Ulm University

Institute of Insurance Science

Optimal Design of Private and Occupational Retirement Plans

Cumulative doctoral thesis

Submitted to the Faculty of Mathematics and Management of Ulm University

For the academic degree of Doctor of Management and Economics Dr. rer. pol.

By

Manuel Matthias Rach

Born in Filderstadt

2020

Dean in office:	Prof. Dr. Martin Müller
Supervisor and first evaluator:	Prof. Dr. An Chen
Second evaluator:	apl. Prof. Dr. Hans-Joachim Zwiesler
Third evaluator:	Prof. Dr. Joël Wagner
Date of the oral examination:	March 16, 2020

Acknowledgments

First and foremost, I want to express my gratitude to my supervisor Prof. Dr. An Chen. Her love for research has been contagious and inspiring for me and her patient guidance and encouraging support have, at all stages of this thesis, been invaluable to me. Without her, I would certainly not be where I am today. I want to extend my thanks to her for giving me the opportunity to assist her in many lectures and seminars as a teaching assistant at the Institute of Insurance Science, all of which I enjoyed very much.

Additionally, I want to thank apl. Prof. Dr. Hans-Joachim Zwiesler for reviewing this thesis. I am especially grateful to him for providing me the opportunity to assist him in the joint research study with Aon on behalf of the BMAS and BMF, where I learned a lot from his experience and knowledge. Needless to say that I enjoyed assisting him in his lectures just as much.

Furthermore, I want to thank Prof. Dr. Gerlinde Fellner-Röhling for being a member of the examination board and Prof. Dr. Joël Wagner for reviewing this thesis.

Moreover, I want to gratefully acknowledge all my friends and colleagues from the Institute of Insurance Science and the Institut für Finanz- und Aktuarwissenschaften (ifa) for the great atmosphere and many fruitful discussions in our joint workshops. Special thanks are given to Dr. Thai Nguyen and Dr. Peter Hieber for many helpful suggestions.

In addition, I would like to acknowledge the financial support given by the DFG for the research project "Zielrente: die Lösung zur alternden Gesellschaft in Deutschland".

I would like to extend my thanks to participants at many conferences and anonymous referees for their valuable suggestions and comments which helped improve this dissertation significantly.

Finally, I am deeply grateful to my parents and my brother for their endless support and loving encouragement which made all this possible in the first place.

Ulm, May 2020 Manuel Rach

Contents

Research Papers

Overview of Research Papers

Research papers included in this dissertation

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

2. Chen, A., Rach, M., and Sehner, T. (2020). On the optimal combination of annuities and tontines. *ASTIN Bulletin: The Journal of the IAA*, 50(1):95–129.

3. Chen, A., Hieber P., and Rach, M. (2019). Optimal retirement products under subjective mortality beliefs. Submitted to special issue of *Insurance: Mathematics and Economics* on behavioral insurance (under review).

4. Chen, A., Nguyen, T., and Rach, M. (2019). Optimal collective investment: The impact of sharing rules, management fees and guarantees. Submitted to *Journal of Banking and Finance* (revise and resubmit).

5. Chen, A., Nguyen, T., and Rach, M. (2019). A collective investment problem in a stochastic volatility environment: The impact of sharing rules. Submitted to *Annals of Operations Research* (under review).

Co-Authorship

An Chen

An Chen is a full professor and head of the Institute of Insurance Science at Ulm University. She received her doctoral degree from the University of Bonn, Germany, in 2007 and has been a professor in Ulm since 2012.

Thai Nguyen

Thai Nguyen is a PostDoc researcher at the Institute of Insurance Science at Ulm University. He received his doctoral degree from the University of Rouen, France, in 2014.

Peter Hieber

Peter Hieber is a PostDoc researcher at the Institute of Insurance Science at Ulm University. He received his doctoral degree from the Technical University of Munich, Germany, in 2013.

Thorsten Sehner

Thorsten Sehner is a fellow Ph.D. student at the Institute of Insurance Science at Ulm University. He received his Master of Science degree from Ulm University in 2018.

Research Context and Summary of Research Papers

1 Field of Research

This cumulative thesis contributes to the field of optimal retirement planning, optimal retirement product design and optimal asset allocation in the context of collective investment problems.

In Germany and many other developed countries, the retirement system can be divided into three pillars: mandatory state pension systems, occupational pensions and private pensions. An aging society and the ongoing low interest rate environment lead to a reduced retirement income from the first pillar, forcing individuals to plan appropriately for their retirement in the second and third pillar. To deal with the current societal challenges, an enormous variety of retirement plans is available and new retirement plans are developed both in the second and third pillar.

The first three research papers in this thesis focus on the third pillar. In particular, we consider novel retirement products which shift the longevity risk (risk of outliving one's financial resources in retirement) towards policyholders. Various such innovative products have recently appeared in the academic literature and the question arises how beneficial they are for policyholders and insurers. While the first paper in this thesis addresses both the policyholder's and the insurer's perspective, the second and third paper focus exclusively on the policyholder's point of view, where the third paper takes account of subjective beliefs the policyholder might have. To assess the benefits of retirement plans from a policyholder's perspective, expected utility has been frequently used in the literature to find optimal payoff structures or to compare different retirement plans (cf. Yaari's famous pioneering article Yaari (1965) as well as Yagi and Nishigaki (1993), Mitchell (2002), Davidoff et al. (2005), Milevsky and Huang (2011), Milevsky and Salisbury (2015, 2016), Peijnenburg et al. (2016), Huang et al. (2017) and Chen et al. (2019)). The first three papers in this thesis contribute to this literature and the field of optimal retirement planning by introducing, modeling and analyzing the attractiveness of novel types of retirement plans.

The main application of the last two papers lies in the second pillar. In occupational pensions, a shift towards lower guarantees can be observed. In many countries, occupational pension schemes no longer promise guarantees to employees but instead provide a retirement income which depends substantially on the performance of the financial market experienced during the accumulation phase (cf. OECD (2016)). In the academic literature, it is common that employees are, during the accumulation phase, modeled as investors in a financial market. Using an expected utility framework, it is then possible to derive the optimal continuous-time trading strategy and the resulting optimal wealth level. For a single investor, this procedure is well-documented, covered in many textbooks (for example Korn (2014)) and goes back to Merton (1969, 1971). Utility maximization problems under portfolio insurance (minimum guarantees) is also covered in the literature, see, for example, Grossman and Vila (1989), Basak (1995), Grossman and Zhou (1993, 1996), Browne (1999), Tepla (2001), Jensen and Sørensen (2001), Deelstra et al. (2004), El Karoui et al. (2005), Gabih et al. (2009) and Chen et al. (2018). The last two papers of this thesis contribute to the literature on optimal asset allocation by analyzing novel types of optimization problems where individuals with heterogeneous risk preferences are tied together in their investment decision and invest collectively under portfolio insurance.

2 Motivation and Objectives

Ongoing low interest rates along with growing life expectancies are challenging modern societies all over the world. Life insurers, policyholders and governments are searching for possible ways to tackle these issues.

One of these recent developments has been the resurrection of the so-called *tontine*, named after its inventor Lorenzo de Tonti. Tontines used to be a popular source of retirement income from the 17th to the 19th century (see, for example, Milevsky (2014, 2015), Milevsky and Salisbury (2015, 2016) and Li and Rothschild (2019)). The idea of a tontine is that a group of policyholders shares the mortality risk and that the insurance company only serves as an administrator. Along with tontines, so-called pooled annuity funds or group self-annuitization schemes have appeared. These innovative retirement products basically follow a tontine-like structure and many efforts have been made to explore the potential and optimal design of these products in today's world, see, for example, Piggott et al. (2005), Valdez et al. (2006), Stamos (2008), Qiao and Sherris (2013), Donnelly et al. (2013, 2014) and Donnelly (2015). In such products, the unsystematic mortality risk can (initially) be diversified by a large enough pool, whereas the systematic mortality risk, which affects all policies in the same direction, cannot.

When priced actuarially fair, life annuities give retirees greater lifetime utility than tontines (see also Milevsky and Salisbury (2015)). When realistic safety loadings or risk margins are taken into consideration, tontines can be preferred to annuities (cf. Milevsky and Salisbury (2015) and Chen et al. (2019)). Tontines and annuities can be seen as two extreme cases of retirement products, as annuities leave the insurer with all the longevity risk, while tontines leave the policyholders with (almost) all of the longevity risk. As a consequence, there have also been suggestions for combining the advantages of both products, for example, in Chen et al. (2019) and Weinert and Gründl (2017). While Chen et al. (2019) present a new retirement product called tonuity, which is a tontine at early retirement ages but switches to an annuity at a predetermined switching time, Weinert and Gründl (2017) focus on how the policyholders can optimally invest fractions of their wealth in tontines and annuities. In this context, the first two research questions of this thesis are

1. How can we design new retirement products which lead to a better risk sharing between policyholders and insurers than tontines, where policyholders carry most of the longevity risk, and annuities, where insurers carry all the longevity risk?

2. Can tontines still be a beneficial supplementary product in addition to annuities for some policyholders, especially in the presence of risk loadings? How attractive is the newly introduced tonuity compared to a portfolio consisting of an annuity and a tontine?

In the first paper, we come up with a novel approach of combining annuities and tontines by forming a tontine with minimum guaranteed payments. We follow Donnelly and Young (2017) on the design of the product and extend their findings by showing that a Milevsky and Salisbury (2015)-tontine with minimum guarantee can be attractive to both policyholders and insurers. In the second paper, we aim to further analyze the attractiveness of the innovative retirement plan tonuity introduced in Chen et al. (2019). In particular, we want to compare it to two additional combinations of annuities and tontines: One which we call antine, working like a reversed tonuity, and a portfolio consisting of a tontine and an annuity, where the individual may choose her own payoff structure of both products.

The third research question, which is still in the context of tontines, is motivated by a comment made by Adam Smith, who pointed out already back in 1776 that an over- or underestimation of one's life expectancy affects the perceived attractiveness of a retirement product (cf. Smith (1776)). Today, this question is still highly relevant. For example, Wu et al. (2015) point out that annuities look overpriced for an investor who is pessimistic about her longevity. Milevsky and Salisbury (2015) also briefly mention that subjective mortality beliefs might impact the willingness to purchase tontines. This brings us to the third research question:

3. Given the main result in Milevsky and Salisbury (2015), that an annuity always yields a higher expected lifetime utility level than a tontine if both products are fairly priced, is it possible that some individuals prefer a tontine over an annuity under subjective mortality beliefs? And if so, what are the main driving factors behind this result?

This question is answered in the third research paper: We find that, indeed, tontines can be preferred over annuities under subjective mortality beliefs. Surprisingly, the main driving factor behind the tontine's relative attractiveness is not that an individual believes that she lives longer than her peers, but instead that she underestimates the remaining lifetimes of the other policyholders relative to the remaining lifetimes assumed by the insurer.

Another development in the light of the recent societal challenges, apart from the resurrection of tontines, is the question how occupational pension schemes shall be designed. In industrialized countries, defined benefit (DB) and defined contribution (DC) pension schemes are still the two main types of occupational retirement plans. In a DB scheme, the sponsoring companies basically promise their employees a guaranteed pension payment. In a DC scheme, on the other hand, sponsoring companies, and often also their employees, pay deterministic contributions to an external pension fund where the capital is invested in financial assets. The benefit at retirement therefore depends on the performance of the investment returns experienced during the membership, i.e. the market risk is carried completely by the employees instead of the employers. In the last years, in most developed countries there has been a shift from DB towards DC plans, as guarantees are difficult to maintain given the current societal challenges. Broeders and Chen (2010) name, among further causes, also the changing pension regulation and the asset-liability-mismatch risk as reasons behind this shift. Further factors responsible for this conversion can also be found, for example, in Aaronson and Coronado (2005). Naturally, the question will be asked whether giving up deterministic guarantees is the correct reform implemented in occupational pension schemes, whether such a drastic risk transfer from the employers to the employees might worsen the benefits of the employees too substantially, and whether hybrid plans combining the advantages of DC and DB plans are the future of occupational pension schemes. A simple example of such a hybrid plan is a DC pension plan with minimum guarantees that ensures that the employees obtain a certain benefit when they retire. In Germany, for instance, such guarantees used to be prescribed by law in all pension schemes, and “pure” DC-like schemes, without any guarantees, did not exist at all until 2018 when the “Betriebsrentenstärkungsgesetz” came into effect and the new occupational pension scheme “Zielrente” was introduced.

It is common that individual pension plan members contribute to a collectively organized pension fund instead of handling the investments of their contributions on their own. Such collectively administered pension funds often cannot reflect each plan member's risk preferences accordingly (see, for example, Alserda et al. (2019) and Frijns (2010)). Modeling the individual plan members with heterogeneous risk preferences as investors in a financial market who are joint together and invest collectively allows us to analyze the following two research questions:

4. How do guarantees, sharing rules and management fees affect the retirement benefits of investors with heterogeneous risk preferences who are tied together in their investment

decisions? What consequences do the results imply in the context of occupational pension schemes? How shall minimum guarantees be prescribed by law in these schemes? Can we come up with a better guarantee design than the prescribed fixed guarantees?

5. How do the results change if the need for guarantees is directly modeled in the utility function? How do the results depend on the financial market assumed? How is each individual affected if a more realistic market with stochastic instead of constant volatility is considered?

In the fourth paper, the fourth research question is tackled. We consider a collective of individuals with heterogeneous risk preferences who are tied together in their investment decisions and delegate a fund manager to invest their initial wealth in a Black-Scholes financial market, under portfolio insurance. Concerning the portfolio insurance constraint, we consider two cases: a deterministic and a flexible state-dependent guarantee. We include management fees and assume that the investors in the collective can be divided into two groups, each with a different willingness to pay the fee. Finally, the fifth paper answers the last research question. In this paper, we set ourselves in a more realistic financial market with stochastic volatility in the sense of Heston (1993). Additionally, we no longer incorporate portfolio insurance as a constraint but instead directly include the demand for guarantees in the utility functions of the individuals under consideration.

The following section provides a detailed summary of each of the five research papers.

3 Summary of Research Papers

Research Paper 1: Options on tontines: An innovative way of combining annuities and tontines

In this paper, we follow Donnelly and Young (2017) and present an innovative way of combining annuities and tontines by designing a tontine with a minimum guarantee. While Donnelly and Young (2017) extend the product design in Donnelly et al. (2014) by a guaranteed benefit, we extend the tontine design from Milevsky and Salisbury (2015). Further extending Donnelly and Young (2017), who analyze the fair price of the guarantee from the policyholder's perspective, we analyze the attractiveness of this new product both from the policyholder's and from the insurer's perspective. We compare it to conventional annuities and tontines (with no guarantees). The paper is joint with An Chen and is published in *Insurance: Mathematics and Economics*.

As the title of this paper suggests, the new product involves options on tontines. We design the product in such a way that the payoff at each time is given by a guaranteed retirement income and a call option on a tontine multiplied by a surplus participation rate which lies between 0 and 1. The main goal of this design is to achieve a better risk-sharing between the policyholders and the insurers. Following standard actuarial techniques (see, for example, Olivieri and Pitacco (2011)), we assume that the insurer uses a risk-neutral pricing measure to price annuities, tontines and the newly formed product. By choosing a risk-neutral pricing measure, a safety loading is implicitly included in the premium of each retirement product. After setting the initial contract value of tontines with minimum guarantees equal to the initial investment of the policyholder, we are able to determine the fair participation rate for a given guarantee, tontine payment and initial investment. As expected, a high guarantee level implies a low fair surplus participation rate, which consequently results in a payoff close to an annuity. A low guarantee, on the other hand, leads to a high surplus participation rate and a more volatile payoff. Therefore, by varying the guarantee and the resulting fair surplus participation rate, the insurer is able to provide not only one but a whole range of products to policyholders with different risk aversions. The fairness condition provides a reasonable foundation to compare various tontines with minimum guarantees.

To determine the attractiveness of the new product from a policyholder's perspective, we consider an expected utility framework, as it is common in this stream of literature. For a given tontine payoff, we determine the utility-maximizing amount of guarantee for a policyholder along with the corresponding fair surplus participation rate. As our results show, a natural tontine as introduced in Milevsky and Salisbury (2015) with a rather low minimum guarantee can yield the highest expected utility level to the majority of risk-averse individuals under reasonable parameter choices. The natural tontine is designed in such a way that its payoff to a single policyholder remains constant over time if deaths in the pool occur exactly as expected. In addition to serving different risk appetites, our product also manages to serve different liquidity needs of policyholders. As an example, we determine the optimal contracts for individuals whose liquidity needs remain constant over time and whose liquidity needs increase with age.

To examine the benefits of the insurer from selling tontines with minimum guarantee, we are inspired, for example, by Li and Hardy (2011) and Olivieri and Pitacco (2019) and focus on the present value of net losses (or present value of unexpected cash flows). In particular, insurers are modeled differently than policyholders as we do not assume expected utility preferences for the insurers. Based on the present value of future losses, we determine and compare the loss probability and the conditional expected loss faced by three different insurance companies, where each of the three sells exclusively one of the three products: annuity, tontine (with no guarantee) and the new tontine with minimum guarantee. Our results show that a natural tontine with minimum guarantee can outperform a conventional annuity in the expected conditional loss, while yielding an (almost) identical loss probability. It indicates that tontines with minimum guarantees can be a desirable product from an insurer's perspective.

In total, the first paper yields one possible answer to the first research question. Of course, there exist many ways to combine annuities and tontines to achieve a better risk sharing between policyholders and insurers, as can already be observed in the literature provided above. The combination proposed in the first paper is new to the literature and the analyses carried out directly show that the product can be more attractive than annuities and tontines (with no guarantees) to the policyholders and more attractive than annuities to insurers. Therefore, it has the desirable feature of leading to a better risk sharing between policyholders and insurers.

Research Paper 2: On the optimal combination of annuities and tontines

In the second paper, we present another way of combining annuities and tontines, a product which we call antine. The introduction of this new retirement plan is, however, only a small novelty in this paper, as the antine is constructed in a very similar way as the tonuity introduced in Chen et al. (2019): While the tonuity starts with tontine-like payments at early retirement ages and switches to annuity-like payments at older ages, the antine works exactly the other way around (annuity first, tontine afterwards). The main novelty in this paper is the comparison of the relative attractiveness of the tonuity, the antine and a portfolio consisting of a conventional tontine and annuity under safety loadings. This paper is joint work with An Chen and Thorsten Sehner and is published in the *ASTIN Bulletin: The Journal of the International Actuarial Association*.

In this article, we again include risk capital charges in the premium calculation and compare the three retirement plans tonuity, antine and the portfolio (consisting of a tontine and an annuity) from the perspective of a policyholder with constant relative risk aversion (CRRA) and no bequest motive. Unlike the first paper, this paper solely considers the policyholder's perspective and not the insurer's. Safety loadings are taken into account by using the expected value premium principle for premium calculations. The results obtained in this article could, however, be extended to other premium principles (for instance, the famous variance or standard deviation principles). Extending Chen et al. (2019), we explicitly derive the utility-maximizing payoff of the tonuity and the antine under safety loadings. The utility-maximizing payoffs of the annuity and tontine in the portfolio can only be determined numerically. Our numerical analyses show that the optimal payoffs of the tontine and the annuity are carefully chosen in such a way that the annuity payments increase at ages where the tontine income is most volatile, i.e. the payment streams of the two products supplement each other. By allowing the individual to choose the payoff structure of the annuity and tontine in the portfolio freely, our article differs substantially from Weinert and Gründl (2017) who maximize the utility over the amount of the initial wealth invested in a given tontine. An additional difference to Weinert and

Gründl (2017) is that the tontine they consider is the Sabin (2010) tontine, while we focus on the Milevsky and Salisbury (2015) tontine. As the two payoffs are chosen jointly, the resulting optimal annuity payoff differs substantially from optimal annuity payoffs in the literature (cf. Yaari (1965), Milevsky and Huang (2011), Milevsky and Salisbury (2015)). The optimal payoff of the tontine in the portfolio, on the other hand, coincides roughly with optimal tontine designs discussed in the literature (cf. Milevsky and Salisbury (2015) and Chen et al. (2019)). Having determined the optimal income streams in the portfolio, we can also implicitly determine the fractions of wealth initially invested in the annuity and the tontine, respectively.

We show that the expected lifetime utility generated by the optimal portfolio is always at least as high as that of the optimal tonuity or antine. This result can be proven by showing that any tonuity and antine payoff can be replicated by a policyholder who invests simultaneously in an annuity and a tontine and chooses their payoff structures accordingly. We further derive theoretical assumptions regarding the loadings of the tontine and the annuity under which a pure investment in either one of them is optimal and under which a simultaneous investment in both is optimal. If the annuity loading is smaller than or equal to the tontine loading, investing 100% in the annuity is optimal. If the annuity loading drastically exceeds the tontine loading, investing 100% in the tontine is optimal. By "drastically", we mean an unrealistically high annuity loading of potentially more than 100%, depending on the actual parameters used. That is, under realistic assumptions on the loadings, a simultaneous investment in both annuity and tontine is optimal. In our numerical analyses, we find that the expected lifetime utility obtained from the tonuity gets fairly close to that of the portfolio. Additionally, we observe that the tonuity is likely to deliver a higher expected lifetime utility than the antine. In fact, a nontrivial antine (that is, an antine which is not an annuity or a tontine) is already outperformed by an annuity or a tontine in our parameter setup.

All in all, the second paper delivers a partial answer to the first and a full answer to the second research question. First, we see that, in the presence of risk loadings, a tontine can be a beneficial supplement to annuities. Additionally, this paper confirms again that the tonuity can be an attractive retirement product from a policyholder's perspective. Since the optimal payoffs of the annuity and the tontine in the portfolio are rather complicated, the tonuity could provide a simpler alternative to many policyholders as it only requires one switch from tontine to annuity and delivers nearly the same level of expected utility as the optimal portfolio.

Research Paper 3: Optimal retirement products under subjective mortality beliefs

In the third research paper, we aim to find out whether the result in Milevsky and Salisbury (2015) (annuities yield a higher expected lifetime utility than tontines under actuarially fair premiums) still holds if we include subjective mortality beliefs in the model. By subjective mortality beliefs, we mean a systematic over-or underestimation of one's lifetime compared to the survival curves the insurer uses for pricing. We include two sources of subjective mortality beliefs from the perspective of a single policyholder: (i) Subjective mortality beliefs for herself, regarding her own remaining lifetime, and (ii) subjective mortality beliefs concerning other policyholders in the tontine, that is, her peers' remaining lifetimes. This paper is submitted to a special issue of *Insurance: Mathematics and Economics* on behavioral insurance.

It is well-documented in the literature that individuals tend to have subjective beliefs about their life expectancy (see, for example, Hurd and McGarry (2002), O'Brien et al. (2005), Greenwald and Associates (2012), Bucher-Koenen et al. (2013), Elder (2013) and Wu et al. (2015)). An over- or underestimation of one's own and others' remaining lifetimes strongly affects the perceived attractiveness of a certain retirement product. Based on the empirical findings, there seems to be a clear tendency for people at younger ages to underestimate their life expectancy, while both under- and overestimations are documented at older ages in various studies. According to the literature review in Wu et al. (2015), there is also a clear tendency that women tend to underestimate their overall life expectancy more than men, and younger cohorts more than older cohorts.

Based on these diverging findings about subjective mortality beliefs at older ages, we allow for both over- and underestimation of one's own and others' survival curves relative to the survival curve used by the insurer, which is used as a benchmark. We follow Milevsky and Salisbury (2015) and Chen et al. (2019) and derive the (perceived) optimal payout functions of an annuity and a tontine under subjective mortality beliefs in an expected utility framework. Similarly as Milevsky and Salisbury (2015), we set ourselves in the actuarially fair pricing framework. In our model, we assume that there are three different survival curves for any x-year old in play: the one used by the insurer (denoted by ${}_tp_x$ for any $t \geq 0$), the one assumed by an individual for herself (denoted by ${}_t\tilde{p}_x$) and the one assumed by the individual for other policyholders (denoted by ${}_t\hat{p}_x$).

We prove that the expected lifetime utility of a tontine increases, the more a policyholder underestimates the survival curve of her peers, while an annuity is not affected by this underestimation. Further, we prove that there exists a critical pool size for which, once it is exceeded, the tontine is always preferred over the annuity under mild assumptions regarding the subjective mortality beliefs. The main driver behind this result is that the policyholder believes that her peers live shorter than the insurer has assumed. Numerically, we find that the subjective belief that a policyholder lives longer (${}_t\hat{p}_x < {}_t\tilde{p}_x$) or shorter (${}_t\hat{p}_x > {}_t\tilde{p}_x$) than her peers only plays a negligible role for this result. In the numerical analyses, we also show that the critical pool size for which the tontine is preferred might already be as small as 3. Additionally, we show that annuities can be perceived as overpriced (too expensive) from the perspective of a policyholder who underestimates her life expectancy, lowering the resulting expected lifetime utility. A tontine's perceived price, on the other hand, is hardly affected by subjective beliefs concerning the policyholder's subjective beliefs for herself.

Thus, the third paper delivers an answer to the third research question: It is, in fact, possible that tontines generate a higher expected lifetime utility than annuities, even if both products are fairly priced. The main assumption for this result is an underestimation of the survival probabilities of one's peers compared to the actuarial bases applied by the insurer. Subjective mortality beliefs have recently been considered as a possible explanation for the annuity puzzle (see, for example, Poppe-Yanez (2017), Caliendo et al. (2017) and O'Dea et al. (2019)). Our article contributes to this literature by the inclusion of tontines and a comparison of annuities and tontines under subjective mortality beliefs.

Research Paper 4: Optimal collective investment: The impact of sharing rules, management fees and guarantees

In the fourth research paper, we consider a collective of investors in a financial market who are tied together in their investment decision to invest collectively in the financial market. The investors can be seen as employees who, in an occupational pension context, delegate a fund manager to invest their total wealth collectively. As the title of the article suggests, we include different designs of portfolio insurance constraints and management fees in the optimization problem and analyze different sharing rules. By sharing rule, we mean the fraction of the (state-dependent) terminal wealth that a single investor obtains from the collective investment at maturity. The fourth article is a joint project with An Chen and Thai Nguyen and has been revised and resubmitted to the *Journal of Banking and Finance*.

To assess the effects of management fees, we assume that there are two types of investors in the collective: Group 1 has access to the complete and arbitrage-free market and, thus, investors in this group can invest on their own. They could still prefer fund delegation over investing on their own because asset management costs time and energy (cf. Kim et al. (2016)). Group 2 has limited access to the financial market and relies on fund delegation. Therefore, investors in Group 1 are ready to pay a lower fee for fund delegation than those in Group 2. Assuming that the fund manager charges an average fee, in our analysis, investors in Group 1 receive less than what they are entitled to and those in Group 2 more than they are entitled to.

Additionally, each of the individuals in the collective may demand a certain guarantee. The fund manager then sets up a collective investment strategy such that all these individual guarantees are met. Inspired by Dumas (1989), Karatzas et al. (1990), Xia (2004), Pazdera et al. (2016) and Branger et al. (2018), we assume that the fund manager uses a collective utility function which is given by a weighted sum of the individual utility functions and carries a maximum operator which implies that the fund manager aims to achieve the highest utility level for a given vector of weights. The fund manager then maximizes this utility function under two types of portfolio insurance constraints: In the first case, the optimal terminal wealth needs to meet a deterministic guaranteed payment which is already known from the beginning. In the other case, we advocate a flexible guarantee framework, that is, the optimal terminal wealth needs to meet a flexible guarantee payment which could be, for example, a market index. In our case, the flexible guarantee consists of a (smaller) deterministic guaranteed payment known from the beginning and a state-dependent payment becoming known at maturity, which is chosen as a fraction of the optimal terminal wealth resulting from the individual optimization problem. Having determined the collective optimal terminal wealth for both types of portfolio insurance schemes, we also derive the optimal dynamic strategy explicitly.

Once the optimal terminal wealth and trading strategy are obtained, we use a state-dependent sharing rule to return to each investor her share of the total terminal payoff. The sharing rule is designed state-dependently because the main objective of the fund manager is to meet the individual guarantee levels required by each person. Additionally, three sharing rules are considered to distribute the bonus exceeding the collective guarantee. One of these takes into account individual guarantee and wealth levels and is chosen to satisfy the concept of financial fairness (see, for example, Bühlmann and Jewell (1979), Schumacher (2018), Boelaars and Broeders (2019) and Orozco-Garcia and Schmeiser (2019)). The second one only accounts for the initial wealth levels and is used, for example, in Jensen and Nielsen (2016), and the last one only takes account of the pool size.

In the deterministic guarantee framework, we find that all investors in Group 1 suffer a loss through fund delegation. In Group 2, on the other hand, there can be a few investors that benefit from the collective investment. These are the ones with a rather high risk aversion, since the guarantee imposes too drastic losses on investors with low risk aversion. As one would expect, the first of the sharing rules mentioned above is the fairest one. Under the second and third one, investors who demand low guarantees finance the relatively higher guarantees of the remaining investors. Under the third sharing rule, investors with low initial wealth additionally benefit from those with a higher wealth. Under the state-dependent guarantee framework, we find the following: In Group 1, still all the investors suffer a loss through fund delegation. These are, however, lower than in the deterministic guarantee framework, especially for investors with a low risk aversion. In this sense, a state-dependent guarantee which depends on the market performance serves each investor with a different risk aversion in a better way than a deterministic guarantee. In Group 2, there can still be investors who benefit from the collective investment. Regarding the sharing rules, the third sharing rule performs worst for Group 1 and best for Group 2 while the first and second yield (almost) identical results.

In total, the fourth research paper provides answers to the fourth research question. It shows that deterministic guarantees embedded in DC schemes deteriorate the benefits of employees and that a more flexible guarantee can lead to higher expected utility levels than the frequently prescribed deterministic guarantees. Regarding the management fees, we find that even in Group 2 not all the members benefit when receiving more than they are entitled to because the guarantee constraint imposes too drastic utility losses.

Research Paper 5: A collective investment problem in a stochastic volatility environment: The impact of sharing rules

In the fifth research paper, we set ourselves in a similar setting as in the fourth paper, i.e. we again consider a collective of investors who can be seen as employees in a collectively administered pension fund. The most important differences to the fourth paper are that we consider a more realistic financial market setting allowing for a stochastic volatility of the stock and that guarantees are no longer a constraint to the optimization problem but instead directly taken into account in the utility functions. In this setting, we want to figure out whether it is possible to achieve individual optimal solutions in a collectively administrated fund. Further, we analyze the effects of the stochastic volatility in comparison to the Black-Scholes model since the assumption of a constant volatility is a well-known weakness of the Black-Scholes model. In particular, we analyze different sharing rules under a Black-Scholes and a Heston model and compare the results obtained in both markets. This paper is again joint with An Chen and Thai Nguyen and has been submitted to the *Annals of Operations Research*.

Similarly as in the fourth paper, each of the individuals in the collective may demand a certain subsistence level / guarantee. The fund manager's investment strategy again aims at meeting all these individual guarantee levels. Due to the guarantees embedded in the individual utility functions, the utility function used by the fund manager can be considered as a generalization of the collective utility function used in the fourth research paper. Extending Branger et al. (2018), we consider two financial market models: We briefly start with the classical Black-Scholes model and then extend the analysis to a more general model where the volatility of the stock is itself a stochastic process driven by a Brownian motion. To be precise, we model the stochastic volatility process in the sense of Heston (1993), that is, we use a square-root process as in the interest rate model in Cox et al. (1985).

Since the stochastic volatility process involves risks that are not traded, the financial market is incomplete. This makes the optimization problem in the stochastic volatility market more complex than in the constant volatility case. Optimization problems in general incomplete markets including stochastic volatility models have been considered extensively in the literature. For common utility functions (power utility in particular) the solution is available in closed form using dynamic programming (see, for example, Pham (2002), Fleming and Hernández-Hernández (2003), Chacko and Viceira (2005), Kraft (2005) and Liu (2006)). In our collective setting, this approach seems not to be possible, which is why we (artificially) complete the financial market using derivatives. This approach is also well-documented in the literature and applied, for instance, in Liu and Pan (2003), Branger et al. (2008, 2017), Escobar et al. (2018) and Chen et al. (2018). Due to the (artificial) market completeness, we can determine the optimal terminal wealth levels and the dynamic trading strategies explicitly in both financial markets using the static martingale approach (see, for example, Cox and Huang (1989)).

In both financial markets, we show that it is possible that each individual in the collective receives her individual optimal terminal wealth under the use of a state-dependent sharing rule and the imposition of financial fairness (defined similarly as in the fourth paper). Our comparison of sharing rules shows that losses occur to all investors in the collective if a financially fair linear sharing rule is applied. If the linear sharing rule does not fulfill the fairness condition, some individuals in the collective are better off, but the majority of investors is worse off than in the individual optimization problem. Our comparison between the constant and stochastic volatility framework reveals that all individuals are worse off in the stochastic volatility model. That is, all the losses are larger and all the gains are smaller compared to the constant volatility model.

To make a long story short, this paper answers the fifth research question. Under the different modeling approach for the guarantees, we see that individual optima are achievable in the collective, given that a financial fairness criterion and a state-dependent sharing rule are used. That is, in contrast to the fourth paper, this paper's message is that guarantees do not necessarily deteriorate the benefits of the employees and that it is possible to achieve an investment strategy and a sharing rule under which no losses occur, even with guarantees. Under more practical linear sharing rules, a market with stochastic volatility leads to (on average) higher losses for each individual in the collective, since the stochastic volatility market involves more risks than the classical Black-Scholes market with constant volatility. When using linear sharing rules, the financial fairness criterion can still account for a fair distribution of the terminal wealth, although individual optimal solutions are no longer available. Under linear sharing rules which do not reflect a financial fairness criterion, some investors benefit at the cost of others.

References

Aaronson, S. and Coronado, J. (2005). Are firms or workers behind the shift away from DB pension plan? *Available at SSRN:* `https://ssrn.com/abstract=716383`.

Alserda, G. A., Dellaert, B. G., Swinkels, L., and van der Lecq, F. S. (2019). Individual pension risk preference elicitation and collective asset allocation with heterogeneity. *Journal of Banking & Finance*, 101:206–225.

Basak, S. (1995). A general equilibrium model of portfolio insurance. *The Review of Financial Studies*, 8(4):1059–1090.

Boelaars, I. and Broeders, D. (2019). Fair pensions. *Available at SSRN:* `https://ssrn.com/abstract=3374456`.

Branger, N., Chen, A., Mahayni, A., and Nguyen, T. (2018). Optimal collective investment. *Working paper. Available at* `https://www.researchgate.net/publication/324910837`.

Branger, N., Muck, M., Seifried, F. T., and Weisheit, S. (2017). Optimal portfolios when variances and covariances can jump. *Journal of Economic Dynamics and Control*, 85:59–89.

Branger, N., Schlag, C., and Schneider, E. (2008). Optimal portfolios when volatility can jump. *Journal of Banking & Finance*, 32(6):1087–1097.

Broeders, D. and Chen, A. (2010). Pension regulation and the market value of pension liabilities: A contingent claims analysis using parisian options. *Journal of Banking & Finance*, 34(6):1201–1214.

Browne, S. (1999). Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark. *Finance & Stochastics*, 3(3):275–294.

Bucher-Koenen, T., Kluth, S., et al. (2013). Subjective life expectancy and private pensions. In *Annual Conference 2013 (Duesseldorf): Competition Policy and Regulation in a Global Economic Order*, number 79806. Verein für Socialpolitik/German Economic Association.

Bühlmann, H. and Jewell, W. S. (1979). Optimal risk exchanges. *ASTIN Bulletin: The Journal of the IAA*, 10(3):243–262.

Caliendo, F. N., Gorry, A., and Slavov, S. (2017). Survival ambiguity and welfare. Technical report, National Bureau of Economic Research.

Chacko, G. and Viceira, L. M. (2005). Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. *The Review of Financial Studies*, 18(4):1369–1402.

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

Chen, A., Nguyen, T., and Stadje, M. (2018). Optimal investment under VaR-regulation and minimum insurance. *Insurance: Mathematics & Economics*, 79:194–209.

Cox, J. C. and Huang, C.-F. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–408.

Davidoff, T., Brown, J. R., and Diamond, P. A. (2005). Annuities and individual welfare. *American Economic Review*, 95(5):1573–1590.

Deelstra, G., Grasselli, M., and Koehl, P.-F. (2004). Optimal design of the guarantee for defined contribution funds. *Journal of Economic Dynamics and Control*, 28(11):2239–2260.

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

Donnelly, C., Guillén, M., and Nielsen, J. P. (2013). Exchanging uncertain mortality for a cost. *Insurance: Mathematics and Economics*, 52(1):65–76.

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

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

Dumas, B. (1989). Two-person dynamic equilibrium in the capital market. *The Review of Financial Studies*, 2(2):157–188.

El Karoui, N., Jeanblanc, M., and Lacoste, V. (2005). Optimal portfolio management with American capital guarantee. *Journal of Economic Dynamics and Control*, 29(3):449–468.

Elder, T. E. (2013). The predictive validity of subjective mortality expectations: Evidence from the health and retirement study. *Demography*, 50(2):569–589.

Escobar, M., Ferrando, S., and Rubtsov, A. (2018). Dynamic derivative strategies with stochastic interest rates and model uncertainty. *Journal of Economic Dynamics and Control*, 86:49–71.

Fleming, W. H. and Hernández-Hernández, D. (2003). An optimal consumption model with stochastic volatility. *Finance and Stochastics*, 7(2):245–262.

Frijns, J. M. (2010). Dutch pension funds: Aging giants suffering and inconsistent risk management. *Rotman International Journal of Pension Management*, 3(2):16–21.

Gabih, A., Sass, J., and Wunderlich, R. (2009). Utility maximization under bounded expected loss. *Stochastic Models*, 25(3):375–407.

Greenwald and Associates (2012). 2011 risks and process of retirement rurvey report of findings. *Society of Actuaries. Prepared by Mathew Greenwald and Associates, Inc., Employee Benefit Research Institute.*

Grossman, S. J. and Vila, J.-L. (1989). Portfolio insurance in complete markets: A note. *Journal of Business*, 62(4):473–476.

Grossman, S. J. and Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3):241–276.

Grossman, S. J. and Zhou, Z. (1996). Equilibrium analysis of portfolio insurance. *The Journal of Finance*, 51(4):1379–1403.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Huang, H., Milevsky, M. A., and Salisbury, T. S. (2017). Retirement spending and biological age. *Journal of Economic Dynamics and Control*, 84:58–76.

Hurd, M. D. and McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, 112(482):966–985.

Jensen, B. A. and Nielsen, J. A. (2016). How suboptimal are linear sharing rules? *Annals of Finance*, 12(2):221–243.

Jensen, B. A. and Sørensen, C. (2001). Paying for minimum interest rate guarantees: Who should compensate who? *European Financial Management*, 7(2):183–211.

Karatzas, I., Lehoczky, J. P., and Shreve, S. E. (1990). Existence and uniqueness of multi-agent equilibrium in a stochastic, dynamic consumption/investment model. *Mathematics of Operations Research*, 15(1):80–128.

Kim, H. H., Maurer, R., and Mitchell, O. S. (2016). Time is money: Rational life cycle inertia and the delegation of investment management. *Journal of Financial Economics*, 121(2):427–447.

Korn, R. (2014). *Moderne Finanzmathematik–Theorie und praktische Anwendung: Band 1–Optionsbewertung und Portfolio-Optimierung*. Springer Fachmedien Wiesbaden.

Kraft, H. (2005). Optimal portfolios and Heston's stochastic volatility model: an explicit solution for power utility. *Quantitative Finance*, 5(3):303–313.

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

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

Liu, J. (2006). Portfolio selection in stochastic environments. *The Review of Financial Studies*, 20(1):1–39.

Liu, J. and Pan, J. (2003). Dynamic derivative strategies. *Journal of Financial Economics*, 69(3):401–430.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Milevsky, M. A. (2014). Portfolio choice and longevity risk in the late seventeenth century: a re-examination of the first english tontine. *Financial History Review*, 21(3):225–258.

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

Milevsky, M. A. and Huang, H. (2011). Spending retirement on planet vulcan: The impact of longevity risk aversion on optimal withdrawal rates. *Financial Analysts Journal*, 67(2):45–58.

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

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

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

O'Brien, C., Fenn, P., and Diacon, S. (2005). How long do people expect to live? Results and implications. CRIS Research report 2005–1.

O'Dea, C., Sturrock, D., et al. (2019). Survival pessimism and the demand for annuities. Technical report, Institute for Fiscal Studies.

OECD (2016). *OECD Pensions Outlook 2016*. OECD Publishing, Paris. Available at `http://dx.doi.org/10.1787/pens_outlook-2016-en`.

Olivieri, A. and Pitacco, E. (2011). *Introduction to Insurance Mathematics: Technical and Financial Features of Risk Transfers*. Springer International Publishing Switzerland.

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

Orozco-Garcia, C. and Schmeiser, H. (2019). Is fair pricing possible? An analysis of participating life insurance portfolios. *Journal of Risk and Insurance*, 86(2):521–560.

Pazdera, J., Schumacher, J. M., and Werker, B. J. (2016). Cooperative investment in incomplete markets under financial fairness. *Insurance: Mathematics and Economics*, 71:394–406.

Peijnenburg, K., Nijman, T., and Werker, B. J. (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, 70:18–35.

Pham, H. (2002). Smooth solutions to optimal investment models with stochastic volatilities and portfolio constraints. *Applied Mathematics & Optimization*, 46(1):55–78.

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

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

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

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

Schumacher, J. M. (2018). Linear versus nonlinear allocation rules in risk sharing under financial fairness. *ASTIN Bulletin: The Journal of the IAA*, 48(3):995–1024.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell, London.

Stamos, M. Z. (2008). Optimal consumption and portfolio choice for pooled annuity funds. *Insurance: Mathematics and Economics*, 43(1):56–68.

Tepla, L. (2001). Optimal investment with minimum performance constraints. *Journal of Economic Dynamics and Control*, 25(10):1629–1645.

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

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

Wu, S., Stevens, R., and Thorp, S. (2015). Cohort and target age effects on subjective survival probabilities: Implications for models of the retirement phase. *Journal of Economic Dynamics and Control*, 55:39–56.

Xia, J. (2004). Multi-agent investment in incomplete markets. *Finance and Stochastics*, 8(2):241–259.

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

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

Research Papers

1 Options on tontines: An innovative way of combining annuities and tontines

Source:
This article was published in *Insurance: Mathematics and Economics*, 89, Chen, A. and Rach, M., Options on tontines: An innovative way of combining annuities and tontines, 189–192,

URL: https://doi.org/10.1016/j.insmatheco.2019.10.004

DOI: 10.1016/j.insmatheco.2019.10.004

Options on tontines: An innovative way of combining tontines and annuities

An Chen* and Manuel Rach*

* Institute of Insurance Science, Ulm University, Helmholtzstr. 20, 89069 Ulm, Germany.
E–mails: an.chen@uni-ulm.de; manuel.rach@uni-ulm.de

Abstract

Increases in the life expectancy, the low interest rate environment and the tightening solvency regulation have led to the rebirth of tontines. Compared to annuities, where insurers bear all the longevity risk, policyholders bear most of the longevity risk in a tontine. Following Donnelly and Young (2017), we come up with an innovative retirement product which contains the annuity and the tontine as special cases: a tontine with a minimum guaranteed payment. The payoff of this product consists of a guaranteed payoff and a call option written on a tontine. Extending Donnelly and Young (2017), we consider the tontine design described in Milevsky and Salisbury (2015) for designing the new product and find that it is able to achieve a better risk sharing between policyholders and insurers than annuities and tontines. For the majority of risk-averse policyholders, the new product can generate a higher expected lifetime utility than annuities and tontines. For the insurer, the new product is able to reduce the (conditional) expected loss drastically compared to an annuity, while the loss probability remains fairly the same. In addition, by varying the guaranteed payments, the insurer is able to provide a variety of products to policyholders with different degrees of risk aversion and liquidity needs.

Keywords: Annuity, tontine, option pricing, optimal retirement products, net loss analysis

JEL: G22, G13

1 Introduction

Annuities provide a life-long payment stream to policyholders and thus, yield efficient protection against longevity risk. Therefore, they are thought to be very desirable retirement products from a policyholder's perspective (Yaari (1965), Mitchell (2002), Davidoff et al. (2005), Peijnenburg et al. (2016)). In practice, however, annuitization rates remain low (see for instance Hu and Scott (2007), Brown (2007), Benartzi et al. (2011), Milevsky (2013) and the references therein). The tightening solvency regulation, the ongoing low interest rate environment, and the increases in life expectancies in OECD countries (cf. OECD (2016))[1] might drive up the price of annuities further and, consequently, annuitization rates are not going to increase in the near future. Insurers and customers are searching for new, more attractive retirement products. Products which have attracted vast attention from academics and practitioners in this context are tontines and pooled annuity funds. For tontines, we refer the interested reader, for instance, to Sabin (2010), Milevsky (2015), Milevsky and Salisbury (2015, 2016) and Li and Rothschild (2019), for pooled annuity funds to Piggott et al. (2005), Valdez et al. (2006), Stamos (2008), Qiao and Sherris (2013), Donnelly et al. (2014), Donnelly (2015) and Donnelly and Young (2017). In these products, in contrast to annuities, a pool of policyholders shares the longevity risk. Although tontines provide an attractive alternative to annuities, particularly when it comes to longevity risk sharing, they might leave the policyholders with rather volatile payments at advanced retirement ages. If many policyholders in the pool live very long, the payments of the tontine might not be sufficient to provide sustainable retirement income at old ages. All in all, both annuities and tontines are well-formed sources of retirement income which both have their own advantages and disadvantages.

As a direct transition from annuities to tontines might be a too drastic step for many policyholders and insurers, we are searching for attractive alternatives for both parties. Naturally, the question arises whether the advantages of annuities and tontines can be combined to form a product which is cheaper than annuities and shifts the longevity risk not completely, but only partially towards the policyholder. Can we come up with an innovative retirement product which leads to a better risk sharing between the policyholders and the insurers than annuities, where insurers bear all the longevity risk, and than tontines, where policyholders bear most of the longevity risks? The newly introduced "tonuity" in Chen et al. (2019), whose payment starts with a tontine-like payment and switches to an annuity-like payment after a fixed switching time, can be considered as an attempt in this direction. Following the idea of Donnelly and Young (2017) who came up with an innovative retirement product providing a

[1]Note that although the life expectancy tends to increase across OECD countries, not all developed countries show such a trend. For example, in the United States, life expectancies have decreased in recent years (Murphy et al. (2018)) and the United Kingdom has experienced a drastic slowdown in life expectancy improvements from 2011 on (Evans (2018)).

minimum guaranteed payment in a mortality risk-sharing scheme, we introduce a similar class of products which we call tontines with minimum guarantees. The small novelty in our paper is that the options are written on the Milevsky and Salisbury (2015) tontine. The payoff of this product can be seen as a sequence of guaranteed annuity payments and call options on a tontine multiplied by a surplus participation rate. For each time, if the underlying tontine payment performs worse than the guaranteed level, the policyholder ends up with the guarantee payment, otherwise additionally endowed with a surplus participation.

We assume that the insurer uses risk-neutral pricing techniques to price annuities, tontines and the newly formed product involving options on tontines. By choosing the risk-neutral pricing measure prudently, a safety loading is implicitly included in the premium of each retirement product. Following standard actuarial techniques (see, for example, Olivieri and Pitacco (2011)), we set the initial contract value of tontines with minimum guarantees equal to the initial investment of the policyholder. Assuming the tontine payoff and the initial investment of the policyholder to be fixed, this enables us to determine the fair participation rate on the surplus for a given guarantee. As expected, a higher guarantee level implies a lower fair surplus participation rate, which consequently makes the payoff of the new product closer to an annuity. If, on the other hand, the guarantee is lowered, the surplus participation rate increases and the payoff becomes more volatile. The analysis of the fair combinations of the guarantee and participation rate shows that tontines with minimum guarantees allow the insurer to provide a rich variety of products to various policyholders. The fairness condition provides a reasonable foundation to compare various tontines with minimum guarantees. Note that Donnelly and Young (2017) also analyze fair valuation. However, the focuses of their and our paper do differ. We go beyond the fair valuation and further study (quantify) the attractiveness of new products, both from the policyholder's and the insurer's perspective. Fair valuation only serves as a constraint in the utility maximization problem of the policyholders.

To concretely examine the attractiveness of the new product for a given policyholder, we consider an expected utility framework. From a relatively new online book of William F. Sharpe (see Sharpe (2017)), it has been pointed out that using expected utility theory (advocated by traditional economists) is at least instructive to see whether the traditional approaches are helpful, also in the context of life insurance and retirement products. Expected utility has been frequently used in the literature on optimal retirement products to evaluate annuities (full guarantees) (see, e.g., Yaari's famous pioneering article Yaari (1965) or Yagi and Nishigaki (1993), Mitchell (2002), Davidoff et al. (2005) and Peijnenburg et al. (2016)), tontines (with no guarantees), and to *compare* these products by computing the lifetime utility resulting from holding these products (see, e.g., Milevsky and Salisbury (2015, 2016)).[2] Assuming constant

[2] For instance, in Milevsky and Salisbury (2015) and Chen et al. (2019), it is shown that annuities deliver a higher expected lifetime utility level than tontines when actuarially fair premiums are applied, while tontines

relative risk aversion (CRRA) preferences,[3] we determine the optimal fair combination of the guaranteed amount and the surplus participation rate for tontines with minimum guarantees which maximizes the expected lifetime utility of the considered policyholder. By allowing the individuals to choose their own optimal guarantee and surplus participation, the new product is able to serve policyholders with different risk aversion levels.

In our numerical analyses, we follow the suggestion made by Milevsky and Salisbury (2015) for an implementation of tontines in today's world and use a so-called *natural* tontine as the underlying for our new product. In a natural tontine, the payments from the insurer to the pool of policyholders decrease in exact proportion to the survival probabilities. In other words, its payoff to a single policyholder remains constant over time if deaths in the pool occur exactly as expected. An additional nice feature of this tontine design is that it is optimal for log-utility maximizers and nearly optimal for CRRA utility maximizers with relative risk aversions different from one (cf. Milevsky and Salisbury (2015)). As our results show, such a natural tontine with a rather low minimum guarantee can yield the highest expected utility level to the majority of risk-averse individuals under a reasonable set of parameter choices. In our parameter setup, the only exceptions are individuals with a rather low relative risk aversion, for whom a pure natural tontine (with no guarantee), is optimal. Our results are consistent with Milevsky and Salisbury (2015) and supplement it by the inclusion of a minimum guarantee in the payoff of natural tontines. In addition to serving policyholders with various risk aversion levels, the new product also manages to serve different liquidity needs of policyholders. For instance, we determine the optimal contracts for individuals who have increasing liquidity needs at more advanced retirement ages.

While Chen et al. (2019) analyze the attractiveness of tonuities solely from the policyholder's perspective, this article also aims to assess the benefits of the insurer from selling the new product. We are inspired, for example, by Bauer and Weber (2008), Li and Hardy (2011), Cairns (2013), Kling et al. (2014) and Olivieri and Pitacco (2019) and model insurers' benefits

can be preferred over annuities when appropriate safety loadings are taken into account. Note that this result is different from Døskeland and Nordahl (2008) in which life insurance contracts considered are exposed *solely to equity risk*. They find that the expected utility achieved from a product with a guarantee cannot be higher than a product with no guarantee. As also noted in Chen et al. (2015), the conclusion drawn from Døskeland and Nordahl (2008) does not hold generally, when other non-financial risks are included in the products.

[3] As pointed out by Sharpe (2017) and many others, power utility (preferences with constant relative risk aversion) is the most frequently used utility function to capture the preferences of individuals. For example, Levy (1994) mentions in the conclusion that "we find strong evidence for the DARA (decreasing absolute risk aversion) hypothesis, but the IRRA (increasing relative risk aversion) hypothesis is strongly rejected. Investors tend to show decreasing relative risk aversion (DRRA) or, at best, constant relative risk aversion (CRRA) but by no means IRRA. This evidence enhances Arrow's assertion that the DARA is observed in daily behavior of investors and that the IRRA has less intuitive evidence." In addition, in the book of Campbell and Viceira (2002), it is pointed out that the long-run behavior of the economy suggests that the long-run risk aversion cannot strongly depend on wealth, which motivates economically the use of the power utility. Note that of the literature on optimal retirement spending mentioned above, the majority also uses CRRA preferences.

differently than the policyholders' (not using expected utility). We use the (random) present value of future losses to examine the risks contained in retirement products. We then determine the loss probability and the conditional expected loss faced by three different insurers: One sells annuities exclusively, another one sells only pure tontines (with no guarantees), and the last one sells only tontines with minimum guarantees. Our results show that tontines with minimum guarantees can lead to a smaller loss probability than pure tontines and, on the other side, a drastically lower conditional expected loss than annuities. The latter result shows particularly that tontines with minimum guarantees can be a desirable complementary alternative retirement product to annuities. An insurer could introduce the new product additionally to the conventional annuity products to lower its potential losses.

The remainder of the article is structured as follows: Section 2 describes the basic model setup, in particular the assumptions regarding the mortality model and the design of annuities and tontines. Section 3 presents the new tontine with a minimum guarantee along with the definition of fairness. In Section 4, we examine the attractiveness of the new product from the policyholder's point of view, and in Section 5 from the insurer's point of view. Section 6 concludes the article.

2 Model setup

In this section, we describe the model setup used throughout the article. We follow, for example, Yaari (1965) and Milevsky and Salisbury (2015) and consider a continuous-time setting, extending the tontine design described in Milevsky and Salisbury (2015).

2.1 A simple stochastic mortality model

For simplicity, we ignore financial market risk in this article and solely focus on the mortality risk. We assume a stochastic mortality risk model which is rather simple, but allows us to distinguish two sources of mortality risk: Unsystematic (or idiosyncratic) and systematic (or aggregate) mortality risk. The unsystematic mortality risk stems from the fact that lifetimes of individuals are unknown but still follow some mortality law. This risk component can be diversified away through pooling, i.e., this risk tends to disappear for large enough portfolios. The systematic mortality risk stems from the fact that we are not able to determine the actual "true" mortality law with certainty. This risk component hits all policies in the same direction. In the context of retirement products, this component can be identified as the so-called longevity risk, that is, the risk of an overall unanticipated decline in mortality rates (see, for example, Pitacco et al. (2009)). When it is present, even with a large portfolio there is a residual part of risk that cannot be eliminated.

For any x-year-old policyholder, the best-estimate t-year survival probability is denoted by ${}_tp_x$ which can be computed from some continuous-time mortality law. We incorporate uncertainty in this mortality law in a similar way as, for example, Lin and Cox (2005) by applying a random shock ϵ to the best-estimate survival probabilities. Given the shock ϵ, the shocked survival curve is given by ${}_tp_x^{1-\epsilon}$. This shock covers the systematic mortality risk described above. We assume that ϵ is a continuous random variable taking values in $(-\infty, 1)$ almost surely whose density is denoted by $f_\epsilon(\cdot)$. The special case without longevity shock is simply obtained by setting $\epsilon = 0$.

2.2 Annuity and tontine

An **annuity** contract continuously provides a deterministic payment stream $\{c(t)\}_{t\geq 0}$ to a policyholder until death. We denote by ζ_ϵ the remaining future lifetime of the policyholder, with ϵ being the random longevity shock as defined in Section 2.1. The payment stream of the annuity can then be expressed as

$$b_A(t) := \mathbb{1}_{\{\zeta_\epsilon > t\}}\, c(t)\,. \tag{1}$$

While in an annuity, the longevity risk is borne by the insurance company, in a **tontine** contract it is shared among a homogeneous pool of $n \geq 1$ policyholders who are of the same age.[4] The unsystematic mortality risk can initially be diversified by a large enough pool size n. The systematic risk, on the other hand, cannot be diversified, as this type of risk affects all the policyholders in the same direction. At older ages, as the pool size decreases, the remaining policyholders are left with both systematic and unsystematic risk (see also Chen et al. (2019)). We follow the specific tontine design described by Milevsky and Salisbury (2015): Denoting by $N_\epsilon(t)$ the number of policyholders alive at time t, each policyholder receives $nd(t)/N_\epsilon(t)$, where $d(t)$ is a deterministic payment stream specified at the beginning of the contract. In total, this yields the following continuous payment stream to a policyholder for each $t \geq 0$:[5]

$$b_{OT}(t) := \mathbb{1}_{\{\zeta_\epsilon > t\}} \frac{nd(t)}{N_\epsilon(t)}\,. \tag{2}$$

In particular, as in an annuity product, the tontine provides an income for life, where the income provided by the tontine (2) depends on the number of survivors $N_\epsilon(t)$ and is, thus, more volatile than that of an annuity (1). Note that, conditional on the considered policyholder still being alive and given ϵ, the remaining number of individuals being alive at time $t > 0$

[4] That is, the insurer carries the longevity risk of the last survivor in the pool.

[5] For the case $N_\epsilon(t) = 0$ we define $b_{OT}(t) := 0$.

follows a binomial distribution, that is, $(N_\epsilon(t) - 1 \mid \zeta_\epsilon > t, \epsilon) \sim \text{Bin}\left(n-1, {}_tp_x^{1-\epsilon}\right)$, where we make use of the (conditional) independence of the pool members.

3 A tontine with a minimum guarantee

In the following, we present a new product which aims to combine the advantages of annuities and tontines. In many developed countries, individuals are not willing to give up guaranteed payments. For example, in all occupational pension schemes in Germany, a minimum guarantee used to be prescribed by law until very recently in 2018. Only from this year on, a "pure defined contribution"-like scheme with no guarantees for the beneficiaries, the so called "Zielrente", has existed in Germany. As the unstable payments of the tontine are one of its main disadvantages from a policyholder's point of view, we propose that the insurance company selling the tontine provide a minimum guarantee to the policyholders. Let $g(t) \geq 0$ for all $t \geq 0$ define the minimum level of guarantee required by the policyholder. A tontine with a minimum guarantee $g(t)$ delivers the payoff

$$b_{OTG}(t) := \mathbb{1}_{\{\zeta_\epsilon > t\}} \left(g(t) + \alpha \max\left\{ \frac{nd(t)}{N_\epsilon(t)} - g(t), 0 \right\} \right). \tag{3}$$

Note that this payoff depends substantially on the guarantee $g(t)$ and the surplus participation rate α. It contains the annuity and the tontine as the following boundary cases:

- $\alpha = 0$: In this case, equation (3) simplifies to an annuity with payoff $\mathbb{1}_{\{\zeta_\epsilon > t\}} g(t)$ to a single policyholder.
- $g(t) = 0$ for all $t \geq 0$: In this case, (3) simplifies to a pure tontine (with no guarantee) with payoff $\mathbb{1}_{\{\zeta_\epsilon > t\}} \frac{n\alpha d(t)}{N_\epsilon(t)}$ to a single policyholder. Note that α simply works like a rescaling parameter in this case. In this specific pure tontine, the predetermined withdrawal plan is described by $(\alpha d(t))$. The reason why we include α in the payoff (3) is that it resembles the common feature of a surplus participation in life and pension insurance.

In the following, we only consider pre-specified payout functions $d(t) \geq 0$ for some $t \geq 0$. That is to say, the only choice variables left in the payoff (3) are the guarantee $g(t)$ and the surplus participation rate α. In subsequent sections, we will not determine any optimal tontine payout functions and assume that $d(t)$ is fixed by the insurer. Note that the payoff (3) consists of the guarantee and a call option written on a tontine. More specifically, the underlying of this option is the tontine payoff paid to a surviving individual policyholder, i.e., $nd(t)/N_\epsilon(t)$. It is a function of the number of the surviving policyholders, with initially in total n policyholders agreeing to enter the same insurance pool. If the tontine payoff outperforms the guarantee, i.e., $N_\epsilon(t) \leq \frac{nd(t)}{g(t)}$, the policyholder obtains the guarantee payment $g(t)$ and additionally participates in the surplus with a participation rate $\alpha \in [0,1]$, otherwise she obtains the guarantee payment. With this specific design of the retirement product, we aim to come up with a product which leads to a better risk sharing between the policyholders and the insurers than annuities, where insurers bear all the longevity risk, and tontines, where policyholders bear most of the longevity risks. In addition, various combinations of $(g(t),\alpha)$ allow policyholders with different risk aversions and liquidity needs to choose their desirable products (see Section 4). From the insurer's viewpoint, issuing the new innovative products shall expose them to less longevity risk than annuities, and to more longevity risk than tontines (see Section 5).

3.1 Fair tontines with minimum guarantees

Following standard actuarial techniques (cf. Olivieri and Pitacco (2011)), a contract which provides the payoff (3) for a single up-front premium of W_0 is called *fair* if the pair $(g(t),\alpha)$ is chosen such that

$$V_0\left(\{b_{OTG}(t)\}_{t\geq 0}\right) = W_0\,, \tag{4}$$

where $V_0(\cdot)$ denotes the initial market value computed under a risk-neutral pricing measure Q. More discussions about the choice of Q are provided in the following subsection. Lemma 3.1 delivers a criterion for the fairness of contracts.

Lemma 3.1. *For a guaranteed payout function $g(t)$, the participation rate α^* which makes the pair $(g(t),\alpha^*)$ a fair contract is given by*

$$\alpha^* = \frac{W_0 - V_0\left(\{g(t)\mathbb{1}_{\{\zeta_\epsilon>t\}}\}_{t\geq 0}\right)}{V_0\left(\left\{\max\left\{\frac{nd(t)}{N_\epsilon(t)}-g(t),0\right\}\mathbb{1}_{\{\zeta_\epsilon>t\}}\right\}_{t\geq 0}\right)}\,. \tag{5}$$

For this fair participation rate α^*, it holds $\alpha^* \in [0,1]$ if and only if

$$W_0 \geq V_0\left(\{g(t)\mathbb{1}_{\{\zeta_\epsilon>t\}}\}_{t\geq 0}\right) \tag{6}$$

$$W_0 \leq V_0\left(\{g(t)\mathbb{1}_{\{\zeta_\epsilon>t\}}\}_{t\geq 0}\right) + V_0\left(\left\{\max\left\{\frac{nd(t)}{N_\epsilon(t)}-g(t),0\right\}\mathbb{1}_{\{\zeta_\epsilon>t\}}\right\}_{t\geq 0}\right). \tag{7}$$

Proof. The fairness condition (4) can be written as

$$W_0 = V_0\left(\{g(t)\mathbb{1}_{\{\zeta_\epsilon>t\}}\}_{t\geq 0}\right) + \alpha V_0\left(\left\{\max\left\{\frac{nd(t)}{N_\epsilon(t)}-g(t),0\right\}\mathbb{1}_{\{\zeta_\epsilon>t\}}\right\}_{t\geq 0}\right). \tag{8}$$

Solving equality (8) for α delivers (5). From (5), it is then straightforward to see that $\alpha^* \geq 0$ is equivalent to (6) and that $\alpha^* \leq 1$ is equivalent to (7). □

That is, if $\alpha > 0$, the guarantee $g(t)$ has to be chosen smaller than in the case where $\alpha = 0$ for the contract to be fair. If a policyholder would like to participate in the surpluses generated by the tontine, she has to give up a small part of her guarantee in exchange. If $\alpha = 1$, we obtain

$$\begin{aligned} b_{OTG}(t) &= \mathbb{1}_{\{\zeta_\epsilon > t\}} \left(g(t) + \max\left\{ \frac{nd(t)}{N_\epsilon(t)} - g(t), 0 \right\} \right) \\ &= \mathbb{1}_{\{\zeta_\epsilon > t\}} \left(\frac{nd(t)}{N_\epsilon(t)} + \max\left\{ g(t) - \frac{nd(t)}{N_\epsilon(t)}, 0 \right\} \right). \end{aligned} \tag{9}$$

In this case, a contract with $V_0\left(\left\{\frac{nd(t)}{N_\epsilon(t)}\mathbb{1}_{\{\zeta_\epsilon > t\}}\right\}_{t\geq 0}\right) \geq W_0$ and a positive guarantee cannot be fair. That is, the tontine payoff in a tontine with a positive minimum guarantee should be set smaller than that in a pure tontine (with no guarantee) by the insurer. From (9) we also see that the payoff of the new product can be rewritten as that of a tontine plus a put option on the tontine in case $\alpha = 1$. Note that a fair participation rate greater than 1 or smaller than zero can theoretically result. In the context of surplus participation, though, it makes sense for us to focus on fair participation rates $\alpha^* \in [0, 1]$.

3.2 Risk-neutral valuation

As the market for insurance is incomplete, there is no unique price resulting from an arbitrage-free pricing. Various methods have been introduced to price mortality- or longevity-linked securities (see e.g. Cairns et al. (2006), Bauer et al. (2010), and the references therein). We rely on the so-called risk-neutral pricing method. The insurer chooses, for pricing purposes, a risk-neutral probability measure Q among the infinitely many risk-neutral probability measures existing in incomplete arbitrage-free markets. The probability measure Q then accounts for both unsystematic and systematic mortality risk. By fixing a risk-neutral probability Q, we are assuming a given Sharpe ratio/risk premium for longevity risk.[6] [7] Strictly speaking, when determining the pricing measure Q, a real-life insurance company should take account of all of

[6]Note that we do not deal with equilibrium pricing of the new product, the tontine with a minimum guarantee. It would imply that we first need to determine the optimal demand of the agents for this product for a given Sharpe ratio. The equilibrium Sharpe ratio then results from the market clearing condition for the considered longevity derivative market. In the present paper, we stay in a framework where an exogenously given pricing measure Q is used for pricing the retirement products. The more challenging task of equilibrium pricing is left for further research.

[7]Bauer et al. (2010) point out that the risk premium for a longevity derivative shall be smaller than or equal to the risk premium within an annuity policy. They find that the Sharpe ratios implied by UK pension annuities are significantly smaller than those from the equity market.

its insurance business, in particular examine whether there exist some natural hedges between issued various products. An insurer who sells annuities and simultaneously traditional life insurance contracts is exposed to less longevity risks than an insurer selling purely annuities. In the present paper, our main purpose is to compare various retirement products (tontines with a minimum guarantee versus annuities or pure tontines), we assume for simplicity that the insurers under consideration sell purely retirement products (cf. Section 5). In the following, let us denote the best-estimate survival curve under the pricing measure Q by ${}_t\tilde{p}_x$. We assume that the insurer is prudent when charging premiums for contracts. As it is assumed that only retirement products are considered, a prudent way of pricing means a higher survival probability under the pricing measure Q, i.e.,

$$ {}_t\tilde{p}_x \geq {}_tp_x \,, \tag{10} $$

where ${}_tp_x$ denotes the real-world survival curve. Note that the choice of the pricing measure Q (or ${}_t\tilde{p}_x$) can depend on the pool size n. A larger pool size will lead to less longevity risks included in the retirement products, as unsystematic risks can be partly eliminated by increasing the pool size. This point will be further elaborated in the numerical section. For simplicity, we further assume that the shock ϵ follows the same distribution under both measures. Assumption (10) implies that the overall survival probability is also greater under Q than under P, that is,

$$ \begin{aligned} {}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x) =& \mathbb{E}_Q\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] = \int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi \\ \geq& \int_{-\infty}^{1} {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi = \mathbb{E}\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] = {}_tp_x \cdot m_\epsilon(-\ln {}_tp_x), \end{aligned} $$

where $m_\epsilon(\cdot)$ is the moment generating function of the shock ϵ, whose existence is assumed in the following. The higher the survival probability under Q, the more conservative and prudent the insurer is when setting prices for retirement products. Denoting by r the risk-free interest rate, the initial value of a tontine with a minimum guarantee can be determined as

$$ \begin{aligned} & V_0\left(\{b_{OTG}(t)\}_{t\geq 0}\right) = \mathbb{E}_Q\left[\int_0^\infty e^{-rt} b_{OTG}(t)\mathrm{d}t\right] \\ & = \alpha \int_0^\infty e^{-rt}\mathbb{E}_Q\left[{}_t\tilde{p}_x^{1-\epsilon}\mathbb{E}_Q\left[\max\left\{\frac{nd(t)}{N_\epsilon(t)} - g(t), 0\right\} \;\middle|\; \zeta_\epsilon > t, \epsilon\right]\right] dt \\ & \quad + \int_0^\infty e^{-rt} g(t) {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x)\mathrm{d}t \\ & = \alpha \int_0^\infty e^{-rt}\mathbb{E}_Q\left[\sum_{k=0}^{n-1} \max\left\{\frac{nd(t)}{k+1} - g(t), 0\right\} \binom{n-1}{k} \left({}_t\tilde{p}_x^{1-\epsilon}\right)^{k+1} \left(1 - {}_t\tilde{p}_x^{1-\epsilon}\right)^{n-1-k}\right] \mathrm{d}t \\ & \quad + \int_0^\infty e^{-rt} g(t) {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x)\mathrm{d}t \end{aligned} $$

$$
\begin{aligned}
&= \alpha \int_0^\infty e^{-rt} \sum_{k=0}^{n-1} \max\left\{\frac{nd(t)}{k+1} - g(t), 0\right\} \binom{n-1}{k} \mathbb{E}_Q\left[\left({}_t\tilde{p}_x^{1-\epsilon}\right)^{k+1}\left(1 - {}_t\tilde{p}_x^{1-\epsilon}\right)^{n-1-k}\right] \mathrm{d}t \\
&\quad + \int_0^\infty e^{-rt} g(t)\, {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x) \mathrm{d}t \\
&= \alpha \int_0^\infty e^{-rt} \sum_{k=0}^{n-1} \max\left\{\frac{nd(t)}{k+1} - g(t), 0\right\} \binom{n-1}{k} \int_{-\infty}^1 \left({}_t\tilde{p}_x^{1-\varphi}\right)^{k+1}\left(1 - {}_t\tilde{p}_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, \mathrm{d}\varphi\, \mathrm{d}t \\
&\quad + \int_0^\infty e^{-rt} g(t)\, {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x) \mathrm{d}t.
\end{aligned}
$$

Using equation (5), we are now able to derive the fair α^* for any given tontine payoff $d(t)$ and any guarantee $g(t)$. Note that, for α^* to be between 0 and 1, inequalities (6) and (7) need to be fulfilled. Furthermore, the initial values of the annuity and pure tontine (with no guarantee) can be obtained as boundary cases by setting $\alpha = 0$ and $g(t) = 0$, respectively.

Let us consider a numerical example. Based on Milevsky and Salisbury (2015), we consider the so-called natural tontine, whose payoff is given by $d(t) := \mathbb{E}_Q\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] d_0 = {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x) d_0$, where d_0 is a constant. If the number of surviving policyholders in the tontine evolves exactly as expected, that is, $N_\epsilon(t) = n\mathbb{E}_Q\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right]$ for any t, this tontine payoff to one policyholder is constant over time and given by

$$
\frac{nd(t)}{N_\epsilon(t)} = \frac{n\mathbb{E}_Q\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] d_0}{n\mathbb{E}_Q\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right]} = d_0 .
$$

Due to its neat structure, this tontine design is recommended by Milevsky and Salisbury (2015) for an implementation of tontines in today's world. Note that this tontine design is only referred to as natural tontine in this specific model setup. It is typical for pooled annuity funds (see the literature on page 1 of this article) to also have this property. For example, the payments of the pooled annuity fund already considered in Piggott et al. (2005) remain constant over time as well if deaths in the pool occur exactly as expected (when disregarding financial market risk).

For our base case, we assume a constant guarantee, that is, $g(t) = g$ for all $t \geq 0$. The base case parameters are summarized in Table 1.

Guarantee $g(t) = g = 7.5$	Pool size $n = 150$	Natural tontine $d_0 = 10$
Premium $W_0 = 100$	Initial age $x = 65$	Risk-free rate $r = 0.04$
Modal ages at death (P and Q) $m = 80$, $\widetilde{m} = 84$	Dispersion coefficient $\beta = 10$	Longevity shock $\epsilon \sim \mathcal{N}_{(-\infty,1]}(-0.0035, 0.0814^2)$

Table 1: Base case parameter setup.

A few remarks regarding our parameter choice:

- The best-estimate survival probabilities ${}_tp_x$ and ${}_t\tilde{p}_x$ are assumed to follow the well-known Gompertz-law (Gompertz (1825)) as used, for example, in Gumbel (1958), Milevsky and Salisbury (2015) and Chen et al. (2019). In other words, we assume that

$$ {}_tp_x = e^{e^{\frac{x-m}{\beta}}\left(1-e^{\frac{t}{\beta}}\right)}, \quad {}_t\tilde{p}_x = e^{e^{\frac{x-\widetilde{m}}{\beta}}\left(1-e^{\frac{t}{\beta}}\right)} $$

  with $\beta > 0$ being the dispersion coefficient and $m > 0$, $\widetilde{m} > 0$ being the modal ages at death. The modal age at death used for pricing is larger than for the real-world measure for (10) to be fulfilled. Our parameter choice concerning the Gompertz parameters is rather typical and results in an expected remaining lifetime of $\mathbb{E}_P[\zeta_\epsilon] = \int_0^\infty \int_{-\infty}^1 {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\,\mathrm{d}\varphi\,\mathrm{d}t \approx 14.180$ under P and $\mathbb{E}_Q[\zeta_\epsilon] = \int_0^\infty \int_{-\infty}^1 {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\,\mathrm{d}\varphi\,\mathrm{d}t \approx 17.040$ under Q. As the average life expectancy across OECD countries was 80.5 (77.8 for men and 83.1 for women) in 2013 (OECD (2016)), we believe that this is a reasonable parameter choice. For the parameters of the random longevity shock ϵ we follow Chen et al. (2019).

- To check whether the fair surplus participation α^* is between 0 and 1, we take a look at conditions (6) and (7). It is easy to verify that both conditions are satisfied. Note that the choice of our tontine payoffs leads to an initial value of the tontines which is larger than the initial investment, that is, $V_0\left(\left\{\frac{nd(t)}{N_\epsilon(t)}\mathbb{1}_{\{\zeta_\epsilon>t\}}\right\}_{t\geq 0}\right) > W_0$. By this assumption, we include pure tontine products in our analysis: If we set $g(t) = 0$ and compute the fair participation rate α^* for the resulting product, we obtain $\alpha^* < 1$. Consequently, the fair participation rate α^* lies between 0 and 1 for all nonnegative choices of the guarantee.

Before we perform a sensitivity analysis for the fair surplus participation α^*, let us demonstrate the effect of the pricing measure Q on the initial value of the annuity and the pure tontine. Let us denote by $V_0^A(\Xi)$ and $V_0^{OT}(\Xi)$ the initial values of the annuity and the pure tontine taken under the probability measure $\Xi \in \{P, Q\}$. Then, the ratio

$$ \mathrm{SL}_X = \frac{V_0^X(Q) - V_0^X(P)}{V_0^X(P)} \tag{11} $$

can be considered as the safety loading the insurer charges for the two retirement products $X \in \{A, OT\}$. We consider an annuity with constant payoff $g(t) = g$ and a pure natural tontine with payoffs as given in Table 1. The resulting safety loadings are given in Table 2. The higher the safety loading, the more prudently the insurer charges the premium. If the price of a product is influenced more substantially by the prudence assumption, more longevity risk is contained in the product. While both safety loadings increase in $\widetilde{m}$, we can see that the use

$\widetilde{m}$	Annuity (SL_A)	Natural tontine (SL_{OT})
80	0	0
84	0.143	0.002
88	0.283	0.006

Table 2: Safety loadings SL_X as given in (11) for the annuity and the pure natural tontine. The parameters are taken from Table 1 (particularly, $m = 80$), where the guarantee is used as annuity payment.

of the risk-neutral measure Q mainly increases the safety loading of the annuity. Concerning the natural tontine, we observe that it is hardly affected by the use of the risk-neutral measure, compared to the annuity. That is, using a prudent life table, the safety loading of an annuity is much larger than that of a natural tontine. In other words, much more longevity risk is involved in annuities than in natural tontines.

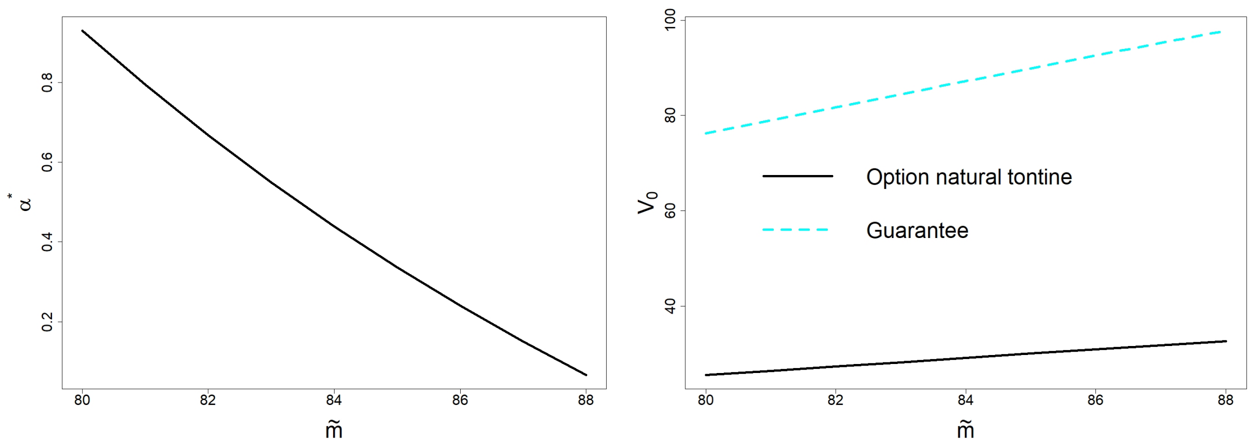
Table 3 provides a sensitivity analysis of the fair factor α^* depending on the choice of the constant guarantee $g = g(t)$. For each guarantee, we determine α^* such that the fairness condition (4) is fulfilled. Table 3 shows that there is a negative relation between the guaranteed

Guarantee	Fair participation rate
$g = 6.5$	$\alpha^* = 0.60$
$g = 7$	$\alpha^* = 0.53$
$g = 7.5$	$\alpha^* = 0.44$
$g = 8$	$\alpha^* = 0.30$
$g = 8.5$	$\alpha^* = 0.07$

Table 3: Fair $\alpha^* = \alpha^*(g)$ as given in (5) depending on the choice of the constant guarantee $g(t) = g$. The parameters are taken from Table 1.

payment g and the resulting fair participation rate α^*. Since we assume the tontine payoff to be given, the fair share of the surplus α^* needs to decrease if the guarantee increases. It is a natural result as the contract value of the new product increases both in the guarantee and in the participation rate.

Figure 1 shows the fair α^* depending on the modal age at death used for pricing $\widetilde{m}$ along with the initial value of the guarantee (see numerator in (5)) and the option on the tontine (see denominator in (5)). We observe in Figure 1 (a) that the fair participation rate α^* decreases in $\widetilde{m}$. That is, the more prudent the insurer is (higher $\widetilde{m}$), the lower the fair participation rate α^* becomes. For a more prudent insurer, a substantially higher fee will be charged for providing the guarantee part (see Figure 1 (b) and also Table 2). In other words, the initial value of the guaranteed payments with $g(t) = g = 7.5$ goes up if a more prudent pricing measure is used ($\widetilde{m}$ increases). Let us now consider the initial value of the option on the tontine. Recall that



(a) Fair participation rate (b) Initial values

Figure 1: Optimal participation rate α^*, the initial value of the guarantee (see numerator in (5)) and the option on the tontine (see denominator in (5)) depending on the modal age at death $\widetilde{m}$ of the pricing measure $\widetilde{m}$. The remaining parameters are chosen as in Table 1 (particularly, $g = 7.5$).

it can be written as

$$V_0\left(\left\{\max\left\{\frac{nd(t)}{N_\epsilon(t)} - g(t), 0\right\}\mathbb{1}_{\{\zeta_\epsilon > t\}}\right\}_{t\geq 0}\right)$$
$$= \int_0^\infty e^{-rt}\sum_{k=0}^{n-1}\max\left\{\frac{nd(t)}{k+1} - g(t), 0\right\}\binom{n-1}{k}\int_{-\infty}^1 \left({}_t\tilde{p}_x^{1-\varphi}\right)^{k+1}\left(1 - {}_t\tilde{p}_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\,\mathrm{d}\varphi\,\mathrm{d}t.$$

In Figure 1 (b), we see that the initial value of the option on the natural tontine increases in $\widetilde{m}$. Note that an increase in $\widetilde{m}$ affects both the individual's survival $\mathbb{1}_{\{\zeta_\epsilon > t\}}$ and the number of surviving pool members $N_\epsilon(t)$. For the natural tontine, the payout function $d(t) = \mathbb{E}_Q\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] d_0$ is directly affected by an increase in $\widetilde{m}$, which offsets the effect of $\widetilde{m}$ on $N_\epsilon(t)$. Consequently, the effect on $\mathbb{1}_{\{\zeta_\epsilon > t\}}$ dominates for the natural tontine and the initial option value increases in $\widetilde{m}$.

As already pointed in the previous texts, it makes sense to choose the pricing measure Q which depends on the pool size n. Theoretically, when n goes to infinity, the pricing Q accounts exclusively for the systematic mortality risk, while for a finite n, it contains both the unsystematic and systematic mortality risk. In our context, if we use an n value smaller than our benchmark case, we could choose a higher ${}_t\tilde{p}_x$ (by choosing a higher modal age at death $\widetilde{m}$). On the contrary, we could choose a lower $\widetilde{m}$ leading to a lower ${}_t\tilde{p}_x$. To examine the effect of the pricing measure Q dependent on the pool size n, we carry out some numerical analyses for two additional cases: $n = 50$, $\widetilde{m} = 86$ and $n = 300$, $\widetilde{m} = 82$.

Analogously to Table 3, Table 4 provides a sensitivity analysis of the fair participation rate α^* depending on the constant guarantee g assuming different specifications of the pricing measure Q which depends on the pool size n.[8] We observe that under $n = 50$ and $\widetilde{m} = 86$ a lower fair

Guarantee	Fair participation rate
$n = 50$, $\widetilde{m} = 86$	
$g = 6.5$	$\alpha^* = 0.46$
$g = 7$	$\alpha^* = 0.37$
$g = 7.5$	$\alpha^* = 0.24$
$g = 8$	$\alpha^* = 0.05$
$n = 300$, $\widetilde{m} = 82$	
$g = 6.5$	$\alpha^* = 0.76$
$g = 7$	$\alpha^* = 0.72$
$g = 7.5$	$\alpha^* = 0.67$
$g = 8$	$\alpha^* = 0.58$

Table 4: Fair $\alpha^* = \alpha^*(g)$ as given in (5) depending on the choice of the constant guarantee $g(t) = g$. We consider two cases: $n = 50$, $\widetilde{m} = 86$ (more prudent insurer with rather small pool size) and $n = 300$, $\widetilde{m} = 82$ (less prudent insurer with larger pool size). The remaining parameters are taken from Table 1.

surplus participation results than in Table 3 (using $n = 150$ and $\widetilde{m} = 84$). Under $n = 300$ and $\widetilde{m} = 82$, on the other hand, a higher surplus participation rate is obtained. This effect of $\widetilde{m}$ on the fair participation rate has already been observed in Figure 1. For a more prudent insurer, a substantially higher fee will be charged for providing the guarantee part. To keep the contract fair, as a compensation, a lower participation rate results for the same guarantee g and a higher $\widetilde{m}$.

In the following two sections, we analyze the attractiveness of the new product to the individual policyholders and to the insurers, respectively. The analyses will be done under the real-world measure P over the possible fair combinations of $(g(t), \alpha^*)$. In order to avoid unnecessary repetition, we decide to use the fair combinations given in Table 3 for the subsequent analyses.

4 Policyholder's perspective

In this section, we analyze how tontines with guarantees can serve different individuals' preferences. Let $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$ be a CRRA utility function with a risk aversion parameter $\gamma > 0$,

[8] Note that Table 4 only covers constant guarantee values up to 8 and not 8.5 like Table 3. The reason for this that the parameter choice $n = 50$, $\widetilde{m} = 86$, $g = 8.5$ would deliver a negative fair surplus participation rate which we do not consider in our analysis.

$\gamma \neq 1$ and ρ be the subjective discount factor of the policyholder. Further, define $\chi(t) := g(t) + \alpha \max\left\{\frac{nd(t)}{N_\epsilon(t)} - g(t), 0\right\}$ as the payoff to a living policyholder. We introduce the policyholder's expected discounted lifetime utility as

$$\begin{aligned} U\big(\{\chi(t)\}_{t\geq 0}\big) &:= \mathbb{E}\left[\int_0^\infty e^{-\rho t} u\left(\chi(t)\right) \mathbb{1}_{\{\zeta_\epsilon > t\}} \mathrm{d}t\right] \\ &= \int_0^\infty e^{-\rho t} \sum_{k=0}^{n-1} u\left(g(t) + \alpha^* \max\left\{\frac{nd(t)}{k+1} - g(t), 0\right\}\right) \binom{n-1}{k} \\ &\quad \cdot \int_{-\infty}^1 \left({}_tp_x^{1-\varphi}\right)^{k+1} \left(1 - {}_tp_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, \mathrm{d}\varphi\, \mathrm{d}t, \end{aligned} \tag{12}$$

where equation (12) can be obtained in a similar way as the initial value of the product. The policyholder now aims to maximize her expected discounted lifetime utility over all possible combinations $(g(t), \alpha^*)$, where

$$\begin{aligned} \alpha^* &= \alpha^*(\{g(t)\}_{t\geq 0}) \\ &= \frac{W_0 - \int_0^\infty e^{-rt} g(t) {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x) \mathrm{d}t}{\int_0^\infty e^{-rt} \sum_{k=0}^{n-1} \max\left\{\frac{nd(t)}{k+1} - g(t), 0\right\} \binom{n-1}{k} \int_{-\infty}^1 \left({}_t\tilde{p}_x^{1-\varphi}\right)^{k+1} \left(1 - {}_t\tilde{p}_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\, \mathrm{d}\varphi\, \mathrm{d}t} \end{aligned} \tag{13}$$

is the fair participation rate as in (5). To be more precise, we consider the following optimization problem:

$$\max_{(g(t),\alpha^*)} U\big(\{\chi(t)\}_{t\geq 0}\big) \quad \text{subject to (13).}$$

Note that the maximum operator is not differentiable. Therefore, no implicit or explicit solution for the optimal choice of $g(t)$ can be derived. Consequently, the only possible way to determine the optimal guarantee $g(t)$ is to calculate the expected utility (12) for all possible choices of $g(t)$ and then find the maximum among these. Therefore, we only consider certain parameterized choices of the guarantee $g(t)$, specifically the following three cases:

- Constant guarantee (CG): We set $g(t) = g$, that is, the policyholders require the same minimum guarantee at all times during the retirement phase. The initial value of this guarantee is given by

$$V_0\left(\{g(t)\mathbb{1}_{\{\zeta_\epsilon > t\}}\}_{t\geq 0}\right) = g \int_0^\infty e^{-rt} {}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x) \mathrm{d}t.$$

- Continuously increasing guarantee (CIG): We set $g_0(t) = g_0 e^{\delta t}$, where $\delta \geq 0$ is a guaranteed interest rate issued by the insurer. The initial value of this guarantee is given by

$$V_0\left(\{g_0(t)\mathbb{1}_{\{\zeta_\epsilon>t\}}\}_{t\geq 0}\right) = g_0\int_0^\infty e^{(\delta-r)t}{}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x)\mathrm{d}t.$$

- Discretely increasing guarantee (DIG): The guarantee is chosen as

$$g_1(t) = \begin{cases} g_1, & \text{for } t \in [0, z-x) \\ \kappa g_1, & \text{for } t \geq z-x \end{cases},$$

where $\kappa > 1$ is a deterministic factor and $z \geq x$ is some predetermined age at which the policyholder would like to increase her minimum guarantee. The initial value of this guarantee is given by

$$\begin{aligned}&V_0\left(\{g_1(t)\mathbb{1}_{\{\zeta_\epsilon>t\}}\}_{t\geq 0}\right)\\ &= g_1\left(\int_0^{z-x} e^{-rt}{}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x)\mathrm{d}t + \kappa\int_{z-x}^\infty e^{-rt}{}_t\tilde{p}_x m_\epsilon(-\log {}_t\tilde{p}_x)\mathrm{d}t\right).\end{aligned}$$

The latter two guarantee choices are chosen to represent the observed pattern that individuals tend to have higher liquidity needs at old ages than at (relatively) young ages in the retirement phase (see, for example, Weinert and Gründl (2017)).[9] The base case parameters δ, κ and z are given in Table 5. In Table 6, we display the optimal (initial) levels of guarantee along with the resulting fair participation rate depending on the risk aversion γ for the three different guarantee designs.

Guaranteed interest rate	Guarantee increasing factor	Age of increase
$\delta = r = 0.04$	$\kappa = 1.5$	$z = 75$

Table 5: Additional base case parameters to Table 1.

γ	0.5	2	4	6	8	10
CG (g^*, α^*)	0, 0.86	0, 0.86	0.52, 0.85	2.05, 0.82	2.96, 0.80	3.76, 0.78
CIG (g_0^*, α^*)	0, 0.86	0, 0.86	0.11, 0.86	0.44, 0.85	0.63, 0.85	0.84, 0.84
DIG (g_1^*, α^*)	0, 0.86	0, 0.86	0.30, 0.86	1.23, 0.84	1.77, 0.82	2.25, 0.81

Table 6: Optimal guarantees along with the fair participation rate for the natural tontines with minimum guarantees depending on the risk aversion γ. The parameters are taken from Tables 1 and 5 and we assume that $\rho = r$.

In Table 6, we make similar observations for all three guarantee designs. As the risk aversion increases, the optimal guarantee increases and the fair participation rate decreases. Note that

[9] Note that other patterns of liquidity need are also observed in the literature. Therefore, our analysis of increasing guarantees should be seen as one example for non-constant liquidity needs instead of a thorough analysis of all the liquidity needs that have been observed. For a detailed analysis of retirement consumption behavior, see, for example, Blanchett (2013).

the fair participation rate never becomes 0, which would correspond to an annuity being the best source of retirement income. In other words, the annuity is outperformed by a natural tontine with a minimum guarantee for all risk aversions considered. In this sense, a natural tontine with a minimum guarantee can serve each individual's risk appetite better and actually forms a whole range of retirement plans for various risk aversions and liquidity needs. Compared to pure tontines (with no guarantees), tontines with minimum guarantees still perform better for the majority of individuals. Only to individuals with a rather low risk aversion, that is, $\gamma \in \{1/2, 2\}$, the pure natural tontine is the optimal source of retirement income in our parameter setup. In total, our results are consistent with Milevsky and Salisbury (2015), who "propose the natural tontine as a reasonable structure for designing tontine products in practice." Our analyses confirm this suggestion and supplement it by the inclusion of a minimum guarantee in the payoff of natural tontines.

5 Insurer's perspective

While Chen et al. (2019) consider the attractiveness of their tonuity exclusively from the policyholder's perspective, we want to analyze whether the insurer can benefit from the new tontines with guarantees as well. As it is less natural to assume a utility function for the insurance company, we consider other important quantities of interest from the insurer's perspective. Inspired by, for example, Bauer and Weber (2008), Li and Hardy (2011), Cairns (2013), Kling et al. (2014) and Olivieri and Pitacco (2019), we consider the (random) present value of future losses.[10] To assess the risks contained in retirement products, we apply two risk measures to this random variable, which are both described in detail below.

5.1 A single cohort

Let L_X denote the (random) present value of future liabilities of retirement product $X \in \{A, OT, OTG\}$. We consider three different insurance companies, each selling one type of contract. The first one sells annuities exclusively, the second one sells pure natural tontines (with no guarantees) and the last one sells only natural tontines with minimum guarantees. It could be the situation that the first insurer attracts individuals with high risk aversion, the second one serves individuals with low risk aversion and the third one is attractive to individuals with medium risk aversion. Each insurer holds a portfolio of n policyholders with the same age and risk aversion who all purchase the same type of retirement plan. At each time $t \geq 0$, the insurers pay (1), (2) and (3), each multiplied by the number of policyholders still alive $N_\epsilon(t)$,

[10]The provided literature considers similar or comparable quantities of interest to the insurer. For example, Olivieri and Pitacco (2019) consider the present value of future profits which is simply the present value of future losses multiplied by (-1).

respectively. Consequently, the (random) present value of future liabilities L_X can, for each retirement plan (or each insurer), be expressed in the following way:

- Insurer one with annuities:

$$L_A = \int_0^\infty N_\epsilon(t) e^{-rt} c(t) \mathrm{d}t \,. \tag{14}$$

- Insurer two with pure tontines (with no guarantees):

$$L_{OT} = \int_0^\infty e^{-rt} n h(t) \mathbb{1}_{\{N_\epsilon(t)>0\}} \mathrm{d}t \,, \tag{15}$$

  where $h(t)$ denotes the deterministic payment stream of the pure tontine contract, fixed at the beginning of the contract.

- Insurer three with tontines with minimum guarantees:

$$L_{OTG} = \int_0^\infty e^{-rt} \left(N_\epsilon(t) g(t) + \alpha^* \max\left\{ n d(t) - N_\epsilon(t) g(t), 0 \right\} \mathbb{1}_{\{N_\epsilon(t)>0\}} \right) \mathrm{d}t \,. \tag{16}$$

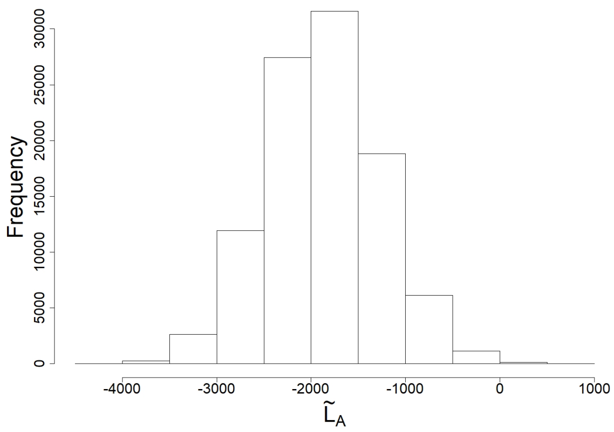
Then, we consider the (random) present value of future losses defined by

$$\widetilde{L}_X := L_X - nW_0 = L_X - n\mathbb{E}_Q\left[L_X\right] \tag{17}$$

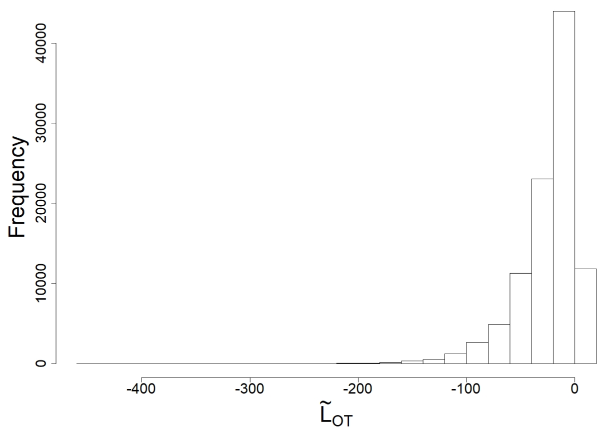
for $X \in \{A, OT, OTG\}$. In particular, all three insurance companies receive the same premium W_0 per contract at time 0. In Figure 2, we show the histograms of 100000 realizations of $\widetilde{L}_X$ of an annuity, a pure natural tontine, and a natural tontine with a constant minimum guarantee. The annuity payoff $c(t)$ and the payoff of the pure tontine $h(t)$ are chosen such that the up-front premium W_0 is equal to their initial market values, respectively. The constant guarantee of the tontine with a minimum guarantee is the utility maximizing guarantee for a risk aversion of $\gamma = 6$, that is, $g = 2.05$ (see Table 6). The resulting fair participation rate is $\alpha^* = 0.82$ (see also Table 3). The remaining parameters are chosen as in Table 1. We observe the following:

- Panel (a) and (b): The pure natural tontine creates lower losses than the annuity. Furthermore, the pure tontine creates losses more frequently than the annuity. The reason for these patterns is the decreasing payoff structure of the natural tontine, which leads to a distribution which is skewed to the left. If more policyholders survive than expected, only low payments are made to the remaining policyholders. The trade-off for this is the rather limited “upside potential” of the natural tontine compared to the annuity.

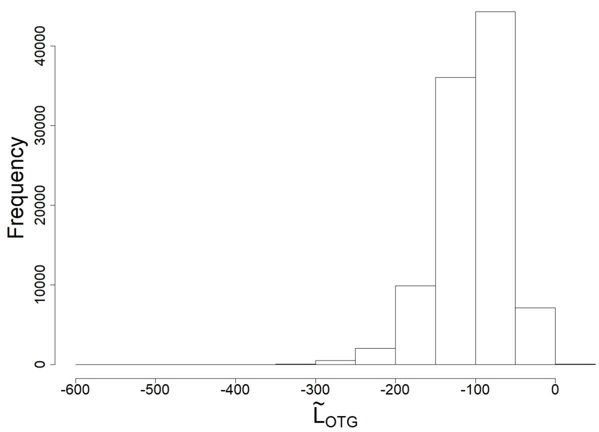
- Panel (c): The natural tontine with a minimum guarantee combines the advantages of the annuity and the pure natural tontine. It creates lower and less frequent losses than



(a) Annuity



(b) Pure natural tontine



(c) Natural tontine with a minimum guarantee

Figure 2: Histograms of the present value of future losses for a constant annuity, a pure natural tontine, and a natural tontine with a constant minimum guarantee. The parameters are chosen as in Table 1, except for the constant guarantee, which is $g = 2.05$. The fair participation rate is $\alpha^* = 0.82$ (see also Table 6).

an annuity. Compared to the pure natural tontine, the natural tontine with a guarantee creates higher losses, but losses occur less frequently.

To further assess the performance of the different retirement plans we now introduce two risk measures: We consider the overall loss probability $P\left(\widetilde{L}_X > 0\right)$ and the *conditional* (expected) present value of future losses per contract $\frac{1}{n}\mathbb{E}\left[\widetilde{L}_X \,\middle|\, \widetilde{L}_X > 0\right]$ which determines the average loss per contract in case a loss occurs. In Table 7, we provide both risk measures for the three different retirement products. Table 7 supports the conclusions drawn from Figure 2. The natural tontine with a minimum guarantee performs better than the annuity in both risk measures. For the given parameters, it particularly manages to reduce the conditional expected loss by more than 95% compared to the annuity, while simultaneously yielding a slightly lower loss probability. Compared to the pure natural tontine, the natural tontine with

	$P\left(\widetilde{L}_X > 0\right)$	$\frac{1}{n}\mathbb{E}\left[\widetilde{L}_X \,\middle|\, \widetilde{L}_X > 0\right]$
Annuity	0.0012	1.23
Pure natural tontine	0.1169	0.01
Natural tontine with a guarantee	0.0009	0.06

Table 7: Risk measures for a constant annuity, a pure natural tontine, and a natural tontine with a constant minimum guarantee. The parameters are chosen as in Table 1, except for the constant guarantee, which is $g = 2.05$. The fair participation rate is $\alpha^* = 0.82$ (see also Table 6).

a guarantee performs worse concerning the conditional expected loss and better concerning the loss probability.

5.2 Multiple cohorts

To further investigate the flexibility of tontines with minimum guarantees, we now assume that each insurer holds a portfolio of k different cohorts. We denote the initial size of each cohort by n_i and the size of each cohort at time $t \geq 0$ by $N_\epsilon^{(i)}(t)$ for all $i = 1, \dots, k$, $k \geq 1$. We assume that, given the shock ϵ, the cohorts are independent of each other. Insurer one only allows the cohorts to differ in *size*, that is, the (random) present value of future liabilities of the annuity is the following slight generalization of (14):

$$L_A^c = \sum_{i=1}^{k} \int_0^\infty N_\epsilon^{(i)}(t) e^{-rt} c(t) \mathrm{d}t.$$

Insurer two allows the cohorts to differ in *size* and consequently, also has to allow them to differ in the *payout functions*, denoted by $h_i(t)$, as each $h_i(t)$ depends on the pool size n_i. The (random) present value of future liabilities of the overall portfolio of pure tontines is thus the following generalization of (15):

$$L_{OT}^c = \sum_{i=1}^{k} \int_0^\infty e^{-rt} n_i h_i(t) \mathbb{1}_{\left\{N_\epsilon^{(i)}(t) > 0\right\}} \mathrm{d}t.$$

Insurer three allows the cohorts to differ in *size and guarantee*. We denote the guarantee of each cohort by $g^{(i)}(t)$ and the resulting fair participation rate by α_i^*. Note that the tontine payout function of the tontines with minimum guarantees $d(t)$ does not vary between the cohorts, that is, all the cohorts purchase the new products building on similar underlying tontines. Note, however, that the tontines are only identical in their payout function and differ in the mortality experienced. The (random) present value of future liabilities of the overall portfolio of tontines with minimum guarantees can then be generalized from (16) to

$$L^c_{OTG} = \sum_{i=1}^{k} \int_0^{\infty} e^{-rt} \left(N^{(i)}_\epsilon(t) g^{(i)}(t) + \alpha^*_i \max\left\{ n_i d(t) - N^{(i)}_\epsilon(t) g^{(i)}(t), 0 \right\} \mathbb{1}_{\left\{ N^{(i)}_\epsilon(t) > 0 \right\}} \right) \mathrm{d}t.$$

Similarly as in equation (17) for one cohort, the present value of future losses is given by

$$\widetilde{L}^c_X := L^c_X - W_0 \sum_{i=1}^{k} n_i \,, \tag{18}$$

where $X \in \{A, OT, OTG\}$. Table 8 provides the similar risk measures as Table 7 for the following parameter setup:

- There are $k = 2$ cohorts, each with size $n_1 = n_2 = 150$.
- Cohort one has guarantee $g^{(1)} = 2.05$, the second cohort has $g^{(2)} = 3.76$, that is, the utility maximizing guarantees of the natural tontine for risk aversions of $\gamma = 6$ and $\gamma = 10$, respectively (see Table 6). The corresponding fair participation rates are given by $\alpha^* = 0.82$ and $\alpha^* = 0.78$, respectively (see also Table 6).
- All the remaining parameters are taken from Table 1.

	$P\left(\widetilde{L}_X > 0\right)$	$\frac{1}{n}\mathbb{E}\left[\widetilde{L}_X \,\middle|\, \widetilde{L}_X > 0\right]$
Annuity	0.0003	1.03
Pure natural tontine	0.0346	0.01
Natural tontines with guarantees	0.0003	0.08

Table 8: Risk measures for a constant annuity, a pure natural tontine, and a natural tontine with a constant minimum guarantee for two cohorts with size $n_1 = n_2 = 150$. The guarantees of the cohorts are $g^{(1)} = 2.05$ and $g^{(2)} = 3.76$, and the fair participation rates are $\alpha^* = 0.82$ and $\alpha^* = 0.78$, respectively (see also Table 6). All the remaining parameters are taken from Table 1.

Primarily, we are interested in whether the tontines with guarantees still manage to outperform the annuity. Concerning the loss probability, the annuity now delivers an (almost) identical value as the natural tontines with guarantees. However, the natural tontines with guarantees still reduce the expected conditional loss by more than 92%. Apart from these observations, the relation between the pure natural tontine and the natural tontine with a minimum guarantee does not change compared to the case with one cohort.

6 Conclusion

In this article, we follow Donnelly and Young (2017) and present a new approach of combining annuities and tontines by considering options on tontines, where the tontine design is based on Milevsky and Salisbury (2015). The resulting product can be seen as a tontine with a minimum guarantee: For each point in time, it consists of a guaranteed annuity-like component and a call option written on a tontine multiplied by a surplus participation rate. In this sense, the new product eliminates the risk that policyholders have to face volatile and, potentially, extremely low payments at old ages, which is one of the main disadvantages of tontines from the policyholders' perspective. Extending Donnelly and Young (2017), we analyze the new product in an expected utility framework. We show that a natural tontine with a minimum guarantee outperforms both an annuity and the comparable pure tontine (with no guarantee) for different risk aversion parameters. By allowing individuals with different types of risk aversion to choose their optimal guarantee, we show that the new product provides a whole range of retirement plans which are able to serve different risk appetites and liquidity needs. Further extending Donnelly and Young (2017), we provide an analysis which shows that a tontine with a minimum guarantee can also be attractive from an insurer's point of view: Compared to annuities, they can reduce the conditional expected loss drastically while yielding an (almost) identical loss probability. Consequently, they provide a new type of retirement plan which can possibly contain less risk than annuities for the insurer.

As we, in this article, solely use natural tontines in our illustrative examples, a possible extension of our article could be the consideration of additional tontine designs. For example, Milevsky and Salisbury (2015) also consider the historical tontine design, which they call *flat* tontine, whose payout function is, in contrast to the natural tontine, constant over time. That is, the payments to a single policyholder increase over time in this tontine design. Such a payment profile would be of special interest to individuals who have higher liquidity needs at older ages than at younger ages, which is often observed among retirees (see, for example, Weinert and Gründl (2017)).

References

Bauer, D., Börger, M., and Ruß, J. (2010). On the pricing of longevity-linked securities. *Insurance: Mathematics and Economics*, 46(1):139–149.

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

Benartzi, S., Previtero, A., and Thaler, R. H. (2011). Annuitization puzzles. *Journal of Economic Perspectives*, 25(4):143–64.

Blanchett, D. (2013). Estimating the true cost of retirement. *Available at:* `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.694.7210`.

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

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

Cairns, A. J., Blake, D., and Dowd, K. (2006). Pricing death: Frameworks for the valuation and securitization of mortality risk. *ASTIN Bulletin: The Journal of the IAA*, 36(1):79–120.

Campbell, J. Y. and Viceira, L. M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors.* Oxford University Press, New York.

Chen, A., Hentschel, F., and Klein, J. K. (2015). A utility-and CPT-based comparison of life insurance contracts with guarantees. *Journal of Banking & Finance*, 61:327–339.

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

Davidoff, T., Brown, J. R., and Diamond, P. A. (2005). Annuities and individual welfare. *American Economic Review*, 95(5):1573–1590.

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

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

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

Døskeland, T. M. and Nordahl, H. A. (2008). Optimal pension insurance design. *Journal of Banking & Finance*, 32(3):382–392.

Evans, A. (2018). Changing trends in mortality: An international comparison. *National Office of Statistics. Available at* `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/articles/changingtrendsinmortalityaninternationalcomparison/2000to2016`.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, 115:513–583.

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

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

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

Levy, H. (1994). Absolute and relative risk aversion: An experimental study. *Journal of Risk and uncertainty*, 8(3):289–307.

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

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

Lin, Y. and Cox, S. H. (2005). Securitization of mortality risks in life annuities. *Journal of risk and Insurance*, 72(2):227–252.

Milevsky, M. A. (2013). Life annuities: An optimal product for retirement income. *CFA Institute Research Foundation Monograph. Available at SSRN:* `https://ssrn.com/abstract=2571379`.

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

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

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

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

Murphy, S. L., Xu, J., Kochanek, K. D., and Arias, E. (2018). Mortality in the United States, 2017. *NCHS Data Brief, no 328. Hyattsville, MD: National Center for Health Statistics.*

OECD (2016). *OECD Factbook 2015-2016: Economic, Environmental and Social Statistics.* OECD Publishing, Paris. Available at `http://dx.doi.org/10.1787/factbook-2015-en`.

Olivieri, A. and Pitacco, E. (2011). *Introduction to Insurance Mathematics: Technical and Financial Features of Risk Transfers*. Springer International Publishing Switzerland.

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

Peijnenburg, K., Nijman, T., and Werker, B. J. (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, 70:18–35.

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

Pitacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press, Oxford.

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

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

Sharpe, W. F. (2017). *Retirement Income Scenarios.* Available at `http://retirementincomescenarios.blogspot.be/`.

Stamos, M. Z. (2008). Optimal consumption and portfolio choice for pooled annuity funds. *Insurance: Mathematics and Economics*, 43(1):56–68.

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

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

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

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

2 On the optimal combination of annuities and tontines

Source:

Chen, A., Rach, M., and Sehner, T. (2020). On the optimal combination of annuities and tontines. *ASTIN Bulletin: The Journal of the IAA*, 50(1):95–129.

URL: http://dx.doi.org/10.1017/asb.2019.37

DOI: 10.1017/asb.2019.37

ON THE OPTIMAL COMBINATION OF ANNUITIES AND TONTINES

BY

AN CHEN, MANUEL RACH AND THORSTEN SEHNER

ABSTRACT

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

KEYWORDS

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

JEL codes: G22, J32.

Astin Bulletin 50(1), 95-129. doi:10.1017/asb.2019.37

1. Introduction

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

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

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

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

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

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

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

2. Model setup

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

2.1. Mortality model

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

2.2. Retirement products

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

2.2.1. *Annuity and tontine*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.2.2. *Portfolio*

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

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

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

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

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

2.2.3. *Tonuity*

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

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

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

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

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

$$\begin{aligned} P_0^{[\tau]} &= \mathbb{E}\left[\int_0^\infty e^{-rt} b_{[\tau]}(t)\, dt\right] \\ &= \int_0^\tau e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_tp_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, d\varphi\, d_{[\tau]}(t)\, dt \end{aligned}$$

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

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

2.2.4. *Antine*

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

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

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

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

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

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

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

3. Optimal payoff and expected utility

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

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

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

where

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

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

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

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

3.1. Tonuity

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

$$\max_{\left(c_{[\tau]}(t),\, d_{[\tau]}(t)\right)_{t\in[0,\infty)}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\left(\mathbb{1}_{\{0\le t<\min\{\tau,T_\epsilon\}\}} u\left(\frac{n}{N_\epsilon(t)} d_{[\tau]}(t)\right) + \mathbb{1}_{\{\tau\le t<T_\epsilon\}} u\left(c_{[\tau]}(t)\right)\right) dt\right] \tag{3.19}$$

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

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

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

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

and

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

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

$$\lambda_{[\tau]} = \left(\frac{1}{v}\left(\int_0^\tau (1+C_{OT})^{1-\frac{1}{\gamma}}\, e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_tp_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^1 \left(1-\left(1-{}_tp_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\,d\varphi\right)^{1/\gamma-1}}\,dt + \int_\tau^\infty (1+C_A)^{1-\frac{1}{\gamma}}\, e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}\,{}_tp_x\, m_\epsilon(-\ln {}_tp_x)\,dt\right)\right)^\gamma \tag{3.22}$$

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

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

TABLE 1

BASE CASE PARAMETER SETUP.

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

The expected discounted lifetime utility is then given by

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

Proof. See Appendix A.1. □

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

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

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

Note the following remarks about our choice of parameters:

- To determine the value of the initial wealth v, we follow the estimation of Royal London (2018). They state that an average (British) employee needs to invest around 260 thousand pounds sterling, which approximately corresponds to 300 thousand euros, in the private pension provision to keep her standard of living in the retirement phase beginning at the age of 65 years.

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

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

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

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

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

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

As the only randomness in the optimal payoff stems from the uncertain future number of living policyholders in the pool, it is clear that, after the switch to the annuity at time $\tau^* = 27$, that is, when the policyholder turns 92 years, the two examined quantiles coincide and equal the constant annuity payment. As long as the tontine defines the tonuity, that is, while the individual

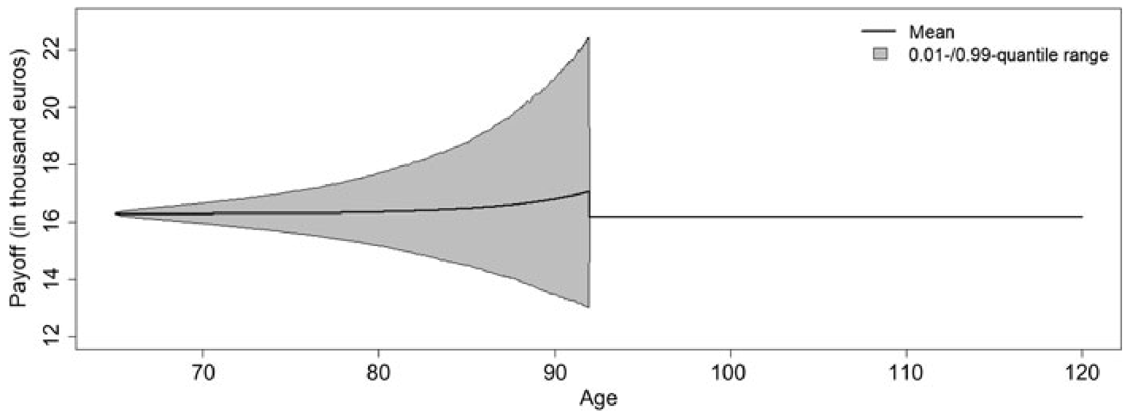


FIGURE 1: Mean and 0.01-/0.99-quantile range of the optimal payoff of the tonuity with $\tau^* = 27$ depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05. The plot is based on the assumption that the considered policyholder is always alive.

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

3.2. Antine

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

$$\max_{\left(c_{[\sigma]}(t),\, d_{[\sigma]}(t)\right)_{t\in[0,\infty)}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\left(\mathbb{1}_{\{0\le t<\min\{\sigma,T_\epsilon\}\}} u\left(c_{[\sigma]}(t)\right) + \mathbb{1}_{\{\sigma\le t<T_\epsilon\}} u\left(\frac{n}{N_\epsilon(t)} d_{[\sigma]}(t)\right)\right) dt\right] \tag{3.27}$$
$$\text{subject to} \quad v = \widetilde{P}_0^{[\sigma]} = (1+C_A)\, P_0^{A,\sigma} + (1+C_{OT})\, P_0^{OT,\sigma}.$$

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

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

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

and

$$d^*_{[\sigma]}(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}\left(\kappa_{n,\gamma,\epsilon}({}_tp_x)\right)^{1/\gamma}}{\lambda_{[\sigma]}^{1/\gamma}\,(1+C_{OT})^{1/\gamma}\left(\int_{-\infty}^{1}\left(1-\left(1-{}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\,d\varphi\right)^{1/\gamma}} \quad \textit{for all } t\in[\sigma,\infty), \tag{3.29}$$

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

$$\begin{aligned}\lambda_{[\sigma]} = \Bigg(\frac{1}{v}\Bigg(\int_0^\sigma (1+C_A)^{1-\frac{1}{\gamma}}\,e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}\,{}_tp_x\,m_\epsilon(-\ln {}_tp_x)\,dt \\ +\int_\sigma^\infty (1+C_{OT})^{1-\frac{1}{\gamma}}\,e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t} \\ \cdot\frac{\left(\kappa_{n,\gamma,\epsilon}({}_tp_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^{1}\left(1-\left(1-{}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\,d\varphi\right)^{1/\gamma-1}}\,dt\Bigg)\Bigg)^{\gamma}\end{aligned} \tag{3.30}$$

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

$$\begin{aligned}U_{[\sigma]} = \mathbb{E}\Bigg[\int_0^\infty e^{-\rho t}\Bigg(\mathbb{1}_{\{0\le t<\min\{\sigma,T_\epsilon\}\}}u\left(c^*_{[\sigma]}(t)\right) \\ +\mathbb{1}_{\{\sigma\le t<T_\epsilon\}}u\left(\frac{n}{N_\epsilon(t)}d^*_{[\sigma]}(t)\right)\Bigg)\,dt\Bigg] = \frac{\lambda_{[\sigma]}}{1-\gamma}v.\end{aligned} \tag{3.31}$$

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

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

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

Similar to the case of the tonuity, we subsequently briefly analyze the optimal payoff of the antine graphically when applying the base case parameter setup specified in Table 1. The corresponding optimal switching time of the antine is given by $\sigma^* = 0$ as the highest lifetime utility is attained at this time

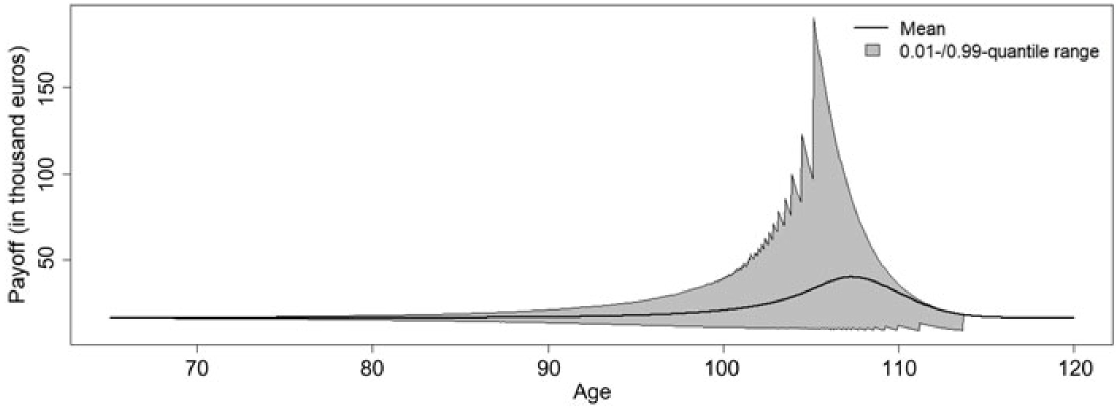


FIGURE 2: Mean and 0.01-/0.99-quantile range of the optimal payoff of the antine with $\sigma^* = 0$ depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05. The plot is based on the assumption that the considered policyholder is always alive.

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

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

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

3.3. Portfolio

In contrast to the optimization problems for the tonuity and the antine, the optimization problem for the case with a portfolio consisting of an annuity and a tontine cannot be solved explicitly. It can be written as:

$$\max_{(c_{AT}(t),\, d_{AT}(t))_{t\in[0,\infty)}} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\mathbb{1}_{\{T_\epsilon>t\}}u\left(c_{AT}(t)+\frac{n}{N_\epsilon(t)}d_{AT}(t)\right)\,dt\right] \quad \text{subject to} \quad v=\widetilde{P}_0^{AT}=\widetilde{P}_0^{A,AT}+\widetilde{P}_0^{OT,AT}. \tag{3.33}$$

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

$$\begin{aligned}\mathcal{L}&=\int_0^\infty e^{-\rho t}\mathbb{E}\left[\mathbb{1}_{\{T_\epsilon>t\}}u\left(c_{AT}(t)+\frac{n}{N_\epsilon(t)}d_{AT}(t)\right)\right]dt+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right)\\
&=\int_0^\infty e^{-\rho t}\mathbb{E}\left[{}_tp_x^{1-\epsilon}\mathbb{E}\left[u\left(c_{AT}(t)+\frac{n}{N_\epsilon(t)}d_{AT}(t)\right)\middle|T_\epsilon>t,\epsilon\right]\right]dt\\
&\quad+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right)\\
&=\int_0^\infty e^{-\rho t}\mathbb{E}\Bigg[{}_tp_x^{1-\epsilon}\sum_{k=0}^{n-1}u\left(c_{AT}(t)+\frac{n}{k+1}d_{AT}(t)\right)\binom{n-1}{k}\left({}_tp_x^{1-\epsilon}\right)^k\\
&\qquad\cdot\left(1-{}_tp_x^{1-\epsilon}\right)^{n-1-k}\Bigg]dt+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right)\\
&=\int_0^\infty e^{-\rho t}\sum_{k=0}^{n-1}u\left(c_{AT}(t)+\frac{n}{k+1}d_{AT}(t)\right)\binom{n-1}{k}\int_{-\infty}^1\left({}_tp_x^{1-\varphi}\right)^{k+1}\left(1-{}_tp_x^{1-\varphi}\right)^{n-1-k}\\
&\quad\cdot f_\epsilon(\varphi)\,d\varphi\,dt+\lambda_{AT}\left(v-\widetilde{P}_0^{A,AT}-\widetilde{P}_0^{OT,AT}\right),\end{aligned} \tag{3.34}$$

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

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial c_{AT}(t)}&=e^{-\rho t}\sum_{k=0}^{n-1}u'\left(c_{AT}(t)+\frac{n}{k+1}d_{AT}(t)\right)\binom{n-1}{k}\\
&\qquad\cdot\int_{-\infty}^1\left({}_tp_x^{1-\varphi}\right)^{k+1}\left(1-{}_tp_x^{1-\varphi}\right)^{n-1-k}\\
&\qquad\cdot f_\epsilon(\varphi)\,d\varphi-\lambda_{AT}(1+C_A)e^{-rt}{}_tp_x\,m_\epsilon(-\ln{}_tp_x)\stackrel{!}{=}0,\end{aligned} \tag{3.35}$$

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

and

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

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

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

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

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

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

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

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

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

Proof. See Appendix A.2. □

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

TABLE 2

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

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

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

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

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

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

$$c^*_{AT}(t) + \frac{n}{N_\epsilon(t)} d^*_{AT}(t). \tag{3.40}$$

For a better understanding of the curve progressions in Figure 3(a), we additionally depict, in Figure 3(b), the related optimal payoff functions $c^*_{AT}(t)$ and $d^*_{AT}(t)$ of the portfolio. Here, we can see that, particularly in the first half of the considered age range, the optimal tontine payoff function $d^*_{AT}(t)$ decreases in age and hence behaves similarly to what, for example, Milevsky and Salisbury (2015) find for optimally designed tontines in their framework. In the second half, the values for $d^*_{AT}(t)$ are rather close to zero and do not seem to differ significantly anymore. Note that the payoff of the tontine in Figure 3(b) is *not* the payoff to a single individual and has to be multiplied by $\frac{n}{N_\epsilon(t)}$. Especially at extremely old ages, $d^*_{AT}(t)$ is scaled by the factor n which explains why the steep decline in annuity payments occurs at extremely old ages and why $d^*_{AT}(t)$ is not exactly zero at these old ages. Note that this payoff structure can only occur because the payoff $nd^*_{AT}(t)$ is guaranteed to the pool of policyholders by the insurer. Concerning the optimal annuity payoff, we note that, after quite a long time of playing no role for the optimal portfolio payoff at all, this payoff is first drastically increasing and then decreasing, opposed to, for instance, a constant

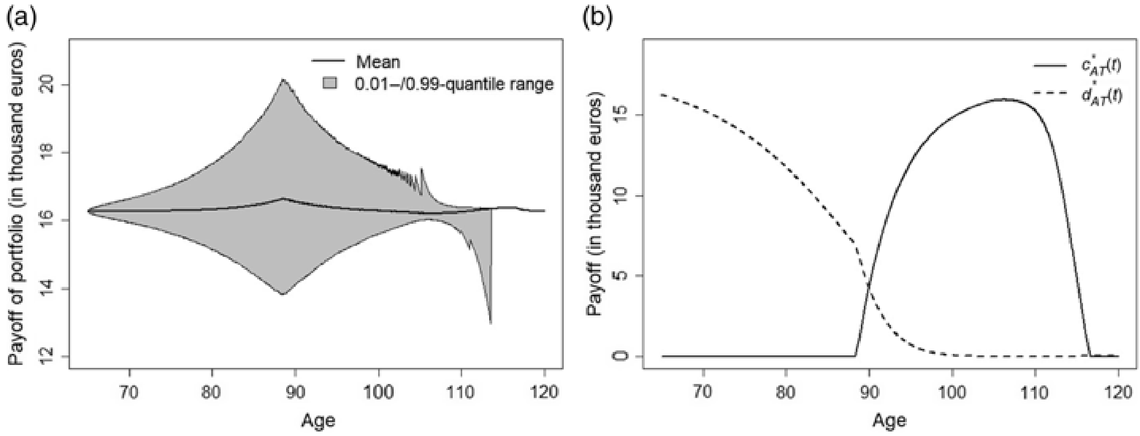


FIGURE 3: (a) Mean and 0.01-/0.99-quantile range of the optimal payoff of the portfolio consisting of an annuity and a tontine depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05. The plot is based on the assumption that the considered policyholder is always alive. (b) Optimal payoff functions of the portfolio consisting of an annuity and a tontine depending on age. The parameters are chosen as in Table 1 and the constant discretization step size of the age range as 0.05.

annuity which turns out to be optimal in the tonuity and the antine if $\rho = r$. We therefore detect in Figure 3(b) that the policyholder defers the annuitization for some time before aiming at obtaining annuity payments at an increasing (and later on decreasing) rate, whereas the tontinization is desired right at the beginning of her retirement with a decreasing (and later on slightly increasing) rate. The reason for this structure is that the annuity provides secure payments at times when the tontine provides the most volatile payments (see also Figure 2). Right after contract initiation, the number of the tontine members is fairly stable. At extremely advanced retirement ages, the tontine payments are, again, stable, because there is only a very low probability for any other survivor to be left. Therefore, at these ages, no stable annuity payments are required. These are only necessary for ages between (in this example approximately) 89 and 116 years, where most policyholders die and, consequently, the payments of the tontine are most volatile. It is important to bear in mind that, as we consider a very general approach for the portfolio optimization problem, the individual can freely decide on the payoff structures of both the annuity and the tontine in the portfolio. Hence, at the optimum, it is possible to obtain such payoff functions as $c^*_{AT}(t)$ in Figure 3(b). In our result, it seems that $c^*_{AT}(t)$ and $d^*_{AT}(t)$ complement each other. As these payoffs are completely determined by solely maximizing the benefits of the policyholder, they might not be provided in practice. However, our analysis might still be interesting for those insurers who aim to offer products which best suit the needs of the customers to improve their competitiveness on the market. Concerning Figure 3(a), we overall make the following further observations: For a long time, there is always a certain degree of uncertainty in the optimal portfolio payoff. As the tontine, which involves randomness, prevails at earlier ages, Figure 3(a) actually resembles Figure 1 in the first several years following the withdrawal from working life. Afterward, the payoff uncertainty declines over the years as the annuity, which stabilizes the payments, becomes more and more significant. From the age of around 107 years until the age of around 114 years, we see that the chances to receive quite low payments, compared to the expectation, grow again. The decrease in $c^*_{AT}(t)$ and the slight increase in $d^*_{AT}(t)$ in this age range, together with the presumption that, in contrast to what is expected, other policyholders in the tontine pool could nevertheless be alive at such high ages, can give reasons for this observation. After the age of around 114 years, it is however highly probable that the event that all the other policyholders are dead occurs, which leads to a virtually deterministic payoff for the remaining time.

4. Comparison of Retirement Plans

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

4.1. Theoretical findings

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\widetilde{P}_0^{AT} = &\mathbb{E}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon>t\}}c_{AT}(t)\,dt\right] + C_A\mathrm{Var}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon>t\}}c_{AT}(t)\,dt\right]\\ &+\mathbb{E}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon>t\}}\frac{n}{N_\epsilon(t)}d_{AT}(t)\,dt\right] + C_{OT}\\ &\cdot\mathrm{Var}\left[\int_0^\infty e^{-rt}\mathbb{1}_{\{T_\epsilon>t\}}\frac{n}{N_\epsilon(t)}d_{AT}(t)\,dt\right],\end{aligned} \tag{4.45}$$

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

4.2. Numerical findings

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

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

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

or equivalently,

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

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

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

TABLE 3

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

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

The other parameters are taken from Table 1.

TABLE 4

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

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

The other parameters are taken from Table 1.

TABLE 5

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

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

The other parameters are taken from Table 1.

obtain 0.06%. Since we would like to keep our numerical analysis as simple as possible, we have decided to round the values in Chen *et al.* (2019) instead of taking the exact values. Note that the parameters in Chen *et al.* (2019) also differ slightly from ours and therefore, we only focus on the rougher magnitude when it comes to the risk loadings. Additionally, the certainty equivalents of the portfolio consisting of an annuity and a tontine along with the optimal fraction of initial wealth invested in the annuity are provided for the different pool sizes n and different tontine risk loadings C_{OT}. The optimal fraction of initial wealth invested in the annuity can be determined as the gross premium of the annuity (computed from the optimal annuity payoff) divided by the total initial wealth. For the remaining parameters besides the pool size and the tontine risk loadings, we use the base case parameter setup as given in Table 1. Table 4 provides similar sensitivity analyses of the certainty equivalents for the risk loading of the annuity C_A. Finally, Table 5 provides similar sensitivity analyses of the certainty equivalents for the risk aversion parameter γ. Note that we have not included sensitivity analyses with respect to the subjective discount rate ρ in these numerical studies as the effects of ρ seem negligible for the given parameters. Overall, we make the following observations which are all drawn *only within our exemplary numbers.* To support the readability of our results, we have highlighted the main observations in italic letters:

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

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

- *The role of the annuity component becomes more prominent within the retirement plans if the risk aversion level of the policyholder increases.* This feature can be observed in Table 5: If the risk aversion increases, the optimal switching time of the tonuity decreases. Furthermore, concerning the portfolio, individuals with a larger risk aversion tend to invest higher fractions of their initial wealth in the annuity.
- *Referring to the portfolio, the policyholder does not invest all her initial wealth exclusively in an annuity or a tontine.* From the fact that the fractions of initial wealth invested in the annuity are above 0 and below 1, within our exemplary numbers, neither a pure annuity nor a pure tontine is an attractive retirement plan for the majority of the policyholders. Instead, partial annuitization combined with partial tontinization turns out to deliver the highest expected lifetime utility for our parameter choices. This can also be seen from Proposition 3.3 and from Table 2, where the critical annuity loading C_A^{crit} is always larger than the considered values for C_A.

5. Conclusion

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

While in this article we exclusively focus on the policyholder's perspective, an interesting topic for future research might be an analysis of the three considered retirement plans from the insurer's perspective. As it is less natural to assume a utility function for the insurance company, we could consider other important quantities of interest from the insurer's perspective, like the (random) present value of future losses (cf. Bauer and Weber, 2008; Li and Hardy, 2011; Cairns, 2013; Kling *et al.*, 2014; Olivieri and Pitacco, 2019). In a recent article, Chen and Rach (2019), the authors analyze the attractiveness of options on tontines also from the insurer's point of view. The results obtained there suggest that insurers can also benefit from selling hybrid products between conventional annuities and tontines. Compared to annuities, it is likely that the tonuity, antine and the portfolio of an annuity and a tontine also reduce potential losses on the insurer's side, as all these retirement plans carry less risks for the insurer than a traditional annuity.

Acknowledgments

Manuel Rach acknowledges the financial support given by the DFG for the research project "Zielrente: die Lösung zur alternden Gesellschaft in Deutschland".

Notes

1. In addition to the traditional tontine, many innovative, longevity-risk-sharing retirement products, for instance, pooled annuity funds or group self-annuitization schemes, have been developed in recent years (see, e.g., Bernhardt and Donnelly, 2019 and the references therein).
2. Note that problem (3.33) is more general than the following optimization problem:

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

where $c^*_{[0]}(t)$ is the optimal annuity payoff and $d^*_{[\infty]}(t)$ is the optimal tontine payoff as obtained from Theorem 3.1 for the boundary cases $\tau=0$ and $\tau=\infty$, respectively. Note that for the annuity case, that is, $\tau=0$, the budget constraint is $v=\widetilde{P}^A_0$. Similarly, for the tontine case, that is, $\tau=\infty$, the budget constraint is $v=\widetilde{P}^{OT}_0$. In other words, in problem (5.48), where we maximize over a fraction of initial wealth, we do not have to consider a budget constraint, as $\alpha\widetilde{P}^A_0+(1-\alpha)\widetilde{P}^{OT}_0=\alpha v+(1-\alpha)v=v$ holds for all $\alpha\in[0,1]$. Note that this describes already the first difference between optimization problem (5.48) and the optimization problem (3.33). Furthermore, as the optimization problem in (5.48) is, by assuming the payoff functions $c^*_{[0]}(t)$ and $d^*_{[\infty]}(t)$ to be given in advance, (a lot) less general than the portfolio problem (3.33), we observe that under problem (5.48), it is no longer possible to replicate the payoff of any tonuity and antine. In this problem, the individual can only decide upon the fraction α and not upon the payment structure of the products in the portfolio. As a consequence, it is possible that the portfolio of an annuity and a tontine can be outperformed by other combinations, for example, by a tonuity.

References

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

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

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

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

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

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

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

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

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

GOMPERTZ, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513–583.

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

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

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

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

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

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

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

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

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

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

MILEVSKY, M.A. and SALISBURY, T.S. (2015) Optimal retirement income tontines. *Insurance: Mathematics and Economics,* **64**, 91–105.

MILEVSKY, M.A. and SALISBURY, T.S. (2016) Equitable retirement income tontines: Mixing cohorts without discriminating. *ASTIN Bulletin: The Journal of the IAA*, **46**(3), 571–604.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

APPENDIX A. PROOFS

A.1. Proof of Theorem 3.1

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

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

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

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

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

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

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

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

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

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

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

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

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

As a consequence, we obtain

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

The expected discounted lifetime utility is then given by

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

□

A.2. Proof of Proposition 3.3

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

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

Regarding condition (3.36), we obtain:

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

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

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

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

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

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

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

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

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

$$= e^{-\rho t}\left(d^*_{[\infty]}(t)\right)^{-\gamma}\kappa_{n,\gamma,\epsilon}({}_tp_x) - \lambda_{[\infty]}(1+C_{OT})e^{-rt}\int_{-\infty}^{1}\left(1-\left(1-{}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\,d\varphi$$
$$= 0.$$

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

□

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

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

1. Initialize n, γ, r, ρ, v, x, m, β, tol, C_A, C_{OT}.
2. Specify a grid of time points $t_1, \ldots, t_N$, where $t_1 = 0$ and t_N lies sufficiently far ahead in the future.
3. Choose upper and lower bounds λ_u and λ_l for $\lambda = \lambda_{AT}$ and set $\lambda = \frac{1}{2}(\lambda_u + \lambda_l)$.
4. While $\left|v - \widetilde{P}_0^{A,AT} - \widetilde{P}_0^{OT,AT}\right| > tol$
   - (a) Choose upper and lower bounds $d_u(t_i)$ and $d_l(t_i)$ for $i = 1, \ldots, N$. Set $d_{AT}(t_i) = \frac{1}{2}(d_u(t_i) + d_l(t_i))$.
   - (b) While $\max_i \left|\frac{\partial \mathcal{L}}{\partial d_{AT}(t_i)}\right| > tol$
     - i. Choose upper and lower bounds $c_u(t_i)$ and $c_l(t_i)$ for $i = 1, \ldots, N$. Set $c_{AT}(t_i) = \frac{1}{2}(c_u(t_i) + c_l(t_i))$.
     - ii. While $\max_i \left|\frac{\partial \mathcal{L}}{\partial c_{AT}(t_i)}\right| > tol$
       - A. For all i with $\frac{\partial \mathcal{L}}{\partial c_{AT}(t_i)} > 0$, set $c_l(t_i) = c_{AT}(t_i)$.
       - B. For all i with $\frac{\partial \mathcal{L}}{\partial c_{AT}(t_i)} < 0$, set $c_u(t_i) = c_{AT}(t_i)$.
       - C. Set $c_{AT}(t_i) = \frac{1}{2}(c_u(t_i) + c_l(t_i))$.
     - iii. For all i with $\frac{\partial \mathcal{L}}{\partial d_{AT}(t_i)} > 0$, set $d_l(t_i) = d_{AT}(t_i)$.
     - iv. For all i with $\frac{\partial \mathcal{L}}{\partial d_{AT}(t_i)} < 0$, set $d_u(t_i) = d_{AT}(t_i)$.
     - v. Set $d_{AT}(t_i) = \frac{1}{2}(d_u(t_i) + d_l(t_i))$.
   - (c) Once $c_{AT}(t_i)$ and $d_{AT}(t_i)$ are computed for $i = 1, \ldots, N$, we can linearly interpolate between these values. This enables us to compute $\widetilde{P}_0^{A,AT}$ and $\widetilde{P}_0^{OT,AT}$ by numerical integration.
   - (d) If $v - \widetilde{P}_0^{A,AT} - \widetilde{P}_0^{OT,AT} > 0$, set $\lambda_u = \lambda$.

(e) If $v - \widetilde{P}_0^{A,AT} - \widetilde{P}_0^{OT,AT} < 0$, set $\lambda_l = \lambda$.
(f) Set $\lambda = \frac{1}{2}(\lambda_u + \lambda_l)$.

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

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

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

3 Optimal retirement products under subjective mortality beliefs

Source:
Chen, A., Hieber P., and Rach, M. (2019). Optimal retirement products under subjective mortality beliefs. Submitted to special issue of *Insurance: Mathematics and Economics* on behavioral insurance (under review). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3287699

Optimal retirement products under subjective mortality beliefs

An Chen*, Peter Hieber* and Manuel Rach*

* Institute of Insurance Science, Ulm University, Helmholtzstr. 20, 89069 Ulm, Germany.
E–mails: an.chen@uni-ulm.de; peter.hieber@uni-ulm.de; manuel.rach@uni-ulm.de

Abstract

Many empirical studies confirm that policyholder's subjective mortality beliefs deviate from the information given by publicly available mortality tables. In this study, we look at the effect of subjective mortality beliefs on the perceived attractiveness of retirement products, focusing on conventional annuities and tontines (where a pool of policyholders shares the longevity risk). Given actuarially fair pricing with no subjective mortality beliefs (that is, the insurer's and the policyholder's perceptions coincide), annuities yield higher lifetime utility than tontines (see also Milevsky and Salisbury (2015)). Staying in an actuarially fair pricing framework, we find that this result might be reversed if the policyholder's subjective survival probabilities for her peers are lower than the ones used by the insurance company. We prove that, assuming such subjective beliefs, there exists a critical tontine pool size from which on the tontine is always preferred over the annuity.

Keywords: Behavioral insurance, subjective mortality beliefs, optimal retirement product design, tontine, annuity

JEL: G22, D81

1 Introduction

The current low interest rate environment and the increasing life expectancy are challenging the life insurance industry. Both insurance companies and retirees are searching for "desirable" retirement products. A product which has recently gained a lot of attention in this context is the so-called *tontine*. It used to be a popular source of retirement income from the 17th to the 19th century (see for example Milevsky (2014, 2015), Milevsky and Salisbury (2015, 2016), Weinert and Gründl (2017), Chen et al. (2019) and Li and Rothschild (2019)).[1] When priced actuarially fair, life annuities give retirees greater lifetime utility than tontines (see also Milevsky and Salisbury (2015)). When realistic safety loadings or risk margins are taken into consideration, tontines can be preferred to annuities (cf. Milevsky and Salisbury (2015) and Chen et al. (2019)). In the present article, we remain in the actuarially fair pricing framework and analyze the optimal retirement decision under subjective mortality beliefs. In particular, we aim to find out whether, under these subjective beliefs, tontines generate a higher lifetime utility level than annuities.

It is well-documented in the literature that individuals tend to have subjective beliefs about their own and others' life expectancy. An over- or underestimation of this will strongly affect the perceived attractiveness of a retirement product, as it has already been noted by Adam Smith back in 1776 (cf. Smith (1776)). An extensive literature review of empirical findings regarding these subjective mortality beliefs is provided in Section 2 of this article. In the majority of these articles, people were asked to estimate their life expectancy or probability of survival towards a certain age and the results were compared to some reference data as, for example, the estimates of the Government Actuary's department. Based on the empirical findings presented in Section 2, people at younger ages tend to be pessimistic about their future lifetime, while, at older ages, various studies document both under- and overestimations of the life expectancy and survival probabilities. In our model setup, we assume that the insurer uses the best-estimate survival curve for pricing and we fix this survival curve as reference curve for the subjective mortality beliefs of policyholders. That is, optimism and pessimism regarding one's future lifetime are assessed with respect to the best-estimate survival curve which the insurer uses for pricing. Additionally, people seem to have subjective beliefs about the life expectancy of their peers, which might differ from the perception about their own life expectancy. Therefore, we, in our model, allow a single individual to (additionally) believe that she lives relatively longer or shorter than her peers. To be precise, we assume in our model that there are three different survival curves for any x-year old policyholder: For any time $t \geq 0$, we

[1] In recent years, many products with a tontine-like structure have appeared. They are often called pooled annuity funds or group self-annuitization, and many efforts have been made in recent years to explore the potential and optimal design of these products in today's world, see, for example, Piggott et al. (2005), Valdez et al. (2006), Stamos (2008), Qiao and Sherris (2013), Donnelly et al. (2013, 2014) and Donnelly (2015).

consider the t-year survival curve ${}_tp_x$ used by the insurer and two subjective survival curves (${}_t\tilde{p}_x$, ${}_t\hat{p}_x$), representing the policyholder's subjective mortality beliefs. Here, we distinguish between the survival curves the individual assumes for herself (denoted by ${}_t\tilde{p}_x$) and the ones assumed by the individual for other policyholders (denoted by ${}_t\hat{p}_x$).

In the present article, we analyze the impact of subjective mortality beliefs on the optimal retirement decision in an expected utility framework, taking tontines and annuities as examples. If actuarially fair premiums are adopted to price annuities and tontines, annuities are preferred to tontines (Milevsky and Salisbury (2015)). Following Milevsky and Salisbury (2015) and Chen et al. (2019), we derive the optimal payout functions of an annuity and a tontine under subjective mortality beliefs. We prove that a tontine's attractiveness increases, the more a policyholder underestimates the survival probabilities of other policyholders, while the annuity remains unaffected by this underestimation. Moreover, we find that policyholders who systematically underestimate others' survival curves compared to the survival curves assumed by the insurer (${}_t\hat{p}_x < {}_tp_x$) prefer a tontine with a sufficiently large pool size over an annuity, even if both products are fairly priced. We show that, under a certain condition, there is a critical pool size N_0 beyond which tontines are preferred over annuities. In our numerical analysis, we also find that whether the policyholder believes that she lives longer (${}_t\hat{p}_x < {}_t\tilde{p}_x$) or shorter (${}_t\hat{p}_x > {}_t\tilde{p}_x$) than her peers does not seem to have a substantial impact on the choice between a tontine and an annuity, particularly when the pool size of the tontine is large. Instead, the subjective belief regarding one's own survival curve (${}_t\tilde{p}_x$) influences the optimal payoff structure of a tontine.

It is also well-acknowledged in the literature that *annuities* seem overpriced for an individual who is pessimistic about her life expectancy (see, for example, Wu et al. (2015)). Our model is consistent with this observation as a pessimistic individual who underestimates her own survival curve relative to that used by the insurer (${}_t\tilde{p}_x < {}_tp_x$) has a different perception regarding the premium charged. An analysis in an expected utility framework supports this effect and shows that a lower expected utility level results if the policyholder perceives the premiums charged for annuities as "too high". Conversely, the policyholder's utility increases if she perceives the premium charged as "too low". We conduct the same analysis for *tontines* and find that the perception of the premium of a tontine, on the other hand, is hardly affected by subjective mortality beliefs, particularly when the pool size is large. As a consequence, subjective mortality beliefs concerning one's own mortality affect the expected lifetime utility of tontines to a rather low extent, in contrast to annuities.

The results of this paper have an interesting implication with respect to the annuity puzzle which is a term used to describe the discrepancy between the theoretical demand for annuities (see, for instance, Yaari (1965) and Peijnenburg et al. (2016)) and the fact that only few households voluntarily purchase an annuity (see, for example, Hu and Scott (2007), Inkmann et al. (2010) and Lockwood (2012)).[2] In this article, we connect to the literature on the annuity puzzle by briefly showing that annuities are perceived as overpriced from a pessimistic policyholder's perspective (who underestimates her survival curve), see, for example, Wu et al. (2015). This result can, to some extent, help explain the observed low demand for annuities. Our article differs from the majority of the literature on the annuity puzzle by the inclusion of tontines and analyses how subjective mortality beliefs affect the relative attractiveness of annuities and tontines.

The remainder of this article is organized as follows: In Section 2, we provide a literature review regarding subjective mortality beliefs. In Section 3, we describe the general model setup used throughout this article. After that, we derive the optimal payout function of the annuity and tontine for a risk-averse policyholder under subjective mortality beliefs in Section 4. We also derive closed-form expressions for the individual's expected discounted lifetime utility from each product, which will enable us to compare the attractiveness of the different products. In Section 5, we analyze the effects of subjective mortality beliefs on the optimal retirement decision. Section 6 concludes the article. Proofs are collected in the Appendix.

2 Subjective mortality beliefs

The phenomenon that people tend to have their own, subjective beliefs regarding their own and others' life expectancy is not new in the literature. Such a phenomenon is of major relevance in life insurance. For example, Bauer et al. (2014) analyze the effects of differing perceptions of mortality on the life settlement market. Individuals might systematically over- or underestimate their own and others' life expectancy, affecting their willingness to buy retirement products like annuities and tontines. Important empirical findings regarding subjective mortality beliefs include, but are not limited to, the following:

- Bucher-Koenen et al. (2013) find that, in Germany, "men as well as women are pessimistic about their life expectancy. Women (men) underestimate their life span by about 7 (6.5) years compared to the official records by the German statistical office." The sample consists of an equal share of males and females aged 26-60.

[2] There is already vast literature exploring the main drivers for this puzzle. For reviews of this literature, we refer the interested reader, for example, to Milevsky (2013) and Benartzi et al. (2011). Further studies related to our article in the context of behavioral insurance are, for example, Salisbury and Nenkov (2016), Chen et al. (2016, 2018), Poppe-Yanez (2017), Caliendo et al. (2017) and O'Dea et al. (2019). Note that there is more than one puzzle in life insurance, see for example Gottlieb (2012). However, the puzzle dealing with underannuitization is probably the most famous one.

- According to O'Brien et al. (2005), individuals in Great Britain underestimate their life expectancies "by 4.62 years (males), 5.95 years (females) compared with the estimates of the Government Actuarys Department". Additionally, "people are optimistic: they think they will live longer, on average, than people of their own age and sex: by 1.19 years (males), 0.76 years (females)." The sample covers ages from 16 to 99. While the underestimation is larger for young than old people, there is still an underestimation of 2.83 years for males and 4.62 years for females at ages 60-69 which is the range of typical retirement ages.

- In Greenwald and Associates (2012) the following results about American citizens are established: "When asked to estimate how long the average person their age and sex can expect to live, more than six in ten retirees (62 percent) and half of pre-retirees (57 percent) provide a response that is below the average. Only about one-quarter overestimate average life expectancy (19 percent of retirees and 28 percent of pre-retirees)."[3] Additionally, a similar observation as in O'Brien et al. (2005) is made: "Despite the tendency to underestimate population life expectancy, half of retirees (50 percent) and pre-retirees (53 percent) appear to believe that the response they provide for their personal life expectancy is within one year of average life expectancy. Three in ten think their estimate of personal life expectancy exceeds average life expectancy (31 percent of retirees and 32 percent of pre-retirees)."

- Wu et al. (2015) find that "respondents are pessimistic about overall life expectancy but optimistic about survival at advanced ages, and older respondents are more optimistic than younger." To be precise, "younger cohorts underestimate survival (the 50-54 age group underestimates life expectancy by more than eight years) while older cohorts tend to overestimate, especially males (Ludwig and Zimper (2013)). (Males in the 70-74 age group overestimate life expectancy by only one year, and females underestimate it by one year.)" These observations are based on the Retirement Plans and Retirement Incomes: Pilot Survey, conducted in May 2011 for Australian citizens.

- Elder (2013) analyzes the Health and Retirement study (HRS) which is a longitudinal survey of American citizens above age 50. The most important finding for our article is that that both men and women under age 65 underestimate the probability of survival to age 75, but overestimate the probability of survival to age 85. A similar observation is made in Hurd and McGarry (2002) who analyze the HRS as well.

There seems to be a clear tendency for younger people to underestimate their life expectancy and survival probabilities, while both under- and overestimations can be observed at older ages.

[3] "Respondents were classified as retirees if they described their employment status as retiree, had retired from a previous career, or were not currently employed and were either age 65 or older or had a retired spouse. All other respondents were classified as pre-retirees."

Additional literature on this subject can, for example, be found in Wu et al. (2015). Furthermore, Payne et al. (2013) emphasize that individuals' responses to questions assessing their subjective mortality and longevity beliefs drastically depend on the framing of the question. Therefore, we consider a general model for subjective mortality beliefs which allows for both under- and overestimations of the life expectancy. We assume that the insurer uses the best-estimate survival curve for pricing purposes. That is, if an individual under- or overestimates her own or others' life expectancy, this means, in our model, that the individual uses a lower or higher survival curve, respectively, than the insurer does.

3 Model setup

In this section, we describe the basic model setup used throughout the remainder of our article. In particular, we explain how the mortality and the subjective mortality beliefs are modeled and how the two retirement products under consideration, the annuity and the tontine, are designed. We ignore financial market risk to solely focus on the longevity risk. We start by introducing the rather simple stochastic mortality model applied throughout the article before getting to the retirement products and their actuarially fair premiums (determined within the described mortality model).

3.1 A simple stochastic mortality model

There are two different kinds of mortality risk: Unsystematic, or idiosyncratic, mortality risk, stems from the lifetimes of people being unknown but still following a certain mortality law. Systematic, or aggregate mortality risk stems from the fact that we cannot certainly determine the actual ("true") mortality law. In the context of retirement products, this risk is also called longevity risk. Further explanations regarding these two different aspects of mortality risk can also be found, for instance, in Piggott et al. (2005). For any x-year-old policyholder, the best-estimate t-year survival probability applied by the insurer is denoted by ${}_tp_x$, which can be computed from some continuous-time mortality law, obtained from, e.g., publicly available mortality tables. The insurer uses this survival curve for pricing the insurance contracts introduced below.

As pointed out above, individuals tend to have their own, subjective estimates of others' and their own life expectancy. These subjective mortality beliefs will be incorporated by assuming that the individual considered and the insurance company use different mortality tables. While we denote by ${}_tp_x$ the insurer's best-estimate survival curve, we assume that an individual believes her own survival curve to be ${}_t\tilde{p}_x$ and the survival curve for the other policyholders to be ${}_t\hat{p}_x$. As mentioned in the introduction, we do not assume any "fixed" relation between these survival curves as both under- and overestimation of the actual life expectancy can be observed among retirees. From the policyholder's perspective, these subjective survival curves are applied to evaluate the attractiveness of the retirement products.

We incorporate uncertainty in this mortality law in a similar way as, for example, Lin and Cox (2005) by applying a random shock ϵ to the survival curves such that the shocked survival curves are given by ${}_tp_x^{1-\epsilon}$, ${}_t\tilde{p}_x^{1-\epsilon}$ and ${}_t\hat{p}_x^{1-\epsilon}$, respectively. For simplicity, we assume that the same shock ϵ will be applied to the three survival curves. This shock covers the systematic mortality risk described above. We assume that ϵ is a continuous random variable with density $f_\epsilon(\cdot)$ and support on $(-\infty, 1)$. The special case in which no shock is applied is obtained by setting $\epsilon = 0$.

3.2 Retirement products

We consider two different retirement products, the annuity and the tontine. Following Yaari (1965), we assume a continuous-time stream of income for the retirement products. In an **annuity** contract, any policyholder continuously receives an annuity payment $c(t)$ until death. Denoting by ζ_ϵ the remaining future lifetime of the policyholder, where ϵ is the random longevity shock as defined in the previous section, the payment stream of the annuity can be written as

$$b_A(t) := \mathbb{1}_{\{\zeta_\epsilon > t\}}\, c(t)\,. \tag{1}$$

The actuarially fair premium charged by the insurer (using the survival curve ${}_tp_x$) can be obtained as

$$\begin{aligned} P_0^A &= \mathbb{E}\left[\int_0^\infty e^{-rt} b_A(t)\mathrm{d}t\right] \\ &= \int_0^\infty e^{-rt}\mathbb{E}\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right]\, c(t)\,\mathrm{d}t \\ &= \int_0^\infty e^{-rt} c(t) \int_{-\infty}^1 {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\,\mathrm{d}\varphi\,\mathrm{d}t \\ &= \int_0^\infty e^{-rt}\, {}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)\, c(t)\,\mathrm{d}t, \end{aligned} \tag{2}$$

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

While in an annuity, the longevity risk is borne by the insurance company, in a **tontine** contract it is shared among a pool of $n \geq 1$ homogeneous policyholders, who are assumed to be identical copies of each other.[4] While the unsystematic mortality risk can initially be diversified by choosing the pool size n large enough, the systematic risk cannot, as this type of risk affects the population as a whole. At older ages, however, the pool size will decrease, leaving the remaining policyholders with both systematic and unsystematic mortality risk. Denoting by $N_\epsilon(t)$ the number of policyholders alive at time t, each policyholder receives $n/N_\epsilon(t)$ multiplied by a payment stream $d(t)$ specified at the beginning of the contract. Following Milevsky and Salisbury (2015), this yields the following continuous payment stream for each $t > 0$:

$$b_T(t) := \begin{cases} \mathbb{1}_{\{\zeta_\epsilon > t\}} \frac{nd(t)}{N_\epsilon(t)}, & \text{if } N_\epsilon(t) > 0, \\ 0, & \text{else} \end{cases}. \tag{3}$$

Note that, in contrast to the annuity payment (1), the tontine payment depends substantially on the number of surviving policyholders $N_\epsilon(t)$. In the special case where the pool consists of only one member, that is, $n = 1$, and $c(t) = d(t)$, the tontine payoff (3) and the annuity payoff (1) coincide. Note that, given a surviving individual, the number of pool members follows a binomial distribution, that is, $(N_\epsilon(t) - 1 \mid \zeta_\epsilon > t, \epsilon) \sim \text{Bin}\left(n-1, {}_tp_x^{1-\epsilon}\right)$ from the insurer's point of view. In particular, the insurer assumes the same survival probabilities for all individuals in the pool. The actuarially fair premium of this contract can then be obtained as

$$\begin{aligned} P_0^T &= \mathbb{E}\left[\int_0^\infty e^{-rt} b_T(t) \mathrm{d}t\right] \\ &= \int_0^\infty e^{-rt} \mathbb{E}\left[{}_tp_x^{1-\epsilon} \mathbb{E}\left[\frac{nd(t)}{N_\epsilon(t)} \;\middle|\; \zeta_\epsilon > t, \epsilon\right]\right] \mathrm{d}t \\ &= \int_0^\infty e^{-rt} \mathbb{E}\left[\sum_{k=0}^{n-1} \frac{n}{k+1} \binom{n-1}{k} \left({}_tp_x^{1-\epsilon}\right)^{k+1} \left(1 - {}_tp_x^{1-\epsilon}\right)^{n-1-k}\right] d(t)\, \mathrm{d}t \\ &= \int_0^\infty e^{-rt} \mathbb{E}\left[\sum_{k=1}^{n} \binom{n}{k} \left({}_tp_x^{1-\epsilon}\right)^{k} \left(1 - {}_tp_x^{1-\epsilon}\right)^{n-k}\right] d(t)\, \mathrm{d}t \\ &= \int_0^\infty e^{-rt} \mathbb{E}\left[1 - \left(1 - {}_tp_x^{1-\epsilon}\right)^n\right] d(t)\, \mathrm{d}t \\ &= \int_0^\infty e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_tp_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, \mathrm{d}\varphi\, d(t)\, \mathrm{d}t\,. \end{aligned} \tag{4}$$

As already pointed out in Chen et al. (2019), the actuarially fair premium for tontines (4) differs from the formula in Milevsky and Salisbury (2015), where it is assumed that the payoff to the pool $d(t)$ will always be provided by the insurer even when there are no policyholders left. The premium in (4) converges to the one in Milevsky and Salisbury (2015) if the pool size n tends to infinity.

[4] Strictly speaking, the insurer carries the longevity risk of the last living person in the pool.

3.3 Subjective perception of the premium

The insurer charges a premium based on its own best-estimate survival probabilities (see (2) and (4)). Let us now examine how this premium is perceived from the policyholder's point of view. As the policyholder has different mortality beliefs than the insurer, she might perceive a product as over- or underpriced. We denote the expected value operator under the subjective beliefs of a policyholder by $\widetilde{\mathbb{E}}[\cdot]$. The subjective premium of the annuity, using the policyholder's subjective survival curve, is given by

$$\begin{aligned}
\widetilde{P}_0^A &= \widetilde{\mathbb{E}}\left[\int_0^\infty e^{-rt} b_A(t)\mathrm{d}t\right] \\
&= \int_0^\infty e^{-rt}\widetilde{\mathbb{E}}\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] c(t)\,\mathrm{d}t \\
&= \int_0^\infty e^{-rt} c(t) \int_{-\infty}^1 {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\,\mathrm{d}\varphi\,\mathrm{d}t \\
&= \int_0^\infty e^{-rt} {}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\, c(t)\,\mathrm{d}t\,.
\end{aligned} \tag{5}$$

The policyholder perceives the premium charged by the insurer for the annuity as "too high" if the charged premium is higher than her perceived premium, that is, $P_0^A > \widetilde{P}_0^A$ (which is the case if ${}_tp_x > {}_t\tilde{p}_x$). Conversely, the premium charged by the insurer is perceived as "too low" if $P_0^A < \widetilde{P}_0^A$ (which is the case if ${}_tp_x < {}_t\tilde{p}_x$).

For the tontine, a single individual uses the survival curve ${}_t\tilde{p}_x$ for herself and ${}_t\hat{p}_x$ for the other policyholders in the pool. Using the policyholder's subjective survival curves for herself and others, the subjective premium of the tontine is given by

$$\begin{aligned}
\widetilde{P}_0^T &= \widetilde{\mathbb{E}}\left[\int_0^\infty e^{-rt} b_T(t)\mathrm{d}t\right] \\
&= \int_0^\infty e^{-rt}\widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon}\widetilde{\mathbb{E}}\left[\frac{nd(t)}{N_\epsilon(t)}\,\middle|\, \zeta_\epsilon > t, \epsilon\right]\right]\mathrm{d}t \\
&= \int_0^\infty e^{-rt}\widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon}\sum_{k=0}^{n-1}\frac{n}{k+1}\binom{n-1}{k}\left({}_t\hat{p}_x^{1-\epsilon}\right)^k\left(1-{}_t\hat{p}_x^{1-\epsilon}\right)^{n-1-k}\right] d(t)\,\mathrm{d}t \\
&= \int_0^\infty e^{-rt}\widetilde{\mathbb{E}}\left[\frac{{}_t\tilde{p}_x^{1-\epsilon}}{{}_t\hat{p}_x^{1-\epsilon}}\sum_{k=1}^{n}\binom{n}{k}\left({}_t\hat{p}_x^{1-\epsilon}\right)^k\left(1-{}_t\hat{p}_x^{1-\epsilon}\right)^{n-k}\right] d(t)\,\mathrm{d}t \\
&= \int_0^\infty e^{-rt}\widetilde{\mathbb{E}}\left[\frac{{}_t\tilde{p}_x^{1-\epsilon}}{{}_t\hat{p}_x^{1-\epsilon}}\left(1-\left(1-{}_t\hat{p}_x^{1-\epsilon}\right)^n\right)\right] d(t)\,\mathrm{d}t \\
&= \int_0^\infty e^{-rt}\int_{-\infty}^1 \frac{{}_t\tilde{p}_x^{1-\varphi}}{{}_t\hat{p}_x^{1-\varphi}}\left(1-\left(1-{}_t\hat{p}_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\,\mathrm{d}\varphi\, d(t)\,\mathrm{d}t\,.
\end{aligned} \tag{6}$$

In (6), the term $\left(1-(1-{}_t\hat{p}_x^{1-\varphi})^n\right)$ is increasing in ${}_t\hat{p}_x$. If the policyholder does not strongly over- or underestimate her own survival curve compared to the remaining policyholders, the term $\frac{{}_t\tilde{p}_x^{1-\epsilon}}{{}_t\hat{p}_x^{1-\epsilon}}$ is close to 1. Hence, in this case, the subjective tontine premium is also overpriced (underpriced) from the perspective of a policyholder who underestimates (overestimates) her life expectancy compared to the insurer. Note however, that the magnitude of over- or underestimation is rather small and these effects mostly vanish as n gets larger and hardly play a role for a large enough pool size.

To illustrate the patterns described above, we now consider a numerical example. Unless stated otherwise, we always use the parameters summarized in Table 1. The survival curves

Net premium	Pool size	Risk-free rate
$P_0^A = P_0^T = 1$	$n = 100$	$r = 0.02$
Initial age	Gompertz parameters	Longevity shock
$x = 65$	$m = 88.721,\ \beta = 10$	$\epsilon \sim \mathcal{N}_{(-\infty,1]}(-0.0035, 0.0814^2)$

Table 1: Base case parameters.

${}_tp_x$, ${}_t\tilde{p}_x$ and ${}_t\hat{p}_x$ are assumed to follow the well-known Gompertz law (Gompertz (1825)) as used, for example, in Gumbel (1958), Milevsky and Salisbury (2015) and Chen et al. (2019). The Gompertz law is parameterized as

$${}_tp_x = e^{e^{\frac{x-m}{\beta}}\left(1-e^{\frac{t}{\beta}}\right)}$$

with dispersion coefficient $\beta > 0$ and modal age at death m. The use of the Gompertz law should be seen solely as one possible example for the survival curves which allows a rather simple incorporation of subjective mortality beliefs. Note that it is not the primal goal of this article to derive an accurate fit to real-life mortality data when using the Gompertz law, but instead to provide a rather simple example for subjective mortality beliefs and their effects on the optimal retirement decision. To demonstrate the effects of subjective mortality beliefs, we vary the modal age at death to obtain the subjective survival curves as

$${}_t\tilde{p}_x = e^{e^{\frac{x-\widetilde{m}}{\beta}}\left(1-e^{\frac{t}{\beta}}\right)}, \quad {}_t\hat{p}_x = e^{e^{\frac{x-\widehat{m}}{\beta}}\left(1-e^{\frac{t}{\beta}}\right)}$$

for subjective modal ages at death $\widetilde{m}$, $\widehat{m} > 0$. Note that we allow for the modal age at death to vary between the insurer's and the policyholder's perceived survival curves, while we use, for simplicity, the same dispersion coefficient for all three curves. That is, we have tacitly assumed that the modal age at death does not depend on the dispersion coefficient and vice versa.[5] The

[5]Note that this is only one possible example of subjective mortality beliefs, i.e. we could with some additional effort analyze more evolved constructions. However, as the empirical studies in Section 2 lead to different conclusions and depend on the framing of the questions, we prefer to choose a rather simple example of subjective mortality beliefs for our numerical illustrations. In principle, it is possible to allow for simultaneous changes in

reason for this is that the survival probability is increasing in the modal age at death for all choices of x, β and t. Thus, the modal age at death allows us to easily control the subjective mortality beliefs. For the Gompertz parameters, we follow Milevsky and Salisbury (2015), for the parameters of the shock we follow Chen et al. (2019). As mentioned earlier already, the insurer uses best-estimate survival probabilities. For example, if a policyholder underestimates (overestimates) her own and others' survival curves, this underestimation (overestimation) is relative to the best-estimates used by the insurer, that is, we assume that ${}_tp_x \geq (\leq) {}_t\tilde{p}_x$ and ${}_tp_x \geq (\leq) {}_t\hat{p}_x$ for our numerical analyses. Using Gompertz-law and only allowing for the modal age at death to differ between the survival curves, this results in the relations $m \geq (\leq)\widetilde{m}$ and $m \geq (\leq)\widehat{m}$. We assume that the insurer uses the modal age $m = 88.721$, which results in an expected remaining lifetime of $\mathbb{E}[\zeta_\epsilon] = \int_0^\infty \int_{-\infty}^1 {}_tp_x^{1-\varphi} f_\epsilon(\varphi) \mathrm{d}\varphi \mathrm{d}t \approx 20.707$ from the insurer's point of view (for an $x = 65$-year old).

In Table 2, we provide the premium charged by the insurer and the subjective premiums as expected by the policyholders for a constant annuity. The constant annuity payoff $c(t) = c$ is chosen such that $P_0^A = 1$ with the premium P_0^A given in (2). Note that the parameter $\widehat{m}$ does not affect the premium of the annuity in any way and can therefore be omitted in Table 2. It can be seen that the subjective premium of the annuity increases in the modal age at

Subjective modal age	Annuity premium $\widetilde{P}_0^A$
$\widetilde{m} = 80.5$	0.7428
$\widetilde{m} = 83$	0.8197
$\widetilde{m} = 88.721$	1
$\widetilde{m} = 92$	1.1038
$\widetilde{m} = 95$	1.1979

Table 2: Subjective premium $\widetilde{P}_0^A$ (see (5)) of a constant annuity, given a fair premium $P_0^A = 1$. The parameters are as in Table 1, in particular, the insurance company's modal age at death is $m = 88.721$.

death $\widetilde{m}$ assumed by the policyholder. That is, an annuity seems more and more overpriced the stronger an individual underestimates her survival curve. Conversely, an annuity appears underpriced to a policyholder who overestimates her survival curve.

Table 3 provides the premium charged by the insurer and the subjective premium as expected

several parameters of the Gompertz law or to change the underlying mortality law. The former can be carried out by using the so-called *Compensation Law of Mortality* (CLaM) taking into account that the lifetimes of individuals with higher mortality hazard rates are also more volatile (for further details see, e.g., Gavrilov and Gavrilova (1991, 2001) and Milevsky (2018)). In our Gompertz framework, this would imply that low modal ages should be paired with high dispersion coefficients, as explicitly stated in Milevsky (2018). Choosing, for example, $(m, \beta) = (88.721, 10)$ and $(\widehat{m}, \hat{\beta}) = (80.5, 11)$, we still obtain ${}_t\hat{p}_x < {}_tp_x$ on the set $x + t < 150$. That is, the parameters could be chosen in such a way that all our qualitative results in the numerical part still remain valid. Hence, we have decided to choose a rather simple way to incorporate the subjective mortality beliefs by changing the modal age at death.

by the policyholders for a so-called natural tontine as introduced in Milevsky and Salisbury (2015). The payoff of the natural tontine is, in our model setup, given by $d(t) = \mathbb{E}[{}_tp_x^{1-\epsilon}] \cdot d_0$, where d_0 is a constant. Note that the tontine payoff to a single individual in the pool remains constant over time if deaths in the pool occur exactly as expected.[6] The constant d_0 is chosen such that $P_0^T = 1$ with the premium P_0^T given in (4). As in this example, we want to focus on the case where the policyholder has a different perception about the survival than the insurer does, we assume that an individual's survival curve assumed for herself coincides with that assigned to others, that is, ${}_t\tilde{p}_x = {}_t\hat{p}_x$. In Table 3, we can observe similar patterns as in Table 2.

Subjective modal age	Tontine premium $\widetilde{P}_0^T$		
	$n = 10$	$n = 100$	$n = 1\,000$
$\widetilde{m} = \widehat{m} = 80.5$	0.9472	0.9873	0.9966
$\widetilde{m} = \widehat{m} = 83$	0.9704	0.9944	0.9988
$\widetilde{m} = \widehat{m} = 88.721$	1	1	1
$\widetilde{m} = \widehat{m} = 92$	1.0068	1.0005	1.0000
$\widetilde{m} = \widehat{m} = 95$	1.0097	1.0006	1.0000

Table 3: Subjective premium $\widetilde{P}_0^T$ (see (6)) of the natural tontine, given a fair premium $P_0^T = 1$. The parameters are as in Table 1, in particular, the insurance company's modal age at death is $m = 88.721$.

Note however, that the tontine's perceived premium $\widetilde{P}_0^T$ is much less affected by subjective mortality beliefs than the premium of the annuity. The effect vanishes almost completely for large tontine pool sizes.

To summarize, we find that tontines' perceived prices change much less substantially than annuities' due to a subjective survival curve perception. A pessimistic policyholder, assuming her own survival curve to be lower than the one used by the insurer, will perceive an annuity as overpriced while tontines are perceived as almost fairly priced. As it is usual in this stream of literature (for example Yaari (1965), Yagi and Nishigaki (1993), Mitchell (2002), Davidoff et al. (2005), Milevsky and Salisbury (2015), Peijnenburg et al. (2016) and Chen et al. (2019)), the attractiveness of a retirement product is frequently examined in an expected utility framework. To confirm our results in such a framework, we, in the following subsections, consider an expected utility framework to figure out how subjective mortality beliefs affect the relative attractiveness of annuities and tontines and whether a tontine is preferable over an annuity under certain subjective mortality beliefs. We analyze whether the arguments about the premium perception in this section still hold true if the policyholder's expected lifetime utility is used for the product comparison. We start by deriving optimal payoff functions of annuities and tontines in Section 4 and then compare the resulting attractiveness of both products in

[6]Due to its nice structure, Milevsky and Salisbury (2015) recommend this tontine design for an implementation of tontines in today's world which is also why we have decided to choose this design for our numerical demonstration.

Section 5.

4 Optimal payoff and expected utility

In this section, we derive the optimal payoff and the corresponding expected lifetime utility of the annuity and the tontine under subjective mortality beliefs. Our results can be seen as a generalization of the theorems given in Milevsky and Salisbury (2015) and Chen et al. (2019). We first introduce the policyholder's expected discounted lifetime utility as

$$U\big(\{\alpha(t)\}_{t\geq 0}\big) := \widetilde{\mathbb{E}}\left[\int_0^\infty e^{-\rho t}\cdot u\left(\alpha(t)\right)\cdot \mathbb{1}_{\{\zeta_\epsilon > t\}}\mathrm{d}t\right], \tag{7}$$

where $\{\alpha(t)\}_{t\geq 0}$ denotes the insurance product's payoff, $u(x)=\frac{x^{1-\gamma}}{1-\gamma}$ is a CRRA utility function with a risk aversion parameter $\gamma > 0$, $\gamma \neq 1$ and ρ is the subjective discount factor of the policyholder. Note that the expected discounted lifetime utility is taken under the policyholder's subjective expectation $\widetilde{\mathbb{E}}[\,\cdot\,]$. Assuming that ${}_t\tilde{p}_x^{1-\epsilon} = \widetilde{\mathbb{E}}\big[\mathbb{1}_{\{\zeta_\epsilon > t\}}\,\big|\,\epsilon\big]$, we incorporate the subjective mortality beliefs in the individual's decision. In the following two subsections, we first consider an annuity with payoff $\alpha(t) = b_A(t)$, then a tontine with payoff $\alpha(t) = b_T(t)$.

4.1 Annuity

We assume that the individual aims to maximize her expected discounted lifetime utility under the constraint that her initial wealth equals the premium charged by the insurer. The expected discounted lifetime utility of an annuity is given by

$$\begin{aligned} U\big(\{b_A(t)\}_{t\geq 0}\big) &= \int_0^\infty e^{-\rho t}\widetilde{\mathbb{E}}\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] u(c(t))\,\mathrm{d}t \\ &= \int_0^\infty e^{-\rho t}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\, u(c(t))\,\mathrm{d}t. \end{aligned}$$

To be more precise, we solve the following optimization problem to determine the optimal annuity payment $c(t)$:

$$\begin{aligned} \max_{c(t)} U\big(\{b_A(t)\}_{t\geq 0}\big) &= \max_{c(t)} \int_0^\infty e^{-\rho t}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\, u(c(t))\,\mathrm{d}t \\ \text{subject to} \quad v = P_0^A &= \int_0^\infty e^{-rt}{}_t p_x \cdot m_\epsilon(-\ln {}_t p_x)\, c(t)\,\mathrm{d}t, \end{aligned} \tag{8}$$

where v is the investor's initial wealth and P_0^A is the premium charged by the insurer. The solution of optimization problem (8) is given in Theorem 4.1.

Theorem 4.1 *For an annuity contract, the solution to problem* (8) *is given by the optimal payout function*

$$c^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}}{\lambda_A^{1/\gamma}} \left(\frac{{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)}{{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)} \right)^{1/\gamma}, \tag{9}$$

where λ_A *is the optimal Lagrangian multiplier given by*

$$\lambda_A = \left(\frac{1}{P_0^A} \int_0^\infty e^{\left(\frac{1}{\gamma}-1\right)rt - \frac{1}{\gamma}\rho t} \int_{-\infty}^1 {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi \left(\frac{\int_{-\infty}^1 {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi}{\int_{-\infty}^1 {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi} \right)^{1/\gamma} \mathrm{d}t \right)^\gamma.$$

The optimal level of expected utility is then given by

$$U_A := \frac{\lambda_A}{1-\gamma} P_0^A. \tag{10}$$

Proof. See Appendix A.1. □

Note that if there are no subjective mortality beliefs and if $r = \rho$, the optimal annuity payment reduces to the constant $\lambda_A^{-\frac{1}{\gamma}}$ which is in line with Yaari (1965). In all the other cases, the optimal annuity payoff is not constant and may increase or decrease over time. This implies that constant annuities are sub-optimal for individuals whose subjective discount rate differs from the risk-free interest rate, consistent with, for example, Yagi and Nishigaki (1993). Before we analyze the effects of subjective mortality beliefs on the optimal payoff $c^*(t)$ and the optimal level of expected utility U_A in Section 5, we derive the optimal tontine payoff in the following subsection.

4.2 Tontine

This section is dedicated to determining the optimal withdrawal payment $d(t)$ for the tontine. The expected discounted lifetime utility of a tontine is given by

$$U\big(\{b_T(t)\}_{t\geq 0}\big) = \int_0^\infty e^{-\rho t}\, u(d(t))\, \widetilde{\mathbb{E}} \left[\mathbb{1}_{\{\zeta_\epsilon > t\}} \left(\frac{n}{N_\epsilon(t)} \right)^{1-\gamma} \right] \mathrm{d}t,$$

where $(N_\epsilon(t) - 1 \mid \zeta_\epsilon > t, \epsilon) \sim \mathrm{Bin}(n-1, {}_t\hat{p}_x^{1-\epsilon})$ from the policyholder's perspective. The policyholder has a different expectation about others' survival, with the probability $\widetilde{\mathbb{E}}\left[{}_t\hat{p}_x^{1-\epsilon}\right]$. Based on this, we obtain

$$\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x) := \widetilde{\mathbb{E}} \left[\mathbb{1}_{\{\zeta_\epsilon > t\}} \left(\frac{n}{N_\epsilon(t)} \right)^{1-\gamma} \right]$$

$$
\begin{aligned}
&= \widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon} \widetilde{\mathbb{E}}\left[\left(\frac{n}{N_\epsilon(t)} \right)^{1-\gamma} \,\middle|\, \zeta_\epsilon > t, \epsilon \right] \right] \\
&= \widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon} \sum_{k=0}^{n-1} \left(\frac{n}{k+1} \right)^{1-\gamma} \binom{n-1}{k} \left({}_t\hat{p}_x^{1-\epsilon} \right)^k \left(1 - {}_t\hat{p}_x^{1-\epsilon} \right)^{n-1-k} \right] \\
&= \widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon} \sum_{k=1}^{n} \left(\frac{k}{n} \right)^{\gamma} \binom{n}{k} \left({}_t\hat{p}_x^{1-\epsilon} \right)^{k-1} \left(1 - {}_t\hat{p}_x^{1-\epsilon} \right)^{n-k} \right] \\
&= \sum_{k=1}^{n} \binom{n}{k} \left(\frac{k}{n} \right)^{\gamma} \int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} \left({}_t\hat{p}_x^{1-\varphi} \right)^{k-1} \left(1 - {}_t\hat{p}_x^{1-\varphi} \right)^{n-k} f_\epsilon(\varphi) \,\mathrm{d}\varphi \,. \qquad (11)
\end{aligned}
$$

That is, we solve the following optimization problem:

$$
\begin{aligned}
\max_{d(t)} U\left(\{b_T(t)\}_{t\geq 0}\right) &= \max_{d(t)} \int_0^\infty e^{-\rho t} u(d(t))\, \kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x) \,\mathrm{d}t \\
\text{subject to} \quad v = P_0^T &= \int_0^\infty e^{-rt} \int_{-\infty}^{1} \left(1 - \left(1 - {}_tp_x^{1-\varphi} \right)^n \right) f_\epsilon(\varphi) \,\mathrm{d}\varphi \, d(t) \,\mathrm{d}t,
\end{aligned}
\qquad (12)
$$

The solution to problem (12) is given in Theorem 4.2.

Theorem 4.2 *For a tontine, the solution to problem* (12) *is given by the optimal payout function*

$$
d^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}} \left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x) \right)^{1/\gamma}}{\lambda_T^{1/\gamma} \left(\int_{-\infty}^{1} \left(1 - \left(1 - {}_tp_x^{1-\varphi} \right)^n \right) f_\epsilon(\varphi) d\varphi \right)^{1/\gamma}} \,, \qquad (13)
$$

where λ_T *is the optimal Lagrangian multiplier given by*

$$
\lambda_T = \left(\frac{1}{P_0^T} \int_0^\infty e^{\left(\frac{1}{\gamma}-1\right) rt - \frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x) \right)^{1/\gamma}}{\left(\int_{-\infty}^{1} \left(1 - \left(1 - {}_tp_x^{1-\varphi} \right)^n \right) f_\epsilon(\varphi) \mathrm{d}\varphi \right)^{1/\gamma - 1}} \mathrm{d}t \right)^{\gamma} .
$$

The expected discounted lifetime utility is then given by

$$
U_T := \frac{\lambda_T}{1-\gamma} P_0^T \,. \qquad (14)
$$

Proof. See Appendix A.2. □

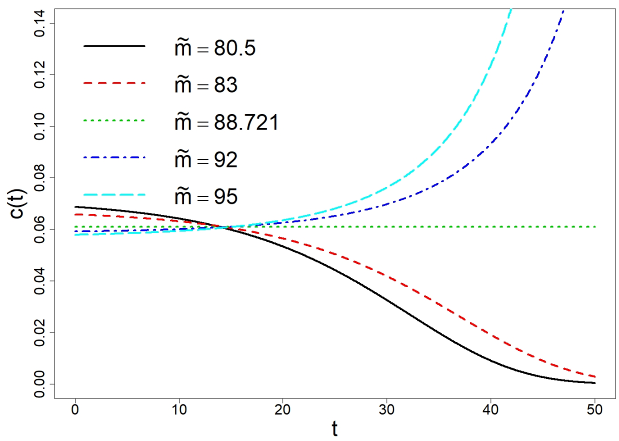
Although the optimal payout structure $d^*(t)$ is much more complex than the optimal annuity payment $c^*(t)$ from (9), the optimal expected utility in (14) differs from (10) only through the Lagrangian multiplier. In the following section, we will have a closer look at the effect of the subjective mortality beliefs on the optimal payoff and expected utility of both annuities and tontines.

5 Effects of subjective mortality beliefs

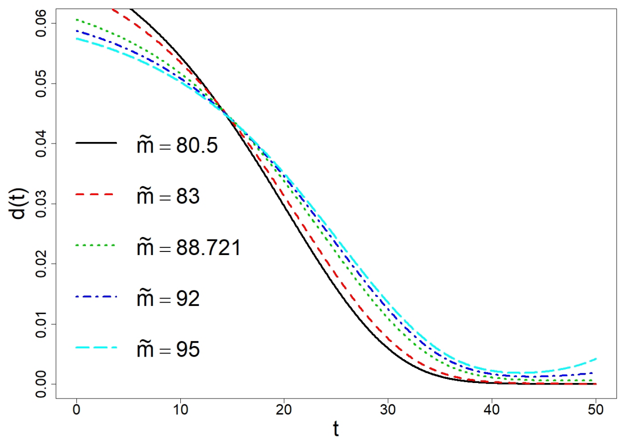
In this section, we analyze the effect of the subjective mortality beliefs on the optimal retirement decision. As explained in Section 3.1, we denote by ${}_tp_x$ the insurer's best-estimate survival curve and by $({}_t\tilde{p}_x, {}_t\hat{p}_x)$ the policyholder's subjective survival curve used for herself and the remaining policyholders in the pool, respectively.

5.1 Subjective mortality beliefs concerning oneself

We start by analyzing the effects of the individual's subjective mortality beliefs about herself on the optimal payoff and the optimal level of expected utility of the two products. In Figure 1, we illustrate the effects of the subjective survival curve ${}_t\tilde{p}_x$ on $c^*(t)$ and $d^*(t)$. In the following analysis, we always consider a policyholder with a risk aversion of $\gamma = 3$ and a subjective discount factor of $\rho = r = 0.02$. We basically make the same observation in the two panels in



(a) Annuity



(b) Tontine, $m = \widehat{m} = 88.721$

Figure 1: Optimal payoff of the annuity and the tontine for different choices of the survival curve ${}_t\tilde{p}_x = \exp\left(e^{\frac{x-\widetilde{m}}{\beta}}\left(1 - e^{\frac{t}{\beta}}\right)\right)$ being a function of the modal age at death $\widetilde{m}$, where the parameters are chosen as in Table 1 with $\gamma = 3$ and $\rho = r$.

Figure 1: If the individual believes that she lives shorter than the insurer has estimated, that is, ${}_t\tilde{p}_x < {}_tp_x$ (here $\widetilde{m} = 80.5$ and $\widetilde{m} = 83$, respectively), the individual will buy a product which provides a higher payment at the early retirement ages and a lower payment at the more advanced retirement ages (compared to the case with ${}_tp_x = {}_t\tilde{p}_x$). For the annuity, this leads to decreasing payoffs. For the tontine, a more steeply declining payoff results (compared to the case with ${}_tp_x = {}_t\tilde{p}_x$). In the reverse case, that is, for a policyholder that is optimistic with respect to her survival curve (${}_t\tilde{p}_x > {}_tp_x$, here $\widetilde{m} = 92$ and $\widetilde{m} = 95$, respectively), the individual buys a product which provides a lower payment at the early retirement ages and a higher payment at the more advanced retirement ages. For the annuity, increasing payoffs

result. For the tontine, less steeply declining payoffs (compared to the case with ${}_tp_x = {}_t\tilde{p}_x$), which slightly increase at very old ages, are obtained. Thus, living shorter in expectation has the same effect as being less patient about the future: Individuals tend to consume more at earlier retirement ages. Older ages are given less importance than earlier ages and therefore, lower payments result at older ages. If, on the other hand, the individual expects to live longer than the insurer assumes, the opposite is true.

We want to verify whether the perceived overpricing (underpricing) of annuities leads to a lower (higher) utility level for the policyholders. For this purpose, we introduce problem (15). A policyholder who assumes her own subjective survival curve to be ${}_t\tilde{p}_x$ wants to choose the optimal retirement product following the optimization problem:

$$
\begin{aligned}
\max_{c(t)} U\big(\{b_A(t)\}_{t\geq 0}\big) &= \max_{c(t)} \int_0^\infty e^{-\rho t}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\, u(c(t))\,\mathrm{d}t \\
\text{subject to} \quad v = \widetilde{P}_0^A &= \int_0^\infty e^{-rt}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\, c(t)\,\mathrm{d}t.
\end{aligned}
\tag{15}
$$

Note that, in contrast to the optimization problem (8), the constraint in the optimization problem (15) is given in terms of the policyholder's subjective premium $\widetilde{P}_0^A$ instead of the insurer's fair premium P_0^A.

For a more thorough analysis of the retirement products, we introduce certainty equivalents CE defined as the level of constant retirement benefits that yield the same expected utility as the annuity and tontine, respectively. In other words, we determine $\mathrm{CE} > 0$ such that

$$
U\big(\{\mathrm{CE}\}_{t\geq 0}\big) = U\big(\{\alpha(t)\}_{t\geq 0}\big)\,, \tag{16}
$$

or equivalently,

$$
\mathrm{CE} = \left((1-\gamma)\left(\int_0^\infty e^{-\rho t}{}_t\tilde{p}_x m_\epsilon(-\ln {}_t\tilde{p}_x)\mathrm{d}t\right)^{-1} \cdot U\big(\{\alpha(t)\}_{t\geq 0}\big)\right)^{\frac{1}{1-\gamma}},
$$

where $U\left(\{\alpha(t)\}_{t\geq 0}\right)$ is the expected discounted lifetime utility of the individual as defined in (7).

Comparing the results of optimization problems (8) and (15), we can see whether "perceived" overpricing (underpricing) leads to a lower (higher) certainty equivalent CE for the policyholder. Table 4 presents the results for an annuity. The column "CE with premium P_0^A" gives the certainty equivalent obtained via optimization problem (8). The third column "CE with premium $\widetilde{P}_0^A$" gives the CE obtained via optimization problem (15). If the third column gives a higher value, problem (15) leads to a higher CE to the policyholder than problem (8) does.

Subjective modal age	CE with premium P_0^A Problem (8)	CE with premium $\widetilde{P}_0^A$ Problem (15)
$\widetilde{m} = 80.5$	0.0629	0.0822
$\widetilde{m} = 83$	0.0619	0.0745
$\widetilde{m} = 88.721$	0.0611	0.0611
$\widetilde{m} = 92$	0.0613	0.0553
$\widetilde{m} = 95$	0.0618	0.0510

Table 4: Certainty equivalents (CE) of the annuity for problems (8) and (15), where the payoff is the (subjective) utility maximizing payoff (9). The parameters are as in Table 1 with risk aversion $\gamma = 3$ and subjective discount factor $\rho = r$.

The reason for this is that the policyholder perceives the premium charged by the insurer in problem (8) as too expensive. We observe that this is the case if the policyholder underestimates her survival curve ($\widetilde{m} < m$), in line with the results in Table 2. The reverse results hold for a policyholder that is optimistic with respect to her survival curve ($\widetilde{m} > m$). These results confirm the intuition about the premium perception analyzed in Section 3.3.

As the price of tontines is hardly influenced by ${}_t\tilde{p}_x$ (see Section 3.3), we leave out the analogous problem to (15) for tontines at this point. Instead, we analyze the main driving factor for a tontine's attractiveness in the following subsection.

5.2 Subjective mortality beliefs concerning others

In the previous subsection, we have analyzed how the policyholder's survival curve for herself ${}_t\tilde{p}_x$ influences the annuity and the tontine. Note, however, that the survival curve the policyholder assumes for everyone else ${}_t\hat{p}_x$ does not affect the payoffs or the expected utility of the annuity in any way. Therefore, we need to analyze the influence of ${}_t\hat{p}_x$ on the expected utility of the tontine if we want to find out which product is more preferable under subjective mortality beliefs.

Proposition 5.1 *The optimal level of expected utility of the tontine U_T decreases in the survival curve for other policyholders ${}_t\hat{p}_x$ for all $\gamma > 0$, $\gamma \neq 1$. As a consequence, the certainty equivalent of the tontine is also decreasing in ${}_t\hat{p}_x$.*

Proof. See Appendix A.3. □

The expected utility decreases in the survival curve of the other policyholders in the tontine pool, so the more the individual under consideration underestimates ${}_t\hat{p}_x$, the higher the expected utility of the tontine. If the individual believes that less individuals will survive a certain time point $t > 0$, for the same payoff to the pool $d(t)$, the surviving individual will receive (at least on average) a higher payout. As a consequence, the "perceived" expected utility of the individual will be raised by this underestimation of the survival curve about the other policyholders in the pool.

This leads us to our main result, the comparison between the certainty equivalents (CE) of annuity and tontine under subjective mortality beliefs. We are able to provide analytical results if systematic mortality risk is ignored, that is if $\epsilon \equiv 0$, see Theorem 5.2. For reasonable parameter sets, these results still hold true if systematic mortality is introduced, see our later numerical results.

Theorem 5.2 (Certainty equivalent comparison under subjective beliefs)

*(a) If beliefs do not differ between policyholder and insurance company, that is if ${}_tp_x = {}_t\tilde{p}_x = {}_t\hat{p}_x$, we find that the CE of a tontine **never** (that is for any portfolio size $n \in \mathbb{N}$) **exceeds** the CE of an annuity.*

*(b) Consider the case **with systematic mortality risk**. If*

$$\int_{-\infty}^{1} {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi > \left(\frac{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} \left(\frac{1}{{}_t\hat{p}_x^{1-\varphi}}\right)^{1-\gamma} f_\epsilon(\varphi)\mathrm{d}\varphi}{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi} \right)^{\frac{1}{\gamma-1}}, \tag{17}$$

*there exists a pool size $N_0 \in \mathbb{N}$ such that the subjective CE of a tontine is (for any portfolio size $n \geq N_0$) **higher** than the subjective CE of an annuity.*

*(c) Consider the case **without systematic mortality risk** ($\epsilon \equiv 0$). In this case, assumption (17) simplifies to ${}_t\hat{p}_x < {}_tp_x$.*

Remark 5.3 (Theorem 5.2) *Part (a) of Theorem 5.2 is not an unknown result and has already been proven in Milevsky and Salisbury (2015) in a scholar setting. The minor extensions of our article compared to Milevsky and Salisbury (2015) are the inclusion of systematic mortality risk by the shock ϵ and the use of a different tontine premium. Parts (b) and (c) of Theorem 5.2 have, to the best of our knowledge, not been proven in the literature.*

Proof:

(a) Consider the following optimization problem (with no subjective mortality beliefs):

$$\max_{\alpha\in[0,1]} \mathbb{E}\left[\int_0^\infty e^{-\rho t}\mathbb{1}_{\{\zeta_\epsilon>t\}} u\left(\alpha c^*(t) + (1-\alpha)\frac{n}{N_\epsilon(t)}d^*(t)\right)\mathrm{d}t\right] \tag{18}$$

where $c^*(t)$ is the optimal annuity payoff (9) and $d^*(t)$ is the optimal tontine payoff (13). In particular, it holds $v = P_0^A = P_0^T$, i.e. the individual maximizes her utility over the fraction of wealth invested in the annuity when having access to both optimal annuities and tontines ($v = \alpha P_0^A + (1-\alpha)P_0^T$). Note that this optimization problem differs from Problems (8) and (12), as , in (18), we maximize over a fraction of initial wealth. Therefore, we do not have to consider a budget constraint, since the individual can only split her initial wealth v between the optimal annuity with initial value $P_0^A = v$ and the optimal tontine with initial value $P_0^T = v$. The objective function to Problem (18) can be written as

$$
\begin{aligned}
\mathcal{F} &= \int_0^\infty e^{-\rho t}\mathbb{E}\left[\mathbb{1}_{\{\zeta_\epsilon > t\}} u\left(\alpha c^*(t) + (1-\alpha)\frac{n}{N_\epsilon(t)} d^*(t)\right)\right] \mathrm{d}t \\
&= \int_0^\infty e^{-\rho t}\mathbb{E}\left[{}_tp_x^{1-\epsilon}\mathbb{E}\left[u\left(\alpha c^*(t) + (1-\alpha)\frac{n}{N_\epsilon(t)} d^*(t)\right)\middle|\zeta_\epsilon > t, \epsilon\right]\right] \mathrm{d}t \\
&= \int_0^\infty e^{-\rho t}\mathbb{E}\left[{}_tp_x^{1-\epsilon}\sum_{k=0}^{n-1} u\left(\alpha c^*(t) + (1-\alpha)\frac{n}{k+1} d^*(t)\right)\binom{n-1}{k}\right. \\
&\qquad\qquad \left.\left({}_tp_x^{1-\epsilon}\right)^k \left(1 - {}_tp_x^{1-\epsilon}\right)^{n-1-k}\right] \mathrm{d}t \\
&= \int_0^\infty e^{-\rho t}\sum_{k=0}^{n-1} u\left(\alpha c^*(t) + (1-\alpha)\frac{n}{k+1} d^*(t)\right)\binom{n-1}{k} \\
&\qquad\qquad \int_{-\infty}^1 \left({}_tp_x^{1-\varphi}\right)^{k+1} \left(1 - {}_tp_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\,\mathrm{d}\varphi\,\mathrm{d}t.
\end{aligned}
$$

We determine the first-order derivative to find a solution to this optimization problem. The first-order condition with respect to α is given by

$$
\begin{aligned}
\frac{\partial\mathcal{F}}{\partial\alpha} &= \int_0^\infty e^{-\rho t}\sum_{k=0}^{n-1} u'\left(\alpha c^*(t) + (1-\alpha)\frac{n}{k+1} d^*(t)\right)\left(c^*(t) - \frac{n}{k+1} d^*(t)\right)\binom{n-1}{k} \\
&\int_{-\infty}^1 \left({}_tp_x^{1-\varphi}\right)^{k+1} \left(1 - {}_tp_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\,d\varphi\,\mathrm{d}t \stackrel{!}{=} 0.
\end{aligned} \tag{19}
$$

Using (9) and some effort, we can verify that $\alpha^* = 1$ fulfills the first-order condition (19). We still need to verify that $\alpha^* = 1$ is a maximum and that it is the only maximum of the objective function. We can do this by taking a look at the second-order derivative:

$$
\begin{aligned}
\frac{\partial^2\mathcal{F}}{\partial\alpha^2} &= \int_0^\infty e^{-\rho t}\sum_{k=0}^{n-1} u''\left(\alpha c^*(t) + (1-\alpha)\frac{n}{k+1} d^*(t)\right)\left(c^*(t) - \frac{n}{k+1} d^*(t)\right)^2\binom{n-1}{k} \\
&\int_{-\infty}^1 \left({}_tp_x^{1-\varphi}\right)^{k+1} \left(1 - {}_tp_x^{1-\varphi}\right)^{n-1-k} f_\epsilon(\varphi)\,d\varphi\,\mathrm{d}t < 0,
\end{aligned}
$$

since $u''\left(\alpha c^*(t) + (1-\alpha)\frac{n}{k+1} d^*(t)\right) < 0$ for all $\alpha \in [0,1]$. If the second-order derivative is strictly negative, this implies that the first-order derivative is strictly decreasing in α. Hence, $\frac{\partial\mathcal{F}}{\partial\alpha}$ can only be equal to zero for exactly one value of α, which we have already found above $(\alpha^* = 1)$. From this, we also see that the first-order derivative has to be greater than zero for all $\alpha < 1$. Consequently, the expected utility in Problem (18) is increasing in α until it reaches its maximum at $\alpha = 1$. Particularly, a 100% investment of initial wealth in the optimal annuity delivers a higher expected lifetime utility than a 100% investment in the optimal tontine.

(b) For the limiting ($n \to \infty$) tontine , we follow the proof of Theorem 4.2. Note first that by the conditional law of large numbers (see, for example, Majerek et al. (2005) and Hanbali et al. (2019)), we obtain, given the longevity shock ϵ, that the share of survivors under subjective beliefs equals:

$$\lim_{n\longrightarrow\infty}\left(\frac{N_\epsilon(t)}{n}\Big|\epsilon\right) = \lim_{n\longrightarrow\infty}\frac{{}_t\tilde{p}_x^{1-\epsilon} + (n-1)\cdot {}_t\hat{p}_x^{1-\epsilon}}{n}\Big|_\epsilon \longrightarrow {}_t\hat{p}_x^{1-\epsilon}.$$

We obtain, applying the dominated convergence theorem:

$$\begin{aligned}\kappa_{\infty,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x) &:= \lim_{n\to\infty}\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\\ &= \lim_{n\to\infty}\widetilde{\mathbb{E}}\left[\mathbb{1}_{\{\zeta_\epsilon>t\}}\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma}\right]\\ &= \lim_{n\to\infty}\widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon}\mathbb{E}\left[\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma}\Big|\epsilon,\, \zeta_\epsilon > t\right]\right]\\ &= \widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon}\left(\frac{1}{{}_t\hat{p}_x^{1-\epsilon}}\right)^{1-\gamma}\right].\end{aligned}$$

For the annuity, we obtain the Lagrangian multiplier (see Theorem 4.1):

$$\lambda_A = \left(\frac{1}{P_0^A}\int_0^\infty e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}\int_{-\infty}^1 {}_tp_x^{1-\varphi}f_\epsilon(\varphi)\mathrm{d}\varphi\left(\frac{\int_{-\infty}^1 {}_t\tilde{p}_x^{1-\varphi}f_\epsilon(\varphi)\mathrm{d}\varphi}{\int_{-\infty}^1 {}_tp_x^{1-\varphi}f_\epsilon(\varphi)\mathrm{d}\varphi}\right)^{1/\gamma}\mathrm{d}t\right)^\gamma.$$

In the limit $n \to \infty$, we obtain the Lagrangian multiplier for the limiting tontine (see Theorem 4.2):

$$\lambda_{T,n\to\infty} = \left(\frac{1}{P_0^T}\int_0^\infty e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}\cdot\left(\kappa_{\infty,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\right)^{\frac{1}{\gamma}}\mathrm{d}t\right)^\gamma.$$

Following Theorem 4.1 and 4.2, we know that the certainty equivalents of annuity and tontine can be written as functions of the Lagrangian multipliers λ_A and λ_T, that is for $i \in \{T, A\}$:

$$\mathrm{CE} = \left(P_0^i\cdot\lambda_i\cdot\left(\int_0^\infty e^{-\rho t}\cdot {}_t\tilde{p}_x\, m_\epsilon(-\ln {}_t\tilde{p}_x)\,\mathrm{d}t\right)^{-1}\right)^{\frac{1}{1-\gamma}}. \tag{20}$$

Comparing the certainty equivalents of annuity and limiting tontine is thus equivalent to comparing $\lambda_A^{\frac{1}{1-\gamma}}$ and $\lambda_{T,n\to\infty}^{\frac{1}{1-\gamma}}$. Using assumption (17), we obtain:

$$\left(\int_{-\infty}^{1} {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi\right)^{\gamma} \cdot \frac{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi}{\int_{-\infty}^{1} {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi} = \int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi \cdot \left(\int_{-\infty}^{1} {}_tp_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi\right)^{\gamma-1}$$

$$\begin{cases} < \int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi \cdot \frac{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi}\left(\frac{1}{{}_t\hat{p}_x^{1-\varphi}}\right)^{1-\gamma} f_\epsilon(\varphi)\mathrm{d}\varphi}{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi}, & \text{if } \gamma\in(0,1) \\ > \int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi \cdot \frac{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi}\left(\frac{1}{{}_t\hat{p}_x^{1-\varphi}}\right)^{1-\gamma} f_\epsilon(\varphi)\mathrm{d}\varphi}{\int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi} f_\epsilon(\varphi)\mathrm{d}\varphi}, & \text{if } \gamma>1 \end{cases}$$

$$= \int_{-\infty}^{1} {}_t\tilde{p}_x^{1-\varphi}\left(\frac{1}{{}_t\hat{p}_x^{1-\varphi}}\right)^{1-\gamma} f_\epsilon(\varphi)\mathrm{d}\varphi = \kappa_{\infty,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\,.$$

This is equivalent to

$$\lambda_A \begin{cases} < \lambda_{T,n\to\infty}, & \text{if } \gamma\in(0,1) \\ > \lambda_{T,n\to\infty}, & \text{if } \gamma>1 \end{cases}$$

which is again equivalent to $\lambda_A^{\frac{1}{1-\gamma}} < \lambda_{T,n\to\infty}^{\frac{1}{1-\gamma}}$. From (20), we can immediately conclude that the *certainty equivalent of the limiting tontine exceeds the certainty equivalent of the annuity.*

Denoting by $\mathrm{CE}_{T,n}$ an optimal tontine's certainty equivalent with pool size n, we can use basic properties of a converging series $\mathrm{CE}_{T,n} \longrightarrow \mathrm{CE}_{T,n\to\infty}$ that there exists a pool size $N_0 \in \mathbb{N}$ such that the CE of a tontine $\mathrm{CE}_{T,n}$ is (for any portfolio size $n \geq N_0$) higher than the CE of an annuity (this basic convergence result can be found in any mathematical textbook covering the convergence of a sequence of real numbers, like, for example, Schulz (2011)). ☐

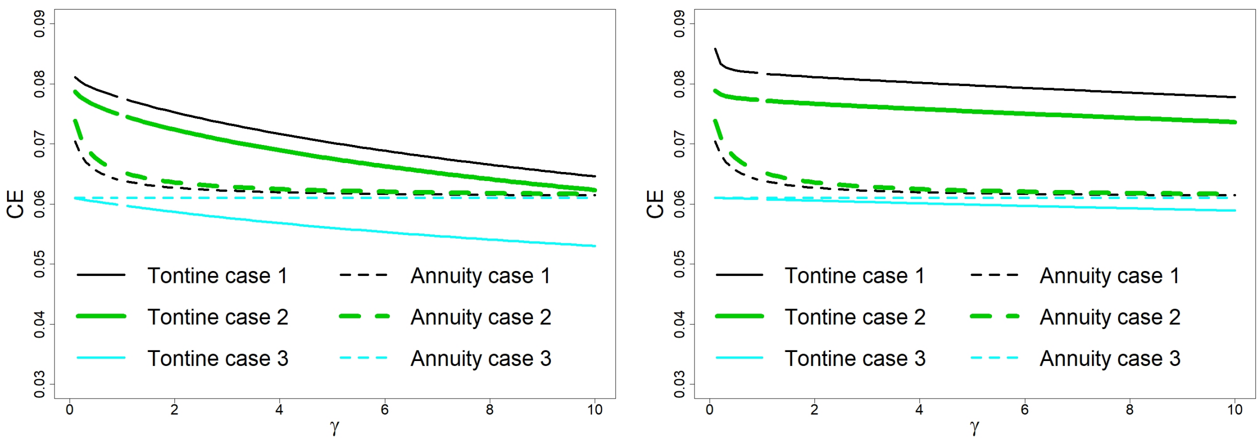
We now analyze for which individuals a tontine might be preferable to an annuity, where the individuals are distinguished by their relative risk aversion. For our numerical analysis, we focus, for example, on the findings of Greenwald and Associates (2012) and O'Brien et al. (2005) who state that people tend to underestimate their own and others' life expectancy, that is, they assign a value ${}_t\tilde{p}_x < {}_tp_x$ and ${}_t\hat{p}_x < {}_tp_x$, respectively. We consider the parameter setup summarized in Table 1 along with the following three cases of subjective mortality beliefs:

- Case 1: $\widetilde{m} = 82$, $\widehat{m} = 80.5$: In this case, the policyholder underestimates others' life expectancy by 6.183 years and her own by 5.128 years compared to the insurer. In particular, the individual believes that she lives in expectation 1.055 years longer than her peers.

- Case 2: $\widetilde{m} = 80.5$, $\widehat{m} = 82$: In this case, the policyholder underestimates others' life expectancy by 5.128 years and her own by 6.183 years compared to the insurer. In particular, the individual believes that she lives in expectation 1.055 years shorter than her peers.

- Case 3: $\widetilde{m} = \widehat{m} = 88.721$: In this case, there are no subjective mortality beliefs, that is, ${}_t\tilde{p}_x = {}_tp_x = {}_t\hat{p}_x$. This corresponds to the setting analyzed in Milevsky and Salisbury (2015) and we mainly include this case to emphasize the importance of our results.

In Figure 2, the corresponding certainty equivalents are given for the annuity and the tontine. The risk aversion parameters are equidistantly placed in the interval $[0.1, 10]$. Here, we consider two different tontines, one with $n = 10$ policyholders and another one with $n = 100$ members. We use very small pool sizes to emphasize that tontines with a low number of policyholders can already be preferred to annuities. We make the following observation from Figure 2:



(a) Portfolio size $n = 10$ (b) Portfolio size $n = 100$

Figure 2: Certainty equivalent for different investors with the three cases explained above and the remaining parameters chosen as in Table 1 with $\rho = r = 0.02$.

- In both panels, we can see that the tontine is preferred by all individuals whose relative risk aversion falls in the interval $[0.1, 10]$ in case 1 and 2, whereas in case 3 all the policyholders prefer the annuity over the tontine. The reason behind the tontine's superiority for the cases with subjective mortality beliefs is the underestimation of the survival curve used for the remaining policyholders ${}_t\hat{p}_x$. As we have already seen in Proposition 5.1, a decrease of this survival curve leads to a higher certainty equivalent of the tontine, while the certainty equivalent of the annuity remains completely unchanged.

- In both Panel (a) and (b), the individual's survival curve assumed for herself ${}_t\tilde{p}_x$ has almost no effect on the tontine's superiority over the annuity. This can be seen from the fact that both in case 1 (${}_t\tilde{p}_x > {}_t\hat{p}_x$) and case 2 (${}_t\tilde{p}_x < {}_t\hat{p}_x$) the tontine is preferred to the annuity for all risk aversion parameters.

- One last effect we can observe here is that the number of policyholders in a tontine also largely impacts the attractiveness of tontines. In both figures, the investors prefer the tontine over the annuity. Comparing Panel (a) with $n = 10$ investors to Panel (b) with $n = 100$ investors, we can see that the certainty equivalent of the tontine is significantly increased when the pool is larger. We have already argued at the beginning of this article that the unsystematic risk in a tontine can be diversified by a sufficiently large pool size and it is well-known already that the attractiveness of a tontine increases with its pool size (see for example Milevsky and Salisbury (2015) and Chen et al. (2019)).

We conclude our numerical analysis by providing the critical values N_0 from Theorem 5.2 for our base case parameter setup. We consider a policyholder with a risk aversion $\gamma = 3$. We can check numerically that condition (17) is fulfilled. Table 5 provides the critical pool sizes N_0 under the three cases considered in Figure 2. Under case 1, the critical pool size N_0 that

Case	N_0
Case 1	2
Case 2	3
Case 3	-

Table 5: Critical pool size N_0 as described in Theorem 5.2. The parameters are as in Table 1 with risk aversion $\gamma = 3$ and subjective discount factor $\rho = r$.

leads to larger certainty equivalent of the tontine compared to the annuity is already equal to 2. Under case 2, the critical pool size equals 3. Note that for case 3 no critical pool size N_0 exists, due to Theorem 5.2 (a).[7]

To summarize the results in our parameter setup, we see that tontines can become more attractive to policyholders than annuities, if policyholders assume a smaller survival curve for their peers than the insurer does. This effect is more pronounced if a larger pool size is considered, as the attractiveness of tontines increases in the pool size.

6 Conclusion

In this article, we study the effects of subjective mortality beliefs on the optimal design of annuities and tontines and their (relative) attractiveness to risk-averse policyholders. In an

[7]Figures of the certainty equivalents of the annuity and the tontine depending on n are available from the authors upon request.

actuarially fair pricing framework, subjective mortality beliefs have a substantial impact on the choice between a tontine and an annuity. If individuals underestimate others' life expectancy compared to the life expectancy assumed by the insurer, the tontine becomes more attractive than the annuity. In particular, if subjective beliefs are present and satisfy a certain condition, there exists a critical pool size from which on the tontine is always preferred over the annuity. The reason for this is that the policyholder believes that less individuals will survive in the pool and, consequently, the share distributed to her will increase. Since annuitization rates remain low and are unlikely to increase in the current low interest environment, this result is of high relevance for the life insurance market as it shows that, under subjective mortality beliefs, a tontine might be an attractive alternative to a conventional annuity. Additionally, we find that policyholders who assume a lower survival curve for themselves than the insurer does perceive annuities as overpriced and it lowers their lifetime utility, consistent with, for example, Wu et al. (2015). Conversely, policyholders who overestimate their life expectancy perceive underpricing of the annuities, which increases their lifetime utility. The premium of a tontine, on the other hand, is only slightly affected by subjective mortality beliefs. In contrast to annuities, an individual's expected lifetime utility from the tontine is only marginally influenced by the individual's subjective survival curve assumed for herself.

An interesting generalization of our subjective beliefs model would be the inclusion of "money illusion", that is, the empirically observed tendency to think in nominal rather than in real monetary terms (see, for example, Basak and Yan (2010)). Although the real terms matter, people tend to think in nominal terms. We leave this question for future research.

References

Basak, S. and Yan, H. (2010). Equilibrium asset prices and investor behaviour in the presence of money illusion. *The Review of Economic Studies*, 77(3):914–936.

Bauer, D., Russ, J., and Zhu, N. (2014). Adverse selection in secondary insurance markets: Evidence from the life settlement market. In *Proceedings of the NBER Insurance Workshop 2014*.

Benartzi, S., Previtero, A., and Thaler, R. H. (2011). Annuitization puzzles. *Journal of Economic Perspectives*, 25(4):143–64.

Bucher-Koenen, T., Kluth, S., et al. (2013). Subjective life expectancy and private pensions. In *Annual Conference 2013 (Duesseldorf): Competition Policy and Regulation in a Global Economic Order*, number 79806. Verein für Socialpolitik/German Economic Association.

Caliendo, F. N., Gorry, A., and Slavov, S. (2017). Survival ambiguity and welfare. Technical report, National Bureau of Economic Research.

Chen, A., Haberman, S., and Thomas, S. (2016). Cumulative prospect theory, deferred annuities and the annuity puzzle. *Available at SSRN:* `https://ssrn.com/abstract=2862792`.

Chen, A., Haberman, S., and Thomas, S. (2018). The implication of the hyperbolic discount model for the annuitisation decisions. *Journal of Pension Economics & Finance*, pages 1–20.

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

Davidoff, T., Brown, J. R., and Diamond, P. A. (2005). Annuities and individual welfare. *American Economic Review*, 95(5):1573–1590.

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

Donnelly, C., Guillén, M., and Nielsen, J. P. (2013). Exchanging uncertain mortality for a cost. *Insurance: Mathematics and Economics*, 52(1):65–76.

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

Elder, T. E. (2013). The predictive validity of subjective mortality expectations: Evidence from the health and retirement study. *Demography*, 50(2):569–589.

Gavrilov, L. A. and Gavrilova, N. S. (1991). *The Biology of Life Span: A Quantitative Approach*. Harwood Academic Publishers, United Kingdom.

Gavrilov, L. A. and Gavrilova, N. S. (2001). The reliability theory of aging and longevity. *Journal of theoretical Biology*, 213(4):527–545.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, 115:513–583.

Gottlieb, D. (2012). Prospect theory, life insurance, and annuities. The Wharton School Research Paper No. 44. *Available at SSRN:* `https://ssrn.com/abstract=2119041`.

Greenwald and Associates (2012). 2011 risks and process of retirement rurvey report of findings. *Society of Actuaries. Prepared by Mathew Greenwald and Associates, Inc., Employee Benefit Research Institute.*

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

Hanbali, H., Denuit, M., Dhaene, J., and Trufin, J. (2019). A dynamic equivalence principle for systematic longevity risk management. *Insurance: Mathematics and Economics*, 86:158–167.

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

Hurd, M. D. and McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, 112(482):966–985.

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

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

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

Lockwood, L. M. (2012). Bequest motives and the annuity puzzle. *Review of economic dynamics*, 15(2):226–243.

Ludwig, A. and Zimper, A. (2013). A parsimonious model of subjective life expectancy. *Theory and Decision*, 75(4):519–541.

Majerek, D., Nowak, W., and Zieba, W. (2005). Conditional strong law of large number. *Int. J. Pure Appl. Math*, 20(2):143–156.

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

Milevsky, M. A. (2014). Portfolio choice and longevity risk in the late seventeenth century: a re-examination of the first english tontine. *Financial History Review*, 21(3):225–258.

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

Milevsky, M. A. (2018). Swimming with wealthy sharks: longevity, volatility and the value of risk pooling. *Journal of Pension Economics & Finance*, pages 1–30.

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

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

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

O'Brien, C., Fenn, P., and Diacon, S. (2005). How long do people expect to live? Results and implications. CRIS Research report 2005–1.

O'Dea, C., Sturrock, D., et al. (2019). Survival pessimism and the demand for annuities. Technical report, Institute for Fiscal Studies.

Payne, J. W., Sagara, N., Shu, S. B., Appelt, K. C., and Johnson, E. J. (2013). Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect. *Journal of Risk and Uncertainty*, 46(1):27–50.

Peijnenburg, K., Nijman, T., and Werker, B. J. (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control*, 70:18–35.

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

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

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

Salisbury, L. C. and Nenkov, G. Y. (2016). Solving the annuity puzzle: The role of mortality salience in retirement savings decumulation decisions. *Journal of Consumer Psychology*, 26(3):417–425.

Schulz, F. (2011). *Analysis 1*. Oldenbourg Verlag, München.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell, London.

Stamos, M. Z. (2008). Optimal consumption and portfolio choice for pooled annuity funds. *Insurance: Mathematics and Economics*, 43(1):56–68.

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

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

Wu, S., Stevens, R., and Thorp, S. (2015). Cohort and target age effects on subjective survival probabilities: Implications for models of the retirement phase. *Journal of Economic Dynamics and Control*, 55:39–56.

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

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

A Proofs

A.1 Proof of Theorem 4.1

We obtain the following Lagrangian function for our optimization problem:

$$\begin{aligned}\mathcal{L} &= \int_0^\infty e^{-\rho t}\widetilde{\mathbb{E}}\left[\mathbb{1}_{\{\zeta_\epsilon > t\}}\right] u(c(t))\,\mathrm{d}t + \lambda_A\left(P_0^A - \int_0^\infty e^{-rt}{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)\,c(t)\,\mathrm{d}t\right)\\ &= \int_0^\infty e^{-\rho t}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\,u(c(t))\,\mathrm{d}t + \lambda_A\left(P_0^A - \int_0^\infty e^{-rt}{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)\,c(t)\,\mathrm{d}t\right).\end{aligned}$$

Rearranging the first order condition delivers

$$c^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}}}{\lambda_A^{1/\gamma}}\left(\frac{{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)}{{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)}\right)^{1/\gamma}.$$

Now we can use the budget constraint to determine the Lagrangian multiplier λ_A. We have

$$\begin{aligned}P_0^A &= \int_0^\infty e^{-rt}{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)\,c^*(t)\,\mathrm{d}t\\ &= \int_0^\infty e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)\left(\frac{{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)}{{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)}\right)^{1/\gamma}\frac{1}{\lambda_A^{1/\gamma}}\mathrm{d}t\,.\end{aligned}$$

As a consequence, we obtain

$$\lambda_A = \left(\frac{1}{P_0^A}\int_0^\infty e^{\left(\frac{1}{\gamma}-1\right)rt-\frac{1}{\gamma}\rho t}{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)\left(\frac{{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)}{{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)}\right)^{1/\gamma}\mathrm{d}t\right)^\gamma.$$

The expected discounted lifetime utility is then given by

$$\begin{aligned}U_A &= \widetilde{\mathbb{E}}\left[\int_0^\infty e^{-\rho t}\mathbb{1}_{\{t<\zeta_\epsilon\}}\,u(c^*(t))\,\mathrm{d}t\right]\\ &= \int_0^\infty e^{-\rho t}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\,u(c^*(t))\,\mathrm{d}t\\ &= \frac{1}{1-\gamma}\int_0^\infty e^{-\rho t}{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)\frac{e^{\frac{1-\gamma}{\gamma}(r-\rho)t}}{\lambda_A^{\frac{1-\gamma}{\gamma}}}\left(\frac{{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)}{{}_tp_x \cdot m_\epsilon(-\ln {}_tp_x)}\right)^{\frac{1-\gamma}{\gamma}}\mathrm{d}t\end{aligned}$$

$$= \frac{\left(\lambda_A^{\frac{1}{\gamma}}\right)^{\gamma-1}}{1-\gamma} \int_0^\infty e^{-\frac{1}{\gamma}\rho t + \frac{1-\gamma}{\gamma} r t} {}_t p_x \cdot m_\epsilon(-\ln {}_t p_x) \left(\frac{{}_t\tilde{p}_x \cdot m_\epsilon(-\ln {}_t\tilde{p}_x)}{{}_t p_x \cdot m_\epsilon(-\ln {}_t p_x)} \right)^{\frac{1}{\gamma}} \mathrm{d}t$$

$$= \frac{\left(\lambda_A^{\frac{1}{\gamma}}\right)^{\gamma-1}}{1-\gamma} \lambda_A^{\frac{1}{\gamma}} P_0^A = \frac{\lambda_A}{1-\gamma} P_0^A .$$

□

A.2 Proof of Theorem 4.2

We obtain the following Lagrangian function for our optimization problem:

$$\begin{aligned} \mathcal{L} &= \int_0^\infty e^{-\rho t}\, u(d(t))\, \widetilde{\mathbb{E}} \left[\mathbb{1}_{\{\zeta_\epsilon > t\}} \left(\frac{n}{N_\epsilon(t)} \right)^{1-\gamma} \right] \mathrm{d}t \\ &\quad + \lambda_T \left(P_0^T - \int_0^\infty e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, \mathrm{d}\varphi\, d(t)\, \mathrm{d}t \right) \\ &= \int_0^\infty e^{-\rho t}\, u(d(t))\, \kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\, \mathrm{d}t \\ &\quad + \lambda_T \left(P_0^T - \int_0^\infty e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, \mathrm{d}\varphi\, d(t)\, \mathrm{d}t \right) . \end{aligned}$$

with $\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)$ defined as in (11). The first order condition is equivalent to

$$d^*(t) = \frac{e^{\frac{(r-\rho)t}{\gamma}} \left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\right)^{1/\gamma}}{\lambda_T^{1/\gamma} \left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \mathrm{d}\varphi \right)^{1/\gamma}} .$$

Now we can use the budget constraint to determine the Lagrangian multiplier λ_T:

$$\begin{aligned} P_0^T &= \int_0^\infty e^{-rt} \int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi)\, \mathrm{d}\varphi\, d^*(t)\, \mathrm{d}t \\ &= \int_0^\infty e^{\left(\frac{1}{\gamma}-1\right) r t - \frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\right)^{1/\gamma}}{\lambda_T^{1/\gamma} \left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \mathrm{d}\varphi \right)^{1/\gamma - 1}} \mathrm{d}t . \end{aligned}$$

As a consequence, we obtain

$$\lambda_T = \left(\frac{1}{P_0^T} \int_0^\infty e^{\left(\frac{1}{\gamma}-1\right) r t - \frac{1}{\gamma}\rho t} \frac{\left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\right)^{1/\gamma}}{\left(\int_{-\infty}^1 \left(1 - \left(1 - {}_t p_x^{1-\varphi}\right)^n\right) f_\epsilon(\varphi) \mathrm{d}\varphi \right)^{1/\gamma - 1}} \mathrm{d}t \right)^\gamma .$$

The expected discounted lifetime utility is then given by

$$
\begin{aligned}
U_T &:= \widetilde{\mathbb{E}}\left[\int_0^\infty e^{-\rho t}\mathbb{1}_{\{t<\zeta_\epsilon\}}\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma} u(d^*(t))\,\mathrm{d}t\right] \\
&= \int_0^\infty e^{-\rho t}\,\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\,u(d^*(t))\,\mathrm{d}t \\
&= \frac{1}{1-\gamma}\int_0^\infty e^{-\rho t}\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\frac{e^{\frac{1-\gamma}{\gamma}(r-\rho)t}\left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\right)^{\frac{1-\gamma}{\gamma}}}{\lambda_T^{\frac{1-\gamma}{\gamma}}\left(\int_{-\infty}^1\left(1-\left(1-{}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\mathrm{d}\varphi\right)^{\frac{1-\gamma}{\gamma}}}\mathrm{d}t \\
&= \frac{\left(\lambda_T^{\frac{1}{\gamma}}\right)^{\gamma-1}}{1-\gamma}\int_0^\infty e^{-\frac{1}{\gamma}\rho t+\frac{1-\gamma}{\gamma}rt}\frac{\left(\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)\right)^{\frac{1}{\gamma}}}{\left(\int_{-\infty}^1\left(1-\left(1-{}_tp_x^{1-\varphi}\right)^n\right)f_\epsilon(\varphi)\mathrm{d}\varphi\right)^{\frac{1-\gamma}{\gamma}}}\mathrm{d}t \\
&= \frac{\left(\lambda_T^{\frac{1}{\gamma}}\right)^{\gamma-1}}{1-\gamma}\lambda_T^{\frac{1}{\gamma}}P_0^T = \frac{\lambda_T}{1-\gamma}P_0^T.
\end{aligned}
$$

□

A.3 Proof of Proposition 5.1

Note that the optimal level of expected utility of the tontine U_T given in (14) depends on ${}_t\hat{p}_x$ only in terms of

$$
\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x) = \widetilde{\mathbb{E}}\left[{}_t\tilde{p}_x^{1-\epsilon}\widetilde{\mathbb{E}}\left[\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma}\;\middle|\;\zeta_\epsilon > t, \epsilon\right]\right].
$$

To figure out the behavior of U_T depending on ${}_t\hat{p}_x$ it thus suffices to determine the behavior of $\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)$:

- It is increasing in ${}_t\hat{p}_x$ for $\gamma > 1$. This can be seen as follows: We know that $(N_\epsilon(t) - 1 \mid \zeta_\epsilon > t, \epsilon) \sim \mathrm{Bin}(n-1, {}_t\hat{p}_x^{1-\epsilon})$. It is shown in Milevsky and Salisbury (2015) that for any random variable $N(p)$ with $N(p) - 1 \sim \mathrm{Bin}(n-1, p)$, it holds that

$$
\frac{d}{dp}\widetilde{\mathbb{E}}[f(N(p))] = \frac{1}{p}\widetilde{\mathbb{E}}\left[(N(p)-1)(f(N(p)) - f(N(p)-1))\right].
$$

  Therefore, we obtain

$$
\begin{aligned}
&\frac{d}{d{}_t\hat{p}_x^{1-\epsilon}}\widetilde{\mathbb{E}}\left[\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma}\;\middle|\;\zeta_\epsilon > t, \epsilon\right] \\
&\quad = n^{1-\gamma}\frac{d}{d{}_t\hat{p}_x^{1-\epsilon}}\widetilde{\mathbb{E}}\left[N_\epsilon(t)^{\gamma-1}\;\middle|\;\zeta_\epsilon > t, \epsilon\right]
\end{aligned}
$$

$$= \frac{n^{1-\gamma}}{{}_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}[\underbrace{(N_\epsilon(t)-1)}_{\geq 0}\underbrace{(N_\epsilon(t)^{\gamma-1} - (N_\epsilon(t)-1)^{\gamma-1})}_{\geq 0} \mid \zeta_\epsilon > t, \epsilon] \geq 0.$$

This implies that $\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)$ is increasing in ${}_t\hat{p}_x$ as ϵ can only take values between $-\infty$ and 1 and thus, ${}_t\hat{p}_x^{1-\epsilon}$ is increasing in ${}_t\hat{p}_x$. Since $1-\gamma < 0$, the utility decreases as ${}_t\hat{p}_x$ increases.

- Now let us consider the case $\gamma \in (0,1)$: We obtain

$$\begin{aligned}
&\frac{d}{d_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}\left[\left(\frac{n}{N_\epsilon(t)}\right)^{1-\gamma} \,\middle|\, \zeta_\epsilon > t, \epsilon\right] \\
&= \frac{n^{1-\gamma}}{{}_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}[(N_\epsilon(t)-1)(N_\epsilon(t)^{\gamma-1} - (N_\epsilon(t)-1)^{\gamma-1}) \mid \zeta_\epsilon > t, \epsilon] \\
&= \frac{n^{1-\gamma}}{{}_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}[(N_\epsilon(t)-1)N_\epsilon(t)^{\gamma-1} - (N_\epsilon(t)-1)^{\gamma} \mid \zeta_\epsilon > t, \epsilon] \\
&= \frac{n^{1-\gamma}}{{}_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}[((N_\epsilon(t)-1)N_\epsilon(t)^{\gamma-1} - (N_\epsilon(t)-1)^{\gamma})\mathbb{1}_{\{N_\epsilon(t)=1\}} \mid \zeta_\epsilon > t, \epsilon] \\
&\quad + \frac{n^{1-\gamma}}{{}_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}[(N_\epsilon(t)-1)(N_\epsilon(t)^{\gamma-1} - (N_\epsilon(t)-1)^{\gamma-1})\mathbb{1}_{\{N_\epsilon(t)\geq 2\}} \mid \zeta_\epsilon > t, \epsilon] \\
&= \frac{n^{1-\gamma}}{{}_t\hat{p}_x^{1-\epsilon}} \widetilde{\mathbb{E}}[\underbrace{(N_\epsilon(t)-1)}_{>0}\underbrace{(N_\epsilon(t)^{\gamma-1} - (N_\epsilon(t)-1)^{\gamma-1})}_{\leq 0}\mathbb{1}_{\{N_\epsilon(t)\geq 2\}} \mid \zeta_\epsilon > t, \epsilon] \leq 0\,.
\end{aligned}$$

This implies that $\kappa_{n,\gamma,\epsilon}({}_t\hat{p}_x, {}_t\tilde{p}_x)$ is decreasing in ${}_t\hat{p}_x$. Since $1-\gamma > 0$, the utility decreases as ${}_t\hat{p}_x$ increases.

The certainty equivalent defined in (16) increases in the expected utility, so it decreases in ${}_t\hat{p}_x$ as well. □

4 Optimal collective investment: The impact of sharing rules, management fees and guarantees

Source:
Chen, A., Nguyen, T., and Rach, M. (2019). Optimal collective investment: The impact of sharing rules, management fees and guarantees. Submitted to *Journal of Banking and Finance* (revise and resubmit). Latest revised and resubmitted version of 2020 available at `http://dx.doi.org/10.2139/ssrn.3249094`

Optimal collective investment: The impact of sharing rules, management fees and guarantees

An Chen*, Thai Nguyen* and Manuel Rach*

* Institute of Insurance Science, Ulm University, Helmholtzstr. 20, 89069 Ulm, Germany.
E–mails: an.chen@uni-ulm.de; thai.nguyen@uni-ulm.de; manuel.rach@uni-ulm.de

Abstract

Many pension beneficiaries are reluctant to give up guaranteed payments, as they believe that their benefits will deteriorate through this. In the present article, we design and solve an optimal collective investment problem under a guarantee constraint. In this problem, we incorporate heterogeneous risk preferences of individual plan members, whose importance has recently been addressed in Alserda et al. (2019). We distinguish between two types of investors (with different willingness to pay management fees) and study the impact of guarantees, sharing rules and management fees on the individual investors' benefits. Regarding the guarantees, we find that requiring deterministic guarantees deteriorates the benefits of investors with different risk aversions in both groups. A flexible guarantee which consists of a deterministic and a state-dependent component turns out to serve each investor's risk appetite in a much better way than the deterministic guarantee. To decently achieve fairness among various investors, careful considerations shall be given to the design of the sharing rule.

Keywords: Collective investment problems, guarantee design, risk sharing

JEL: G11, G23

1 Introduction

Different hybrid pension schemes which aim to combine the advantages of the traditional defined benefit (DB) and defined contribution (DC) schemes have appeared across the developed countries in the last years.[1] An overview over a variety of existing plans can, for instance, be found in Turner (2014). An important example would be the DB Underpin pension plan, often also called floor-offset plan, which provides a guaranteed minimum benefit within a DC plan to the employees. This type of pension schemes has already been studied extensively in recent years, for example in Zhu et al. (2018) or Chen and Hardy (2009). Another example of a hybrid pension plan of increasing importance (Kravitz (2016)) would be the Cash Balance plan which is by law defined to be a DB plan although it actually works more like a DC plan. For further details regarding Cash Balance plans we refer interested readers to Hardy et al. (2014). As many employees are reluctant to give up guarantees, most of these hybrid schemes include a minimum guarantee to the beneficiaries. An additional feature of pension schemes is that investors often do not handle the investment of their contribution on their own, but instead have a fund manager delegate their investment decisions. Naturally, the question will be asked how the well-being of all the individual investors (beneficiaries) is affected by having their capital invested in such a collective fund. The present article answers this question in a *collective* utility-based continuous-time framework under a *portfolio insurance* constraint (a deterministic or a flexible guarantee).

The recent work of Alserda et al. (2019) emphasizes the importance of taking account of heterogeneous risk preferences of pension plan members when setting up an investment strategy for a collectively organized pension fund. In the present article, we design and solve a theoretical collective investment problem, incorporating the heterogeneous risk aversions of the plan members, as in Alserda et al. (2019). In addition, we allow the plan members to have different access to the arbitrage-free and complete market and to require various guarantee levels, as we aim to capture the fact that investors with various risk aversions require different degrees of security in their investment outcomes. Typically, a more risk-averse investor requires a more secure payment. We further analyze the effects of management fees, assuming that the collective of investors can be split into two groups: one group has free access, and the other group limited access to the arbitrage-free and complete market. Accordingly, the investors in the two groups are willing to pay different fees for fund delegation and we assume that an average fee is charged by the fund manager. In our analysis, one group pays too much, and the other group too little for what they are entitled to. We observe that those who are entitled to more than

[1] In a DB scheme, the sponsoring companies basically promise their employees a guaranteed pension payment. In a DC scheme, on the other hand, sponsoring companies, and often also their employees, pay deterministic contributions to an external pension fund where the capital is invested in financial assets. The benefit at retirement therefore depends on the performance of the investment returns experienced during the membership, which implies that the market risk is carried completely by the employees instead of the employers.

they should obtain are more likely to benefit from the fund delegation. The magnitudes by which the individual investors' benefits are influenced depend highly on the sharing rules. The term sharing rule refers to the rule applied by the fund manager to distribute the total terminal wealth among the individual investors.[2] We use state-dependent sharing rules to return to each investor the individual guarantee plus a bonus whose magnitude depends on the performance of the financial market, which is very similar to the payoff of a unit-linked life insurance, as described, for example, in Brennan and Schwartz (1976).

We assume that the fund manager uses a collective utility function defined by a weighted sum of the individual utility functions.[3] The fund manager then maximizes this collective utility function under two types of portfolio insurance constraints: In the first case, the optimal terminal wealth needs to meet a deterministic guaranteed payment. To gain more flexibility, in the other case, we assume that the optimal terminal wealth needs to meet a flexible guarantee payment which consists of a (smaller) deterministic guaranteed payment and a state-dependent payment which depends on the market state at maturity. It is a common goal for an investor to exceed a certain, state-dependent benchmark, like a market index (see, e.g., Grossman and Zhou (1993), Browne (1999) and Tepla (2001)). For both optimization problems, we derive the collective optimal terminal wealth and optimal dynamic trading strategies.

We find that a deterministic guarantee framework deteriorates the well-beings of the majority of the investors. Using state-dependent guarantees (which, in our case, consist of a deterministic and a state-dependent component) allows for a lot of flexibility and enables investors with a low risk aversion to (almost) obtain their individual optimum, while simultaneously allowing the more risk-averse investors to demand high fixed guarantees. Under both guarantee schemes, investors who have no access to the financial market benefit from the fund delegation if the fee they are willing to pay largely exceeds the fee charged by the fund manager. If the fees of both groups are close to each other, the vast majority of investors in both groups suffer losses in utility. For this result, the sharing rule is also an important factor. We find that the sharing rule should carefully reflect each investor's initial contribution to the fund and guarantee level, otherwise some investors benefit at the cost of other investors.

While the effects of heterogeneous risk preferences in a pool are analyzed in detail in Alserda et al. (2019), our analysis focuses on the effects of portfolio insurance, sharing rules and man-

[2] The sharing rule could also be seen as an allocation rule used by the fund manager to allocate capital to the different individuals in the collective. In the following, we will follow the literature on joint decisions under uncertainty (provided below) and rely on the widely used term "sharing rule".

[3] We are aware that the fund manager might have objectives that deviate from those of the collective and do not consider it as her primal objective to maximize the benefits for the collective. Similar to Kim et al. (2016), we ignore the agency problem between the fund manager and the collective. It seems a reasonable assumption for a collectively organized pension fund.

agement fees on the heterogeneous investors in the collective. Alserda et al. (2019) point out that the utility of every individual plan member worsens through collective investments, as the applied investment strategies in the collective pension fund are safer than implied by members' preferences. The deterministic guarantee framework considered in this article provides a rational explanation why a safer investment strategy is adopted and quantifies the utility loss for each individual in the fund when the safer investment strategy is applied. The flexible guarantee framework could provide a solution to the above issue as the heterogeneous risk preferences are better incorporated in the collective investment problem when allowing the individual to choose a flexible guarantee reflecting her own risk appetite. Hence, our results are highly relevant in the context of occupational pension schemes as it is a heavily disputed question in today's world how pension schemes should be designed and how their investment strategies should be determined (see, for example, Lucas and Zeldes (2009) in the context of public pensions).

Our paper contributes to the literature on pensions and optimal asset allocation under portfolio insurance by the analysis of a collective DC scheme with guarantees. Note that the literature on pensions often focuses on a single beneficiary (e.g. Broeders and Chen (2010) and Broeders et al. (2011)) or assumes homogeneous risk preferences of plan members (e.g. Beetsma et al. (2013)). This assumption is also often made in the literature analyzing different designs of portfolio insurance in financial markets (see, for example, Grossman and Vila (1989), Basak (1995), Grossman and Zhou (1993, 1996), Browne (1999), Tepla (2001), Jensen and Sørensen (2001), Deelstra et al. (2004), El Karoui et al. (2005), Gabih et al. (2009) and Chen et al. (2018)). Our paper is also closely related to literature in which a collective of investors faces a joint decision under uncertainty in a financial market (without portfolio insurance), for example Wilson (1968), Amershi and Stoeckenius (1983), Huang and Litzenberger (1985), Dumas (1989), Weinbaum (2009), Pazdera et al. (2016), Jensen and Nielsen (2016), Schumacher (2018), Branger et al. (2018a) and Branger et al. (2018b). To the best of our knowledge, the existing literature on collective decisions under uncertainty has not considered portfolio insurance constraints. In this sense, our article combines the literature streams on collective investment and portfolio insurance by considering collective optimal asset allocation problems under portfolio insurance.

The remainder of the paper is structured as follows. In Section 2 we start with reviewing the individual optimization problem, which serves as a comparison basis throughout the paper. As a next step, we assume that the two groups of investors are tied together in their investment decision and solve the optimal collective investment problem under the deterministic guarantee framework in Section 3. In Section 4, we introduce the flexible guarantee payment and solve the optimal collective investment problem under the flexible guarantee framework. In both Sections 3 and 4, we also answer the questions of how the investment strategies for the two optimization problems can be derived and provide numerical analyses for investors' well-beings. In Chapter 5, a discussion of the practical relevance and implications of our findings can be found. Section 6 extends our baseline model to stochastic interest rates. Section 7 concludes the paper and is followed by the appendix where some of the proofs are shown.

2 Model setup and review of individual optimization problem

We consider a collective of n individual investors (pension beneficiaries). Investor i has an initial wealth x_i delegated to a fund manager for investment at time 0. Reasons for fund delegation in general can be various: Individual investors might believe that fund managers will perform better due to their professional skills. It might also be the case that individuals display investment inertia and prefer to delegate their investment decisions because asset management costs time and energy (cf. Kim et al. (2016)). Delegation to a collective fund manager is also motivated by risk sharing of non-marketable risks and economies of scale. In an occupational pension context, it is common that individual pension plan members do not administrate investments themselves, but instead contribute to a collectively organized pension fund. Usually, the pension fund cannot fully reflect each individual's risk preferences (see, for example, Alserda et al. (2019) and Frijns (2010)). In this sense, the investors are tied together and invest collectively.

We assume throughout this paper that each investor's preferences are modeled by a CRRA utility function $U_i(x) = \frac{x^{1-\gamma_i}}{1-\gamma_i}$ where $\gamma_i \neq 1$, $\gamma_i > 0$ for $i = 1, \dots, n$. As pointed out by Sharpe (2017), power utility (preferences with constant relative risk aversion (CRRA)) is the most frequently used utility function to capture the preferences of individuals. We allow the risk aversion parameters γ_i to be different for all investors to capture the heterogeneity in risk preferences among pension plan members observed in Alserda et al. (2019).

2.1 Financial market and two groups of investors

We consider a finite-horizon $[0, T]$-economy whose uncertainty is represented by a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, P)$. Let $W = \{W_t\}_{t\in[0,T]}$ be a standard Brownian motion under P and assume that the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ is the augmented filtration generated by W. For our analysis, we assume a Black-Scholes economy. To be more precise, there are two assets traded in the market, a risk-free asset B earning a constant interest rate $r \in \mathbb{R}$ (which might be negative) and a stock S following a geometric Brownian motion with instantaneous rate of return $\mu > r$ and volatility $\sigma > 0$:

$$\begin{aligned} dB_t &= rB_t \mathrm{d}t, \;\; B_0 = 1, \\ dS_t &= \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t, \;\; S_0 = s. \end{aligned} \tag{1}$$

In this economy, the state price density process is uniquely determined by

$$\xi_t = \exp\left(-rt - \frac{1}{2}\eta^2 t - \eta W_t\right), \quad \eta = \frac{\mu - r}{\sigma}, \tag{2}$$

or, equivalently,

$$d\xi_t = -\xi_t \left(r\mathrm{d}t + \eta \mathrm{d}W_t\right), \quad \xi_0 = 1.$$

The random variable ξ_t reflects the state of the financial market at time t. Low values imply a good performance of the market, high values imply a bad state. Additionally, for any contingent T-claim with payoff V_T, the time-t-price can be determined as

$$V_t = \mathbb{E}\left[\frac{\xi_T}{\xi_t} V_T \,\middle|\, \mathcal{F}_t\right] \quad \text{for all } t \leq T.$$

Note that the state price density has been frequently used in the literature, see for example Dybvig and Ross (1987), Basak (1995), Jensen and Sørensen (2001), Boyle and Tian (2007) and Chen et al. (2018).

Coming back to the n investors, we assume that they can be divided into two groups: Group 1 with $n_1 < n$ investors who have access to the complete and arbitrage-free market and can in principle invest on their own, and Group 2 with $n - n_1$ investors who do not have full access to the complete and arbitrage-free market. The investors in Group 1 might still delegate their asset management decisions, as they might want to save time and energy for other joyful things (see Kim et al. (2016)). The investors in Group 2 might be more ready to delegate their investment decision, as they benefit from professional skills of the fund manager. Without loss of generality, we assume that investors $i = 1, \ldots, n_1$ are in Group 1 and that the remaining investors $i = n_1 + 1, \ldots, n$ are in Group 2. If both groups delegate their funds to the same fund manager, Group 1 probably is ready to pay a lower fee than Group 2 for fund delegation. We assume for simplicity that the fund manager charges a common fee for each. Using the common fee can cause the presence of adverse selection, i.e. Group 1 might have less incentives to stay in the collective. In the present article, we leave out this issue and assume that the two groups of investors are tied together and the fund is collectively administered by a fund manager, which is the case, e.g., in a collectively organized pension fund. Details on this setup are provided in the following sections. In the remainder of this section, we will first review the optimal asset allocation problem of the fund manager investing on behalf of a single investor, which will then serve as a benchmark in later sections.

2.2 Optimal investment for a single investor

The single-investor case is well-known and covered in many textbooks, like, for example, Korn (2014). Our setting is slightly different by incorporating delegation fees. Let us assume $\epsilon_1 \in (0,1)$ and $\epsilon_2 \in (0,1)$ be the fee fractions that these two groups of investors are willing to pay for fund delegation if the fund manager invests for them separately, where we assume $\epsilon_1 < \epsilon_2$. We use $\pi^{(i)} = \{\pi_t^{(i)}\}_{t\in[0,T]}$ to denote the fraction of wealth that is, on behalf of investor i, invested in the risky asset. The remaining fraction is invested in the risk-free asset. After subtracting the proportional fee $\epsilon_i = \epsilon_1$ for $i = 1, \ldots, n_1$ and $\epsilon_i = \epsilon_2$ for $i = n_1 + 1, \ldots, n$, the fraction of initial wealth invested in financial assets is then given by $X_0^{(i)} = (1-\epsilon_1)x_i$ for $i = 1, \ldots, n_1$ and $X_0^{(i)} = (1-\epsilon_2)x_i$ for $i = n_1 + 1, \ldots, n$. The investment strategy $\pi^{(i)}$ shall be chosen from the following admissible set for each investor $i = 1, \ldots, n$:

$$\mathcal{A}((1-\epsilon_i)x_i) := \Bigg\{ \pi^{(i)} \;\Bigg|\; X_0^{(i)} = (1-\epsilon_i)x_i,\ \pi^{(i)} \text{ is progressively measurable,} \\ X_t^{(i)} \geq 0 \text{ for all } t \geq 0,\ \int_0^T \left(\pi_t^{(i)}\right)^2 dt < \infty \Bigg\}.$$

Here, $X_t^{(i)}$ is the investor's wealth process under the self-financing condition which satisfies the following stochastic differential equation:

$$dX_t^{(i)} = \left(r + \pi_t^{(i)}(\mu - r)\right) X_t^{(i)}\mathrm{d}t + \sigma \pi_t^{(i)} X_t^{(i)} \mathrm{d}W_t\,, \quad X_0^{(i)} = (1-\epsilon_i)x_i\,. \tag{3}$$

The optimization problem which the fund manager solves on behalf of investor i is then given by[4]

$$\max_{\{\pi_t^{(i)}\}_{t\in[0,T]} \in \mathcal{A}((1-\epsilon_i)x_i)} \mathbb{E}\left[\frac{\left(X_T^{(i)}\right)^{1-\gamma_i}}{1-\gamma_i} \right] \text{ subject to (3).} \tag{4}$$

In a complete market, we can solve this dynamic optimization problem by using the static martingale approach (see for example Cox and Huang (1989)), that is, by first finding the

[4] As this article deals with fund delegation, we assume that the individuals have set aside a fraction of their initial wealth for consumption prior to the investment and only focus on the terminal wealth obtained at maturity.

optimal terminal wealth $X_T^{(i,*)}$ of the following static optimization problem:

$$\max_{X_T^{(i)}} \mathbb{E}\left[\frac{\left(X_T^{(i)}\right)^{1-\gamma_i}}{1-\gamma_i}\right] \quad \text{subject to} \quad \mathbb{E}\left[\xi_T X_T^{(i)}\right] = (1-\epsilon_i)x_i\,. \tag{5}$$

It is well-known (cf. for example Karatzas and Shreve (1998)) that in a complete financial market, any contingent claim whose initial market value is given by $(1-\epsilon_i)x_i$ can be replicated by a self-financing investment strategy starting with the same initial investment $(1-\epsilon_i)x_i$. Hence, solving the dynamic problem (4) and the static problem (5) results in the same optimal terminal payoff which is given by

$$\begin{aligned} X_T^{(i,*)} &= I_i(\lambda_i \xi_T) \\ &= (1-\epsilon_i)x_i \exp\left(\left(r+\frac{1}{2}\eta^2\right)\left(1-\frac{1}{\gamma_i}\right)T - \frac{1}{2}\eta^2\left(1-\frac{1}{\gamma_i}\right)^2 T\right)\xi_T^{-\frac{1}{\gamma_i}}, \end{aligned} \tag{6}$$

where $I_i(\cdot) = (\cdot)^{-\frac{1}{\gamma_i}}$ is the inverse marginal utility function and λ_i is the Lagrangian multiplier which is determined such that the budget constraint is fulfilled. As usual in the literature on optimal asset allocation, in (6) and subsequent sections, we express the optimal wealth and investment strategy at time $t \in (0,T]$ in terms of the state price density ξ_t. Note that there is a one-to-one negative relation between the state price density and the stock price by the following equation:

$$\xi_t = \exp\left(-rt - \frac{1}{2}\eta^2 t + \frac{\eta}{\sigma}\left(\mu - \frac{1}{2}\sigma^2\right)t\right)\left(\frac{S_t}{S_0}\right)^{-\frac{\eta}{\sigma}}, \tag{7}$$

which shows the state price density reflects the market state in the exact opposite way as the stock price. Hence, all results can also be represented in terms of the stock price.

The optimal dynamic investment strategy associated with the payoff in (6) is given by the so-called Merton strategy (see Merton (1971))

$$\pi_t^{(i)} = \frac{\mu - r}{\gamma_i \sigma^2}\,,$$

a constant fraction in the risky asset which corresponds to the adjusted Sharpe-ratio divided by the individual relative risk aversion coefficient γ_i. The optimal individual terminal wealth (6) solely serves as a benchmark and a comparison basis to the collective investment problem below.

3 Collective optimization with deterministic guarantees

In this section, we consider the collective optimization problem faced by the fund manager for a collective of heterogeneous investors. In addition, we incorporate guarantee constraints in the collective optimization problem.

3.1 Collective utility function

We assume that the n investors considered delegate a fund manager to invest the total initial wealth $x = \sum_{i=1}^{n} x_i$ collectively. The fund manager charges a proportional fee $\bar{\epsilon} \in [\epsilon_1, \epsilon_2]$ from each investor in the collective, independently of the group which this investor belongs to. That is, the fund manager invests $(1-\bar{\epsilon})x$ in the capital market and retains $\bar{\epsilon}x$. For now, we allow $\bar{\epsilon}$ to be a constant between ϵ_1 and ϵ_2 and specify it only in the numerical section. To capture the risk preferences of each individual in the collective, we assume that the fund manager uses the following utility function

$$U_B(v) = \max_{v_1 \geq 0, \ldots, v_n \geq 0,\, v = \sum_{i=1}^{n} v_i} \sum_{i=1}^{n} \beta_i U_i(v_i)\,, \tag{8}$$

where $B = (\beta_1, \ldots, \beta_n)$ is a vector consisting of strictly positive numbers adding up to 1. This collective utility function has been widely considered in the literature e.g. Dumas (1989), Xia (2004), Karatzas et al. (1990), Pazdera et al. (2016) and Branger et al. (2018b). The utility function carries a maximum operator, meaning that the fund manager aims to achieve the highest utility level (or the highest total wealth level) for a given set of weighting factors. By choosing a specific set of the weighting factors β_i, $i = 1, \ldots, n$, the fund manager can decide how she weighs each individual investor in the collective investment problem.

It has been shown (for example in Branger et al. (2018b)) that U_B is a strictly increasing concave utility function whose inverse marginal utility is given by

$$I_B(\cdot) := (U_B')^{-1}(\cdot) = \sum_{i=1}^{n} I_i\left(\frac{\cdot}{\beta_i}\right). \tag{9}$$

This analytical expression of the inverse marginal utility allows us to write down the optimal solution to the collective terminal wealth in analytical form in the following section. From now on, let $G_T^{\text{det}} \in [0, (1-\bar{\epsilon})xe^{rT}]$ be the deterministic guarantee that needs to be met by the fund manager. An upper bound for the guarantee level is given to make sure that the optimization problem (10) is feasible. Both the guarantee specification and the sharing rule used to redistribute the total terminal payoff will be introduced in Section 3.4. For now it is not necessary to make further assumptions regarding these quantities. It is, though, important to emphasize

here that for the fund manager's investment decision only the total level of guarantee G_T^{det} matters. The task of meeting all the individual guarantees will be discussed by designing an appropriate sharing rule in Section 3.4.
Under the portfolio insurance with a deterministic guarantee, we obtain the following optimization problem:

$$\begin{aligned} \max_{X_T} \mathbb{E}\left[U_B\left(X_T\right)\right] \quad \text{subject to} \quad & \mathbb{E}\left[\xi_T X_T\right] = (1-\bar{\epsilon})x, \\ & X_T \geq G_T^{\text{det}} \text{ a.s.} \end{aligned} \tag{10}$$

Before we present the general solution to Problem (10) in the following subsection, we observe the following cases:

- The case $G_T^{\text{det}} = (1-\bar{\epsilon})xe^{rT}$ results in a 100% investment in the risk-free asset, because this is the only admissible investment strategy.
- The case $G_T^{\text{det}} = 0$ leads to an optimization problem with no guarantee constraint, as covered, for example, in Branger et al. (2018b). Note that it is possible in this optimization problem to achieve the individual optimal solutions (6) for each investor i. The main assumption for this result is the choice of the weights β_i which we will discuss in Section 3.5.
- Note that, for $\beta_i = 1$ for some i and $\beta_j = 0$ for all $j \neq i$, Problem (10) coincides with the optimization problem considered in Jensen and Sørensen (2001) as the fund manager only takes into account one single investor in the collective.
- If we assume $\beta_i = 1$ for some i, $\beta_j = 0$ for all $j \neq i$ and $G_T^{\text{det}} = 0$, then we end up with Problem (5) where the initial budget is given now by $(1-\bar{\epsilon})x$ instead of $(1-\epsilon_i)x_i$.

3.2 Optimal terminal wealth

In order to avoid redundancy, we leave out the derivation of the solution of the optimization problem (10) in this section. It can be treated as a special case of Proposition 4.1 where we deal with a flexible guarantee. Therefore, we directly jump to the solution of Problem (10):

$$X_T^* = \max\left\{I_B\left(\lambda\xi_T\right), G_T^{\text{det}}\right\}, \tag{11}$$

where $I_B(\cdot) = \left(U_B'\right)^{-1}(\cdot) = \sum_{i=1}^n I_i\left(\frac{\cdot}{\beta_i}\right)$ is the inverse marginal utility function of the fund manager given in (9) and λ is determined to make the budget constraint binding. The optimal collective terminal wealth (11) is a decreasing function in the state price density ξ_T beyond the total guarantee and stays at the guarantee level once it is reached. Having determined the optimal terminal payoff, we can use the martingale representation theorem to determine the corresponding optimal investment strategy.

3.3 Investment strategy

In this setting, we provide a semi-analytical solution for the investment strategy. To do this, we first introduce Lemma 3.1, where we point out that the guarantee can be expressed in terms of $I_B(\lambda\bar{\xi})$ for some fixed value $\bar{\xi}$:[5]

Lemma 3.1. *For any level of the guarantee* $G_T^{det} \in [0,(1-\bar{\epsilon})xe^{rT})$ *we can find a unique value* $\bar{\xi} \in [0,\infty]$ *such that*

$$G_T^{det} = I_B\left(\lambda\bar{\xi}\right). \tag{12}$$

In particular, we can make the following decomposition:

$$X_T^* = \max\left\{I_B\left(\lambda\xi_T\right), G_T^{det}\right\} = I_B\left(\lambda\xi_T\right)\mathbb{1}_{\left\{\xi_T<\bar{\xi}\right\}} + G_T^{det}\mathbb{1}_{\left\{\xi_T\geq\bar{\xi}\right\}}.$$

Proof: See Appendix A.1.

We can now explicitly determine the optimal wealth at $t \in [0,T)$ and the corresponding self-financing investment strategy.

Proposition 3.2. *The optimal wealth at time* $t \in [0,T)$ *is given by*

$$\begin{aligned} X_t^* = &\sum_{i=1}^n I_i\left(\frac{\lambda}{\beta_i}\xi_t\right)k_i(t)\left(1-\Phi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\right) \\ &+ G_T^{det}e^{-r(T-t)}\Phi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\right), \end{aligned} \tag{13}$$

where $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution,*

$$k_i(t) := e^{\left(1-\frac{1}{\gamma_i}\right)\left(-r-\frac{1}{2}\eta^2\right)(T-t)+\frac{1}{2}\eta^2\left(1-\frac{1}{\gamma_i}\right)^2(T-t)} \tag{14}$$

and

$$d(t,\xi_t,\bar{\xi}) := \frac{\ln\xi_t - \ln\bar{\xi} - r(T-t) - \frac{1}{2}\eta^2(T-t)}{\eta\sqrt{T-t}}, \tag{15}$$

[5] Note that Lemma 3.1 does not cover the case $G_T^{\text{det}} = (1-\bar{\epsilon})xe^{rT}$. However, for this case we already know that the corresponding trading strategy is just investing 100% in the risk-free asset.

with $\bar{\xi}$ determined from (12). *The optimal fraction of wealth invested in the risky asset is obtained by*

$$\pi_t^* = \frac{1}{\sigma X_t^*}\left[\sum_{i=1}^{n}\frac{1}{\gamma_i}\left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}}\eta k_i(t)\left(1-\Phi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\right)\right.$$
$$+\sum_{i=1}^{n}\left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}}k_i(t)\varphi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\frac{1}{\sqrt{T-t}} \tag{16}$$
$$\left.-G_T^{det}e^{-r(T-t)}\varphi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\right)\frac{1}{\sqrt{T-t}}\right],$$

where $\varphi(\cdot)$ is the density of the standard normal distribution.

Proof: See Appendix A.2.

As mentioned before, the optimal wealth (13) and the optimal investment strategy (16) are expressed in terms of ξ_t. Note that we can easily come up with an expression of these optima in terms of the stock price S_t by using equation (7). A numerical analysis of the strategy given in (16) is given in Section 3.5. Before coming to this, let us briefly introduce the sharing rule used to redistribute the collective terminal wealth to each investor in the following subsection.

3.4 Sharing rules

In the above subsection, if the fund manager follows the suggested optimal investment strategy, it is ensured that the fund value will always lie above or at the guaranteed level G_T^{det}. A natural choice of G_T^{det} is to add up the individual guarantees

$$G_T^{\text{det}} = \sum_{i=1}^{n}(1-\bar{\epsilon})x_i e^{g_i T},$$

where $g_i \leq r$ denotes the level of guaranteed interest rate that investor i requires which might be negative. A negative value of g_i implies $e^{g_i T} < 1$, which economically means that investor i is only interested in obtaining a fraction of her net initial wealth $((1-\bar{\epsilon})x_i)$ as a guarantee from the fund manager. Here, we assume that the fund manager wants to serve each investor i at least with her guaranteed amount corresponding to the initial wealth subtracted by fees accumulated with an interest rate guarantee g_i. This choice of individual guarantees results in $G_T^{\text{det}} \leq (1-\bar{\epsilon})xe^{rT}$. In order to ensure that each individual investor obtains a terminal amount at least as high as her desired guarantee level, the fund manager needs to fix a reasonable sharing rule. Note first that a simple linear sharing rule from which each investor receives a fixed percentage of the terminal wealth cannot do the job. In what follows, we use

$X_T^{(i)}$ to denote investor i's terminal wealth obtained from the collective fund and we assume that the total terminal wealth X_T^* is redistributed to the participants in the following way:

- If $X_T^* = G_T^{\text{det}}$, each participant obtains her guarantee. We obtain $X_T^{(i)} = (1-\bar{\epsilon})x_i e^{g_i T}$.
- If $X_T^* > G_T^{\text{det}}$, each investor receives the above individual guarantee plus a fraction of the surplus $X_T^* - G_T^{\text{det}}$. We assume that the surplus part is shared proportionally among the participants by the proportions $(\alpha_1, \ldots, \alpha_n)$, where $\sum_{i=1}^n \alpha_i = 1$. Then investor i's payoff is given by $X_T^{(i)} = (1-\bar{\epsilon})x_i e^{g_i T} + \alpha_i \left(X_T^* - G_T^{\text{det}}\right)$.

By using this state-dependent sharing rule, we assume that the fund manager's primal goal is to provide the desired guarantees to all the investors. In total, we can write the individual payoff as:

$$\begin{aligned} X_T^{(i)} &= (1-\bar{\epsilon})x_i e^{g_i T} + \alpha_i \left(X_T^* - G_T^{\text{det}}\right) \\ &= (1-\bar{\epsilon})x_i e^{g_i T} + \alpha_i \left(I_B(\lambda\xi_T) - G_T^{\text{det}}\right)^+, \end{aligned} \tag{17}$$

where $(v)^+ := \max\{v, 0\}$, $v \in \mathbb{R}$. The payoff to an investor in the collective is thus similar to the payoff of a unit-linked life insurance, as described, for example, in Brennan and Schwartz (1976): It is given by the individual guarantee and a bonus that is shared among all the participants. The bonus in (17) can be seen as a call option with $I_B(\lambda\xi_T)$ as underlying and strike G_T^{det}.

The question now is how the vector $(\alpha_1, \ldots, \alpha_n)$ shall be chosen. We consider the following choices of proportional sharing rules:[6]

- **Sharing rule 1 (SR 1)**: We choose the vector $(\alpha_1, \ldots, \alpha_n)$ in such a way that the financial fairness condition as defined in Bühlmann and Jewell (1979) and Schumacher (2018) is fulfilled. While these articles consider financial fairness in a risk exchange setting, Orozco-Garcia and Schmeiser (2019) also analyze financial fairness in a life insurance setting and Boelaars and Broeders (2019) in the context of collective defined contribution schemes. In our setting with management fees, this results in a "pseudo financial fairness" condition:

$$\begin{aligned} (1-\bar{\epsilon})x_i &= \mathbb{E}\left[\xi_T\left((1-\bar{\epsilon})x_i e^{g_i T} + \alpha_i\left(X_T^* - G_T^{\text{det}}\right)\right)\right] \\ &= (1-\bar{\epsilon})x_i e^{(g_i - r)T} - \alpha_i G_T^{\text{det}} e^{-rT} + \alpha_i(1-\epsilon)x. \end{aligned}$$

[6]Following the terminology of Schumacher (2018), we limit ourselves to proportional sharing rules, i.e. each investor receives a fixed percentage of the state-dependent terminal wealth. Proportional sharing rules are within the class of linear sharing rules which would also allow a fixed amount to be added to or subtracted from the fixed percentage.

We use "pseudo" as the initial market value of what the individuals are entitled to is equal to the net investments $(1-\bar{\epsilon})x_i$ after paying the delegation fee. Note that, as $\epsilon_1 < \bar{\epsilon} < \epsilon_2$, the members in Group 1 receive less than what they are entitled to, while those in Group 2 receive more than what they are entitled to. The above sharing rule can be rearranged to

$$\alpha_i = \frac{x_i - x_i e^{(g_i - r)T}}{x - \sum_{j=1}^{n} x_j e^{g_j T} e^{-rT}}. \tag{18}$$

This is a generalization of an often-used sharing rule provided below where only the case without guarantee is considered (see Jensen and Nielsen (2016) and Branger et al. (2018a)). In contrast to this literature, we apply the sharing rule (18) to the bonus exceeding the collective guarantee. Additionally, we want to emphasize that the use of our pseudo financially fair sharing rule only affects how the total terminal wealth is shared and is not related to an efficiency criterion.

- **Sharing rule 2 (SR 2)**: A popular sharing rule used in practice is a simpler form of SR 1 (see also Jensen and Nielsen (2016) or Branger et al. (2018a)):

$$\alpha_i(x_i) = \frac{x_i}{x}. \tag{19}$$

- **Sharing rule 3 (SR 3)**: An additional sharing rule which does not necessarily fulfill the financial fairness condition is:

$$\alpha_i(n) = \frac{1}{n}. \tag{20}$$

Both SR 2 and 3 are frequently used in practice, as they are very easy-to-communicate. A few special cases can be listed:

- If all the individuals have the same initial wealth level, that is, $x_i = \frac{x}{n}$, then the sharing rules (20) and (19) coincide.
- If all the individuals have the same initial wealth level, that is, $x_i = \frac{x}{n}$ and all the individual guarantee levels are identical, that is, $g_i = g \leq r$ for all $i = 1, \ldots, n$, then the sharing rules SR 1, SR 2 and SR 3 coincide. Therefore, in our numerical analysis, we consider cases with different initial wealth levels and guarantees.

While it is rather intuitive that Group 1 most likely suffers utility losses through the collective investment under most sharing rules, it is not clear how each individual investor in Group 2 is affected by the collective investment, the guarantee and the sharing rules.

3.5 Numerical analyses

In this section we perform numerical analyses to illustrate the impact of the guarantee and sharing rules on the investors in the collective. In particular, our aim is to find out how each of the investor's well-being is influenced by requiring a deterministic guarantee amount, compared to her individual optimal solution under different sharing rules. To assess utility losses and gains, we follow Jensen and Sørensen (2001) and Jensen and Nielsen (2016) by considering the *wealth equivalent*, which is defined as the initial wealth level that is needed for an investor to achieve the same level of expected utility in the individual optimal investment problem as can be achieved with the initial wealth x_i in the collective optimization problem. A straightforward calculation shows that the optimal level of expected utility in the benchmark optimization problem is given by

$$\begin{aligned} U_i^*((1-\epsilon_i)x_i) &= \mathbb{E}\left[\frac{(X_T^{(i,*)})^{1-\gamma_i}}{1-\gamma_i}\right] \\ &= \frac{1}{1-\gamma_i}\left((1-\epsilon_i)x_i\right)^{1-\gamma_i} e^{\left(r+\frac{1}{2}\frac{(\mu-r)^2}{\gamma_i\sigma^2}\right)(1-\gamma_i)T}. \end{aligned}$$

See also Jensen and Sørensen (2001). We denote by $\overline{U}_i((1-\bar{\epsilon})x_i)$ the expected utility obtained from the collective optimization problem with initial wealth x_i and fee $\bar{\epsilon}$. Following Jensen and Sørensen (2001) and Jensen and Nielsen (2016), the mathematical definition of the wealth equivalent WE_i is

$$U_i^*((1-\epsilon_i)\mathrm{WE}_i) = \overline{U}_i((1-\bar{\epsilon})x_i) \quad \Leftrightarrow \quad \mathrm{WE}_i = \left(\frac{\overline{U}_i((1-\bar{\epsilon})x_i)}{U_i^*(1-\epsilon_i)}\right)^{\frac{1}{1-\gamma_i}}. \tag{21}$$

To make our analysis independent of the initial wealth levels, we divide the wealth equivalent by the initial wealth to obtain

$$y_i = \frac{\mathrm{WE}_i}{x_i}. \tag{22}$$

Collective investments are beneficial for an individual investor if $y_i \geq 1$, otherwise the individual investor suffers a loss through the collective investment.

Throughout all the numerical analyses we use the following parameters:

- Following Branger et al. (2018b), we choose the weights β_i as

$$\beta_i = \frac{1/\lambda_i}{\sum_{j=1}^n 1/\lambda_j}, \tag{23}$$

  where λ_i is the Lagrangian multiplier in (6). Note that the weights β_i may not be chosen arbitrarily as they shall neutralize the units of the different utility functions. Let us ignore the max-operator in U_B and only focus on the units:

$$U_B(v) = \sum_{i=1}^{n} \beta_i U_i(v_i) = \frac{1}{\sum_{j=1}^{n} 1/\lambda_j} \sum_{i=1}^{n} \frac{1}{\lambda_i} \cdot \frac{v_i^{1-\gamma_i}}{1-\gamma_i} \, ,$$

where the monetary terms taken to a power cancel out the "unnecessary" units: $1/\lambda_i = (1-\epsilon_i)x_i^{\gamma_i} \exp(\dots)$ and $v_i^{1-\gamma_i}$. Note that the constant in the power of v leads to the unit being in monetary terms again in all summands. Treating all the other terms as constants, there is no problem by adding up the terms in U_B as they all have the same unit.[7] That is, for our theoretical results, we have tacitly assumed that the weights are chosen in such a way that the units in the utility functions do not cause any problems. As shown in Branger et al. (2018b), under this choice of weights individual optimal solutions are achievable in Problem (10) if the collective guarantee is equal to zero. In other words, in the following, we compare the collective optimization problem with individual fees and no guarantees to a collective optimization problem with an average fee and guarantee constraints.

- The parameters of the financial market are $r = 0.02$, $\mu = 0.07$, $\sigma = 0.12$, $T = 1$.[8]

We consider the following parameter setup:

- The number of participants in the pool is $n = 100$. Note that in the existing literature, frequently only two investors are considered (see for example Dumas (1989), Weinbaum (2009) and Jensen and Nielsen (2016)). While the assumed number of employees is reasonable for small or medium size plans, there are usually more than 100 members in most large pension plans.

- To capture the heterogeneity of risk preferences in pension plans (see for example Alserda et al. (2019)), we assume that the risk aversion parameters of the investors in the collective differ rather drastically.[9] We follow Chiappori and Paiella (2011) who observe that relative risk aversion (RRA) follows a right-skewed distribution.[10] To keep our analysis simple and tractable, we capture this right-skewness by assuming an exponential distribution which

[7]Note that a rather simple choice like $\beta_i = 1/n$ for all $i = 1, \dots, n$ would imply that different units are added up in the utility function U_B which would lead to inconsistent results.

[8]For simplicity, we consider a rather short investment horizon. Note that our results do not change under longer maturities.

[9]Note that our model setup is only one possible way of representing a collective of individuals with heterogeneous risk preferences. For example, Atmaz and Basak (2018) focus on the equilibrium price in the presence of belief heterogeneity using a different modeling approach. To be precise, they assume that an investor's type is normally distributed and the representative investor maximizes the weighted average of individual investors' utility.

[10]Note that there exist many findings in the literature that differ from Chiappori and Paiella (2011): For

is shifted by a constant number to be in line with estimates on the RRA in Chiappori and Paiella (2011). To be more precise, we design the two groups in the following way:

– Shaw (1996) finds that "more educated individuals are more likely to be risk takers". Therefore, Group 1 is the one with (on average) the lower RRA and Group 2 is the one with (on average) the higher RRA. Chiappori and Paiella (2011) "find a small but significant negative correlation between wealth and risk aversion", a finding which is also supported by Shaw (1996). Therefore, we assume that Group 1 is, in total, more wealthy than Group 2. For simplicity, we assume, similar to Jensen and Sørensen (2001), that $x_i = 1$ for the first group. That is, for Group 1, the wealth equivalent can be seen not only as an amount of capital but also as a percentage of the initial wealth. The guaranteed interest rates of Group 1 are given by $g_i = 0$ for all $i = 1, \ldots, n_1$ and the fee ϵ_1 equals 0.25%. The investors in Group 2 have a lower initial wealth level than Group 1 and demand a higher guarantee. Their initial wealth levels are given by $x_i = 0.5$ and their guaranteed interest rates are given by $g_i = 0.01$ for all $i = n_1 + 1, \ldots, n$. The fee that Group 2 is willing to pay is equal to $\epsilon_2 = 1\%$. We assume that the fund manager charges the average fee in such a way that the fees in the two collective problems we compare are identical, i.e. as

$$\begin{aligned} \bar{\epsilon}x &= \epsilon_1 \sum_{i=1}^{n_1} x_i + \epsilon_2 \sum_{i=n_1+1}^{n} x_i \\ \Leftrightarrow \ \bar{\epsilon} &= \frac{\epsilon_1 \sum_{i=1}^{n_1} x_i + \epsilon_2 \sum_{i=n_1+1}^{n} x_i}{x} = 0.5\% \end{aligned}$$

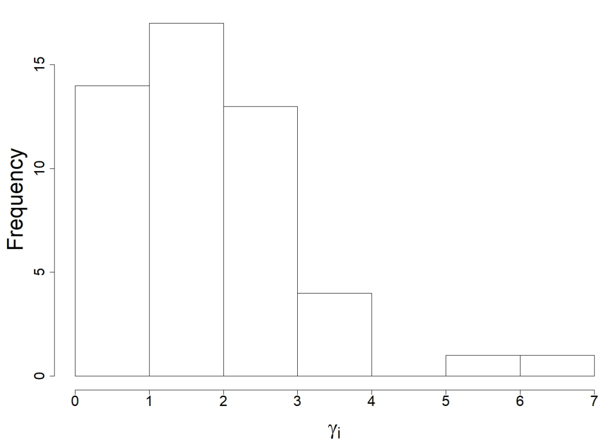
which is in line with the values provided in Malkiel (2013). To emphasize the effects of the fee structure on the well-being of each investor, we also consider the case $\epsilon_2 = 2.5\%$, i.e. $\bar{\epsilon} = 1\%$.

– We assume that the RRA coefficient of any individual does not fall below 0.35. This assumption is based on Table 1 in Conine et al. (2017) who provide an excellent literature review on estimates of RRA in the past. For Group 1 we have then simulated $n_1 = n/2$ exponentially distributed random variables with parameter $\lambda_1 = 1/1.35$. For Group 2, we have simulated $n - n_1 = n/2$ realizations of an exponentially distributed random variable with parameter $\lambda_2 = 1/3$. To all these realizations, we have added 0.35. The mean RRA in the collective is then 2.5 and the median is 1.72, in line with Chiappori and Paiella (2011). Furthermore, Chiappori and Paiella (2011) report that 25% of the population have a RRA above 3. In our

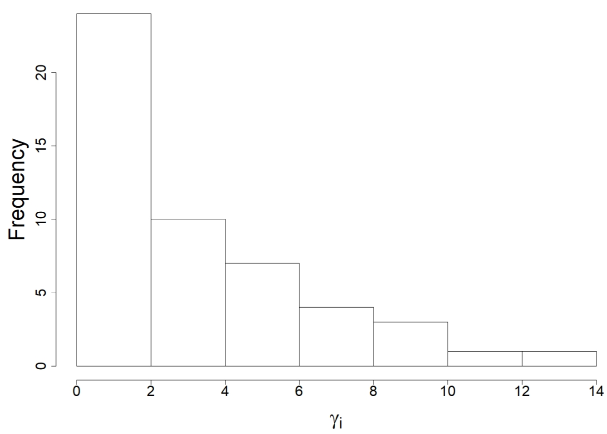
example, Barsky et al. (1997) find that 65% of their data shows a RRA above 3.76, 24% below 2 and 12% between 2 and 3.76. Davies (1981) describes 3, 4 and 5 as most reasonable estimates of RRA and Azar (2010) estimates 3.01 and 3.74 as interval bounds for the RRA. As we have decided to follow Chiappori and Paiella (2011), our numerical analyses should only be seen as examples and other, different compositions of the RRAs can also be implemented.

sample, there are exactly 25% (i.e. 25 individuals) with a RRA above 3 as well.

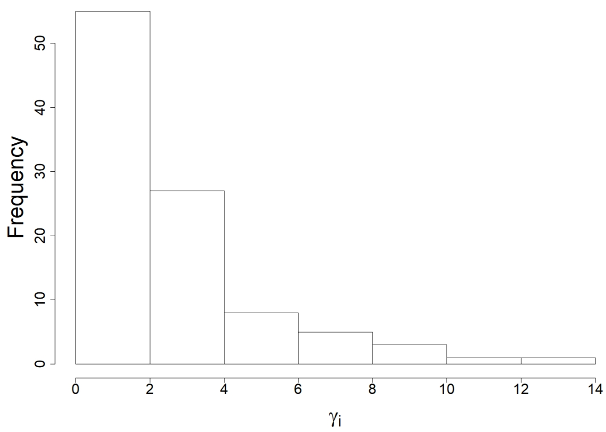
Histograms of the RRA of the two groups and the collective (both groups put together) are given in Figure 1. The actual γ_i values are available from the authors upon request.



(a) Group 1



(b) Group 2

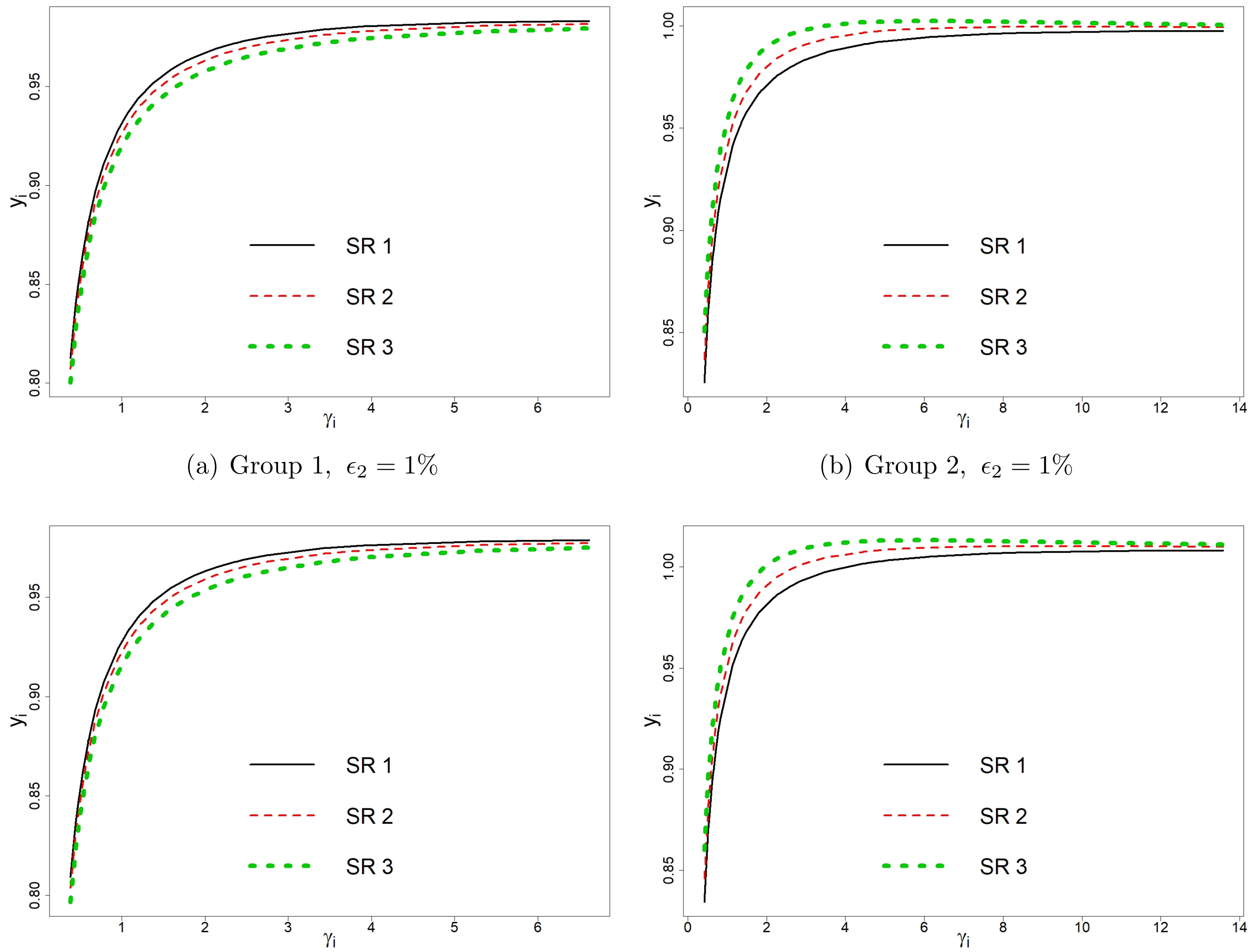


(c) Collective (both groups)

Figure 1: Histograms of the RRA coefficients in Group 1, Group 2 and the collective (both groups put together). For Group 1 we have simulated $n_1 = n/2$ exponentially distributed random variables with parameter $\lambda_1 = 1/1.35$. For Group 2, we have simulated $n - n_1 = n/2$ realizations of an exponentially distributed random variable with parameter $\lambda_2 = 1/3$. Then, we have added 0.35 to all the realizations.

The wealth equivalents for the two groups are provided in Figure 2. We observe the following:

- In both Groups, investors with a RRA below 3 suffer drastic losses due to the guarantee constraint. Recall that 75% of the collective considered have such a RRA below 3.
- **Group 1 / Panels (a) and (c):** Naturally, the members of the first group have utility losses as the average fee is higher than the fee they are willing to pay. All the investors



(a) Group 1, $\epsilon_2 = 1\%$ (b) Group 2, $\epsilon_2 = 1\%$

(c) Group 1, $\epsilon_2 = 2.5\%$ (d) Group 2, $\epsilon_2 = 2.5\%$

Figure 2: Comparison of the wealth equivalents divided by initial wealth of the members of the two groups ordered by RRA. For each group, the sharing rules 1 (18), 2 (19) and 3 (20) are compared.

in this group suffer utility losses since their wealth equivalent divided by initial wealth is below 1. The lower the RRA is, the more severe the loss in utility is. The first sharing rule leads to the highest wealth equivalents. This is due to the fact that the first sharing rule takes the guarantees into account. Since Group 2 demands a higher guarantee than Group 1, Group 1 has to (at least partially) finance the guarantees of Group 2 in the second and third sharing rule. The second sharing rule outperforms the third sharing rule because the second one takes account of the initial wealth levels. As Group 1 is more wealthy than Group 2, Group 1 suffers the largest losses under SR 3, because this sharing rule does not reflect the investors' initial wealth levels.

- **Group 2 / Panels (b) and (d)**: Whether Group 2 benefits from the collective (and thus cheaper) fee depends substantially on the magnitude of the fee. In Panel (d), investors with a RRA above 4 (16 investors) benefit from the collective investment under the first sharing rule, investors with a RRA above 2.8 (21 investors) benefit from the collective investment under the second sharing rule and investors with a RRA above 2 (26 investors) benefit from the collective investment under the third sharing rule. The main reason for this should be the lower fee $\bar{\epsilon}$ which returns more to Group 2 than they would be entitled to without the collective investment. Note that mainly investors with a rather high RRA benefit from the collective investment as the guarantee leads to drastic losses for investors with a low RRA. In Panel (b), under SR 1 and SR 2, all the investors are worse off than in the benchmark case since y_i is below 1 for all investors. Under SR 3, investors with a RRA above 3.3 (19 investors) benefit from the collective investment. Regarding the sharing rules, the order of attractiveness is exactly opposite to Group 1. By demanding higher guarantees, Group 2 benefits from Group 1 under the second sharing rule. Being the less wealthy group, Group 2 additionally benefits from the more wealthy Group 1 under the third sharing rule.

Let us now take a look at the investment strategy. In Figure 3 we show the investment strategy at time $t = T/2$ given in (16) depending on the stock return S_t/S_0 for the fee $\epsilon_2 = 1\%$, i.e. $\bar{\epsilon} = 0.5\%$.[11] We have used equation (7) to express the strategy in (16) in terms of the stock return S_t/S_0.[12] As only 25% of the investors in the collective have a RRA above 3,

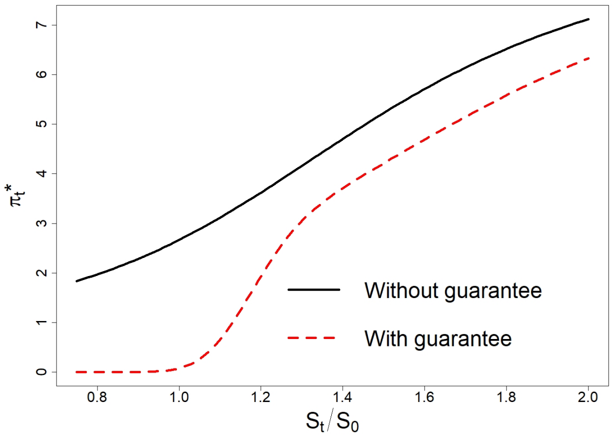


Figure 3: Optimal fraction of wealth invested in the risky asset at time $t = T/2$ depending on the stock return at that time for $\bar{\epsilon} = 0.5\%$. We show the optimal collective investment strategy with and without guarantee.

both optimal investment strategies suggest investing more than 100% of wealth in the risky asset if the market performs well. Naturally, this fraction is lower under a portfolio insurance constraint than with no constraint.

[11] A figure with $\epsilon_2 = 2.5\%$, i.e. $\bar{\epsilon} = 1\%$ looks almost identical to Figure 3, which is why we omit it here.

[12] Note that the investment strategy for the case with no guarantee can be obtained by letting G_T^{det} tend to 0.

In the numerical analysis, we have seen that investors are likely to suffer a loss from deterministic guarantees. Especially investors with a RRA below 3 (which are 75% of the collective in our case) suffer large losses from the guarantee. Investors in the collective can only be better off compared to their benchmark optimization problem at the cost of other investors' well-being. As the deterministic guarantee framework provides a rationale why a safe investment strategy is adopted, our results are to some extent consistent with the results of Alserda et al. (2019) who observe that current asset allocations of pension plans are "safer than implied by members' preferences". In the following, we introduce a state-dependent guarantee scheme which is more flexible and better suited for each investor's risk appetite.

4 Optimization problem with state-dependent guarantees

We learn from the above analyses that a deterministic guarantee is likely to deteriorate the benefits of all investors in the pool. Inspired by Grossman and Zhou (1993), Browne (1999), Tepla (2001) and Deelstra et al. (2004), we now consider the case where the fund manager aims to exceed a more flexible, state-dependent guarantee which becomes known at maturity T. As argued in these articles, it is often the goal of an investor to exceed a given state-dependent benchmark, which might, for example, be a market index. In our case, we assume that individuals (who do not invest on their own behalf) fix a deterministic guarantee part plus a fraction of their own optimal terminal wealth as benchmark, where they can control the magnitude of both components. This leads to the following total guarantee which the fund manager needs to exceed:

$$\begin{aligned} G_T^{\mathrm{Fle}}(\xi_T) &:= \sum_{i=1}^{n}\left((1-\bar{\epsilon})p_i x_i e^{g_i T} + \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)X_T^{(i,*)}\right) \\ &= (1-\bar{\epsilon})G + (1-\bar{\epsilon})\sum_{i=1}^{n}\frac{(1-p_i)}{(1-\epsilon_i)}X_T^{(i,*)}, \end{aligned} \tag{24}$$

where $G = \sum_{i=1}^{n} p_i x_i e^{g_i T}$ is the sum of the deterministic guarantee components in (24) (divided by $(1-\bar{\epsilon})$), $p_i \in [0,1]$, $g_i < r$ and $X_T^{(i,*)}$ is the individual unrestricted terminal wealth of investor i as given in (6) for all $i = 1, \ldots, n$.[13] Note that $X_T^{(i,*)}$ contains a component $(1-\epsilon_i)$. The state-dependent part in (24) represents an unrestricted terminal wealth starting with an initial wealth $(1-\bar{\epsilon})x_i$. In other words, we lower the deterministic part of the guarantee and add a state-dependent part which corresponds to a fraction of the optimal unrestricted terminal wealth. Now each investor can choose for herself the fraction of the deterministic guarantee

[13] The flexible guarantee we present can be seen as one possible example of the many flexible guarantees that can be realized in today's world which could help increase welfare by lowering deterministic guarantees.

which she is willing to give up. The lower the deterministic part of the guarantee is, the higher (lower) the total flexible guaranteed payment gets if the market performs well (badly). So a lower deterministic guarantee leads to a more risky, state-dependent guaranteed payoff. There are two special cases that deserve further attention: If $p_i = 0$ for all i, then every investor wants to obtain her individual optimal terminal wealth. For $p_i = 1$ for all i, we are back to the optimization problem with a deterministic guarantee level. In addition, the choice of $G_T^{\text{Fle}}(\xi_T)$ ensures that the initial value of this total guarantee is not higher than the total initial wealth $(1-\bar{\epsilon})x$ available for investment, that is, $\mathbb{E}\left[\xi_T G_T^{\text{Fle}}\right] \leq (1-\bar{\epsilon})x$.

We now proceed in the same manner as in Section 3: We start by deriving the optimal terminal wealth followed by the optimal investment strategy and the optimal wealth at each time $t < T$. Although these two subsections are more technical than the case with deterministic guarantees, the methodology is exactly the same. We then come to a short discussion of sharing rules and carry out the numerical analyses similar to Section 3.

4.1 Optimal terminal wealth

The optimization problem under the state-dependent guarantee (24) can be formulated as follows:

$$\begin{aligned} \max_{X_T} \mathbb{E}\left[U_B(X_T)\right] \quad \text{s.t. } & \mathbb{E}\left[\xi_T X_T\right] = (1-\bar{\epsilon})x, \\ & X_T \geq G_T^{\text{Fle}}(\xi_T) \text{ a.s.} \end{aligned} \tag{25}$$

In the following proposition, we determine the optimal solution of this optimization problem.

Proposition 4.1. *Assuming that $p_i > 0$ for at least one i, the solution to the optimal terminal wealth is given by*

$$\widehat{X}_T^* = \max\left\{G_T^{Fle}(\xi_T), I_B(\hat{\lambda}\xi_T)\right\}, \tag{26}$$

where $\hat{\lambda}$ makes the budget constraint binding.

Proof. For any single event $\omega \in \Omega$ (any scenario of the financial market) and $\tilde{\lambda} > 0$ given, consider the following static optimization problem:[14]

$$\widehat{X}_T^*(\tilde{\lambda})(\omega) := \underset{X \geq G_T^{\text{Fle}}(\omega)}{\operatorname{argmax}} \left(U_B(X) - \tilde{\lambda}\xi_T(\omega)X\right).$$

[14]Throughout this proof we use the notation $G_T^{\text{Fle}}(\omega) := G_T^{\text{Fle}}(\xi_T(\omega))$.

Due to the concavity of $U_B(\cdot)$, we get $\widehat{X}_T^*(\tilde{\lambda})(\omega) = I_B(\tilde{\lambda}\xi_T(\omega))$ if $I_B(\tilde{\lambda}\xi_T(\omega)) \geq G_T^{\text{Fle}}(\omega)$ and $\widehat{X}_T^*(\tilde{\lambda})(\omega) = G_T^{\text{Fle}}(\omega)$ if $I_B(\lambda\xi_T(\omega)) < G_T^{\text{Fle}}(\omega)$, or equivalently

$$\widehat{X}_T^*(\tilde{\lambda})(\omega) = \max\left\{I_B(\tilde{\lambda}\xi_T(\omega)), G_T^{\text{Fle}}(\omega)\right\}.$$

The solution of the above static problem defines $\widehat{X}_T^*(\tilde{\lambda}) := \max\{I_B(\tilde{\lambda}\xi_T), G_T^{\text{Fle}}(\xi_T)\}$ which is an $\mathcal{F}_T$-measurable random variable. Next, we will show that $\widehat{X}_T^*(\tilde{\lambda})$ is an admissible terminal wealth for problem (25) for some $\tilde{\lambda} > 0$. Indeed, consider $\Psi(\tilde{\lambda}) := \mathbb{E}\left[\xi_T \max\left\{I_B\left(\tilde{\lambda}\xi_T\right), G_T^{\text{Fle}}(\xi_T)\right\}\right]$. It is clear that $\Psi(\tilde{\lambda})$ is continuous and decreasing in $\tilde{\lambda}$, that $\lim_{\tilde{\lambda}\to 0}\Psi(\tilde{\lambda}) = \infty$ and that $\lim_{\tilde{\lambda}\to\infty}\Psi(\tilde{\lambda}) = \mathbb{E}\left[\xi_T(1-\bar{\epsilon})G\right] < (1-\bar{\epsilon})x$. Therefore, by the theorem of intermediate values, there exists $\hat{\lambda} > 0$ such that $(1-\bar{\epsilon})x = \Psi(\hat{\lambda}) = \mathbb{E}\left[\xi_T\widehat{X}_T^*(\hat{\lambda})\right]$, that is, $\widehat{X}_T^*(\hat{\lambda})$ is admissible. Now we show that $\widehat{X}_T^*(\hat{\lambda})$ is an optimal solution of Problem (25). To this end, let Y_T be an arbitrary admissible terminal payoff, that is, $\mathbb{E}\left[\xi_T Y_T\right] = (1-\bar{\epsilon})x$ and $Y_T \geq G_T^{\text{Fle}}(\xi_T)$. We have

$$\begin{aligned}
\mathbb{E}\left[U_B(Y_T)\right] &= \mathbb{E}\left[U_B(Y_T)\right] + \hat{\lambda}((1-\bar{\epsilon})x - \mathbb{E}\left[\xi_T Y_T\right]) \\
&= \mathbb{E}\left[U_B(Y_T) - \hat{\lambda}\xi_T Y_T\right] + \hat{\lambda}(1-\bar{\epsilon})x \\
&\leq \mathbb{E}\left[\max_{X \geq G_T^{\text{Fle}}(\xi_T)}\left(U_B(X) - \hat{\lambda}\xi_T X\right)\right] + \hat{\lambda}(1-\bar{\epsilon})x \\
&= \mathbb{E}\left[U_B(\widehat{X}_T^*(\hat{\lambda})) - \hat{\lambda}\xi_T\widehat{X}_T^*(\hat{\lambda})\right] + \hat{\lambda}(1-\bar{\epsilon})x \\
&= \mathbb{E}\left[U_B(\widehat{X}_T^*(\hat{\lambda}))\right],
\end{aligned}$$

where we have used $(1-\bar{\epsilon})x = \mathbb{E}\left[\xi_T\widehat{X}_T^*(\hat{\lambda})\right]$. Hence, we can conclude that $\widehat{X}_T^*(\hat{\lambda})$ is an optimal solution. □

If $p_i > 0$ for at least one i, a unique solution to optimization problem (25) always exists and is given by (26), since the present value of the guarantee $G_T^{\text{Fle}}(\xi_T)$ is smaller than the capital initially invested. If $p_i = 0$ for all i, on the other hand, then the present value of the guarantee is given by $\mathbb{E}\left[\xi_T G_T^{\text{Fle}}(\xi_T)\right] = (1-\bar{\epsilon})x$. As a consequence, the solution to problem (25) has to, necessarily, be given by $\widehat{X}_T^* = G_T^{\text{Fle}}(\xi_T) = \sum_{i=1}^n \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)} X_T^{(i,*)}$.[15] Hence, problem (25) admits a unique solution for all possible choices of the p_i's.

Having determined the optimal terminal payoff, we can again compute the corresponding investment strategy in a similar way as in Section 3. Let us, however, point out that some further assumptions need to be made to conduct similar analyses as for the case with a deterministic guarantee.

[15]This can be seen as follows: If $\widehat{X}_T^* < G_T^{\text{Fle}}(\xi_T)$ with positive probability, then $\widehat{X}_T^*$ cannot be admissible to problem (25). If, on the other hand, $\widehat{X}_T^* > G_T^{\text{Fle}}(\xi_T)$ with positive probability, this would be a contradiction to the budget constraint.

4.2 Investment strategy

Note that the optimal terminal wealth for problem (25) can be derived *explicitly* for all choices of p_i's. In this section, we show that the optimal investment strategies can be determined for all the choices of p_i's, too. However, an *explicit* derivation, similar to Section 3.3, is only possible if we impose further conditions on the p_i's. From Section 3.3, we learn that the investment strategy for the optimization problem with a deterministic guarantee can be determined explicitly, as we can find a unique critical value of the state price density $\bar{\xi}$ which allows us to decompose the optimal terminal wealth into two parts. A natural question arises whether similar analyses can be carried out for the case with a flexible guarantee. The following lemma deals with this question.

Lemma 4.2. *If $p_i > 0$ for at least one i and*

$$\sum_{i=1}^{n} \frac{1}{\gamma_i} \xi_T^{-\frac{1}{\gamma_i}} \left(\left(\frac{\hat{\lambda}}{\beta_i} \right)^{-\frac{1}{\gamma_i}} - (1-p_i) \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)} \lambda_i^{-\frac{1}{\gamma_i}} \right) > 0 \quad a.s., \tag{27}$$

there exists a unique positive number $\hat{\xi}$ such that

$$\widehat{X}_T^* = I_B\left(\hat{\lambda}\xi_T\right) \mathbb{1}_{\{\xi_T < \hat{\xi}\}} + G_T^{Fle}(\xi_T) \mathbb{1}_{\{\xi_T \geq \hat{\xi}\}}.$$

Proof: See Appendix A.3.

The assumption that $p_i > 0$ for at least one i is no restriction at all, because the investment strategy for the case where $p_i = 0$ for all $i = 1, \ldots, n$ is already known.[16] It is, however, important to note that condition (27) is satisfied if

$$\left(\frac{\hat{\lambda}}{\beta_i} \right)^{-\frac{1}{\gamma_i}} - \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)} (1-p_i) \lambda_i^{-\frac{1}{\gamma_i}} > 0 \quad \text{for all } i = 1, \ldots, n. \tag{28}$$

Equation (28) imposes an implicit condition for the choice of p_i's. In other words, we can see that we are unable to choose the p_i's freely, if we want to determine the self-financing investment strategy in an explicit form. However, as $\hat{\lambda}$ depends on all the different p_i's, we cannot easily come up with an explicit lower bound for all the p_i's. If condition (27) is not fulfilled, this means that there might be more than one intersection between $G_T^{\text{Fle}}(\xi_T)$ and $I_B(\hat{\lambda}\xi_T)$.[17]

[16] As we know from Branger et al. (2018b), we can obtain the resulting collective terminal wealth by letting G_T^{det} tend to 0 in (16) and choosing $\beta_i = \frac{1/\lambda_i}{\sum_{j=1}^{n} 1/\lambda_j}$.

[17] For example, there might be two intersections. In this case, the flexible guarantee is effective in both good and bad market scenarios. Hence, the guarantee would not only protect an individual from bad market scenarios but could also offer some upside potential in good market scenarios compared to fixed guarantees.

In this case, we need to rely on numerical procedures to determine the investment strategy. One possible approach for this will be briefly described in Section 4.4.

Using the result of Lemma 4.2, we can again compute the optimal wealth at $t \in [0,T)$ and the corresponding optimal self-financing investment strategy. Recall that η is defined in (2).

Proposition 4.3. *Under the assumptions of Lemma 4.2, the optimal wealth at* $t \in [0,T)$ *is given by*

$$\begin{aligned}\widehat{X}_t^* &= \sum_{i=1}^n I_i\left(\frac{\hat{\lambda}}{\beta_i}\xi_t\right)k_i(t)\left(1-\Phi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\right)\\ &\quad+(1-\bar{\epsilon})Ge^{-r(T-t)}\Phi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\right)\\ &\quad+\sum_{i=1}^n\frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)I_i\left(\lambda_i\xi_t\right)k_i(t)\Phi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right),\end{aligned} \tag{29}$$

where $\hat{\xi}$ *is determined such that* $G_T^{Fle}(\hat{\xi}) = I_B(\hat{\lambda}\hat{\xi})$, *while* $d(t,\xi_t,\hat{\xi})$ *and* $k_i(t)$ *are defined as in* (15) *and* (14), *respectively. The optimal fraction of wealth invested in the risky asset is then given by*

$$\begin{aligned}\widehat{\pi}_t^* = \frac{1}{\sigma\widehat{X}_t^*}\Bigg[&\sum_{i=1}^n\frac{1}{\gamma_i}\left(\frac{\hat{\lambda}}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}}\eta k_i(t)\left(1-\Phi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\right)\\ &+\sum_{i=1}^n\left(\frac{\hat{\lambda}}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}}k_i(t)\varphi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\frac{1}{\sqrt{T-t}}\\ &-(1-\bar{\epsilon})Ge^{-r(T-t)}\varphi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\right)\frac{1}{\sqrt{T-t}}\\ &+\sum_{i=1}^n\frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)\frac{1}{\gamma_i}\left(\lambda_i\xi_t\right)^{-\frac{1}{\gamma_i}}\eta k_i(t)\Phi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\\ &-\sum_{i=1}^n\frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)\left(\lambda_i\xi_t\right)^{-\frac{1}{\gamma_i}}k_i(t)\varphi\left(d(t,\xi_t,\hat{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\frac{1}{\sqrt{T-t}}\Bigg].\end{aligned} \tag{30}$$

Proof: See Appendix A.4.

Again, the optimal wealth (29) and the investment strategy (30) are expressed in terms of ξ_t. Similar to the deterministic guarantee case, we can easily come up with an expression of these quantities in terms of the stock price S_t by using equation (7).

4.3 Sharing rules

Similar to Section 3, we assume that it is the primal goal of the fund manager to meet individual guarantees. Thus, a linear sharing rule is again not applicable to the total terminal wealth and we need to rely on a state-dependent sharing rule. We suggest that the total terminal wealth $\widehat{X}_T^*$ be redistributed to the participants in the following way:

- Let us denote the payoff which investor i obtains from the collective fund by $\widehat{X}_T^{(i)}$. If the guarantee is met, that is, $\widehat{X}_T^* = G_T^{\text{Fle}}(\xi_T)$, participant i obtains her individual flexible guarantee, that is, $\widehat{X}_T^{(i)} = (1-\bar{\epsilon})p_i x_i e^{g_i T} + \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)X_T^{(i,*)}$.
- If the collective guarantee is exceeded, each investor receives the above individual guarantee plus the fraction of terminal wealth exceeding the total guarantee $G_T^{\text{Fle}}(\xi_T)$ is shared proportionally among the participants. Each investor receives a fixed proportion $\alpha_i \in (0,1)$ of this surplus, where we assume again that $\sum_{i=1}^n \alpha_i = 1$. So the payoff obtained by investor i is, in this case, given by $\widehat{X}_T^{(i)} = (1-\bar{\epsilon})p_i x_i e^{g_i T} + \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)X_T^{(i,*)} + \alpha_i\left(\widehat{X}_T^* - G_T^{\text{Fle}}(\xi_T)\right)$.

In total, the payoff can be written as

$$\begin{aligned}\widehat{X}_T^{(i)} &= (1-\bar{\epsilon})p_i x_i e^{g_i T} + \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)X_T^{(i,*)} + \alpha_i\left(\widehat{X}_T^* - G_T^{\text{Fle}}(\xi_T)\right). \qquad (31)\\ &= (1-\bar{\epsilon})p_i x_i e^{g_i T} + \frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}(1-p_i)X_T^{(i,*)} + \alpha_i\left(I_B(\hat{\lambda}\xi_T) - G_T^{\text{Fle}}(\xi_T)\right)^+.\end{aligned}$$

Investor i's payment consists of the desired flexible guarantee she requires and a fraction of surplus participation which corresponds to an exchange option, that is, the option to exchange the fund value for the flexible guarantee.

Similar to the deterministic guarantee, we choose the vector $(\alpha_1,\ldots,\alpha_n)$ in such a way that the pseudo financial fairness criterion as defined in Section 3 is fulfilled, that is,

$$\begin{aligned}(1-\bar{\epsilon})x_i &= \mathbb{E}\left[\xi_T \widehat{X}_T^{(i)}\right]\\ &= (1-\bar{\epsilon})p_i x_i e^{(g_i-r)T} + (1-\bar{\epsilon})(1-p_i)x_i + \alpha_i \mathbb{E}\left[\xi_T\left(\widehat{X}_T^* - G_T^{\text{Fle}}(\xi_T)\right)\right]. \qquad (32)\end{aligned}$$

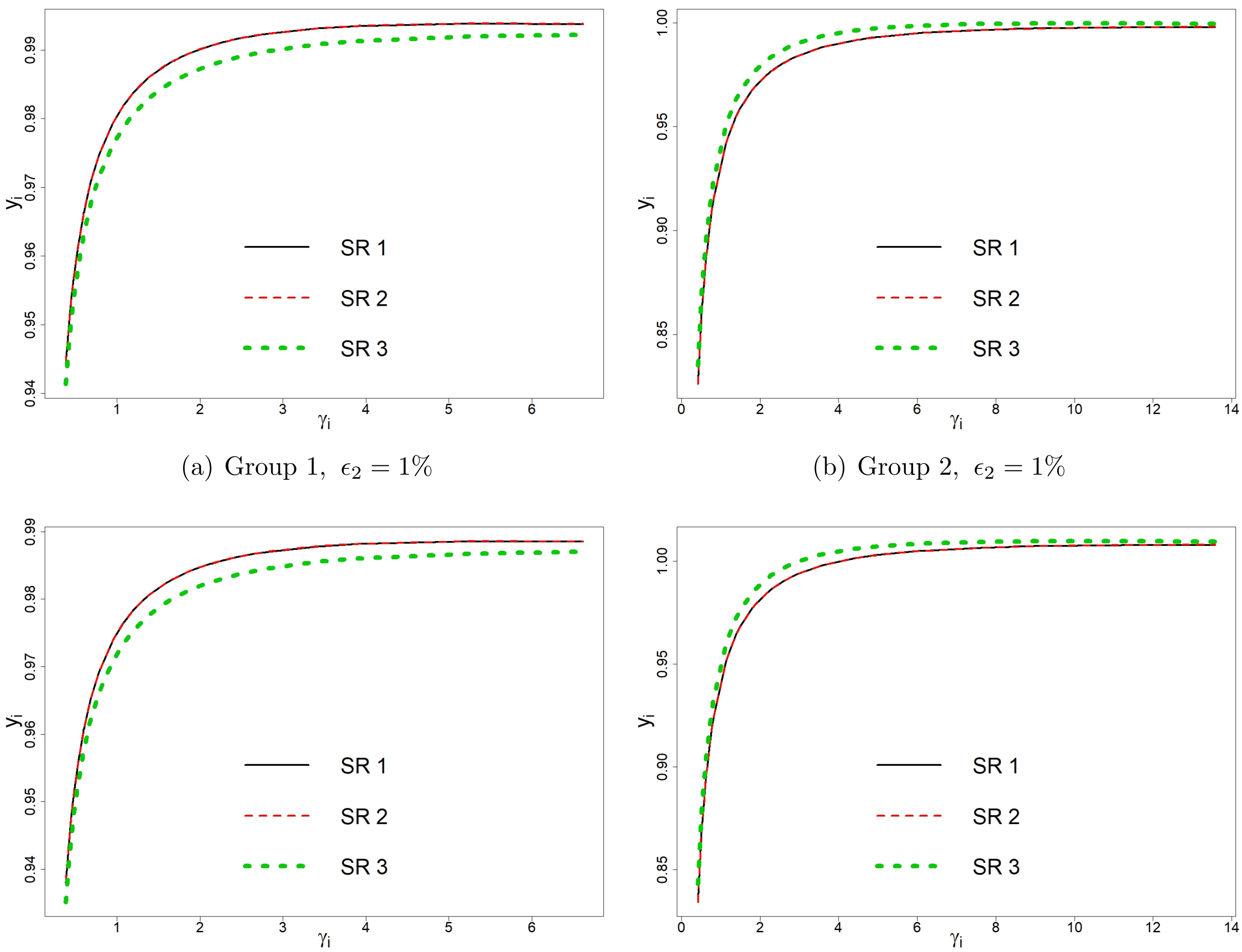
We can now rewrite (32) as

$$\alpha_i = \frac{x_i - p_i x_i e^{(g_i-r)T} - (1-p_i)x_i}{x - Ge^{-rT} - \sum_{i=1}^n (1-p_i)x_i}. \qquad (33)$$

Note that this is a straightforward generalization of the result for the deterministic guarantee given in (18): In the numerator we subtract the present value of the individual guarantee from the individual initial wealth, in the denominator we subtract the present value of the total guarantee from the total initial wealth. Therefore, in this section, we refer to (33) as sharing rule 1 (SR 1) and no longer consider (18). Apart from the sharing rule (33), we again consider the sharing rules 2 and 3 (see (19) and (20)).

4.4 Numerical analyses

We will now have a closer look at the effects the flexible guarantee has on the wealth equivalent of the different investors in the pool. We consider the same parameter setup and the same pool of investors as in Section 3.5. We only slightly change the guarantee design in Group 1: We assume that $p_i = 0.5$ for all $i = 1, \dots, n_1$ and, similar to Section 3.5, that the guaranteed interest rates g_i are equal to 0. The numerical results are given in Figure 4.



(a) Group 1, $\epsilon_2 = 1\%$ (b) Group 2, $\epsilon_2 = 1\%$

(c) Group 1, $\epsilon_2 = 2.5\%$ (d) Group 2, $\epsilon_2 = 2.5\%$

Figure 4: Comparison of the wealth equivalents of the members of the two groups ordered by RRA. For each group, the sharing rules 1 (33), 2 (19) and 3 (20) are compared.

We observe the following:

- **Group 1 / Panels (a) and (c):** At a first glance, the results for this group look identical to Figure 2. However, the losses are now substantially smaller than for the purely deterministic guarantee, especially for investors with a RRA below 3. The second sharing rule performs negligibly better than the first one and the third performs, again, worst, for the same reasons as in Figure 2.

- **Group 2 / Panels (b) and (d)**: The results for Group 2 look almost identical to Figure 2. In Panel (d), under SR 1 and 2, investors with a RRA above 4.4 (16 investors) benefit from the collective investment. Under SR 3, investors with a RRA above 3.3 (19 investors) benefit from the collective investment. On the other hand, in Panel (b), all the investors in Group 2 are worse off than in the benchmark case since y_i is below 1 for all investors under all three sharing rules. In this sense, the flexible guarantee scheme improves the fairness between the two groups because there are now less members in Group 2 who benefit at the cost of Group 1 than in the deterministic guarantee setting. Note that investors with a RRA below 3 in Group 2 still suffer drastic losses due to the deterministic guarantee. Regarding the sharing rules, the order of attractiveness is exactly opposite to Group 1.

In Figure 5, the investment strategy for the flexible guarantee case is plotted and compared to the case with no guarantees. Regardless of whether condition (28) is fulfilled or not, we can rely on numerical procedures to determine the investment strategy.[18] We observe that the

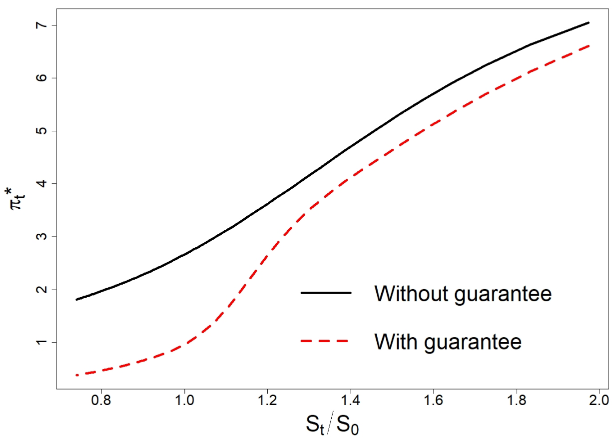


Figure 5: Optimal fraction of wealth invested in the risky asset at time $t = T/2$ depending on the stock return at that time for $\bar{\epsilon} = 0.5\%$. We show the optimal collective investment strategy with and without guarantee.

investment strategy with guarantee is now more risky than in Figure 3. It is, of course, still

[18]If the condition (28) is not fulfilled, we can rely on Monte Carlo estimation procedures to determine one path of the wealth process and then compute the derivative of the wealth for each point in time numerically. Then we combine Itô's formula with (3) to obtain a corresponding realization of the fraction invested in the risky asset at each point in time. A detailed pseudo code is provided in Appendix A.5.

dominated by the investment strategy with no guarantees. As in the case with a purely deterministic guarantee, the investment strategy does not seem to depend largely on the average fee $\bar{\epsilon}$.

In total, we observe that the new, state-dependent guarantee is much more flexible than a deterministic guarantee. It manages to take into account each investor's risk preference and allows for each investor to receive almost what would be optimal for this investor. Especially less risk-averse investors are able to obtain (almost) their unrestricted optimum. Hence, by allowing the plan members to choose a flexible guarantee, we are able to better incorporate their heterogeneous risk preferences in the collective investment problem. An additional advantage of the flexible guarantee is (in our parameter setup) that SR 1 and 2 deliver nearly identical results, i.e. the fund manager can rely on the rather simple SR 2 which is very easy to communicate. This framework could provide a solution to the issue described in Alserda et al. (2019) that strongly heterogeneous risk preferences of individual plan members are not taken into account in the investment strategies of the collective pension fund, which potentially lowers individuals' utility levels drastically.

5 Practical relevance

Hybrid pension schemes are of increasing importance in today's world and DC and/or hybrid plans will be the future trend of occupational plans. As our results show, fixed guarantees in collectively administered pension plans induce utility losses for all individual investors within our numerical examples. Our results in the flexible guarantee framework show that even small reductions in the deterministic guarantee can already reduce these losses in utility substantially, especially for investors with a low risk aversion. Reducing the fixed guarantees and (partially) replacing them by a state-dependent component could thus help reduce utility losses for pension beneficiaries. For the flexible guarantee, for example, a market index could be used.

It is well believed that the shift towards DC schemes is beneficial to the employers, as they do not have to set up high reserves to ensure the pension payments. This is especially the case in the current low interest rate environment, as more capital is required to achieve the same investment goal with lower interest rates. In contrast, it is also well believed that pure DC schemes or releasing the guarantees fully will deteriorate the benefits of the pension beneficiaries. Our results show that moving away from the real guarantees does good to the beneficiaries as well. If employees are reluctant to give up full guarantees, more flexible guarantees which are state-dependent shall lead to a higher utility level than fixed guarantees. With such guarantees, fund delegation can still occur without drastic losses in utility. Therefore, we could imagine that flexible guarantees work better on a psychological level to employees than releasing the guarantees completely. Our results show that adding some flexibility in the design of the guarantee can put forward the reform in the occupational plans. It can lead to a better risk-sharing between the employees and the employers.

The shift towards DC schemes and hybrid schemes like DC with minimum guarantees enhances the importance of investment strategies. As the optimal collective investment strategies do differ from the individual optimal investment strategies, the fund manager needs to carefully consider the investment strategies. It has been pointed out in Alserda et al. (2019) and Frijns (2010) that collectively organized pension funds often cannot reflect each individual's risk preferences and our findings support these observations on a quantitative level.

If the guarantee and wealth levels in the collectively administered pension fund differ, the sharing rule applied by the fund manager should carefully reflect each investor's initial contribution and guarantee level, otherwise unfair distributions of the terminal wealth can result. This implication can be drawn from our numerical analysis with different sharing rules. which has shown that the impact of sharing rules can be substantial. In this sense, when fixing a sharing rule, a compromise between the feature of being simple and interpretable and financial fairness shall be accomplished.

Regarding the fairness, not only the sharing rule matters. The design of the management fees and the financial expertise along with the willingness to pay fees of the plan members also play an important role. The current setting with two groups that we consider is rather simple and in real-world situations, there might be (by far) more than two groups in the collective. Based on the findings of this article, it would, in this case, be likely that the group with the lowest financial expertise benefits most and the group with the highest financial expertise suffers most from the collective investment if the willingness to pay fees differs largely. In reality, we could also think of the individual plan members filling out forms prior to the joint investment to ensure a clear communication between the fund manager and the individual investors.

Concerning the individual beneficiaries, depending on their own risk preferences and other features, our analysis lastly suggests that pension plans could be designed in a more individualized way. While it is typically the case that the pension fund prescribes the same constant guarantee level to all plan members (using e.g. a guaranteed interest rate), we might also think of pension plans where all individuals demand their own (fixed of flexible) guarantee level. By setting up the collective investment strategy and the sharing rule accordingly, each plan members' guarantee shall be satisfied, given that it fulfills the corresponding budget constraint.

6 Optimal collective investment under stochastic interest rates

As typically life and pension insurance contracts are long-term contracts, incorporating interest rate fluctuation is of utmost importance and also more realistic. In this section, we will discuss how our setting and results can be adjusted when a stochastic interest rate model is applied. For simplicity, we disregard management fees in this section, i.e. we assume that $\epsilon_1 = \epsilon_2 = \bar{\epsilon} = 0$. Throughout this section, they could be incorporated in the exact same way as in the previous sections as a constant proportion of the initial investment charged by each individual.

Consider a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ generated by two independent Brownian motions W^S and W^r. For the instantaneous short interest rate r, we consider a Vasicek model (Vasicek (1977)). That is, we assume that it is given by an Ornstein-Uhlenbeck process

$$\mathrm{d}r_t = a(b - r_t)\mathrm{d}t + \sigma_r \mathrm{d}W_t^r, \tag{34}$$

where the rate of mean reversion a, the mean level b and the volatility $\sigma_r > 0$ are constant. We assume furthermore that the market price of risk associated with r is given by a constant $\lambda_r > 0$. The first risky asset is a zero-coupon bond with maturity T whose time t risk-neutral price is denoted by $P(t,T)$. It is shown e.g. in Hull and White (1990) and Hainaut (2009) that under $\mathbf{P}$ the dynamics of $P(t,T)$ are given by

$$\frac{\mathrm{d}P(t,T)}{P(t,T)} = r_t \mathrm{d}t - \sigma_r B(t,T)(\mathrm{d}W_t^r + \lambda_r \mathrm{d}t), \tag{35}$$

where $B(t,H) := a^{-1}(1 - e^{-a(H-t)})$. The second risky asset (a stock) is modeled by a geometric Brownian motion which is correlated with the instantaneous rate:

$$\mathrm{d}S_t = r_t S_t \mathrm{d}t + \sigma_S S_t(\mathrm{d}W_t^S + \lambda_S \mathrm{d}t) + \sigma_{Sr} S_t(\mathrm{d}W_t^r + \lambda_r \mathrm{d}t), \quad S_0 > 0, \tag{36}$$

where σ_S and σ_{Sr} are positive constants.[19] Given the couple of market prices of risk $\lambda :=$

[19]Note that the stock dynamics can be rewritten as

$$\mathrm{d}S_t = S_t(r_t + \nu_S)\mathrm{d}t + S_t\sigma(\sqrt{1-\rho^2}\mathrm{d}W_t^S + \rho \mathrm{d}W_t^r),$$

where $\sigma^2 := \sigma_S^2 + \sigma_{Sr}^2$ is the stock volatility, $\rho := \sigma_{Sr}/\sqrt{\sigma_S^2 + \sigma_{Sr}^2}$ is the correlation coefficient between the stock and the interest rate and $\nu_S = \sigma_S\lambda_S + \sigma_{Sr}\lambda_r$.

(λ_r, λ_S) , the pricing kernel is uniquely defined by

$$\xi_t := \exp\left\{ -\int_0^t r_s \mathrm{d}s - \frac{1}{2}\int_0^t (\lambda_r^2 + \lambda_S^2)\mathrm{d}s - \int_0^t (\lambda_r \mathrm{d}W_s^r + \lambda_S \mathrm{d}W_s^S)\right\}. \tag{37}$$

Now let (π_t^S, π_t^P) be the vector of proportions of wealth invested in the stock and the zero-coupon bond, respectively. The wealth process starting with $X_0 > 0$ can be written by

$$\begin{aligned} \mathrm{d}X_t &= (X_t - \pi_t^S X_t - \pi_t^P X_t) r_t \mathrm{d}t + \frac{\pi_t^S X_t}{S_t}\mathrm{d}S_t + \frac{\pi_t^P X_t}{P(t,T)}\mathrm{d}P(t,T) \\ &= X_t\Big((r_t + \nu_S \pi_t^S + \nu_P \pi_t^P)\mathrm{d}t + \sigma_S \pi_t^S \mathrm{d}W_t^S + (\sigma_{Sr}\pi_t^S - \sigma_r \pi_t^P B(t,T))\mathrm{d}W_t^r \Big), \end{aligned} \tag{38}$$

where $\nu_S = \sigma_S \lambda_S + \sigma_{Sr}\lambda_r$ and $\nu_P := -\sigma_r B(t,T)\lambda_r$ are the risk premiums of the stock and the zero-coupon bond, respectively. Given the market price of risk, the market is complete and the individual optimization problem (5) can be solved as in Section 2. The solution is provided in Theorem 6.1 where we omit the i index for simplicity.

Theorem 6.1. *In a Vasicek stochastic interest rate environment, the optimal investment strategy of the individual problem is given by*

$$\pi_t^S = \frac{\lambda_S}{\gamma \sigma_S}; \quad \pi_t^P = \frac{1}{B(t,T)}\left(\frac{\sigma_{Sr}}{\sigma_S \sigma_r}\frac{\lambda_S}{\gamma} - \frac{\lambda_r}{\sigma_r \gamma} + \frac{\gamma - 1}{\gamma}B(t,T)\right). \tag{39}$$

Proof. Note first that given the market price of risk, the market is complete. The optimal terminal wealth can be expressed by $X_T^* = (y\xi_T)^{-1/\gamma}$, where y is the Lagrangian multiplier defined by the budget constraint $\mathbb{E}[\xi_T X_T^*] = \mathbb{E}[\xi_T (y\xi_T)^{-1/\gamma}] = x$. By Lemma 6.4 we obtain

$$x = \mathbb{E}[\xi_T (y\xi_T)^{-1/\gamma}] = (y)^{-1/\gamma}\mathbb{E}[(\xi_T)^{\widetilde{\gamma}}] = (y)^{-1/\gamma}\exp\{\widetilde{\gamma}\mu_0^\xi + \widetilde{\gamma}^2(\sigma_0^\xi)^2/2\}, \tag{40}$$

where $\widetilde{\gamma} := (\gamma - 1)/\gamma$ and μ_0^ξ and $(\sigma_0^\xi)^2$ are defined in Lemma 6.4. Hence, $y = x^{-\gamma}\exp\{(\gamma - 1)\mu_0^\xi \gamma \widetilde{\gamma}^2 (\sigma_0^\xi)^2/2\}$. Analogously, the optimal wealth is computed by

$$X_t^* = \mathbb{E}[\xi_T \xi_t^{-1} X_T^* | \mathcal{F}_t] = \xi_t^{-1/\gamma}(y)^{-1/\gamma}\mathbb{E}\left[\left(\frac{\xi_T}{\xi_t}\right)^{\widetilde{\gamma}} \Big| \mathcal{F}_t \right] = \xi_t^{-1/\gamma}(y)^{-1/\gamma}\exp\{\widetilde{\gamma}\mu_t^\xi + \widetilde{\gamma}^2(\sigma_t^\xi)^2/2\}. \tag{41}$$

Note that the optimal wealth at time t can be expressed as $X_t^* = F(t, \xi_t, r_t)$ for a deterministic function F, for which, by applying the multi-dimensional Itô-formula, we obtain

$$\mathrm{d}X_t^* = F_t \mathrm{d}t + F_\xi \mathrm{d}\xi_t + F_r \mathrm{d}r_t + \frac{1}{2}\Big(F_{\xi\xi}\mathrm{d} < \xi, \xi >_t + 2F_{\xi r}\mathrm{d} < \xi, r >_t + F_{rr}\mathrm{d} < r, r >_t \Big).$$

Note that $F_\xi = (-1/\gamma)X_t^*/\xi_t$, $F_r = -X_t^*\widetilde{\gamma}B(t,T)$ and $\mathrm{d}\xi_t := -\xi_t(r_t \mathrm{d}t + \lambda_r \mathrm{d}W_t^r + \lambda_S \mathrm{d}W_t^S)$.

Therefore we can represent

$$\mathrm{d}X_t^* = [\quad]\,\mathrm{d}t + (1/\gamma)X_t^*(\lambda_S \mathrm{d}W_t^S + \lambda_r \mathrm{d}W_t^r) + \widetilde{\gamma} X_t^* \sigma_r B(t,T)\mathrm{d}W_t^r.$$

Identifying the last dynamics with (38), we obtain (39). □

Analogously we obtain the following optimal investment strategy for the collective problem.

Theorem 6.2. *In a Vasicek stochastic interest rate environment, the optimal investment strategy of the collective problem with no guarantee constraint is given by*

$$\pi_t^S = \sum_{i=1}^n \frac{\lambda_S}{\gamma_i \sigma_S}; \quad \pi_t^P = \frac{1}{B(t,T)} \sum_{i=1}^n \left(\frac{\sigma_{Sr}}{\sigma_S \sigma_r} \cdot \frac{\lambda_S}{\gamma_i} - \frac{\lambda_r}{\sigma_r \gamma_i} + \frac{\gamma_i - 1}{\gamma_i} B(t,T) \right) \frac{X_t^{i^*}}{X_t^*}, \tag{42}$$

where $X_t^* := \sum_{i=1}^n X_t^{i^*}$ *with*

$$X_t^{i^*} = \xi_t^{-1/\gamma} (\widetilde{y}/\beta_i)^{-1/\gamma_i} \exp\{\widetilde{\gamma}_i \mu_t^\xi + \widetilde{\gamma}_i^2 (\sigma_t^\xi)^2/2\}, \tag{43}$$

$\widetilde{\gamma}_i := (\gamma_i - 1)/\gamma_i$ *, and* $\widetilde{y}$ *is the Lagrangian multiplier which satisfies the budget constraint*

$$\mathbb{E}\left[\xi_T \sum_{i=1}^n \left(\frac{\widetilde{y}}{\beta_i}\xi_T\right)^{-\frac{1}{\gamma_i}}\right] = \sum_{i=1}^n (\widetilde{y}/\beta_i)^{-1/\gamma_i} \exp\{\widetilde{\gamma}_i \mu_0^\xi + \widetilde{\gamma}_i^2 (\sigma_0^\xi)^2/2\} = x = \sum_{i=1}^n x_i.$$

We now turn our attention to the optimization problem (10) under a fixed guarantee G_T^{det} whose optimal terminal wealth is given by

$$X_T^{det^*} = \max\left\{I_B(\widetilde{y}\xi_T), G_T^{\text{det}}\right\} = I_B(\widetilde{y}\xi_T)\,\mathbb{1}_{\{\xi_T < \bar{\xi}\}} + G_T^{\text{det}}\mathbb{1}_{\{\xi_T \geq \bar{\xi}\}},$$

where $\bar{\xi}$ is the constant defined by $G_T^{\text{det}} = I_B\left(\widetilde{y}\bar{\xi}\right)$, see Lemma 3.1.

Theorem 6.3. *Assume* $G_T^{det} \leq x \exp\left(-\mu_0^\xi - \frac{1}{2}\left(\sigma_0^\xi\right)^2\right)$ *, where* μ_0^ξ *and* σ_0^ξ *are given in Lemma 6.4. Then, the optimal wealth at time* $t \in [0,T)$ *is given by*

$$\begin{aligned} X_t^{det^*} = &\sum_{i=1}^n \left(\frac{\widetilde{y}}{\beta_i}\xi_t\right)^{-1/\gamma_i} \exp\{\widetilde{\gamma}_i \mu_t^\xi + \widetilde{\gamma}_i^2 (\sigma_t^\xi)^2/2\} \left(1 - \Phi\left(m^\xi(t,\xi_t,\bar{\xi}) + \widetilde{\gamma}_i \sigma_t^\xi\right)\right) \\ &+ G_T^{det} \exp\{\mu_t^\xi + (\sigma_t^\xi)^2/2\} \Phi\left(m^\xi(t,\xi_t,\bar{\xi}) + \sigma_t^\xi\right), \end{aligned} \tag{44}$$

where $\widetilde{\gamma}_i := (\gamma_i - 1)/\gamma_i$ *and* $m^\xi(t,\xi_t,\bar{\xi}) := \frac{\ln \xi_t - \ln \bar{\xi} + \mu_t^\xi}{\sigma_t^\xi}$ *. The optimal fractions of wealth invested* *in the risky asset can then be obtained as*

$$\begin{aligned}\pi_t^{S^*} =&\frac{\lambda_S}{\sigma_S X_t^{det^*}}\Bigg(\sum_{i=1}^{n}\frac{1}{\gamma_i}\left(\frac{\widetilde{y}}{\beta_i}\xi_t\right)^{-1/\gamma_i}\exp\{\widetilde{\gamma}_i\mu_t^\xi+\widetilde{\gamma}_i{}^2(\sigma_t^\xi)^2/2\}\Big(1-\Phi\left(m^\xi(t,\xi_t,\bar{\xi})+\widetilde{\gamma}_i\sigma_t^\xi\right)\\&-\frac{1}{\sigma_t^\xi}\sum_{i=1}^{n}\left(\frac{\widetilde{y}}{\beta_i}\xi_t\right)^{-1/\gamma_i}\exp\{\widetilde{\gamma}_i\mu_t^\xi+\widetilde{\gamma}_i{}^2(\sigma_t^\xi)^2/2\}\varphi\left(m^\xi(t,\xi_t,\bar{\xi})+\widetilde{\gamma}_i\sigma_t^\xi\right)\\&+\frac{1}{\sigma_t^\xi}G_T^{det}\exp\{\mu_t^\xi+(\sigma_t^\xi)^2/2\}\varphi(m^\xi(t,\xi_t,\bar{\xi})+\sigma_t^\xi)\Bigg)\end{aligned}\tag{45}$$

and

$$\begin{aligned}\pi_t^{P^*} =&\frac{\pi_t^{S^*}}{B(t,T)}\left(\frac{\sigma_{Sr}}{\sigma_r}-\frac{\sigma_S}{\sigma_r}\frac{\lambda_r}{\lambda_S}\right)\\&+\frac{B(t,T)}{X_t^{det^*}B(t,T)}\Bigg(\sum_{i=1}^{n}\widetilde{\gamma}_i\left(\frac{\widetilde{y}}{\beta_i}\xi_t\right)^{-1/\gamma_i}\exp\{\widetilde{\gamma}_i\mu_t^\xi+\widetilde{\gamma}_i{}^2(\sigma_t^\xi)^2/2\}(1-\Phi(m^\xi(t,\xi_t,\bar{\xi})))\\&-\frac{1}{\sigma_t^\xi}\sum_{i=1}^{n}\left(\frac{\widetilde{y}}{\beta_i}\xi_t\right)^{-1/\gamma_i}\exp\{\widetilde{\gamma}_i\mu_t^\xi+\widetilde{\gamma}_i{}^2(\sigma_t^\xi)^2/2\}\varphi(m^\xi(t,\xi_t,\bar{\xi}))\\&+G_T^{det}\frac{1}{\sigma_t^\xi}\exp\{\mu_t^\xi+(\sigma_t^\xi)^2/2\}\left(\Phi(m^\xi(t,\xi_t,\bar{\xi})+\sigma_t^\xi)+\varphi(m^\xi(t,\xi_t,\bar{\xi})+\sigma_t^\xi)\right)\Bigg).\end{aligned}\tag{46}$$

Proof. To ensure the feasibility of the optimization problem, we assume $\mathbb{E}\left[\xi_T G_T^{\text{det}}\right] \leq x$. This can be rearranged to $G_T^{\text{det}} \leq x/\mathbb{E}\left[\xi_T\right] = x\exp\left(-\mu_0^\xi-\frac{1}{2}\left(\sigma_0^\xi\right)^2\right)$, where we have used Lemma 6.4.The optimal wealth at time $t\in[0,T)$ is given by the martingale property $X_t^{det^*} = \mathbb{E}[(\xi_T/\xi_t)X_T^{det^*}|\mathcal{F}_t]$. Using Lemma 6.4 we can show directly that for any $q\in\mathbb{R}$,

$$\mathbb{E}[(\xi_T/\xi_t)^q\mathbb{1}_{\{\xi_T<\bar{\xi}\}}|\mathcal{F}_t] = \exp\{q\mu_t^\xi+q^2(\sigma_t^\xi)^2/2\}\left(1-\Phi\left(m^\xi(t,\xi_t,\bar{\xi})+q\sigma_t^\xi\right)\right),\tag{47}$$

and (44) follows directly from (47) by taking $q=\widetilde{\gamma}_i$, $i=1,\cdots n$ and $q=1$. To compute the optimal strategy, it remains to look at the dynamics of $X_t^{det^*}$ using the two-dimensional Itô-formula. Identifying the resulting dynamics of $X_t^{det^*}$ with (38) we obtain the optimal fractions $(\pi_t^{S^*},\pi_t^{P^*})$ given by (45) and (46). □

Let us remark that $\bar{\xi}\nearrow+\infty$ when $G_T^{\text{det}}\searrow 0$, which implies that $m^\xi(t,\xi_t,\bar{\xi})\to-\infty$ a.s. and the result in Theorem 6.2 is recovered.

Lemma 6.4. *Under* $\mathbf{P}$, $\frac{\xi_T}{\xi_t}|_{\mathcal{F}_t}$ *has lognormal distribution with mean* μ_t^ξ *and variance* $(\sigma_t^\xi)^2$

given by

$$\mu_t^\xi = -\left(\frac{1}{2}(\lambda_r^2 + \lambda_S^2) + b\right)(T-t) - (r_t - b)B(t,T) \tag{48}$$

$$(\sigma_t^\xi)^2 = \int_t^T (\sigma_r B(u,T) + \lambda_r)^2 \mathrm{d}u + \lambda_S^2(T-t). \tag{49}$$

Proof. It can be derived by using the dynamics of r_t and Fubini's theorem. □

7 Conclusion

In our paper, we study the effects of differently structured guarantees, sharing rules and management fees imposed on a collective of investors with heterogeneous risk preferences being tied together in their investment decision. The investors are divided into two groups, where Group 1 has access to the financial market and is thus ready to pay only a low fee for fund delegation and Group 2 has no access to the financial market and is willing to pay a higher fee. The analyses conducted in this paper are of high relevance for occupational pension schemes as the question how pension schemes should be designed in today's world is currently one of the most disputed ones. In both the deterministic and the flexible guarantee framework, each investor may demand her own individual guarantee. By adding these guarantees, the fund manager responsible for the joint investment can make sure that each individual guarantee is met by choosing a proper sharing rule.
Within our numerical examples, we have seen that deterministic guarantees are not beneficial to the majority of investors in the collective and that only a minority of investors in the collective might benefit from the lower average fee charged by the fund manager. While the obvious solution to this problem is to remove the guarantee completely, our analysis reveals that a proper design of the guarantee might also be a solution. We have considered a flexible guarantee consisting of a deterministic component and a fraction of the (state-dependent) individual optimal terminal wealth. By designing the guarantee like this, the fund manager is able to repay to the (relatively) less risk-averse investors (almost) their optimal terminal wealth, while the more risk-averse investors are still able to demand high guarantees. As a consequence, the utility levels of the investors in the pool, especially the ones with relatively low risk aversion, can be increased significantly. It seems that providing the individual investors a flexible guarantee allows the fund manager to better identify the individuals' risk appetite, which might help resolve the issue of not being able to incorporate the strongly heterogeneous risk preferences of individual plan members in a collective investment problem described in Alserda et al. (2019). Regarding the sharing rule, we find that it should carefully take into account the investors' initial contributions and guarantee levels, otherwise the terminal wealth

is likely to be distributed in an unfair way.

An interesting question might now be how our results change if no CRRA, but a different type of utility function would be used for the investors in the collective. In particular, the use of modified power utility functions, which would result in a hyperbolic absolute risk aversion (HARA), would lead to a completely new situation where the guarantee constraint would be taken into account directly in the utility function. To make the setting more practical, we can also include mortality risk into the individual benefits. As mortality risk cannot be completely hedged by only trading the financial market, the new setting will be no longer a complete market. Optimal stochastic control using Hamilton-Jacobi-Bellman equations can be applied to solve the resulting optimization problem. Another possibility is to separate the mortality and financial risks in individual payoffs and consider an equivalent non-concave optimization problem in a complete market. Non-concavity can be treated by using concavification techniques, see e.g. Chen et al. (2019) and Nguyen and Stadje (2018). We leave these questions for future research.

References

Alserda, G. A., Dellaert, B. G., Swinkels, L., and van der Lecq, F. S. (2019). Individual pension risk preference elicitation and collective asset allocation with heterogeneity. *Journal of Banking & Finance*, 101:206–225.

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

Atmaz, A. and Basak, S. (2018). Belief dispersion in the stock market. *The Journal of Finance*, 73(3):1225–1279.

Azar, S. A. (2010). Bounds to the coefficient of relative risk aversion. *Banking and Finance Letters*, 2(4):391–398.

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, 112(2):537–579.

Basak, S. (1995). A general equilibrium model of portfolio insurance. *The Review of Financial Studies*, 8(4):1059–1090.

Beetsma, R. M., Romp, W. E., and Vos, S. J. (2013). Intergenerational risk sharing, pensions, and endogenous labour supply in general equilibrium. *The Scandinavian Journal of Economics*, 115(1):141–154.

Boelaars, I. and Broeders, D. (2019). Fair pensions. *Available at SSRN:* https://ssrn.com/abstract=3374456.

Boyle, P. and Tian, W. (2007). Portfolio management with constraints. *Mathematical Finance*, 17(3):319–343.

Branger, N., Chen, A., Gatzert, N., and Mahayni, A. (2018a). Optimal investments under linear sharing rules. *Working paper. Available from the authors upon request.*

Branger, N., Chen, A., Mahayni, A., and Nguyen, T. (2018b). Optimal collective investment. *Working paper. Available at* https://www.researchgate.net/publication/324910837.

Brennan, M. J. and Schwartz, E. S. (1976). The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics*, 3(3):195–213.

Broeders, D. and Chen, A. (2010). Pension regulation and the market value of pension liabilities: A contingent claims analysis using parisian options. *Journal of Banking & Finance*, 34(6):1201–1214.

Broeders, D., Chen, A., and Koos, B. (2011). A utility-based comparison of pension funds and life insurance companies under regulatory constraints. *Insurance: Mathematics and Economics*, 49:1–10.

Browne, S. (1999). Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark. *Finance & Stochastics*, 3(3):275–294.

Bühlmann, H. and Jewell, W. S. (1979). Optimal risk exchanges. *ASTIN Bulletin: The Journal of the IAA*, 10(3):243–262.

Chen, A., Hieber, P., and Nguyen, T. (2019). Constrained non-concave utility maximization: An application to life insurance contracts with guarantees. *European Journal of Operational Research*, 273(3):1119–1135.

Chen, A., Nguyen, T., and Stadje, M. (2018). Optimal investment under VaR-regulation and minimum insurance. *Insurance: Mathematics & Economics*, 79:194–209.

Chen, K. and Hardy, M. R. (2009). The DB underpin hybrid pension plan: fair valuation and funding. *North American Actuarial Journal*, 13(4):407–424.

Chiappori, P.-A. and Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, 9(6):1021–1052.

Conine, T. E., McDonald, M. B., and Tamarkin, M. (2017). Estimation of relative risk aversion across time. *Applied Economics*, 49(21):2117–2124.

Cox, J. C. and Huang, C.-F. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83.

Davies, J. B. (1981). Uncertain lifetime, consumption, and dissaving in retirement. *Journal of political Economy*, 89(3):561–577.

Deelstra, G., Grasselli, M., and Koehl, P.-F. (2004). Optimal design of the guarantee for defined contribution funds. *Journal of Economic Dynamics and Control*, 28(11):2239–2260.

Dumas, B. (1989). Two-person dynamic equilibrium in the capital market. *The Review of Financial Studies*, 2(2):157–188.

Dybvig, P. H. and Ross, S. A. (1987). Arbitrage. In *The New Palgrave: A Dictionary of Economics* (eds. J. Eatwell, M. Milgate and P. Neumann), pages 100–106. Macmillan, London.

El Karoui, N., Jeanblanc, M., and Lacoste, V. (2005). Optimal portfolio management with American capital guarantee. *Journal of Economic Dynamics and Control*, 29(3):449–468.

Frijns, J. M. (2010). Dutch pension funds: Aging giants suffering and inconsistent risk management. *Rotman International Journal of Pension Management*, 3(2):16–21.

Gabih, A., Sass, J., and Wunderlich, R. (2009). Utility maximization under bounded expected loss. *Stochastic Models*, 25(3):375–407.

Grossman, S. J. and Vila, J.-L. (1989). Portfolio insurance in complete markets: A note. *Journal of Business*, 62(4):473–476.

Grossman, S. J. and Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3):241–276.

Grossman, S. J. and Zhou, Z. (1996). Equilibrium analysis of portfolio insurance. *The Journal of Finance*, 51(4):1379–1403.

Hainaut, D. (2009). Dynamic asset allocation under var constraint with stochastic interest rates. *Annals of Operations Research*, 172(1):97–117.

Hardy, M. R., Saunders, D., and Zhu, X. (2014). Market-consistent valuation and funding of cash balance pensions. *North American Actuarial Journal*, 18(2):294–314.

Huang, C.-f. and Litzenberger, R. (1985). On the necessary condition for linear sharing and separation: a note. *Journal of Financial and Quantitative Analysis*, 20(3):381–384.

Hull, J. and White, A. (1990). Pricing interest-rate-derivative securities. *The review of financial studies*, 3(4):573–592.

Jensen, B. A. and Nielsen, J. A. (2016). How suboptimal are linear sharing rules? *Annals of Finance*, 12(2):221–243.

Jensen, B. A. and Sørensen, C. (2001). Paying for minimum interest rate guarantees: Who should compensate who? *European Financial Management*, 7(2):183–211.

Karatzas, I., Lehoczky, J. P., and Shreve, S. E. (1990). Existence and uniqueness of multi-agent equilibrium in a stochastic, dynamic consumption/investment model. *Mathematics of Operations research*, 15(1):80–128.

Karatzas, I. and Shreve, S. E. (1998). *Methods of Mathematical Finance*, volume 39. Springer, New York.

Kim, H. H., Maurer, R., and Mitchell, O. S. (2016). Time is money: Rational life cycle inertia and the delegation of investment management. *Journal of Financial Economics*, 121(2):427–447.

Korn, R. (2014). *Moderne Finanzmathematik–Theorie und praktische Anwendung: Band 1–Optionsbewertung und Portfolio-Optimierung*. Springer Fachmedien Wiesbaden.

Kravitz (2016). 2016 national cash balance research report. *Technical report*, Kravitz, Inc.

Lucas, D. J. and Zeldes, S. P. (2009). How should public pension plans invest? *American Economic Review*, 99(2):527–32.

Malkiel, B. G. (2013). Asset management fees and the growth of finance. *Journal of Economic Perspectives*, 27(2):97–108.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Nguyen, T. and Stadje, M. (2018). Optimal investment for participating insurance contracts under var-regulation. *arXiv preprint arXiv:1805.09068*.

Orozco-Garcia, C. and Schmeiser, H. (2019). Is fair pricing possible? An analysis of participating life insurance portfolios. *Journal of Risk and Insurance*, 86(2):521–560.

Pazdera, J., Schumacher, J. M., and Werker, B. J. (2016). Cooperative investment in incomplete markets under financial fairness. *Insurance: Mathematics and Economics*, 71:394–406.

Schumacher, J. M. (2018). Linear versus nonlinear allocation rules in risk sharing under financial fairness. *ASTIN Bulletin: The Journal of the IAA*, 48(3):995–1024.

Sharpe, W. F. (2017). *Retirement Income Scenarios.* Available at `http://retirementincomescenarios.blogspot.be/`.

Shaw, K. L. (1996). An empirical analysis of risk aversion and income growth. *Journal of Labor Economics*, 14(4):626–653.

Tepla, L. (2001). Optimal investment with minimum performance constraints. *Journal of Economic Dynamics and Control*, 25(10):1629–1645.

Turner, J. A. (2014). Hybrid pensions: risk sharing arrangements for pension plan sponsors and participants. *Society of Actuaries*.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188.

Weinbaum, D. (2009). Investor heterogeneity, asset pricing and volatility dynamics. *Journal of Economic Dynamics and Control*, 33(7):1379–1397.

Wilson, R. (1968). The theory of syndicates. *Econometrica: Journal of the Econometric Society*, 36(1):119–132.

Xia, J. (2004). Multi-agent investment in incomplete markets. *Finance and Stochastics*, 8(2):241–259.

Zhu, X., Hardy, M., and Saunders, D. (2018). Valuation of an early exercise defined benefit underpin hybrid pension. *Scandinavian Actuarial Journal*, 2018(9):823–844.

A Proofs

A.1 Proof of Lemma 3.1

As $I_i\left(\frac{\lambda\xi_T}{\beta_i}\right)$ is strictly decreasing in ξ_T for all $i = 1, \ldots, n$, $I_B(\lambda\xi_T)$ is also strictly decreasing in ξ_T. Additionally, for all $i = 1, \ldots, n$, we clearly have $I_i\left(\frac{\lambda\xi_T}{\beta_i}\right) \to 0$ as $\xi_T \to \infty$ and $I_i\left(\frac{\lambda\xi_T}{\beta_i}\right) \to \infty$ as $\xi_T \to 0$. Therefore, $I_B(\lambda\xi_T)$ behaves in the exact same way. We can conclude that $I_B(\cdot)$ is bijective on $(0, \infty)$. Therefore, for any positive level of the guarantee G_T^{det} we can find a unique value $\bar{\xi} > 0$ such that

$$G_T^{\text{det}} = I_B\left(\lambda\bar{\xi}\right).$$

Since $I_B(\lambda\xi_T)$ is strictly decreasing in ξ_T, it is clear that $I_B(\lambda\xi_T) > G_T^{\text{det}}$ for all $\xi_T < \bar{\xi}$ and $I_B(\lambda\xi_T) \leq G_T^{\text{det}}$ for all $\xi_T \geq \bar{\xi}$. Hence, we can make the following decomposition:

$$X_T^* = \max\left\{I_B(\lambda\xi_T), G_T^{\text{det}}\right\} = I_B(\lambda\xi_T)\,\mathbb{1}_{\left\{\xi_T<\bar{\xi}\right\}} + G_T^{\text{det}}\,\mathbb{1}_{\left\{\xi_T\geq\bar{\xi}\right\}}.$$

$\square$

A.2 Proof of Proposition 3.2

Note that $\{\xi_t X_t^*\}_{t\in[0,T]}$ is a martingale under P. Assuming $\bar{\xi} > 0$ fulfills (12), we get:

$$
\begin{aligned}
X_t^* &= \mathbb{E}\left[\frac{\xi_T}{\xi_t}X_T^* \,\middle|\, \mathcal{F}_t\right] \\
&= \mathbb{E}\left[\frac{\xi_T}{\xi_t}I_B\left(\lambda\xi_T\right)\mathbb{1}_{\{\xi_T<\bar{\xi}\}} \,\middle|\, \mathcal{F}_t\right] + \mathbb{E}\left[\frac{\xi_T}{\xi_t}G_T^{\text{det}}\mathbb{1}_{\{\xi_T\geq\bar{\xi}\}} \,\middle|\, \mathcal{F}_t\right] \\
&= \underbrace{\mathbb{E}\left[\frac{\xi_T}{\xi_t}I_B\left(\lambda\frac{\xi_T}{\xi_t}\xi_t\right)\mathbb{1}_{\left\{\frac{\xi_T}{\xi_t}<\frac{\bar{\xi}}{\xi_t}\right\}} \,\middle|\, \mathcal{F}_t\right]}_{=(\mathrm{I})} + \underbrace{\mathbb{E}\left[\frac{\xi_T}{\xi_t}G_T^{\text{det}}\mathbb{1}_{\left\{\frac{\xi_T}{\xi_t}\geq\frac{\bar{\xi}}{\xi_t}\right\}} \,\middle|\, \mathcal{F}_t\right]}_{=(\mathrm{II})}.
\end{aligned}
\tag{50}
$$

Defining $d(t,\xi_t,\bar{\xi})$ as in equation (15) and $k_i(t)$ as in (14), we can now compute the conditional expectations (I) and (II) from (50):

$$
\begin{aligned}
(\mathrm{I}) &= \mathbb{E}\left[\sum_{i=1}^n\left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}}\left(\frac{\xi_T}{\xi_t}\right)^{1-\frac{1}{\gamma_i}}\mathbb{1}_{\left\{\frac{\xi_T}{\xi_t}<\frac{\bar{\xi}}{\xi_t}\right\}} \,\middle|\, \mathcal{F}_t\right] \\
&= \sum_{i=1}^n\left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}}\int_{d(t,\xi_t,\bar{\xi})}^{\infty} e^{\left(1-\frac{1}{\gamma_i}\right)\left(-r(T-t)-\frac{1}{2}\eta^2(T-t)-z\eta\sqrt{T-t}\right)}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\
&= \sum_{i=1}^n I_i\left(\frac{\lambda}{\beta_i}\xi_t\right)k_i(t)\left(1-\Phi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\left(1-\frac{1}{\gamma_i}\right)\right)\right).
\end{aligned}
\tag{51}
$$

The second expectation in (50) can be computed in the following way:

$$
\begin{aligned}
(\mathrm{II}) &= G_T^{\text{det}}\int_{-\infty}^{d(t,\xi_t,\bar{\xi})} e^{-r(T-t)-\frac{1}{2}\eta^2(T-t)-\eta\sqrt{T-t}z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\
&= G_T^{\text{det}}e^{-r(T-t)}\Phi\left(d(t,\xi_t,\bar{\xi})+\eta\sqrt{T-t}\right).
\end{aligned}
\tag{52}
$$

Adding (51) and (52), we end up with the wealth process as given in (13). From Itô's lemma and the dynamic wealth process of the fund manager (see equation (3) without the (i)-indexes), we know that

$$
\frac{\partial X_t^*}{\partial W_t}\mathrm{d}W_t = \pi_t\sigma X_t^*\mathrm{d}W_t \quad \Leftrightarrow \quad \pi_t^* = \frac{\partial X_t^*/\partial W_t}{\sigma X_t^*}.
$$

This directly delivers the investment strategy as given in (16). □

A.3 Proof of Lemma 4.2

We need to find out whether the two functions $I_B\left(\hat{\lambda}\xi_T\right)$ and $G_T^{\text{Fle}}(\xi_T)$ have a unique intersection $\hat{\xi} > 0$. We assume that $p_i > 0$ for at least one i. For $\xi_T \to \infty$ we have

$G_T^{\text{Fle}}(\xi_T) \to (1-\bar{\epsilon})G > 0$ and $I_B(\hat{\lambda}\xi_T) \to 0$. For $\xi_T \to 0$, on the other hand, we have $G_T^{\text{Fle}}(\xi_T) \to \infty$ and $I_B(\hat{\lambda}\xi_T) \to \infty$. Therefore, the two functions $G_T^{\text{Fle}}(\xi_T)$ and $I_B(\hat{\lambda}\xi_T)$ need to have at least one intersection. If they did not intersect, we would necessarily have $G_T^{\text{Fle}}(\xi_T) > I_B(\hat{\lambda}\xi_T)$ a.s. because of the limiting behavior analyzed before and since both functions are strictly decreasing in ξ_T. This would, however, result in $\widehat{X}_T^* = G_T^{\text{Fle}}(\xi_T)$ a.s. which would be a contradiction to the budget constraint, since $\mathbb{E}[\xi_T G_T^{\text{Fle}}(\xi_T)] < (1-\bar{\epsilon})x = \mathbb{E}[\xi_T \widehat{X}_T^*]$. The uniqueness of this intersection can be shown by looking at the function

$$\begin{aligned} h(\xi_T) &= I_B\left(\hat{\lambda}\xi_T\right) - G_T^{\text{Fle}}(\xi_T) \\ &= \sum_{i=1}^n \xi_T^{-\frac{1}{\gamma_i}} \left(\left(\frac{\hat{\lambda}}{\beta_i}\right)^{-\frac{1}{\gamma_i}} - (1-p_i)\frac{(1-\bar{\epsilon})}{(1-\epsilon_i)}\lambda_i^{-\frac{1}{\gamma_i}} \right) - (1-\bar{\epsilon})G. \end{aligned}$$

Under assumption (27), $h(\cdot)$ is strictly decreasing which can be seen by taking the first-order derivative. This implies that the intersection $\hat{\xi}$ is uniquely determined. Since $h(\cdot)$ is strictly decreasing, we observe that $I_B\left(\hat{\lambda}\xi_T\right) > G_T^{\text{Fle}}(\xi_T)$ for all $\xi_T < \hat{\xi}$ and $I_B\left(\hat{\lambda}\xi_T\right) \le G_T^{\text{Fle}}(\xi_T)$ for all $\xi_T \ge \hat{\xi}$. Hence, we can make the following decomposition:

$$\widehat{X}_T^* = \max\left\{G_T^{\text{Fle}}(\xi_T), I_B(\hat{\lambda}\xi_T)\right\} = I_B\left(\hat{\lambda}\xi_T\right)\mathbb{1}_{\{\xi_T<\hat{\xi}\}} + G_T^{\text{Fle}}(\xi_T)\mathbb{1}_{\{\xi_T\ge\hat{\xi}\}}.$$

□

A.4 Proof of Proposition 4.3

Having computed $\hat{\xi}$ by solving a fixed-point problem, we can determine the investment strategy explicitly. First we note that $\{\xi_t \widehat{X}_t^*\}_{t\in[0,T]}$ is a martingale under P, so we get:

$$\begin{aligned} \widehat{X}_t^* &= \mathbb{E}\left[\frac{\xi_T}{\xi_t}\widehat{X}_T^* \,\middle|\, \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\frac{\xi_T}{\xi_t} I_B\left(\hat{\lambda}\xi_T\right)\mathbb{1}_{\{\xi_T<\hat{\xi}\}} \,\middle|\, \mathcal{F}_t\right] + \mathbb{E}\left[\frac{\xi_T}{\xi_t} G_T^{\text{Fle}}(\xi_T)\mathbb{1}_{\{\xi_T\ge\hat{\xi}\}} \,\middle|\, \mathcal{F}_t\right] \\ &= \underbrace{\mathbb{E}\left[\frac{\xi_T}{\xi_t} I_B\left(\hat{\lambda}\frac{\xi_T}{\xi_t}\xi_t\right)\mathbb{1}_{\left\{\frac{\xi_T}{\xi_t}<\frac{\hat{\xi}}{\xi_t}\right\}} \,\middle|\, \mathcal{F}_t\right]}_{=(\text{I})} + \underbrace{\mathbb{E}\left[\frac{\xi_T}{\xi_t} G_T^{\text{Fle}}(\xi_T)\mathbb{1}_{\left\{\frac{\xi_T}{\xi_t}\ge\frac{\hat{\xi}}{\xi_t}\right\}} \,\middle|\, \mathcal{F}_t\right]}_{=(\text{II})} \end{aligned} \tag{53}$$

Defining $k_i(t)$ as in (14) and $d(t,\xi_t,\hat{\xi})$ as in (15), we can now compute the conditional expectations (I) and (II) from (53). The first one can be computed in the exact same way as in the case with the deterministic guarantee:

$$\text{(I)} = \sum_{i=1}^{n} I_i \left(\frac{\hat{\lambda}}{\beta_i} \xi_t \right) k_i(t) \left(1 - \Phi \left(d(t, \xi_t, \hat{\xi}) + \eta\sqrt{T-t} \left(1 - \frac{1}{\gamma_i} \right) \right) \right). \tag{54}$$

The second expectation in (53) needs to be split into two parts:

$$\begin{aligned}
\text{(II)} &= (1-\bar{\epsilon})G\,\mathbb{E}\left[\frac{\xi_T}{\xi_t} \mathbb{1}_{\left\{ \frac{\xi_T}{\xi_t} \geq \frac{\hat{\xi}}{\xi_t} \right\}} \,\middle|\, \mathcal{F}_t \right] + \sum_{i=1}^{n} (1-\bar{\epsilon})(1-p_i)\mathbb{E}\left[\frac{\xi_T}{\xi_t} I_i \left(\lambda_i \frac{\xi_T}{\xi_t} \xi_t \right) \mathbb{1}_{\left\{ \frac{\xi_T}{\xi_t} \geq \frac{\hat{\xi}}{\xi_t} \right\}} \,\middle|\, \mathcal{F}_t \right] \\
&= (1-\bar{\epsilon})Ge^{-r(T-t)}\Phi\left(d(t, \xi_t, \hat{\xi}) + \eta\sqrt{T-t} \right) \\
&\quad + \sum_{i=1}^{n} (1-\bar{\epsilon})(1-p_i) I_i \left(\lambda_i \xi_t \right) k_i(t) \Phi \left(d(t, \xi_t, \hat{\xi}) + \eta\sqrt{T-t} \left(1 - \frac{1}{\gamma_i} \right) \right).
\end{aligned} \tag{55}$$

Adding (54) and (55), we end up with the wealth process given in (29). From Itô's lemma and the dynamic wealth process of the fund manager (see equation (3) without the (i)-indexes), we know that

$$\frac{\partial X_t^*}{\partial W_t} \mathrm{d}W_t = \pi_t \sigma X_t^* \mathrm{d}W_t \quad \Leftrightarrow \quad \pi_t^* = \frac{\partial X_t^* / \partial W_t}{\sigma X_t^*}.$$

This directly delivers the investment strategy as given in (30). □

A.5 Pseudo code for the numerical calculation of the optimal investment strategy with flexible guarantees

1. Simulate one path of the Brownian motion $\{W_t\}_{t \in [0,T]}$.
2. Create N realizations of the state price density and the optimal terminal wealth given the path of the Brownian motion, that is, $\xi_{T,i} \mid W_t$ and $\widehat{X}_{T,i}^* \mid W_t$ for all t, $i = 1, \ldots, N$.
3. Then, one realization of $\widehat{X}_t^*(W_t)$ is (approximately) given by

$$\widehat{X}_t^*(W_t) = \xi_t^{-1}(W_t) \frac{1}{N} \sum_{i=1}^{N} \left(\xi_{T,i} \widehat{X}_{T,i}^* \mid W_t \right)$$

   for all t, so we obtain one path of $\widehat{X}_t^*$.
4. From Itô's lemma we know that

$$\pi_t = \frac{\partial \widehat{X}_t^* / \partial W_t}{\sigma \widehat{X}_t^*},$$

so we need to determine the derivative of $\widehat{X}_t^*$ numerically. We compute N realizations of the state price density and the optimal terminal payoff given $W_t \pm h$, that is, $\xi_{T,i} \mid W_t \pm h$ and $\widehat{X}_{T,i}^* \mid W_t \pm h$ for all t, $i = 1, \ldots, N$, where $h > 0$ is an arbitrarily small number.

5. Then we can compute

$$\widehat{X}_t^*(W_t \pm h) = \xi_t^{-1}(W_t \pm h) \frac{1}{N} \sum_{i=1}^{N} \left(\xi_{T,i} \widehat{X}_{T,i}^* \mid W_t \pm h \right)$$

and with these values, using the central difference method, the derivative can, for all t, be determined as

$$\frac{\partial \widehat{X}_t^*}{\partial W_t} = \frac{\widehat{X}_t^*(W_t + h) - \widehat{X}_t^*(W_t - h)}{2h} .$$

5 A collective investment problem in a stochastic volatility environment: The impact of sharing rules

Source:

Chen, A., Nguyen, T., and Rach, M. (2019). A collective investment problem in a stochastic volatility environment: The impact of sharing rules. Submitted to *Annals of Operations Research* (under review). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3424804

A collective investment problem in a stochastic volatility environment: The impact of sharing rules

An Chen*, Thai Nguyen* and Manuel Rach*

* Institute of Insurance Science, Ulm University, Helmholtzstr. 20, 89069 Ulm, Germany.
E–mails: an.chen@uni-ulm.de; thai.nguyen@uni-ulm.de; manuel.rach@uni-ulm.de

Abstract

It is typical in collectively administered pension funds that employees delegate fund managers to invest their contributions. In addition, many pension funds still need to sustain guarantees (prescribed by law) in spite of the current low interest environment. In this paper, we consider an optimal collective investment problem for a pool of investors who (implicitly) demand minimum guarantees by deriving utility from the wealth exceeding their guarantees in two financial market settings, one with a stochastic and one with a constant volatility. We find that individual investors' well-being will not be worsened through the collective investment in both financial markets, as individual optimal solutions are attainable if a financially fair state-dependent sharing rule is applied. When more prevailing sharing rules like linear rules are applied, this holds no longer. Furthermore, the degree of sub-optimality imposed by linear sharing rules is more pronounced in the stochastic volatility market than in the constant volatility market.

Keywords: Collective investment problems, stochastic volatility, portfolio insurance, sharing rules

JEL: G11, G23

1 Introduction

There exist various reasons for fund delegation in today's world, one of the most prominent examples being a collectively administered pension fund. There are two main types of occupational pensions schemes: In a defined benefit (DB) scheme, the sponsoring companies promise their employees a guaranteed pension payment. In a defined contribution (DC) scheme, the benefit at retirement depends on the performance of the investment returns experienced during the plan membership. Consequently, in a DC scheme, the market risk is carried completely by the employees instead of the employers.[1] Very recently, people have started to believe that hybrid pension plans combining the DC and DB plan might meet the requirements of employees and employers even better. A key component of such hybrid schemes is to provide safety by offering a minimum guarantee (which is lower than in pure DB schemes), and, simultaneously, to let employees participate in potential upside scenarios of the markets.[2] In such pensions, fund managers shall take account of the guarantee requirement in their portfolio planning, while simultaneously capturing the members' risk preferences in the investment strategy to provide acceptable bonuses to the members in well-performing markets. Another reason for fund delegation, besides collectively administered pension funds, is explained, for example, in Kim et al. (2016): Many households display investment inertia because handling investments costs time and energy. The authors find that delegation can be beneficial for individual investors.

In this article, we consider an optimal investment problem of a fund manager who invests on behalf of a *collective* of individuals requiring a minimum guaranteed payment in a stochastic volatility framework. In a utility maximization framework, it is common in the literature to assume that individuals implicitly satisfy their guarantee requirements by deriving utility only from the residual wealth exceeding the guarantee,[3] see, for example, Basak (2002), Balder and Mahayni (2010) and Zieling et al. (2014).[4] Each of the individuals in the collective may demand a certain guarantee. We allow individuals with various degrees of risk aversion to choose a different guarantee level. The utility function used by the fund manager is itself

[1] For further details on DB and DC schemes, see also OECD (2018).

[2] An overview over existing hybrid schemes can, for example, be found in Turner (2014). A literature review on dynamic hybrid pension products is provided by Hambardzumyan and Korn (2019).

[3] Note that, in a Black-Scholes setting, the resulting optimal investment strategy for an *individual* investor with such utility preferences is a so-called constant proportion portfolio insurance (CPPI) strategy, a rather popular type of portfolio insurance strategies (see, for example, Black and Jones (1987), Black and Perold (1992), Basak (2002) and, more recently, Temocin et al. (2018), and for the relevance of CPPI strategies in practice see Pain and Rand (2008)). While this type of strategy is optimal for a single individual with risk preferences as described above, we find that this result holds no longer for a collective of individuals who jointly invest their initial wealth. Further details regarding this result and CPPI strategies are provided in Section 3.2 of this article.

[4] While we follow this first approach in this article, a second popular approach would be to impose a minimum guarantee constraint in the utility maximization problem, as done, for example, in Jensen and Sørensen (2001) and Hambardzumyan and Korn (2019).

defined by an optimization problem in such a way that the weighted sum of the individual utility functions is maximized for a given vector of positive weights. Due to the inclusion of guarantee requirements, this is a generalization of a popular utility funciton (with no minimum subsistence level) in the literature (see, for example, Dumas (1989), Karatzas et al. (1990), Xia (2004), Pazdera et al. (2016) and Branger et al. (2018b)). The fund manager then sets up a collective investment strategy such that all these individual guarantees are met. We not only consider the Black-Scholes setting with a constant volatility, but also move beyond normally distributed returns and describe the evolution of the stock with a more general stochastic volatility model in the sense of Heston (1993). A stochastic volatility model is more realistic than a model with constant volatility, for it allows to explain stylized facts often observed in financial markets such as heavy tails, volatility clustering, and the smile of implied volatilities (see Cont and Tankov (2004)). In such a stochastic volatility model, which leads to bigger tail risks, appropriate fund management under portfolio insurance becomes even more important than in a model with normally distributed returns, because the probability of extremal market scenarios increases (see Chen et al. (2018)). In this article, we are particularly interested in finding out the influence of such a more realistic financial market modeling on the expected utility of the individual investors, and comparing it to the constant volatility framework.

We show that, under both constant and stochastic volatility, individual optimal solutions are achievable if a state-dependent sharing rule is applied to the optimal collective terminal wealth and the financial fairness condition in the sense of Bühlmann and Jewell (1979) and Schumacher (2018) is imposed. In other words, individual welfare does not deteriorate in both financial markets if an appropriate state-dependent sharing rule is applied. Under more prevailing sharing rules, like linear ones, this result holds no longer, as linear sharing rules impose a certain suboptimality to the collective (see, for example, Jensen and Nielsen (2016)). Then, either all the individuals in the collective suffer a loss or an unfair distribution of the terminal wealth, where some individuals benefit at the cost of others, results. To assess the losses imposed by linear sharing rules in both financial markets, we compare the state-dependent sharing rule to two linear sharing rules: one satisfying the financial fairness and one not. If the linear sharing rule does not fulfill the fairness condition, some individuals in the collective are better off, but the majority of investors is largely worse off than in the individual optimization problem. When a financially fair linear sharing rule is applied, all the individuals suffer a (relatively) small loss. In this sense, a financially fair linear sharing rule performs better from a fund manager's point of view who wants to consider all the individuals in the collective in a fair way. A comparison between the constant and stochastic volatility framework reveals that the degree of sub-optimality imposed by linear sharing rules is larger under stochastic volatility.

Individuals' utility optimization in incomplete stochastic volatility markets has been considered extensively in the literature (see, for example, Pham (2002), Fleming and Hernández-Hernández

(2003), Chacko and Viceira (2005), Kraft (2005) and Liu (2006)). For common utility functions (for example power utility), the solution is available in closed form by applying a separation technique in the Hamilton-Jacobi-Bellman (HJB) equation resulting from the dynamic programming principle. Unfortunately, such a separation technique seems impossible in our collective utility maximization framework. Consequently, we rely on another way of solving the optimal investment problem: We complete the financial market using derivatives. This approach is well-documented in the literature and applied, for instance, in Liu and Pan (2003), Branger et al. (2008, 2017), Escobar et al. (2018) and Chen et al. (2018). Following this approach, we can determine the optimal terminal wealth levels and the dynamic trading strategies explicitly for our collective utility maximization problem in the stochastic volatility framework using the static martingale approach (see, for example, Cox and Huang (1989)). Solving the collective optimization problem under a constant volatility less complicated, as the constant volatility market is complete without adding derivatives. In this sense, our article contributes to the literature on utility maximization in incomplete stochastic volatility markets by the consideration of a *collective* utility maximization problem.

The remainder of the paper is organized in the following way: Section 2 introduces the utility preferences assumed for the individuals in the collective and, particularly, the collective utility function used for modeling the fund manager's preferences. Section 3 briefly presents the solution to the optimal collective investment problem in a constant volatility framework. Section 4 deals with the collective optimization problem in the Heston model which allows for stochastic volatility. In Section 5, we show that individual optimal solutions can be achieved through the collective investment under financial fairness and a state-dependent sharing rule. In Section 6, we discuss different sharing rules and compare the well-being of the investors in the collective under constant and stochastic volatility. Section 7 concludes the article and is followed by the appendix with one proof.

2 Risk preferences

In this section, we describe the basic assumptions regarding the preferences of the individuals. To model individual preferences, we mainly take account of the fact that the individuals are interested in obtaining a minimum payment and building their utility on the (residual) wealth exceeding the minimal guarantee.

2.1 Individual preferences

We consider a collective of n individuals on a financial market. Each of the investors assigns her initial wealth x_i to the fund manager for investment in financial assets at time 0. We use a special type of HARA utility function of the form $U_{i,G^i}(v) = \frac{\left(v-G^i\right)^{1-\gamma_i}}{1-\gamma_i}$ with $\gamma_i \neq 1$, $\gamma_i > 0$ for $i = 1, \ldots, n$ to model each investor's preferences, where G^i is investor i's subsistence level. This subsistence level will be referred to as a minimum guarantee that investor i is interested in achieving from now on. The corresponding relative risk aversion is given by $\frac{\gamma_i v}{v-G^i}$ which is increasing in G^i and γ_i. In the special case that $G^i = 0$, we obtain constant relative risk aversion (CRRA) utility functions. Note that each investor derives her utility only from the difference between the total terminal wealth and the guarantee. This preference representation for individuals interested in sustaining a minimum guaranteed income is common in the literature, see, for example, Basak (2002), Balder and Mahayni (2010) and Zieling et al. (2014). The resulting inverse marginal utility function is denoted by

$$I_{i,G^i}(\cdot) := (U'_{i,G^i})^{-1}(\cdot) = G^i + (\cdot)^{-\frac{1}{\gamma_i}}.$$

2.2 Collective utility function

From now on, we assume that the n investors delegate a fund manager to collectively invest their total initial wealth $x = \sum_{i=1}^n x_i$ on their behalf. Reasons for fund delegation can be different: For example, in an occupational pension context, it is common that beneficiaries do not administrate their contributions themselves. Instead, contributions are collectively managed by a pension fund manager. Another reason for fund delegation could be professional skills and knowledge of the fund manager, leading individual investors to believe that investment delegation is more beneficial for them than handling investments on their own. Further, Kim et al. (2016) observe that many individuals display investment inertia as managing money costs time and energy and show that delegation is valuable.

We assume that the fund manager does not charge any additional fees, so the total wealth x is completely invested in financial assets. The fund manager's primal goal is to provide individual guarantees, as it is the case, for example, in many occupational pension schemes that are not of the pure DC type.[5] In many other real-life fund delegation situations, fund managers might be more interested in maximizing their own compensations from advising individual investors regarding the suitability of financial products. As in Kim et al. (2016), we assume that the fund manager behaves more on behalf of individual investors, that is, the fund manager's utility function reflects the individuals' utility and, more importantly, the fund manager aims to meet individual guarantees. We denote by G the time-T-value of the collective guarantee which the fund manager needs to meet. Unless stated otherwise, we will always assume that this guarantee is equal to the sum of the individual guarantees, that is, $G = \sum_{i=1}^n G^i$. Ad-

[5] For instance, in all German pension schemes, some sort of guarantee had been prescribed until very recently in 2018 when a new pension scheme was introduced along with the "Betriebsrentenstärkungsgesetz".

ditionally, we want to emphasize that the case with no guarantee is included throughout this article, as all the individual guarantees can always be set equal to zero.

Concerning the collective utility function (which the fund manager uses), we are inspired, for example, by Dumas (1989), Karatzas et al. (1990), Xia (2004), Pazdera et al. (2016) and Branger et al. (2018b). We assume that the fund manager uses the following (collective) utility function which depends on the collective and individual guarantees:

$$U_{B,G} : (G, \infty) \to (0, \infty), \quad v \mapsto U_{B,G}(v) = \max_{\substack{v_1 \geq G^1, \ldots, v_n \geq G^n \\ v = \sum_{i=1}^n v_i}} \sum_{i=1}^n \beta_i U_{i,G^i}(v_i)\,, \tag{1}$$

where $B = (\beta_1, \ldots, \beta_n)$ is a vector consisting of strictly positive numbers adding up to 1. The vector B controls how each individual investor is weighted in the collective investment problem. Note that the utility of the fund manager is only defined for values exceeding the collective guarantee. Lemma 2.1 states that $u_{B,G}$ is, in fact, a utility function, as it has already been shown in Branger et al. (2018b) for the case where all the individual guarantees are equal to zero.

Lemma 2.1. *$U_{B,G}$ is a strictly increasing and concave function on (G, ∞) for all $G = \sum_{i=1}^n G^i$ with $G^i \geq 0$, whose inverse marginal utility is given by*

$$I_{B,G}(\cdot) := (U'_{B,G})^{-1}(\cdot) = \sum_{i=1}^n I_{i,G^i}\left(\frac{\cdot}{\beta_i}\right). \tag{2}$$

Proof. The collective utility function given in (1) is itself defined through an optimization problem whose Lagrangian is given by

$$\mathcal{L} = \sum_{i=1}^n \beta_i U_{i,G^i}(v_i) + y\left(v - \sum_{i=1}^n v_i\right).$$

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial v_i} = \beta_i U'_{i,G^i}(v_i) - y = 0 \quad \Leftrightarrow \quad v_i = I_{i,G^i}\left(\frac{y}{\beta_i}\right) \quad \text{for all } i = 1, \ldots, n.$$

This results in

$$v = \sum_{i=1}^n I_{i,G^i}\left(\frac{y}{\beta_i}\right). \tag{3}$$

Now we define the function

$$I_{B,G} : (0,\infty) \to (G,\infty), \quad y \mapsto I_{B,G}(y) = \sum_{i=1}^{n} I_{i,G^i}\left(\frac{y}{\beta_i}\right).$$

Note that this function is strictly decreasing on $(0,\infty)$, that $\lim_{y\to 0} I_{B,G}(y) = \infty$ and that $\lim_{y\to\infty} I_{B,G}(y) = G$. Hence, for any $v \in (G,\infty)$ there exists a unique value $y \in (0,\infty)$ such that $I_{B,G}(y) = v$ for which the optimization problem (1) attains its maximum at $v_i = I_{i,G^i}\left(\frac{y}{\beta_i}\right)$. This maximum collective utility level is given by

$$U_{B,G}(v) = \sum_{j=1}^{n} \beta_j U_{j,G^j}(v_j) = \sum_{j=1}^{n} \beta_j U_{j,G^j}\left(I_{j,G^j}\left(\frac{y}{\beta_j}\right)\right). \tag{4}$$

The first order derivative of $U_{B,G}$ can now be determined as

$$U'_{B,G}(v) = \sum_{j=1}^{n} \beta_j U'_{j,G^j}\left(I_{j,G^j}\left(\frac{y}{\beta_i}\right)\right) I'_{j,G^j}\left(\frac{y}{\beta_j}\right)\frac{dy}{\beta_j dv} = y\sum_{j=1}^{n} I'_{j,G^j}\left(\frac{y}{\beta_j}\right)\frac{dy}{\beta_j dv} = y,$$

where the last equality can be obtained from taking the derivative with respect to v on both sides of (3). This leads to

$$(U'_{B,G})^{-1}(y) = v = \sum_{i=1}^{n} I_{i,G^i}\left(\frac{y}{\beta_i}\right),$$

which completes the proof. $\square$

3 Constant volatility model

We start our analysis by a brief consideration of the classic Black-Scholes model which assumes a constant volatility of the risky asset. It will serve as a comparison basis to the stochastic volatility case specified in Section 4.

3.1 Financial market

We consider a financial market consisting of a risk-free asset B and a risky asset S. The risk-free asset B is assumed to earn a constant interest rate r, that is,

$$dB_t = rB_t dt, \ \ B_0 = 1. \tag{5}$$

Let $\{W_t\}_{t\in[0,T]}$ be a standard Brownian motion on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ satisfying the usual hypothesis. The risky asset S follows a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \;\; S_0 = s.$$

Here we assume that $\mu - r > 0$ and $\sigma > 0$. In this complete market, the state price density process is uniquely determined by the following stochastic differential equation:

$$d\xi_t = -\xi_t \left(r dt + \chi dW_t \right), \quad \xi_0 = 1, \quad \chi = \frac{\mu - r}{\sigma}.$$

The value ξ_t can be interpreted as the state of the economy at time t: The better the market performs, the lower ξ_t gets. In the following sections, we will use this property to analyze the performance of the (state-dependent) terminal wealth and investment strategy under different market states. From now on, let $G_t^i = e^{-r(T-t)}G^i$ denote the time-t-value of the fixed level of guarantee investor i requires, where $G^i = G_T^i \in (0, x_i e^{rT})$. We assume an upper bound for the guarantee to ensure the feasibility of our optimization problems.

If we denote by $\{\pi_t\}_{t\in[0,T]}$ the fraction of total wealth which is invested in the risky asset by the fund manager and assuming a self-financing trading strategy, the dynamics of the total wealth $\{X_t\}_{t\in[0,T]}$ are described by the following stochastic differential equation:

$$dX_t = \left(r + \pi_t(\mu - r) \right) X_t dt + \sigma \pi_t X_t dW_t \,, \quad X_0 = x. \tag{6}$$

The trading strategy $\{\pi_t\}_{t\in[0,T]}$ is chosen from the following set of admissible strategies:

$$\mathcal{A}(x) := \Bigg\{ \{\pi_t\}_{t\in[0,T]} \;\Bigg|\; X_0 = x, \; \{\pi_t\}_{t\in[0,T]} \text{ is progressively measurable,} \\ X_t \geq 0 \text{ for all } t \geq 0, \; \int_0^T \pi_s^2 ds < \infty \text{ a.s.} \Bigg\}.$$

3.2 Collective optimization problem

In a constant volatility framework, the collective optimization problem can be written down as

$$\max_{(\pi_t)_{t\in[0,T]}} \mathbb{E}\left[U_{B,G}(X_T)\right] \quad \text{subject to (6)}\,. \tag{7}$$

Due to the market completeness, this problem can be solved using the static martingale approach (Cox and Huang (1989)), that is, by solving the static optimization problem

$$\max_{X_T} \mathbb{E}\left[U_{B,G}(X_T)\right] \quad \text{subject to} \quad \mathbb{E}\left[\xi_T X_T\right] = x \tag{8}$$

for the optimal terminal wealth X_T and then determining the optimal trading strategy from the optimal wealth. To ensure the feasibility of the optimization problem (8), we shall examine the following two integrability conditions:

$$\mathbb{E}\left[\xi_T I_{B,G}(\lambda\xi_T)\right] < \infty, \tag{9}$$

$$\mathbb{E}\left[U_{B,G}\left(I_{B,G}(\lambda\xi_T)\right)\right] < \infty, \tag{10}$$

for all $\lambda > 0$. Condition (9) ensures that the initial market value of the terminal wealth is finite for all possible values of the Lagrangian multiplier. Condition (10) ensures that the value function is finite for all possible values of the Lagrangian multiplier. Note that both conditions are fulfilled in our Black-Scholes financial market with (modified) power utility functions.

Due to the nice property of the collective utility function, particularly the explicit representation of the inverse marginal utility of $U_{B,G}$, we obtain the solution of the optimization problem (8) as

$$X_T^* = I_{B,G}(\lambda\xi_T) = \sum_{i=1}^{n} I_{i,G^i}\left(\frac{\lambda}{\beta_i}\xi_T\right) = \sum_{i=1}^{n}\left(G^i + \left(\frac{\lambda}{\beta_i}\xi_T\right)^{-\frac{1}{\gamma_i}}\right), \tag{11}$$

where λ is the Lagrangian multiplier which can be uniquely determined from the budget constraint $\mathbb{E}[\xi_T X_T^*] = x$.

Remark 3.1. *For* $n = 1$, *Problem* (8) *is reduced to the individual optimization problem (taking individual* i *with an initial wealth* $x_i > G^i e^{-rT}$ *as an example):*

$$\max_{X_T^i} \mathbb{E}\left[U_{i,G_i}(X_T^i)\right] \quad \textit{subject to} \quad \mathbb{E}\left[\xi_T X_T^i\right] = x_i\,. \tag{12}$$

The individual optimal solution for individual i is given by

$$X_T^{(i,*)} = I_{i,G^i}(\lambda_i\xi_T), \tag{13}$$

where λ_i *can be determined explicitly from the budget constraint* $\mathbb{E}[\xi_T X_T^{(i,*)}] = x_i$.

From equation (11), we can derive the optimal wealth for any $t \in [0,T)$ as follows:

$$\begin{aligned}
X_t^* &= \mathbb{E}\left[\frac{\xi_T}{\xi_t} X_T^* \,\middle|\, \mathcal{F}_t\right] \\
&= G_t + \mathbb{E}\left[\frac{\xi_T}{\xi_t} \sum_{i=1}^n \left(\frac{\lambda}{\beta_i}\xi_T\right)^{-\frac{1}{\gamma_i}} \,\middle|\, \mathcal{F}_t\right] \\
&= G_t + \sum_{i=1}^n \mathbb{E}\left[\left(\frac{\xi_T}{\xi_t}\right)^{1-\frac{1}{\gamma_i}} \left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}} \,\middle|\, \mathcal{F}_t\right] \\
&= G_t + \sum_{i=1}^n \left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}} k_i(t),
\end{aligned} \tag{14}$$

where $k_i(t) := \mathbb{E}\left[\left(\frac{\xi_T}{\xi_t}\right)^{1-\frac{1}{\gamma_i}}\right] = e^{\left(1-\frac{1}{\gamma_i}\right)\left(-r-\frac{1}{2}\chi^2\right)(T-t)+\frac{1}{2}\chi^2\left(1-\frac{1}{\gamma_i}\right)^2(T-t)}$. Applying Itô's formula to (14) and comparing it to the wealth dynamics in (6), we obtain the self-financing investment strategy by equating the coefficients of dW_t:

$$\pi_t^* = \sum_{i=1}^n \frac{k_i(t)}{X_t^*} \left(\frac{\lambda}{\beta_i}\xi_t\right)^{-\frac{1}{\gamma_i}} \frac{\chi}{\sigma\gamma_i}\,, \tag{15}$$

where the terms $\frac{\chi}{\sigma\gamma_i}$ are the individual Merton portfolios (Merton (1971)). For the special case $n=1$, expression (15) simplifies as follows for an individual investor i:

$$\pi_t^{(i,*)} = \frac{\chi}{\sigma\gamma_i} \cdot \frac{k_i(t)\left(\lambda_i\xi_t\right)^{-\frac{1}{\gamma_i}}}{X_t^{(i,*)}} = \frac{\chi}{\sigma\gamma_i} \cdot \frac{X_t^{(i,*)} - G_t^i}{X_t^{(i,*)}} =: m_i \cdot \frac{X_t^{(i,*)} - G_t^i}{X_t^{(i,*)}}\,, \tag{16}$$

where $X_t^{(i,*)}$ can be obtained from (14) by setting $n=1$. The strategy in (16) is a CPPI strategy, where the multiplier m_i is the Merton portfolio of investor i. It is a well-known result that the strategy given in (16) is optimal for investors with modified power utility preferences (see, for example, Basak (2002)). The idea behind a CPPI strategy is simple: To ensure that the guarantee level G^i is met, the fraction of wealth $\frac{G_t^i}{X_t^{(i,*)}}$ is invested in the risk-free asset. The remainder $\frac{X_t^{(i,*)}-G_t^i}{X_t^{(i,*)}}$, where $X_t^{(i,*)} - G_t^i$ is the so-called *cushion*, multiplied by m_i is then the fraction of wealth invested in the risky asset. Note, however, that the collective optimal solution obtained in (15) is not a CPPI strategy. In this sense, there is a clear difference between the optimal individual and collective investment strategy. There are going to be losses in the collective expected utility if a CPPI investment strategy is applied by a fund manager. It would then be interesting to analyze the suboptimality induced by using CPPI strategy on the individual investors. We leave this analysis for future research.

4 Collective investment under stochastic volatility

As motivated in the introduction, the assumption of a constant volatility may not reflect a realistic financial market. In this section, we describe the evolution of the stock with a more general stochastic volatility model in the sense of Heston (1993). In this stochastic volatility setting, we solve the collective optimal investment problems, based on which we then study the welfare implications to the individual investors.

4.1 Financial market

We assume that the volatility of the risky asset is itself driven by a stochastic process. While the risk-free asset B remains as in (5), the drift and the volatility of the risky asset are now given by μ_t and $\sqrt{V_t}$. Further, let $\{W_t^{(1)}\}_{t\in[0,T]}$ and $\{W_t^{(2)}\}_{t\in[0,T]}$ be two independent Brownian motions in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The risky asset and its volatility are then assumed to follow the Heston model (Heston (1993)):

$$\begin{aligned} dS_t &= S_t \left(\mu_t dt + \sqrt{V_t} dW_t^{(1)}\right), \\ dV_t &= \kappa(\overline{V} - V_t)dt + \delta\sqrt{V_t}\left(\rho dW_t^{(1)} + \sqrt{1-\rho^2} dW_t^{(2)}\right), \end{aligned}$$

where $\rho \in (-1, 1)$ is a correlation coefficient, $\overline{V} > 0$ is the long-run mean for the variance, $\kappa > 0$ is the speed of mean reversion and $\delta > 0$ is the volatility of the variance.[6] In particular, the variance process follows a square-root process as used in the interest rate model in Cox et al. (1985). To ensure that the variance is almost surely positive at all times, we assume $2\kappa\overline{V} \geq \delta^2$ and $V_0 > 0$ (see Cox et al. (1985)).

The variance process contains the second source of risk which is not traded in the market and cannot be hedged. Therefore, the underlying financial market is incomplete. In other words, the market price for the second source of risk is not uniquely determined. A typical way to proceed is to choose a market price of risk for both sources of randomness and make the considered market artificially complete. This can be done by adding a derivative written on the risky asset to the financial market. This approach is well-known in the literature, see, for example, Liu and Pan (2003), Branger et al. (2008), Branger et al. (2017), Escobar et al. (2018) and Chen et al. (2018). We start by assuming

$$\frac{\mu_t - r}{\sqrt{V_t}} = \eta_1 \sqrt{V_t}, \tag{17}$$

[6]The perfect negative/positive correlation implies that the variance risk is fully hedgeable through trading in the underlying asset. In this case, we return to a complete market setting. The solution to this problem is then more similar to the constant volatility case.

and define the volatility risk premium as $\eta_2\sqrt{V_t}$, where η_1 and η_2 are constants. We set

$$\widetilde{W}_t^{(i)} = W_t^{(i)} + \int_0^t \eta_i \sqrt{V_s} ds$$

for $i = 1, 2$. By Girsanov's theorem, $\widetilde{W}_t^{(1)}$ and $\widetilde{W}_t^{(2)}$ are independent Brownian motions under the probability measure $\mathbb{P}^{(\eta)}$, $\eta := (\eta_1, \eta_2)$, which is defined by

$$\zeta_t := \frac{d\mathbb{P}^{(\eta)}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = \exp\left(-\int_0^t \eta_1 \sqrt{V_s} dW_s^{(1)} - \int_0^t \eta_2 \sqrt{V_s} dW_s^{(2)} - \frac{\eta_1^2 + \eta_2^2}{2} \int_0^t V_s ds\right)$$

for any $t \in [0, T]$. It is a density, as shown in Chen et al. (2018). Note that for the equivalent martingale measure $\mathbb{P}^{(\eta)}$, the process

$$\xi_t^{(\eta)} = e^{-rt} \frac{d\mathbb{P}^{(\eta)}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = e^{-rt}\zeta_t, \quad t \in [0, T],$$

is the corresponding pricing kernel or stochastic discounting process. The value $\xi_t^{(\eta)}$ has a similar interpretation as the state price density ξ_t in the constant volatility framework: The lower $\xi_t^{(\eta)}$ gets, the better the market performs. In the following sections, we express the optimal wealth and investment strategy in terms of $\xi_t^{(\eta)}$ and can thus easily interpret their performance under different market states. In this market, for any derivative O on the risky asset with some maturity $T_1 \leq T$, the no-arbitrage-price at time $t \leq T_1$ is now given by

$$O_t := \mathbb{E}\left[\frac{\xi_{T_1}^{(\eta)}}{\xi_t^{(\eta)}} O_{T_1} \;\middle|\; \mathcal{F}_t\right].$$

Now let g be a smooth function such that $O_t = g(t, S_t, V_t)$ (which exists as $\{S_t, V_t\}_{t\in[0,T]}$ is a Markov process). Since $e^{-rt}O_t$ is a martingale under $\mathbb{P}^{(\eta)}$, Itô's formula leads to

$$dO_t = rO_t dt + g_S S_t \sqrt{V_t} d\widetilde{W}_t^{(1)} + \delta g_V \sqrt{V_t}\left(\rho d\widetilde{W}_t^{(1)} + \sqrt{1-\rho^2} d\widetilde{W}_t^{(2)}\right),$$

where g_S and g_V are the first order partial derivatives of g with respect to the risky asset and the variance process. Now let π_t denote the fraction of wealth invested in the risky asset S and ϕ_t denote the fraction of wealth invested in the derivative O. The remainder $1 - \pi_t - \phi_t$ is invested in the risk-free asset. Assume that $\{\pi_t, \phi_t\}_{t\in[0,T]}$ is self-financing. This yields the following dynamics for the collective wealth process $\{X_t\}_{t\in[0,T]}$:

$$
\begin{aligned}
dX_t &= X_t \left(rdt + \left(\pi_t + \phi_t \frac{g_S S_t + \delta \rho g_V}{O_t} \right) \sqrt{V_t} d\widetilde{W}_t^{(1)} + \phi_t \frac{g_V \delta \sqrt{1-\rho^2}}{O_t} \sqrt{V_t} d\widetilde{W}_t^{(2)} \right) \\
&= X_t \left(rdt + \Theta_t^{(1)} \sqrt{V_t} d\widetilde{W}_t^{(1)} + \Theta_t^{(2)} \sqrt{V_t} d\widetilde{W}_t^{(2)} \right) \\
&= X_t \left(\left(r + \eta_1 \Theta_t^{(1)} V_t + \eta_2 \Theta_t^{(2)} V_t \right) dt + \Theta_t^{(1)} \sqrt{V_t} dW_t^{(1)} + \Theta_t^{(2)} \sqrt{V_t} dW_t^{(2)} \right), \qquad (18)
\end{aligned}
$$

where $X_0 = x$, $\Theta_t^{(1)}$ is the hedge demand and $\Theta_t^{(2)}$ is the speculative demand, following, for example, Liu and Pan (2003) and Chen et al. (2018). The hedge demand reflects the fund manager's position in the risk which is hedgeable by the risky asset. The speculative demand reflects the fund manager's position in the risk which cannot be hedged by trading in the risky asset. In our optimization problem, the hedge and speculative demand will be determined explicitly.

Before proceeding to the following sections, let us first state Lemma 4.1 which is of major importance when determining the solutions of our optimization problems.

Lemma 4.1. *Consider the following notation:*

$$
\begin{aligned}
\eta_+ &:= \rho\eta_1 + \sqrt{1-\rho^2}\eta_2, \\
\eta_- &:= \sqrt{1-\rho^2}\eta_1 - \rho\eta_2, \\
q_i &:= 1 - \frac{1}{\gamma_i}, \\
a_i &:= q_i\, \eta_+ \delta^{-1}, \\
b_i &:= q_i \left(\eta_+ \kappa \delta^{-1} + \frac{\eta_+^2}{2} + (1-q_i)\frac{\eta_-^2}{2} \right),
\end{aligned}
$$

and assume that

$$
\kappa^2 + q_i \left(2\eta_+ \kappa\delta + \eta_+^2 \delta^2 + \frac{1}{\gamma_i} \eta_-^2 \delta^2 \right) \geq 0 \quad \text{for all } i = 1, \ldots, n. \qquad (19)
$$

Then it holds

$$
k_i^{(\eta)}(t) := \mathbb{E}\left[\left(\frac{\xi_T^{(\eta)}}{\xi_t^{(\eta)}} \right)^{q_i} \,\middle|\, \mathcal{F}_t \right] = e^{-rq_i(T-t) + \kappa \overline{V} a_i (T-t)} \Psi_i(T-t, a_i, b_i, V_t) \qquad (20)
$$

for all $i = 1, \ldots, n$ *, where*

$$
\begin{aligned}
\Psi_i(s, a_i, b_i, V_t) &= \exp\left(-A_i(s) - V_t(B_i(s) - a_i)\right), \\
A_i(s) &= -\frac{2\kappa\overline{V}}{\delta^2} \ln\left(\frac{2\theta_i e^{(\theta_i+\kappa)\frac{s}{2}}}{\delta^2 a_i(e^{\theta_i s} - 1) + \theta_i(e^{\theta_i s} + 1) + \kappa(e^{\theta_i s} - 1)}\right), \\
B_i(s) &= \frac{a_i\left(\theta_i + \kappa + e^{\theta_i s}(\theta_i - \kappa)\right) + 2b_i(e^{\theta_i s} - 1)}{\delta^2 a_i(e^{\theta_i s} - 1) + \theta_i(e^{\theta_i s} + 1) + \kappa(e^{\theta_i s} - 1)}, \\
\theta_i &= \sqrt{\kappa^2 + 2b_i\delta^2}.
\end{aligned}
$$

Proof. See Appendix A. □

4.2 Collective optimization problem

The collective optimization problem under stochastic volatility can be expressed as:

$$
\max_{(\pi_t, \phi_t)_{t\in[0,T]}} \mathbb{E}\left[U_{B,G}(X_T)\right] \quad \text{subject to (18)}. \tag{21}
$$

Let us first mention that optimization for a single investor in a market with stochastic volatility has been considered extensively in the literature, see, for example, Pham (2002), Fleming and Hernández-Hernández (2003), Chacko and Viceira (2005), Kraft (2005) and Liu (2006) using the dynamic programming principle. For power utility functions, a closed-form solution can be obtained by applying a separation technique together with a verification step. This verification procedure is essential to make sure that the value function is finite (see, for example, Kraft (2005)). To the best of our knowledge, the optimization problem (21) has not yet been considered in the literature. In our collective framework, such a separation technique seems not possible. Hence, dynamic programming does not allow us to achieve an explicit solution to the value function and the investment strategies. Therefore, below, we solve Problem (21) by relying on a martingale approach which results in an explicit solution. To this end, we complete the market with an additional hedging instrument as discussed in the previous section. In particular, our objective is the following static problem

$$
\max_{X_T} \mathbb{E}\left[U_{B,G}(X_T)\right] \quad \text{subject to} \quad \mathbb{E}\left[\xi_T^{(\eta)} X_T\right] = x. \tag{22}
$$

In order to proceed with the collective utility maximization problem, we shall examine two integrability conditions similar to (9) and (10) to ensure that the optimization problem (22) is well-defined:

$$\mathbb{E}\left[\xi_T^{(\eta)} I_{B,G}(\lambda \xi_T^{(\eta)})\right] < \infty, \tag{23}$$

$$\mathbb{E}\left[U_{B,G}\left(I_{B,G}(\lambda \xi_T^{(\eta)})\right)\right] < \infty, \tag{24}$$

for all $\lambda > 0$. Note that assumption (19) is sufficient for both conditions to be fulfilled: For (23), this is straightforward to see. To show (24), we use (4) to obtain

$$\begin{aligned}\mathbb{E}\left[U_{B,G}\left(I_{B,G}(\lambda \xi_T^{(\eta)})\right)\right] &= \mathbb{E}\left[\sum_{i=1}^n \beta_i U_{i,G^i}\left(I_{i,G^i}\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)\right)\right] \\ &= \sum_{i=1}^n \frac{\beta_i}{1-\gamma_i}\mathbb{E}\left[\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)^{-\frac{1-\gamma_i}{\gamma_i}}\right],\end{aligned}$$

which leads to (19) again. Thus, roughly speaking, the verification result needed in the context of HJB now boils down to the integrability assumption (24) in the martingale approach.
The solution of Problem (22) can be obtained from the Lagrangian approach as

$$X_T^* = I_{B,G}(\lambda \xi_T^{(\eta)}) = \sum_{i=1}^n I_{i,G^i}\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right) = \sum_{i=1}^n \left(G^i + \left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_i}}\right), \tag{25}$$

where λ is determined from the budget constraint.

Remark 4.2. *For $n = 1$, Problem* (22) *is reduced to the individual optimization problem (taking individual i with an initial wealth $x_i > G^i e^{-rT}$ as an example):*

$$\max_{X_T^i} \mathbb{E}\left[U_{i,G_i}(X_T^i)\right] \quad \textit{subject to} \quad \mathbb{E}\left[\xi_T^{(\eta)} X_T^i\right] = x_i\,. \tag{26}$$

The individual optimal solution for individual i is given by

$$X_T^{(i,*)} = I_{i,G^i}(\lambda_i \xi_T^{(\eta)}), \tag{27}$$

where λ_i can be determined explicitly from the budget constraint and is given by

$$\lambda_i = \left(\frac{x_i - G_0^i}{\mathbb{E}\left[\left(\xi_T^{(\eta)}\right)^{1-\frac{1}{\gamma_i}}\right]}\right)^{-\gamma_i}.$$

Using Lemma 4.1, we can now determine the optimal strategy of Problem (22) explicitly.

Proposition 4.3. *Consider the optimization problem* (22)*. Using the notations in Lemma 4.1,* *the optimal wealth at time* $t \in [0,T)$ *is given by*

$$X_t^* = X_t^*\left(\xi_t^{(\eta)}, V_t\right) = \sum_{i=1}^n G_t^i + \left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t).$$

The optimal hedge and speculative demand are then given by

$$\Theta_t^{(1,*)} = \frac{1}{X_t^*}\sum_{i=1}^n \left(\frac{\eta_1}{\gamma_i} - \delta\rho H_i(T-t)\right)\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t), \tag{28}$$

$$\Theta_t^{(2,*)} = \frac{1}{X_t^*}\sum_{i=1}^n \left(\frac{\eta_2}{\gamma_i} - \delta\sqrt{1-\rho^2} H_i(T-t)\right)\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t), \tag{29}$$

where $H_i(s) := B_i(s) - a_i$. *In particular, the optimal fraction of wealth invested in the derivative and the risky asset at time* $t \in [0,T)$ *are then given by*

$$\phi_t^* = \frac{\sum_{i=1}^n \left(\frac{\eta_2}{\gamma_i} - \delta\sqrt{1-\rho^2}H_i(T-t)\right)\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t)}{X_t^* \frac{g_V \delta\sqrt{1-\rho^2}}{O_t}},$$

$$\pi_t^* = \frac{1}{X_t^*}\sum_{i=1}^n \left(\frac{\eta_1}{\gamma_i} - \delta\rho H_i(T-t)\right)\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t) - \phi_t^* \frac{g_S S_t + \delta\rho g_V}{O_t}.$$

Proof. Using Lemma 4.1, we obtain

$$\begin{aligned} X_t^* &= \mathbb{E}\left[\frac{\xi_T^{(\eta)}}{\xi_t^{(\eta)}} I_{B,G}(\lambda\xi_T^{(\eta)}) \;\middle|\; \mathcal{F}_t\right] = \sum_{i=1}^n \mathbb{E}\left[\frac{\xi_T^{(\eta)}}{\xi_t^{(\eta)}}\left(G^i + \left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_i}}\right) \;\middle|\; \mathcal{F}_t\right] \\ &= \sum_{i=1}^n G_t^i + \left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t). \end{aligned} \tag{30}$$

Next, we compute the optimal fractions of wealth invested in the risky asset π_t^* and the derivative ϕ_t^*. Observe that

$$\begin{aligned} \frac{\partial X_t^*}{\partial \xi_t^{(\eta)}} &= -\sum_{i=1}^n \frac{1}{\gamma_i}\left(\frac{\lambda}{\beta_i}\right)^{-\frac{1}{\gamma_i}} \left(\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}-1} k_i^{(\eta)}(t)\,, \\ \frac{\partial X_t^*}{\partial V_t} &= -\sum_{i=1}^n \left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t) H_i(T-t)\,. \end{aligned}$$

We use Itô's formula to represent the dynamics of the wealth process as

$$dX_t^* = (\text{drift})dt + \frac{\partial X_t^*}{\partial \xi_t^{(\eta)}} d\xi_t^{(\eta)} + \frac{\partial X_t^*}{\partial V_t} dV_t\,.$$

Recall that

$$d\xi_t^{(\eta)} = -\xi_t^{(\eta)}\left(rdt + \eta_1\sqrt{V_t}dW_t^{(1)} + \eta_2\sqrt{V_t}dW_t^{(2)}\right),$$
$$dV_t = \kappa(\overline{V} - V_t)dt + \delta\sqrt{V_t}(\rho dW_t^{(1)} + \sqrt{1-\rho^2}dW_t^{(2)}).$$

This leads to

$$\frac{\partial X_t^*}{\partial \xi_t^{(\eta)}} d\xi_t^{(\eta)} = \sum_{i=1}^n \frac{1}{\gamma_i}\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t)\left(rdt + \eta_1\sqrt{V_t}dW_t^{(1)} + \eta_2\sqrt{V_t}dW_t^{(2)}\right),$$
$$\frac{\partial X_t^*}{\partial V_t} dV_t = -\sum_{i=1}^n \left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t)H_i(T-t)\left(\kappa(\overline{V} - V_t)dt + \delta\sqrt{V_t}(\rho dW_t^{(1)} + \sqrt{1-\rho^2}dW_t^{(2)})\right).$$

From this, together with (18), we obtain

$$\Theta_t^{(1,*)} X_t^* \sqrt{V_t} dW_t^{(1)} = \sum_{i=1}^n \left(\frac{\eta_1}{\gamma_i} - \delta\rho H_i(T-t)\right)\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t)\sqrt{V_t}dW_t^{(1)},$$
$$\Theta_t^{(2,*)} X_t^* \sqrt{V_t} dW_t^{(2)} = \sum_{i=1}^n \left(\frac{\eta_2}{\gamma_i} - \delta\sqrt{1-\rho^2} H_i(T-t)\right)\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t)\sqrt{V_t}dW_t^{(2)}.$$

From the definitions of $\Theta_t^{(1)}$ and $\Theta_t^{(2)}$ as given in (18), it is straightforward to derive formulas for π_t^* and ϕ_t^*. $\square$

Remark 4.4. *As pointed out in Chen et al. (2018) (Appendix F), individual i's optimal hedge and speculative demand in the case without guarantees are given by $\frac{\eta_1}{\gamma_i} - \delta\rho H_i(T-t)$ and $\frac{\eta_2}{\gamma_i} - \delta\sqrt{1-\rho^2}H_i(T-t)$, respectively. Hence, the collective optimal hedge and speculative demand are given as the sum of the optimal individual demands without guarantees, weighted by $\left(\frac{\lambda}{\beta_i}\xi_t^{(\eta)}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)}(t)/X_t^*$, which is individual i's surplus (see (30) for $n = 1$) divided by the collective optimal wealth.*

Let us consider a numerical example. The base case parameter choice is summarized in Table 1.[7] Concerning the choice of the volatility parameters and the correlation, we follow Liu and Pan (2003) whose choice of parameters is "in the generally agreed region" of the empirical studies by Andersen et al. (2002), Pan (2002) and Eraker et al. (2003). Concerning the risk premiums, we also follow Liu and Pan (2003). In particular, we assume that volatility risk is negatively priced, as supported by the findings of Benzoni (1998), Chernov and Ghysels (2000), Pan (2002) and Bakshi and Kapadia (2003). Note that the negative price of volatility risk leads the investor to seek a short position in volatility risk. In other words, under negatively priced volatility risk, an investor seeks a short position in derivatives with positive exposure to the

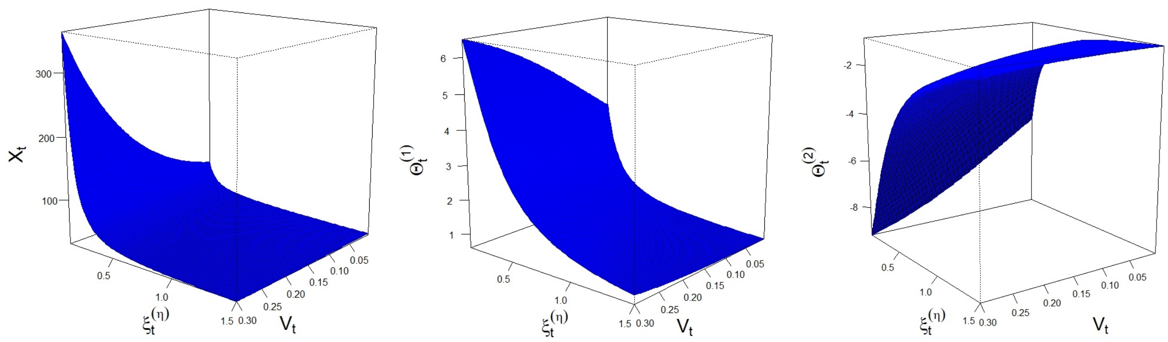
[7] For simplicity, we assume a rather short investment horizon. Our main qualitative arguments do not change under a longer maturity.

Volatility parameters $V_0 = \overline{V} = 0.13^2$, $\kappa = 5$, $\delta = 0.25$	Risk premiums $\eta_1 = 4$, $\eta_2 = -6$	Correlation $\rho = -0.4$
Risk-free rate, maturity $r = 0.02$, $T = 1$	Pool size $n = 30$	Weights β_i $\beta_i = \frac{(\lambda_i)^{-1}}{\sum_{i=1}^{n}(\lambda_i)^{-1}}$, $i = 1, \ldots, n$
Degrees of risk aversion $\gamma_i = \frac{1}{2} + \frac{9.5(i-1)}{n-1}$, $i = 1, \ldots, n$	Initial wealth $x_i = 1$, $i = 1, \ldots, n$	Guarantee $G_i = 0.5$, $i = 1, \ldots, n$

Table 1: Base case parameters.

volatility risk. The choice of the weights β_i is motivated by Section 5, where we show that the collective terminal wealth (25) rewrites to the sum of individual terminal wealths (27) under these weights.

In Figure 1 the optimal wealth at time $t = 1/2$, the hedge demand and the speculative demand are plotted as functions of the pricing kernel $\xi_t^{(\eta)}$ and the instantaneous variance V_t at $t = 1/2$. We see that the optimal wealth and the hedge demand are increasing in the



(a) Optimal wealth (b) Hedge demand (c) Speculative demand

Figure 1: The optimal wealth X_t^*, the hedge demand $\Theta_t^{(1,*)}$ and the speculative demand $\Theta_t^{(2,*)}$ as functions of the pricing kernel $\xi_t^{(\eta)}$ and the variance V_t at $t = T/2$ for the base case parameter setup summarized in Table 1.

variance V_t and that the speculative demand is decreasing in V_t. The reason for the increase of the wealth and the hedge demand in this example is assumption (17) along with the positive choice of η_1 which imply that an increase in the volatility at time t yields a higher rate of return per unit of volatility. The decrease of the speculative demand can be explained by the negatively priced volatility risk ($\eta_2 < 0$). Regarding the pricing kernel, we observe that well-performing markets (a low value of $\xi_t^{(\eta)}$) lead to a higher wealth and hedge demand and a lower speculative demand. Further, we see that the hedge demand is positive (long position) while the speculative demand is negative (short position) in all scenarios. The reason for this "reverse" behavior of the hedge and speculative demand in our parameter setup is the negative volatility risk premium which results from choosing η_2 smaller than zero.

5 Achieving individual optimal solutions

In this section, we address the question how the optimal terminal wealth (25) can be shared among the individuals in the collective. As it is the fund manager's primal goal to meet individual guarantees G^i, we assume that the fund manager starts by distributing to each individual her guarantee G^i. Further, let $(\alpha_i)_{i=1,\dots,n}$ be any (possibly state-dependent) sharing rule satisfying $\alpha_i \geq 0$ for all $i = 1, \dots, n$ and $\sum_{i=1}^n \alpha_i = 1$. This sharing rule is applied to the wealth exceeding the collective guarantee and determines the fraction of terminal surplus each individual receives. That is, for any collective terminal wealth $X_T > G$, investor i receives $X_T^i = G^i + \alpha_i(X_T - G)$. Based on the optimal collective terminal wealth (25), a natural candidate for the terminal wealth which investor i obtains is given by

$$X_T^i = I_{i,G^i}\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right) = G^i + \left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_i}}, \tag{31}$$

that is,

$$\alpha_i(\xi_T^{(\eta)}) = \frac{\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_i}}}{X_T^* - G}. \tag{32}$$

Naturally, the question arises whether a *fair* way of sharing the surplus can be achieved by a specific choice of the weights β_i, as the sharing rule in (32) is not necessarily fair. Without fairness, there might be some individuals in the collective who profit from the collective investment and some who suffer losses (compared to their individual investment). Let us now assume that the financial fairness condition as considered in Bühlmann and Jewell (1979) or, more recently, also in Schumacher (2018) is fulfilled. To be precise, we assume that the initial market value of the terminal payoff received by each investor i equals the initial contribution of this investor, that is,

$$x_i = \mathbb{E}\left[\xi_T^{(\eta)} I_{i,G^i}\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)\right]. \tag{33}$$

In our setting, it is then possible to return to each investor her individual optimum as obtained from Problem (22) for $n = 1$.

Proposition 5.1. *We assume that each investor in the collective receives the terminal wealth given in* (31)*. If we further impose the financial fairness condition* (33)*, each investor in the collective obtains her individual optimum as given in* (27)*.*

Proof. Let us introduce the notation $k_i^{(\eta)} = k_i^{(\eta)}(0)$ as defined in equation (20). The financial fairness condition delivers

$$x_i = \mathbb{E}\left[\xi_T^{(\eta)} I_{i,G^i}\left(\frac{\lambda}{\beta_i}\xi_T^{(\eta)}\right)\right] = G_0^i + \left(\frac{\lambda}{\beta_i}\right)^{-\frac{1}{\gamma_i}} k_i^{(\eta)} \quad \Leftrightarrow \quad \frac{\beta_i}{\lambda} = \left(\frac{x_i - G_0^i}{k_i^{(\eta)}}\right)^{\gamma_i}.$$

Using the fact that the β_i add up to 1, we obtain

$$\lambda = \frac{1}{\sum_{i=1}^n \left(\frac{x_i - G_0^i}{k_i^{(\eta)}}\right)^{\gamma_i}}, \quad \beta_i = \frac{\left(\frac{x_i - G_0^i}{k_i^{(\eta)}}\right)^{\gamma_i}}{\sum_{j=1}^n \left(\frac{x_j - G_0^j}{k_j^{(\eta)}}\right)^{\gamma_j}}. \tag{34}$$

For the case $n = 1$, the budget constraint of Problem (22) can be written as

$$x_i = \mathbb{E}\left[\xi_T^{(\eta)} I_{i,G^i}\left(\lambda_i \xi_T^{(\eta)}\right)\right] = G_0^i + \lambda_i^{-\frac{1}{\gamma_i}} k_i^{(\eta)}.$$

Plugging this expression into the two expressions given in (34), we obtain

$$\lambda = \frac{1}{\sum_{i=1}^n (\lambda_i)^{-1}}, \quad \beta_i = \frac{(\lambda_i)^{-1}}{\sum_{i=1}^n (\lambda_i)^{-1}}. \tag{35}$$

Consequently, each investor's terminal payoff in (31) simplifies to $I_{i,G^i}(\lambda_i \xi_T^{(\eta)})$. $\square$

Proposition 5.1 states that it is possible to achieve the individual optimal terminal wealth for all the individuals in the collective. The main assumptions for this result are the financial fairness and the use of a state-dependent sharing rule. This result is also valid under constant volatility and can be proven in the exact same way by simply replacing $\xi_T^{(\eta)}$ with ξ_T. In fact, this result has already been proven in Branger et al. (2018b) in a Black-Scholes market for CRRA utility functions (i.e. with all the individual guarantees being equal to zero).

Under the financial fairness condition, the sharing rule (32) can be simplified to the following:

$$\alpha_i(\xi_T^{(\eta)}) = \frac{\left(\lambda_i \xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_i}}}{\sum_{j=1}^n \left(\lambda_j \xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_j}}} \tag{36}$$

for all $i = 1, \ldots, n$. A disadvantage of this sharing rule is, however, that it depends on the market state at maturity and is, thus, not easy to communicate. Therefore, in the following section, we consider two examples of more prevailing sharing rules which are easier to communicate.

6 Linear sharing rules and welfare analysis

In practice, sharing rules that are easier to communicate, like linear (or affine) sharing rules, are applied. We aim at finding out how these sharing rules affect individuals' benefits in a stochastic volatility setting and compare the results to a constant volatility setting. Following the existing literature, we consider two further sharing rules in addition to the sharing rule defined in (36):[8]

- **Linear sharing rule (without financial fairness):** One of the most frequently used sharing rules in practice is the linear (or affine) sharing rule $(\widetilde{\alpha}_i)_{i=1,\dots,n}$ defined by

$$\widetilde{\alpha}_i = \frac{x_i}{x}, \quad i = 1, \dots, n. \tag{37}$$

  The shares that the individuals obtain from the total surpluses correspond to the shares of their initial investment in the fund. This simple sharing rule is known at time 0 and is thus easier to communicate than the state-dependent sharing rule (36). In addition, it can be shown that this linear sharing rule does not necessarily fulfill the financial fairness condition. It has been documented that this linear sharing rule is suboptimal, see, for example, Jensen and Nielsen (2016) and Branger et al. (2018a). In our numerical analysis, we will quantify the possible utility loss for the individuals.

- **Financially fair linear sharing rule:** A slight modification of (37) delivers the fair sharing rule $(\widehat{\alpha}_i)_{i=1,\dots,n}$ which is defined as

$$\widehat{\alpha}_i = \frac{x_i - G_0^i}{x - G_0}, \quad i = 1, \dots, n. \tag{38}$$

  It is straightforward to check that this sharing rule results in a financially fair payoff to each individual.

Note that none of the two linear sharing rules manages to deliver the individually optimal solutions to all the individuals in the collective. In the following, we compare the well-being of the investors in the financial market under the sharing rules introduced above. To measure the well-being of the investors, we rely on the certainty equivalent return introduced in Subsection 6.1. For our analysis, we assume that the weights β_i are given as in (35).

[8] The sharing rules considered here are inspired by previous works concerning a collective of individuals facing a joint decision under uncertainty. For example, Wilson (1968) and Huang and Litzenberger (1985) analyze the Pareto optimality of sharing rules. Weinbaum (2009) considers two individuals with different utility functions who are tied together by a social planner who uses a weighted sum of the individual utility functions and then characterizes the optimal sharing rule implicitly. Jensen and Nielsen (2016) consider a similar social planner whose sharing rule is initially fixed to be linear, though. Branger et al. (2018a) consider a rather similar setting as Jensen and Nielsen (2016) but generalize the analysis to n investors instead of two.

6.1 Certainty equivalent

For any investor i and a given terminal payoff X_T^i, the certainty equivalent *wealth* is denoted by $\mathrm{CE}_i = \mathrm{CE}_i(X_T^i)$. It is defined as the deterministic wealth level which yields the same expected utility as some terminal wealth X_T^i:

$$U_{i,G^i}(\mathrm{CE}_i) = \mathbb{E}[U_{i,G^i}(X_T^i)].$$

This results in

$$\mathrm{CE}_i = \mathbb{E}\left[(\alpha_i(X_T^* - G))^{1-\gamma_i}\right]^{\frac{1}{1-\gamma_i}} + G^i$$

for some sharing rule $(\alpha_i)_{i=1,\dots,n}$. Under the state-dependent sharing rule (36), we can compute the certainty equivalent of investor i in the Heston model as

$$\begin{aligned}
\mathrm{CE}_i^* &= G^i + \mathbb{E}\left[\left(\lambda_i \xi_T^{(\eta)}\right)^{-\frac{1-\gamma_i}{\gamma_i}}\right]^{\frac{1}{1-\gamma_i}} \\
&= G^i + \lambda_i^{-\frac{1}{\gamma_i}} \mathbb{E}\left[\left(\xi_T^{(\eta)}\right)^{1-\frac{1}{\gamma_i}}\right]^{\frac{1}{1-\gamma_i}} \\
&= G^i + \lambda_i^{-\frac{1}{\gamma_i}} \left(k_i^{(\eta)}\right)^{\frac{1}{1-\gamma_i}} .
\end{aligned} \tag{39}$$

Under the linear sharing rule (37), for example, we can compute the certainty equivalent of investor i in the Heston model as

$$\begin{aligned}
\widetilde{\mathrm{CE}}_i &= \mathbb{E}\left[(\widetilde{\alpha}_i(X_T^* - G))^{1-\gamma_i}\right]^{\frac{1}{1-\gamma_i}} + G^i \\
&= \widetilde{\alpha}_i \mathbb{E}\left[(X_T^* - G)^{1-\gamma_i}\right]^{\frac{1}{1-\gamma_i}} + G^i \\
&= \widetilde{\alpha}_i \mathbb{E}\left[\left(\sum_{j=1}^{n}\left(\lambda_j \xi_T^{(\eta)}\right)^{-\frac{1}{\gamma_j}}\right)^{1-\gamma_i}\right]^{\frac{1}{1-\gamma_i}} + G^i,
\end{aligned} \tag{40}$$

and an analogous calculation can be carried out for the linear sharing rule (38). Inspired by, for example, Zieling et al. (2014) and Branger et al. (2018a), we now consider the certainty equivalent *return*. It is defined as the deterministic rate of return $y_i = y_i(X_T^i)$ which delivers the same utility as some state-dependent terminal wealth, that is,

$$U_{i,G^i}\left(x_i e^{y_i T}\right) = \mathbb{E}[U_{i,G^i}(X_T^i)] \quad \Leftrightarrow \quad y_i = \frac{1}{T}\ln\left(\frac{\mathrm{CE}_i(X_T^i)}{x_i}\right). \tag{41}$$

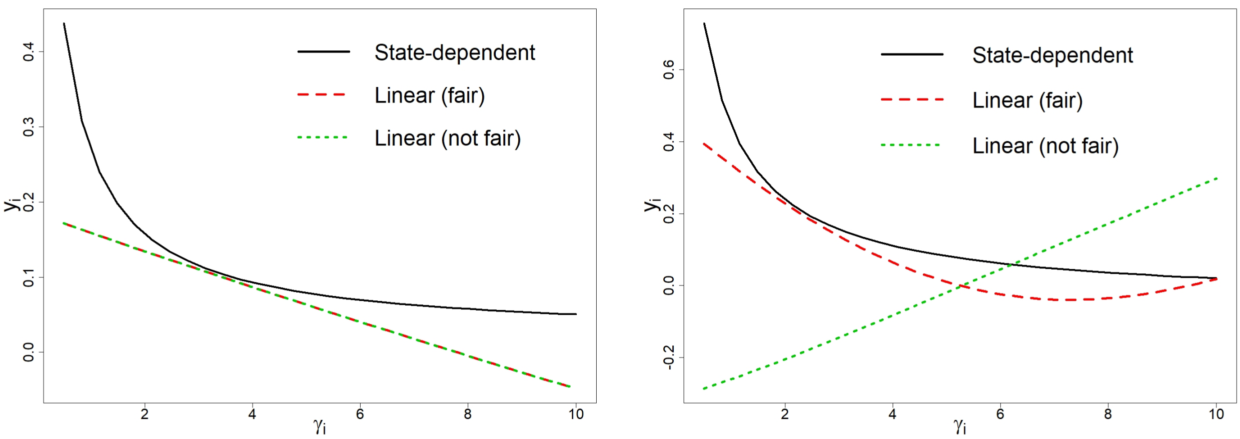
The certainty equivalent return is easier to interpret than the certainty equivalent wealth, particularly when individuals own different wealth levels.

6.2 Numerical analysis

In Figure 2, we compare the certainty equivalent returns defined in (41). Panel (a) demonstrates the certainty equivalent returns for the base case, where the parameters are listed in Table 1. In particular, we have used in the base case that the initial wealth levels and the required guarantees for all the individuals are identical, which implies that both linear sharing rules are identical. In addition to the base case, we show in Panel (b) the case of a guarantee which increases in γ_i. The guarantees for this case are chosen as

$$G^i = p_i x_i e^{g_i T}, \quad p_i = \frac{i-1}{n-1}, \quad g_i = -0.015 + 0.03\frac{i-1}{n-1}, \quad i = 1, \dots, n. \tag{42}$$

This illustrates a more realistic choice for the minimum guarantees: the more risk-averse an individual is, the higher is the minimum guarantee chosen. In Figure 2, we observe the following:



(a) Base case (b) Base case with increasing guarantee

Figure 2: Certainty equivalent returns (41) of the investors in the collective. In Panel (a), the base case is used while in Panel (b) we show the case given in (42).

- In Panel (a), we observe that imposing the linear sharing rule leads to losses for all the individuals, which can be considered as the (negative) deviation of the certainty equivalent returns from the state-dependent case (or individuals' optimal certainty equivalent returns). High losses arise for the least and the most risk-averse individuals. The benefits of some investors are hardly influenced by the linear sharing rule. The main reason for this result is probably that the optimal collective investment strategy is closest to the individual optimal investment strategies with γ_i around 3.5. For those who are least or most risk-averse, the optimal collective investment strategy differs significantly from their individual optimal investment strategies.
- In Panel (b), we make the following observations:[9]

[9]Note that in the determination of the optimal collective investment strategies, only the *total* guarantee

– **Fair linear sharing rule:** Similarly as in Panel (a), compared to the (fair) state-dependent sharing rule, the application of a fair linear sharing rule causes losses to all the investors. Different from Panel (a), individuals having medium and small risk aversions suffer most. The reason that highly risk-averse investors' losses are reduced compared to Panel (a) is the additional effect caused by their required high minimum guarantees.

– **Unfair linear sharing rule:** As the fund manager will first meet all the individual guarantees and then split the surpluses, individuals who require low guarantees implicitly finance the guarantees of individuals who demand high guarantees. For the unfair linear sharing rule, this argument seems to dominate. As a consequence, individuals demanding low guarantees suffer drastic losses, whereas investors demanding high guarantees are better off compared to their individual optimal solution. The certainty equivalents can even become negative for individuals with low risk aversion and low guarantees.

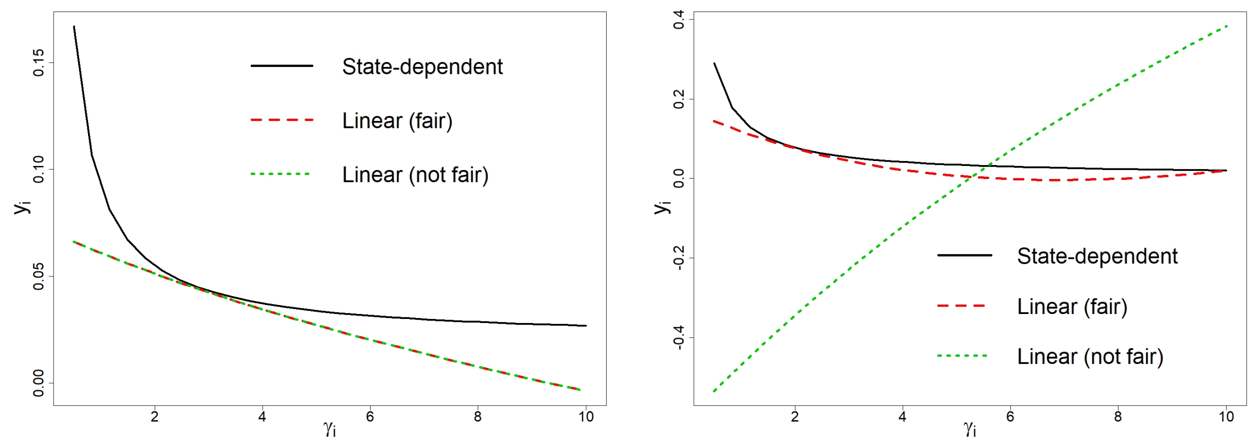
Due to the drastic losses occurring under unfair linear sharing rules (compared to moderate losses under a fair sharing rule), fairness shall certainly be taken into consideration if a linear sharing rule is applied in practice.

In Figure 3 we compare the certainty equivalents of the investors in the collective under constant volatility. We use the parameters from Table 1. The parameters of the Black-Scholes market are specified as $\mu = 0.0876$ and $\sigma = 0.13$. Note that we obtain $\frac{\mu-r}{\sigma} = \eta_1\sqrt{V}$ under these parameters. We consider the Black-Scholes analogue of the sharing rule defined in (36) (that is, we replace $\xi_T^{(\eta)}$ by ξ_T) and the linear sharing rules (37) and (38). We observe from Figure 3 that the certainty equivalent returns under constant volatility exhibit similar patterns as in the stochastic volatility case (cf. Figure 2). However, the certainty equivalent returns under constant volatility seem to be overall lower than in the stochastic volatility case. The "unfairness" caused by the (unfair) linear sharing rule (37) seems to weaken slightly.

In conclusion, by the use of the unfair linear sharing rule, individuals requiring high guarantees benefit largely from the collective, while those who require low guarantees suffer substantially from the collective. The fair linear sharing rule, on the other hand, causes only moderate losses to *all* the individuals in the collective. Thus, from a fund manager's perspective, to serve each individual in the collective fairly, the use of a financially fair sharing rule shall be preferred.

Given the widespread use of linear sharing rules, it is interesting to find out whether the sub-optimality of linear sharing rules will be amplified in the more realistic stochastic volatility

level plays a role, while the individuals' certainty equivalent returns do additionally depend on the *individual* guarantees.



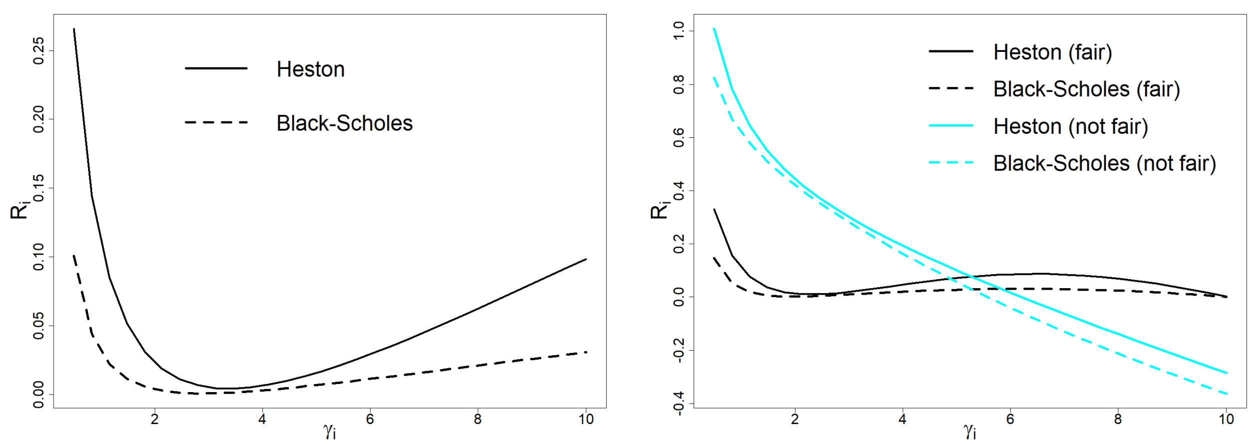
(a) Base case (b) Base case with increasing guarantee

Figure 3: Certainty equivalent returns (41) of the investors in the collective for constant volatility case (where $\xi_T^{(\eta)}$ is replaced by ξ_T). In Panel (a), the base case is used while in Panel (b) we show the case (42).

setting. For this purpose, we consider the quantity

$$R_i := y_i^* - y_i^\ell \tag{43}$$

for all individuals $i = 1, \ldots, n$, where y_i^ℓ is the certainty equivalent return under a linear sharing rule. They are provided in Figure 4. In both Panels, we observe that the curves



(a) Base case (b) Base case with increasing guarantee

Figure 4: Relative losses R_i as defined in (43) of the investors in the collective. In Panel (a), the base case is used while in Panel (b) we show the case (42).

resulting from the Heston model lie above those from the Black-Scholes model. In other words, the imposition of a suboptimal linear sharing rule leads to larger losses (and, in Panel (b), smaller gains) for all the individuals in the collective under the Heston model compared to the Black-Scholes model. Let us, for example, consider the base case: In the current parameter choice, an individual with parameter $\gamma_i \approx 6$ obtains a certainty equivalent return that is 3% lower than the optimal one in the Heston model and 1% lower than the one in the Black-Scholes model. Among the individuals in the collective, the difference between the two models is the largest for those with a very low risk aversion. This behavior can be explained by the thicker tails of the Heston model which delivers more extreme market scenarios than the Black-Scholes model. Thus, in a more realistic financial market setting with stochastic volatility, the sub-optimality of the linear sharing rule is intensified.

7 Conclusion

In this article, we solve a collective investment problem of a fund manager who invests for a collective of individuals who measure their utility from the terminal wealth exceeding a deterministic minimum guarantee, both in a Black-Scholes model and a Heston model. We have shown that all the investors in the collective receive their individually optimal terminal wealth levels under financial fairness when the fund manager uses a specific state-dependent sharing rule. Using a financially fair linear sharing rule leads to moderate losses for all investors in the collective. However, imposing a linear sharing rule which is not financially fair makes some individuals better and some worse off, compared to the financially fair linear sharing rule. As ignoring the financial fairness condition can lead to drastic losses for some individuals, a financially fair linear sharing rule performs better from a fund manager's perspective if she wants to take account of all individuals in the collective in a fair way. Our results show that losses imposed by linear sharing rules are larger under stochastic volatility than under constant volatility.

It would be interesting to analyze how individual utility is affected if the fund manager is restricted to some commonly applied investment strategies like, for example, CPPI strategies. We leave this analysis for future research.

References

Andersen, T. G., Benzoni, L., and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance*, 57(3):1239–1284.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, 16(2):527–566.

Balder, S. and Mahayni, A. (2010). How good are portfolio insurance strategies? In *Alternative Investments and Strategies* (eds. R. Kiesel, M. Scherer and R. Zagst), pages 227–257. World Scientific Publishing Co. Ltd., London.

Basak, S. (2002). A comparative study of portfolio insurance. *Journal of Economic Dynamics and Control*, 26(7-8):1217–1241.

Benzoni, L. (1998). Pricing options under stochastic volatility: an econometric analysis. *Manuscript, University of Minnesota*.

Black, F. and Jones, R. W. (1987). Simplifying portfolio insurance. *The Journal of Portfolio Management*, 14(1):48–51.

Black, F. and Perold, A. (1992). Theory of constant proportion portfolio insurance. *Journal of Economic Dynamics and Control*, 16(3-4):403–426.

Branger, N., Chen, A., Gatzert, N., and Mahayni, A. (2018a). Optimal investments under linear sharing rules. *Working paper. Available from the authors upon request.*

Branger, N., Chen, A., Mahayni, A., and Nguyen, T. (2018b). Optimal collective investment. *Working paper. Available at* `https://www.researchgate.net/publication/324910837`.

Branger, N., Muck, M., Seifried, F. T., and Weisheit, S. (2017). Optimal portfolios when variances and covariances can jump. *Journal of Economic Dynamics and Control*, 85:59–89.

Branger, N., Schlag, C., and Schneider, E. (2008). Optimal portfolios when volatility can jump. *Journal of Banking & Finance*, 32(6):1087–1097.

Bühlmann, H. and Jewell, W. S. (1979). Optimal risk exchanges. *ASTIN Bulletin: The Journal of the IAA*, 10(3):243–262.

Chacko, G. and Viceira, L. M. (2005). Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. *The Review of Financial Studies*, 18(4):1369–1402.

Chen, A., Nguyen, T., and Stadje, M. (2018). Optimal investment under VaR-regulation and minimum insurance. *Insurance: Mathematics & Economics*, 79:194–209.

Chernov, M. and Ghysels, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56(3):407–458.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton, Florida.

Cox, J. C. and Huang, C.-F. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83.

Cox, J. ., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–408.

Dumas, B. (1989). Two-person dynamic equilibrium in the capital market. *The Review of Financial Studies*, 2(2):157–188.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269–1300.

Escobar, M., Ferrando, S., and Rubtsov, A. (2018). Dynamic derivative strategies with stochastic interest rates and model uncertainty. *Journal of Economic Dynamics and Control*, 86:49–71.

Fleming, W. H. and Hernández-Hernández, D. (2003). An optimal consumption model with stochastic volatility. *Finance and Stochastics*, 7(2):245–262.

Hambardzumyan, H. and Korn, R. (2019). Dynamic hybrid products with guarantees – An optimal portfolio framework. *Insurance: Mathematics and Economics*, 84:54–66.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of Financial Studies*, 6(2):327–343.

Huang, C.-f. and Litzenberger, R. (1985). On the necessary condition for linear sharing and separation: a note. *Journal of Financial and Quantitative Analysis*, 20(3):381–384.

Jeanblanc, M., Yor, M., and Chesney, M. (2009). *Mathematical Methods for Financial Markets*. Springer, London.

Jensen, B. A. and Nielsen, J. A. (2016). How suboptimal are linear sharing rules? *Annals of Finance*, 12(2):221–243.

Jensen, B. A. and Sørensen, C. (2001). Paying for minimum interest rate guarantees: Who should compensate who? *European Financial Management*, 7(2):183–211.

Karatzas, I., Lehoczky, J. P., and Shreve, S. E. (1990). Existence and uniqueness of multi-agent equilibrium in a stochastic, dynamic consumption/investment model. *Mathematics of Operations Research*, 15(1):80–128.

Kim, H. H., Maurer, R., and Mitchell, O. S. (2016). Time is money: Rational life cycle inertia and the delegation of investment management. *Journal of Financial Economics*, 121(2):427–447.

Kraft, H. (2005). Optimal portfolios and Heston's stochastic volatility model: an explicit solution for power utility. *Quantitative Finance*, 5(3):303–313.

Liu, J. (2006). Portfolio selection in stochastic environments. *The Review of Financial Studies*, 20(1):1–39.

Liu, J. and Pan, J. (2003). Dynamic derivative strategies. *Journal of Financial Economics*, 69(3):401–430.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

OECD (2018). *OECD Pensions Outlook 2018*. OECD Pensions Outlook, OECD Publishing, Paris. Available at `https://doi.org/10.1787/pens_outlook-2018-en`.

Pain, D. and Rand, J. (2008). Recent developments in portfolio insurance. *Bank of England. Quarterly Bulletin*, 48(1):37–46.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3–50.

Pazdera, J., Schumacher, J. M., and Werker, B. J. (2016). Cooperative investment in incomplete markets under financial fairness. *Insurance: Mathematics and Economics*, 71:394–406.

Pham, H. (2002). Smooth solutions to optimal investment models with stochastic volatilities and portfolio constraints. *Applied Mathematics & Optimization*, 46(1):55–78.

Schumacher, J. M. (2018). Linear versus nonlinear allocation rules in risk sharing under financial fairness. *ASTIN Bulletin: The Journal of the IAA*, 48(3):995–1024.

Temocin, B. Z., Korn, R., and Selcuk-Kestel, A. S. (2018). Constant proportion portfolio insurance in defined contribution pension plan management. *Annals of Operations Research*, 266(1-2):329–348.

Turner, J. A. (2014). Hybrid pensions: risk sharing arrangements for pension plan sponsors and participants. *Society of Actuaries*.

Weinbaum, D. (2009). Investor heterogeneity, asset pricing and volatility dynamics. *Journal of Economic Dynamics and Control*, 33(7):1379–1397.

Wilson, R. (1968). The theory of syndicates. *Econometrica: Journal of the Econometric Society*, 36(1):119–132.

Xia, J. (2004). Multi-agent investment in incomplete markets. *Finance and Stochastics*, 8(2):241–259.

Zieling, D., Mahayni, A., and Balder, S. (2014). Performance evaluation of optimized portfolio insurance strategies. *Journal of Banking & Finance*, 43:212–225.

A Proof of Lemma 4.1

To compute the conditional expectation in (20), let us additionally introduce the following notation:

$$\begin{aligned} Z_t^+ &:= \rho W_t^{(1)} + \sqrt{1-\rho^2} W_t^{(2)}, \\ Z_t^- &:= \sqrt{1-\rho^2} W_t^{(1)} - \rho W_t^{(2)}, \\ \zeta_t^\pm &:= \exp\left(-\eta_\pm \cdot \int_0^t \sqrt{V_s} dZ_s^\pm - \frac{\eta_\pm^2}{2} \int_0^t V_s ds \right). \end{aligned}$$

Note that Z_t^+ and Z_t^- are independent Brownian motions because their covariation is equal to zero. Under this notation, we have

$$\int_s^t \sqrt{V_\nu} dZ_\nu^+ = \frac{1}{\delta} \left(V_t - V_s - \kappa \overline{V}(t-s) + \kappa \int_s^t V_\nu d\nu \right), \tag{44}$$

and we can write $\xi_t^{(\eta)} = e^{-rt} \zeta_t^+ \zeta_t^-$. This leads us to

$$\mathbb{E}\left[\left(\frac{\xi_T^{(\eta)}}{\xi_t^{(\eta)}} \right)^{q_i} \middle| \mathcal{F}_t \right] = e^{-rq_i(T-t)} \mathbb{E}\left[\left(\frac{\zeta_T^+ \zeta_T^-}{\zeta_t^+ \zeta_t^-} \right)^{q_i} \middle| \mathcal{F}_t \right].$$

Conditioning on the path $\{Z_s^+\}_{s\in[t,T]}$, which is the Brownian motion driving the volatility, the process ζ_t^- follows a log-normal distribution. Hence, using (44), this term can be expressed as

$$\begin{aligned} & e^{-rq_i(T-t)} \mathbb{E}\left[\left(\frac{\zeta_T^+}{\zeta_t^+} \right)^{q_i} \mathbb{E}\left[\left(\frac{\zeta_T^-}{\zeta_t^-} \right)^{q_i} \middle| \{Z_s^+\}_{s\in[t,T]} \right] \middle| \mathcal{F}_t \right] \\ &= e^{-rq_i(T-t)} \mathbb{E}\left[\left(\frac{\zeta_T^+}{\zeta_t^+} \right)^{q_i} e^{-q_i(1-q_i)\frac{\eta_-^2}{2} \int_t^T V_s ds} \middle| \mathcal{F}_t \right] \\ &= e^{-rq_i(T-t)} \mathbb{E}\left[e^{(V_t + \kappa \overline{V}(T-t)) a_i} e^{-a_i V_T - b_i \int_t^T V_s ds} \middle| \mathcal{F}_t \right] \\ &= e^{-rq_i(T-t)} e^{(V_t + \kappa \overline{V}(T-t)) a_i} \mathbb{E}\left[e^{-a_i V_T - b_i \int_t^T V_s ds} \middle| \mathcal{F}_t \right]. \end{aligned} \tag{45}$$

The expectation in (45) is the Laplace transform of $(V_T, \int_t^T V_s ds)$ at (a_i, b_i) . An explicit formula for this Laplace transform and the necessary conditions for this representation are given in Proposition 5.1 in Kraft (2005) and Proposition 6.3.4.1 in Jeanblanc et al. (2009). It is shown in Chen et al. (2018) that assumption (19) is sufficient for the Laplace transform to be well-defined at (a_i, b_i) for all $i = 1, \ldots, n$. Therefore, we can simplify (45) to (20). □

Curriculum Vitae

Der Inhalt dieser Seite wurde aus Gründen des Datenschutzes entfernt.

Der Inhalt dieser Seite wurde aus Gründen des Datenschutzes entfernt.