The modern tontine

An innovative instrument for longevity risk management in an aging society

Jan-Hendrik Weinert[1] · **Helmut Gründl[1]**

Abstract

We investigate whether a historical pension concept, the tontine, yields enough innovative potential to extend and improve the prevailing privately funded pension solutions in a modern way. The tontine basically generates an age-increasing cash flow, which can help to match the increasing financing needs at old ages. In contrast to traditional pension products, however, the tontine generates volatile cash flows, which means that the insurance character of the tontine cannot be guaranteed in every situation. By employing Multi Cumulative Prospect Theory (MCPT) we answer the question to what extent tontines can be a complement to or a substitute for traditional annuities. We find that it is only optimal to invest in tontines for a certain range of initial wealth. In addition, we investigate in how far the tontine size, the volatility of individual liquidity needs and expected mortality rates contribute to the demand for tontines.

Keywords Life insurance · Tontines · Annuities · Asset allocation · Retirement Welfare · Aging society

✉ Jan-Hendrik Weinert
  weinert@finance.uni-frankfurt.de

1 International Center for Insurance Regulation, Faculty of Economics and Business Administration, Goethe University Frankfurt, Frankfurt, Germany

1 Introduction

Through changes in social, financial and regulatory conditions, both life insurance policyholders and life insurers are facing big challenges. One of the large social challenges in most of the Western countries is the demographic change caused by declining birth rates and an increasing longevity of the population.[1] Therefore, the old-age-dependency ratio[2] rises. As a result, pay-as-you-go retirement systems are under pressure while funded retirement products gain relevance. In addition, the liquidity need increases for elderly people, which is mainly driven by increasing medical expenses at old ages. According to a study by Standard Life [52], the liquidity need of persons older than 85 years is six times higher than for persons below 65 years of age. As a consequence, the demand for funded retirement products that help to diminish the pension provision gap in an aging society can be expected to increase. In this context it is, however, surprising to see that life care annuities that combine life annuities with long-term care insurance have relatively low market volumes,[3] and that product innovation for the decumulation phase is limited in Europe.[4]

Insurance companies are exposed to changing financial and regulatory conditions. Traditional pension, health and long-term care insurance products often entail minimum return guarantees, which providers try to ensure by investing extensively in fixed-income securities.[5] However, the current low-interest environment clearly shows large solvency risks caused by the issuance of lifetime guarantees. The possible way out of this problem, i.e., to invest extensively in more profitable asset classes like stocks, is however restricted due to its higher risk and its limited ability to cover granted guarantees. Furthermore, providers of pension products are exposed to the longevity risk of their customers, which can only be partially passed on to them.[6] Therefore, alternative solutions for providing pensions, long-term care and health insurance are needed.

In the European Union and many other parts of the world, the regulatory conditions for providers of private pension products change substantially with the introduction of risk-based solvency regulation. The market-consistent valuation of

---

[1] See for example Statistisches Bundesamt [54]. for a prognosis for the demographic change until the year 2060 in Germany and United States Census Bureau [56] for a prognosis for the demographic change until the year 2050 in USA.

[2] The old-age-dependency ratio is the ratio between the number of persons aged 65 and above and the number of persons aged between 15 and 64.

[3] In Germany, e.g., in 2018 the number of long-term care riders on life insurance products (not only annuities) is relatively small (0.6 million) compared with the number of annuity contracts (38.5 million), or compared with, e.g., the number of accident insurance riders (5.4 million), although the number of long-term care riders has gone up by 9.7% in 2018; see Gesamtverband der Deutschen Versicherungswirtschaft e.V. [20], pp. 15 and 17.

[4] See European Insurance and Occupational Pensions Authority [15], p. 49, and Financial Conduct Authority [17], p. 15.

[5] See Berdin and Gründl [5].

[6] See, e.g., the legal provisions for profit participation for German life insurance products. There, 90% of mortality gains must be passed on to the policyholders, whereas mortality losses must be completely borne by the insurer; see §7 MindZV (profit participation regulation).

investments as well as of technical provisions immediately reveals the addressed high risks of traditional life and health insurance products involved and therefore can cause severe financial imbalance for life insurers. In a traditional insurance context, managing these risks requires considerable equity capital backing or other comprehensive risk management activities like re-insurance or securitization. Such risk management ultimately has to be funded by higher insurance premiums, which might make pension planning unattractive.

The changes in social, financial and regulatory conditions therefore lead to the quest for innovative instruments for private pension planning. A product innovation should optimally reduce investment guarantees and risks related to longevity, and nevertheless be able to provide reliable insurance performance. At the same time it should meet the concerns of increasing liquidity needs at old ages.

Against this backdrop, we transfer the idea of the historic tontine to a modern context and analyze whether it can help to solve the aforementioned problems. A tontine provides a mortality-driven, age-increasing payout structure. Although an insurer can easily replicate such a payout structure, the tontine has the big advantage of its simplicity and low costs. While traditional insurance products entail large safety and administrative cost loadings,[7] a tontine can be offered at low additional costs.[8] This is because a tontine is a simple redistribution mechanism of the invested funds without guarantees and the need for an active management. The investment strategy of the tontinized wealth can be decided on an individual basis according to the individual risk aversion, without the issuance of guarantees. Due to the linkage between tontine returns and individual survival prospects, tontine payments are very low in younger years and increase sharply for very high ages.

In this article, we will assess tontines as to their ability to preserve people's standard of living at old ages. Therefore, we take the perspective of a retiree and build a portfolio of traditional life annuities and tontines and evaluate the resulting income stream relative to the standard of living, i.e. the liquidity need at a certain age, as a reference point. To evaluate an income stream relative to a reference point we employ Multi Cumulative Prospect Theory (MCPT), as applied by Ruß and Schelling [48], which is based on Cumulative Prospect Theory (CPT), originated by Kahneman and Tversky [33] and enhanced by Tversky and Kahneman [55].[9] CPT is also widely used in asset pricing (e.g. Barberis and Huang [3]) and portfolio selection (e.g. He and Zhou [25]) literature. Finding an optimal, i.e. lifetime-utility maximizing combination of traditional annuities and a tontine investment will answer the question to what extent tontines can be a complement to or a substitute for traditional annuities.

The main result in our base case scenario is that it is optimal to invest a fraction of wealth in a tontine. The utility loss through low tontine payoffs before the age of 80 is outweighed by the utility gains through high payoffs at older ages.

---

[7] According to Bundesanstalt für Finanzdienstleistungsaufsicht [10] the average acquisition and administrative costs for German life insurers are 10.7% of the gross premiums.

[8] See Weinert [59] for a cost analysis of tontines.

[9] Ruß and Schelling [48] apply Cumulative Prospect Theory in a multi-period context.

However, tontines are only in demand among relatively wealthy people. By assuming average mortality rates, there is a positive demand for tontines by people with initial wealth (at the age of 62) of more than EUR 571,000. For less wealthy people the sole use of traditional annuities is optimal because annuities are needed for at least partly covering the liquidity need before the age of 80, whereas any tontine investment extracts payments from these years of retirement. The additional income generated by tontines in later years cannot offset the loss of utility in the early years of retirement.

For people with an initial wealth of more than EUR 751,000 the traditional annuity is sufficient to cover the liquidity need at all ages. Extracting income before the age of 80, when the likelihood of being alive is high, reduces utility more sharply than utility gets increased through high payments at older ages, in which tontine payments are not needed for meeting the liquidity needs. These results in principle also apply when using different probability weights under the MCPT or when applying a von-Neumann-Morgenstern utility maximization.

Through the pooling effect, a greater tontine size reduces the volatility of tontine payoffs, and therefore extends the range of initial wealth endowments for which tontine investments become advantageous.

The higher the volatility of the liquidity need, the lower is the demand for tontines, because annuities are then the better tool for closing possible liquidity gaps in the early years of retirement. The higher the liquidity need is in the later years of retirement, the higher is the demand for tontines.

Employing lower-than-average mortality rates can be a consequence either of using subjective probabilities or belonging to a specific population group. More wealthy people, for whom we find it optimal to purchase tontines, also face lower mortality rates.[10] We find that tontines become even more favorable for people with lower mortality rates, especially because they can enjoy the higher tontine payoffs for a longer time with higher probabilities. In addition, the range of wealth endowments for which tontine purchases are advantageous widens.

The remainder of the article is organized as follows: Sect. 2 introduces the general concept of tontines. Section. 3 reviews the relevant literature on tontines and increasing liquidity needs at old ages. Section. 4 introduces our model framework specifying the underlying mortality dynamics, the tontine model as well as the old-age liquidity need curve and the valuation of annuities. Finally, we propose a Cumulative Prospect Theory based valuation of lifetime utility of tontines and annuities. In Sect. 5, we first describe the data and the calibration we adopt and provide findings for the optimal individual wealth allocation and discuss our results. Section 6 provides implications and our conclusion.

---

[10] See Demakakos et al. [13].

2 Tontines

The Italian Lorenzo de Tonti invented a product to consolidate the French public-sector deficit in the early 1650s, which was introduced in 1689.[11] His ideas were based on the pooling of persons by considering their mortality risk. The innovation was that, in exchange for a lump sum payment to the French government, one received the right to a yearly, lifelong pension, which increased over time because the yields were distributed among a lower number of surviving beneficiaries. The last survivor thus received the pensions of all others who died before. As Manes [38] notes, the valuation of the original tontine was inaccurate, retirees were grouped in broad age classes, and so the contract terms were not fair in an actuarial sense. In this article, we build upon a fair tontine based on Sabin [49] that allows participants to be of any age, of any gender, and to invest a desired amount of money[12] in the tontine. Furthermore, the tontine is revolving, which means that new members can join the tontine at any age and take on the position of deceased members. Apart from that it is not allowed to leave the tontine before passing away. The tontine is a fair lottery for every member. Expected individual tontine payments equal the individual investment in the tontine, yielding an unconditional expected profit of zero. Expected tontine payments depend on the individual stake in the tontine and on the individual survival probability. On the one hand, if a member dies, he or she loses the entire stake, while, on the other, if he or she survives he or she receives some fraction of the stake of deceased members. To be fair, expected gains and losses are equal in each period. Because the survival probability declines with age and tontine payments are only paid if one survives, the probability of receiving tontine payments decreases. To counterbalance the otherwise induced reduction in expected tontine payments, the size of the payments one receives has to increase. Through this mortality-driven feature, the expected conditional tontine payments increase with age. Mortality, therefore, is the crucial factor for determining the tontine benefit structure. For example, a man born in 1981 has a life expectancy of 73 years, while a man born in 2020 has an increased life expectancy of 81 years.[13] This difference of 8 years translates directly into different benefit patterns, especially at old ages. Furthermore, the whole composition of the demographic structure of the tontine members changes on the basis of the population mortality, which also impacts the actual benefit structure. Therefore, it is important to model and forecast the development of mortality and demographic structure of the tontine members. We use the one-factor model by Lee and Carter [36] to forecast mortality, which is the standard approach to model mortality rates.[14]

---

[11] See McKeever [40] for an overview of the history of tontines.

[12] According to Sabin [49] a fair tontine is a tontine in which the distribution to surviving participants is made in unequal portions according to a plan that provides each participant with a fair bet.

[13] See DESA [14]. For the year 2070, the male life expectancy is estimated to further grow to 88 years.

[14] See, e.g. Hunt and Blake [31], Nigri et al. [44] Renshaw and Haberman [47] and Renshaw and Haberman [46].

While in a traditional annuity[15] longevity risk is transferred from the insured to the insurer (and covered by its risk management instruments), in a tontine the risk that a single participant might live longer than expected is fully borne and shared by the other tontine holders who in this case receive lower cash flows than expected. Therefore, no equity capital backing is needed to cover longevity risk, and the tontine can be offered without a risk-cost loading. However, the tontine has the disadvantage that because individual shares in the tontine as well as times of death of tontine members are random, both the amount and timing of tontine payments are uncertain.

Because the tontine members carry, pool and share the total risk among each other while the offering provider does not bear it, a tontine can be offered at a cheaper price than a comparable traditional life insurance product. In addition, it generates age-increasing benefits and is therefore able to meet increasing monetary requirements at old ages.

In this sense, the tontine is aimed at average individuals who want to run private old-age provision and it might compete with specific long-term care insurance schemes. However, such schemes depend on the degree of care and contain exceptions, which means that soft factors and uninsured aspects are not covered. In contrast, the tontine allows for open use of funds, such as the age-appropriate conversion of an apartment (e. g. ground level bathroom or stair lift). Other possible uses of a tontine comprise the financing of costly items to maintain the standard of living (e. g. increased taxi driving with reduced vision, the use of high-quality meals-on-wheels services or shopping delivery services) or the provision of high-quality nursing care services beyond the statutory level.

Moreover, due to the absence of guarantees, the tontine enables participation in stock market developments. However, the tontine generates volatile payments, which means that the insurance character of a tontine might not be ensured in every situation.

So far, tontines have been considered an alternative or historic predecessor of traditional pension insurance. However, tontines were considered to be inferior compared to traditional pension insurance, because the latter provides less volatile payments for an equal expected return.[16] In contrast to this point of view we will regard tontines as a complement to traditional pension products.

3 Literature review

The literature provides several contributions on the suitability of tontines for pension planning. Sabin [49] designs a fairly priced tontine with regard to age, gender and entry date that is equivalent to a common annuity scheme. His results exhibit a more cost-efficient payout pattern compared to a typical insurer-provided annuity

---

[15] In the following we use "annuity" synonymously for the traditional life insurance product.
[16] See Sabin [49].

not just on average, but for virtually every member who lived more than just a few years.

Forman and Sabin [18] construct a fair transfer plan (FTP) to guarantee a fair bet for all participating investors of a tontine by accounting for each age, life expectancy and investment level. They show that a fairly designed tontine is superior to defined benefit plans in terms of funding and sponsoring of the pension system. They illustrate that a fairly developed tontine model would improve the situation of pension providers while serving the retirement income demand of the tontine participants.

Milevsky and Salisbury [42] and Milevsky and Salisbury [43] derive expected-lifetime-utility maximizing tontine designs by accounting for sensitivity of both the tontine size and the longevity risk aversion for each tontine member. In the case without transaction costs, Milevsky and Salisbury [42] find that, due to higher volatility of the payments, the tontine provides a lower utility than a traditional life annuity.

Chen et al. [11] propose a retirement product called tonuity, in which a tontine and an annuity are combined to serve best the policyholder needs in a way that the policyholder owns a tontine in the early years of retirement, and the tontine is then converted into a deferred annuity. As tontine payouts are relatively stable in the early years of retirement at low costs because of the absence of guarantees, its payouts are highly volatile in the later years of retirement. Therefore, the tontine is converted into a stable deferred annuity at an optimal switching time. Chen et al. [12] compare the "tonuity" with a bundled product called "antine" that starts with annuity payments and switches to tontine payments in later years, and with a product bundle consisting of a traditional annuity and a tontine. In an expected-lifetime-utility framework they find that the product bundle is superior to the bundled products "tonuity" and "antine".

Milevsky and Salisbury [42] also encourage the idea of both tontine and annuities to co-exist. Their proposed optimal tontine structure allows anyone of any age to participate in the scheme, making the mix of cohorts possible.[17] Along with our results, Milevsky and Salisbury [42] support the conclusion that introducing properly designed tontines could help to maximize lifetime utility. The idea of combining both the benefits of tontines and conventional annuities is likewise encouraged by Chen et al. [11]. However, they propose the tonuity as a single product, which offers a tontine-like pay-off in the first years of retirement and then switches to a secure payment of an annuity. In their case, they do not focus on initial endowment wealth as a determinant but suggest that retirees with medium risk aversion will prefer the combination of tontines and annuities.

In an expected-lifetime-utility context, Bernhardt and Donnelly [6] derive the optimal investment in a tontine that allows to bequeath payments in the case of death.

Weinert [60] extends the prevailing tontine scheme by the possibility of a premature surrender and determines the fair surrender value.

---

[17] See Milevsky and Salisbury [43].

4 Model framework

We first model mortality dynamics in Germany for the upcoming decades and derive possible population pyramids in a second step. These, in turn, are the basis for the composition of the fair revolving tontine. We then estimate the risk-free benefit profile of a standard annuity and compare it to the risky benefit profile of a tontine. We assume that each individual i has initial wealth endowment W_i and that there is no further source of income in the future. W_i can be seen as the sum of both discounted future earnings until retirement and savings up to the investment date. W_i will be completely converted into pension installments. For the expected tontine benefits, we provide a closed-form solution while we determine the realized benefits by performing a Monte Carlo simulation. We then analyze to what extent tontine and traditional annuity are able to satisfy an empirically estimated, increasing old-age liquidity need function for different settings. In our analysis, we assume a risk-fee rate of zero. On the one hand, it describes the current low interest rate environment that exerts pressure on the profitability and solvency positions of insurers.[18] On the other hand, it allows us to mainly focus on the effects stemming from mortality risk. For this reason, we also refrain from modeling stochastic investment returns. A constant risk-free rate of zero is likewise assumed in the models by Milevsky and Salisbury [42]. Furthermore, we estimate an optimal portfolio consisting of annuity and tontine, maximizing expected utility according to a Cumulative Prospect Theory framework. We provide results for different demographic scenarios and mortality dynamics and show the capability of tontines as instruments for retirement planning from a policyholder perspective. Finally we incorporate subjective beliefs about individual mortality to account for different perceptions about individual life expectations, which leads to a changing optimal asset allocation for retirement planning.

4.1 Mortality model

In a first step, we project mortality rates for a forecast horizon of $t = 1, \ldots, T$ years. Our starting point is the one-factor-model for estimating mortality rates by Lee and Carter [36]. According to the Lee-Carter Model the force of mortality $\mu_{x,t}$ of a person aged x in year t is specified[19] as

$$\ln\left(\mu_{x,t}\right) = \alpha_x + \beta_x \cdot \kappa_t \Leftrightarrow \mu_{x,t} = e^{\alpha_x + \beta_x \cdot \kappa_t} \tag{1}$$

where α_x and β_x are time constant parameters for a male[20] aged x that determine the shape and the sensitivity of the mortality rate to changes in κ_t, which is a time-varying parameter that captures the changes in the mortality rates over time. As proposed by Lee and Carter [36], the parameter κ_t can be modeled and estimated using

---

[18] See European Insurance and Occupational Pensions Authority [16].

[19] We assume that the force of mortality is constant over each year of age and year, yielding that force of mortality and the central death rate coincide.

[20] We use y for a female aged y years analogously.

ARIMA processes. κ_t is assumed to follow a random walk with drift. Thus the time variable parameter is given by $\kappa_t = \delta + \kappa_{t-1} + \varepsilon_t$, where δ is the drift parameter and the error terms are independent, normally distributed $\varepsilon_t \sim N(0, \sigma_k)$. The one year death probability $q_{x,t}$ of a person aged x in year t is[21] $q_{x,t} = 1 - \exp(-\mu_{x,t})$.

4.2 Demographic Structure

Based on the predicted mortality rates $q_{y,t}$ for the one-year death probability of a woman aged y in period t, and $q_{x,t}$ for the one-year death probability of a man aged x in period t, we determine the demographic structure of an economy in every period t. $Q_{y,t}$ $(Q_{x,t})$ is the total quantity of female (male) people of a cohort aged y (x) at time t. Equation (2) shows the updating process. Newborns, or people in their first year of life $(y, x = 1)$ are determined by the sum of the age-specific fertility rate AGZ_y times the quantity of females of the respective age y in each period t, which is weighted by the fraction of newborn females f_0 and males $m_0 = 1 - f_0$. From the second year of being alive, the number of people is the probability to survive one year of someone who was one year younger in the year before, times the number of people who were one year younger in the year before:

$$Q_{y,t} = \begin{cases} f_0 \cdot \sum_{y=2}^{\Omega} Q_{y,t} \cdot AGZ_y \\ Q_{y-1,t-1} \cdot (1 - q_{y-1,t-1}) \end{cases} \quad Q_{x,t} = \begin{cases} m_0 \cdot \sum_{y=2}^{\Omega} Q_{y,t} \cdot AGZ_y & \text{for } y, x = 1 \\ Q_{x-1,t-1} \cdot (1 - q_{x-1,t-1}) & \text{for } y, x = 2 \dots \Omega \end{cases}$$

$$(2)$$

We determine these quantities for all cohorts $y, x = 1, \dots, \Omega$ in all periods $t = 1, \dots, T$ for males and females to estimate the corresponding population pyramids. $F_{y,t}$ $(F_{x,t})$ in Eq. (3) shows the Cumulative Distribution Function (CDF) of a person aged y (x) in each t, where $\kappa_{y,t} = \frac{Q_{y,t}}{\sum_{y=1}^{\Omega} Q_{y,t}}$ and $\kappa_{x,t} = \frac{Q_{x,t}}{\sum_{x=1}^{\Omega} Q_{x,t}}$ (for $y, x = 1, \dots, \Omega$) are the fractions of each cohort of females and males of the total female and male population in each period t.

$$F_{y,t} = \begin{cases} 0 & : y < 0 \\ \sum_1^y \kappa_{y,t} & : 0 \leq y < \Omega \\ 1 & : y > \Omega \end{cases} \quad F_{x,t} = \begin{cases} 0 & : x < 0 \\ \sum_1^x \kappa_{x,t} & : 0 \leq x < \Omega \\ 1 & : x > \Omega \end{cases} \quad (3)$$

The fractions of females and males aged y and x of the total population in each t are $f_{y,t} = \frac{\kappa_{y,t}}{\kappa_{y,t} + \kappa_{x,t}}$ and $m_{x,t} = 1 - f_{y,t}$. The predicted mortality rates and demographic structures are the basis for calculating the tontine composition and tontine benefits, and are the starting point to analyze the impact of the demographic development on the tontine and its possible application as an alternative retirement planning product.

---

[21] See for example Milevsky [41].

4.3 Tontine model

We model a fair revolving tontine based on Sabin [49] that allows tontine participants to be of any gender and age, and to invest a desired one-time initial amount of money B_i at tontine entrance. B_i then is tied in the tontine and cannot be withdrawn before the tontine member's death. Furthermore, there is no possibility to inject additional capital for any individual in future periods. While the original model considers infinitesimal points in time, resulting in only one member being able to die at one point in time, we adjust the model to a yearly time frame allowing for multiple deaths. We further assume the number of the tontine members N as fixed: every time a participant dies, the tontine is refilled to N. A new entrant i is randomly drawn from the period-corresponding demographic structure. We assume entrants at least to be of a certain age $\underline{y}, \underline{x}$ and also assume an upper limit of entering the tontine of age $\overline{y}, \overline{x}$ so $0 \le \underline{y}, \underline{x} < \overline{y}, \overline{x} \le \Omega$. The random age of an individual i entering the tontine in t is expressed by

$$y_{E,i,t}, \quad x_{E,i,t} = F_t^{-1}(z) \tag{4}$$

where F_t^{-1} is the inverse function of F_t with $Pr(y) = f_{y,t}$ and $Pr(x) = m_{x,t}$ and with $z \in \left(F_t^{-1}\left(\underline{y}\right), F_t^{-1}\left(\overline{y}\right) \right)$ or $z \in \left(F_t^{-1}\left(\underline{x}\right), F_t^{-1}\left(\overline{x}\right) \right)$ where z is uniformly distributed on $\mathcal{U}\left(F_t^{-1}\left(\underline{y}\right), F_t^{-1}\left(\overline{y}\right) \right)$ or $\mathcal{U}\left(F_t^{-1}\left(\underline{x}\right), F_t^{-1}\left(\overline{x}\right) \right)$. We further assume the establishment of the tontine in $t = 0$ and refrain from investing the tied capital to streamline the model and to be able to quantify solely the interrelation of mortality benefits and demographic change. For better readability, we denote the one-year death probability of individual i with the beforehand assigned characteristics as $q_{i,t}$ in t. The index i allows to identify each individual with its specific characteristics in each period. Furthermore we denote the age of a person as x in the following, irrespective of the gender.

Let $\left\{ A_{i,t} \right\}$ be the event that i dies in t with $P\left(A_{i,t}\right) = q_{i,t}$ and $\left\{ A_{i,t}^{\mathsf{c}} \right\}$ be the event that i survives in t with $P\left(A_{i,t}^{\mathsf{c}}\right) = 1 - q_{i,t}$. Let $\left\{ A_{0,t}^{\mathsf{c}} \right\}$ be the event that at least someone dies in t and $\left\{ A_{0,t} \right\}$ be the event that no one dies in t. Using the inclusion-exclusion principle,[22] the probability that at least someone dies in t is

$$P\left(A_{0,t}^{\mathsf{c}}\right) = P\left(\bigcup_{i=1}^{N} A_{i,t} \right) = \sum_{j=1}^{N} \left((-1)^{j+1} \sum_{\substack{I \subseteq \{1, \dots N\} \\ |I| = j}} P\left(\bigcap_{i \in I} A_{i,t} \right) \right)$$

and the probability that no one dies in t is

---

[22] See for example Graham et al. [21].

$$P(A_{0,t}) = 1 - P\left(\bigcup_{i=1}^{N} A_{i,t}\right).$$

$\left\{A_{k,t} \mid A_{0,t}^{\mathrm{c}}\right\}$ denotes the event that k dies in t conditioned that at least someone dies in t. Using the law of total probability yields for the probability that k dies in t conditioned that at least someone dies in t, $\rho_{k,t}$

$$\rho_{k,t} = P\left(A_{k,t} \mid A_{0,t}^{\mathrm{c}}\right) = \frac{P\left(A_{k,t}\right)}{P\left(A_{0,t}^{\mathrm{c}}\right)} = \frac{q_{k,t}}{\sum_{j=1}^{N}\left((-1)^{j+1} \sum_{\substack{I \subseteq \{1, \ldots N\} \\ |I| = j}} P\left(\bigcap_{i \in I} A_{i,t}\right)\right)}.$$

(5)

If member i dies, his or her balance account B_i is distributed to the survivors. To be fair, this reallocation takes place according to the specific characteristics of the surviving members: Older members and those with a larger stake in the tontine have to receive more. If member $k \neq i$ dies, member i receives a fraction of k's balance $a_{i,k,t}B_k$, where

$$0 \leq a_{i,k,t} \leq 1 \text{ for } i, k = 1, \ldots, N \text{ and } i \neq k.$$

(6)

k's balance is forfeited entirely, so

$$a_{k,k,t} = -1 \text{ for } k = 1, \ldots, N.$$

(7)

Equation (7) states that the dying members' stake in the tontine is distributed among the surviving members. In sum, the amount lost by k equals the sum of the distributed benefits to the surviving members, so

$$\sum_{i=1}^{N} a_{i,k,t} = 0 \text{ for } k = 1, \ldots, N.$$

(8)

The unconditional expected benefit received by member i in t is the return in case no one dies and the return if at least someone dies, weighted with their corresponding probabilities, thus

$$E[r_{i,t}] = E[r_{i,t} \mid A_{0,t}]P(A_{0,t}) + E\left[r_{i,t} \mid A_{0,t}^{\mathrm{c}}\right]P\left(A_{0,t}^{\mathrm{c}}\right)$$

(9)

Since return is generated solely by mortality, there cannot be any return if no one dies. Thus $E[r_{i,t} \mid A_{0,t}] = 0$ and the expected return reduces to the second term of the right-hand side of Eq. (9). The expected return conditioned that at least someone dies is the sum of the conditional death probability weighted fractions of the balance accounts over all k members in t, thus

$$E\left[r_{i,t} \mid A_{0,t}^{c}\right] = \sum_{k=1}^{N} \rho_{k,t} a_{i,k,t} B_k. \tag{10}$$

To achieve a fair tontine, each member's expected benefit is zero in each year. This is because the expected loss of the own balance account in the case of the own death has to be offset by the expected gains one receives from other members' deaths, so $E\left[r_{i,t} \mid A_{0,t}^{c}\right] \overset{!}{=} 0$. The older i is, the higher the death probability $q_{i,t}$, causing that $\rho_{i,t}$ increases as well (assuming that the composition of the tontine does not change) and leads to a higher expected loss in case of the i-th death. This has to be compensated by an increase in the fractions $a_{i,k,t}$ one receives in the case of other members' death to counterbalance the aforementioned effect and to create a fair bet.

   To satisfy the conditions of a fair bet for every tontine member $i = 1, \ldots, N$, one has to search for a set of $a_{i,k,t}$ that yield an expected benefit of zero for every tontine member and which fulfills conditions (5), (6), (7) and (8), yielding $E[r_{i,t}] = E[r_{i,t}|A_{0,t}^{c}] = 0$.

   As Sabin [49] shows, such a set of $a_{i,k,t}$ exists only if no member is exposed to more than half of the total risk of the tontine. This can be achieved by introducing a ceiling of the amounts to invest B_i. Choosing N large enough additionally reduces the threat of a single individual holding too large a fraction of risky exposure of the tontine. Here, we implement an algorithm[23] for the determination of the set of $a_{i,k,t}$ that is proposed by Sabin [49], which approximately assigns constant $a_{i,k,t}$, irrespective of k for $k \neq i$ and which provides best results for large N. In the following, we assume that the resulting $a_{i,k,t}$ satisfy conditions (5)–(8) and that no member holds more than half of the risky exposure of the tontine, formally meaning that

$$\rho_{i,t} B_i \leq \frac{1}{2} \sum_{k=1}^{N} \rho_{k,t} B_k \text{ for } i = 1 \ldots N. \tag{11}$$

The expected return, conditioned that i survives in t, is

$$E\left[r_{i,t}|A_{i,t}^{c}\right] = \sum_{\substack{k=1 \\ k \neq i}}^{N} \rho_{k,t} a_{i,k,t} B_k. \tag{12}$$

Because Eq. (10) is solved to be zero, to yield a fair bet, $\sum_{\substack{k=1 \\ k \neq i}}^{N} \rho_{k,t} a_{i,k,t} B_k = -\rho_{i,t} a_{i,i} B_i$ and Eq. (12) is

$$E\left[r_{i,t}|A_{i,t}^{c}\right] = v_{i,t} = q_{i,t} B_i. \tag{13}$$

---

[23] For further algorithms to construct a tontine, see Sabin [50].

This is an interesting property since the individual expected return in case of the own survival is solely driven by the own mortality $q_{i,t}$ and the own investment in the tontine B_i, and does not depend on the tontine composition.[24]

The unconditional realized benefit for i in t is

$$r_{i,t} = \sum_{\substack{k=1 \\ k \neq i}}^{N} a_{i,k,t} B_k \mathbf{1}_{\left\{ A_{k,t} \cap A_{i,t}^c \right\}} - B_i \mathbf{1}_{\left\{ A_{i,t} \right\}}$$

where the indicator function $\mathbf{1}_{\{\ldots\}}$ takes on the value of 1 if the respective event occurs and 0 otherwise and therefore $\mathbf{1}_{\{\ldots\}} \sim Ber_{P(\ldots)}$. The realized return conditioned that i survives is

$$r_{i,t} | A_{i,t}^{\mathbf{c}} = \sum_{\substack{k=1 \\ k \neq i}}^{N} a_{i,k,t} B_k \mathbf{1}_{\left\{ A_{k,t} \right\}}. \tag{14}$$

Since the $a_{i,k,t}$ are approximately constant for i for large N, $\overline{a}_{i,k,t} \approx \dfrac{q_{i,t} B_i}{\sum_{\substack{k=1 \\ k \neq i}}^{N} q_{k,t} B_k}$, and

Eq. (14) is

$$r_{i,t} | A_{i,t}^{\mathbf{c}} \approx q_{i,t} B_i \frac{\displaystyle\sum_{\substack{k=1 \\ k \neq i}}^{N} B_k \mathbf{1}_{\left\{ A_{k,t} \right\}}}{\displaystyle\sum_{\substack{k=1 \\ k \neq i}}^{N} B_k q_{k,t}}. \tag{15}$$

Although the volatility of the tontine converges toward zero for $N \to \infty$, for a finite and realistic tontine size, the payouts are volatile.[25] As Appendix *Tontine Volatility* shows, the tontine payouts are approximately normally distributed, with

$$v_{i,t} = q_{i,t} B_i$$

and

---

[24] As long as no individual holds more than the mortality weighted tontine exposure (Eq. 11), a tontine exists (i.e. no negative death probabilities exist).

[25] For example, the tontines offered by Le Conservateur have to comprise at least 200 members to be launched. See http://www.conservateur.fr/.

$$\sigma_{i,t} = \frac{q_{i,t}Bi}{\sqrt{M-1}} \sqrt{\sum_{m=1}^{M} \left(\frac{\sum_{\substack{k=1 \\ k \neq i}}^{N} B_k \mathbf{1}_{\{A_{k,t}\}}^m}{\sum_{\substack{k=1 \\ k \neq i}}^{N} B_k q_{k,t}} - 1 \right)^2}. \tag{16}$$

with M Monte Carlo simulation paths.

4.4 Annuity model

The considered individual has pension wealth W_i, no other wealth, no other source of income and the share of wealth which is not tontinized gets annuitized. The individual can choose to hold any positive proportion of his wealth in an annuity or in a tontine. The individual has no heirs and no desire to leave a bequest. We refer to an annuity applied by Milevsky [41] which requires a lump sum investment of $W_i - B_i$. The annuity then pays a stable income stream on a yearly basis until the participants' death, starting as an immediate annuity.

The conditional survival probability of a person aged x in t of surviving τ more years is defined as

$$_{\tau}p_x^t = \prod_{j=0}^{\tau-1} \left(1 - q_{x+j}^{t+j} \right).$$

The annuity provider is risk-neutral. Therefore, because we refrain from interest rates in this model, the lump sum price $\bar{a}_x^t$ in t of an immediate annuity which provides an income of 1 EUR per year until death is

$$\bar{a}_x^t = \sum_{\tau=1}^{\Omega} {}_{\tau}p_x^{t+\tau}.$$

To simplify, we denote the price of the immediate annuity as $\bar{a}_i^t$, where the individual characteristics can be identified via i. A lump sum investment of $W_i - B_i$ in the annuity provides a stable, lifelong and yearly income stream of

$$R_{i,t} = \frac{W_i - B_i}{\bar{a}_i^t}. \tag{17}$$

The individual can decide on the allocation of tontinization (B_i) and annuitization ($W_i - B_i$) of the wealth W_i.

4.5 Old age liquidity need function

According to Worldbank [61], worldwide life expectancy at birth has increased between 1960 and 2015 from 52.5 to above 71.7 years. The increasing lifetime will

cause the number of people over 80 years to almost double to 9 millions in Germany by the year 2060 according to forecasts by the Statistisches Bundesamt [54]. In the future, it is therefore very probable that very high ages of 100 years and even more will be reached by a large number of people. According to the medicalisation thesis motivated by Gruenberg [22], the additional years that people live due to demographic change are increasingly spent in bad health condition and disability. Jagger et al. [32] find a significant increase in life expectancy between 1991 and 2011 in England. However, cognitive impairment, poor health and disability increase with age and elderly disability increased in that period. In those additional years of life, the demand for care products and medical service increases over-proportionately. Coming from 2.6 million nursing cases in Germany in 2013, Kochskämper [34] estimates between 1.5 and 1.9 million additional nursing cases in Germany in the year 2060 due to demographic change. By the year 2030, the demand for stationary permanent care will increase by 220,000 places in Germany.

While previous research finds a systematic decrease in the consumption level at retirement,[26] incorporating the nursing care costs and medical expenses into consumption yields a so called retirement smile.[27] When people retire they are mostly still healthy and have therefore time to spend on lifestyle. As they age, first physical constraints appear, they become more and more home-bound, and thus consumption declines while supplementary and medical costs are still at a low level. As people become very old they rely more on assisted living requirements, and costly long-term nursing care is needed. Therefore the typical monetary demand for a retiree is U-shaped. Based on own empirical research[28] we model an old-age liquidity need function, which accounts for demand for nursing care and medical service. The determination of the liquidity need function is based on data available on consumer spending in the SOEP from 1984 to 2013. We determine the age-specific expenditure pattern for the spending categories of food, living, health, care, leisure, refurbishment and miscellaneous and finally aggregate them. The considered age ranges from 60 to 95. From the age of 95 on, we extrapolate until the age of 105 due to the limited data basis for those ages. The modeling of the nursing care costs is based on the costs of an inpatient, permanent care, which occurs in nursing homes, even though a large share of care is performed by family members and nursing care services. The inpatient care costs reflect the actual potential resources needed in a more appropriate manner because in home care, the time spent by family members is not taken into account. Moreover, many refuse inpatient care only because of lacking resources. Figure 1a shows our estimates of the average nursing care costs from the age of 60 to 90 per year. In addition, the average liquidity need without nursing care

---

[26] See for example Hamermesh [24], Mariger [39], Banks et al. [2], Bernheim et al. [7] or Haider and Stephens Jr [23].

[27] In an empirical work, Blanchett [9] derives the trend of retirement consumption. He observes a shift in the expenditures towards increasing health care, entertainment and food.

[28] The data used in this publication was made available to us by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.

costs is shown. If we aggregate both components, we obtain the characterized retirement smile, which is presented in Figure 1b.

In our model, we map the old-age liquidity need via a polynomial liquidity need function D_t of order 2 which is calibrated based on our results, and assume an extrapolation up to the age of $x = 105$. The desired consumption level is driven by age x in t so

$$D_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \epsilon_t \tag{18}$$

where the parameters β_0, β_1 and β_2 are fitted using our empirical data, and ϵ_t is the error term in t. For simplicity, tontine payments and liquidity need are assumed to be independent.[29]

4.6 Multi cumulative prospect theory valuation

So far, we have on the one hand an income stream which is composed of both certain annuity and volatile tontine payments, and on the other an age-increasing liquidity need which the income stream should cover. We aim to design the payout pattern of tontine and annuity such that the liquidity need can be served optimally. Therefore, we evaluate the income stream relative to the liquidity need as a reference point. An income stream larger than the liquidity need is considered a gain and utility-generating, while an income stream lower than the liquidity need generates a loss, which provides disutility.[30] In this sense, we look at the utility of the income relative to the liquidity need, rather than the absolute level of income. Since the liquidity need increases with age, a payout which is able to meet the demand in early years might not be sufficient in the later years of retirement. Therefore, to evaluate an income stream relative to a reference point, the Cumulative Prospect Theory (CPT), originated by Kahneman and Tversky [33] and enhanced by Tversky and Kahneman [55] is highly suitable for our purpose.[31] Although CPT is a descriptive rather than a normative theory, it allows us to capture the aforementioned properties and to determine an optimal, CPT-utility maximizing fraction to be invested in the tontine. To capture the life-cycle dynamics of the repeating payments until death, we use the Multi Cumulative Prospect Theory (MCPT), as applied by Ruß and Schelling [48], where the CPT-utility is determined in every period t by considering a changing reference point which is represented by the respective liquidity need D_t, and is finally aggregated with respect to survival prospects. The total utility of person i over his or her stochastic remaining lifespan is the sum of the CPT-utilities of the gains and losses $Z_{i,t}$ of the payouts generated by the portfolio of tontine and annuity in relation

---

[29] In future work, it would be interesting to explore how the consumption pattern and portfolio decisions respond to late-life care expenses. See Yogo [62] and Koijen et al. [35].

[30] In this article we assume that the cash flows generated in the respective period are used to maximize the utility of the respective period. Thus, there is no intertemporal consumption optimization where generated funds are saved for future periods in which they can better meet liquidity needs. See Gemmo et al. [19], who study a tontine in an intertemporal life cycle model.

[31] In this sense, Schmidt [51] uses CPT to determine insurance demand.
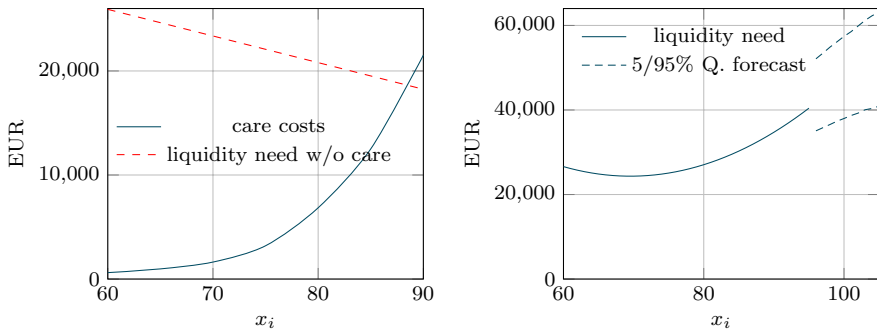
(a) Average nursing care costs and liquidity need without nursing care per year

(b) Aggregated liquidity need per year

Fig. 1 The retirement smile

to the liquidity need at each point in time t, deflated by a subjective discount factor $\delta \leq 1$. The conditional survival probability $_\tau p_x$ of an x-year-old of surviving τ more years is incorporated in the CPT-utility. It is combined with the density of the respective retirement payouts and the resulting joint probability is valued according to the CPT, thus

$$MCPT(i, t) = \sum_{\tau=1}^{T-x} \delta^\tau CPT\left(Z_{i,t+\tau}\right). \tag{19}$$

The CPT-utility in each period is

$$CPT\left(Z_{i,t+\tau}\right) = \int_{-\infty}^{0^-} v(z)d\left(w^-\left(F_{i,t+\tau}(z)\right)\right) + \int_{0^+}^{\infty} v(z)d\left(-w^+\left(1 - F_{i,t+\tau}(z)\right)\right) \tag{20}$$

in a continuous context.[32] The probability weighting function $w^+(F)$ for gains and $w^-(F)$ for losses is

$$w^+(F) = \frac{F^\gamma}{\left(F^\gamma + (1-F)^\gamma\right)^{1/\gamma}}, \quad w^-(F) = \frac{F^\kappa}{\left(F^\kappa + (1-F)^\kappa\right)^{1/\kappa}} \tag{21}$$

where γ and κ are the probability weighting parameters. The value function $v(z)$ is given by

$$v(z) = \begin{cases} z^a & z \geq 0 \\ -\lambda |z|^b & z < 0 \end{cases} \tag{22}$$

---

[32] See for example Hens and Rieger [26], Ågren[1] or Ruß and Schelling [48] who use the CPT in a continuous context.
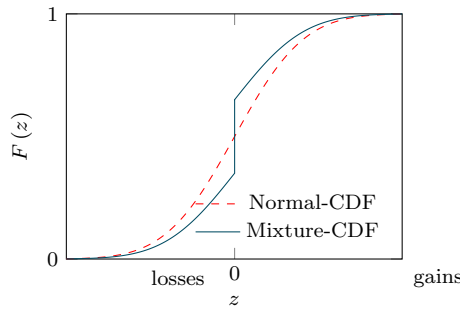
Fig. 2 Mixture CDF of survival probability and retirement payout distribution vs. Normal CDF of retirement payout

where $a, b \in (0, 1)$ and $\lambda > 1$. The mixture cumulative distribution function, to account for the joint probability of conditional survival and the payout size, is

$$F_{i,t+\tau}(z) = \left(1 -_\tau p_x\right)\mathbf{1}_{[0,\infty)} +_\tau p_x \int_{-\infty}^{z} dF_{\mathcal{N}_{i,t+\tau}}(u) \tag{23}$$

where $F_{\mathcal{N}_{i,t+\tau}}$ is the CDF of a normal distribution with the first moment

$$v_{i,t+\tau} = q_{i,t+\tau}B_i + \frac{W_i - B_i}{a_{i,0}} - D_{i,t+\tau} \tag{24}$$

and the standard deviation $\sigma_{i,t+\tau}$ resulting from Eq. (16) for the normally distributed gains and losses.

Figure 2 shows the illustration of Eq. (23): dying leads to $z = 0$ to which the death probability $1 -_\tau p_x$ is assigned. This explains the jump in the CDF.

By incorporating the mixture CDF in the analysis, Eq. (20) becomes

$$\begin{aligned} CPT\left(Z_{i,t+\tau}\right) =_\tau p_x \Bigg[& \int_{-\infty}^{0^-} v(z)w^{-\prime}\left(F_{i,t+\tau}(z)\right)f_{\mathcal{N}_{i,t+\tau}}(z)dz \\ & + \int_{0^+}^{\infty} v(z)w^{+\prime}\left(1 - F_{i,t+\tau}(z)\right)f_{\mathcal{N}_{i,t+\tau}}(z)dz \Bigg] \\ & + \left(1 -_\tau p_x\right)\Big[v(0^-)w^{-\prime}\left(F_{i,t+\tau}(0^-)\right) \\ & + v\left(0^+\right)w^{+\prime}\left(1 - F_{i,t+\tau}\left(0^+\right)\right)\Big] \end{aligned} \tag{25}$$

with

$$w'(F) = \frac{F^\theta}{\left(F^\theta + (1-F)^\theta\right)^{1/\theta}} \cdot \left[\frac{(\theta - 1)F^\theta + (F + \theta(1-F))(1-F)^{\theta-1}}{F^{\theta+1} + F(1-F)^\theta}\right] \tag{26}$$

where $\theta \in (\gamma, \kappa)$. The first two lines of Eq. (25) are the utility in case of survival, whereas the third and fourth line are the utility in case of death. Since the utility in case of death is zero because of $v(0) = 0$, Eq. (25) reduces to just the first two lines.

Finally, we numerically maximize Eq. (19) subject to the optimal level of tontine investment B_i.

$$\max_{B_i} MCPT(i)$$
$$\text{s. t. } B_i \leq W_i, B_i \geq 0 \tag{27}$$

4.7 Variation: stochastic liquidity need in MCPT valuation

Because Eqs. (24) and (16) are assumed to be normally distributed, it follows that also the combined payout of tontine and annuity is normally distributed with

$$Z_{i,t} \sim \mathcal{N}\left(v_{i,t}, \sigma_{i,t}^2\right).$$

To cover effects stemming from uncertainty about the future liquidity need, we assume that the liquidity need itself is normally distributed with mean $E(D_{i,t})$ and standard deviation $\sigma_{D_{i,t}}$, therefore

$$v_{i,t} \sim \mathcal{N}\left(v_{i,t0}, \sigma_{v_{i,t}}^2\right)$$

where $v_{i,t0} = q_{i,t}B_i + \frac{W_i - B_i}{a_{i,0}} - E(D_{i,t})$ and $\sigma_{v_{i,t}}^2 = \sigma_{D_{i,t}}^2$ is calibrated based on own empirical research. Tontine payments and liquidity need are assumed to be independent. If we write

$$Z_{i,t}' = \left(Z_{i,t}' - v_{i,t}\right) + v_{i,t}$$

then

$$\left(Z_{i,t}' - v_{i,t}\right) \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right)$$

and

$$Z_{i,t}' \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right) + \mathcal{N}\left(v_{i,t0}, \sigma_{v_{i,t}}^2\right) = \mathcal{N}\left(v_{i,t0}, \sigma_{i,t}^2 + \sigma_{v_{i,t}}^2\right) \tag{28}$$

because the sum of two normally distributed random variables is also normally distributed. While the mean expected payout remains the same, the volatility increases to the sum of the variance of the tontine payment and the variance of the liquidity need.

4.8 Variation: subjective mortality

To account for subjective beliefs about one's own mortality risk, we adjust the objective forecasted mortality. It is important to understand the difference compared to the probability adjustment which CPT undertakes: while CPT accounts for a deviating perception of objective probabilities, the subjective mortality adjustment

modifies the average probabilities subject to own perceptions about the individual health status. People who believe to live longer than the aggregate average because they feel very healthy or have an active lifestyle perceive to have lower death probabilities and thus believe to have a longer expected remaining lifetime. Therefore, the optimistic subjective death probability $q'_{x,t}$ is lower than the average, objective death probability $q_{x,t}$. Likewise, the pessimistic subjective death probability $q'_{x,t}$ for people who believe to live shorter than the overall average (because of severe illness or the awareness of a poor lifestyle) is higher than the actual death probability $q_{x,t}$. Bissonnette et al. [8] show that, within different groups (e.g. gender, ethnic background or education), people with similar characteristics are only slightly optimistic regarding their survival prospects compared to the average mortality within the subgroup, whereas the actual subgroup mortality itself differs tremendously from the overall population mortality. The authors conclude that the individual perceptions are very precise. Therefore, it is important to incorporate subjective survival probabilities in our analysis, because people who believe to live longer tend to live longer, and thus different retirement planning solutions are needed for different individuals. To account for subjective mortality in our model, we adjust the actual mortality rates $q_{x,t}$ by an individual mortality multiplier d, therefore the subjective mortality rate $q'_{x,t}$ is

$$q'_{x,t} = \begin{cases} d \cdot q_{x,t} & \text{if } d \cdot q_{x,t} \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{29}$$

where d is the realization of a random variable D and determines the subjective survival probability. For $0 < d < 1$ the individual expects to live longer than the average, if $d = 1$ the individual self assesses his or her lifetime of being average and if $d > 1$, the individual expects to live shorter than the actual mortality table predicts. Furthermore, $q_\Omega = 1$ which means that there is a limiting age Ω when the individual dies with certainty. A simple modeling approach for $d \sim D$ is shown in Appendix *Modeling Subjective Mortality*, where D is modeled using a Gamma Distribution. Since the insurance company offering tontines and annuities uses average objective mortality rates, pricing is undertaken on the basis of average mortalities. The subjective beliefs only influence the subjective determination of individual utility.

5 Calibration and results

We calibrate[33] the Lee-Carter model based on data from the *Human Mortality Database*,[34] and forecast mortality rates for $T = 100$ years, beginning from 2011, which is denoted by $t = 1$ in the analysis. The fractions of female and male newborns to

---

[33] See appendix Calibration of the Lee-Carter model for the fitted values for α_x, α_y, β_x, β_y, κ_t^x, and κ_t^y for 1–105 years of age and years 1956–2011. To forecast κ_t, an ARIMA process is used.

[34] Data from 2011, Source: http://www.mortality.org.

update population pyramids are calculated based on German birth statistics.[35] The age-specific birth rates are based on German birth statistics[36] and describe the number of newborns per year of a woman in each cohort. The maximum attainable age is set to be $\Omega = 105$ which means that at the age of $x = 105$ one dies with certainty. We consider initial wealth W_i as an independent variable and measure its influence on the optimal investment behavior under various scenarios. The parameters of the polynomial liquidity need function D_t are $\beta_0 = 163,984.686$, $\beta_1 = -4,000.634$ and $\beta_2 = 28.589$ and fit our empirically estimated old-age liquidity need function based on SOEP data.

5.1 Base case

In Table 1, we report the parameters used in the MCPT[37] analysis, which constitute the base case. We consider i as a male individual aged 62 in the year the tontine is set up ($t = 1$), and vary his initial endowment W_i. Furthermore, we assume that the remaining tontine members $k = 1 \ldots N, k \neq i$ behave optimally, i.e. the individual amounts B_k invested in the tontine are the $MCPT_k$-utility maximizing amounts and lie between 0 and approximately EUR 50,000. We therefore assume them to be uniformly distributed on $[0, 50,000]$.[38] Based on $M = 10,000$ simulations, we calculate the realized tontine returns for individual i in every period in which he is alive and thereby determine the moments of the normal approximation of tontine returns for member i. We set the subjective discount factor $\delta = 1$, because we assume that the future states are as important as present states for an individual who aims to secure the future standard of living.[39] We calibrate the CPT-value function parameters a, b and λ according to the values proposed by Tversky and Kahneman [55], but use the actual (i.e. not weighted) probabilities, i.e. $\gamma = \kappa = 1$.[40]

---

[35] Data from 2000 to 2010, see Statistisches Bundesamt [53].

[36] Data from 2011, German Federal Statistical Office, https://www-genesis.destatis.de/genesis/online.

[37] We use the notation MCPT according to Eq. (19) for the sum of the periodic CPT-utilities and CPT according to Eq. (25) for the periodic utilities.

[38] For every amount invested in the tontine B_k, there exists a corresponding amount of initial wealth W_k for which B_k is optimal. Since we chose the optimal amounts of B_k randomly, we implicitly specify their initial wealth levels W_k. The chosen values for B_k correspond to the range of optimal amounts invested in the tontine for average individuals for the considered range of initial wealth W_i. Thereby we provide the optimal investment decision for hypothetical levels of W_k, which are roughly in the same range as the bandwidth of W_i. This implies that individuals have a sufficiently high wealth W_i that allows them to behave optimally. This assumption could be violated in reality if the required range of wealth does not fully reflect the population. An iterative adjustment of the population pyramid can help here. Our choice of the total population pyramid is therefore simplifying and offers a starting point for further research.

[39] In this sense, Parsonage and Neuburger [45] and Van der Pol and Cairns [57] provide empirical evidence that it is feasible to assume a subjective discount rate of zero for the discounting of future health benefits.

[40] This is in line with the asset pricing literature. See for example Barberis et al. [4] and Levy and Levy [37], who also use the actual probabilities. Cumulative Prospect Theory deviates from expected utility theory mainly in two respects: by using reference points and distorted occurrence probabilities. As we want to concentrate on effects stemming from deviations from reference points, we refrain from distorting the occurrence probabilities in our base case. In Sect. 5.7 we calibrate γ and κ according to Tversky and Kahneman [55].

Table 1 MCPT parameters base case

Parameter	Notation	Value
Forecast horizon (in years)	T	100
Maximum attainable age	Ω	105
Fraction of female newborns	ω^f	48.68%
Fraction of male newborns	ω^m	51.32%
Lower boundary age at tontine entrance	$\underline{x}$	62
Upper boundary age at tontine entrance	$\bar{x}$	100
Size of the tontine	N	10,000
Monte Carlo Paths	M	10,000
Subjective discount factor	δ	1
CPT value function parameters	a, b	0.88
CPT loss sensitivity factor	λ	2.25
CPT w^+ parameter	γ	1
CPT w^- parameter	κ	1

Since the expected tontine return as well as the tontine volatility are driven by the individual survival probability, both increase as i becomes older. Aged 62 in $t = 1$, a person investing $B_i = 30,000$ EUR in the tontine can expect to receive a first-year tontine return of 345.15 EUR which amounts to 1.15% of the initial tontine investment. In $t = 10$, his expected return is roughly 1.6 times higher than in the first year but still amounts to only 1.86% of the initial investment. The payments thus increase slowly in the early retirement years because of the slow increase of death probabilities in the early years. After 20 years, at the age of 81, the expected return is already 4.8 times as large as in the first year and the single payment in this year amounts to 5.51% of the initial investment. For very high ages, the payments increase tremendously: at the age of 91, in $t = 30$, the expected tontine return is almost 16.3 times as large as in the first year and yields a rate of return of 18.73%. Every year of further survival then yields even steeper increasing returns, being 47.69% of the initial investment at the age of 101 in $t = 40$, and finally 100% at the maximum attainable age of 105 in $t = 44$. Neglecting interest rate effects, one can expect to recoup the initial investment in year 26 at the age of 87. Since the standard deviation of the tontine returns depends on the individual mortality, volatility increases similarly with age. In comparison, an immediate fairly priced annuity with a lump sum investment of 30,000 EUR would yield an income of yearly 1458.96 EUR. The investor could recoup the initial investment already after 21 years at the age of 82. This is because the annuity provides stable payments whereas tontine payments increase with age. Figure 3 shows the yearly expected payout patterns of a tontine and an annuity with investment volume normalized to unity.
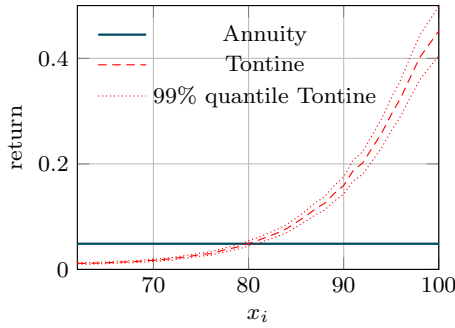
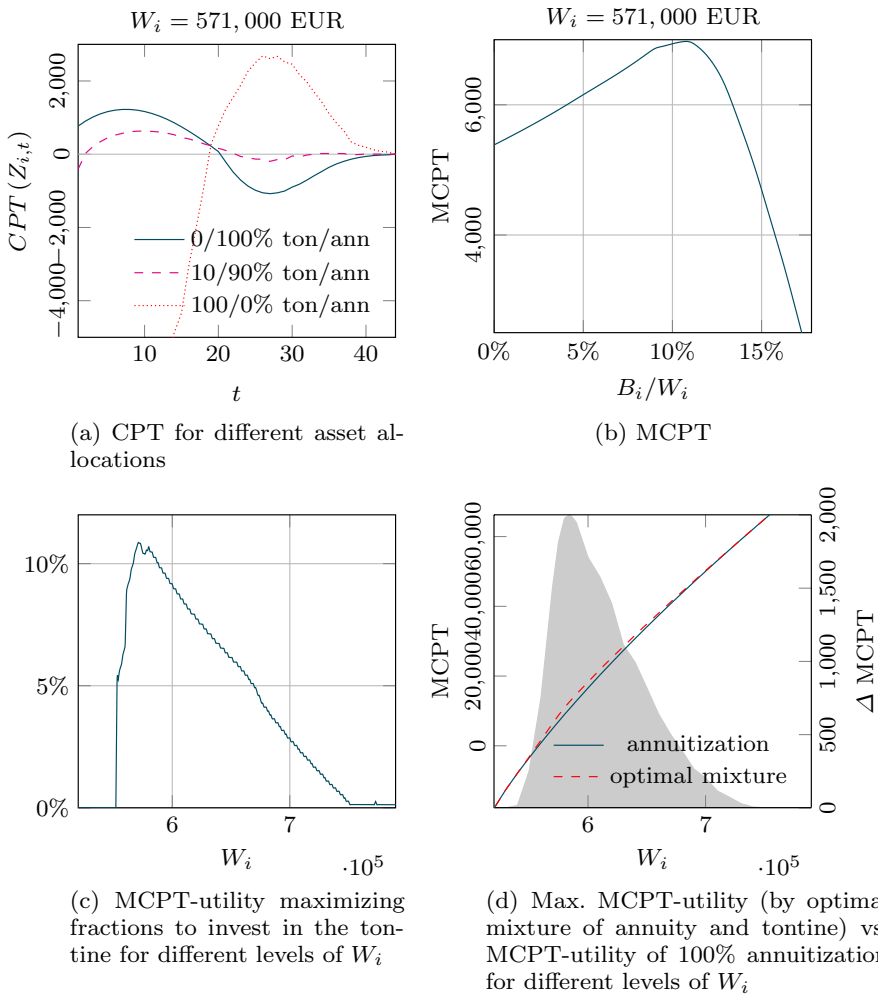**Fig. 3** Normalized payout of annuity vs. tontine for $N = 10,000$

(a) CPT for different asset allocations

(b) MCPT

(c) MCPT-utility maximizing fractions to invest in the tontine for different levels of W_i

(d) Max. MCPT-utility (by optimal mixture of annuity and tontine) vs. MCPT-utility of 100% annuitization for different levels of W_i

Fig. 4 Base case

Based on these considerations, we determine the CPT-utility $CPT(Z_{i,t})$ in each period for different levels of W_i for member i. Figure 4a shows the CPT-utility of member i for an initial wealth endowment of $W_i = 571,000$ EUR for different portfolio compositions at each point in time t. This represents the expected contribution of the CPT-utility on the aggregated MCPT-utility. Since survival probabilities decline with age, the impact of each CPT-utility declines with age and finally converges toward zero. We first consider the case in which person i completely annuitizes his initial wealth (solid line). As annuity payments are constant, an increasing liquidity need $D_{i,t}$ causes declining CPT-utilities in time. In early years, the liquidity need can be met. As the liquidity need rises and exceeds the available funds, CPT-utility decreases and becomes negative. As age increases, the declining survival probability causes a lower CPT-utility which reduces the impact of late periods on MCPT-utility. For the very late years, the low survival probabilities outweigh the negative CPT-utility, yielding that, finally, the impact of very late years on MCPT-utility approaches zero. Second, we consider the complete tontinization of initial wealth (dotted line). Because tontine payments are driven by mortality, payments are very low in the early years and increase in age, thus the liquidity need cannot be met for early ages and can easily satisfy $D_{i,t}$ in later years. Again, very low survival probabilities in later years reduce the impact on MCPT-utility. Furthermore, since tontine and annuity payments proceed adversely, a portfolio of both can help to generate payout patterns which enable to finance the increasing liquidity need appropriately. The dashed line shows the CPT-utilities of a payout pattern of a portfolio consisting of 10% tontine and 90% annuity. While still being able to satisfy the liquidity need in the early years, it is also able to provide almost sufficient funds in the later years. The sum of the CPT-utilities yields the MCPT-utility.

Figure 4b exemplarily shows the MCPT-utility for different fractions of $W_i = 571,000$ EUR being invested in the tontine (B_i). The remaining fraction $W_i - B_i$ is annuitized. Starting from a situation of complete annuitization, i can increase his MCPT-utility by investing a positive fraction in the tontine, and finally maximizes his MCPT-utility if he invests 10.86% of W_i in the tontine. An optimal fraction exists because of two counteracting effects: up to an optimal point, a higher investment in the tontine increases the later years' CPT-utilities more than it decreases the early years' CPT-utilities, yielding an increasing MCPT-utility. Beyond this optimal point, the decrease in CPT-utility in early years outweighs the increase in CPT-utility in the late years, yielding a declining MCPT-utility. These effects are resulting from the fact that up to the age of 80, the annuity provides a higher return than the tontine, while beyond the age of 80, the tontine outperforms the annuity. Therefore, one unit of additional investment in the tontine decreases CPT-utilities until the age of 80 and increases CPT-utilities beyond the age of 80, finally yielding an optimal MCPT-utility maximizing tontine investment level.

Figure 4c shows the optimal, MCPT-utility maximizing fractions to be invested in the tontine for different levels of initial wealth W_i. If $W_i < 553,000$ EUR, it is optimal not to invest in the tontine. This is because even for complete annuitization, annuity payments are so low that the CPT-utility losses in early years, caused by investing in the tontine, are large and cannot be offset by the CPT-utility gains in later years, caused by increasing tontine payouts. As W_i increases, the optimal

fraction to invest in the tontine increases very sharply up to $W_i = 571,000$ EUR and decreases thereafter. In the wealth region $553,000 \leq W_i \leq 751,000$ the reduction in early years' CPT-utilities due to shifting from annuity to tontine investment is over-compensated by the increase in late years' CPT-utilities. This is because the marginal CPT-utility in early years is lower for higher W_i, and therefore more wealth can be shifted from the annuity to the tontine investment. The optimal tontine fraction decreases beyond the peak at $W_i = 571,000$ EUR because for higher W_i marginal CPT-utility decreases for late years' consumption and less wealth in relative terms is needed to increase late years' CPT-utilities. In other words, the CPT-utilities in early years do not decline much, while late years' consumption can be financed with the additional tontine payments. For $W_i > 751,000$ EUR it is again optimal not to invest at all in the tontine. At this wealth level, the annuity payments are sufficient to satisfy the liquidity need in early as well as in later years. An investment in the tontine thus would reduce early consumption possibilities and therefore reduce early years' CPT-utility, while the gain from later consumption would be very small because later years' liquidity need can already be met by the annuity payments. Therefore, the tontine would take away funds in early years in which survival prospects are high and therefore negatively impact utility. In turn, the tontine would provide funds in states when the additional tontine payments are not needed because funds from the annuity payments are already sufficient. In addition, these funds hardly contribute to MCPT-utility because of low survival prospects at high ages[41].

Figure 4d shows the optimal MCPT-utility for different levels of W_i compared to the MCPT-utility under complete annuitization. For $W_i < 553,000$ EUR and $W_i > 751,000$ EUR, complete annuitization provides the highest MCPT-utility. As seen before, in these domains tontine investment reduces the MCPT-utility. For $553,000 \leq W_i \leq 751,000$ EUR, the highest MCPT-utility can be achieved by investing a positive fraction of wealth in the tontine. The highest MCPT-utility increase can be generated at $W_i = 584,000$ EUR. This can be seen in the gray shaded area, which corresponds to the scale on the right hand side of the figure.

The central result in the previous tontine literature is that a tontine has the highest (von-Neumann–Morgenstern) expected utility if its payment profile corresponds to that of a constant annuity.[42] Our approach is different in that we take the age-increasing payout profile of the tontine as given and use this natural property to analyze how well the tontine is suited to meet the needs of an aging individual. For this, we do not optimize the intertemporal consumption pattern. We rather optimize the investment in a tontine to cover the age-increasing liquidity need by using Cumulative Prospect Theory. Thus, our results already differ by assumption from the previous literature, on the one hand, due to the different utility concept, and on the other hand since we do not engineer the naturally resulting tontine payments. Therefore,

---

[41] If we used a positive subjective discount rate δ, the advantage of the tontine investment in the later years would diminish (See Fig. 4a), resulting in lower optimal fractions to invest in the tontine.

[42] See for example Milevsky and Salisbury [42] Milevsky and Salisbury [43], Chen et al. [12] and Chen et al. [11].

our results do not contradict previous results, as we look at the tontine from a different angle.

5.2 Variation: equal treatment of gains and losses

If we set the parameters of the value function of the CPT to $a = b = 0.5$ and the CPT loss sensitivity factor to $\lambda = 1$, we receive a square root utility, by which gains and losses are treated equally. As Fig. 5 shows, the resulting fractions of tontine investment are generally similar compared to the base case setting. It is striking that at $W_i = 591,000$ EUR the optimal fraction to invest in the tontine immediately jumps from 0 to 10.73% and decreases thereafter, until it finally reaches 0 again at $W_i = 950,000$ EUR. Compared to the base case, investment in the tontine is optimal for higher W_i. Figure 6 provides an explanation for these results.

Figure 6a shows the MCPT-utility values for $W_i = 560,000$ EUR for different levels of tontine investment. The highest MCPT-utility can be achieved if no investment in the tontine takes place. If the tontine investment increases, the MCPT-utility first decreases, and at roughly 12% there is a little peak with a local maximum where MCPT-utility slightly increases, but decreases thereafter again[43] (Fig. 6b). As initial wealth reaches the threshold value $W_i = 591,000$ EUR (Fig. 6c), the hump is as large as that the MCPT-utility with 10.73% tontine investment equals the MCPT-utility without tontine investment. Therefore, the individual is indifferent between no tontine investment and 10.73% tontine investment. For a tontine investment between 0 and 10.73%, the MCPT-utility is lower compared to the maximum MCPT-utility. For tontine investments larger than 10.73%, the MCPT-utility decreases as well. As W_i further increases, the peak further moves to the left and surmounts the MCPT-utility without tontine investment (Fig. 6d). Gradually, the local minimum between no tontine investment and optimal tontine investment disappears (Fig. 6e). Finally, as W_i is very large, the slope around the local maximum is very flat and finally disappears when the MCPT-utility maximizing fraction to invest in the tontine hits 0 again (Fig. 6f).

5.3 Variation: tontine size

For an increased tontine size of $N = 100,000$ (compared to $N = 10,000$ in the base case), the volatility of the tontine payments declines (Table 2). As presented in Fig. 7, less volatile tontine payments make it optimal to invest in the tontine for lower W_i than in the base case scenario. Similarly, for higher W_i, investing in the tontine remains beneficial with an increased pool size. This is because less volatile payments generally enhance CPT-utilities. Therefore, it is optimal for both a lower and a higher W_i to invest more in the tontine compared to $N = 10,000$.

---

[43] The reason for this peak lies in the tradeoff between early and later years' consumption possibilities as explained in the previous base case.
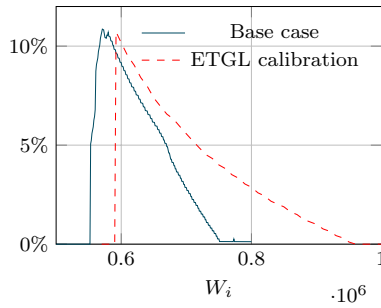
Fig. 5 Equal treatment of gains and losses-calibration (ETGL) maximizing fractions to invest in the tontine for different levels of W_i
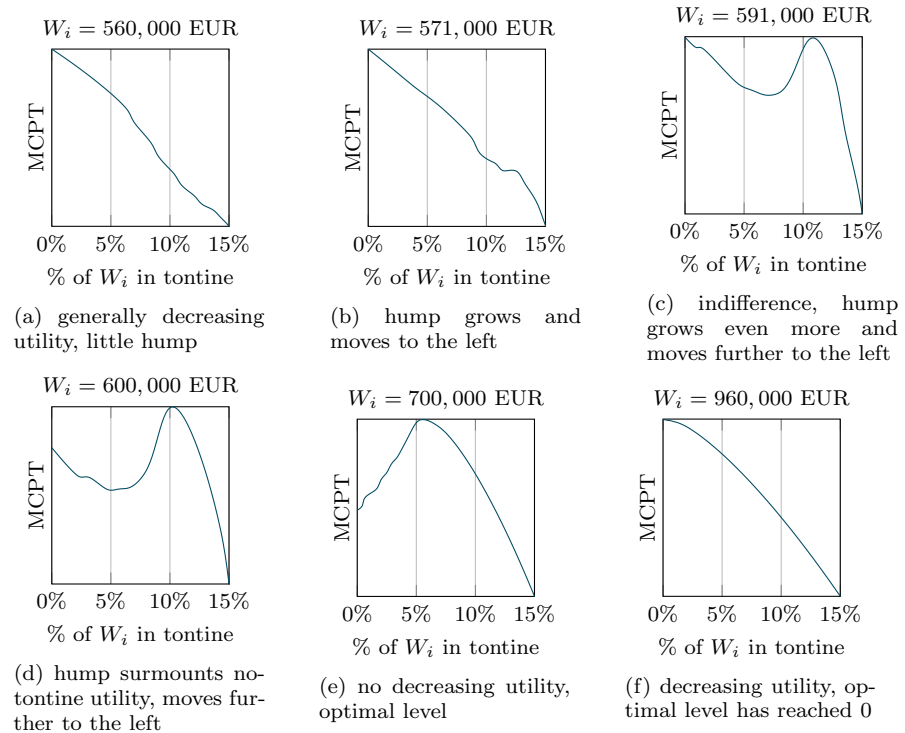

(a) generally decreasing utility, little hump

(b) hump grows and moves to the left

(c) indifference, hump grows even more and moves further to the left

(d) hump surmounts no-tontine utility, moves further to the left

(e) no decreasing utility, optimal level

(f) decreasing utility, optimal level has reached 0

Fig. 6 MCPT-utility for different levels of W_i and increasing fractions of tontine investment relative to W_i for square-root utility calibration of the base case

5.4 Variation: stochastic liquidity need

If we assume a stochastic liquidity need, the fact whether cash flows lead to gains or losses with respect to the liquidity need is affected by the volatility of the tontine payments as well as by the volatility of the liquidity need. A stochastic liquidity
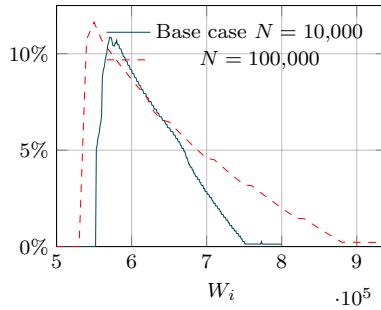
Fig. 7 MCPT-utility maximizing fractions to invest in the tontine for different levels of W_i for $N = 100,000$

Table 2 Properties of normally distributed tontine returns for $B_i = 30,000$ in t in EUR for tontine size $N = 100,000$ vs. $N = 10,000$

t	1	10	20	30	40	44
$v_{i,t}$	345.15	559.72	1652.04	5619.14	14,307.30	30,000
$\sigma_{i,t}^{N=10,000}$	19.05	25.30	63.63	227.70	572.00	1169.77
$\sigma_{i,t}^{N=100,000}$	5.93	8.74	23.00	78.43	187.74	370.75
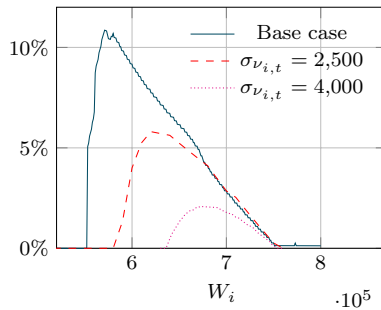
Fig. 8 MCPT-utility maximizing fractions to invest in the tontine for different levels of W_i for stochastic liquidity need

need increases the overall volatility and therefore CPT-utilities decline. First, we set the variance of the liquidity need at $\sigma_{v_{i,t}}^2 = \sigma_{D_{i,t}}^2 = 2500^2$ which could reflect both, uncertainty about health care costs and uncertainty of becoming frail and dependent. As a consequence, the wealth level at which it becomes optimal to invest in the tontine increases compared to the base case (see Fig. 8). The reason for this lies in the fact that in early years the more volatile nature of gains and losses makes it desirable to hold more funds to cover the liquidity need. Every unit taken away from the annuity in the early years causes a huge decline in early years' CPT-utilities. Therefore, it is optimal only for a higher W_i to invest in the tontine. The opposite effect applies to
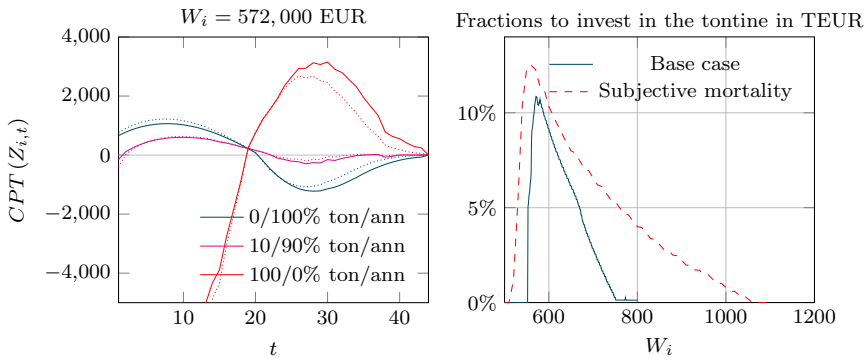
(a) CPT for different asset allocations, actual vs subjective mortality $(d = 0.8)$

(b) MCPT-utility maximizing fractions to invest in the tontine for different levels of W_i for actual and subjective mortality $(d = 0.8)$

Fig. 9 Subjective mortality beliefs

high W_i. The volatile liquidity need brings about situations of high liquidity need in which payments from the tontine can support its coverage. Therefore, it is optimal for a higher W_i to hold some fraction in the tontine. Furthermore, the optimal fraction to invest in the tontine is smaller compared to the base case, because the tontine investment itself adds another layer of volatility to the payments, which decreases utility. As we further increase the volatility of the liquidity need to $\sigma^2_{v_{i,t}} = \sigma^2_{D_{i,t}} = 4,000^2$, we can observe a boost of both effects. A higher W_i is required to start investing in the tontine in order to lower the risk of experiencing a utility-harming drop far below the liquidity need. As the level of W_i is relatively high, the tontine investment loses its efficiency compared to the resulting annuity payments, yielding a lower optimal fraction to be invested in the tontine.

5.5 Variation: subjective mortality

If we adjust the mortality according to Eq. (29) by $d = 0.8$, the individual expects to live longer than average. This means that future periods have a greater impact on MCPT-utility because survival probabilities decline less fast. Therefore, later years' CPT-utilities are higher compared to the base case. This situation is presented in Fig. 9a, where the dotted lines represent the CPT-utility paths for the base case and the solid lines represent the CPT-utility paths for the subjective, improved mortality. As a result, positive and negative subjective CPT-utilities both have a higher impact on total MCPT-utility compared to the base case, indicating that it might be more favorable to invest a higher fraction in the tontine because it is more likely to experience the later years' CPT-utilities. Figure 9b shows that it is optimal to invest in the tontine for lower W_i because, by investing in the tontine, later years' CPT-utilities gain more relevance and are higher although early years' CPT-utilities are reduced. By investing more intensely in the tontine, overall MCPT-utility can be increased. Furthermore, it is optimal to invest in the tontine up to a higher W_i, compared to

(a) Variations of the liquidity need curve

(b) MCPT-utility maximizing fractions to invest in the tontine for different levels of W_i for different shapes of liquidity need
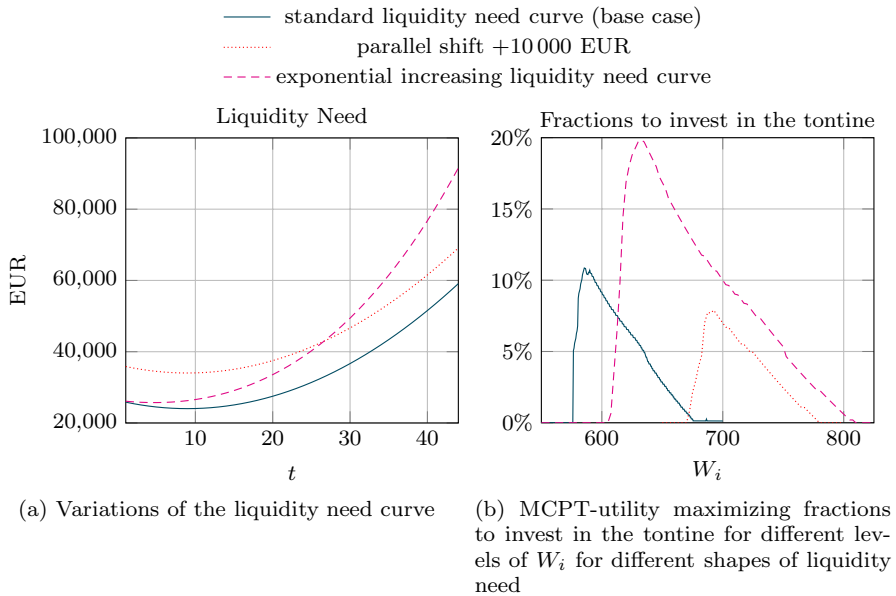
Fig. 10 Variations in liquidity need shape

the base case. This is because marginal CPT-utility in later years increases as survival probabilities increase. Early years' CPT-utility losses can be overcompensated by later years' CPT utility gains. Furthermore, later years' CPT-utility losses also have a higher impact on the MCPT and, therefore, a tontine investment in higher W_i regions can help to mitigate the otherwise resulting underfunding problem. In addition, it is optimal to invest a higher fraction in the tontine for all W_i for which it is optimal to invest in the base case. This is due to the increased probability of experiencing CPT-utilities in the late years.

5.6 Variation: changing liquidity need

In this section we change the shape of the liquidity need. First, we parallel shift the standard liquidity need curve up by 10,000 EUR. Second, we assume an exponential growth of the standard liquidity need curve by $D_t^{exp} = 1.01^t D_t$. Since the standard liquidity need curve represents the average liquidity need unconditioned on the health status, an exponential growth can be interpreted as the liquidity need conditional on bad health. Figure 10a, b show the resulting liquidity need curves and the optimal fractions to invest in the tontine for the different liquidity need curves. If we assume a parallel, upward shift of the liquidity need curve by 10,000 EUR, two characteristics of the optimal investment choice can be observed. First, the optimal investment pattern shifts to the right, which means that it is optimal to invest in the tontine only for higher initial wealth endowment W_i. Second, the optimal fractions to invest in the tontine are lower compared to the base case. The reason for these
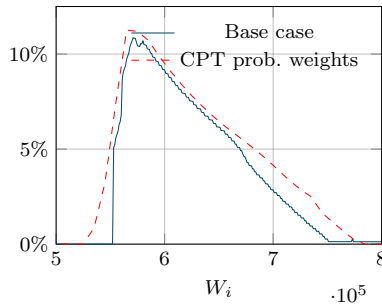
Fig. 11 MCPT-utility maximizing fractions to invest in the tontine for objective ($\gamma = \kappa = 1$) and subjective ($\gamma = 0.61$ and $\kappa = 0.69$) probability weights

two properties lies in the increasing liquidity need in every period. For a relatively low W_i, it is not optimal to invest in the tontine because the loss in CPT-utilities in the early years due to a reduction of annuitized wealth exceeds the CPT-utility gains in later years. Only if there is sufficient initial wealth it is optimal to invest in the tontine. The maximal tontine investment in the parallel shift case is lower compared to the peak in the base case. This is a consequence of a substantially higher liquidity need in the early years which can not be covered by tontines. If we assume an exponentially increasing liquidity need, investment in the tontine starts for a higher W_i compared to the base case and below the parallel shift case at approximately $W_i = 600,000$ EUR. Furthermore, the peak of the optimal amount to be invested in the tontine is almost twice as large compared to the base case. Optimal positive fractions of tontine investment persist longer for high W_i. This is because in the early years the liquidity need in the exponential case is relatively close to the base case and disproportionately increases with age, compared to the base case. Therefore, the CPT-utility decrease in early years is relatively low when investing some fraction in the tontine, while the CPT-utility gains of the tontine investment in the late years are very high. Thus, with large amounts of money invested in the tontine, the early years' CPT-utilities do not suffer much, while later years' CPT-utilities benefit strongly. As a consequence, larger amounts to be invested in the tontine are optimal to satisfy the liquidity need best. To sum up, the tontine is most powerful if the liquidity need is low in the early years and high in the later years of retirement.

5.7 Variation: cumulative prospect theory probability weights

If we calibrate the probability weighting function of the Cumulative Prospect Theory according to the originally proposed values by Tversky and Kahneman [55], $\gamma = 0.61$ and $\kappa = 0.69$, we can observe in Fig. 11 that the optimal fractions to invest in the tontine are on a similar level as using objective probabilities. However, using the distorted weights makes it optimal to begin to invest in the tontine for a slightly lower W_i, and that it is optimal to invest a larger fraction in the tontine for any wealth level, where it is optimal in the base case to invest at least some fraction.

Furthermore, it is still optimal for a higher available wealth W_i to invest in the tontine. The reason for the slightly different tontine investment pattern is that small probabilities are over- and large probabilities are under-weighted. Small gains and losses occur with relatively equal probabilities and thus are hardly distorted. However, large gains and losses are low probability events and are therefore perceived to occur less frequent. Since a large loss harms more than a large gain benefits, the lower perception of those extreme events yields that the optimal tontine investment curve with distorted probabilities lies slightly above the base case optimality curve.

6 Summary and conclusion

The changing social, financial and regulatory framework, such as an increasingly aging society, the current low-interest environment, as well as the implementation of risk-based capital standards in the insurance industry, lead to the search for new product forms for private pension provision. These product forms should reduce or avoid investment guarantees and risks stemming from longevity, provide reliable insurance benefits and reflect in the payout pattern the increasing financial resources required for very high ages. We propose the traditional tontine to serve as such "product innovation", especially in combination with a traditional life annuity.

To assess the effects of tontine investments on policyholders' welfare, we develop a lifetime-utility model based on Multi Cumulative Prospect Theory by which individual old-age liquidity needs and payouts stemming from annuity and tontine investment can be evaluated. The analysis results in an optimal retirement planning decision, based on individual preferences, characteristics and subjective mortality beliefs.

To show the effects of a tontine investment on retirement planning, we model the development of the changing population structure for the next decades in Germany. Based on the changing mortality dynamics, we describe a fair revolving tontine. To assess its advantages and disadvantages compared to a traditional life annuity, we derive a targeted consumption level from empirical data and combine the tontine payout structure with the payout structure of a traditional annuity to optimally cover the desired consumption. Our results reveal that a portfolio of annuity and tontine can provide the highest expected Multi Cumulative Prospect Theory utility. While the annuity pays a stable, constant pension, the tontine provides volatile, age-increasing payouts. However, purchasing tontines is only optimal for individuals within a certain range of wealth. On the one hand, these individuals can cover their liquidity needs in their early years of retirement by traditional annuities. On the other hand, they are wealthy enough to forgo income in those early retirement years by investing in tontines, which in the subsequent old age years generate the required income that could otherwise not be provided by level annuity payments. The higher the liquidity need is in the later years of retirement, the higher is the demand for tontines. In contrast, individuals with a low wealth level need traditional annuity payments in their early years of retirement, and cannot give up income for investing in tontines whose benefits mainly accrue in later years. Individuals with a very high wealth level, in turn, do not need the tontine income to cover their liquidity needs in their later years of retirement, while the tontine investment reduces

their income in the early years of retirement. These results, in principle, also hold if we apply von-Neumann-Morgenstern expected utility maximization.

Through the advantages of pooling, a greater tontine size extends the range of initial wealth endowments for which tontine investments become advantageous. A further result of our analysis is that the more volatile the liquidity need is, the lower is the demand for tontines, because traditional annuities are then more useful to close possible liquidity gaps in the early retirement years that contribute strongly to the overall utility. Finally, the demand for tontines is higher for individuals with lower mortality rates because of the expected longer lifetime in which they receive high tontine payouts.

Future research could incorporate investment risk and analyze its effects on the optimal tontine investment decision. The level of investment risk might significantly change the optimal allocation of retirement wealth. In this context, the integration of reinvestment opportunities of tontine and annuity returns might yield an additional determining factor for the optimal retirement planning decision.

Further research could include a mechanism for financing liquidity gaps or investing excess liquidity during the decumulation phase. This could be done in a two-step approach. In a first step the deviations from reference point, i.e. the liquidity need, could be evaluated and would go into the MCPT-valuation. In a second step, financing the gap or investing surplus would take place, leading to a change of deviations from liquidity needs in future years. These future deviations would then be treated in the same two-step approach. This procedure would, however, not measure utility and disutility of consumption (deviations), but rather utility and disutility from actual investment and financing activities.

Another interesting area for further research is the analysis from an insurer's perspective. It will be interesting to analyze whether the supply of tontines can reduce the insurer's capital requirements, or reduce safety loadings included in annuity prices, since the tontine involves no longevity risk for the provider, and might partially substitute annuities.

Appendix

Tontine volatility

The variance of the return conditional that i survives in t is

$$
Var\left[r_{i,t}|A_{i,t}^c\right] = E\left[r_{i,t}^2|A_{i,t}^c\right] - \left(E\left[r_{i,t}|A_{i,t}^c\right]\right)^2
$$

$$
= E\left[\left(\sum_{\substack{k=1\\k\neq i}}^{N} a_{i,k,t}B_k\mathbf{1}_{\{A_{k,t}\}}\right)^2\right] - \left(q_{i,t}B_i\right)^2.
$$

If we assume constant $\bar{a}_{i,k,t}$'s, the variance can be expressed as

$$
Var\left[r_{i,t}|A_{i,t}^c\right] = q_{i,t}^2 B_i^2 \frac{\sum_{\substack{k=1\\k\neq i}}^{N} B_k^2 E\left[\mathbf{1}_{\{A_{k,t}\}}^2\right]}{\sum_{\substack{k=1\\k\neq i}}^{N} B_k^2 q_{k,t}^2 + \sum_{\substack{k=1\\k\neq i}}^{N}\sum_{\substack{k=1\\j\neq i}}^{N} B_k B_j q_{k,t} q_{j,t}}
$$

$$
+ q_{i,t}^2 B_i^2 \frac{\sum_{\substack{k=1\\k\neq i}}^{N}\sum_{\substack{k=1\\j\neq i}}^{N} B_k B_j E\left[\mathbf{1}_{\{A_{k,t}\}}\right] E\left[\mathbf{1}_{\{A_{j,t}\}}\right]}{\sum_{\substack{k=1\\k\neq i}}^{N} B_k^2 q_{k,t}^2 + \sum_{\substack{k=1\\k\neq i}}^{N}\sum_{\substack{k=1\\j\neq i}}^{N} B_k B_j q_{k,t} q_{j,t}} - q_{i,t}^2 B_i^2.
$$

For $N \to \infty$

$$
Var\left[r_{i,t+1}|A_{i,t}^c\right] = q_{i,t}^2 B_i^2 - q_{i,t}^2 B_i^2 = 0
$$

because $E\left[\mathbf{1}_{\{A_{k,t}\}}\right] = q_{k,t}$. Nevertheless, the tontine we employ is large, but still risky[44]. For a fixed tontine size N, the tontine payouts are still volatile and follow a Poisson binomial distribution in each t which we approximate by a normal distribution[45] $\mathcal{N}\left(v_{i,t}, \sigma_{i,t}\right)$ for large N where $v_{i,t} = q_{i,t}B_i$ and

---

[44] At a tontine size of $N = 100,000,000$ members, the volatility would be negligible. Of course, a tontine of this size is not realistic. Therefore, in our numerical approach, we employ a tontine size of $N = 10,000$ members which might be practically realizable and which comprises significant volatility.

[45] See for example Volkova [58], Hong et al. [30], Hong and Meeker [29] and Hong [28].

$$\sigma_{i,t} = \frac{q_{i,t}Bi}{\sqrt{M-1}} \sqrt{\sum_{m=1}^{M} \left(\frac{\sum\limits_{\substack{k=1 \\ k \neq i}}^{N} B_k \mathbf{1}^m_{\{A_{k,t}\}}}{\sum\limits_{\substack{k=1 \\ k \neq i}}^{N} B_k q_{k,t}} - 1 \right)^2}.$$

with M Monte Carlo simulation paths.

Modeling subjective mortality

Hoermann and Ruß [27] propose a gamma distribution for modeling $D \sim \Gamma(\alpha, \beta, \gamma)$ with density function

$$f^{\Gamma}_{(\alpha,\beta,\gamma)}(d) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}(d-\gamma)^{\alpha-1}e^{-\frac{d-\gamma}{\beta}},$$

expected value

$$E(D) = \alpha\beta + \gamma$$

and variance

$$Var(D) = \alpha\beta^2$$

for $d \geq \gamma, \gamma \in \mathbb{R}$ and $\alpha, \beta > 0$.

Calibration of the Lee-Carter model

The fitted parameters α, β and κ of the Lee-Carter model re depicted in Figs. 12, 13, 14.
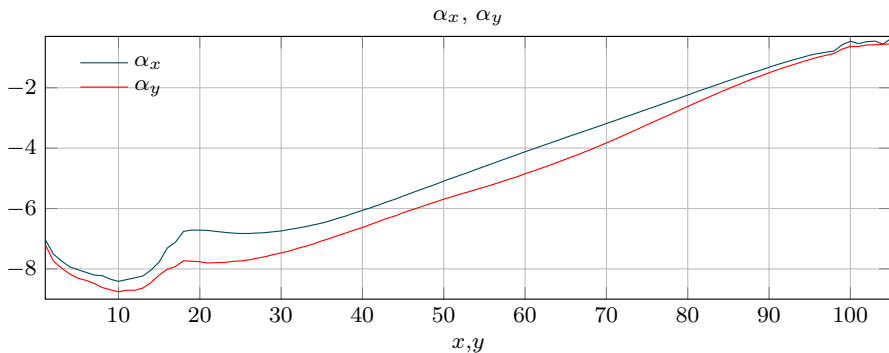
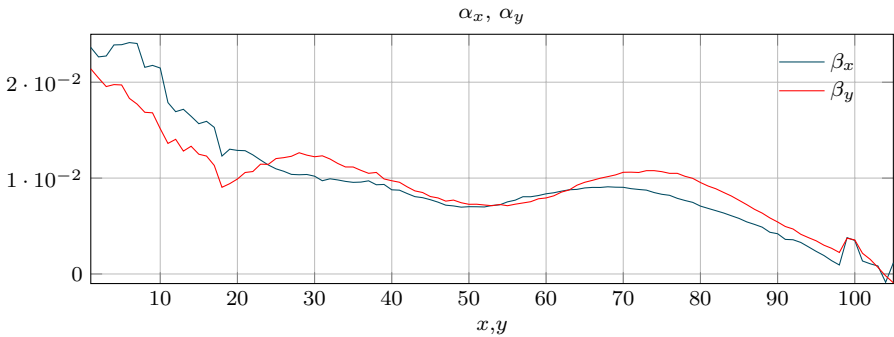**Fig. 12** Lee-Carter fit for male α_x and female α_y for $x, y = 1 \dots 105$

Fig. 13 Lee-Carter fit for male β_x and female β_y for $x, y = 1 \ldots 105$
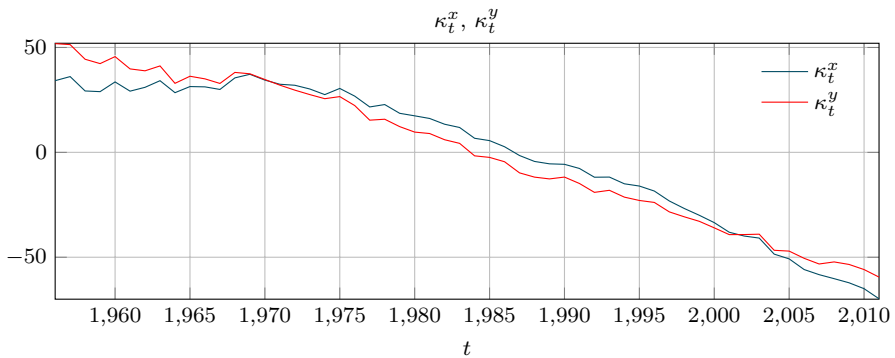

Fig. 14 Lee-Carter fit for male κ_t^x and female κ_t^y for $t = 1956 \ldots 2011$

References

1. Ågren M (2006) Prospect theory and higher moments. Tech. Rep., Working Paper, Department of Economics, Uppsala University
2. Banks J, Blundell R, Tanner S (1998) Is there a retirement-savings puzzle? Am Econ Rev 88(4):769–788. http://www.jstor.org/stable/117005
3. Barberis N, Huang M (2008) Stocks as lotteries: the implications of probability weighting for security prices. Am Econ Rev 98(5):2066–2100
4. Barberis N, Huang M, Santos T (2001) Prospect theory and asset prices. Q J Econ 116(1):1–53
5. Berdin E, Gründl H (2015) The effects of a low interest rate environment on life insurers. Geneva Pap Risk Insur Issues Pract 40(3):385–415
6. Bernhardt T, Donnelly C (2019) Modern tontine with bequest: innovation in pooled annuity products. Insur Math Econ 86:168–188
7. Bernheim BD, Skinner J, Weinberg S (2001) What accounts for the variation in retirement wealth among U.S. households? Am Econ Rev 91(4):832–857. http://www.jstor.org/stable/2677815
8. Bissonnette L, Hurd MD, Michaud PC (2017) Individual survival curves comparing subjective and observed mortality risks. Health Econ 26(12):e285–e303
9. Blanchett D (2013) Estimating the true cost of retirement. https://corporate.morningstar.com/ib/documents/MethodologyDocuments/ResearchPapers/Blanchett_True-Cost-of-Retirement.pdf

10. Bundesanstalt für Finanzdienstleistungsaufsicht (2014) Statistik der Bundesanstalt für Finanz-dienstleistungsaufsicht. BaFin, Stabsstelle Presse- u. Öffentlichkeitsarbeit, Erstversicherungsunternehmen und Pensionsfonds
11. Chen A, Hieber P, Klein JK (2019) Tonuity: a novel individual-oriented retirement plan. ASTIN Bull J IAA 49(1):5–30
12. Chen A, Rach M, Sehner T (2020) On the optimal combination of annuities and tontines. ASTIN Bull 50:95–129
13. Demakakos P, Biddulph JP, Bobak M, Marmot MG (2016) Wealth and mortality at older ages: a prospective cohort study. J Epidemiol Community Health 70(4):346–353
14. DESA U (2019) World population prospects 2019: highlights. United Nations Department for Economic and Social Affairs, New York (US)
15. European Insurance and Occupational Pensions Authority (2019a) Consumer trends report 2019
16. European Insurance and Occupational Pensions Authority (2019b) Financial stability report 2019
17. Financial Conduct Authority (2019) Retirement outcomes review final report
18. Forman JB, Sabin MJ (2014) Tontine pensions. Univ Pa Law Rev 163(3):755–832
19. Gemmo I, Rogalla R, Weinert JH (2020) Optimal portfolio choice with tontines under systematic longevity risk. Ann Actuar Sci 14(2):302–315
20. Gesamtverband der Deutschen Versicherungswirtschaft eV (2019) Die Deutsche Lebensversicherung in Zahlen 2019
21. Graham RL, Grotschel M, Lovász L, Wilf HS (1997) Handbook of combinatorics. Math Intell 19(2):65
22. Gruenberg EM (1977) The failures of success. Milbank Mem Fund Q Health Soc 55(1):3–24. http://www.jstor.org/stable/3349592
23. Haider SJ, Stephens M Jr (2007) Is there a retirement-consumption puzzle? Evidence using subjective retirement expectations. Rev Econ Stat 89(2):247–264
24. Hamermesh DS (1984) Consumption during retirement: the missing link in the life cycle. Rev Econ Stat 66(1):1–7. http://www.jstor.org/stable/1924689
25. He XD, Zhou XY (2011) Portfolio choice under cumulative prospect theory: an analytical treatment. Manag Sci 57(2):315–331
26. Hens T, Rieger MO (2010) Financial economics: a concise introduction to classical and behavioral finance. Springer Science & Business Media, Berlin
27. Hoermann G, Ruß J (2008) Enhanced annuities and the impact of individual underwriting on an insurer's profit situation. Insur Math Econ 43(1):150–157
28. Hong Y (2011) On computing the distribution function for the sum of independent and nonidentical random indicators. Department of Statistics, Virginia Tech, Blacksburg, VA, USA, Technical Report 112
29. Hong Y, Meeker WQ (2010) Field-failure and warranty prediction based on auxiliary use-rate information. Technometrics 52(2):148–159
30. Hong Y, Meeker WQ, McCalley JD (2009) Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. Ann Appl Stat 3(2):857–879. http://www.jstor.org/stable/30244268
31. Hunt A, Blake DP (2015) Identifiability in age/period mortality models. Pensions Institute Discussion Paper, PI-1508
32. Jagger C, Matthews FE, Wohland P, Fouweather T, Stephan BC, Robinson L, Arthur A, Brayne C, Function MRCC, Collaboration A et al (2016) A comparison of health expectancies over two decades in England: results of the cognitive function and ageing study I and II. Lancet 387(10020):779–786
33. Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–291. http://www.jstor.org/stable/1914185
34. Kochskämper S, Pimpertz J (2015) Herausforderungen an die Pflegeinfrastruktur. IW-Trends Vierteljahresschrift zur empirischen Wirtschaftsforschung 42(3):59–75
35. Koijen RS, Van Nieuwerburgh S, Yogo M (2016) Health and mortality delta: assessing the welfare cost of household insurance choice. J Finance 71(2):957–1010
36. Lee RD, Carter LR (1992) Modeling and forecasting US mortality. J Am Stat Assoc 87(419):659–671
37. Levy H, Levy M (2004) Prospect theory and mean-variance analysis. Rev Financ Stud 17(4):1015–1041
38. Manes A (1932) Versicherungswesen, iii. Bd, Personenversicherung, Teubner, Leipzig and Berlin

39. Mariger RP (1987) A life-cycle consumption model with liquidity constraints: theory and empirical results. Econometrica 55(3):533–557. http://www.jstor.org/stable/1913599
40. McKeever K (2009) A short history of tontines. Fordham J Corp Financ Law 15:491–521
41. Milevsky MA (2006) The calculus of retirement income: financial models for pension annuities and life insurance. Cambridge University Press, Cambridge
42. Milevsky MA, Salisbury TS (2015) Optimal retirement income tontines. Insur Math Econ 64:91–105
43. Milevsky MA, Salisbury TS (2016) Equitable retirement income tontines: mixing cohorts without discriminating. ASTIN Bull J IAA 46(3):571–604
44. Nigri A, Levantesi S, Marino M, Scognamiglio S, Perla F (2019) A deep learning integrated lee-carter model. Risks 7(1):33
45. Parsonage M, Neuburger H (1992) Discounting and health benefits. Health Econ 1(1):71–76
46. Renshaw A, Haberman S (2008) On simulation-based approaches to risk measurement in mortality with specific reference to poisson Lee-Carter modelling. Insur Math Econ 42(2):797–816
47. Renshaw AE, Haberman S (2003) Lee-Carter mortality forecasting with age-specific enhancement. Insur Math Econ 33(2):255–272
48. Ruß J, Schelling S (2018) Multi cumulative prospect theory and the demand for cliquet-style guarantees. J Risk Insur 85(4):1103–1125
49. Sabin MJ (2010) Fair tontine annuity. Social Science Research Network Working Paper Series
50. Sabin MJ (2011) A fast bipartite algorithm for fair tontines. Social Science Research Network Working Paper Series
51. Schmidt U (2016) Insurance demand under prospect theory: a graphical analysis. J Risk Insur 83(1):77–89
52. Standard Life (2013) The retirement smile. http://ukgroup.standardlife.com/content/news/new_articles/2013/260613NewDrawdownTransferOption.xml
53. Statistisches Bundesamt (2015) Bevölkerung Deutschlands bis 2060: 13. koordinierte Bevölkerungsvorausberechung. https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060Presse5124204159004.pdf?__blob=publicationFile
54. Statistisches Bundesamt (2015) Bevölkerung Deutschlands bis 2060: 13. koordinierte Bevölkerungsvorausberechnung. http://www.jstor.org/stable/1170050
55. Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncertain 5(4):297–323. https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060Presse5124204159004.pdf?__blob=publicationFile
56. United States Census Bureau (2016) International programs: international data base. https://www.census.gov/data-tools/demo/idb/informationGateway.php
57. Van der Pol MM, Cairns JA (2000) Negative and zero time preference for health. Health Econ 9(2):171–175
58. Volkova AY (1996) A refinement of the central limit theorem for sums of independent random indicators. Theory Probab Appl 40(4):791–794
59. Weinert JH (2017) Comparing the cost of a tontine with a tontine replicating annuity. ICIR Working Paper Series, Goethe University Frankfurt 31/2017
60. Weinert JH (2017) The fair surrender value of a tontine. ICIR Working Paper Series, Goethe University Frankfurt 26/2017
61. Worldbank (2015) Databank. Life expectancy at birth. http://data.worldbank.org/indicator/SP.DYN.LE00.IN
62. Yogo M (2016) Portfolio choice in retirement: health risk and the demand for annuities, housing, and risky assets. J Monet Econ 80:17–34