x)$ is defined as in (3.23). The expected discounted lifetime utility is then given by

$$
\begin{aligned}
U_{[\sigma]} = \mathbb{E}\Bigg[\int_0^{\infty} e^{-\rho t} \Bigg(&\mathbb{1}_{\{0 \le t < \min\{\sigma, T_\epsilon\}\}} u\left(c^*_{[\sigma]}(t)\right) \\
&+ \mathbb{1}_{\{\sigma \le t < T_\epsilon\}} u\left(\frac{n}{N_\epsilon(t)} d^*_{[\sigma]}(t)\right) \Bigg) dt \Bigg] = \frac{\lambda_{[\sigma]}}{1-\gamma} v.
\end{aligned}
$$

(3.31)

Proof. The proof can be carried out in the same way as the proof of Theorem 3.1. □

Note that the optimal annuity and tontine payoffs within the antine, (3.28) and (3.29), structurally coincide with those of the tonuity ((3.21) and (3.20)) and differ only in the intervals on which they are defined. In particular, the optimal annuity payoff is again decreasing, constant, or increasing over time if $\rho > r$, $\rho = r$, or $\rho < r$, respectively.

As Theorem 3.2 holds for any σ, it is again possible, by the same method as before, to numerically find the integer optimal switching time $\sigma^* \in \{0, 1, \dots, 54, 55\}$ for the antine such that the highest lifetime utility is achieved for a specific policyholder. The resulting optimal payoff functions are then denoted by $c^*_{[\sigma^*]}(t)$ and $d^*_{[\sigma^*]}(t)$.

Similar to the case of the tonuity, we subsequently briefly analyze the optimal payoff of the antine graphically when applying the base case parameter setup specified in Table 1. The corresponding optimal switching time of the antine is given by $\sigma^* = 0$ as the highest lifetime utility is attained at this time
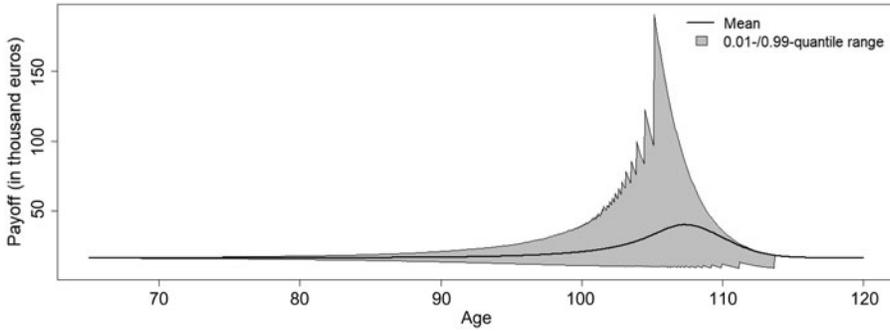
FIGURE 2: Mean and 0.01-/0.99-quantile range of the optimal payoff of the antine with $\sigma^* = 0$ depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05. The plot is based on the assumption that the considered policyholder is always alive.

when considering $\{0, 1, \ldots, 54, 55\}$ as the possible choices for σ. That is, optimally, the antine coincides with a tontine as the (theoretical) switch from the annuity to the tontine occurs right at the outset. In Figure 2, we present, depending on the policyholder's age, the applied optimal antine payoff, that is,

$$\mathbb{1}_{\{0 \leq t < \sigma^*\}} c^*_{[\sigma^*]}(t) + \mathbb{1}_{\{\sigma^* \leq t\}} \frac{n}{N_\epsilon(t)} d^*_{[\sigma^*]}(t) = \frac{n}{N_\epsilon(t)} d^*_{[\sigma^*]}(t). \tag{3.32}$$

As Figure 2 displays features of a pure tontine, we can, for a very long period, detect the same trend behavior as in Figure 1 referring to the tonuity when the tontine defines the tonuity: the payoff uncertainty increases over time, where the possible upward movement grows to a much greater extent, so that, at ages around 105 years, the payoff can potentially even far exceed 100 thousand euros. As a consequence thereof, the average payoff to the policyholder increases until these high ages. However, we can also observe that this average payoff declines afterward. This is due to the fact that the chances to receive lower payments than expected remain, whereas the ones to receive larger payments than expected rapidly diminish. Eventually, the average payoff flattens out and any type of uncertainty in the payoff stops as it is extremely likely that the tontine pool contains only the considered policyholder from the age of around 114 years on and that all the other participants have passed away earlier. Note that, in general, dents in the upper and lower quantile curves can appear at older ages, as is the case with Figure 2, due to the rising effect of a death of another participant in the tontine pool in this age range on the payoff.

3.3. Portfolio

In contrast to the optimization problems for the tonuity and the antine, the optimization problem for the case with a portfolio consisting of an annuity

and a tontine cannot be solved explicitly. It can be written as:

$$\max_{(c_{AT}(t),\, d_{AT}(t))_{t\in[0,\infty)}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\mathbb{1}_{\{T_\epsilon>t\}}u\left(c_{AT}(t)+\frac{n}{N_\epsilon(t)}d_{AT}(t)\right)dt\right]$$

$$\text{subject to}\quad v=\widetilde{P}_0^{AT}=\widetilde{P}_0^{A,AT}+\widetilde{P}_0^{OT,AT}.$$

(3.33)

In this retirement plan, the individual maximizes her utility simultaneously over the payoff functions $c_{AT}(t)$ and $d_{AT}(t)$. Thus, within the utility maximization problem, not only the optimal structures of $c_{AT}(t)$ and $d_{AT}(t)$ are determined, but also implicitly the fractions of initial wealth invested in the annuity and the tontine.[2] The Lagrangian function corresponding to optimization problem (3.33) can be calculated as

$$
\begin{aligned}
\mathcal{L} &= \int_0^\infty e^{-\rho t}\mathbb{E}\left[\mathbb{1}_{\{T_\epsilon>t\}}u\left(c_{AT}(t)+\frac{n}{N_\epsilon(t)}d_{AT}(t)\right)\right]dt+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right) \\
&= \int_0^\infty e^{-\rho t}\mathbb{E}\left[{}_tp_x^{1-\epsilon}\mathbb{E}\left[u\left(c_{AT}(t)+\frac{n}{N_\epsilon(t)}d_{AT}(t)\right)\Big|T_\epsilon>t,\epsilon\right]\right]dt \\
&\quad +\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right) \\
&= \int_0^\infty e^{-\rho t}\mathbb{E}\left[{}_tp_x^{1-\epsilon}\sum_{k=0}^{n-1}u\left(c_{AT}(t)+\frac{n}{k+1}d_{AT}(t)\right)\binom{n-1}{k}\left({}_tp_x^{1-\epsilon}\right)^k\right. \\
&\quad \left.\cdot\left(1-{}_tp_x^{1-\epsilon}\right)^{n-1-k}\right]dt+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right) \\
&= \int_0^\infty e^{-\rho t}\sum_{k=0}^{n-1}u\left(c_{AT}(t)+\frac{n}{k+1}d_{AT}(t)\right)\binom{n-1}{k}\int_{-\infty}^1\left({}_tp_x^{1-\varphi}\right)^{k+1}\left(1-{}_tp_x^{1-\varphi}\right)^{n-1-k} \\
&\quad \cdot f_\epsilon(\varphi)\,d\varphi\,dt+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right),
\end{aligned}
$$

(3.34)

where λ_{AT} is the Lagrangian multiplier. The first-order conditions with respect to $c_{AT}(t)$, $d_{AT}(t)$ and λ_{AT} are given as

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial c_{AT}(t)} &= e^{-\rho t}\sum_{k=0}^{n-1}u'\left(c_{AT}(t)+\frac{n}{k+1}d_{AT}(t)\right)\binom{n-1}{k} \\
&\quad \cdot\int_{-\infty}^1\left({}_tp_x^{1-\varphi}\right)^{k+1}\left(1-{}_tp_x^{1-\varphi}\right)^{n-1-k} \\
&\quad \cdot f_\epsilon(\varphi)\,d\varphi-\lambda_{AT}(1+C_A)e^{-rt}{}_tp_x\,m_\epsilon(-\ln{}_tp_x)\overset{!}{=}0,
\end{aligned}
$$

(3.35)

$$\frac{\partial \mathcal{L}}{\partial d_{AT}(t)} = e^{-\rho t} \sum_{k=0}^{n-1} u'\left(c_{AT}(t) + \frac{n}{k+1}d_{AT}(t)\right) \frac{n}{k+1}\binom{n-1}{k}\int_{-\infty}^{1}\left({}_tp_x^{1-\varphi}\right)^{k+1}$$
$$\cdot \left(1 - {}_tp_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\,d\varphi - \lambda_{AT}(1 + C_{OT})e^{-rt}$$
$$\cdot \int_{-\infty}^{1}\left(1 - \left(1 - {}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\,d\varphi \overset{!}{=} 0 \tag{3.36}$$

and

$$v = \widetilde{P}_0^{AT}. \tag{3.37}$$

As the first derivative of a CRRA utility function of a sum is generally not equal to the corresponding sum of the first derivatives of this utility function, the system of equations given in (3.35)–(3.37) can only be solved numerically for the optimal payoff functions and the Lagrangian multiplier.

Although the system of equations (3.35)–(3.37) can only be solved numerically, there are quite a few general conclusions that we can draw from this system of equations. They are summarized in Proposition 3.3.

Proposition 3.3. *Consider problem* (3.33). *Then the following holds true (with the same notations as in Theorem 3.1):*

1. *The solution to problem* (3.33) *is a 100% investment in the annuity, that is, the solution is $d_{AT}(t) = 0$, $c_{AT}(t) = c_{[0]}^*(t)$ and $\lambda_{AT} = \lambda_{[0]}$, if and only if $C_A \le C_{OT}$.*
2. *If and only if*

$$C_A \ge C_A^{\mathrm{crit}} := (1 + C_{OT})\max_{t \ge 0}\frac{\kappa_{n,\gamma+1,\epsilon}({}_tp_x)\int_{-\infty}^{1}\left(1 - \left(1 - {}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\,d\varphi}{\kappa_{n,\gamma,\epsilon}({}_tp_x)\,{}_tp_x\,m_\epsilon(-\ln {}_tp_x)} - 1, \tag{3.38}$$

the solution to problem (3.33) *is a 100% investment in the tontine, that is, the solution is $d_{AT}(t) = d_{[\infty]}^*(t)$, $c_{AT}(t) = 0$ and $\lambda_{AT} = \lambda_{[\infty]}$.*
3. *Consequently, if and only if*

$$C_{OT} < C_A < C_A^{\mathrm{crit}}, \tag{3.39}$$

the optimal solution to problem (3.33) *is investing in both annuity and tontine.*

Proof. See Appendix A.2.   □

In Table 2, we present the critical annuity loading C_A^{crit} in dependence of the risk aversion γ. We observe that the critical loading increases in the risk aversion. For an investor with a CRRA of 0.5, an annuity loading of 21% prevents her from investing in the annuity at all. The more risk averse the policyholder is, the more she prefers an annuity over a tontine which is reflected in a higher loading this policyholder is willing to pay. From $\gamma = 2$ on, the loading is unrealistically high as it is nearly 100% and even greater for larger values of γ.

TABLE 2

CRITICAL ANNUITY LOADING C_A^{crit} AS DEFINED IN
(3.38) DEPENDING ON THE RISK AVERSION ($C_{OT} = 0.01\%$).

γ	0.5	2	4	6	8
C_A^{crit}	0.21	0.96	2.12	3.93	6.86

The parameters (except for γ) are taken from Table 1.

Hence, this critical magnitude needs to be extremely large such that a pure tontine becomes optimal for the policyholder or so that there is no investment in the annuity at all.

In Appendix B, we provide a pseudocode for the numerical determination of $c_{AT}^*(t)$ and $d_{AT}^*(t)$ under general parameters. Once these two functions are determined, it is possible to compute their initial market value and, consequently, the fractions of wealth initially invested in the annuity and the tontine. The optimal fraction of initial wealth invested in the annuity is the gross premium of the annuity computed from the optimal annuity payoff divided by the initial wealth. From this quantity, we can determine how the initial wealth of the policyholder shall be split optimally between the tontine and the annuity in the beginning. Further details on this are discussed in Section 4.2.

As for the tonuity and the antine, we also show the numerical mean and 0.01-/0.99-quantile range of the (approximately) optimal payoff of the portfolio in Figure 3(a) underlying again the base case parameter setup as given in Table 1. Given the survival of the policyholder, the applied optimal portfolio payoff at any time t is here given by

$$c_{AT}^*(t) + \frac{n}{N_\epsilon(t)} d_{AT}^*(t). \tag{3.40}$$

For a better understanding of the curve progressions in Figure 3(a), we additionally depict, in Figure 3(b), the related optimal payoff functions $c_{AT}^*(t)$ and $d_{AT}^*(t)$ of the portfolio. Here, we can see that, particularly in the first half of the considered age range, the optimal tontine payoff function $d_{AT}^*(t)$ decreases in age and hence behaves similarly to what, for example, Milevsky and Salisbury (2015) find for optimally designed tontines in their framework. In the second half, the values for $d_{AT}^*(t)$ are rather close to zero and do not seem to differ significantly anymore. Note that the payoff of the tontine in Figure 3(b) is *not* the payoff to a single individual and has to be multiplied by $\frac{n}{N_\epsilon(t)}$. Especially at extremely old ages, $d_{AT}^*(t)$ is scaled by the factor n which explains why the steep decline in annuity payments occurs at extremely old ages and why $d_{AT}^*(t)$ is not exactly zero at these old ages. Note that this payoff structure can only occur because the payoff $nd_{AT}^*(t)$ is guaranteed to the pool of policyholders by the insurer. Concerning the optimal annuity payoff, we note that, after quite a long time of playing no role for the optimal portfolio payoff at all, this payoff is first drastically increasing and then decreasing, opposed to, for instance, a constant
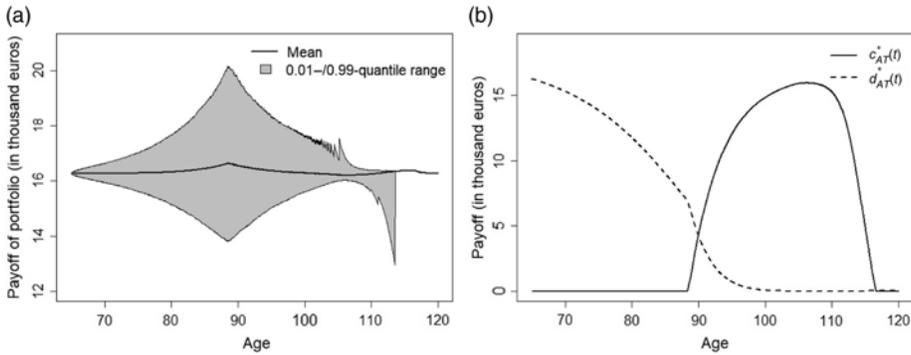
FIGURE 3: (a) Mean and 0.01-/0.99-quantile range of the optimal payoff of the portfolio consisting of an annuity and a tontine depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05. The plot is based on the assumption that the considered policyholder is always alive. (b) Optimal payoff functions of the portfolio consisting of an annuity and a tontine depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05.

annuity which turns out to be optimal in the tonuity and the antine if $\rho = r$. We therefore detect in Figure 3(b) that the policyholder defers the annuitization for some time before aiming at obtaining annuity payments at an increasing (and later on decreasing) rate, whereas the tontinization is desired right at the beginning of her retirement with a decreasing (and later on slightly increasing) rate. The reason for this structure is that the annuity provides secure payments at times when the tontine provides the most volatile payments (see also Figure 2). Right after contract initiation, the number of the tontine members is fairly stable. At extremely advanced retirement ages, the tontine payments are, again, stable, because there is only a very low probability for any other survivor to be left. Therefore, at these ages, no stable annuity payments are required. These are only necessary for ages between (in this example approximately) 89 and 116 years, where most policyholders die and, consequently, the payments of the tontine are most volatile. It is important to bear in mind that, as we consider a very general approach for the portfolio optimization problem, the individual can freely decide on the payoff structures of both the annuity and the tontine in the portfolio. Hence, at the optimum, it is possible to obtain such payoff functions as $c_{AT}^*(t)$ in Figure 3(b). In our result, it seems that $c_{AT}^*(t)$ and $d_{AT}^*(t)$ complement each other. As these payoffs are completely determined by solely maximizing the benefits of the policyholder, they might not be provided in practice. However, our analysis might still be interesting for those insurers who aim to offer products which best suit the needs of the customers to improve their competitiveness on the market. Concerning Figure 3(a), we overall make the following further observations: For a long time, there is always a certain degree of uncertainty in the optimal portfolio payoff. As the tontine, which involves randomness, prevails at earlier ages, Figure 3(a) actually resembles Figure 1 in the first several years following the withdrawal from working life. Afterward, the payoff uncertainty declines over the years as the annuity, which

stabilizes the payments, becomes more and more significant. From the age of around 107 years until the age of around 114 years, we see that the chances to receive quite low payments, compared to the expectation, grow again. The decrease in $c_{AT}^*(t)$ and the slight increase in $d_{AT}^*(t)$ in this age range, together with the presumption that, in contrast to what is expected, other policyholders in the tontine pool could nevertheless be alive at such high ages, can give reasons for this observation. After the age of around 114 years, it is however highly probable that the event that all the other policyholders are dead occurs, which leads to a virtually deterministic payoff for the remaining time.

4. COMPARISON OF RETIREMENT PLANS

In this section, we compare the expected lifetime utilities of the policyholder under the different retirement plans if their payoffs are optimally chosen, as elaborated in Section 3. We start with a theoretical comparison, where the main finding of our article is presented. Afterward, a concise numerical section follows.

4.1. Theoretical findings

Although we cannot determine the optimal payoff functions of the portfolio consisting of an annuity and a tontine analytically, as discussed in Section 3.3, we are able to explicitly compare the optimal expected utility of the portfolio with those resulting from the tonuity and the antine. Our appropriate key result is formulated in Proposition 4.1 which is generally valid and states that the optimal expected utility of the portfolio is always at least as high as that of the tonuity and of the antine. It bases on the fact that, for any switching time, any payoff of a tonuity or an antine can be replicated by a policyholder holding a portfolio of an annuity and a tontine, given the initial premiums of the retirement plans are identical. Note that, as we require no restrictions for the switching times, it is clear that this statement also holds for the optimal switching times τ^* and σ^* which we intensively use for our analyses in the next section.

Proposition 4.1. *We denote by $U_{[\tau]}$, $U_{[\sigma]}$ and U_{AT} the optimal levels of expected utility resulting from problems (3.19), (3.27) and (3.33), respectively. In particular, we assume that the premiums charged for the three retirement plans follow the expected value principle as introduced before. Then, it holds*

$$U_{AT} \geq U_{[\tau]}, \qquad U_{AT} \geq U_{[\sigma]} \tag{4.41}$$

for all possible switching times τ and σ, and for all possible risk loadings C_A and C_{OT}.

Proof. We denote by $\mathcal{A}_{[\tau]}$, $\mathcal{A}_{[\sigma]}$ and $\mathcal{A}_{AT}$ the sets of admissible solutions of optimization problems (3.19), (3.27) and (3.33), respectively, that is, the elements of these sets fulfill the respective budget constraints. Note that the payoffs of any tonuity and antine can be replicated by a portfolio consisting of an annuity and a tontine by choosing the payoffs of the annuity and the tontine appropriately. Let us, as an example, consider a tonuity with a switching time τ and payoffs $d_{[\tau]}(t)$ for $0 \leq t < \tau$ and $c_{[\tau]}(t)$ for $t \geq \tau$ which satisfy the budget constraint $v = (1 + C_{OT})P_0^{OT,\tau} + (1 + C_A)P_0^{A,\tau}$ for fixed v, C_{OT} and C_A. We can define

$$c_{AT}(t) = \begin{cases} 0, & \text{for } 0 \leq t < \tau, \\ c_{[\tau]}(t), & \text{for } t \geq \tau, \end{cases} \qquad d_{AT}(t) = \begin{cases} d_{[\tau]}(t), & \text{for } 0 \leq t < \tau, \\ 0, & \text{for } t \geq \tau, \end{cases}$$
(4.42)

as the payoffs of the portfolio. Having defined the payoffs of the portfolio, it is also clear that the budget constraint of the portfolio is satisfied as

$$v = (1 + C_{OT})P_0^{OT,\tau} + (1 + C_A)P_0^{A,\tau} = \widetilde{P}_0^{OT,AT} + \widetilde{P}_0^{A,AT},$$
(4.43)

with $\widetilde{P}_0^{OT,AT}$ and $\widetilde{P}_0^{A,AT}$ being the gross premiums of an annuity and a tontine with payoffs $d_{AT}(t)$ and $c_{AT}(t)$, respectively. Consequently, the gross premiums of the portfolio and the tonuity are equal if the portfolio generates exactly the same payoff as the tonuity under our choice of the premiums. This whole line of reasoning works similarly for an antine and therefore, we overall obtain

$$\mathcal{A}_{[\tau]} \subseteq \mathcal{A}_{AT}, \quad \mathcal{A}_{[\sigma]} \subseteq \mathcal{A}_{AT}.$$
(4.44)

As a consequence, the optimal level of expected utility of the portfolio is always at least as high as the optimal level of expected utility resulting from a tonuity and from an antine with a given switching time. □

Remark 4.2. *Throughout this article, we assume that the premiums of the three retirement plans are determined by the expected value principle. However, Proposition 4.1 still holds if another premium principle is applied. Important is that the premium principle still leads to the same premium level for the different combinations of an annuity and a tontine. Then, any payoff which is an admissible solution to the tonuity (antine) problem is also admissible to the portfolio problem. For example, the result of Proposition 4.1 remains valid under the famous variance and standard deviation principles. For a review of existing premium principles, we refer, for example, to Young (2014). For illustrative purposes, let us consider again the tonuity in comparison with the portfolio and the variance principle: if the same method to define the gross premiums of combining products as before is used, which distinguishes between the annuity and the tontine parts, then, by means of (2.8), the gross premium of the portfolio is given by*

$$\widetilde{P}_0^{AT} = \mathbb{E}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon > t\}}c_{AT}(t)\, dt\right] + C_A\text{Var}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon > t\}}c_{AT}(t)\, dt\right]$$

$$+ \mathbb{E}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon > t\}}\frac{n}{N_\epsilon(t)}d_{AT}(t)\, dt\right] + C_{OT} \tag{4.45}$$

$$\cdot\text{Var}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon > t\}}\frac{n}{N_\epsilon(t)}d_{AT}(t)\, dt\right],$$

and the one of the tonuity satisfying the budget constraint, that is, $v = \widetilde{P}_0^{[\tau]}$, can be derived similarly by means of (2.10). If the portfolio payoff functions are now chosen as in (4.42), it is clear that the portfolio budget constraint $v = \widetilde{P}_0^{[\tau]} = \widetilde{P}_0^{AT}$ is also fulfilled. By this fact and the same arguments as in the proof of Proposition 4.1, the portfolio still provides the highest (or at least the same) level of expected utility (as the tonuity) under the variance principle.

4.2. Numerical findings

In the following, we aim at numerically studying the policyholder's individual benefits resulting from the purchase of the various retirement plans. Moreover, our goal is also to learn more about the attractiveness of each retirement plan when applying different parameter combinations. Specifically, we are able to confirm the theoretical statement of Proposition 4.1 on the basis of concrete exemplary numbers.

To make results easier to interpret, we introduce the certainty equivalent CE as the level of the deterministic retirement payment that yields the same expected utility as the given retirement plan with payoff $\{\chi(t)\}_{t\geq 0}$. That is, CE is determined by

$$U\left(\{CE\}_{t\geq 0}\right) = U\left(\{\chi(t)\}_{t\geq 0}\right), \tag{4.46}$$

or equivalently,

$$CE = \left((1-\gamma)U\left(\{\chi(t)\}_{t\geq 0}\right)\left(\int_0^\infty e^{-\rho t}{}_tp_x\, m_\epsilon(-\ln {}_tp_x)\, dt\right)^{-1}\right)^{\frac{1}{1-\gamma}}, \tag{4.47}$$

where $U\left(\{\chi(t)\}_{t\geq 0}\right)$ is the expected discounted lifetime utility of the individual as defined in (3.16). Note that the certainty equivalent is an increasing function in $U\left(\{\chi(t)\}_{t\geq 0}\right)$.

Table 3 provides the optimal switching times and the corresponding resulting certainty equivalents for the tonuity and antine for different pool sizes n and different tontine risk loadings C_{OT}. As already described before, it is reasonable to assume that the loading C_{OT} decreases in the pool size n. For the dependence of the loading C_{OT} on n, we take into account the findings of Chen *et al.* (2019) regarding the safety loadings: For $n = 100$, we use 0.1% (Chen *et al.*, 2019 obtain 0.1089%), for $n = 1000$, we use 0.01% (Chen *et al.*, 2019 obtain 0.0133%) and for $n = 500$, we linearly interpolate between these two values and

TABLE 3

CERTAINTY EQUIVALENTS (IN THOUSAND EUROS) OF THE TONUITY, THE ANTINE AND THE PORTFOLIO CONSISTING OF AN ANNUITY AND A TONTINE ALONG WITH THE OPTIMAL SWITCHING TIMES AND THE FRACTION OF WEALTH INVESTED IN THE ANNUITY EMBEDDED IN THE PORTFOLIO, RESPECTIVELY, FOR DIFFERENT POOL SIZES n AND DIFFERENT TONTINE LOADINGS C_{OT}.

n	C_{OT}	Tonuity	Antine	Portfolio
100	0.1%	16.13, $\tau^* = 21$	15.91, $\sigma^* = 0$	16.14, $\widetilde{P}_0^{A,AT}/v = 0.14$
500	0.06%	16.23, $\tau^* = 26$	16.14, $\sigma^* = 0$	16.23, $\widetilde{P}_0^{A,AT}/v = 0.06$
1000	0.01%	16.25, $\tau^* = 27$	16.18, $\sigma^* = 0$	16.26, $\widetilde{P}_0^{A,AT}/v = 0.05$

The other parameters are taken from Table 1.

TABLE 4

CERTAINTY EQUIVALENTS (IN THOUSAND EUROS) OF THE TONUITY, THE ANTINE AND THE PORTFOLIO CONSISTING OF AN ANNUITY AND A TONTINE ALONG WITH THE OPTIMAL SWITCHING TIMES AND THE FRACTION OF WEALTH INVESTED IN THE ANNUITY EMBEDDED IN THE PORTFOLIO, RESPECTIVELY, FOR DIFFERENT ANNUITY LOADINGS C_A.

C_A	Tonuity	Antine	Portfolio
0.02	16.27, $\tau^* = 24$	16.18, $\sigma^* = 0$	16.29, $\widetilde{P}_0^{A,AT}/v = 0.09$
0.03	16.26, $\tau^* = 25$	16.18, $\sigma^* = 0$	16.28, $\widetilde{P}_0^{A,AT}/v = 0.07$
0.04	16.25, $\tau^* = 27$	16.18, $\sigma^* = 0$	16.26, $\widetilde{P}_0^{A,AT}/v = 0.05$
0.05	16.24, $\tau^* = 28$	16.18, $\sigma^* = 0$	16.26, $\widetilde{P}_0^{A,AT}/v = 0.04$

The other parameters are taken from Table 1.

TABLE 5

CERTAINTY EQUIVALENTS (IN THOUSAND EUROS) OF THE TONUITY, THE ANTINE AND THE PORTFOLIO CONSISTING OF AN ANNUITY AND A TONTINE ALONG WITH THE OPTIMAL SWITCHING TIMES AND THE FRACTION OF WEALTH INVESTED IN THE ANNUITY EMBEDDED IN THE PORTFOLIO, RESPECTIVELY, FOR DIFFERENT RISK AVERSIONS γ.

γ	Tonuity	Antine	Portfolio
0.8	16.33, $\tau^* = 36$	16.33, $\sigma^* = 0$	16.33, $\widetilde{P}_0^{A,AT}/v = 0.004$
2	16.30, $\tau^* = 32$	16.29, $\sigma^* = 0$	16.30, $\widetilde{P}_0^{A,AT}/v = 0.02$
4	16.27, $\tau^* = 29$	16.24, $\sigma^* = 0$	16.28, $\widetilde{P}_0^{A,AT}/v = 0.04$
6	16.25, $\tau^* = 27$	16.18, $\sigma^* = 0$	16.26, $\widetilde{P}_0^{A,AT}/v = 0.05$
8	16.23, $\tau^* = 25$	16.12, $\sigma^* = 0$	16.24, $\widetilde{P}_0^{A,AT}/v = 0.07$
10	16.22, $\tau^* = 24$	16.07, $\sigma^* = 0$	16.23, $\widetilde{P}_0^{A,AT}/v = 0.09$

The other parameters are taken from Table 1.

obtain 0.06%. Since we would like to keep our numerical analysis as simple as possible, we have decided to round the values in Chen *et al.* (2019) instead of taking the exact values. Note that the parameters in Chen *et al.* (2019) also differ slightly from ours and therefore, we only focus on the rougher magnitude when it comes to the risk loadings. Additionally, the certainty equivalents of the portfolio consisting of an annuity and a tontine along with the optimal

fraction of initial wealth invested in the annuity are provided for the different pool sizes n and different tontine risk loadings C_{OT}. The optimal fraction of initial wealth invested in the annuity can be determined as the gross premium of the annuity (computed from the optimal annuity payoff) divided by the total initial wealth. For the remaining parameters besides the pool size and the tontine risk loadings, we use the base case parameter setup as given in Table 1. Table 4 provides similar sensitivity analyses of the certainty equivalents for the risk loading of the annuity C_A. Finally, Table 5 provides similar sensitivity analyses of the certainty equivalents for the risk aversion parameter γ. Note that we have not included sensitivity analyses with respect to the subjective discount rate ρ in these numerical studies as the effects of ρ seem negligible for the given parameters. Overall, we make the following observations which are all drawn *only within our exemplary numbers.* To support the readability of our results, we have highlighted the main observations in italic letters:

- *The tonuity and the antine are outperformed by the portfolio consisting of an annuity and a tontine.* As the certainty equivalents of the portfolio are larger than those of the hybrid products tonuity and antine in all presented tables, Proposition 4.1 is numerically confirmed here. Recall that the portfolio allows the policyholder to combine the two retirement products annuity and tontine in a more general way than prescribed by the tonuity and antine. Strictly speaking, the tonuity and the antine can be seen as special cases of the portfolio as it is possible to choose the payoffs of the portfolio in such a way that they equal the payoffs of the tonuity or antine. Note, however, that the certainty equivalent of the tonuity is only negligibly smaller than that of the portfolio in all the cases considered, that is, a policyholder might as well buy a tonuity with a single switch between tontine and annuity instead of purchasing the rather complicated payoff structure of the annuity and the tontine in the portfolio shown in Figure 3(b).
- *Among the novel retirement plans tonuity and antine, the tonuity is the one which performs better.* In our parameter setup, we see that the tonuity always yields a certainty equivalent greater than that of the antine and thus, is the more attractive retirement product to the policyholder. In fact, a nontrivial antine is, in our parameter setup, always outperformed by a tontine. This is probably due to the design of the antine: when the switching time is high, the price of the antine is close to that of the appropriate annuity, but it leaves the individual holding the antine with a volatile payoff in her more advanced retirement ages. On the contrary, if the switching time is low, the payoff of the antine is close to that of the appropriate tontine, but the price of it might be still noticeably higher than that of the tontine.
- *The role of the tontine component becomes more prominent within the retirement plans if the tontine pool size increases.* This feature can be observed in Table 3: If the pool size increases, the optimal switching time of the tonuity increases as well. Furthermore, concerning the portfolio, individuals tend to invest higher fractions of their initial wealth in the annuity if the

pool size declines. All this is mainly due to the well-known fact that a tontine's attractiveness increases with the pool size (see, for example, Milevsky and Salisbury, 2015). Note that further components driving up the attractiveness of the tontine are decreasing tontine loadings (cf. Table 3) and an increasing annuity loading (cf. Table 4).

- *The role of the annuity component becomes more prominent within the retirement plans if the risk aversion level of the policyholder increases.* This feature can be observed in Table 5: If the risk aversion increases, the optimal switching time of the tonuity decreases. Furthermore, concerning the portfolio, individuals with a larger risk aversion tend to invest higher fractions of their initial wealth in the annuity.

- *Referring to the portfolio, the policyholder does not invest all her initial wealth exclusively in an annuity or a tontine.* From the fact that the fractions of initial wealth invested in the annuity are above 0 and below 1, within our exemplary numbers, neither a pure annuity nor a pure tontine is an attractive retirement plan for the majority of the policyholders. Instead, partial annuitization combined with partial tontinization turns out to deliver the highest expected lifetime utility for our parameter choices. This can also be seen from Proposition 3.3 and from Table 2, where the critical annuity loading C_A^{crit} is always larger than the considered values for C_A.

5. CONCLUSION

In this article, we consider three possibilities to combine annuities and tontines and analyze and compare their attractiveness from a policyholder's perspective in an expected utility framework. The three retirement plans we consider are the tonuity previously introduced by Chen *et al.* (2019), a new product which we call antine and a portfolio consisting of an annuity and a tontine. Our theoretical and numerical results show that the portfolio outperforms any tonuity and antine in the sense that it always delivers a higher expected lifetime utility than the two novel products. The reason for this is that a policyholder can choose the payoffs of the annuity and the tontine in the portfolio in such a way that the payoff of any tonuity and antine is replicated with the same initial investment. Consequently, the optimal level of expected utility stemming from the portfolio can never fall below that of any tonuity and antine. Additionally, we derive conditions regarding the loadings of the annuity and the tontine, under which a pure investment in the annuity, the tontine and a combination of both is optimal, respectively. We find that, under reasonable parameters, an investment in both products delivers a higher expected lifetime utility than the single products. In our parameter setup, we further find that the tonuity always delivers a higher expected lifetime utility than the antine.

While in this article we exclusively focus on the policyholder's perspective, an interesting topic for future research might be an analysis of the three considered retirement plans from the insurer's perspective. As it is less natural to assume a utility function for the insurance company, we could consider other

important quantities of interest from the insurer's perspective, like the (random) present value of future losses (cf. Bauer and Weber, 2008; Li and Hardy, 2011; Cairns, 2013; Kling *et al.*, 2014; Olivieri and Pitacco, 2019). In a recent article, Chen and Rach (2019), the authors analyze the attractiveness of options on tontines also from the insurer's point of view. The results obtained there suggest that insurers can also benefit from selling hybrid products between conventional annuities and tontines. Compared to annuities, it is likely that the tonuity, antine and the portfolio of an annuity and a tontine also reduce potential losses on the insurer's side, as all these retirement plans carry less risks for the insurer than a traditional annuity.

Acknowledgments

Notes

1. In addition to the traditional tontine, many innovative, longevity-risk-sharing retirement products, for instance, pooled annuity funds or group self-annuitization schemes, have been developed in recent years (see, e.g., Bernhardt and Donnelly, 2019 and the references therein).

2. Note that problem (3.33) is more general than the following optimization problem:

$$\max_{\alpha \in [0,1]} \mathbb{E}\left[\int_0^\infty e^{-\rho t} \mathbb{1}_{\{T_\epsilon > t\}} u\left(\alpha c_{[0]}^*(t) + (1-\alpha)\frac{n}{N_\epsilon(t)} d_{[\infty]}^*(t)\right) dt\right], \tag{5.48}$$

where $c_{[0]}^*(t)$ is the optimal annuity payoff and $d_{[\infty]}^*(t)$ is the optimal tontine payoff as obtained from Theorem 3.1 for the boundary cases $\tau = 0$ and $\tau = \infty$, respectively. Note that for the annuity case, that is, $\tau = 0$, the budget constraint is $v = \widetilde{P}_0^A$. Similarly, for the tontine case, that is, $\tau = \infty$, the budget constraint is $v = \widetilde{P}_0^{OT}$. In other words, in problem (5.48), where we maximize over a fraction of initial wealth, we do not have to consider a budget constraint, as $\alpha \widetilde{P}_0^A + (1-\alpha)\widetilde{P}_0^{OT} = \alpha v + (1-\alpha)v = v$ holds for all $\alpha \in [0,1]$. Note that this describes already the first difference between optimization problem (5.48) and the optimization problem (3.33). Furthermore, as the optimization problem in (5.48) is, by assuming the payoff functions $c_{[0]}^*(t)$ and $d_{[\infty]}^*(t)$ to be given in advance, (a lot) less general than the portfolio problem (3.33), we observe that under problem (5.48), it is no longer possible to replicate the payoff of any tonuity and antine. In this problem, the individual can only decide upon the fraction α and not upon the payment structure of the products in the portfolio. As a consequence, it is possible that the portfolio of an annuity and a tontine can be outperformed by other combinations, for example, by a tonuity.

References

Bauer, D. and Weber, F. (2008) Assessing investment and longevity risks within immediate annuities. *Asia-Pacific Journal of Risk and Insurance*, **3**(1), 89–111.

Bernhardt, T. and Donnelly, C. (2019) Modern tontine with bequest: Innovation in pooled annuity products. *Insurance: Mathematics and Economics*, **86**, 168–188.

Brown, J.R. (2007) Rational and behavioral perspectives on the role of annuities in retirement planning. Technical report, National Bureau of Economic Research.

CAIRNS, A.J. (2013) Robust hedging of longevity risk. *Journal of Risk and Insurance*, **80**(3), 621–648.

CALIENDO, F.N., GORRY, A. and SLAVOV, S. (2017) Survival ambiguity and welfare. Technical report, National Bureau of Economic Research.

CHEN, A., HABERMAN, S. and THOMAS, S. (2018) The implication of the hyperbolic discount model for the annuitisation decisions. *Journal of Pension Economics & Finance*, 1–20.

CHEN, A., HIEBER, P. and KLEIN, J.K. (2019) Tonuity: A novel individual-oriented retirement plan. *ASTIN Bulletin: The Journal of the IAA*, **49**(1), 5–30.

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

DAVIDOFF, T., BROWN, J.R. and DIAMOND, P.A. (2005) Annuities and individual welfare. *American Economic Review*, **95**(5), 1573–1590.

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.

GOTTLIEB, D. (2012) *Prospect theory, life insurance, and annuities*. The Wharton School Research Paper No. 44.

GUMBEL, E. (1958) *Statistics of Extremes*. New York: Columbia University Press.

HU, W.-Y. and SCOTT, J.S. (2007) Behavioral obstacles in the annuity market. *Financial Analysts Journal*, **63**(6), 71–82.

INKMANN, J., LOPES, P. and MICHAELIDES, A. (2010) How deep is the annuity market participation puzzle? *The Review of Financial Studies*, **24**(1), 279–319.

KLING, A., RUSS, J. and SCHILLING, K. (2014) Risk analysis of annuity conversion options in a stochastic mortality environment. *ASTIN Bulletin: The Journal of the IAA*, **44**(2), 197–236.

LI, J.S.-H. and HARDY, M.R. (2011) Measuring basis risk in longevity hedges. *North American Actuarial Journal*, **15**(2), 177–200.

LI, Y. and ROTHSCHILD, C. (2019) Selection and redistribution in the irish tontines of 1773, 1775, and 1777. *Journal of Risk and Insurance*.

LIN, Y. and COX, S.H. (2005) Securitization of mortality risks in life annuities. *Journal of Risk and Insurance*, **72**(2), 227–252.

MILEVSKY, M.A. (2013) *Life Annuities: An Optimal Product for Retirement Income*. Charlottesville: CFA Institute.

MILEVSKY, M.A. (2015) *King William's Tontine: Why the Retirement Annuity of the Future Should Resemble Its Past*. Cambridge: Cambridge University Press.

MILEVSKY, M.A. and SALISBURY, T.S. (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: The Journal of the IAA*, **46**(3), 571–604.

MITCHELL, O.S. (2002) Developments in decumulation: The role of annuity products in financing retirement. In *Ageing, Financial Markets and Monetary Policy* (eds. A.J. Auerbach and H. Herrmann), pp. 97–125. Berlin, Heidelberg: Springer.

O'DEA, C. and STURROCK, D. (2019) Survival pessimism and the demand for annuities. Technical report, Institute for Fiscal Studies.

OLIVIERI, A. and PITACCO, E. (2019) *Linking annuity benefits to the longevity experience: A general framework*. Available at SSRN: https://ssrn.com/abstract=3326672.

PEIJNENBURG, K., NIJMAN, T. and WERKER, B.J. (2016) The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, **70**, 18–35.

PIGGOTT, J., VALDEZ, E.A. and DETZEL, B. (2005) The simple analytics of a pooled annuity fund. *Journal of Risk and Insurance* **72**(3), 497–520.

POPPE-YANEZ, G. (2017) *Mortality learning and optimal annuitization*. Working paper. Available at https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Programs/Economics/Other%20docs/mortannui.pdf.

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.

ROYAL LONDON (2018) *Will we ever summit the pension mountain?* Technical report, Royal London Mutual Insurance Society Limited.

SABIN, M.J. (2010) *Fair tontine annuity*. Available at SSRN: https://ssrn.com/abstract=1579932.

STATISTA (2019) *Average risk-free investment rate in germany 2015-2019.* Available at https://www.statista.com/statistics/885774/average-risk-free-rate-germany/; accessed on October 23, 2019.

WEINERT, J.-H. and GRÜNDL, H. (2017) *The modern tontine: An innovative instrument for longevity risk management in an aging society.* Available at SSRN: https://ssrn.com/abstract=3088527.

YAARI, M.E. (1965) Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, **32**(2), 137–150.

YAGI, T. and NISHIGAKI, Y. (1993) The inefficiency of private constant annuities. *Journal of Risk and Insurance*, **60**(3), 385–412.

YOUNG, V.R. (2014) Premium principles. *Wiley StatsRef: Statistics Reference Online.*

AN CHEN
Institute of Insurance Science
Ulm University
Helmholtzstraße 20
89069 Ulm, Germany
E-Mail: an.chen@uni-ulm.de

MANUEL RACH   (Corresponding author)
Institute of Insurance Science
Ulm University
Helmholtzstraße 20
89069 Ulm, Germany
E-Mail: manuel.rach@uni-ulm.de

THORSTEN SEHNER
Institute of Insurance Science
Ulm University
Helmholtzstraße 20
89069 Ulm, Germany
E-Mail: thorsten.sehner@uni-ulm.de

APPENDIX A. PROOFS

A.1. Proof of Theorem 3.1

Note that some of the steps within this proof are similar to the results in Chen *et al.* (2019). Thus, a few derivations are shortened here and can be reviewed in the mentioned article. We first recall that $_tp_x^{1-\epsilon} = \mathbb{E}\left[\mathbb{1}_{\{T_\epsilon > t\}}\big|\epsilon\right]$ and that it further holds $(N_\epsilon(t) - 1|T_\epsilon > t, \epsilon) \sim \text{Bin}\left(n - 1, {_tp_x^{1-\epsilon}}\right)$. By means of these observations, we can write the Lagrangian function

for our optimization problem in the following way:

$$
\begin{aligned}
\mathcal{L} &= \int_0^\tau e^{-\rho t} \mathbb{E}\left[\mathbb{1}_{\{T_\epsilon > t\}}\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma}\right] u\left(d_{[\tau]}(t)\right) dt + \int_\tau^\infty e^{-\rho t} \mathbb{E}\left[\mathbb{1}_{\{T_\epsilon > t\}}\right] u\left(c_{[\tau]}(t)\right) dt \\
&\quad + \lambda_{[\tau]}\left(v - (1 + C_{OT})\int_0^\tau e^{-rt}\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\, d_{[\tau]}(t)\, dt \right. \\
&\quad \left. - (1 + C_A)\int_\tau^\infty e^{-rt}{}_t p_x\, m_\epsilon(-\ln {}_t p_x)\, c_{[\tau]}(t)\, dt\right) \\
&= \int_0^\tau e^{-\rho t}\kappa_{n,\gamma,\epsilon}({}_t p_x) u\left(d_{[\tau]}(t)\right) dt + \int_\tau^\infty e^{-\rho t}{}_t p_x\, m_\epsilon(-\ln {}_t p_x)\, u\left(c_{[\tau]}(t)\right) dt \\
&\quad + \lambda_{[\tau]}\left(v - (1 + C_{OT})\int_0^\tau e^{-rt}\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\, d_{[\tau]}(t)\, dt \right. \\
&\quad \left. - (1 + C_A)\int_\tau^\infty e^{-rt}{}_t p_x\, m_\epsilon(-\ln {}_t p_x)\, c_{[\tau]}(t)\, dt\right),
\end{aligned}
$$

where $\kappa_{n,\gamma,\epsilon}({}_t p_x)$ is explicitly calculated in Chen *et al.* (2019) and given by

$$
\kappa_{n,\gamma,\epsilon}({}_t p_x) = \mathbb{E}\left[\mathbb{1}_{\{T_\epsilon > t\}}\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma}\right] = \sum_{k=1}^n \binom{n}{k}\left(\frac{k}{n}\right)^\gamma \int_{-\infty}^1 \left({}_t p_x^{1-\varphi}\right)^k \left(1 - {}_t p_x^{1-\varphi}\right)^{n-k} f_\epsilon(\varphi)\, d\varphi.
$$

By taking partial derivatives with respect to $d_{[\tau]}(t)$ and $c_{[\tau]}(t)$, we obtain the following first-order conditions:

$$
\frac{\partial \mathcal{L}}{\partial d_{[\tau]}(t)} = e^{-\rho t}\kappa_{n,\gamma,\epsilon}({}_t p_x) d_{[\tau]}(t)^{-\gamma} - \lambda_{[\tau]}(1 + C_{OT})e^{-rt}\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi \overset{!}{=} 0, \tag{A1}
$$

$$
\frac{\partial \mathcal{L}}{\partial c_{[\tau]}(t)} = e^{-\rho t}{}_t p_x\, m_\epsilon(-\ln {}_t p_x)\, c_{[\tau]}(t)^{-\gamma} - \lambda_{[\tau]}(1 + C_A)e^{-rt}{}_t p_x\, m_\epsilon(-\ln {}_t p_x) \overset{!}{=} 0. \tag{A2}
$$

Now, by solving (A1) for $d_{[\tau]}(t)$, we get the following optimal tontine payoff:

$$
d_{[\tau]}^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}\left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma}}{\lambda_{[\tau]}^{1/\gamma}(1 + C_{OT})^{1/\gamma}\left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\right)^{1/\gamma}} \quad \text{for all } t \in [0, \tau).
$$

Similarly, by solving (A2) for $c_{[\tau]}(t)$, we obtain the following optimal annuity payoff:

$$
c_{[\tau]}^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}}{\lambda_{[\tau]}^{1/\gamma}(1 + C_A)^{1/\gamma}} \quad \text{for all } t \in [\tau, \infty).
$$

Now, we can use the budget constraint to determine the optimal Lagrangian multiplier $\lambda_{[\tau]}$. We have

$$
\begin{aligned}
v &= (1 + C_{OT}) \int_0^\tau e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \, d\varphi \, d_{[\tau]}^*(t) \, dt \\
&\quad + (1 + C_A) \int_\tau^\infty e^{-rt} {}_t p_x \, m_\epsilon(-\ln {}_t p_x) \, c_{[\tau]}^*(t) \, dt \\
&= \int_0^\tau e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma}}{\lambda_{[\tau]}^{1/\gamma} (1 + C_{OT})^{1/\gamma-1} \left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \, d\varphi\right)^{1/\gamma-1}} \, dt \\
&\quad + \int_\tau^\infty e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} {}_t p_x \, m_\epsilon(-\ln {}_t p_x) \frac{1}{\lambda_{[\tau]}^{1/\gamma} (1 + C_A)^{1/\gamma-1}} \, dt.
\end{aligned}
$$

As a consequence, we obtain

$$
\begin{aligned}
\lambda_{[\tau]} = \Bigg(\frac{1}{v} \Bigg(&\int_0^\tau (1 + C_{OT})^{1-\frac{1}{\gamma}} e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \, d\varphi\right)^{1/\gamma-1}} \, dt \\
&+ \int_\tau^\infty (1 + C_A)^{1-\frac{1}{\gamma}} e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} {}_t p_x \, m_\epsilon(-\ln {}_t p_x) \, dt \Bigg) \Bigg)^\gamma.
\end{aligned}
$$

The expected discounted lifetime utility is then given by

$$
\begin{aligned}
U_{[\tau]} &= \int_0^\tau e^{-\rho t} \mathbb{E}\left[\mathbb{1}_{\{T_\epsilon > t\}} \left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma} \right] u\left(d_{[\tau]}^*(t)\right) \, dt + \int_\tau^\infty e^{-\rho t} \mathbb{E}\left[\mathbb{1}_{\{T_\epsilon > t\}} \right] u\left(c_{[\tau]}^*(t)\right) \, dt \\
&= \int_0^\tau e^{-\rho t} \kappa_{n,\gamma,\epsilon}({}_t p_x) u\left(d_{[\tau]}^*(t)\right) \, dt + \int_\tau^\infty e^{-\rho t} {}_t p_x \, m_\epsilon(-\ln {}_t p_x) \, u\left(c_{[\tau]}^*(t)\right) \, dt \\
&= \frac{1}{1-\gamma} \int_0^\tau e^{-\rho t} \kappa_{n,\gamma,\epsilon}({}_t p_x) \\
&\quad \cdot \frac{e^{(1/\gamma-1)(r-\rho)t} \left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma-1}}{\lambda_{[\tau]}^{1/\gamma-1} (1 + C_{OT})^{1/\gamma-1} \left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \, d\varphi\right)^{1/\gamma-1}} \, dt \\
&\quad + \frac{1}{1-\gamma} \int_\tau^\infty e^{-\rho t} {}_t p_x \, m_\epsilon(-\ln {}_t p_x) \frac{e^{(1/\gamma-1)(r-\rho)t}}{\lambda_{[\tau]}^{1/\gamma-1} (1 + C_A)^{1/\gamma-1}} \, dt \\
&= \frac{\lambda_{[\tau]}^{1-1/\gamma}}{1-\gamma} \Bigg(\int_0^\tau (1 + C_{OT})^{1-\frac{1}{\gamma}} e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t p_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \, d\varphi\right)^{1/\gamma-1}} \, dt \\
&\quad + \int_\tau^\infty (1 + C_A)^{1-\frac{1}{\gamma}} e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} {}_t p_x \, m_\epsilon(-\ln {}_t p_x) \, dt \Bigg) \\
&= \frac{\lambda_{[\tau]}^{1-1/\gamma}}{1-\gamma} \lambda_{[\tau]}^{\frac{1}{\gamma}} v = \frac{\lambda_{[\tau]}}{1-\gamma} v.
\end{aligned}
$$

$\square$

A.2. Proof of Proposition 3.3

1. Clearly, the budget constraint (3.37) is fulfilled when choosing $d_{AT}(t) = 0$ and $c_{AT}(t) = c_{[0]}^*(t)$. Now consider condition (3.35). We plug in $d_{AT}(t) = 0$, $c_{AT}(t) = c_{[0]}^*(t)$ and $\lambda_{AT} = \lambda_{[0]}$:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial c_{AT}(t)} &= e^{-\rho t} \sum_{k=0}^{n-1} u'\left(c_{[0]}^*(t)\right) \binom{n-1}{k} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{k+1} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[0]}(1 + C_A) e^{-rt} {}_t p_x\, m_\epsilon(-\ln {}_t p_x) \\
&= e^{-\rho t} u'\left(c_{[0]}^*(t)\right) \mathbb{E}\left[\sum_{k=0}^{n-1} \binom{n-1}{k} \left({}_t p_x^{1-\epsilon}\right)^{k+1} \left(1 - {}_t p_x^{1-\epsilon}\right)^{n-1-k}\right] \\
&\quad - \lambda_{[0]}(1 + C_A) e^{-rt} {}_t p_x\, m_\epsilon(-\ln {}_t p_x) \\
&= e^{-\rho t} \left(c_{[0]}^*(t)\right)^{-\gamma} \mathbb{E}\left[{}_t p_x^{1-\epsilon} \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} \left({}_t p_x^{1-\epsilon}\right)^{k} \left(1 - {}_t p_x^{1-\epsilon}\right)^{n-1-k}}_{=1}\right] \\
&\quad - \lambda_{[0]}(1 + C_A) e^{-rt} {}_t p_x\, m_\epsilon(-\ln {}_t p_x) \\
&= 0.
\end{aligned}
$$

Regarding condition (3.36), we obtain:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial d_{AT}(t)} &= e^{-\rho t} \sum_{k=0}^{n-1} u'\left(c_{[0]}^*(t)\right) \frac{n}{k+1} \binom{n-1}{k} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{k+1} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[0]}(1 + C_{OT}) e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi \\
&= e^{-\rho t} u'\left(c_{[0]}^*(t)\right) \sum_{j=1}^{n} \binom{n}{j} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{j} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-j} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[0]}(1 + C_{OT}) e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi \\
&= e^{-\rho t} u'\left(c_{[0]}^*(t)\right) \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[0]}(1 + C_{OT}) e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi \\
&= \lambda_{[0]}(C_A - C_{OT}) e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi. \tag{A3}
\end{aligned}
$$

Clearly, (A3) is equal to zero if $C_A = C_{OT}$. That is, the optimum is exactly achieved for $d_{AT}(t) = 0$ if $C_A = C_{OT}$.

Let us now take a look at the second-order derivative:

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial d_{AT}(t)^2} &= e^{-\rho t} \sum_{k=0}^{n-1} u''\left(c_{AT}(t) + \frac{n}{k+1} d_{AT}(t)\right) \left(\frac{n}{k+1}\right)^2 \binom{n-1}{k} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{k+1} \\
&\quad \cdot \left(1 - {}_t p_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, d\varphi < 0,
\end{aligned}
$$

because $u''\left(c_{AT}(t) + \frac{n}{k+1} d_{AT}(t)\right) < 0$ for all values of $c_{AT}(t)$ and $d_{AT}(t)$. This implies that $\frac{\partial \mathcal{L}}{\partial d_{AT}(t)}$ is strictly decreasing in $d_{AT}(t)$. Note that $\lambda_{[0]}$ is greater than zero for all choices of C_A. Therefore, the term in (A3) is smaller than zero for $C_A < C_{OT}$. The expected lifetime utility could thus be increased at $d_{AT}(t) = 0$ if $C_A < C_{OT}$ by choosing $d_{AT}(t)$ even smaller than zero. Since we do not allow for negative payments of the tontine, the optimal portfolio is thus again a 100% investment in the annuity resulting in the payoff $c^*_{[0]}(t)$.

2. We now consider the first-order conditions for $d_{AT}(t) = d^*_{[\infty]}(t)$, $c_{AT}(t) = 0$ and $\lambda_{AT} = \lambda_{[\infty]}$. It is again clear that the budget constraint (3.37) is fulfilled. Regarding (3.35), we obtain

$$
\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial c_{AT}(t)} \\
&= e^{-\rho t} \sum_{k=0}^{n-1} u'\left(\frac{n}{k+1} d^*_{[\infty]}(t)\right) \binom{n-1}{k} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{k+1} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[\infty]}(1 + C_A)e^{-rt} {}_t p_x m_\epsilon(-\ln {}_t p_x) \\
&= e^{-\rho t} u'\left(d^*_{[\infty]}(t)\right) \sum_{k=0}^{n-1} \left(\frac{n}{k+1}\right)^{1-(\gamma+1)} \binom{n-1}{k} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{k+1} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[\infty]}(1 + C_A)e^{-rt} {}_t p_x m_\epsilon(-\ln {}_t p_x) \\
&= e^{-\rho t} \left(d^*_{[\infty]}(t)\right)^{-\gamma} \sum_{j=1}^{n} \left(\frac{j}{n}\right)^{\gamma+1} \binom{n}{j} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{j} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-j} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[\infty]}(1 + C_A)e^{-rt} {}_t p_x m_\epsilon(-\ln {}_t p_x) \\
&= e^{-rt} \frac{\kappa_{n,\gamma+1,\epsilon}({}_t p_x)}{\kappa_{n,\gamma,\epsilon}({}_t p_x)} \lambda_{[\infty]}(1 + C_{OT}) \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[\infty]}(1 + C_A)e^{-rt} {}_t p_x m_\epsilon(-\ln {}_t p_x) \overset{!}{\leq} 0.
\end{aligned}
\tag{A4}
$$

Similarly as in the first part of this proof, we want the derivative to be smaller or equal than zero for all $t \geq 0$. If it is equal to zero, the optimum is reached, if it is below zero, the utility can be increased by choosing negative payoffs for the annuity which we do not allow. Solving inequality (A4) (which has to hold for all $t \geq 0$) for C_A delivers (3.38). We still need to check whether (3.36) is fulfilled. We proceed as before:

$$
\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial d_{AT}(t)} \\
&= e^{-\rho t} \sum_{k=0}^{n-1} u'\left(\frac{n}{k+1} d^*_{[\infty]}(t)\right) \frac{n}{k+1} \binom{n-1}{k} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{k+1} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[\infty]}(1 + C_{OT})e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi \\
&= e^{-\rho t} \left(d^*_{[\infty]}(t)\right)^{-\gamma} \sum_{j=1}^{n} \left(\frac{j}{n}\right)^{\gamma} \binom{n}{j} \int_{-\infty}^{1} \left({}_t p_x^{1-\varphi}\right)^{j} \left(1 - {}_t p_x^{1-\varphi}\right)^{n-j} f_\epsilon(\varphi)\, d\varphi \\
&\quad - \lambda_{[\infty]}(1 + C_{OT})e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^{n}\right) f_\epsilon(\varphi)\, d\varphi
\end{aligned}
$$

$$= e^{-\rho t} \left(d^*_{[\infty]}(t) \right)^{-\gamma} \kappa_{n,\gamma,\epsilon}({}_t p_x) - \lambda_{[\infty]}(1 + C_{OT})e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_t p_x^{1-\varphi} \right)^n \right) f_\epsilon(\varphi)\, d\varphi$$

$$= 0.$$

3. Condition (3.39) and the third part of Proposition 3.3 directly follow from the first and the second part.

$\square$

APPENDIX B. PSEUDOCODE FOR SOLUTION OF PORTFOLIO PROBLEM (3.33)

The pseudocode given below delivers the optimal payoff functions $c^*_{AT}(t)$, $d^*_{AT}(t)$ and the Lagrangian multiplier λ_{AT}. The objective is to simultaneously fulfill Equations (3.35)–(3.37). Our code relies heavily on the bisection method, which we apply repeatedly in three while loops until (3.35)–(3.37) are all (approximately) fulfilled.

1. Initialize n, γ, r, ρ, v, x, m, β, tol, C_A, C_{OT}.
2. Specify a grid of time points $t_1, \ldots, t_N$, where $t_1 = 0$ and t_N lies sufficiently far ahead in the future.
3. Choose upper and lower bounds λ_u and λ_l for $\lambda = \lambda_{AT}$ and set $\lambda = \frac{1}{2}(\lambda_u + \lambda_l)$.
4. While $\left| v - \widetilde{P}_0^{A,AT} - \widetilde{P}_0^{OT,AT} \right| > tol$

   (a) Choose upper and lower bounds $d_u(t_i)$ and $d_l(t_i)$ for $i = 1, \ldots, N$. Set $d_{AT}(t_i) = \frac{1}{2}(d_u(t_i) + d_l(t_i))$.

   (b) While $\max_i \left| \frac{\partial \mathcal{L}}{\partial d_{AT}(t_i)} \right| > tol$

        i. Choose upper and lower bounds $c_u(t_i)$ and $c_l(t_i)$ for $i = 1, \ldots, N$. Set $c_{AT}(t_i) = \frac{1}{2}(c_u(t_i) + c_l(t_i))$.

        ii. While $\max_i \left| \frac{\partial \mathcal{L}}{\partial c_{AT}(t_i)} \right| > tol$

           A. For all i with $\frac{\partial \mathcal{L}}{\partial c_{AT}(t_i)} > 0$, set $c_l(t_i) = c_{AT}(t_i)$.

           B. For all i with $\frac{\partial \mathcal{L}}{\partial c_{AT}(t_i)} < 0$, set $c_u(t_i) = c_{AT}(t_i)$.

           C. Set $c_{AT}(t_i) = \frac{1}{2}(c_u(t_i) + c_l(t_i))$.

        iii. For all i with $\frac{\partial \mathcal{L}}{\partial d_{AT}(t_i)} > 0$, set $d_l(t_i) = d_{AT}(t_i)$.

        iv. For all i with $\frac{\partial \mathcal{L}}{\partial d_{AT}(t_i)} < 0$, set $d_u(t_i) = d_{AT}(t_i)$.

        v. Set $d_{AT}(t_i) = \frac{1}{2}(d_u(t_i) + d_l(t_i))$.

   (c) Once $c_{AT}(t_i)$ and $d_{AT}(t_i)$ are computed for $i = 1, \ldots, N$, we can linearly interpolate between these values. This enables us to compute $\widetilde{P}_0^{A,AT}$ and $\widetilde{P}_0^{OT,AT}$ by numerical integration.

   (d) If $v - \widetilde{P}_0^{A,AT} - \widetilde{P}_0^{OT,AT} > 0$, set $\lambda_u = \lambda$.

(e) If $v - \widetilde{P}_0^{A,AT} - \widetilde{P}_0^{OT,AT} < 0$, set $\lambda_l = \lambda$.

(f) Set $\lambda = \frac{1}{2}(\lambda_u + \lambda_l)$.

5. Once the (approximately) "true" λ, $c_{AT}^*(t_i)$ and $d_{AT}^*(t_i)$ are computed for $i = 1, \ldots, N$, the optimal level of expected utility and the certainty equivalent can be computed by numerical integration.

Since we know how the first-order conditions behave in c, d and λ, we know how to search for the solutions using the bisection method and that it will deliver a correct solution and converge, because it only terminates if the prespecified level of accuracy is reached.

The average computing time mostly depends on the discretization of the time axis and the pool size n. The smaller the discretized time steps are, and the larger the pool size n is, the longer it takes for the code to finish. We found that the discretization of the time axis does, actually, not have a huge impact on the results, and have therefore chosen a discretization of $N = 100$ steps from 0 to a maturity of 55. For this discretization and $n = 100$, the code only needs about 1 min to determine the optimal payoffs and the resulting certainty equivalent for one policyholder. For the current base case pool size $n = 1000$, it takes about 4 min.