UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

INTEREST RATE AND MORTALITY RISK IN MODERN LIFE INSURANCE AND PENSION PRODUCTS

THÈSE DE DOCTORAT

Présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Doctorat en sciences actuarielles

par

Sascha GÜNTHER

Directeur de thèse
Prof. Peter Hieber

Jury
Prof. Valérie Chavez-Demoulin, présidente
Prof. Hansjoerg Albrecher, expert interne
Prof. Annamaria Olivieri, experte externe
Prof. Steven Vanduffel, expert externe

LAUSANNE
2025

UNIL | Université de Lausanne
Faculté des hautes études commerciales
Internef - 261
CH-1015 Lausanne

IMPRIMATUR

La Faculté des hautes études commerciales de l'Université de Lausanne autorise le dépôt de la thèse de doctorat rédigée par

Sascha GÜNTHER

intitulée

Interest Rate and Mortality Risk in Modern Life Insurance and Pension Products

sans se prononcer sur les opinions exprimées dans cette thèse.

Lausanne, le 07.07.2025

Professeure Marianne Schmid Mast, Doyenne

Thesis committee

PROF. **PETER HIEBER**
Thesis supervisor,
Department of Actuarial Science, Université de Lausanne, Switzerland.

PROF. **VALÉRIE CHAVEZ**
President of the jury,
Department of Operations, Université de Lausanne, Switzerland.

PROF. **HANSJÖRG ALBRECHER**
Internal expert,
Department of Actuarial Science, Université de Lausanne, Switzerland.

PROF. **ANNAMARIA OLIVIERI**
External expert,
Department of Economics, Università di Parma, Italy.

PROF. **STEVEN VANDUFFEL**
External expert,
Department of Business, Vrije Universiteit Brussel, Belgium.

Acknowledgements

I would like to start by thanking my supervisor, Peter Hieber, for teaching me how to do research. Even before I arrived in Lausanne, you never made me feel like I was working for you, but rather like I was working with you. Thank you for literally always having an open door and for giving me so many opportunities to gain experience, from organising summer schools together to teaching my own lectures. I am so happy we got to share so many memories, both in and out of work. Please stay as you are!

Thank you to my committee members, Prof. Annamaria Olivieri, Prof. Hansjörg Albrecher and Prof. Steven Vanduffel, for your insightful questions and helpful comments during my defense as well as at various seminars and conferences over the past few years.

Thank you to everyone at our DSA, the PhDnet, and UNIL for making it so easy to feel at home in Lausanne. A special thank you to Alaric and Brandon for all these amazing memories from conferences, holidays and nights out in Lausanne. I could not imagine anyone making these four years more entertaining than you two. Thank you to my office mates, Aleksandr and Andrey, as well as Nikolai, who completes our kayak-owning crew. Thank you to everyone at Extranef. Dina, Eman, Enej, Evgenii, Francesco, José-Miguel, Laura and Pavel, I will miss the many lunch breaks and after-work drinks, without which the PhD would not have been half as enjoyable. Thank you to Ekaterina, Felicia, Maria, Mikele, and many more from the PhDnet for all the fun events we organised together.

Thank you, Anna, Albrecht, Sascha and Vincent, for making my time studying in Dresden so incredibly enjoyable that I did not want to leave university. Thank you for several Zoom calls, holidays, game nights and simply for being great friends.

I would like to thank my whole family for always being so understanding, supportive, and interested in what I do, even though I am not at home as regularly as I used to be. A special thank you to my sister, Sandra, for kick-starting my affinity for math when you taught me how to do plus and minus at the age of five. Last but not least, thank you to my parents. I could write dozens of pages thanking you for everything you have done to shape who I am and what I know. So I just want to say thank you for being who you are.

Abstract

This thesis investigates the valuation, risk analysis and design of life insurance and pension products under interest rate and longevity risk. As populations age and interest rates remain volatile, insurers and pension providers have to develop sustainable products and tools to assess their exposure to changes in mortality rates and in the financial market. This work comprises four independent but related papers, each contributing new tools to address these risks from both theoretical and computational perspectives.

The first two papers focus on the valuation of cliquet-style equity-indexed annuities, which combine downside protection with equity participation. These products are path-dependent and sensitive to the interest rate environment. The first contribution introduces a scenario-matrix method under the Vasicek model, which allows for rapid and precise computation of risk measures without resorting to Monte-Carlo simulation. The second paper extends this approach to the G2++ model, which captures the term structure of interest rates. We demonstrate how to account for the correlation between long-term and short-term interest rates and show that efficient valuation remains feasible under a two-factor model.

The third paper addresses the computation of annuity conversion factors, which are central to defined contribution pension schemes. Using feed-forward neural networks trained on data generated from stochastic interest rate and mortality models, the paper proposes a fast and accurate method for approximating conversion factors. Thereby, we are able to reduce the amount of inner simulations in a nested Monte-Carlo simulation approach.

The final paper contributes to the growing literature on modern tontine design by introducing refundable income tontines, which incorporate partial refunds upon early death. We design a self-funded payment scheme that operates without an insurer and show that its payoff is higher than that of a refundable annuity with a high probability.

Together, these papers offer methodological and practical innovations for developing and analysing retirement products that are exposed to long-term risks in both financial markets and population demographics.

Contents

CHAPTER 1

Introduction

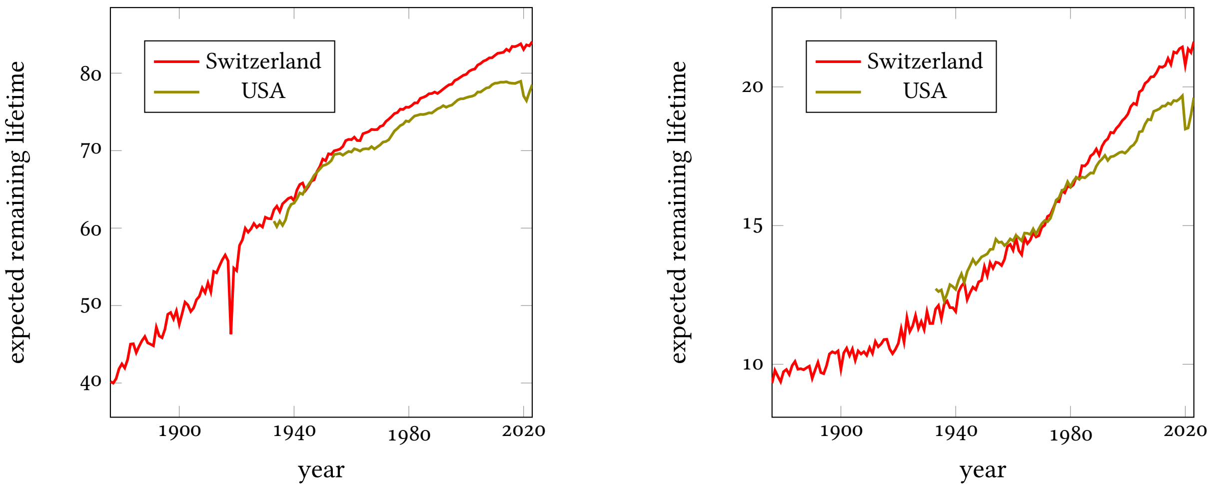
1.1. Motivation and Context

Over the past decades, industrialised nations have experienced significant demographic and economic transformations. Life insurers and pension systems manage liabilities over the next decades. Forecasting future developments over this time horzion comes with a lot of uncertainty. Two of the most important sources of uncertainty are *longevity risk*, due to continued improvements in mortality rates, and *interest rate risk*, most apparent in a prolonged period of historically low interest rates during the last decade. Both risks affect the solvency and sustainability of retirement products and pension schemes, as well as the financial security of retirees.

Life expectancy at birth has steadily increased in most developed countries with survival improvements extending well into advanced ages. Figure 1 shows the period life expectancy at birth and at age 65 in Switzerland and the United States of America. In Switzerland, the life expectancy at birth has more than doubled from around 40 years in 1876 to almost 85 years in 2023. Up until the 1940s, the increase in life expectancy was mostly achieved by reducing the mortality of infants, children and young adults. Since 1945, the life expectancy at age 65, which coincides with the retirement age in Switzerland, has increased by 10 years. While this trend is a result of societal and medical progress, it is also a significant financial burden on pension providers and annuity issuers. Traditional life-contingent products that offer guaranteed lifelong income become increasingly difficult to hedge, particularly when longevity exceeds the assumptions on which premiums and reserves were originally based. Moreover, systematic longevity risk - the uncertainty about future mortality improvements shared by all individuals of a population - cannot be diversified. This makes collective risk-sharing solutions preferable.

Simultaneously, interest rates have continuously declined since the 1980s. Especially, the decade following the 2008 financial crisis was characterised by an ultra-low or even negative interest rate regime. Figure 2 illustrates the interest rates offered by different central banks since the year 2000. This environment has significantly impacted the valuation and risk management of long-term liabilities. Guaranteed returns included in pension and life insurance contracts are often fixed decades in advance. Thus, they become increasingly expensive to hedge or replicate in a low-yield market. The rise in regulatory capital requirements under solvency regimes such as Solvency II further increases the sensitivity of insurers to interest rate movements, especially when liabilities are long-term and discounting plays a critical role. As a consequence, pension systems worldwide are gradually shifting from defined benefit (DB) schemes, where benefits are pre-specified and funded collectively, to defined contribution (DC) frameworks, where individual contributions and investment outcomes determine the retirement income. This transition reflects a general reallocation of risk - from insurers or states to individuals. However, this also introduces new concerns about fairness and risk-sharing in old age.

Innovations in product structuring, such as equity-indexed annuities and modern tontines, offer new ways to balance security, flexibility, and affordability for life insurance and pension



(A) life expectancy at birth (B) life expectancy at age 65

FIGURE 1. Historical expected remaining lifetimes of individuals living in Switzerland and the United States of America at different ages, based on data from Human Mortality Database (2022).

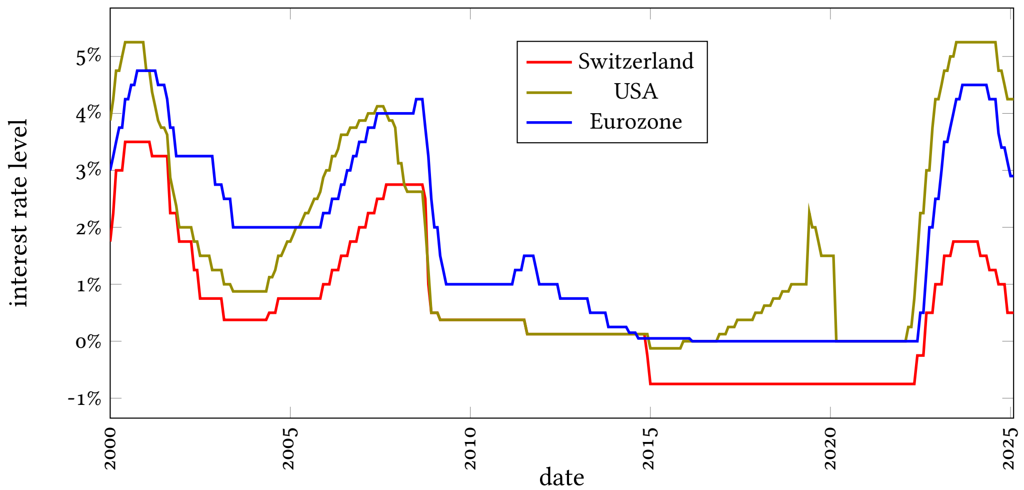


FIGURE 2. Historical interest rates offered by central banks in Switzerland, the United States of America and the Eurozone, based on data from Swiss National Bank (2025).

products. Equity-indexed annuities let policyholders benefit from higher returns of an underlying fund, while modern tontines pool their members' longevity risk without relying on an insurance company to provide guarantees. At the same time, advances in computational power and machine learning techniques open new opportunities to develop efficient algorithms for pricing and risk measurement.

This thesis focuses on the valuation and risk analysis of life insurance and pension products exposed to interest rate and longevity risk while considering both traditional and modern product designs. It contributes to the literature by offering tractable methods for complex valuation problems, proposing alternative product structures and applying advanced numerical and machine learning techniques to core actuarial challenges.

1.2. Life insurance and pension products under risk

Life insurance and pension products are long-term financial instruments designed to provide income security to policyholders who face uncertainty regarding lifespan, health status, and market returns. A common objective in the design of life-contingent products is to provide retirement income that is both sustainable and predictable. This can be achieved through guaranteed lifelong payments, typically offered via life annuities or structured payout schemes linked to market performance.

Participating products, such as equity-indexed annuities (EIAs), are addressing this demand. These instruments offer policyholders the potential to participate in financial market upside. Several variants retain a form of downside protection, often in the form of a minimum return or guaranteed accumulation. In cliquet-style or ratchet-type guarantees, for example, minimum returns are guaranteed annually. This way, they provide long-term reliability to the policyholder. However, these annual guarantees introduce path-dependence which makes their valuation particularly sensitive to interest rate dynamics and market volatility. In a setting of low interest rates and increasing longevity, guaranteed products become more difficult to manage for insurance companies. The capital intensity of these guarantees – due to their sensitivity to discount rates and mortality trends – puts pressure on insurers' balance sheets and solvency positions.

Additionally, longevity-linked products have also received increased attention. Deferred life annuities and longevity bonds have been introduced as tools for managing the tail of the lifespan distribution. Also, tontines have recently attracted interest as viable alternatives to annuities. In their modern form, they provide risk-pooling for the individual/idiosyncratic mortality risk without guarantees on the systematic mortality risk, and allow for fair redistribution of funds among surviving members. Their self-funding structure eliminates any need for risk capital which is why they are attractive under modern solvency capital regulations.

However, these designs are not without their challenges. Managing systematic mortality risk, ensuring intergenerational fairness, and addressing heterogeneity in the pool require careful actuarial modelling. Moreover, integrating refund guarantees, which are increasingly demanded in practice, adds complexity to traditional tontine schemes. Similarly, annuity conversion factors, which translate accumulated wealth into periodic income, must reflect current market conditions and cohort-specific mortality expectations, both of which are subject to significant uncertainty.

To account for these considerations, actuarial research has developed model-based approaches that can accommodate stochastic dynamics in both financial and mortality risk. These include affine term structure models, multi-factor interest rate models, and stochastic mortality frameworks. Many of these allow for analytical tractability and efficient numerical implementation. In addition, machine learning and data-driven techniques allow us to tackle high-dimensional, nonlinear problems, such as the estimation of annuity conversion factors or mortality surfaces.

This thesis addresses several of these challenges and innovations. It focuses on actuarially sound methods to quantify and mitigate financial and biometric risks in modern life insurance and pension products, with particular attention to interest rate modelling, mortality risk, and computational efficiency.

1.2.1. Interest rate risk. Interest rate risk is a central concern in the valuation and management of long-term insurance and pension liabilities. Life insurance products, especially those with minimum return guarantees, are very sensitive to the term structure of interest rates. Even small changes in yields can lead to significant variations in the present value of future cash flows. Hence, modelling, measuring, and managing interest rate risk is important in actuarial science and financial regulation.

Historically, life insurers worked in environments with relatively stable or declining interest rates. However, the prolonged low-yield period from the 2009 global financial crisis until 2022 showed how exposed the balance sheets of insurance companies are to interest rate risk. Products offering minimum return guarantees became difficult to fund without incurring losses. More recently, we have experienced a rapid rise in interest rates in 2022–2023, driven by inflationary pressures and monetary policy tightening. This has put another focus on understanding the impacts of rising versus falling rates on asset-liability management strategies.

EIAs combine equity participation with an annual minimum return, thus integrating a sequence of path-dependent options into the contract structure. The pricing and risk analysis of such products requires stochastic interest rate models, as the value of future guarantees critically depends on the evolution of interest rates over time. The Vasicek model, and its multi-factor extensions such as the G2++ model, provide a tractable framework for capturing interest rate dynamics while allowing for (semi-)closed-form expressions for key quantities of interest.

In Chapters 2 and 3, we address the interest rate risk in equity-indexed annuities from both analytical and simulation-based perspectives. In the first contribution, we consider a Vasicek model to compute values and risk measures of cliquet-style annuities efficiently. In the second paper, we extend the analysis to the G2++ model, a two-factor Gaussian model that captures the yield curve as well as the correlation between long-term and short-term interest rates. Using both, exact simulation and efficient valuation techniques, we assess the impact of stochastic interest rates on the distribution and risk profile of equity-indexed annuities.

We find that the interest rate environment can significantly affect both the expected value and tail risk of the annuity payoff. Moreover, traditional pricing methods that assume constant interest rates or ignore path dependence may significantly underestimate the capital requirements or hedging needs of such products. More broadly, our contributions suggest that modelling interest rates stochastically in life insurance and pension applications is crucial to accurately analyse future liabilities.

1.2.2. Mortality and longevity risk. Mortality and longevity risk are at the core of life insurance. These risks originate from the uncertainty surrounding the lifespan of individuals (the idiosyncratic risk) and populations (the systematic risk). In particular, longevity risk, i.e. the risk that policyholders live longer than expected, significantly impacts the solvency of institutions offering life-contingent benefits.

Historically, mortality rates have consistently improved for decades, driven by advances in healthcare, sanitation, nutrition, and socioeconomic development. However, the pace and structure of these improvements are difficult to forecast accurately, especially at older ages and in the long term. Sudden shifts due to medical innovation, epidemics, or structural population changes introduce systematic mortality risk, which affects all individuals in a cohort and cannot be diversified. This form of risk is particularly problematic for life annuities, deferred pensions, and longevity-linked securities, whose values depend heavily on long-term survival probabilities.

The actuarial literature has responded by developing increasingly sophisticated models to represent mortality dynamics. These range from simple laws for the idiosyncratic risk, such as Gompertz' law and Makeham's law, extrapolating life tables into the future via deterministic scaling, to stochastic mortality frameworks that incorporate cohort effects, volatility, and dependence structures, such as the Lee–Carter or the Li-Lee model. In practice, insurers often use regulatory mortality tables that are updated periodically to reflect evolving longevity trends. These life tables incorporate risk margins to account for the uncertainty of future mortality developments.

In response to these developments, Chapters 4 and 5 focus on computational methods and product innovations that address longevity risk. In Chapter 4, we propose an efficient approach to computing annuity conversion factors, which determine the periodic income that an individual receives upon retirement. They are sensitive to both interest rate assumptions and cohort-specific mortality projections.

In Chapter 5, we focus on quantifying the systematic longevity risk in alternative pension systems. Moreover, we specify how to update the annual payments based on new mortality data. This step is essential to retain fairness between different cohorts.

Taken together, these contributions offer practical tools and theoretical insights into the management of mortality and longevity risk in life insurance and pensions. They highlight the need for flexible valuation methods and adaptive product design.

1.3. Computational challenges in modern actuarial models

The complexity of life insurance and pension products combined with stochastic financial and mortality models poses significant computational challenges for actuaries and risk managers. Modern actuarial problems often involve high-dimensional, path-dependent structures that do not allow for closed-form solutions.

Classical valuation methods, such as analytical pricing formulas rely on restrictive assumptions, e.g., constant interest rates or deterministic mortality. In contrast, products like equity-indexed annuities with cliquet guarantees or dynamic tontine schemes involve payoff structures that are path-dependent, nonlinear, and sensitive to multiple sources of uncertainty. The interaction between stochastic interest rates, market returns, and mortality trends creates a multidimensional modelling environment in which standard tools struggle to provide fast and accurate results.

A commonly used solution is Monte-Carlo simulation, which can be applied to a broad class of problems. However, its convergence can be slow, particularly in the tail of the distribution, which is critical for solvency and risk capital calculations. Moreover, simulation-based approaches can become computationally expensive when nested simulations are required, for instance in computing annuity conversion factors.

To address these challenges, this thesis develops and applies a range of efficient numerical and machine-learning techniques to actuarial problems.

In the context of equity-indexed annuities, the scenario-matrix (SM) method introduced in Chapters 2 and 3 allows for efficient computation of the moment-generating function and tail probabilities of path-dependent payoffs. The method discretises the interest rate process and uses matrix iteration to obtain precise estimates of any moment of the terminal payoff. Unlike Fourier transform or projection techniques, the SM method avoids computationally intensive integral computations and converges rapidly. The tools developed in this thesis are illustrated in the context of EIAs and (two-factor) Vasicek models but can be adapted to other settings.

Chapter 4 introduces feed-forward neural networks as a tool for computing annuity conversion factors. By training these networks on simulated risk factors describing the interest rate and mortality risk, we approximate annuity conversion factors with high accuracy and low computational cost. Importantly, the computational complexity of this problem remains manageable for higher dimensional problems.

These approaches reflect a general shift in actuarial science toward hybrid approaches that combine traditional theory with modern computational techniques. Methods from numerical analysis, stochastic calculus, and machine learning are combined to handle the nonlinearity, path-dependence, and uncertainty in pension and insurance products.

Moreover, computational efficiency is not just a technical concern - it has practical implications for product design, pricing, and regulatory compliance. For example, efficient algorithms enable insurers to explore a wider range of product features during the development phase, meet reporting requirements under time constraints imposed by solvency regulations and reduce the cost for expensive simulations. Our contributions demonstrate how advanced computational techniques can be integrated into actuarial modelling.

1.4. From classical annuities to modern tontines

Classical life annuities provide guaranteed lifelong income in exchange for an upfront premium, thereby eliminating the policyholder's longevity risk. Insurance companies can diversify their portfolio to reduce the idiosyncratic mortality risk. However, they still hold the systematic mortality risk, which require insurers to hold significant reserves and capital, particularly in low interest rate environments. In exchange for covering the systematic mortality risk, they charge a risk premium on top of the fair annuity price.

In this context, there has been renewed interest in tontines - financial arrangements that pool longevity risk among participants without offering formal guarantees. The key attraction of tontines is that they are self-funded. Unlike guaranteed annuities, they do not require hedging of tail risks or capital buffers against systematic changes in the underlying mortality. Payments are conditional on survival, and the pool adjusts automatically to realised mortality. Policyholders exchange the security of having a lifelong guaranteed annuity for a higher expected benefit from a tontine that can operate without a risk premium. Moreover, annuitisation remains unpopular among retirees in many countries - a phenomenon often referred to as the annuity puzzle. Tontines may be one step towards solving this problem.

Despite these advantages, the practical adoption of tontines faces obstacles. One major concern is the perceived unfairness in the event of early death, where the entire remaining balance is "lost" to the pool. In response to this concern, Chapter 5 of this thesis introduces and investigates refundable income tontines - a class of products that combines longevity pooling with refund guarantees. These features allow for a return of the invested capital (or a part thereof) upon early death, either deterministically or stochastically, depending on the product structure.

We show that refund features can be integrated into tontine designs without undermining the core principle of being self-funded. The resulting products may be more attractive to potential participants, especially those concerned about bequests. Importantly, we provide a formal actuarial framework to analyse and price refundable tontines, accounting for systematic mortality risk and heterogeneous pools.

Furthermore, the proposed design allows for dynamic updating of mortality assumptions, enabling tontine funds to adapt over time in response to emerging longevity trends. This flexibility is particularly valuable in an environment where mortality tables may become outdated or misaligned with realised experience.

The analysis also emphasises the role of heterogeneity in tontine pools. When individuals differ in age, health status, or preferences, we are able to remain actuarially fair. Overall, modern tontines represent a promising solution for sustainable pension design. It avoids the capital constraints associated with guaranteed products. As such, it has the potential to complement or even replace traditional pension systems.

1.5. Structure of the thesis

This thesis contains four research articles, which contribute to the actuarial literature on the valuation, simulation, design, and modernisation of life-contingent financial products, with a particular focus on interest rate and longevity risks.

Chapter 2 develops a novel scenario-matrix (SM) method to analyse the interest rate risk in cliquet-style equity-indexed annuities. These contracts offer policyholders both a guaranteed minimum return and participation in the upside of an equity index. The path-dependent nature of the guarantees and their long maturities make them highly sensitive to interest rate dynamics. Our method uses the tractability of the Vasicek interest rate model to derive semi-closed-form expressions for the moment-generating function of the payoff. This in turn enables us to efficiently compute risk measures such as solvency ratios. The SM method avoids computationally intensive techniques such as Fourier inversion or simulation and demonstrates quadratic convergence in numerical experiments.

Building on the previous analysis, Chapter 3 considers the valuation of the same class of annuity products in a more flexible interest rate setting. The G2++ model, a two-factor Gaussian model, captures the initial yield curve and allows for more realistic correlations between long-term and short-term interest rates. The paper provides exact results on the distribution of the G2++ model for simulation purposes. Additionally, it adapts the SM method from Chapter 2 to evaluate the annuity payoff distribution under the G2++ model. The proposed method achieves high accuracy and computational efficiency, allowing for an efficient pricing mechanism and detailed sensitivity analysis.

Chapter 4 addresses the challenge of computing annuity conversion factors, which are critical for translating accumulated wealth at retirement into a lifelong income in pension systems. These factors depend on both interest rate assumptions and mortality projections, making them complex to calculate under stochastic models. The paper proposes the use of feed-forward neural networks trained on simulated data to approximate the conversion function. The approach is model-agnostic and computationally efficient, making it suitable for real-time applications and regulatory reporting.

Chapter 5 introduces and analyses a class of refundable income tontines, which extend classical tontine structures by incorporating capital refund mechanisms. The design allows for deterministic or stochastic refunds upon early death, making the product more attractive to risk-averse individuals concerned with bequest motives. The paper provides a comprehensive actuarial analysis of these products, including valuation, simulation of heterogeneous pools, and assessment of systematic longevity risk. It also lays out how to account for updated mortality tables, which is critical to retain fairness among different cohorts entering the tontine pool. Our findings suggest that tontine pools acquiring as little as 100 new members annually offer higher lifetime payments with more than 95% probability.

CHAPTER 2

Analyzing the interest rate risk of equity-indexed annuities via scenario matrices

This chapter is based on the following article:

S. Günther, P. Hieber (2024): Analyzing the interest rate risk of equity-indexed annuities via scenario matrices. *Insurance: Mathematics and Economics* 114:15–28.

Abstract. The financial return of equity-index annuities depends on an underlying fund or investment portfolio complemented by an investment guarantee. We discuss a so-called cliquet-style or ratchet-type guarantee granting a minimum annual return. Its path-dependent payoff complicates valuation and risk management, especially if interest rates are modelled stochastically. We develop a novel scenario-matrix (SM) method. In the example of a Vasicek-Black-Scholes model, we derive closed-form expressions for the value and moment-generating function of the final payoff in terms of the scenario matrix. This allows efficient evaluation of values and various risk measures, avoiding Monte-Carlo simulation or numerical Fourier inversion. In numerical tests, this procedure proves to converge quickly and outperforms the existing approaches in the literature in terms of computation time and accuracy.

2.1. Introduction

We discuss equity-indexed life insurance policies that guarantee a minimum annual return and allow to participate in returns of an underlying fund or investment portfolio (see, e.g., Briys and Varenne, 1994; Grosen and Jørgensen, 2000; Graf et al., 2011; Bacinello et al., 2011). The payoff assures the policyholder a protected investment with potential additional gains while the insurance company takes a share of the surplus to cover expenses and reach profit targets. Variants that are particularly common in central Europe are cliquet-style guarantees that provide a constant minimum return on the contract's investment. In the literature, these type of guarantees are also referred to as ratchet-type. The path-dependent nature of this product makes its valuation and risk management a complex task.

Commonly, these policies are settled with a long-term maturity. Due to the lack or high cost of long-term bonds, insurance companies can typically not fully eliminate their interest rate risk. The value of long-term guarantees depends substantially on interest rate risk and cliquet-options are prone to change in the interest rate environment. In the past, we have observed periods of low interest rates as well as rapid increases in the interest rate level. That is why it is advisable to consider interest rate risk for these types of products. Technically, modelling interest rates stochastically leads to a dependence between consecutive annual returns, complicating the analysis of cliquet-style guarantees. We contribute to the tractability of these products by introducing a novel scenario-matrix (SM) method.

The valuation and contract design of these types of guarantees is a frequently discussed problem. In Lévy settings, that is, under stationary and independent increments, the contract's value can be expressed in terms of integral equations or Fourier algorithms, see Kassberger et al. (2008), Alonso-García et al. (2017) or, for a non-compounding variant of cliquet-guarantees, Bernard and Li (2013). Korn et al. (2017) use a central-limit type argument to approximate guarantee values in Heston's stochastic volatility model. Miltersen and Persson (2003) consider a compounding type and investigate how to set up a mechanism for sharing the surplus between customer and insurer. Bacinello (2001) and Barbarin and Devolder (2005) outline strategies for choosing contract parameters such as the annual guarantee level and the participation rate.

The literature on cliquet guarantees in a stochastic interest rate environment is rather scarce, probably due to the technical challenges. While the case of a maturity guarantee (e.g., Lin and Tan, 2003; Barbarin and Devolder, 2005; Bernard et al., 2005) can be tackled by standard option pricing theory, the cliquet-style variant requires more elaborate techniques. Existing results mostly rely on simulation (e.g., Zaglauer and Bauer, 2008; Deelstra and Rayée, 2013; Hieber et al., 2019, and many others). The case of a regime switching interest rate where the valuation can use Fourier techniques (e.g., Fan et al., 2015; Ignatieva et al., 2016; Hieber, 2017; Cui et al., 2017) or Erlangization techniques (e.g., Deelstra and Hieber, 2023) is analytically rather tractable. For the special case of two subperiods, Persson and Aase (1997) and Miltersen and Persson (1999) derive premiums in closed-form in a generalisation of the Vasicek-Black-Scholes model. For the multi-period case, Kijima and Wong (2007) represent the value of a cliquet option in terms of a multi-dimensional normal integral. Further, a computationally more efficient approach is the PROJ method by Kirkby (2015), a frame projection approach that relies on density approximations through fast Fourier transformations. This idea is applied by Cui et al. (2017) to a cliquet option where the underlying can incorporate stochastic volatility models with jumps that are approximated by Markov chain processes. Kirkby (2023) uses the same technique for stochastic interest rate models.

In contrast to most of the previous articles, we do not only focus on the valuation problem, but also analyse the (interest rate) risk of cliquet-style guarantees. The 2009-2022 low interest rate period has revealed that this risk can be substantial. Further, the SM method relies neither on simulations nor on Fourier transformation algorithms. To achieve this goal, we focus on a financial market that is modelled by a Vasicek-Black-Scholes model, being less general than some of the related literature. Conditioning on the annual interest rates, this framework allows us to derive many quantities like the conditional moment-generating function of the logarithmic return in closed-form. This allows us to get values and higher moments of the payoff of the cliquet-option in terms of the scenario matrix (SM). In combination with techniques from Carr and Madan (1999) we obtain quantiles and the Value-at-Risk of cliquet-style guarantees. This proves to conveniently well approximate the true prices and risk measures. The final approximation requires the computation of (scenario) matrix multiplications only and turns out to be faster than benchmark techniques like the PROJ method.

The remainder of this article is organised as follows: In Section 2.2, we describe the underlying dynamics of the financial market and the payoff of the cliquet option that we want to consider. The SM method for efficiently estimating the moment-generating function and quantiles of the log-payoff is introduced in Section 2.3. A brief introduction to techniques like Fourier pricing or the PROJ method and its application to cliquet options are given in Section 2.4. These serve as our benchmarks for accuracy and time efficiency, which we consider in numerical experiments in Section 2.5. We also investigate the sensitivity of risk metrics like the expected discounted payoff or the Value-at-Risk with respect to different parameters.

2.2. Equity-indexed annuities and financial market

We look at equity-indexed annuities with a single premium P_0 paid at time 0. Their payoff depends on the financial market. To describe the market, we consider a probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$, where $\mathbb{P}$ is the real world measure and $\mathcal{F}_t = \sigma(W_s^{(1)}, W_s^{(2)}, s \leqslant t)$ is the natural filtration for two independent Brownian motions $W_t^{(1)}$ and $W_t^{(2)}$. The stochastic interest rate process follows

$$\mathrm{d}r_t = \kappa\left(\theta^* - r_t\right)\,\mathrm{d}t + \sigma_r\,\mathrm{d}W_t^{(1)}\,, \qquad r_0 \in \mathbb{R}\,, \tag{1}$$

under the real-world measure $\mathbb{P}$ with mean-reversion speed $\kappa \in \mathbb{R}$, long-term mean interest rate level $\theta^* \in \mathbb{R}$ and volatility $\sigma_r > 0$. The financial market contains a risky asset $\{S_t\}_{t\geq 0}$ with dynamics under the real-world measure $\mathbb{P}$

$$\mathrm{d}S_t = S_t\left(\mu_t\,\mathrm{d}t + \sigma_S\left(\rho\mathrm{d}W_t^{(1)} + \sqrt{1-\rho^2}\,\mathrm{d}W_t^{(2)}\right)\right)\,, \qquad S_0 = 1\,, \tag{2}$$

where μ_t is the drift term, $\sigma_S > 0$ the volatility and $\rho \in [-1, 1]$ the correlation parameter.

Starting from the single premium P_0, every year, the policyholder account is continuously compounded by a rate of return. This rate is determined by the maximum of a participation share $\alpha \in (0, 1)$ of the log-returns of the underlying $\ln(S_t/S_{t-1})$ and the guaranteed rate g. The maturity of the contract is $T > 0$. The compounding nature of the cliquet option leads to the time-T payoff to the policyholders

$$Y_T = P_0 \cdot \prod_{t=1}^{T} \max\left(\mathrm{e}^g, \left(\frac{S_t}{S_{t-1}}\right)^{\alpha}\right)\,. \tag{3}$$

We also introduce the risk-free discounted payoff as

$$Z_T = \mathrm{e}^{-\int_0^T r_s\,\mathrm{d}s} \cdot Y_T\,.$$

For the valuation of equity-indexed annuities, we assume that there exist constant risk premiums $\lambda_r, \lambda_S \in \mathbb{R}$ for both the interest rate (1) and asset process (2) (see also Barbarin and Devolder, 2005; Graf et al., 2011; Hieber et al., 2019). More specifically, we suppose $\theta^* = \theta + \frac{\lambda_r \sigma_r}{\kappa}$ for some risk-neutral mean-reversion level $\theta \in \mathbb{R}$ and $\mu_t = r_t + \lambda_S$. Analogously to Graf et al. (2011), this uniquely determines the valuation measure via the change of measure:

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = \mathrm{e}^{-\lambda_r W_t^{(1)} - \frac{1}{2}\lambda_r^2 t - \frac{\lambda_S - \rho\lambda_r\sigma_S}{\sqrt{1-\rho^2}\sigma_S} W_t^{(2)} - \frac{1}{2}\frac{(\lambda_S - \rho\lambda_r\sigma_S)^2}{(1-\rho^2)\sigma_S^2} t}\,. \tag{4}$$

Note that, given this assumption, the short rate process follows a Vasicek model under both $\mathbb{P}$ and $\mathbb{Q}$. Additionally, the dynamics for the risky asset under $\mathbb{Q}$ are as in (2) with drift term μ_t replaced by the interest rate r_t. The long-term mean interest rate under $\mathbb{Q}$ is θ.

To analyse the interest rate risk of equity-indexed annuities, we want to efficiently compute the expected discounted payoff $\mathbb{E}_{\mathbb{Q}}[Z_T]$, the variance of the discounted payoff $\mathrm{Var}_{\mathbb{P}}[Z_T]$ and the probability that the terminal payoff exceeds certain thresholds $G > 0$, that is $\mathbb{P}(Y_T \leq G)$. The latter probability allows us to analyse contracts with an additional maturity guarantee G, that is payoffs of the form $\max(Y_T, G)$. We can also derive Value-at-Risk and downside-risk related quantities of interest.

2.3. Moment-generating function

To study these products more closely, we first derive the moment-generating function of the logarithm of the discounted payoff random variable Z_T. To simplify the notation, define the logarithm of the discounted annual returns $L_{t-1,t}$ as

$$L_{t-1,t} := \ln\left(\mathrm{e}^{-\int_{t-1}^{t} r_s\,\mathrm{d}s} \cdot \max\left(\mathrm{e}^{g},\, \left(\frac{S_t}{S_{t-1}}\right)^{\alpha}\right)\right) = \max\left(g,\, \alpha \cdot \ln\left(\frac{S_t}{S_{t-1}}\right)\right) - \int_{t-1}^{t} r_s\,\mathrm{d}s\,,$$

so that

$$Z_T = P_0 \cdot \prod_{t=1}^{T} \mathrm{e}^{L_{t-1,t}}\,.$$

The moment-generating function $\mathcal{L}(u) := \mathbb{E}\left[\mathrm{e}^{uX}\right]$ of a random variable X allows us to make inferences about its distribution. The moment-generating function of the logarithm of Z_T can be expressed as

$$\mathcal{L}_T(u) = \mathbb{E}\left[\mathrm{e}^{u\cdot\ln(Z_T)}\right] = \mathbb{E}\left[(Z_T)^u\right] = \mathbb{E}\left[(P_0)^u \prod_{t=1}^{T} \mathbb{E}\left[\mathrm{e}^{u\cdot L_{t-1,t}} \,\middle|\, r_{t-1}, r_t\right]\right]\,. \tag{5}$$

Theorem 2.3.1 provides closed-form expressions for the conditional moment-generating function $\phi_{r_{t-1},r_t}(u) := \mathbb{E}\left[\mathrm{e}^{u\cdot L_{t-1,t}} \,|\, r_{t-1}, r_t\right]$ in equation (5).

Theorem 2.3.1 (Conditional moment-generating function). *Assume that the financial market is described by* (1) *and* (2). *Define*

$$g_\kappa(t) := \frac{1-\mathrm{e}^{-\kappa t}}{\kappa}.$$

The moment-generating function of the discounted annual log-returns $L_{t-1,t}$, conditional on r_{t-1}, r_t, is given by

$$\begin{aligned}
\phi_{r_{t-1},r_t}(u) &:= \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{u\cdot L_{t-1,t}} \,\middle|\, r_{t-1}, r_t\right] \\
&= \mathrm{e}^{-u\mu(r_{t-1},r_t)+u^2\frac{\Sigma^2}{2}} \left(\mathrm{e}^{ug}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}(0,u,r_{t-1},r_t)}{\widetilde{\Sigma}}\right)\right. \\
&\qquad \left. + \,\mathrm{e}^{u\widetilde{\mu}(0,u,r_{t-1},r_t)\alpha+\frac{1}{2}u^2\alpha^2\widetilde{\Sigma}^2}\Phi\left(\frac{-\frac{g}{\alpha}+\widetilde{\mu}(0,u,r_{t-1},r_t)+u\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right)\right),
\end{aligned} \tag{6}$$

where Φ describes the cumulative distribution function of a standard normal random variable and

$$\mu(r_{t-1},r_t) = \frac{g_\kappa(1)}{1+\mathrm{e}^{-\kappa}} \cdot (r_{t-1}+r_t) + \left(1-\frac{2g_\kappa(1)}{1+\mathrm{e}^{-\kappa}}\right)\cdot\theta\,,$$

$$\Sigma^2 = \sigma_r^2\left(\frac{1-g_\kappa(1)}{\kappa^2} - \frac{g_\kappa(1)^2}{\kappa(1+\mathrm{e}^{-\kappa})}\right),$$

$$c = \sigma_r^2\left(\frac{1-g_\kappa(1)}{\kappa^2} - \frac{g_\kappa(1)^2}{2\kappa} - \frac{g_\kappa(1)^3}{2(1+e^{-\kappa})}\right) - \frac{\rho\sigma_S\sigma_r}{\kappa}\left(\frac{2g_\kappa(1)}{1+\mathrm{e}^{-\kappa}}-1\right),$$

$$\widetilde{\mu}(\lambda,u,r_{t-1},r_t) = \lambda + \frac{g_\kappa(1)-2\rho\frac{\sigma_S}{\sigma_r}\mathrm{e}^{-\kappa}}{1+\mathrm{e}^{-\kappa}}\cdot r_{t-1} + \frac{g_\kappa(1)+2\rho\frac{\sigma_S}{\sigma_r}}{1+\mathrm{e}^{-\kappa}}\cdot r_t$$

$$+\left(1-2\frac{g_\kappa(1)+\rho\frac{\sigma_S}{\sigma_r}(1-\mathrm{e}^{-\kappa})}{1+\mathrm{e}^{-\kappa}}\right)\cdot\theta-c\cdot u-\frac{1}{2}\sigma_S^2\,,$$

$$\widetilde{\Sigma}^2=c+\sigma_S^2-2\rho^2\sigma_S^2\frac{g_\kappa(1)}{1+\mathrm{e}^{-\kappa}}-\frac{\rho\sigma_S\sigma_r}{\kappa}\left(\frac{2g_\kappa(1)}{1+\mathrm{e}^{-\kappa}}-1\right).$$

Proof. See the Section 2.A. □

Remark 2.3.2. *(a) In (6), we determined the conditional moment-generating function under the risk neutral measure $\mathbb{Q}$. We can also derive an expression under $\mathbb{P}$ since the financial markets follows a Vasicek-Black-Scholes model under both measures. Considering the change of measure as in (4), we only have to set $\lambda=\lambda_S$ and change θ to $\theta^*=\theta+\frac{\lambda_r\sigma_r}{\kappa}$ to obtain the result under $\mathbb{P}$.*
(b) Note that (6) can be modified to calculate the moment-generating function of the undiscounted annual log-returns $L^*_{t-1,t}:=\max(g,\,\alpha\cdot\ln(S_t/S_{t-1}))$. For this derivation, we do not change the measure as in Section 2.A and we thus find

$$\mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{u\cdot L^*_{t-1,t}}\,\middle|\,r_{t-1},r_t\right]=\mathrm{e}^{ug}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}^*(0,r_{t-1},r_t)}{\widetilde{\Sigma}}\right)$$
$$+\,\mathrm{e}^{u\widetilde{\mu}^*(0,r_{t-1},r_t)\alpha+\frac{1}{2}u^2\alpha^2\widetilde{\Sigma}^2}\Phi\left(\frac{-\frac{g}{\alpha}+\widetilde{\mu}^*(0,r_{t-1},r_t)+u\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right),$$

where $\widetilde{\mu}^*(\lambda,r_{t-1},r_t)=\widetilde{\mu}(\lambda,u,r_{t-1},r_t)+c\cdot u$. Therefore, we are also able to describe the distribution of the undiscounted payoff Y_T. This is, for example, needed to include caps on the ultimate payoff Y_T at maturity. However, for simplicity, we continue to derive our approximation for the discounted payoff Z_T.

(c) We are interested in estimating quantiles of the solvency ratio Y_T/S_T since this provides insightful risk measures. We have

$$\mathbb{E}\left[\mathrm{e}^{u\cdot\ln(Y_T/S_T)}\right]=\mathbb{E}\left[\left(\frac{Y_T}{S_T}\right)^u\right]=\mathbb{E}\left[\prod_{t=1}^{T}\mathbb{E}\left[\mathrm{e}^{u\cdot L^*_{t-1,t}-u\cdot\ln\left(\frac{S_t}{S_{t-1}}\right)}\,\middle|\,r_{t-1},r_t\right]\right].$$

It is straightforward to derive from the proof in Section 2.A that

$$\mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{u\cdot L^*_{t-1,t}-u\cdot\ln\left(\frac{S_t}{S_{t-1}}\right)}\,\middle|\,r_{t-1},r_t\right]=\mathrm{e}^{ug-u\widetilde{\mu}^*(0,r_{t-1},r_t)+\frac{1}{2}u^2\widetilde{\Sigma}^2}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}^*(0,r_{t-1},r_t)+u\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right)$$
$$+\,\mathrm{e}^{u\widetilde{\mu}^*(0,r_{t-1},r_t)(\alpha-1)+\frac{1}{2}u^2(\alpha-1)^2\widetilde{\Sigma}^2}\Phi\left(\frac{-\frac{g}{\alpha}+\widetilde{\mu}^*(0,r_{t-1},r_t)+u(\alpha-1)\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right).$$

2.3.1. Approximation on a grid. Given the conditional moment-generating function (6), we still need to determine the outer expectation in (5), that is to integrate over the annual interest rates $(r_1, r_2, \ldots, r_T)$. We approximate this integral by choosing discrete grid points $-\infty<r^{(1)}<\cdots<r^{(K+1)}=\infty$, $K\in\mathbb{N}$, for the annual interest rates $r_1,\ldots,r_T$. We further set $r^{(0)}=-\infty$ and $r^{(K+1)}=\infty$. We want to choose a grid that contains r_0 and set $r^{(i_0)}=r_0$. A possibility on how to choose this grid is Example 2.3.3.

Example 2.3.3 (Choice of discretisation)**.** In a Vasicek model, the interest rate r_t is normally distributed with variance $\hat{\sigma}_t^2=\frac{\sigma_r^2}{2\kappa}(1-\mathrm{e}^{-2\kappa t})$. We may just choose an equidistant grid around r_0 whose upper and lower bounds exceed the 99.9% and 0.01% quantiles of r_T, for example $r^{(j)}=r_0+\frac{8(j-\frac{K}{2})}{K}\hat{\sigma}_T$, for $j=1,2,\ldots,K$ and uneven K. In this case, $i_0=(K+1)/2$. Note, that centering around r_0 in this way is only reasonable if it is not too far from the long-term mean θ.

We denote the interval around $r^{(j)}$, $1 \leq j \leq K$, by

$$\left[r^{(j)}\right] := \left(\frac{r^{(j-1)} + r^{(j)}}{2}, \frac{r^{(j)} + r^{(j+1)}}{2}\right]. \tag{7}$$

We proceed as follows: Using the same grid for each rate $r_1, \ldots, r_T$ allows us to speed up calculations significantly. For subsequent rates (r_{t-1}, r_t), we apply a two-dimensional trapezoidal rule based on rectangles $r_{t-1} \times r_t = \left[r^{(i)}\right] \times \left[r^{(j)}\right]$, $i, j = 1, 2, \ldots, K$. On each of these rectangles, we assume that the conditional moment-generating function is constant, approximated by $\phi_{r^{(i)}, r^{(j)}}(u)$, that is:

$$\begin{aligned} &\mathbb{E}\left[\mathrm{e}^{u \cdot L_{t-1,t}} \,\middle|\, r_{t-1} \in \left[r^{(i)}\right],\, r_t \in \left[r^{(j)}\right]\right] \\ &\qquad \approx \mathbb{E}\left[\mathrm{e}^{u \cdot L_{t-1,t}} \,\middle|\, r_{t-1} = r^{(i)},\, r_t = r^{(j)}\right] = \phi_{r^{(i)}, r^{(j)}}(u)\,. \end{aligned} \tag{8}$$

The probability mass over the rectangle $r_{t-1} \times r_t \in \left[r^{(i)}\right] \times \left[r^{(j)}\right]$ is approximately the same as the integrated density over $r_{t-1} \times r_t \in r^{(i)} \times \left[r^{(j)}\right]$ and thus we find the probabilities

$$\begin{aligned} \mathbb{P}\Big(r_t \in \left[r^{(j)}\right] \,\Big|\, r_{t-1} \in \left[r^{(i)}\right]\Big) &\approx \mathbb{P}\left(\frac{r^{(j-1)} + r^{(j)}}{2} < r_t \leq \frac{r^{(j)} + r^{(j+1)}}{2} \,\middle|\, r_{t-1} = r^{(i)}\right) \\ &= \Phi\left(\frac{\frac{r^{(j)} + r^{(j+1)}}{2} - \mu_i}{\hat{\sigma}_1}\right) - \Phi\left(\frac{\frac{r^{(j-1)} + r^{(j)}}{2} - \mu_i}{\hat{\sigma}_1}\right) =: p_{ij}\,, \end{aligned} \tag{9}$$

for any combination $i \times j \in \{1, 2, \ldots, K\} \times \{1, 2, \ldots, K\}$, where $\mu_i = \mathrm{e}^{-\kappa} \cdot r^{(i)} + \theta(1 - \mathrm{e}^{-\kappa})$ and $\hat{\sigma}_1^2 = \frac{\sigma_r^2}{2\kappa}(1 - \mathrm{e}^{-2\kappa})$. These probabilities together with the conditional moment-generating functions define the scenario matrix $\big(\mathcal{Q}(u)\big)_{i,j} \approx \big(\phi_{r^{(i)}, r^{(j)}}(u) \cdot p_{ij}\big)_{i,j}$. The scenario probabilities p_{ij} in each row of this matrix sum up to 1, that is $\sum_{i=1}^{K} p_{ij} = 1$. The moment-generating function of the logarithm of the discounted payoff variable Z_T can be expressed in terms of this matrix, see Theorem 2.3.4.

Theorem 2.3.4 (Moment-generating function $\mathcal{L}_T(u)$). *We approximate the moment-generating function $\mathcal{L}_T(u)$ of the logarithm of the time-0 value of the payoff Y_T by $\hat{\mathcal{L}}_T(u)$ through a matrix iteration*

$$\mathcal{L}_T(u) = (P_0)^u \cdot \mathbf{1}_{i_0}' \, \mathcal{Q}(u)^T \mathbf{1} \approx (P_0)^u \cdot \mathbf{1}_{i_0}' \, \hat{\mathcal{Q}}(u)^T \mathbf{1} = \hat{\mathcal{L}}_T(u)\,, \tag{10}$$

where $\mathbf{1} = (1, \ldots, 1)'$ and $'$ denotes the transposition of a vector. $\mathbf{1}_{i_0}$ is a unit vector that has an entry of 1 in the i_0-th position. For a matrix $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{K \times K}$, note that $\big(\boldsymbol{\mathcal{Q}}\big)^T := \boldsymbol{\mathcal{Q}} \cdot \boldsymbol{\mathcal{Q}} \cdots \boldsymbol{\mathcal{Q}}$. The scenario matrix $\mathcal{Q}(u)$ has entries

$$\big(\mathcal{Q}(u)\big)_{i,j} = \mathbb{E}\left[\mathrm{e}^{u \cdot L_{t-1,t}} \,\middle|\, r_{t-1} \in \left[r^{(i)}\right],\, r_t \in \left[r^{(j)}\right]\right] \cdot \mathbb{P}\Big(r_t \in \left[r^{(j)}\right] \,\Big|\, r_{t-1} \in \left[r^{(i)}\right]\Big)\,. \tag{11}$$

Using (8), (9), these entries can be approximated as $\big(\mathcal{Q}(u)\big)_{i,j} \approx \big(\hat{\mathcal{Q}}(u)\big)_{i,j} = \phi_{r^{(i)}, r^{(j)}}(u) \cdot p_{ij}$, where p_{ij} is as in (9) and $\phi_{r^{(i)}, r^{(j)}}(u)$ was defined in Theorem 2.3.1.

Proof. We use the fact that, conditional on the annual interest rates $r_1, \ldots, r_T$, the annual increments of our underlying are independent. Adapting a conditioning idea from Hieber

(2017) for regime switching models to our setting yields

$$\begin{aligned}
\hat{\mathcal{L}}_T(u) =&\ (P_0)^u \cdot \mathbb{E}\left[\mathrm{e}^{-u\cdot\int_0^T r_s\,\mathrm{d}s}\mathrm{e}^{u\cdot\sum_{t=1}^T \max\left(g,\,\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)\right)}\right]\\
=&\ (P_0)^u \cdot \mathbb{E}\left[\mathbb{E}\left[\prod_{t=1}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_1\right]\right]\\
=&\ (P_0)^u \cdot \sum_{i=1}^K \mathbb{E}\left[\mathbb{1}_{\left\{r_1\in\left[r^{(i)}\right]\right\}}\mathrm{e}^{u\cdot L_{0,1}}\right]\mathbb{E}\left[\prod_{t=2}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_1\in\left[r^{(i)}\right]\right]\\
=&\ (P_0)^u \cdot \left(\mathbb{E}\left[\mathbb{1}_{\left\{r_1\in\left[r^{(1)}\right]\right\}}\mathrm{e}^{u\cdot L_{0,1}}\right]\cdots\mathbb{E}\left[\mathbb{1}_{\left\{r_1\in\left[r^{(K)}\right]\right\}}\mathrm{e}^{u\cdot L_{t-1,t}})\right]\right)\\
&\cdot\begin{pmatrix}\mathbb{E}\left[\prod_{t=2}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_1\in\left[r^{(1)}\right]\right]\\ \vdots\\ \mathbb{E}\left[\prod_{t=2}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_1\in\left[r^{(K)}\right]\right]\end{pmatrix}\\
=&\ (P_0)^u \cdot \mathbf{1}_{i_0}' \cdot \mathcal{Q}(u)\cdot\\
&\begin{pmatrix}\sum_{j=1}^K \mathbb{E}\left[\mathbb{1}_{\left\{r_2\in\left[r^{(j)}\right]\right\}}\mathrm{e}^{u\cdot L_{1,2}}\,\middle|\, r_1\in\left[r^{(1)}\right]\right]\mathbb{E}\left[\prod_{t=3}^T \mathrm{e}^{u\, L_{t-1,t}}\,\middle|\, r_2\in\left[r^{(j)}\right]\right]\\ \vdots\\ \sum_{j=1}^K \mathbb{E}\left[\mathbb{1}_{\left\{r_2\in\left[r^{(j)}\right]\right\}}\mathrm{e}^{u\cdot L_{1,2}}\,\middle|\, r_1\in\left[r^{(K)}\right]\right]\mathbb{E}\left[\prod_{t=3}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_2\in\left[r^{(j)}\right]\right]\end{pmatrix}\\
=&\ (P_0)^u \cdot \mathbf{1}_{i_0}'\big(\mathcal{Q}(u)\big)^2\cdot\begin{pmatrix}\mathbb{E}\left[\prod_{t=3}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_2\in\left[r^{(1)}\right]\right]\\ \vdots\\ \mathbb{E}\left[\prod_{t=3}^T \mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_2\in\left[r^{(K)}\right]\right]\end{pmatrix}\\
=&\ \cdots = (P_0)^u\cdot\mathbf{1}_{i_0}'\,\mathcal{Q}(u)^T\mathbf{1}\,,
\end{aligned}$$

The entries of $\mathcal{Q}(u)$ are

$$\big(\mathcal{Q}(u)\big)_{i,j} = \mathbb{E}\left[\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_{t-1}\in\left[r^{(i)}\right],\, r_t\in\left[r^{(j)}\right]\right]\cdot\mathbb{P}\left(r_t\in\left[r^{(j)}\right]\,\middle|\, r_{t-1}\in\left[r^{(i)}\right]\right).$$

Note that these do not depend on the time t since, conditioned on the same preceding and succeeding interest rates $r^{(i)}$ and $r^{(j)}$, the annual returns and discount factors are identically distributed. □

Remark 2.3.5. *In practice, we avoid computing T matrix-matrix products in* (10) *by iteratively computing T matrix-vector products $\hat{\mathcal{Q}}(u)^T\mathbf{1} = \hat{\mathcal{Q}}(u)(\hat{\mathcal{Q}}(u)^{T-1}\mathbf{1})$. Therefore, our algorithm implicitly computes the moment-generating function for the same policy with all maturities in one run. This allows to incorporate mortality and cancellation probabilties that are independent of the financial market prior to maturity T.*

Letting $K \to \infty$ and thus reducing the size of the intervals, our approximation of the interest rates converges weakly to the true process, see Theorem 2.3.6. In Section 2.5, we discuss the choice of the number of gridpoints K.

Theorem 2.3.6 (Weak convergence). *Let $(\bar{r}_t^K)_{t\in\mathbb{N}}$ be a discrete time process that uses the transition probabilities described in* (9) *and let the grid points $r^{(j)}$ be chosen so that $\mathbb{P}(r_1 \in [r^{(j)}]) \in$*

$\mathcal{O}(1/K)$, $j = 1, \dots, K$. Then, for all $t \in \mathbb{N}$, $\bar{r}_t$ converges weakly to the continuous process r_t in (1).

Proof. See the Section 2.B. □

2.3.2. Quantile and tail probability estimation. We are interested in quantiles of the undiscounted payoff variable Y_T. For our cliquet option with maturity T and a constant $G \in \mathbb{R}$, we have:

$$P_T(G) := \mathbb{P}\big(\ln(Y_T) > G\big) = \int_G^\infty p_T(s)\mathrm{d}s\,,$$

where p_T denotes the density of the logarithm of Y_T. To compute these tail probabilities given the moment-generating function of the logarithm of Y_T, we apply the fast Fourier algorithm, see, for example, Carr and Madan (1999).

We introduce a dampening coefficient $\gamma \in (0, \infty)$ to define a dampened probability $d_T(G)$ and its Fourier transformation $\psi_T(v)$ as

$$d_T(G) := \mathrm{e}^{\gamma G} P_T(G), \qquad \psi_T(v) = \int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}vG} d_T(G)\,\mathrm{d}G\,.$$

Then, we can relate these two quantities via

$$P_T(G) = \frac{\mathrm{e}^{-\gamma G}}{2\pi}\int_{-\infty}^\infty \mathrm{e}^{-\mathrm{i}vG}\psi_T(v)\,\mathrm{d}v = \frac{\mathrm{e}^{-\gamma G}}{\pi}\int_0^\infty \mathrm{e}^{-\mathrm{i}vG}\psi_T(v)\,\mathrm{d}v\,, \tag{12}$$

where we can express

$$\begin{aligned}\psi_T(v) &= \int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}vG}\int_G^\infty \mathrm{e}^{\gamma G} p_T(s)\,\mathrm{d}s\,\mathrm{d}G = \int_{-\infty}^\infty p_T(s)\int_{-\infty}^s \mathrm{e}^{(\mathrm{i}v+\gamma)G}\,\mathrm{d}G\,\mathrm{d}s \\ &= \int_{-\infty}^\infty p_T(s)\frac{\mathrm{e}^{(\mathrm{i}v+\gamma)s}}{\gamma+\mathrm{i}v}\,\mathrm{d}s = \frac{\mathcal{L}_T(\gamma+\mathrm{i}v)}{\gamma+\mathrm{i}v}\,.\end{aligned}$$

It is possible to simultaneously calculate these probabilities for various thresholds G_k via fast Fourier transformation.
To this end, let $v_j = \eta(j-1)$ and $G_k = b+\omega(k-1)$ for $\eta, \omega > 0$ and $j, k = 1, \dots, N$, $N \in \mathbb{N}$, such that $\omega\eta = \frac{2\pi}{N}$. A standard Simpson's rule applied to (12) gives

$$\begin{aligned}P_T(G_k) &\approx \frac{\mathrm{e}^{-\gamma G_k}}{\pi}\sum_{j=1}^N \mathrm{e}^{-\mathrm{i}v_j G_k}\psi(v_j)\cdot\frac{\eta}{3}\Big(3+(-1)^j-\delta_{j-1}\Big) \\ &= \frac{\mathrm{e}^{-\gamma G_k}}{\pi}\sum_{j=1}^N \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}(j-1)(u-1)}\mathrm{e}^{-\mathrm{i}bv_j}\frac{\mathcal{L}_T(\gamma+\mathrm{i}v_j)}{\gamma+\mathrm{i}v_j}\cdot\frac{\eta}{3}\Big(3+(-1)^j-\delta_{j-1}\Big)\,,\end{aligned}$$

where δ_n is equal to 1 for $n = 0$ and zero otherwise. This immediately allows to apply fast Fourier transformation.

An application of this technique for calculating exceedance probabilities is to consider a capped index participation that limits the maximum reward for the policyholder at time T at some upper barrier G_k.

2.4. Benchmarks: Markov chain approximation, Fourier pricing and PROJ

Different from our approach, several authors approximate the interest rate process by a Markov chain. Using the discrete grid introduced earlier, this means that the interest rate can only take discrete values $r^{(1)}$, $r^{(2)}$, $\dots$, $r^{(K)}$, summarised in the vector

$\boldsymbol{r} := (r^{(1)}, r^{(2)}, \dots, r^{(K)})' \in \mathbb{R}^K$. The Markov chain approximation $(\check{r}_t)_{t\geq 0}$ is defined by a suitable choice of the intensity matrix $\boldsymbol{Q} = (q_{k,j})_{k,j}$ that defines the transition between the K states. For $k, j = 1, 2, \dots, K$, an example is

$$q_{k,j} = \begin{cases} \frac{\sigma_r^2 - \kappa(\theta - r^{(k)})c_k}{c_{k-1}(c_{k-1}+c_k)} & \text{if } j = k-1\,, \\ \frac{\sigma_r^2 + \kappa(\theta - r^{(k)})c_{k-1}}{c_k(c_{k-1}+c_k)} & \text{if } j = k+1\,, \\ -q_{k,k-1} - q_{k,k+1} & \text{if } j = k\,, \end{cases} \tag{13}$$

where $c_k := r^{(k+1)} - r^{(k)}$. The approximation (13) was introduced by Lo and Skindilias (2014) as a generalisation of Kushner and Dupuis (2001) and Chourdakis (2004). It is designed to match the first two moments of the original interest rate process. However, this formula can only be applied for sufficiently small c_k that is for large enough K. Otherwise, if we get negative entries off the main-diagonal of $\boldsymbol{Q}$, we can resort to an approximation by Piccioni (1987) which only matches the first moment of the original process.

This approximation of the interest rate process is fundamentally different from our approach, since $\check{r}_t$ does not behave like a Vasicek model within one year. This is in contrast to the SM method where the interest rate process between consecutive annual time points follows a Vasicek model. A visualisation of the two approximations is given in Fig. 1 for $K = 8$ and $K = 87$.

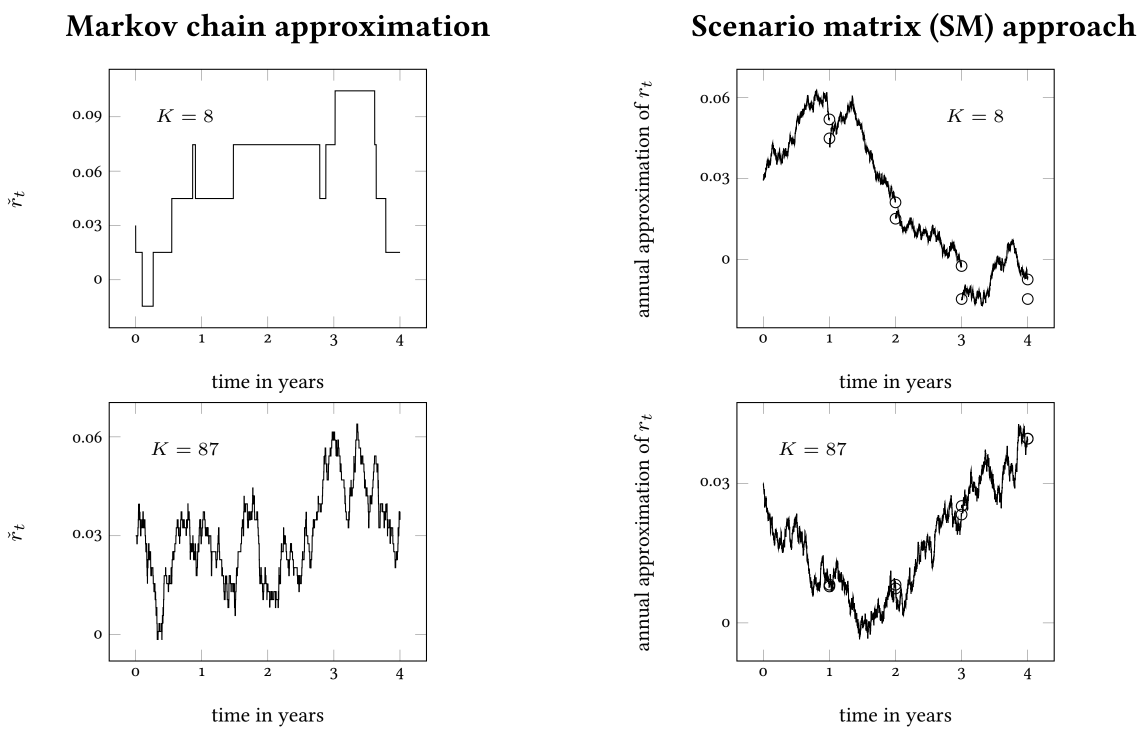


FIGURE 1. Sample paths of Markov chain approximations $\check{r}$ (left column) and annual rounding to the nearest grid point as in the SM method (right column) for $K = 8$ (first row) and $K = 87$ grid points (second row). The parameters have been chosen as in Table 1.

2.4.1. Fourier pricing. Given the intensity matrix Q, the discounted characteristic function $\boldsymbol{\varphi}_t(u,\boldsymbol{r}) := \{\varphi_t^{(k,j)}(u,\boldsymbol{r})\}_{k,j}$ is given explicitly in terms of a matrix exponential:

$$\begin{aligned}\varphi_t^{(k,j)}(u) &:= \mathbb{E}\Big[\mathbb{1}_{\{\check{r}_t=r^{(j)}\}}\,\mathrm{e}^{-\int_0^t \check{r}_s\,\mathrm{d}s+\mathrm{i}u\ln(S_t/S_0)}\,\Big|\,r_0=r^{(k)}\Big]\\ &= \mathbf{1}_k'\,\exp\Big(\boldsymbol{Q}t+\mathrm{diag}\Big(-\boldsymbol{r}+\mathrm{i}u\Big(\boldsymbol{r}+\lambda-\frac{1}{2}\sigma_S^2\Big)-\frac{1}{2}\sigma_S^2u^2\Big)t\Big)\,\mathbf{1}_j\,. \qquad (14)\end{aligned}$$

Given (14), let us introduce the matrices:

$$\big\{\boldsymbol{\mathcal{D}}\big\}_{k,j=1,2,\dots,n} := \mathbb{E}\Big[\mathbb{1}_{\{\check{r}_t=r^{(j)}\}}\mathrm{e}^{-\int_{t-1}^t \check{r}_s\,\mathrm{d}s}\,\Big|\,r_{t-1}=r^{(k)}\Big] = \exp\Big(\boldsymbol{Q}-\mathrm{diag}(\boldsymbol{r})\Big)\,,$$

$$\begin{aligned}\big\{\boldsymbol{\mathcal{C}}(b)\big\}_{k,j=1,2,\dots,n} &:= \mathbb{E}\Big[\mathbb{1}_{\{\check{r}_t=r^{(j)}\}}\mathrm{e}^{-\int_{t-1}^t \check{r}_s\,\mathrm{d}s}\Big[\big(S_T/S_0\big)^\alpha-\mathrm{e}^g\Big]^+\,\Big|\,r_{t-1}=r^{(k)}\Big]\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\boldsymbol{\varphi}_T\big(v-\mathrm{i}\delta\big)\cdot\hat{f}_g\big(\mathrm{i}\delta-v\big)\,\mathrm{d}v\,, \qquad (15)\end{aligned}$$

where $\delta\in(0,\infty)$ is a dampening coefficient. To compute the expectation, we derive

$$\hat{f}_g(u) := \int_{\mathbb{R}}\mathrm{e}^{iuy}\Big[\mathrm{e}^{\alpha y}-\mathrm{e}^g\Big]^+\,\mathrm{d}y = \int_{\frac{g}{\alpha}}^{\infty}\mathrm{e}^{(\alpha+\mathrm{i}u)y}\,\mathrm{d}y-\mathrm{e}^g\int_{\frac{g}{\alpha}}^{\infty}\mathrm{e}^{iuy}\,\mathrm{d}y = \mathrm{e}^{g(1+\frac{\mathrm{i}u}{\alpha})}\frac{\alpha}{\mathrm{i}u(\alpha+\mathrm{i}u)}\,.$$

From Hieber (2017), we then obtain:

$$\mathbb{E}[Z_T] = P_0\cdot\mathbf{1}_k'\Big(\mathrm{e}^g\boldsymbol{\mathcal{D}}(1)+\boldsymbol{\mathcal{C}}(g)\Big)^T\,\mathbf{1}\,. \qquad (16)$$

This requires to compute the Fourier integral (15).

2.4.2. PROJ method. An alternative to the SM method is the projection method (PROJ) introduced by Kirkby (2015), Cui et al. (2017). The method approximates numerical integrals using a B-spline basis, is very general and has a variety of applications. We apply PROJ to compute the expectation:

$$\mathbb{E}\Big[\mathrm{e}^{L_{t-1,t}}\,\Big|\,\check{r}_{t-1}=r^{(k)},\,\check{r}_t=r^{(j)}\Big] = \int_0^\infty h_{k,j}(y)\cdot f(y)\,\mathrm{d}y\,, \qquad (17)$$

where $f(y)$ is the density of the integrated interest rate $y=\int_{t-1}^t\check{r}_s\,\mathrm{d}s$ and $h_{k,j}(y)$ the expectation given $y=\int_{t-1}^t\check{r}_s\,\mathrm{d}s$. PROJ projects the density $f(y)$ on a basis of transformed hat functions $\varphi^{[1]}(y)=(1-|x|)\mathbb{1}_{[-1,1]}(y)$. More specifically, PROJ uses the linear B-spline basis $\{\varphi_{a,n}(y)\}_{n=1,2,\dots,N} := \big\{\sqrt{a}\cdot\varphi^{[1]}\big(a(x-x_n)\big)\big\}_{n=1,2,\dots,N}$ with resolution $a>0$ and grid points $x_n=x_1+(n-1)/a$, $n=1,2,\dots,N$.
The coefficients $\beta_{a,n}$ of the approximation $f(y)\approx\sum_{n=1}^N\beta_{a,n}\cdot\varphi_{a,n}(y)$ by $N\in\mathbb{N}$ basis functions are available as an integral

$$\beta_{a,n} = \big\langle f(y)\,,\varphi_{a,n}(y)\big\rangle = \frac{12a^{\frac{3}{2}}}{\pi}\int_0^\infty\hat{f}(\xi)\cdot\mathrm{e}^{-\mathrm{i}x_n\xi}\cdot\frac{\sin^2\big(\frac{\xi}{2a}\big)}{\xi^2\big(2+\cos\frac{\xi}{a}\big)}\,\mathrm{d}\xi\,, \qquad (18)$$

see Kirkby (2015), Cui et al. (2017) or the Section 2.C for details. The integral (18) can efficiently be solved by discrete Fourier transformation (DFT). From (14), we obtain the Fourier transform $\widehat{f}(\xi):=\int_{-\infty}^{\infty}\mathrm{e}^{-i\xi y}f(y)\,\mathrm{d}y$ of the integrated interest rate process $y=\int_{t-1}^t\check{r}_s\,\mathrm{d}s$:

$$\begin{aligned}\hat{f}(\xi) &= \mathbb{E}\Big[\mathrm{e}^{-\mathrm{i}\xi\int_{t-1}^t\check{r}_s\,\mathrm{d}s}\,\Big|\,\check{r}_{t-1}=r^{(k)}\,\check{r}_t=r^{(j)}\Big]\\ &= \frac{\mathbb{E}\big[\mathbb{1}_{\{\check{r}_t=r^{(j)}\}}\,\mathrm{e}^{-\mathrm{i}\xi\int_{t-1}^t\check{r}_s\,\mathrm{d}s}\,\big|\,\check{r}_{t-1}=r^{(k)}\big]}{\mathbb{P}\big(\check{r}_t=r^{(j)}\,\big|\,\check{r}_{t-1}=r^{(k)}\big)} = \frac{\mathbf{1}_k'\,\exp\big(\boldsymbol{Q}-\mathrm{i}\xi\,\mathrm{diag}(\boldsymbol{r})\big)\,\mathbf{1}_j}{\mathbf{1}_k'\,\exp\big(\boldsymbol{Q}\big)\,\mathbf{1}_j}\,, \qquad (19)\end{aligned}$$

with $\boldsymbol{r} := (r^{(1)}, r^{(2)}, \dots, r^{(K)})$ and $\boldsymbol{Q} = (q_{k,j})_{k,j}$ as in (13). We finally obtain:

$$\int_0^\infty h_{k,j}(y) \cdot f(y)\,\mathrm{d}y \approx \sum_{n=1}^N \beta_{a,n} \cdot \int_{x_{n-1}}^{x_n} h_{k,j}(y) \cdot \varphi_{a,n}(y)\,\mathrm{d}y =: \sum_{n=1}^N \beta_{a,n} \cdot \theta_{a,n}^{k,j}\,.$$

For a small resolution a, the terms $\theta_{a,n}^{k,j}$ are computed by a composite Simpson's rule:

$$\theta_{a,n}^{k,j} \approx \frac{\sqrt{a}}{3}\Big(h_{k,j}(x_n - a/2) + h_{k,j}(x_n) + h_{k,j}(x_n + a/2)\Big)\,. \tag{20}$$

In Section 2.C, we derive the function $h_{k,j}$ explicitly as:

$$\begin{aligned} h_{k,j}(y) &:= \mathbb{E}\Big[e^{L_{t-1,t}}\,\Big|\,\check{r}_{t-1} = r^{(k)},\ \check{r}_t = r^{(j)},\ \int_{t-1}^t \check{r}_s\,\mathrm{d}s = y\Big] \\ &= \mathrm{e}^{g-y}\Phi\left(\frac{\frac{g}{\alpha} - \mu_{k,j}(y)}{\sigma_{k,j}}\right) + \mathrm{e}^{\alpha\mu_{k,j}(y) + \frac{\alpha\sigma_{k,j}^2}{2} - y}\Phi\left(\frac{\alpha\sigma_{k,j}^2 + \mu_{k,j}(y) - \frac{g}{\alpha}}{\sigma_{k,j}}\right)\,, \end{aligned} \tag{21}$$

where $\mu_{k,j}(y) = \kappa\frac{\sigma_S}{\sigma_r}\big(r^{(k)} - r^{(j)}\big) + y\big(1 - \rho\frac{\sigma_S}{\sigma_r}\theta\big) - \big(\rho\frac{\sigma_S}{\sigma_r}\theta + \frac{\sigma_s^2}{2}\big)$ and $\sigma_{k,j}^2 = \sigma_S^2(1-\rho^2)$.

2.5. Numerical results

For numerical experiments, we implemented the methods described in Sections 2.3 and 2.4 in R on a personal computer with 16 GB RAM and 11th Gen Intel(R) Core(TM) i7-1185G7 @ 3.00GHz. Generally, we looked at a contract maturing after $T = 25$ years. All parameters for this implementation are listed in Table 1. They have been chosen in accordance with the existing literature (e.g. Barbarin and Devolder, 2005; Graf et al., 2011; Hieber et al., 2019). However, we adjusted the initial interest rate to be in line with the EURIBOR 1-week rates in May 2023. We also reduced the volatility σ_S to account for the fact that the investment strategy of the insurer will usually focus on less risky investments than the overall stock market. Furthermore, we chose the participation rate α by creating a fair policy assuming a constant interest rate equal to the long-term mean θ. That is, it is designed by the premium equivalence $\mathbb{E}_{\mathbb{Q}}[Z_T] = P_0$ that is assumed to be true for a constant interest rate scenario $r_t = \theta = r_0$ for all t.

contract	initial capital	P_0	1
	guarantee level	g	0.015
	participation rate	α	0.422
	maturity in years	T	25
financial market	initial interest rate	r_0	0.03
	long-term mean interest rate level	θ	0.03
	mean reversion speed	κ	0.3
	volatility of the interest rate process	σ_r	0.015
	constant risk premium of interest rate	λ_r	−0.23
	correlation parameter	ρ	0.15
	volatility of the asset process	σ_S	0.1
	constant risk premium of asset process	λ_S	0.03

TABLE 1. Choice of parameters for the contract and the financial market.

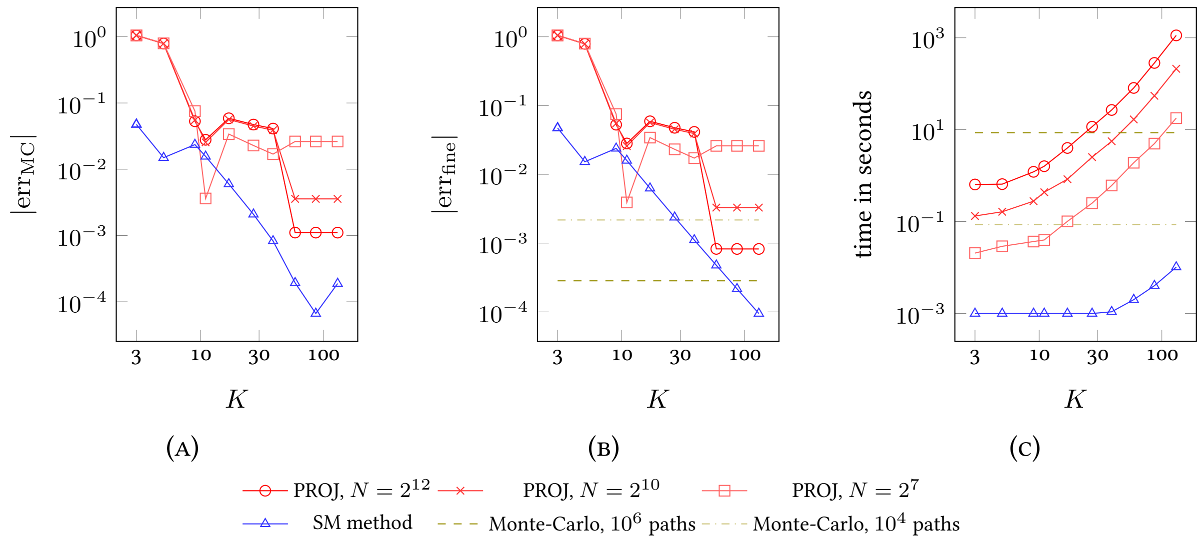


FIGURE 2. Comparison of absolute error and computation time between the SM method, the PROJ method for different values of N, and Monte-Carlo simulations. The parameters have been chosen as in Table 1.
Panel (A): absolute error compared to the Monte-Carlo simulation with 10^6 scenarios in relation to the number of discretisation points K. Panel (B): absolute error of all methods compared to true solution (obtained by a fine approximation with $K = 10\,001$ in the SM method). Panel (C): time spent on calculations for all methods.

2.5.1. Convergence. The left image in Fig. 2 shows the error of the PROJ method and the SM method in relation to a Monte-Carlo-simulation with 10^6 scenarios. We can see that all estimations reach a plateau after using 50 to 100 discretisation points for the interest rate. At this point, our approach entered the 95% confidence interval of the Monte-Carlo simulations and thus outperforms its accuracy afterwards. The middle image shows the absolute error of all methods for different amounts of grid points in comparison to a very fine approximation by our method with $10\,001$ points. The estimations with the SM method almost perfectly exhibit a quadratic convergence. Note that the sudden drop for the SM method at $K = 5$ and for the PROJ method at $K = 11$ are random. In these instances, the grid approximation is by chance very close to the true solution. We can also see that the precision of the PROJ method is not only limited by the number of discretisation points K, but also by the number of terms N used in the fast Fourier transform, since for $N = 2^7$ and $N = 2^{10}$, it is not as close to the finest approximation as a Monte-Carlo simulation with 10^4 paths.

For $K = 87$ grid points, the SM method first outperforms a Monte-Carlo simulation with 10^6 scenarios and has a relative error of 0.013%. The calculations for a maturity of 25 years took approximately 0.004 seconds. Generally, we can see on the right side of Fig. 2 that both methods have a complexity of $\mathcal{O}(K^2)$, while the computation of every matrix entry in the PROJ method requires N times as many calculations for determining the elements of the fast Fourier algorithm compared to one calculation in the SM method.

2.5.2. Analysis of interest rate risk. Expected discounted payoffs for different levels of the mean reversion speed κ, the interest rate volatility σ_r and the correlation parameter ρ are shown in Tables 2 to 4. Note that a Vasicek model with $\sigma_r = 0$ and $\theta = r_0$ is equivalent to assuming a constant interest rate r_0. These values are computed under the risk-neutral measure $\mathbb{Q}$. Low values of κ increase the likelihood for longer periods of interest rates deviating from

σ_r	0	0.005	0.010	0.015	0.020	0.025	0.030
$\mathbb{E}_{\mathbb{Q}}(Z_T)$	0.999	1.002	1.011	1.024	1.043	1.068	1.098
$\mathrm{Q}_{99\%}(Y_T/S_T)$	1.586	1.757	1.989	2.291	2.691	3.207	3.864
$\mathbb{P}(Y_T > S_T)$	11.4%	15.7%	21.1%	27.3%	33.7%	39.9%	45.5%

Table 2. Expected discounted payoff, 99%-quantile of the ratio between the terminal payoff and the value of the underlying at maturity $\mathrm{Q}_{99\%}(Y_T/S_T)$ and the probability of being underfunded, that is $\mathbb{P}(Y_T > S_T)$, for different values of σ_r. The remaining parameters are chosen as in Table 1.

κ	0.10	0.25	0.40	0.55	0.70	0.85	1.00
$\mathbb{E}_{\mathbb{Q}}(Z_T)$	1.108	1.031	1.016	1.011	1.008	1.006	1.005
$\mathrm{Q}_{99\%}(Y_T/S_T)$	4.277	2.478	2.083	1.929	1.846	1.795	1.761
$\mathbb{P}(Y_T > S_T)$	47.4%	30.3%	23.2%	19.8%	17.9%	16.7%	15.8%

Table 3. Expected discounted payoff, 99%-quantile of the ratio between the terminal payoff and the value of the underlying at maturity $\mathrm{Q}_{99\%}(Y_T/S_T)$ and the probability of being underfunded, that is $\mathbb{P}(Y_T > S_T)$, for different values of κ. The remaining parameters are chosen as in Table 1.

ρ	−0.9	−0.6	−0.3	0.0	0.3	0.6	0.9
$\mathbb{E}_{\mathbb{Q}}(Z_T)$	1.011	1.015	1.019	1.022	1.026	1.029	1.032
$\mathrm{Q}_{99\%}(Y_T/S_T)$	1.376	1.640	1.897	2.160	2.430	2.714	3.009
$\mathbb{P}(Y_T > S_T)$	12.7%	18.7%	22.9%	26.0%	28.4%	30.3%	31.9%

Table 4. Expected discounted payoff, 99%-quantile of the ratio between the terminal payoff and the value of the underlying at maturity $\mathrm{Q}_{99\%}(Y_T/S_T)$ and the probability of being underfunded, that is $\mathbb{P}(Y_T > S_T)$, for different values of ρ. The remaining parameters are chosen as in Table 1.

the long-term mean θ. Particularly low interest rate periods reduce the discounting effect while the payoff is still bounded from below by the guaranteed level, whereas the stronger discounting effect of high interest rate periods is partially compensated by larger returns from the underlying fund. For all values, this effect is even more pronounced on the 99%-quantile of the ratio between the undiscounted payoff Y_T and the investment in the underlying S_T since these predominantly include extreme scenarios with long low-interest rate periods. This means, for example, that for an extreme value of $\kappa = 0.1$, the investment in the underlying covers less than $\frac{1}{4.277} \approx 24\%$ of the guaranteed payoff to the policyholder. The quantiles, that is $\mathrm{Q}_{99\%}(Y_T/S_T)$, are determined under the real-world measure $\mathbb{P}$. Also, we can see that the probability for $Y_T > S_T$, i.e. of being underfunded, is affected considerably.

A similar effect can be identified on varying values of σ_r where an increased volatility results in larger deviations from the long-term mean θ and the effects of low and high interest rate periods are the same as described before. However, increasing the volatility does not affect the length of these periods, but their gap to the long-term mean.

When we vary the correlation parameter ρ, the effect is less distinct, but still considerable. Note, however, that values of $\rho = \pm 0.9$ are extreme choices for the correlation parameter. We can observe that negative correlations lead to less risky contracts and lower prices. This is due to the fact that negative values for the correlation parameter ρ result in an overall lower volatility of the underlying asset process in (2) and vice-versa. Thus, we can conclude that modelling interest rates stochastically strongly influences the expected disounted payoff and will usually reveal a higher value than constant interest rates.

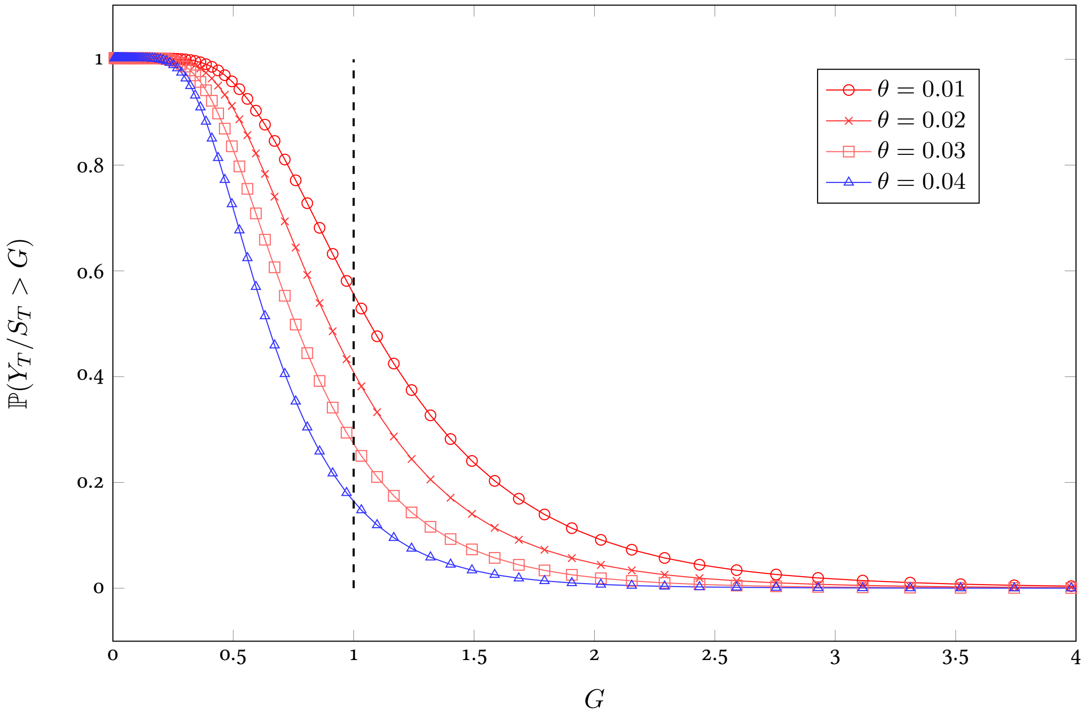


FIGURE 3. Probability that the ratio between the terminal payoff and the value of the underlying at maturity Y_T/S_T exceeds thresholds G for different values of the long-term mean θ. The dashed line at $G = 1$ marks the probability of being underfunded. The remaining parameters have been chosen as in Table 1.

To manage the risk of an equity-indexed annuity, it is necessary to evaluate the probability that the underlying fund can cover the guaranteed expenses at maturity, and quantiles of the potential loss, that is to describe the distribution of the solvency ratio Y_T/S_T. This relative Value-at-Risk is computed using the SM method as explained in Remark 2.3.2. In Fig. 3, we plot the tail function of the solvency ratio Y_T/S_T, that is the probability that Y_T/S_T exceeds thresholds G. Note that the tail function at the threshold $G = 1$ describes the probability that $Y_T > S_T$ and thus that the initial investment can not cover the payment at maturity. Lowering the long-term mean interest rate θ from 0.03 to 0.02 increases this probability from 27.3% to 40.8%. The ratio Y_T/S_T in the 1% least favorable scenarios, that is the 99%-quantile of this quotient, for $\theta = 0.03$ is above 2.76, compared to 1.59 under constant interest rates. In other words, with a probability of 1%, the underlying fund covers only around $1/2$ of the payments to the policyholder. This imbalance intensifies to 3.35 for $\theta = 0.01$ showing how sensitive extreme events are to changes in the underlying assumptions and thus giving another reason for modelling interest rates stochastically.

2.6. Conclusion

This article introduces the scenario-matrix (SM) method for evaluating the risks of equity-indexed annuities with a cliquet-style payoff structure. This approach is fundamentally different from a Markov chain approximation, but has some analogy to the case of regime switching models (see, e.g. Hieber, 2017). In the case of a Vasicek-Black-Scholes model, it outperforms benchmark techniques such as the PROJ method in terms of speed and accuracy and is easy to implement. Approximating the moment-generating function of the log-payoff allows us to describe the whole distribution of the payoff. In contrast to the existing literature, we consider quantile estimation and offer an access to the relative Value-at-Risk.

In numerical studies, we are able to show convergence with quadratic order. The algorithm is computationally less expensive and provides more accurate results than existing approaches. The sensitivity analysis suggests that first, modelling interest rates stochastically considerably influences the risk assessment when compared to constant interest rates. Second, changes in the financial market should be considered in the management of these products.

We introduced the SM method adapted to the special case of a Vasicek-Black-Scholes model that leads to very convenient formulas in Theorem 2.3.1. This technique can, however, be applied and extended to much more general settings. First, it is straightforward to use a different stochastic interest rate model, for example a two-factor Hull-White model. Second, the scenarios might incorporate not only interest rate risk but also a stochastic volatility (see also Cui et al., 2017; Kirkby, 2023). Instead of conditioning on $r^{(k)}$ and $r^{(j)}$, the scenario matrix conditions on pairs $(r^{(k)},\, \sigma^{(k)})$ and $(r^{(j)},\, \sigma^{(j)})$ for an appropriately chosen volatility grid $\sigma^{(k)}$, $k = 1, 2, \ldots, K$. A further possible research direction would be to estimate the scenario matrix immediately from data without specifying a model, see the recent results on data-driven estimations (e.g. Li and Forsyth, 2019).

In future research, it may be interesting to incorporate several additional contract features. First, instead of fixing the maturity time, we may consider early contract terminations and mortality. Second, it could be interesting to analyse the effect of the investment strategy of the underlying fund on the contract payoff and its risk, see also Chen and Hieber (2016) for the case of a maturity guarantee. Third, in this article, we focus on a single contract or a portfolio of homogeneous contracts. In practice, we may be interested in the risk management of a heterogeneous portfolio (see also Hieber et al., 2019). If the SM method cannot be adapted to such settings, the estimates can at least serve as control variates to speed up computations for solvency ratios or absolute Value-at-Risks.

2.A. Proof of Theorem 2.3.1

Definition 2.A.1. *Using the notation* $P_u(t-1,t) = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-u\int_{t-1}^t r_s\,\mathrm{d}s}\right]$, *the u-scaled t-forward measure* $\mathbb{Q}_u^t$ *is defined via*

$$\frac{\mathrm{d}\mathbb{Q}_u^t}{\mathrm{d}\mathbb{Q}} = \frac{\mathrm{e}^{-u\int_{t-1}^t r_s\,\mathrm{d}s}}{P_u(t-1,t)} = \exp\left(-u\sigma_r\int_{t-1}^t g_\kappa(t-s)\,\mathrm{d}W_s^{(1)} + \frac{1}{2}u^2\sigma_r^2\int_{t-1}^t g_\kappa(t-s)^2\,\mathrm{d}s\right). \tag{22}$$

The integrated interest $\int_{t-1}^t r_s\,\mathrm{d}s$ and the log return $\ln(S_t/S_{t-1})$, both conditional on the interest rates r_t and r_{t-1}, under the pricing measure $\mathbb{Q}$ and the u-scaled t-forward measure $\mathbb{Q}_u^t$ respectively, are distributed as

$$\int_{t-1}^t r_s\,\mathrm{d}s\,\big|\, r_t, r_{t-1} \sim \mathcal{N}_{\mathbb{Q}}(\mu, \Sigma^2)\,,$$
$$\ln(S_t/S_{t-1})\,\big|\;\, r_t,\, r_{t-1} \sim \mathcal{N}_{\mathbb{Q}_u^t}(\widetilde{\mu}, \widetilde{\Sigma}^2)\,,$$

where $\mu, \widetilde{\mu}, \Sigma, \widetilde{\Sigma}$ are derived in the following. Using the t-forward measure $\mathbb{Q}_u^t$ as introduced in (22), we know that according to Girsanov's Theorem in the form of (Shreve, 2010, Theorem 5.4.1),

$$\widetilde{W}_t^{(1)} = W_t^{(1)} + u\cdot\sigma_r\int_{t-1}^t g_\kappa(t-s)\,\mathrm{d}s$$

and $W_t^{(2)}$ are Brownian motions under $\mathbb{Q}_u^t$.

Therefore, we can use this change of measure to decouple the integrated interest rate and the annual returns of the underlying stock as follows:

$$\begin{aligned}
\phi_{r_{t-1},r_t}(u) &= P_u(t-1,t)\cdot \mathbb{E}_{\mathbb{Q}}\left[\frac{\mathrm{e}^{-u\int_{t-1}^t r_s\,\mathrm{d}s}}{P_u(t-1,t)}\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_{t-1},r_t\right]\\
&= P_u(t-1,t)\cdot \mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, r_{t-1},r_t\right]\\
&= P_u(t-1,t)\cdot \Bigg(\mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot g}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)\leqslant g\right\}}\,\middle|\, r_{t-1},r_t\right]\\
&\qquad + \mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)> g\right\}}\,\middle|\, r_{t-1},r_t\right]\Bigg).
\end{aligned} \tag{23}$$

To determine the distribution of $\int_{t-1}^t r_s\,\mathrm{d}s$, given r_{t-1}, r_t, under the risk neutral measure $\mathbb{Q}$, we use that for a Vasicek interest rate model, we have (see Shreve, 2010, Chapter 4.4)

$$\begin{aligned}
r_t &= \mathrm{e}^{-\kappa}r_{t-1} + \theta(1-\mathrm{e}^{-\kappa}) + \sigma_r\mathrm{e}^{-\kappa t}\int_{t-1}^t \mathrm{e}^{\kappa s}\,\mathrm{d}W_s^{(1)}\,,\\
\int_{t-1}^t r_s\,\mathrm{d}s &= g_\kappa(1)\,r_{t-1} + \theta\big(1-g_\kappa(1)\big) + \sigma_r\int_{t-1}^t g_\kappa(t-s)\,\mathrm{d}W_s^{(1)}\,,\\
\ln\left(\frac{S_t}{S_{t-1}}\right) &= \int_{t-1}^t r_s\,\mathrm{d}s - \frac{1}{2}\sigma_S^2 + \sigma_S\left(\rho(W_t^{(1)}-W_{t-1}^{(1)}) + \sqrt{1-\rho^2}(W_t^{(2)}-W_{t-1}^{(2)})\right)\,.
\end{aligned}$$

Using Itô's isometry, we can describe the distribution of $\int_{t-1}^t r_s\,\mathrm{d}s$ and r_t under $\mathbb{Q}$ as a two-dimensional normal distribution with parameters

$$\begin{aligned}
\mu_1 &:= \mathbb{E}_{\mathbb{Q}}\left[\int_{t-1}^t r_s\,\mathrm{d}s\right] = g_\kappa(1)r_{t-1} + \theta\big(1-g_\kappa(1)\big)\,,\\
\mu_2 &:= \mathbb{E}_{\mathbb{Q}}\left[r_t\right] = \mathrm{e}^{-\kappa}r_{t-1} + \theta(1-\mathrm{e}^{-\kappa})\,,\\
\sigma_{11}^2 &:= \mathrm{Var}_{\mathbb{Q}}\left(\int_{t-1}^t r_s\,\mathrm{d}s\right) = \frac{\sigma_r^2}{\kappa^2}\left(1-g_\kappa(1)-\frac{\kappa}{2}g_\kappa(1)^2\right)\,,\\
\sigma_{22}^2 &:= \mathrm{Var}_{\mathbb{Q}}(r_t) = \frac{\sigma_r^2}{2}g_\kappa(2)\,,\\
\sigma_{12} &:= \mathrm{Cov}_{\mathbb{Q}}\left(r_t,\int_{t-1}^t r_s\,\mathrm{d}s\right) = \frac{\sigma_r^2}{2}g_\kappa(1)^2\,.
\end{aligned}$$

Therefore, under $\mathbb{Q}$, Eaton (2007, Proposition 3.13) shows that

$$\int_{t-1}^t r_s\,\mathrm{d}s\,\big|\, r_t, r_{t-1} \sim \mathcal{N}(\mu,\Sigma^2)\,, \tag{24}$$

with

$$\mu := \mu_1 + \sigma_{12}\sigma_{22}^{-2}(r_t-\mu_2)\,,\qquad \Sigma^2 := \sigma_{11}^2 - \sigma_{12}^2\sigma_{22}^{-2}\,.$$

To express these values in the form of Theorem 2.3.1, it is helpful to write

$$\sigma_{12}\sigma_{22}^{-2} = \frac{g_\kappa(1)^2}{g_\kappa(2)} = \frac{g_\kappa(1)(1-e^{-\kappa})}{(1+e^{-\kappa})(1-e^{-\kappa})} = \frac{g_\kappa(1)}{1+e^{-\kappa}}. \tag{25}$$

To determine the distribution of $\ln\left(S_t/S_{t-1}\right)$, given r_t, r_{t-1}, under $\mathbb{Q}_u^t$, we calculate

$$\begin{aligned} r_t =&\ \mathrm{e}^{-\kappa} r_{t-1} + \theta(1-\mathrm{e}^{-\kappa}) + \sigma_r \mathrm{e}^{-\kappa t}\left(\int_{t-1}^{t} \mathrm{e}^{\kappa s}\,\mathrm{d}\widetilde{W}_s^{(1)} - \int_{t-1}^{t} \mathrm{e}^{\kappa s} u\sigma_r g_\kappa(t-s)\,\mathrm{d}s\right) \\ =&\ \mathrm{e}^{-\kappa} r_{t-1} + \theta(1-\mathrm{e}^{-\kappa}) - u\frac{\sigma_r^2}{2} g_\kappa(1)^2 + \sigma_r \mathrm{e}^{-\kappa t}\int_{t-1}^{t} \mathrm{e}^{\kappa s}\,\mathrm{d}\widetilde{W}_s^{(1)}\,, \end{aligned}$$

$$\begin{aligned} \int_{t-1}^{t} r_s\,\mathrm{d}s =&\ g_\kappa(1) r_{t-1} + \theta\big(1-g_\kappa(1)\big) + \sigma_r\left(\int_{t-1}^{t} g_\kappa(t-s)\,\mathrm{d}\widetilde{W}_s^{(1)} - \int_{t-1}^{t} u\sigma_r g_\kappa(t-s)^2\,\mathrm{d}s\right) \\ =&\ g_\kappa(1) r_{t-1} + \theta\big(1-g_\kappa(1)\big) - u\frac{\sigma_r^2}{\kappa^2}\left(1-g_\kappa(1)-\frac{\kappa}{2}g_\kappa(1)^2\right) \\ &+ \sigma_r\int_{t-1}^{t} g_\kappa(t-s)\,\mathrm{d}\widetilde{W}_s^{(1)}\,, \end{aligned}$$

$$\begin{aligned} \ln\left(\frac{S_t}{S_{t-1}}\right) =&\ \int_{t-1}^{t} r_s\,\mathrm{d}s - \frac{1}{2}\sigma_S^2 + \sigma_S\bigg(\int_{t-1}^{t}\rho\,\mathrm{d}\widetilde{W}_s^{(1)} - \int_{t-1}^{t}\rho u\sigma_r g_\kappa(t-s)\,\mathrm{d}s \\ &\qquad + \int_{t-1}^{t}\sqrt{1-\rho^2}\,\mathrm{d}W_s^{(2)}\bigg) \\ =&\ g_\kappa(1) r_{t-1} + \theta\big(1-g_\kappa(1)\big) - u\frac{\sigma_r^2}{\kappa^2}\left(1-g_\kappa(1)-\frac{\kappa}{2}g_\kappa(1)^2\right) - \frac{1}{2}\sigma_S^2 \\ &- u\frac{\sigma_S\sigma_r\rho}{\kappa}\big(1-g_\kappa(1)\big) + \sigma_r\int_{t-1}^{t} g_\kappa(t-s)\,\mathrm{d}\widetilde{W}_s^{(1)} + \sigma_S\int_{t-1}^{t}\rho\,\mathrm{d}\widetilde{W}_s^{(1)} \\ &+ \sigma_S\int_{t-1}^{t}\sqrt{1-\rho^2}\,\mathrm{d}W_s^{(2)}. \end{aligned}$$

Again, we can describe the distribution of $\ln\left(S_t/S_{t-1}\right)$ and r_t as a two-dimensional normal distribution with parameters

$$\begin{aligned} \widetilde{\mu}_1 :=&\ \mathbb{E}_{\mathbb{Q}_u^t}\left[\ln\left(\frac{S_t}{S_{t-1}}\right)\right] = g_\kappa(1) r_{t-1} + \theta\big(1-g_\kappa(1)\big) - u\frac{\sigma_r^2}{\kappa^2}\left(1-g_\kappa(1)-\frac{\kappa}{2}g_\kappa(1)^2\right) \\ &\qquad - \frac{1}{2}\sigma_S^2 - u\frac{\sigma_S\sigma_r\rho}{\kappa}\big(1-g_\kappa(1)\big)\,, \end{aligned}$$

$$\widetilde{\mu}_2 := \mathbb{E}_{\mathbb{Q}_u^t}\left[r_t\right] = \mathrm{e}^{-\kappa} r_{t-1} + \theta(1-\mathrm{e}^{-\kappa}) - u\frac{\sigma_r^2}{2} g_\kappa(1)^2\,,$$

$$\widetilde{\sigma}_{11}^2 := \mathrm{Var}_{\mathbb{Q}_u^t}\left(\ln\left(\frac{S_t}{S_{t-1}}\right)\right) = \frac{\sigma_r^2}{\kappa^2}\left(1-g_\kappa(1)-\frac{\kappa}{2}g_\kappa(1)^2\right) + \frac{2\sigma_r\sigma_S\rho}{\kappa}\big(1-g_\kappa(1)\big) + \sigma_S^2\,,$$

$$\widetilde{\sigma}_{22}^2 := \mathrm{Var}_{\mathbb{Q}_u^t}(r_t) = \frac{\sigma_r^2}{2} g_\kappa(2)\,,$$

$$\widetilde{\sigma}_{12} := \mathrm{Cov}_{\mathbb{Q}_u^t}\left(r_t, \ln\left(\frac{S_t}{S_{t-1}}\right)\right) = \frac{\sigma_r^2}{2} g_\kappa(1)^2 + \sigma_r\sigma_S\rho g_\kappa(1).$$

Therefore, we can conclude that, under $\mathbb{Q}_u^t$, we have

$$\ln(S_t/S_{t-1})\,\big|\ r_t,\, r_{t-1} \sim \mathcal{N}(\widetilde{\mu}, \widetilde{\Sigma}^2)\,, \tag{26}$$

with

$$\widetilde{\mu} := \widetilde{\mu}_1 + \widetilde{\sigma}_{12}\widetilde{\sigma}_{22}^{-2}(r_t - \widetilde{\mu}_2)\,, \qquad \widetilde{\Sigma}^2 := \widetilde{\sigma}_{11}^2 - \widetilde{\sigma}_{12}^2\widetilde{\sigma}_{22}^{-2}.$$

Note that, similarly to equation (25), we find

$$\sigma_{12}\sigma_{22}^{-2} = \frac{g_k(1)^2 + s\frac{\sigma_s}{\sigma_r}\rho g_\kappa(1)}{g_\kappa(2)} = \frac{g_k(1) + s\frac{\sigma_s}{\sigma_r}\rho}{1+e^{-\kappa}}.$$

Using (24), we get

$$P_u(t,t-1) = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_{t-1}^t r_s\,\mathrm{d}s}\,\middle|\, r_{t-1}, r_t\right] = \exp\left(-u\mu + \frac{u^2\Sigma^2}{2}\right).$$

With (26), we can calculate

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot g}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)\leqslant g\right\}}\,\middle|\, r_{t-1}, r_t\right] + \mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)> g\right\}}\,\middle|\, r_{t-1}, r_t\right]\\
&= \mathrm{e}^{ug}\mathbb{Q}_u^t\left(\ln\left(\frac{S_t}{S_{t-1}}\right)\leqslant \frac{g}{\alpha}\,\middle|\, r_{t-1}, r_t\right) + \int_{\frac{g}{\alpha}}^\infty \mathrm{e}^{u\alpha x}\sqrt{2\pi\widetilde{\Sigma}^2}^{-1}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\widetilde{\mu}}{\widetilde{\Sigma}}\right)^2}\,\mathrm{d}x\\
&= \mathrm{e}^{ug}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}}{\widetilde{\Sigma}}\right) + \mathrm{e}^{u\alpha\widetilde{\mu}+\frac{1}{2}u^2\alpha^2\widetilde{\Sigma}^2}\Phi\left(-\frac{\frac{g}{\alpha}-\widetilde{\mu}-u\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right).
\end{aligned}$$

Inserting both results into (23), we arrive at

$$\phi_{r_{t-1},r_t}(u) = \mathrm{e}^{-u\mu+\frac{u^2\Sigma^2}{2}}\left(\mathrm{e}^{ug}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}}{\widetilde{\Sigma}}\right) + \mathrm{e}^{u\alpha\widetilde{\mu}+\frac{1}{2}u^2\alpha^2\widetilde{\Sigma}^2}\Phi\left(-\frac{\frac{g}{\alpha}-\widetilde{\mu}-u\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right)\right). \tag{27}$$

□

2.B. Proof of Theorem 2.3.6

We have to show that for all $r^* \in \mathbb{R}$ and all $t \in \mathbb{N}$

$$\lim_{K\to\infty}\left|\mathbb{P}\left(\bar{r}_t^K \leqslant r^*\right) - \mathbb{P}(r_t \leqslant r^*)\right| = 0\,,$$

which we will proof by induction. Let $r^* \in (r^{(j)}, r^{(j+1)})$. We have for $t=1$, that

$$\begin{aligned}
\left|\mathbb{P}(\bar{r}_1^K \leqslant r^*) - \mathbb{P}(r_1 \leqslant r^*)\right| &= \left|\mathbb{P}\Big(r_1 \leqslant \frac{1}{2}\big(r^{(j)} + r^{(j+1)}\big)\Big) - \mathbb{P}\big(r_1 \leqslant r^*\big)\right|\\
&\leqslant \mathbb{P}\left(r_1 \in \left[r^{(j)}, r^{(j+1)}\right]\right) \in \mathcal{O}(1/K)\,.
\end{aligned}$$

Assuming that the statement is valid for $t-1$, we find for $t>1$ that

$$\begin{aligned}
&\left|\mathbb{P}(\bar{r}_t^K \leqslant r^*) - \mathbb{P}(r_t \leqslant r^*)\right|\\
&= \Bigg|\sum_{i=1}^K \mathbb{P}\left(\bar{r}_t^K \leqslant r^* \,\middle|\, \bar{r}_{t-1}^K = r^{(i)}\right)\cdot\mathbb{P}\left(\bar{r}_{t-1}^K = r^{(i)}\right)\\
&\qquad - \mathbb{P}\left(r_t \leqslant r^* \,\middle|\, r_{t-1} \in \left[r^{(i)}\right]\right)\cdot\mathbb{P}\left(r_{t-1} \in \left[r^{(i)}\right]\right)\Bigg|\\
&\leqslant \left|\sum_{i=1}^K \int_{\frac{r^{(i)}+r^{(i-1)}}{2}}^{\frac{r^{(i)}+r^{(i+1)}}{2}} \mathbb{P}\left(r_t \leqslant r^* \,\middle|\, r_{t-1} = x\right) - \mathbb{P}\Big(r_t \leqslant \frac{1}{2}\big(r^{(j)} + r^{(j+1)}\big)\,\Big|\, r_{t-1} = x\Big)\,\mathbb{P}\big(r_{t-1} \in \mathrm{d}x\big)\right|
\end{aligned}$$

$$
\begin{aligned}
&+\left|\sum_{i=1}^{K}\mathbb{P}\left(r_{t-1}\in[r^{(i)}]\right)\left(\mathbb{P}\Big(r_t\leqslant\frac{1}{2}\big(r^{(j)}+r^{(j+1)}\big)\,\Big|\,r_{t-1}=r^{(i)}\Big)\right.\right.\\
&\qquad\left.\left.-\mathbb{P}\Big(r_t\leqslant\frac{1}{2}\big(r^{(j)}+r^{(j+1)}\big)\,\Big|\,r_{t-1}\in\Big[r^{(i)}\Big]\Big)\right)\right|\\
&+\left|\sum_{i=1}^{K}\mathbb{P}\left(r_t\leqslant\frac{1}{2}\big(r^{(j)}+r^{(j+1)}\big)\,\Big|\,r_{t-1}=r^{(i)}\right)\left(\mathbb{P}\left(r_{t-1}\in[r^{(i)}]\right)-\mathbb{P}\left(\bar r_{t-1}^{K}=r^{(i)}\right)\right)\right|\\
&=:|S_1|+|S_2|+|S_3|\in\mathcal{O}(1/K)\,.
\end{aligned}
$$

Here, we used that

$$
\begin{aligned}
&\mathbb{P}\left(\bar r_t^{K}\leqslant r^*\,\middle|\,\bar r_{t-1}^{K}=r^{(i)}\right)\cdot\mathbb{P}\left(\bar r_{t-1}^{K}=r^{(i)}\right)+S_2+S_3\\
&\qquad=\mathbb{P}\Big(r_t\leqslant\frac{1}{2}\big(r^{(j)}+r^{(j+1)}\big)\,\Big|\,r_{t-1}\in\Big[r^{(i)}\Big]\Big)\cdot\mathbb{P}\Big(r_{t-1}\in\Big[r^{(i)}\Big]\Big)\,.
\end{aligned}
$$

Thus, we are able to express the right hand side of the above equation as an integral in S_1. Additionally, S_1 is a sum of K integrals whose integrand is in $\mathcal{O}(1/K)$ by the same reasoning as for $t=1$. Furthermore, its integration area, $[r^{(i)}]$ has a probability weight of $\mathcal{O}(1/K)$ since the grid points were chosen such that $\mathbb{P}(r_1\in[r^{(j)}])\in\mathcal{O}(1/K)$. For finite t, this property transfers from r_1 to r_t. Combining these observations, S_1 is in $K\cdot\mathcal{O}(1/K^2)=\mathcal{O}(1/K)$.
By the same argument, in S_2, $\mathbb{P}(r_1\in[r^{(j)}])$ is in $\mathcal{O}(1/K)$. Also $\mathbb{P}(r_t\leqslant\frac{1}{2}(r^{(j)}+r^{(j+1)})\,|\,r_{t-1}=r^{(i)})-\mathbb{P}(r_t\leqslant\frac{1}{2}(r^{(j)}+r^{(j+1)})\,|\,r_{t-1}=[r^{(i)}])$ is in $\mathcal{O}(1/K)$ since it varies the starting condition r_{t-1} by no more than $r^{(i-1)}-r^{(i+1)}$.
Finally, to see that S_3 is in $\mathcal{O}(1/K)$, we need to apply summation by parts and the induction assumption that $\mathbb{P}(\bar r_{t-1}^{K}\leqslant r^*)-\mathbb{P}(r_{t-1}\leqslant r^*)\in\mathcal{O}(1/K)$. □

2.C. PROJ method: Details

This follows Kirkby (2015), Cui et al. (2017). Given the Haar scaling function $\varphi^{[0]}(y)=\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(y)$, it is easy to show that $\varphi^{[1]}(y)=\varphi^{[0]}(y)\star\varphi^{[0]}(y)$.[1] The Fourier transform of the Haar scaling function is given by:

$$
\hat\varphi^{[0]}(\xi)=\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\xi y}\cdot\varphi^{[0]}(y)\,\mathrm{d}y=\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\xi y}\cdot\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(y)\,\mathrm{d}y=\frac{2\cdot\sin\left(-\frac{\xi}{2}\right)}{-\xi}=\frac{\sin\left(\frac{\xi}{2}\right)}{\frac{\xi}{2}}\,.
$$

This and $\varphi^{[1]}(y)=\varphi^{[0]}(y)\star\varphi^{[0]}(y)$ shows that the Fourier transform of the B-spline basis function is $\hat\varphi^{[1]}(\xi)=4\cdot\sin^2\left(\frac{\xi}{2}\right)/\xi^2$. Obviously the function $\hat\varphi^{[1]}(\xi)$ is symmetric and real-valued. Using the definition of scalar product for complex numbers and the Plancherel theorem, we obtain:

$$
\begin{aligned}
\beta_{a,n}&=\big\langle f(y)\,,\varphi_{a,n}(y)\big\rangle=\int_{-\infty}^{\infty}f(y)\cdot\overline{\varphi_{a,n}(y)}\,\mathrm{d}y\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat f(\xi)\cdot\overline{\hat\varphi_{a,n}(\xi)}\,\mathrm{d}\xi=\frac{\sqrt{a}}{2\pi}\int_{-\infty}^{\infty}\hat f(\xi)\cdot\frac{\mathrm{e}^{-\mathrm{i}x_n\xi}}{a}\overline{\hat\varphi(\xi/a)}\,\mathrm{d}\xi
\end{aligned}
$$

[1]Note that:

$$
\varphi^{[1]}(y)=\varphi^{[0]}(y)\star\varphi^{[0]}(y)=\int_{-\infty}^{\infty}\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(y-x)\cdot\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(x)\,\mathrm{d}x=\begin{cases}\int_{-\frac{1}{2}}^{y+\frac{1}{2}}\,\mathrm{d}x=1+y, & y\in[-1,0]\\ \int_{y-\frac{1}{2}}^{\frac{1}{2}}\,\mathrm{d}x=1-y, & y\in[0,1]\end{cases}\,.
$$

$$= \frac{\sqrt{a}}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\xi)\cdot\frac{\mathrm{e}^{-\mathrm{i}x_n\xi}}{a}\overline{\hat{\varphi}^{[1]}(\xi/a)}\,\mathrm{d}\xi = \frac{12}{2\pi\sqrt{a}}\int_{-\infty}^{\infty}\hat{f}(\xi)\cdot\mathrm{e}^{-\mathrm{i}x_n\xi}\cdot\frac{a^2\cdot\sin^2\left(\frac{\xi}{2a}\right)}{\xi^2\left(2+\cos\frac{\xi}{a}\right)}\,\mathrm{d}\xi$$

$$(28)\qquad = \frac{12a^{\frac{3}{2}}}{\pi}\int_0^{\infty}\hat{f}(\xi)\cdot\mathrm{e}^{-\mathrm{i}x_n\xi}\cdot\frac{\sin^2\left(\frac{\xi}{2a}\right)}{\xi^2\left(2+\cos\frac{\xi}{a}\right)}\,\mathrm{d}\xi\,,$$

We evaluate the integral (28) among grid points $\xi_j = 2\pi(j-1)a/N$ for $j = 1, 2, \ldots, N$. This results in a discrete Fourier transform using the integration steps of length $\triangle\xi_j = 2\pi a/N$:

$$\beta_{a,n} \approx \frac{24a^{\frac{5}{2}}}{N}\cdot\mathcal{D}_n(\boldsymbol{H})\,,\qquad \mathcal{D}_n(\boldsymbol{H}) = \sum_{j=1}^{N}\mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}(j-1)(n-1)}\cdot H_j\,,$$

where $H_1 = 1/24a^2$ and $H_j = \mathrm{e}^{-\mathrm{i}x_1\xi_j}\cdot\hat{f}(\xi_j)\cdot\sin^2(\xi_j/2a)/(\xi_j^2(2+\cos(\xi_j/a)))$, $j = 2, 3, \ldots, N$.

Given the financial market (1) and (2), we can exploit the affine structure of the Vasicek interest model. Given the functions $\eta(r_t) := \rho\frac{\sigma_S}{\sigma_r}(\theta - r_t)$ and $\zeta(r_t, r_0) := \kappa\frac{\sigma_S}{\sigma_r}(r_t - r_0)$, it is straightforward to show that $\rho\sigma_S\,\mathrm{d}W_t^{(1)} = \mathrm{d}\zeta(r_t, r_0) - \rho\cdot\eta(r_t)\,\mathrm{d}t$. This yields:

$$\begin{aligned}\mathrm{d}\ln(S_t) &= \left(r_t + \lambda - \frac{\sigma_S^2}{2}\right)\mathrm{d}t + \sqrt{1-\rho^2}\sigma_S\,\mathrm{d}W_t^{(2)} + \rho\sigma_S\mathrm{d}W_t^{(1)}\\ &= \left(r_t + \lambda - \frac{\sigma_S^2}{2}\right)\mathrm{d}t + \sqrt{1-\rho^2}\sigma_S\,\mathrm{d}W_t^{(2)} + \mathrm{d}\zeta(r_t, r_0) - \rho\eta(r_t)\,\mathrm{d}t\end{aligned}$$

which can be integrated to:

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = \zeta(r_t, r_{t-1}) + \lambda + \int_{t-1}^{t}\big(r_s + \lambda - \eta(r_s)\big)\,\mathrm{d}s - \frac{\sigma_S^2}{2} + \sqrt{1-\rho^2}\sigma_S\left(W_t^{(2)} - W_{t-1}^{(2)}\right).$$

We condition on $\left(\int_{t-1}^{t} r_s\,\mathrm{d}s,\ r_{t-1},\ r_t\right) = (y,\ r^{(k)},\ r^{(j)})$ and use the affinity of η to show that the log-returns $\ln(S_t/S_{t-1})$ are normally distributed with mean and variance:

$$(29)\qquad \mu_{k,j}(y) := \mathbb{E}\big[\ln(S_t/S_{t-1})\big] = \zeta(r^{(k)}, r^{(j)}) + y\Big(1 - \rho\frac{\sigma_S}{\sigma_r}\theta\Big) - \Big(\rho\frac{\sigma_S}{\sigma_r}\theta + \frac{\sigma_s^2}{2}\Big)\,,$$

$$(30)\qquad \sigma_{k,j}^2 := \mathrm{Var}\left(\ln(S_t/S_{t-1})\right) = \sigma_S^2\left(1-\rho^2\right)\,.$$

We represent the value of an equity-indexed annuity in the integral form (17) over the integrated interest rate $y = \int_{t-1}^{t} r_s\,\mathrm{d}s$, where the expected payoff function $h(y)$ is conditioned on $r_{t-1} = r^{(k)}$ and $r_t = r^{(j)}$, denoted by $h_{k,j}(y)$. Similarly, the density of $y = \int_{t-1}^{t} r_s\,\mathrm{d}s$ conditional on $r_{t-1} = r^{(k)}$ and $r_t = r^{(j)}$ is denoted $f_{k,j}(y)$. Analogous to our proof of Theorem 2.3.1, we can show that:

$$\begin{aligned}h_{k,j}(y) &:= \mathbb{E}_{\mathbb{P}}\Big[\mathrm{e}^{L_{t-1,t}}\,\Big|\,r_{t-1} = r^{(k)},\ r_t = r^{(j)},\ \int_{t-1}^{t} r_s\,\mathrm{d}s = y\Big]\\ (31)\qquad &= \mathrm{e}^{g-y}\Phi\left(\frac{\frac{g}{\alpha} - \mu_{k,j}(y)}{\sigma_{k,j}}\right) + \mathrm{e}^{\alpha\mu_{k,j}(y) + \frac{\alpha\sigma_{k,j}^2}{2} - y}\Phi\left(\frac{\alpha\sigma_{k,j}^2 + \mu_{k,j}(y) - \frac{g}{\alpha}}{\sigma_{k,j}}\right).\end{aligned}$$

CHAPTER 3

Efficient simulation and valuation of equity-indexed annuities under a two-factor G2++ model

This chapter is based on the following article:

S. Günther, P. Hieber (2024): Efficient simulation and valuation of equity-indexed annuities under a two-factor G2++ model. *European Actuarial Journal* 14: 905–928.

Abstract. Equity-indexed annuities (EIAs) with investment guarantees are pension products sensitive to changes in the interest rate environment. A flexible and common choice for modelling this risk factor is a Hull-White model in its G2++ variant. We investigate the valuation of EIAs in this model setting and extend the literature by introducing a more efficient framework for Monte-Carlo simulation. In addition, we build on previous work by adapting an approach based on scenario matrices to a two-factor G2++ model. This method does not rely on simulations or on Fourier transformations. In numerical studies, we demonstrate its fast convergence and the limitations of techniques relying on the independence of annual returns and the central limit theorem.

3.1. Introduction

Interest rate levels significantly impact the management of equity-indexed annuities (EIAs), guaranteeing minimum annual returns while allowing participation in higher returns from an underlying fund. The period of low interest rates spanning from 2009 to 2022 presented serious challenges for life insurance companies and has been followed by the frequent succession of rising interest rates since 2022. The uncertainty around future interest rate developments and the difficulties arising from those motivate the adoption of stochastic interest rate models. However, the stochastic modelling of interest rates introduces its own set of technical challenges, particularly for variants like cliquet-style or ratchet-type EIAs with a compounding structure.

For deterministic interest rates, Korn et al. (2017) employ a central limit theorem (CLT) argument to approximate the distribution of the final payoff by a normal distribution. Bernard and Li (2013) derive an integral expression for the non-compounding variant of EIAs within a Lévy interest rate model. The case of a regime-switching interest rate model is considered by Hieber (2017), who employs Fourier techniques. Attempts at a Black-Scholes-Vasicek model are presented by Kijima and Wong (2007), who develop an analytical yet numerically challenging solution, representing the contract value as a multi-dimensional normal integral. A computationally more efficient approach to this setting is presented in Günther and Hieber (2024), where the problem is decomposed into a set of scenarios by conditioning on preceding and succeeding annual interest rates, allowing for efficient closed-form approximations.

In this article, we consider a two-factor interest rate model, being more general than academic articles dealing with one-factor Hull-White or Vasicek interest rate models (see, for example, Kijima and Wong, 2007; Günther and Hieber, 2024). We use the two-factor Hull-White model in the so-called G2++ variant (see Section 4.2 in Brigo and Mercurio, 2006 and Section 3.2 for more details). This model has several advantages:

- It can perfectly replicate an existing initial yield curve.
- It allows for negative interest rates.
- The normal distribution of the short rate leads to some analytical tractability of the model.
- In Germany and Austria, it is recognised for the capital market model for pan-European PRIIP-regulation. In Germany, the G2++ model is known as the PIA ("Produkt Informationsstelle Altersvorsorge") model, and used to classify risk/return profiles of second and third pillar state subsidised pension products. For this reason, the model is well understood in the insurance industry.

The G2++ model was first introduced into the academic actuarial literature by Graf and Korn (2020) in terms of a framework for Monte-Carlo simulation in the chance-risk classification of insurance products. Berninger and Pfeiffer (2021) show how to calibrate this model under the real-world measure by introducing time-dependent risk premia for more stable long-term interest rate forecasts. The modelling of capital requirements under a G2++ model for premium risk in non-life insurance is considered by Cotticelli and Savelli (2023). Barigou and Delong (2022) apply neural networks to annuity pricing based on this model.

The contribution of this article is two-fold: First, we want to contribute to the tractability of the Hull-White G2++ model, providing the joint distribution of integrated interest rate and a correlated risky asset as well as its moment-generating function. This allows for analytic results within this model framework but also an unbiased Monte-Carlo simulation on an annual grid. Secondly, we want to suggest an alternative to Monte-Carlo simulation to analyse products in the Hull-White G2++ model. We, therefore, apply the scenario-matrix approach suggested in Günther and Hieber (2024) to the specifics of a two-factor short-rate model. We compare our results to the Monte-Carlo simulation approach as explained in Graf and Korn (2020).

The article is organised as follows. We introduce the setting of the financial market in Section 3.2. There, we also describe the joint distribution of the two factors of the interest rate process, the integrated interest rate and the annual log-returns as a multivariate normal distribution. This result can replace the computationally less efficient Euler-Maruyama scheme that has been suggested by Graf and Korn (2020). We describe the specific procedure in Section 3.3 alongside our scenario-matrix approach. In Section 3.4, we compare the two approaches.

3.2. Equity indexed annuity, financial market and G2++ model

A cliquet-style equity-indexed annuity (EIA) guarantees the policyholder a minimum annual return, specified by a continuously compounding guaranteed rate g, on their investment. However, in successful years, they participate on higher returns of the underlying fund $S = \{S_t\}_{t\geq 0}$. In this article, we work with logarithmic annual returns $R_t = \ln(S_t/S_{t-1})$. The log-return of the equity-indexed annuity is then the maximum of g and a share $\alpha \in [0, 1]$ of R_t, that is $\max(g,\, \alpha \cdot R_t)$. To keep things simple, we look at single-premium EIAs with a premium P_0 paid at time 0 and a maturity of T years. This can easily be extended to a regular premium payment EIA that is simply a portfolio of single-premium EIAs. The terminal EIA payoff Y_T at maturity T is composed of the product of its annual returns, i.e.:

$$Y_T = P_0 \cdot \prod_{t=1}^{T} \mathrm{e}^{\max(g,\, \alpha\cdot R_t)} = \prod_{t=1}^{T} \max\left(\mathrm{e}^g, \left(\frac{S_t}{S_{t-1}}\right)^{\alpha}\right) . \tag{32}$$

This article aims to look at the financial risk of cliquet-style EIAs. Due to the long-term nature of these products, interest rate risk plays a significant role. Therefore, we introduce a financial market with a 2-factor interest rate model that describes the risk to vary from an initially given yield curve. The model is named G2++ model in the following (see Brigo and Mercurio, 2006, Section 4.2). We next introduce this model for a given pricing measure $\mathbb{Q}$, equivalent to the real-world measure $\mathbb{P}$.

Model description under the risk-neutral pricing measure. We consider a financial market with two risky securities: A risk-free bank account whose (stochastic) rate of return is the risk-free interest $\{r_t\}_{t\geq 0}$ and a risky asset with price $\{S_t\}_{t\geq 0}$. The risk drivers of these investments are three standard Brownian motions $W_t^{(1)}$, $W_t^{(2)}$ and $W_t^{(3)}$.

The initial market forward yield curve is adjusted to a Nelson-Siegel-Svensson yield curve (see Svensson, 1994; Korn et al., 2017), where the t-year forward rate is given by:

$$\begin{aligned} f^M(0,t) = \beta_0 + \beta_1 \frac{\theta_1}{t}\left(1 - \mathrm{e}^{-\frac{t}{\theta_1}}\right) + \beta_2 \frac{\theta_1}{t}\left(1 - \mathrm{e}^{-\frac{t}{\theta_1}} \frac{t+\theta_1}{\theta_1}\right) \\ + \beta_3 \frac{\theta_2}{t}\left(1 - \mathrm{e}^{-\frac{t}{\theta_1}} \frac{t+\theta_2}{\theta_2}\right) . \end{aligned} \tag{33}$$

Estimates of the parameters $\beta_0,\, \beta_1,\, \beta_2,\, \beta_3 \in \mathbb{R}$ and $\theta_1,\, \theta_2 > 0$, calibrated to German government bonds are published regularly, for example, by the German Bundesbank (see Bundesbank, 2023). The risk-free bank account evolves as:

$$B_t = B_0 \cdot e^{\int_0^t r_s \,\mathrm{d}s} ,$$

in a way that its expected value behaves according to the t-year forward rate (33):

$$P(s,t) = \mathbb{E}\Big[\, B_t/B_s \Big] = \mathbb{E}\Big[\mathrm{e}^{-\int_s^t r_u \,\mathrm{d}u}\Big] = \mathrm{e}^{-\int_s^t f^M(0,u)\,\mathrm{d}u} , \tag{34}$$

see the Appendix 3.A for more details. The interest rate risk is described by a G2++ model that is composed of three elements: A term ψ_t linked to the expected risk-free rate and two mean-reverting processes x_t and y_t describing the variation around this term:

$$\begin{aligned} r_t &= \psi_t + x_t + y_t , \\ \mathrm{d}x_t &= -a\, x_t \,\mathrm{d}t + \nu \,\mathrm{d}W_t^{(1)} , \\ \mathrm{d}y_t &= -b\, y_t \,\mathrm{d}t + \eta\Big(\rho_r \,\mathrm{d}W_t^{(1)} + \sqrt{1-\rho_r^2}\,\mathrm{d}W_t^{(2)}\Big) , \end{aligned} \tag{35}$$

with two independent Brownian motions $W_t^{(1)}$ and $W_t^{(2)}$ as well as mean-reverting speeds $a > 0, b > 0$ and volatilities $\nu > 0, \eta >$ of the respective processes x and y. The initial values of the two factors x_0 and y_0 are given. In order to satisfy (34), the factor ψ_t is chosen as:

$$\psi_t = f^M(0,t) + \frac{\nu^2}{2} \cdot g_a(t)^2 + \frac{\eta^2}{2} \cdot g_b(t)^2 + \rho\nu\eta \cdot g_a(t) \cdot g_b(t) , \tag{36}$$

where $g_k(s) = \frac{1}{k}\left(1 - \mathrm{e}^{-ks}\right)$, see also the Appendix 3.A for technical details and derivations. For details on the calibration of the interest rate process, we refer to Graf and Korn (2020).

Additionally, the financial market contains a risky asset S_t (the underlying fund in (32)) following a generalised Black-Scholes model with risk-neutral $\mathbb{Q}$-dynamics

$$\frac{\mathrm{d}S_t}{S_t} = r_t\,\mathrm{d}t + \sigma\Big(\rho_1\,\mathrm{d}W_t^{(1)} + \rho_2\,\mathrm{d}W_t^{(2)} + \sqrt{1-\rho_1^2-\rho_2^2}\,\mathrm{d}W_t^{(3)}\Big)\,, \tag{37}$$

where $\rho_1,\ \rho_2 \in [-1,1]$ are the correlation parameters to steer the dependence between the risky asset and the risk-free interest rate. We impose that $\rho_1^2+\rho_2^2 \leq 1$. $W_t^{(3)}$ is a third Brownian motion that is independent of $W_t^{(1)}$ and $W_t^{(2)}$ and allows to model additional financial risk.

The advantage of the model (35) $-$ (37) is the fact that despite its complexity, it is analytically very tractable. It turns out that the interest risk factors x_t and y_t as well as the integrated interest rate $\int_{t-1}^t r_s\,\mathrm{d}s$ are normally distributed. Lemma 3.2.1 proves this fact and provides the groundwork to later simulate the process without bias on an annual grid.

Lemma 3.2.1 (Distribution of $(x_t,\ y_t,\ \int_{t-1}^t r_s\,\mathrm{d}s,\ \ln(S_t/S_{t-1}))$**).**
Conditional on the factors $x_{t-1},\ y_{t-1}$, the vector

$$\boldsymbol{v}_t = \Big(x_t,\ y_t,\ \int_{t-1}^t r_s\,\mathrm{d}s,\ \ln(S_t/S_{t-1})\Big)'$$

is normally distributed with mean $\boldsymbol{\mu} = (\mu_1,\ \mu_2,\ \mu_3,\ \mu_4)'$ and covariance matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})_{i,j}$. We find:

$$\boldsymbol{\mu} = \mathbb{E}_{\mathbb{Q}}\big[v_t \,\big|\, \mathcal{F}_{t-1}\big] = \begin{pmatrix} \mathrm{e}^{-a}\cdot x_{t-1} + \lambda_x(1-\mathrm{e}^{-a}) \\ \mathrm{e}^{-b}\cdot y_{t-1} + \lambda_y(1-\mathrm{e}^{-b}) \\ \mu_3 \\ \mu_3 + \lambda_S - \frac{1}{2}\sigma^2 \end{pmatrix}.$$

The conditional covariance $\boldsymbol{\Sigma} = \mathrm{Var}_{\mathbb{Q}}\big(v_t \,\big|\, \mathcal{F}_{t-1}\big)$ is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \nu^2\cdot g_{2a}(1) & \rho_r\nu\eta\cdot g_{a+b}(1) & \Sigma_{13} & \Sigma_{13}+\rho_1\sigma\nu\cdot g_a(1) \\ \rho_r\nu\eta\cdot g_{a+b}(1) & \eta^2\cdot g_{2b}(1) & \Sigma_{23} & \Sigma_{23}+\widetilde{\rho}_2\sigma\eta\cdot g_b(1) \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} & \Sigma_{33}+\Delta \\ \Sigma_{13}+\rho_1\sigma\nu\cdot g_a(1) & \Sigma_{23}+\widetilde{\rho}_2\sigma\eta\cdot g_b(1) & \Sigma_{33}+\Delta & \Sigma_{33}+\Delta+\sigma^2 \end{pmatrix},$$

where $g_k(s) = \frac{1}{k}\left(1-\mathrm{e}^{-ks}\right)$ and

$$\mu_3 = \int_{t-1}^t \psi_s\mathrm{d}s + g_a(1)\cdot x_{t-1} + g_b(1)\cdot y_{t-1} + \lambda_x(1-g_a(1)) + \lambda_y(1-g_b(1)),$$

$$\Sigma_{13} = \frac{\nu^2}{2}g_a(1)^2 + \frac{\rho_r\nu\eta}{b}\big(g_a(1)-g_{a+b}(1)\big)\,, \qquad \Sigma_{23} = \frac{\eta^2}{2}g_b(1)^2 + \frac{\rho_r\nu\eta}{a}\big(g_b(1)-g_{a+b}(1)\big)\,,$$

$$\Sigma_{33} = \frac{\nu^2}{a^2}\big(1-2g_a(1)+g_{2a}(1)\big) + \frac{\eta^2}{b^2}\big(1-2g_b(1)+g_{2b}(1)\big) + 2\frac{\rho_r\nu\eta}{ab}\big(1-g_a(1)-g_b(1)+g_{a+b}(1)\big)\,,$$

$$\Delta := \frac{\rho_1\sigma\nu}{a}\big(1-g_a(1)\big) + \frac{\widetilde{\rho}_2\sigma\eta}{b}\big(1-g_b(1)\big)\,, \qquad \widetilde{\rho}_2 := \rho_1\rho_r + \rho_2\sqrt{1-\rho_r^2}\,.$$

Proof. See the Appendix 3.C. □

Before using Lemma 3.2.1 to value EIAs, we introduce the model (35) $-$ (37) under the real-world measure $\mathbb{P}$ and discuss the choice of pricing measure $\mathbb{Q}$.

Model description under the real-world measure. To describe the process under the real-world measure $\mathbb{P}$, retaining a G2++ model under both $\mathbb{P}$ and $\mathbb{Q}$, we introduce constant risk-premia λ_x, λ_y and λ_S for both the interest rate and the equity process. Thus, under $\mathbb{P}$, we have

$$
\begin{aligned}
\mathrm{d}x_t &= a(\lambda_x - x_t)\,\mathrm{d}t + \nu\,\mathrm{d}W_t^{(1)}\,, \\
\mathrm{d}y_t &= b(\lambda_y - y_t)\,\mathrm{d}t + \eta\Big(\rho_r\,\mathrm{d}W_t^{(1)} + \sqrt{1-\rho_r^2}\,\mathrm{d}W_t^{(2)}\Big)\,, \\
\frac{\mathrm{d}S_t}{S_t} &= \big(r_t + \lambda_S\big)\mathrm{d}t + \sigma\Big(\rho_1\,\mathrm{d}W_t^{(1)} + \rho_2\,\mathrm{d}W_t^{(2)} + \sqrt{1-\rho_1^2+\rho_2^2}\,\mathrm{d}W_t^{(3)}\Big)\,.
\end{aligned}
\tag{38}
$$

3.3. Efficient valuation of EIAs

Technically, the stochastic interest rate model introduces dependence between subsequent annual returns of the underlying fund and thus also of the EIA. For this reason, the valuation of cliquet-style EIAs gets numerically more involved than a simple maturity EIA or a cliquet-style EIA with independent annual returns. Section 3.3.1 adapts the scenario-matrix approach by Günther and Hieber (2024) to the G2++ model, while Section 3.3.2 provides Monte-Carlo simulations as a benchmark. In both approaches, we look at the discounted payoff Z_T of the cliquet-style EIA:

$$
Z_T = \mathrm{e}^{-\int_0^T r_s\,\mathrm{d}s} \cdot Y_T\,.
$$

$$
\begin{array}{cc|c|c|c|}
 & & \multicolumn{3}{c}{\overbrace{\hspace{9em}}^{r_t}} \\
 & & x^{(k_1)}, y^{(l_1)} & x^{(k_2)}, y^{(l_2)} & x^{(k_3)}, y^{(l_3)} \\
\cline{3-5}
\multirow{3}{*}{$r_{t-1}\left\{\begin{array}{c}\\ \\ \\ \end{array}\right.$} & x^{(k_1)}, y^{(l_1)} & \mathcal{P}_{1,1}\cdot\phi_{1,1} & \mathcal{P}_{1,2}\cdot\phi_{1,2} & \mathcal{P}_{1,3}\cdot\phi_{1,3} \\
\cline{3-5}
 & x^{(k_2)}, y^{(l_2)} & \mathcal{P}_{2,1}\cdot\phi_{2,1} & \mathcal{P}_{2,2}\cdot\phi_{2,2} & \mathcal{P}_{2,3}\cdot\phi_{2,3} \\
\cline{3-5}
 & x^{(k_3)}, y^{(l_3)} & \mathcal{P}_{3,1}\cdot\phi_{3,1} & \mathcal{P}_{3,2}\cdot\phi_{3,2} & \mathcal{P}_{3,3}\cdot\phi_{3,3} \\
\cline{3-5}
\end{array}
$$

FIGURE 1. Illustration of the scenario matrix $\mathcal{Q}$ as a product of the scenario probabilities $\mathcal{P}_{i,j}$ and the annual discounted payoff conditional on the same scenarios $\phi_{i,j}$.

3.3.1. Scenario-matrix approach.

Under a stochastic interest rate model, consecutive increments of the equity process are not independent. The idea of the scenario-matrix (SM) approach as introduced in Günther and Hieber (2024) is to first condition on the stochastic interest rates at annual time points. Conditional on these rates, subsequent annual returns are independent. We adapt this technique to the G2++ model and condition on the factors x_{t-1}, y_{t-1}, x_t and y_t. We can express the moment-generating function as:

$$
\mathbb{E}\Big[(Z_T)^u\Big] = \mathbb{E}\left[(P_0)^u \cdot \prod_{t=1}^{T} \mathbb{E}\Big[\mathrm{e}^{u\cdot L_{t-1,t}}\,\Big|\, x_{t-1},\, y_{t-1},\, x_t,\, y_t\Big]\right]\,,
\tag{39}
$$

where $L_{t-1,t}$ are defined as the annual discounted log-returns on the EIA account:

$$L_{t-1,t} := \ln\left(\mathrm{e}^{-\int_{t-1}^{t} r_s\,\mathrm{d}s} \cdot \max\left(\mathrm{e}^{g},\ \left(\frac{S_t}{S_{t-1}}\right)^{\alpha}\right)\right) = \max\left(g,\ \alpha\cdot\ln\left(\frac{S_t}{S_{t-1}}\right)\right) - \int_{t-1}^{t} r_s\,\mathrm{d}s\,.$$

The inner conditional expectation

$$\phi_{x_{t-1},y_{t-1},x_t,y_t}(u,t) := \mathbb{E}\left[\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, x_{t-1},\ y_{t-1},\ x_t,\ y_t\right] \tag{40}$$

in (39) can be solved analytically, as shown in Theorem 3.3.2. To approximate the outer expectation over all possible states of $x_{t-1},\ y_{t-1},\ x_t$ and y_t, we split the problem into a finite set of scenarios. Figure 1 illustrates each scenario as one box, which will ultimately be represented by one matrix entry. Each scenario is determined by the state of the interest rate process at times $t-1$ (in rows) and t (in columns), specifically by the combination of the processes x and y. We are now interested in the probabilities $\mathcal{P}_{i,j}$ of each scenario and the corresponding moment-generating function conditional on that scenario $\phi_{i,j}$. Therefore, we first define an equidistantly spaced grid for the process x. We set $x^{(0)} = -\infty$ and $x^{(N+1)} = +\infty$ and choose

$$-\delta\cdot\sqrt{\mathrm{Var}(x_1)} =: x^{(1)} < x^{(2)} < \cdots < x^{(N)} := \delta\cdot\sqrt{\mathrm{Var}(x_1)}\,, \tag{41}$$

with grid spacing $\epsilon_x = (x^{(k)} - x^{(k-1)})/2$ for $k = 2,3,\ldots,N$. We discuss the choice of the parameter δ that defines the width of the discretisation interval in Section 3.4.1. To account for the correlation between the two processes, we define an individual grid $y^{(i,1)},\ldots,y^{(i,N)}$ for every grid point x_i. To this end, we consider the distribution of the bivariate normal distribution (x_t, y_t) from Lemma 3.2.1. We determine the distribution of y_1 conditional on x_1 via its conditional mean $\mu_{y|x}$ and variance $\sigma^2_{y|x}$ as [1]

$$\mu_{y|x} = \frac{\Sigma_{12}}{\Sigma_{11}}\cdot x, \qquad \sigma^2_{y|x} = \Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}}\,.$$

We use this to choose an equidistant spacing for

$$\mu_{y|x^{(i)}} - \delta\cdot\sigma_{y|x^{(i)}} =: y^{(i,1)} < y^{(i,2)} < \cdots < y^{(i,N)} := \mu_{y|x^{(i)}} + \delta\cdot\sigma_{y|x^{(i)}}\,, \tag{42}$$

with grid spacing $\epsilon_y = (y^{(i,k)} - y^{(i,k-1)})/2$ for $k = 2,3,\ldots,N$. Given these grid points, there are N^2 distinct pairs of the form $\left(x^{(k_i)}, y^{(l_i)}\right)$, where $k_i = \lceil i/N\rceil, l_i = (k_i, i \pmod N)$ and $i = 1,2,\ldots,N^2$. Figure 2 illustrates how this two-step approach ensures that the grid points adapt to the joint distribution of x and y. Given these pairs, we can describe scenarios $\mathcal{S}_{ij},\ i,\ j = 1,2,\ldots,N^2$ that take all possible combinations of factors x and y at the beginning and the end of any year t, i.e.

$$\mathcal{S}_{ij} = \left\{x_{t-1}\in\left[x^{(k_i)}\right],\ y_{t-1}\in\left[y^{(l_i)}\right],\ x_t\in\left[x^{(k_j)}\right],\ y_t\in\left[y^{(l_j)}\right]\right\}. \tag{43}$$

[1] Let $X\sim\mathcal{N}(\mu_x,\sigma_x^2)$ and $Y\sim\mathcal{N}(\mu_y,\sigma_y^2)$ be two normally distributed random variables with covariance $\mathrm{Cov}(X,Y) = \sigma_{xy}$. Then we can write Y as

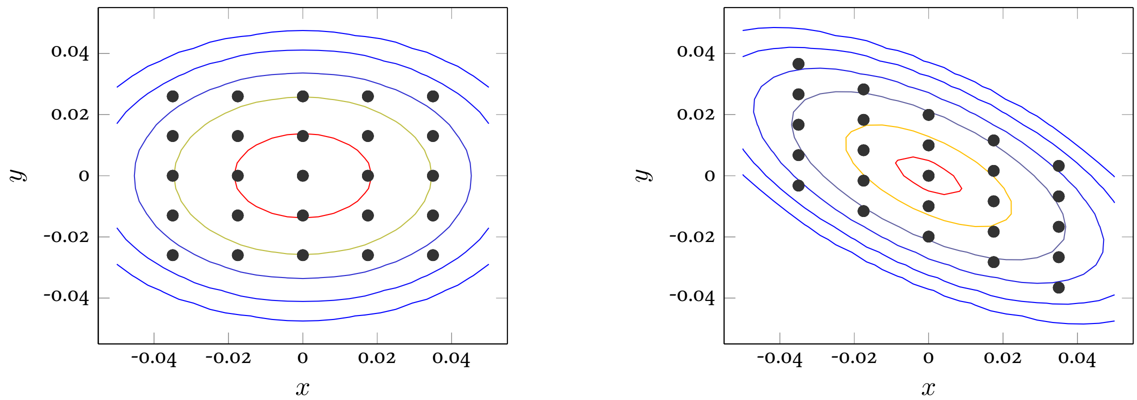
$$Y = \frac{\sigma_{xy}}{\sigma_x^2}\cdot X + \left(\mu_y - \frac{\sigma_{xy}}{\sigma_x^2}\cdot\mu_x + \sqrt{\sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}}\cdot Z\right)\,,$$

where Z is a standard normal random variable that is independent of X. It is straightforward to see, that

$$\mathbb{E}(Y\mid X=x) = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}\cdot(x-\mu_x)\,,$$

$$\mathrm{Var}(Y\mid X=x) = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x}\,.$$

In our context, we have $\mu_y = \mu_x = 0$, $\sigma_x = \mathrm{Var}(x_1) = \Sigma_{11}$, $\sigma_y = \mathrm{Var}(y_1) = \Sigma_{22}$ and $\sigma_{xy} = \mathrm{Cov}(x_1,y_1) = \Sigma_{12}$. For a more general version of this short proof, see (Eaton, 2007, Proposition 3.13).



(A) $\rho_r = 0$ (B) $\rho_r = -0.645$

FIGURE 2. Grid points for $\delta = 2$, $N = 5$ and different levels of ρ_r in black. Coloured lines represent contour levels of the bivariate normal distribution of (x_1, y_1). The remaining parameters are chosen as in Table 1.

Here, $\left[x^{(k_i)}\right]$ describes the surrounding interval $[x^{(k_i-1)} + \epsilon_x,\ x^{(k_i+1)} - \epsilon_x]$ while $\left[y^{(l_i)}\right]$ describes the interval $[y^{(l_i-1)} + \epsilon_y,\ y^{(l_i+1)} - \epsilon_y]$. Thus, all scenarios together span the two-dimensional Euclidean space, that is $\cup_{i,j=1}^{N^2} \mathcal{S}_{ij} = \mathbb{R}^2$.

To each scenario $\mathcal{S}_{ij}$, we assign the probability $\mathbb{P}(\mathcal{S}_{ij})$ of its occurrence. Working with sufficiently small grid lengths ϵ_x and ϵ_y, these probabilities can well be approximated by:

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_{ij}) \;&\approx\; \mathbb{P}\Big(x_t \in \left[x^{(k_j)}\right],\ y_t \in \left[y^{(l_j)}\right] \,\Big|\, x_{t-1} = x^{(k_i)},\ y_{t-1} = y^{(l_i)}\Big) \\
&= \mathbb{P}\Big(x^{(k_j)} - \epsilon_x < x_t < x^{(k_j)} + \epsilon_x, \\
&\qquad\qquad y^{(l_j)} - \epsilon_y < y_t < y^{(l_j)} + \epsilon_y \,\Big|\, x_{t-1} = x^{(k_i)},\ y_{t-1} = y^{(l_i)}\Big)\,. \qquad (44)
\end{aligned}
$$

Using the results on the distribution of the vector (x_t, y_t) in Lemma 3.2.1, we can easily compute these probabilities. Note that the probabilities are independent of time t and we can store them in a matrix $\mathcal{P} = \left(\mathcal{P}_{ij}\right)_{i,j=1,\ldots,N^2}$. The scenario matrix $\mathcal{Q}(u,t) = \left(q_{ij}(u,t)\right)_{i,j=1,\ldots,N^2}$ is an $N^2 \times N^2$ matrix where each entry refers to a scenario $\mathcal{S}_{ij}$. The entry is the product of the probability of this scenario and the conditional expectations $\mathbb{E}[\mathrm{e}^{u \cdot L_{t-1,t}} \mid \mathcal{S}_{ij}]$, i.e.:

$$
\begin{aligned}
q_{ij}(u,t) &= \mathbb{P}(\mathcal{S}_{ij}) \cdot \mathbb{E}\left[\mathrm{e}^{u \cdot L_{t-1,t}} \,\middle|\, \mathcal{S}_{ij}\right] \\
&\approx \mathbb{P}\Big(x_t \in \left[x^{(k_j)}\right],\ y_t \in \left[y^{(l_j)}\right] \,\Big|\, x_{t-1} = x^{(k_i)},\ y_{t-1} = y^{(l_i)}\Big) \\
&\quad \cdot \phi_{x^{(k_i)}, y^{(l_i)}, x^{(k_j)}, y^{(l_j)}}(u,t) \\
&=: \mathcal{P}_{i,j} \cdot \phi_{i,j}\,.
\end{aligned}
$$

Figure 1 illustrates the above approximation as a point-wise matrix product of scenario probabilities $\mathcal{P}_{i,j}$ and the conditional moment-generating functions $\phi_{i,j}$. Given the scenario matrix, it is easy to verify that:

$$
\mathbb{E}\left[(Z_T)^u\right] = (P_0)^u \cdot \mathbf{1}_{i_0}' \left(\prod_{t=1}^{T} \mathcal{Q}(u,t)\right) \mathbf{1}\,,
$$

where $\mathbf{1} = (1, \ldots, 1)'$ is used to sum over all possible ultimate states and $'$ denotes the transposition of a vector. $\mathbf{1}_{i_0}$ is a unit vector that has an entry of 1 in the i_0-th position that indicates the initial condition of the interest rate process, i.e. $x^{(i_0)} = y^{(i_0)} = 0$. In other words, the moments of the discounted payoff Z_T from an EIA are simply given by a matrix multiplication of the annual scenario matrices $\mathcal{Q}(1,t)$, $t = 1,\ldots,T$.

The following Algorithm 3.3.1 serves as a step-by-step guide to compute the expected discounted payoff via the SM approach.

Algorithm 3.3.1. *We aim to compute the expected discounted value* $\mathbb{E}\big[Z_T\big]$ *of an EIA. We initiate the algorithm by different initial states* $(x_0, y_0) = (x^{(k_i)}, y^{(l_i)})$, $i = 1, 2, \ldots, N^2$ *as follows.*

- *We create grids* $x^{(k_i)}, y^{(l_i)}$, $i = 1, \ldots, N$ *as described in* (41) *and* (42).

At time $t = 0$*, the account contains exactly the initial investment. Thus, we set* $A_0 = \mathbf{1}$.

- *We compute the matrix* $\mathcal{P} = \big(\mathcal{P}_{ij}\big)_{i,j=1,\ldots,N^2} = \mathbb{P}(\mathcal{S}_{ij})$ *as defined in* (44)*. This matrix is independent of the annual time step* t.

Now, we iterate over all annual time points until maturity $t = 1, \ldots, T$*:*

- *We compute the matrix* $\mathcal{C}(t)$ *with entries* $\mathcal{C}_{ij}(t) = \phi_{x^{(k_i)},y^{(l_i)},x^{(k_j)},y^{(l_j)}}(1,t)$ *as described in Theorem 3.3.2. We then arrive at the iteration matrix* $\mathcal{Q}(t)$ *via*

$$\mathcal{Q}(t) = \mathcal{P} \odot \mathcal{C}(t)\,, \tag{45}$$

  where $\odot$ *denotes an element-wise product of two matrices.*
- *We update the expected discounted account values* A *as*

$$A_t = \mathcal{Q}(t) \cdot A_{t-1} \tag{46}$$

  and iterate the time t *by 1.*

After T *iterations, we arrive at the expected discounted payoff by picking only the initial state* $(x_0, y_0) = (x^{(i_0)}, y^{(i_0)}) = (0, 0)$*. Thus, we compute*

$$\mathbb{E}[Z_T] = \mathbf{1}_{i_0} \cdot A_T\,.$$

Theorem 3.3.2 provides a closed-form solution for the conditional expected moment-generating function $\phi_{x_{t-1},y_{t-1},x_t,y_t}(u,t)$ in (40).

Theorem 3.3.2 (Conditional moment-generating function). *Assume that the financial market is described by* (35) *and* (37)*. The moment-generating function of the discounted annual log-returns* $L_{t-1,t}$*, conditional on* $x_{t-1}, y_{t-1}, x_t, y_t$*, is given by*

$$\begin{aligned}\phi_{x_{t-1},y_{t-1},x_t,y_t}(u,t) &:= \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, x_{t-1},\, y_{t-1},\, x_t,\, y_t\right] \\ &= \mathrm{e}^{-\bar{\mu}(r_{t-1},r_t)+\frac{\bar{\Sigma}^2}{2}}\left(\mathrm{e}^{g}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}(u,r_{t-1},r_t)}{\widetilde{\Sigma}}\right)\right. \\ &\qquad\left.+\,\mathrm{e}^{\widetilde{\mu}(u,r_{t-1},r_t)\alpha+\frac{1}{2}\alpha^2\widetilde{\Sigma}^2}\,\Phi\left(\frac{-\frac{g}{\alpha}+\widetilde{\mu}(u,r_{t-1},r_t)+\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right)\right),\end{aligned} \tag{47}$$

where Φ *describes the cumulative distribution function of a standard normal random variable and*

$$\begin{aligned}\bar{\mu} &:= \mu_3 + \begin{pmatrix}\Sigma_{13} & \Sigma_{23}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}x_t-\mu_1\\ y_t-\mu_2\end{pmatrix},\\ \bar{\Sigma}^2 &:= \Sigma_{33} - \begin{pmatrix}\Sigma_{13} & \Sigma_{23}\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}\Sigma_{13}\\ \Sigma_{23}\end{pmatrix},\\ \widetilde{\mu} &:= \widehat{\mu}_3 + \begin{pmatrix}\widehat{\Sigma}_{13} & \widehat{\Sigma}_{23}\end{pmatrix}\begin{pmatrix}\widehat{\Sigma}_{11} & \widehat{\Sigma}_{12}\\ \widehat{\Sigma}_{12} & \widehat{\Sigma}_{22}\end{pmatrix}^{-1}\begin{pmatrix}x_t-\widehat{\mu}_1\\ y_t-\widehat{\mu}_2\end{pmatrix},\end{aligned}$$

$$\widetilde{\Sigma}^2 := \widehat{\Sigma}_{33} - \begin{pmatrix} \widehat{\Sigma}_{13} & \widehat{\Sigma}_{23} \end{pmatrix} \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{12} & \widehat{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Sigma}_{13} \\ \widehat{\Sigma}_{23} \end{pmatrix} .$$

The values for Σ_{ij}, $1 \le i, j \le 3$ are taken from Lemma 3.2.1 and we have

$$\widehat{\mu}_1 = \mu_1 - u \cdot \sigma_{13} , \qquad \widehat{\mu}_2 = \mu_2 - u \cdot \sigma_{23} ,$$

$$\widehat{\mu}_3 = \mu_3 + \lambda_S - \frac{1}{2}\sigma^2 - u \cdot \left(\Sigma_{33} + \sigma \left(\rho_1 \frac{\nu}{a}(1 - g_a(1)) + \widetilde{\rho}_2 \frac{\eta}{b}(1 - g_b(1)) \right) \right) ,$$

$$\widehat{\Sigma}_{11} = \Sigma_{11} , \qquad \widehat{\Sigma}_{22} = \Sigma_{22} , \qquad \widehat{\Sigma}_{12} = \Sigma_{12}$$

$$\widehat{\Sigma}_{33} = \sigma^2 + \Sigma_{33} + \sigma \cdot \left(\rho_1 \frac{\nu}{a}(1 - g_a(1)) + \widetilde{\rho}_2 \frac{\eta}{b}(1 - g_b(1)) \right) ,$$

$$\widehat{\Sigma}_{13} = \Sigma_{13} + \sigma \rho_1 \nu g_a(1) , \qquad \widehat{\Sigma}_{23} = \Sigma_{23} + \sigma \widetilde{\rho}_2 \eta g_b(1) .$$

PROOF. See the Appendix 3.D. □

3.3.2. Monte-Carlo Simulations. Monte-Carlo simulations are one of the most common techniques in the risk analysis of insurance products. By repeatedly imitating the behaviour of the financial market they enable us to approximate the distribution of the payoff and to derive risk measures from it.

3.3.2.1. *Euler-Maruyama scheme.* In the G2++ model, Graf and Korn (2020) suggest using the Euler-Maruyama scheme to simulate the financial market on monthly time points $t_i = i/12$. They advise to create a set of standard normally distributed variables $(R_1^{(k)}(i),\ R_2^{(k)}(i),\ R_3^{(k)}(i))$, for $i = 1, \dots, T{\cdot}12$ and $k = 1, \dots, 10\,000$. Based on this realisation of random variables, they simulate for $i = 1, \dots, T \cdot 12$, starting from $x^{(k)}(0) = y^{(k)}(0) = 0$ and $S^{(k)}(0) = 1$:

$$\begin{aligned}
x^{(k)}(t_{i+1}) &= x^{(k)}(t_i) + a \cdot (\lambda_x - x^{(k)}(t_i))/12 + \nu \sqrt{\frac{1}{12}} \cdot R_1^{(k)}(i+1) , \\
y^{(k)}(t_{i+1}) &= y^{(k)}(t_i) + b \cdot (\lambda_y - y^{(k)}(t_i))/12 \\
&\qquad + \eta \sqrt{\frac{1}{12}} \left(\rho_r \cdot R_1^{(k)}(i+1) + \sqrt{1 - \rho_r^2} \cdot R_2^{(k)}(i+1) \right) , \\
S^{(k)}(t_{i+1}) &= S^{(k)}(t_i) \cdot \exp \left(r^{(k)}(t_{i+1}) + \frac{1}{12} \big(\lambda_S - \frac{1}{2}\sigma^2 \big) + \sigma \sqrt{\frac{1}{12}} \Big(\rho_1 \cdot R_1^{(k)}(i+1) \right. \\
&\qquad \left. + \rho_2 \cdot R_2^{(k)}(i+1) + \sqrt{1 - \rho_1^2 - \rho_2^2} \cdot R_3^{(k)}(i+1) \Big) \right) .
\end{aligned}$$

Monthly time steps are necessary to retain a sufficiently accurate approximation of the distribution of the interest rate process. Simulated values for the entire interest rate process can be obtained through $r^{(k)}(t_{i+1}) = x^{(k)}(t_{i+1}) + y^{(k)}(t_{i+1}) + \psi(t_{i+1})$, where the initial yield curve $\psi(t)$ is deterministic.
Approximations for the integrated interest rate are further determined by a Riemann sum:

$$\int_0^{t_i} r_s^{(k)} \,\mathrm{d}s \approx \sum_{j=1}^{i} \frac{r^{(k)}(t_j)}{12} .$$

3.3.2.2. *Exact and unbiased simulation.* The disadvantage of the Euler-Maruyama scheme is the fact that it might still lead to a discretisation bias (although small). However, knowing the exact distribution of the factors x_t and y_t as well as the risky asset S_t in Lemma 3.2.1 allows us to alternatively simulate the processes exactly and without bias on an annual grid. To this end, we would simulate vectors $(\widetilde{R}_1^{(k)}(i), \widetilde{R}_2^{(k)}(i), \widetilde{R}_3^{(k)}(i), \widetilde{R}_4^{(k)}(i))$, for $i = 1, \dots, T$ and $k = 1, \dots, 10\,000$ using the covariance matrix Σ from Lemma 3.2.1 and mean 0 in all dimensions. Note that in contrast to the Euler-Maruyama scheme, the variables $\widetilde{R}^{(k)}(i)$ are now correlated to each other. Based on this, we compute:

$$
\begin{aligned}
x^{(k)}(i+1) &= \mathrm{e}^{-a} \cdot x^{(k)}(i) + \lambda_x(1-\mathrm{e}^{-a}) + \widetilde{R}_1^{(k)}(i+1)\,,\\
y^{(k)}(i+1) &= \mathrm{e}^{-b} \cdot y^{(k)}(i) + \lambda_y(1-\mathrm{e}^{-b}) + \widetilde{R}_2^{(k)}(i+1)\,,\\
\int_0^{i+1} r_s^{(k)}\,\mathrm{d}s &= \int_0^{i} r_s^{(k)}\,\mathrm{d}s + \int_i^{i+1} \psi_s\,\mathrm{d}s + g_a(1)\cdot x^{(k)}(i) + g_b(1)\cdot y^{(k)}(i)\\
&\quad + \lambda_x(1-g_a(1)) + \lambda_y(1-g_b(1)) + \widetilde{R}_3^{(k)}(i+1)\,,\\
S^{(k)}(i+1) &= S^{(k)}(i)\cdot \exp\Bigg(\int_i^{i+1} \psi(s)\,\mathrm{d}s + g_a(1)\cdot x^{(k)}(i) + g_b(1)\cdot \mathrm{e}^{-b}\cdot y^{(k)}(i)\\
&\quad + \lambda_x(1-g_a(1)) + \lambda_y(1-g_b(1)) + \lambda_S - \frac{1}{2}\sigma^2 + \widetilde{R}_3^{(k)}(i+1) + \widetilde{R}_4^{(k)}(i+1)\Bigg)\,.
\end{aligned}
$$

In addition to being more accurate, avoiding monthly time steps reduces the computational complexity by around 85% compared to the Euler-Maryuama scheme.

3.4. Numerical results

TABLE 1. Choice of parameters for the contract and the financial market.

contract	initial capital	P_0	1
	guarantee level	g	0.5%
	participation rate	α	46.5%
	maturity in years	T	25
financial market	Nelsson-Siegel-Svensson yield curve	β_0	0.0154617
		β_1	0.0247537
		β_2	−0.2622674
		β_3	0.2604199
		τ_1	0.0480792
		τ_2	0.0583035
	mean reversion speed of x_t	a	0.3912
	mean reversion speed of y_t	b	0.0785
	volatility of x_t	ν	0.0201
	volatility of y_t	η	0.0135
	correlation between x_t and y_t	ρ_r	−0.6450
	correlation between S_t and x_t	ρ_1	−0.15
	correlation between S_t and y_t	$\widetilde{\rho}_2$	0.15
	volatility of the asset process	σ	0.1

For numerical tests, we implemented our code in the statistical software R on a personal computer with 16 GB RAM and 11th Gen Intel(R) Core(TM) i7-1185G7 @ 3.00GHz. The parameters of the financial market are described in Table 1. We take the estimators for the Nelsson-Siegel-Svensson yield curve as published by Bundesbank (2023) on the 29th of November 2023. They obtain these results by calibrating the parameters to German government bonds. Graf and Korn (2020) suggest fitting the remaining parameters a, b, ν, η and ρ_r that define the dynamics of the interest rate process to market prices of interest rate caps and interest rate swaptions. We assume that the general dynamics of the interest rate processes have not changed significantly in recent years. Therefore, we adopt these parameters from Graf et al. (2021).
Additionally, we assume a small negative dependence of 0.15 between the short-term process x_t and the equity process and an inverse behaviour for the long-term process y_t. The asset volatility σ of 0.1 accounts for the insurer's tendency to use an investment strategy that is less risky than the overall stock market.
We choose the participation rate α so that a contract with guarantee level $g = 0.5\%$ over a 25-year maturity is fair, i.e. it is determined by the premium equivalence $\mathbb{E}[Z_T] = 1$ under the risk-neutral measure $\mathbb{Q}$, when calculating with a deterministic interest rate. In particular, we assume that the interest rate r_t is given by the deterministic part $\psi(t)$. Then we can compute the value at time $t = 0$ of the product as

$$\begin{aligned}\mathbb{E}[Z_T] &= \mathbb{E}\left[\prod_{t=1}^{T} \mathrm{e}^{-\int_{t-1}^{t} r_s \, \mathrm{d}s} \cdot \max\left(\mathrm{e}^{g}, \left(\frac{S_t}{S_{t-1}}\right)^{\alpha}\right)\right] \\ &= \prod_{t=1}^{T} \mathrm{e}^{-\int_{t-1}^{t} \psi(s) \, \mathrm{d}s} \cdot \left(\mathrm{e}^{g}\Phi\left(\frac{\frac{g}{\alpha} - \int_{t-1}^{t} \psi(s) \, \mathrm{d}s + 0.5\sigma^2}{\sigma}\right)\right. \\ &\quad \left. + \, \mathrm{e}^{\alpha\left(\int_{t-1}^{t} \psi(s) \, \mathrm{d}s - 0.5\sigma^2\right) + \frac{1}{2}\alpha^2\sigma^2}\Phi\left(\frac{-\frac{g}{\alpha} + \int_{t-1}^{t} \psi(s) \, \mathrm{d}s - 0.5\sigma^2 + \alpha\sigma^2}{\sigma}\right)\right),\end{aligned}$$

which is a modification of Theorem 3.3.2. Using this closed-form solution, we determine a value of the participation rate α that satisfies $\mathbb{E}_{\mathbb{Q}}[Z_T] = 1$.

3.4.1. Convergence. Figure 3 displays the relative error between the SM method and the true solution, which has been obtained via a Monte-Carlo simulation with 10^6 scenarios. The error is shown concerning the number of grid points N for processes x and y, with both axes in logarithmic scale. The results are presented for varying values δ, which defines the width of the discretisation interval of the processes x_t and y_t (cf. Equation (41)).

The selection of the parameter δ is a critical and challenging task since it needs to balance two phenomena. On the one hand, a small value of δ means that significant parts of the distributions of the short- and long-term interest rate processes x and y are not considered in the computations. On the other hand, a larger value of the width of the discretisation interval δ means that it takes longer for the grid to reach a sufficiently fine resolution. Thus, the approximations take longer to approach the true solution. To summarise, increasing the value of δ can increase the accuracy of our approximation, but requires a higher number of grid points N to reach it.
Empirically, we find that for $\delta = 8.25$ and $\delta = 9$, the SM method quickly converges within the 95%-confidence interval of the Monte-Carlo simulation, achieving an error of less than 0.05%. Note that the increase in the error in the left panel of Figure 3 is a result of the approximations crossing the true solution from above, as shown in the right panel of Figure 3. Therefore, for some values of the number of grid points N, the approximation is very close

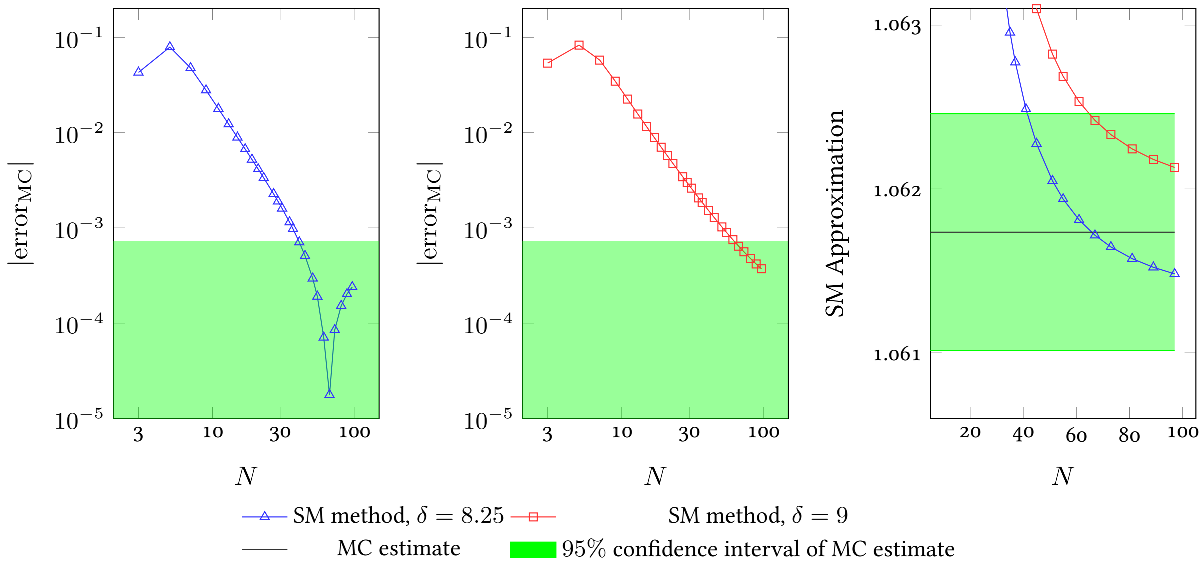


FIGURE 3. Panels 1 and 2: Error as a function of the number of grid points for different values of parameter δ that determines the width of the discretisation interval as in Equation (41). Panel 3: Approximation via scenario-matrix method as a function of the number of grid points N in blue and red compared to the approximation via Monte-Carlo simulation in black. Shaded green: 95%-confidence interval of Monte-Carlo simulation. The parameters are chosen as in Table 1.

to the true solution while it is still converging to a different value. This value is within the confidence interval of the true solution. Notably, the convergence rate is quadratic for all choices of the width of the discretisation interval δ, but larger values lead to a delayed start of the convergence. This confirms the results for a one-factor Vasicek model in Günther and Hieber (2024).

However, the two-factor G2++ model is more computationally demanding. First, fitting an initial yield curve means that the conditional payoff in Theorem 3.3.2 is time-dependent. Therefore, we need to compute a new matrix at each annual time step. Second, computing the transition probabilities in (44) involves the cumulative distribution function of a bivariate normal distribution. Existing techniques for approximating these probabilities are significantly slower than for univariate normal distributions. Therefore, this step consumes most of the computational budget of the SM approach. Finally, the total number of grid points is N^2, resulting in N^4 elements in the scenario matrix $\mathcal{Q}$. The complexity of the algorithm is therefore $\mathcal{O}(N^4)$. The computational efficiency is thus equivalent to that of Monte-Carlo simulations, i.e. for a K-fold increase in the computational budget, the error can be reduced by $\sqrt{K}$.

3.4.2. Approximate normality of payoff. Korn et al. (2017) introduce an approach for incorporating global caps on the terminal payoff. They use the CLT, asserting that for i.i.d. consecutive returns R_i and sufficiently large maturity T, a payoff of the form

$$\sum_{t=1}^{T} g(R_i)$$

is approximately normally distributed with mean $T \cdot \mathbb{E}[g(R_i)]$ and variance $T \cdot \mathrm{Var}[g(R_i)]$, where g is a function that represents the annual guarantee. To approximate the value of a

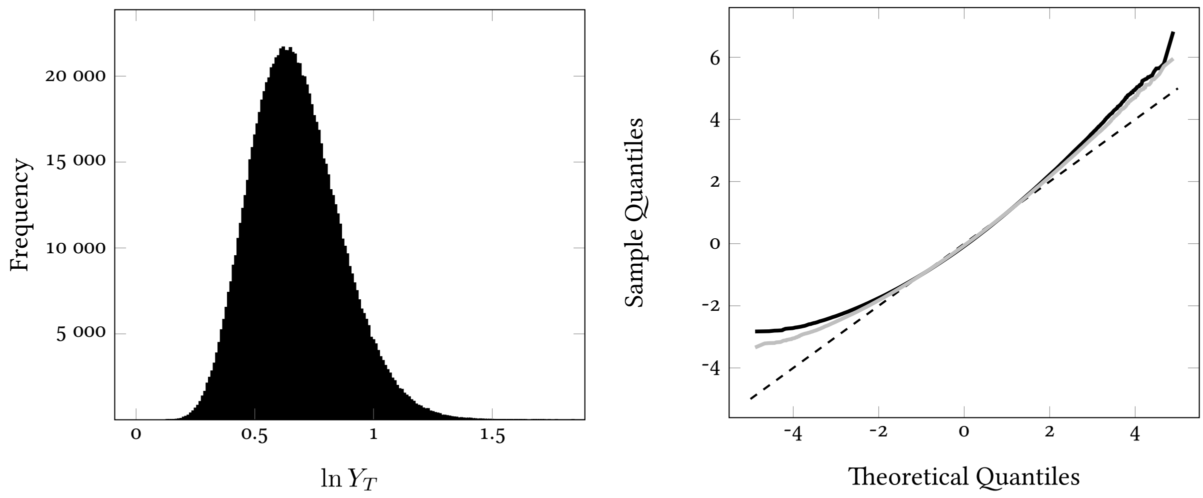


FIGURE 4. Histogram and quantile-quantile plot of the logarithm of the final payoff $\ln Y_T$. Stochastic interest rates in black, constant interest rates in grey. Samples from Monte-Carlo simulations with 10^6 scenarios. The parameters are chosen as in Table 1.

contract with a global cap C, we can thus compute

$$\mathbb{E}\Big[\min\Big(C, \sum_{t=1}^{T} g(R_i)\Big)\Big] \approx \mathbb{E}\big[\min(C, X)\big]$$

for a normal random variable X with the corresponding mean and variance. Korn et al. (2017) suggest to use this approximation as control variates to increase the speed of convergence in Monte-Carlo simulations. We can extend this approach to multiplicative payoffs by exploiting that

$$\prod_{t=1}^{T} g(R_i) = \exp\Big(\sum_{t=1}^{T} \ln(g(R_i))\Big)$$

is approximately log-normal under the same conditions. Considering the definition of our financial market, consecutive returns are neither independent nor identically distributed. However, Lyapunov's version of the CLT does not require identical distributions. Also for very long maturities T, the dependence between the annual returns is negligible.
To assess the distribution of the final payoff $\ln Y_T$, we examine the histograms and quantile-quantile plot presented in Figure 4, contrasting stochastic interest rates from the G2++ model in black with constant interest rates in grey. Note that under constant interest rates, annual returns are independent and satisfy the setting of Korn et al. (2017).
However, the empirical samples from Monte-Carlo simulations do not follow a normal distribution for either interest rate model. Furthermore, this deviation is particularly evident in the tails of the distribution, which are crucial for calculating the value of a global cap. We conclude that a maturity of $T = 25$ years is not sufficiently long for the convergence of the CLT to take effect. The behaviour is even more apparent for the G2++ interest rate model which we can attribute to the correlation between consecutive annual returns. As a result, these findings strongly advise against approximating the terminal payoff by a normal random variable. Therefore, they should only be used with caution as a control variable in Monte-Carlo simulations.

3.5. Conclusion

In this article, we investigate the valuation of equity-indexed annuities under the two-factor G2++ model. By describing the joint distribution of the (integrated) interest rate and the equity process as a multivariate normal distribution in Lemma 3.2.1, we are able to improve the accuracy and computational efficiency of existing Monte-Carlo simulation frameworks.

We further show how the SM approach introduced in Günther and Hieber (2024) can be adapted to the challenges of a two-factor model. This method also relies on our knowledge of the joint distribution of the interest rate and equity process. In addition, it requires a discretisation of the interest rate processes to approximate the annual evolution in different scenarios. We discuss the difficulty of choosing an appropriate discretisation while maintaining good convergence properties. In general, we find empirical evidence of quadratic convergence, confirming previous results. However, a two-factor model significantly increases the computational complexity, which mainly depends on the computation of the cumulative distribution function of a bivariate normal distribution. Therefore, efficient techniques to approximate these probabilities could significantly improve the efficiency of the SM approach.

Lastly, we analyse the idea of Korn et al. (2017) to approximate the terminal (logarithmic) payoff by a normal distribution in more detail. Under stochastic interest rates, the independence assumption of the Central Limit Theorem is violated. While such an approximation can be useful as a control variate in Monte-Carlo simulations, it should not be used to approximate prices.

3.A. Initial yield curve

From (34), we find for the initial market forward yield curve $f^M(0,t)$ that:

$$f^M(0,t) = -\frac{\partial \ln P(0,t)}{\partial t} = \psi(t) - \frac{\partial \ln\left(\mathbb{E}\left[-\mathrm{e}^{\int_s^t x_u + y_u \,\mathrm{d}u}\right]\right)}{\partial t}. \tag{48}$$

Since $\int_s^t x_u + y_u \,\mathrm{d}u$ is normally distributed, the second part in the above sum only depends on its mean and variance, which can be found in (Brigo and Mercurio, 2006, Lemma 4.2.1). This leads us to

$$\psi(t) = f^M(0,t) + \frac{\nu^2}{2}\cdot g_a(t)^2 + \frac{\eta^2}{2}\cdot g_b(t)^2 + \rho\nu\eta\cdot g_a(t)g_b(t)$$

and

$$\int_{t-1}^t \psi(s)\,\mathrm{d}s = \int_{t-1}^t f^M(0,s) + C_t$$

with a time-dependent constant

$$\begin{aligned} C_t = {} & \frac{\nu^2}{2a^2}\left(1 - 2\mathrm{e}^{-a(t-1)}g_a(1) + \mathrm{e}^{-2a(t-1)}g_{2a}(1)\right) \\ & + \frac{\eta^2}{2b^2}\left(1 - 2\mathrm{e}^{-b(t-1)}g_b(1) + \mathrm{e}^{-2b(t-1)}g_{2b}(1)\right) \\ & + \rho\frac{\nu\eta}{ab}\left(1 - \mathrm{e}^{-a(t-1)}g_a(1) - \mathrm{e}^{-b(t-1)}g_b(1) + \mathrm{e}^{-(a+b)(t-1)}g_{a+b}(1)\right). \end{aligned}$$

For $t > 1$[2], we obtain integrated rates:

$$\begin{aligned} \int_{t-1}^t f^M(0,s)\mathrm{d}s = {} & \beta_0 + (\beta_1+\beta_2)\theta_1\left(\ln\left(\frac{t}{t-1}\right) - E_1\left(\frac{t-1}{\theta_1}\right) + E_1\left(\frac{t}{\theta_1}\right)\right) \\ & - \beta_2\theta_1\left(e^{\frac{1}{\theta_1}} - 1\right)e^{-\frac{t}{\theta_1}} \end{aligned}$$

(49)
$$+\beta_3\theta_2\left(\ln\left(\frac{t}{t-1}\right)-E_1\left(\frac{t-1}{\theta_2}\right)+E_1\left(\frac{t}{\theta_2}\right)\left(e^{\frac{1}{\theta_2}}-1\right)e^{-\frac{t}{\theta_2}}\right),$$

introducing the integral $E_1(x) := \int_x^\infty \frac{\mathrm{e}^{-t}}{t}\,\mathrm{d}t$.

3.B. Useful integrals

For the later derivations, let us give some useful integrals related to the function $g_k(s) = \frac{1}{k}(1-e^{-ks})$ for $k>0$. For any $a,b>0$, we find that:

$$\begin{aligned}
I_1(a) &= \int_0^1 \mathrm{e}^{-a(1-s)}\,\mathrm{d}s = \frac{1}{a}\left[\mathrm{e}^{-a(1-s)}\right]_0^1 = \frac{1}{a}\left(1-\mathrm{e}^{-a}\right) = g_a(1)\,,\\
I_2(a,b) &= \int_0^1 \mathrm{e}^{-a(1-s)}\cdot\mathrm{e}^{-b(1-s)}\,\mathrm{d}s = \int_0^1 \mathrm{e}^{-(a+b)(1-s)}\,\mathrm{d}s = I_1(a+b) = g_{a+b}(1)\,,\\
I_3(a) &= \int_0^1 g_a(1-s)\,\mathrm{d}s = \frac{1}{a}\int_0^1 1-\mathrm{e}^{-a(1-s)}\,\mathrm{d}s = \frac{1}{a}\big(1-g_a(1)\big)\,,\\
I_4(a,b) &= \int_0^1 \mathrm{e}^{-a(1-s)}\cdot g_b(1-s)\,\mathrm{d}s = \frac{1}{b}\big(g_a(1)-g_{a+b}(1)\big)\,,\\
I_5(a,b) &= \int_0^1 g_a(1-s)\cdot g_b(1-s)\,\mathrm{d}s = \frac{1}{ab}\big(1-g_a(1)-g_b(1)+g_{a+b}(1)\big)\,.
\end{aligned}$$

We find that:

$$\begin{aligned}
I_4(a,b)+I_4(b,a) &= \frac{1}{b}\big(g_a(1)-g_{a+b}(1)\big)+\frac{1}{a}\big(g_b(1)-g_{a+b}(1)\big)\\
&= \frac{1}{ab}\left(2-\mathrm{e}^{-a}-\mathrm{e}^{-b}\right)-\frac{a+b}{ab(a+b)}\left(1-\mathrm{e}^{-(a+b)}\right)\\
&= \frac{1}{ab}\left(1-\mathrm{e}^{-a}-\mathrm{e}^{-b}+\mathrm{e}^{a+b}\right) = g_a(1)\cdot g_b(1)\,.
\end{aligned}$$

From this, we also get $I_4(a,a) = g_a(1)^2/2$.

3.C. Joint distribution of interest rate and asset process

Proof. To determine closed-form solutions for $x_t,\ y_t,\ \int_{t-1}^t r_s\,\mathrm{d}s$ and $\ln(S_t/S_{t-1})$, we refer to (Brigo and Mercurio, 2006, Section 4.2), and find that

(50) $$x_t = \mathrm{e}^{-a}x_{t-1}+\nu\int_0^1 \mathrm{e}^{-a(1-s)}\mathrm{d}W_s^{(1)}\,,$$

(51) $$y_t = \mathrm{e}^{-b}y_{t-1}+\eta\left(\rho_r\int_0^1 \mathrm{e}^{-b(1-s)}\mathrm{d}W_s^{(1)}+\sqrt{1-\rho_r^2}\int_0^1 \mathrm{e}^{-b(1-s)}\mathrm{d}W_s^{(2)}\right),$$

$$\int_{t-1}^t r_s\,\mathrm{d}s = \int_{t-1}^t \psi(s)\mathrm{d}s + g_a(1)x_{t-1}+g_b(1)y_{t-1}+\nu\int_0^1 g_a(1-s)\mathrm{d}W_s^{(1)}$$

[2] For $t=1$, we find:

$$\begin{aligned}
\int_0^1 f^M(0,s)\,\mathrm{d}s =& \beta_0+(\beta_1+\beta_2)\theta_1\left(-\ln(\theta_1)+E_1\Big(\frac{1}{\theta_1}\Big)+\gamma\right)-\beta_2\theta_1^2e^{-\frac{1}{\theta_1}}\\
&+\beta_3\theta_2\left(\theta_2e^{-\frac{1}{\theta_2}}-\ln(\theta_2)+E_1\Big(\frac{1}{\theta_2}\Big)+\gamma\right),
\end{aligned}$$

where $\gamma\approx 0.577215665$ is the Euler-Mascheroni constant.

$$+\eta\left(\rho_r\int_0^1 g_b(1-s)\mathrm{d}W_s^{(1)}+\sqrt{1-\rho_r^2}\int_0^1 g_b(1-s)\mathrm{d}W_s^{(2)}\right)\,,\tag{52}$$

$$\ln\left(\frac{S_t}{S_{t-1}}\right)=\int_{t-1}^t r_s\,\mathrm{d}s+\lambda_S-\frac{1}{2}\sigma^2+\sigma\bigg(\int_{t-1}^t\rho_1\mathrm{d}W_s^{(1)}+\int_{t-1}^t\rho_2\mathrm{d}W_s^{(2)}$$
$$+\int_{t-1}^t\sqrt{1-\rho_1^2-\rho_2^2}\mathrm{d}W_s^{(3)}\bigg)\,.\tag{53}$$

We introduce $\widetilde{\rho}_2:=\rho_1\rho_r+\rho_2\sqrt{1-\rho_r^2}$ as the correlation between the Brownian motion driving the equity process S_t and the Brownian motion driving the process y_t. We use Itô's isometry and Appendix 3.B to derive the entries of the (symmetric) covariance matrix Σ:

$$\Sigma_{11}:=\mathrm{Var}_{\mathbb{Q}}(x_t)=\nu^2\cdot g_{2a}(1)\,,\qquad \Sigma_{22}:=\mathrm{Var}_{\mathbb{Q}}(y_t)=\eta^2\cdot g_{2b}(1)\,,$$
$$\Sigma_{33}:=\mathrm{Var}_{\mathbb{Q}}\left(\int_{t-1}^t r_s\,\mathrm{d}s\right)=\frac{\nu^2}{a^2}\big(1-2g_a(1)+g_{2a}(1)\big)+\frac{\eta^2}{b^2}\big(1-2g_b(1)+g_{2b}(1)\big)$$
$$+2\frac{\rho_r\nu\eta}{ab}\big(1-g_a(1)-g_b(1)+g_{a+b}(1)\big)\,,$$
$$\Sigma_{12}:=\mathrm{Cov}_{\mathbb{Q}}\left(x_t,\,y_t\right)=\rho_r\nu\eta\cdot g_{a+b}(1)\,,$$
$$\Sigma_{13}:=\mathrm{Cov}_{\mathbb{Q}}\left(x_t,\,\int_{t-1}^t r_s\,\mathrm{d}s\right)=\frac{\nu^2}{2}g_a(1)^2+\frac{\rho_r\nu\eta}{b}\big(g_a(1)-g_{a+b}(1)\big)\,,$$
$$\Sigma_{23}:=\mathrm{Cov}_{\mathbb{Q}}\left(y_t,\,\int_{t-1}^t r_s\,\mathrm{d}s\right)=\frac{\eta^2}{2}g_b(1)^2+\frac{\rho_r\nu\eta}{a}\big(g_b(1)-g_{a+b}(1)\big)\,,$$
$$\Sigma_{14}:=\mathrm{Cov}_{\mathbb{Q}}\left(x_t,\,\ln\left(\frac{S_t}{S_{t-1}}\right)\right)=\Sigma_{13}+\rho_1\sigma\nu\cdot g_a(1)\,,$$
$$\Sigma_{24}:=\mathrm{Cov}_{\mathbb{Q}}\left(y_t,\,\ln\left(\frac{S_t}{S_{t-1}}\right)\right)=\Sigma_{23}+\widetilde{\rho}_2\sigma\eta\cdot g_b(1)\,,$$
$$\Sigma_{34}:=\mathrm{Cov}_{\mathbb{Q}}\left(\int_{t-1}^t r_s\,\mathrm{d}s,\,\ln\left(\frac{S_t}{S_{t-1}}\right)\right)=\Sigma_{33}+\frac{\rho_1\sigma\nu}{a}\big(1-g_a(1)\big)+\frac{\widetilde{\rho}_2\sigma\eta}{b}\big(1-g_b(1)\big)\,,$$
$$\Sigma_{44}:=\mathrm{Var}_{\mathbb{Q}}\left(\ln\left(\frac{S_t}{S_{t-1}}\right)\right)=\Sigma_{34}+\sigma^2\,.$$

□

3.D. Proof of Theorem 3.3.2

Definition 3.D.1. *Using the notation $P_u(t-1,t)=\mathbb{E}_{\mathbb{Q}}\big[\mathrm{e}^{-u\int_{t-1}^t r_s\,\mathrm{d}s}\big]$, the u-scaled t-forward measure $\mathbb{Q}_u^t$ is defined via*

$$\frac{\mathrm{d}\mathbb{Q}_u^t}{\mathrm{d}\mathbb{Q}}=\frac{\mathrm{e}^{-u\int_{t-1}^t r_s\,\mathrm{d}s}}{P_u(t-1,t)}$$
$$=\exp\bigg(-u\int_{t-1}^t \nu g_a(t-s)+\rho_r\eta g_b(t-s)\,\mathrm{d}W_s^{(1)}-u\int_{t-1}^t\sqrt{1-\rho_r^2}\eta g_b(t-s)\,\mathrm{d}W_s^{(2)}$$
$$+\frac{1}{2}u^2\int_{t-1}^t\big(\nu g_a(t-s)+\rho_r\eta g_b(t-s)\big)^2\,\mathrm{d}s+\frac{1}{2}u^2\int_{t-1}^t(1-\rho_r^2)\eta^2 g_b(t-s)^2\,\mathrm{d}s\bigg).\tag{54}$$

Using the t-forward measure $\mathbb{Q}_u^t$ as introduced in (54), we know that according to Girsanov's Theorem in the form of (Shreve, 2010, Theorem 5.4.1),

$$\widetilde{W}_t^{(1)} = W_t^{(1)} + u \cdot \int_{t-1}^{t} \nu \cdot g_a(t-s) + \rho_r \eta \cdot g_b(t-s)\,\mathrm{d}s\,,$$
$$\widetilde{W}_t^{(2)} = W_t^{(2)} + u \cdot \int_{t-1}^{t} \sqrt{1-\rho_r^2}\eta \cdot g_b(t-s)\,\mathrm{d}s$$

and $W_t^{(3)}$ are Brownian motions under $\mathbb{Q}_u^t$. Therefore, we can use this change of measure to decouple the integrated interest rate and the annual returns of the underlying stock as follows:

$$\begin{aligned}\phi_{r_{t-1},r_t}(u) &= P_u(t-1,t) \cdot \mathbb{E}_{\mathbb{Q}}\left[\frac{\mathrm{e}^{-u\int_{t-1}^t r_s\,\mathrm{d}s}}{P_u(t-1,t)}\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, x_{t-1},\, x_t,\, y_{t-1},\, y_t\right]\\
&= P_u(t-1,t)\cdot \mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot L_{t-1,t}}\,\middle|\, x_{t-1},\, x_t,\, y_{t-1},\, y_t\right]\\
&= P_u(t-1,t)\cdot\Bigg(\mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot g}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)\leqslant g\right\}}\,\middle|\, x_{t-1},\, x_t,\, y_{t-1},\, y_t\right]\\
&\qquad + \mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)> g\right\}}\,\middle|\, x_{t-1},\, x_t,\, y_{t-1},\, y_t\right]\Bigg)\,.\end{aligned} \tag{55}$$

The integrated interest $\int_{t-1}^t r_s\,\mathrm{d}s$ and the log return $\ln(S_t/S_{t-1})$, both conditional on the processes x_t and y_t, under the pricing measure $\mathbb{Q}$ and the u-scaled t-forward measure $\mathbb{Q}_u^t$ respectively, are distributed as

$$\int_{t-1}^t r_s\,\mathrm{d}s\,\big|\, x_t, y_t \sim \mathcal{N}_{\mathbb{Q}}(\bar{\mu}, \bar{\Sigma})\,,$$
$$\ln(S_t/S_{t-1})\,\big|\;x_t, y_t \sim \mathcal{N}_{\mathbb{Q}_u^t}(\widetilde{\mu}, \widetilde{\Sigma}^2)\,,$$

where $\mu, \widetilde{\mu}, \Sigma, \widetilde{\Sigma}$ are derived in the following.
We use Lemma 3.2.1 to describe the distribution of x_t, y_t and $\int_{t-1}^t r_s\,\mathrm{d}s$ under $\mathbb{Q}$ as a three-dimensional normal distribution

$$\begin{pmatrix} x_t \\ y_t \\ \int_{t-1}^t r_s\,\mathrm{d}s\end{pmatrix} \sim \mathcal{N}_{\mathbb{Q}}(\mu,\Sigma),\quad \mu := \begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix},\quad \Sigma := \begin{pmatrix}\Sigma_{11} & \Sigma_{12} & \Sigma_{13}\\ \Sigma_{12} & \Sigma_{22} & \Sigma_{23}\\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33}\end{pmatrix}.$$

Then, under $\mathbb{Q}$, Eaton (2007, Proposition 3.13) shows that

$$\int_{t-1}^t r_s\,\mathrm{d}s\,\big|\, x_t, y_t \sim \mathcal{N}_{\mathbb{Q}}(\bar{\mu}, \bar{\Sigma}^2)\,, \tag{56}$$

with

$$\begin{aligned}\bar{\mu} &:= \mu_3 + (\Sigma_{13}\quad \Sigma_{23})\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix} x_t-\mu_1\\ y_t-\mu_2\end{pmatrix}\\
&= \mu_3 + \frac{1}{\Sigma_{11}\Sigma_{22}-\Sigma_{12}^2}\left((\Sigma_{22}\Sigma_{13}-\Sigma_{12}\Sigma_{23})(x_t-\mu_1) + (\Sigma_{11}\Sigma_{23}-\Sigma_{12}\Sigma_{13})(y_t-\mu_2)\right)\,,\\
\bar{\Sigma}^2 &:= \Sigma_{33} - (\Sigma_{13}\quad \Sigma_{23})\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}\Sigma_{13}\\ \Sigma_{23}\end{pmatrix}\\
&= \Sigma_{33} - \frac{1}{\Sigma_{11}\Sigma_{22}-\Sigma_{12}^2}\left(\Sigma_{22}\Sigma_{13}^2 - 2\Sigma_{12}\Sigma_{13}\Sigma_{23} + \Sigma_{11}\Sigma_{23}^2\right)\,.\end{aligned}$$

To determine the distribution of $\ln\left(S_t/S_{t-1}\right)$, given x_t, y_t, under $\mathbb{Q}_u^t$, we calculate

$$\begin{aligned}
x_t &= \mathrm{e}^{-a}x_{t-1} - u\cdot \mathrm{Cov}_{\mathbb{Q}}\left(x_t, \int_{t-1}^t r_s\,\mathrm{d}s\right) + \nu\int_0^1 e^{-a(1-s)}\mathrm{d}\widetilde{W}_s^{(1)} \\
y_t &= \mathrm{e}^{-b}y_{t-1} - u\cdot \mathrm{Cov}_{\mathbb{Q}}\left(y_t, \int_{t-1}^t r_s\,\mathrm{d}s\right) + \eta\bigg(\rho_r\int_0^1 e^{-b(1-s)}\mathrm{d}\widetilde{W}_s^{(1)} \\
&\quad + \sqrt{1-\rho_r^2}\int_0^1 e^{-b(1-s)}\mathrm{d}\widetilde{W}_s^{(2)}\bigg)\,, \\
\int_{t-1}^t r_s\,\mathrm{d}s &= \int_{t-1}^t \psi(s)\mathrm{d}s + g_a(1)x_{t-1} + g_b(1)y_{t-1} - u\cdot\mathrm{Var}_{\mathbb{Q}}\left(\int_{t-1}^t r_s\,\mathrm{d}s\right) \\
&\quad + \nu\int_0^1 g_a(1-s)\mathrm{d}\widetilde{W}_s^{(1)} \\
&\quad + \eta\left(\rho_r\int_0^1 g_b(1-s)\mathrm{d}\widetilde{W}_s^{(1)} + \sqrt{1-\rho_r^2}\int_0^1 g_b(1-s)\mathrm{d}\widetilde{W}_s^{(2)}\right)\,, \\
\ln\left(\frac{S_t}{S_{t-1}}\right) &= \int_{t-1}^t r_s\,\mathrm{d}s + \lambda_S - \frac{1}{2}\sigma^2 - u\cdot\sigma\left(\rho_1\frac{\nu}{a}(1-g_a(1)) + \widetilde{\rho}_2\frac{\eta}{b}(1-g_b(1))\right) \\
&\quad + \sigma\bigg(\int_{t-1}^t \rho_1\mathrm{d}\widetilde{W}_s^{(1)} + \int_{t-1}^t \rho_2\mathrm{d}\widetilde{W}_s^{(2)} + \int_{t-1}^t \sqrt{1-\rho_1^2-\rho_2^2}\mathrm{d}W_s^{(3)}\bigg)\,.
\end{aligned}$$

Again, we can describe the distribution of $\ln\left(S_t/S_{t-1}\right)$ and r_t as a two-dimensional normal distribution with parameters

$$\begin{pmatrix} x_t \\ y_t \\ \int_{t-1}^t r_s\,\mathrm{d}s \end{pmatrix} \sim \mathcal{N}_{\mathbb{Q}_u^t}(\widehat{\mu},\widehat{\Sigma}),\quad \widehat{\mu} := \begin{pmatrix} \widehat{\mu}_1 \\ \widehat{\mu}_2 \\ \widehat{\mu}_3 \end{pmatrix},\quad \widehat{\Sigma} := \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} & \widehat{\Sigma}_{13} \\ \widehat{\Sigma}_{12} & \widehat{\Sigma}_{22} & \widehat{\Sigma}_{23} \\ \widehat{\Sigma}_{13} & \widehat{\Sigma}_{23} & \widehat{\Sigma}_{33} \end{pmatrix},$$

where

$$\begin{aligned}
\widehat{\mu}_1 &= \mu_1 - u\cdot\sigma_{13}\,, \qquad \widehat{\mu}_2 = \mu_2 - u\cdot\sigma_{23}\,, \\
\widehat{\mu}_3 &= \mu_3 + \lambda_S - \frac{1}{2}\sigma^2 - u\cdot\left(\Sigma_{33} + \sigma\left(\rho_1\frac{\nu}{a}(1-g_a(1)) + \widetilde{\rho}_2\frac{\eta}{b}(1-g_b(1))\right)\right)\,, \\
\widehat{\Sigma}_{11} &= \Sigma_{11}\,, \qquad \widehat{\Sigma}_{22} = \Sigma_{22}\,, \qquad \widehat{\Sigma}_{12} = \Sigma_{12} \\
\widehat{\Sigma}_{33} &= \sigma^2 + \Sigma_{33} + \sigma\cdot\left(\rho_1\frac{\nu}{a}(1-g_a(1)) + \widetilde{\rho}_2\frac{\eta}{b}(1-g_b(1))\right)\,, \\
\widehat{\Sigma}_{13} &= \Sigma_{13} + \sigma\rho_1\nu g_a(1)\,, \qquad \widehat{\Sigma}_{23} = \Sigma_{23} + \sigma\widetilde{\rho}_2\eta g_b(1)\,.
\end{aligned}$$

Therefore, we can conclude that, under $\mathbb{Q}_u^t$, we have

$$\ln(S_t/S_{t-1})\,\big|\; x_t, y_t \sim \mathcal{N}(\widetilde{\mu},\widetilde{\Sigma}^2)\,, \tag{57}$$

with

$$\begin{aligned}
\widetilde{\mu} &:= \widehat{\mu}_3 + \begin{pmatrix} \widehat{\Sigma}_{13} & \widehat{\Sigma}_{23} \end{pmatrix} \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{12} & \widehat{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_t - \widehat{\mu}_1 \\ y_t - \widehat{\mu}_2 \end{pmatrix}\,, \\
\widetilde{\Sigma}^2 &:= \widehat{\Sigma}_{33} - \begin{pmatrix} \widehat{\Sigma}_{13} & \widehat{\Sigma}_{23} \end{pmatrix} \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{12} & \widehat{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Sigma}_{13} \\ \widehat{\Sigma}_{23} \end{pmatrix}\,.
\end{aligned}$$

Using (56), we get

$$P_u(t,t-1) = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_{t-1}^{t} r_s\,\mathrm{d}s}\,\middle|\, r_t\right] = \exp\left(-u\mu + \frac{u^2\bar{\Sigma}^2}{2}\right).$$

With (57), we can calculate

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot g}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)\leqslant g\right\}}\,\middle|\, r_t\right] + \mathbb{E}_{\mathbb{Q}_u^t}\left[\mathrm{e}^{u\cdot\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)}\mathbb{1}_{\left\{\alpha\ln\left(\frac{S_t}{S_{t-1}}\right)> g\right\}}\,\middle|\, r_t\right]\\
&= \mathrm{e}^{ug}\mathbb{Q}_u^t\left(\ln\left(\frac{S_t}{S_{t-1}}\right)\leqslant\frac{g}{\alpha}\,\middle|\, r_t\right) + \int_{\frac{g}{\alpha}}^{\infty}\mathrm{e}^{u\alpha x}\sqrt{2\pi\widetilde{\Sigma}^2}^{-1}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\widetilde{\mu}}{\widetilde{\Sigma}}\right)^2}\,\mathrm{d}x\\
&= \mathrm{e}^{ug}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}}{\widetilde{\Sigma}}\right) + \mathrm{e}^{u\alpha\widetilde{\mu}+\frac{1}{2}u^2\alpha^2\widetilde{\Sigma}^2}\Phi\left(-\frac{\frac{g}{\alpha}-\widetilde{\mu}-u\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right).
\end{aligned}$$

Inserting both results into (55), we arrive at

$$\phi_{r_{t-1},r_t}(u) = \mathrm{e}^{-u\bar{\mu}+\frac{u^2\bar{\Sigma}^2}{2}}\left(\mathrm{e}^{ug}\Phi\left(\frac{\frac{g}{\alpha}-\widetilde{\mu}}{\widetilde{\Sigma}}\right) + \mathrm{e}^{u\alpha\widetilde{\mu}+\frac{1}{2}u^2\alpha^2\widetilde{\Sigma}^2}\Phi\left(-\frac{\frac{g}{\alpha}-\widetilde{\mu}-u\alpha\widetilde{\Sigma}^2}{\widetilde{\Sigma}}\right)\right). \tag{58}$$

CHAPTER 4

Efficiently Computing Annuity Conversion Factors via Feed-Forward Neural Networks

This chapter is based on the following article:

M. Aragona, S. Günther and P. Hieber (2025): Efficiently computing annuity conversion factors via feed-forward neural networks. *Annals of Actuarial Science* 19(2): 304-316.

Abstract. Many pension plans and private retirement products contain annuity factors, converting the funds at some future time into lifelong income. In general model settings like for example the Li-Lee mortality model, analytical values for the annuity factors are not available and one has to rely on numerical techniques. Their computation typically requires nested simulations as they depend on the interest rate level and the mortality tables at the time of retirement. We exploit the flexibility and efficiency of feed-forward neural networks to value the annuity factors at the time of retirement. In a numerical study, we compare our deep learning approach to (least-squares) Monte-Carlo (LSMC) which can be represented as a special case of the neural network (NN).

4.1. Introduction

Pension plans and other retirement products are characterised by long time horizons of 30-40 years. In many of them, the account value at retirement is transformed into a lifelong income using annuity factors. Sometimes, these factors are guaranteed when initiating the contract. The interest and longevity risk contained is, however, very difficult to analyse or quantify. During such long time horizons, substantial changes in the interest rate level or life expectancy can fundamentally change the value of annuity conversion factors.

This article proposes a computationally efficient algorithm for computing annuity conversion factors[1] based on deep learning. Feed-forward neural networks predict the factors based on the realisation of the main risk factors at the time of retirement. A naive simulation of annuity factors suffers from the so-called "simulation-within-simulation" or nested simulation problem. Based on the realisation of risk factors at the time of retirement (outer simulation), another simulation has to be carried out to compute the value of the annuity factor (inner simulation). This approach is often unfeasible or inefficient in practical applications. To circumvent the issue of nested simulations, several ideas were proposed to more efficiently carry out the outer simulation step. Feng et al. (2022) suggest reusing simulated random variables

[1]The techniques presented in this paper can similarly be used also for the pricing and risk management of mortality-linked derivatives, see for example Boyer and Stentoft (2013), Boyer and Stentoft (2017) and Bacinello et al. (2021) for examples using LSMC techniques or Cheridito et al. (2020) and Ha and Bauer (2022) for examples using NN. This includes so-called K-options or put spreads (see, for example, Cairns (2011), Li et al. (2019), Feng et al. (2022) and many others).

of the inner step ("green nested simulation") and achieve a significant reduction of the simulation budget. Cairns (2011) and Dowd et al. (2011) approximate the transformed (spot) survival probabilities by a second-order Taylor expansion. This immediately expresses the annuity factor as a function of the realisation of the risk factors at the time of retirement. A similar approach using affine mortality models is Biffis and Millossovich (2006). Closely related, least-squares Monte-Carlo (LSMC) algorithms were used to regress annuity factors on the realisation of the underlying risk factors, see, for example, Bauer et al. (2010), Floryszczak et al. (2016), Ha and Bauer (2022), Bacinello et al. (2024). The link between Taylor expansions and LSMC to mortality-linked products is given by Bacinello et al. (2021), one of the main references of this article. Hainaut and Akbaraly (2023) combine LSMC with a k-means clustering of the output variables (=local least squares). Similar to LSMC, many authors replace the inner simulation by feed-forward neural networks (NN). This seems to be especially promising if computations are based on many risk factors (see the examples of 12-20 risk factors by Krah et al., 2020; Castellani et al., 2021; Jonen et al., 2023). An example of a smaller network size is Fernandez-Arjona (2021).

In this article, we contribute to the comparison between LSMC and NN techniques. In previous research, it turned out that we need complex problems (12-20 risk factors) and complex neural network structures (three layers) to slightly beat a simple LSMC with polynomial basis functions (see, for example, Castellani et al., 2021). Despite its simplicity, an LSMC with only 2-3 degree polynomials performs very well in many applications. Some authors see disadvantages of LSMC and find that LSMC algorithms do "not always produce satisfactory results" (Feng et al., 2022), or "polynomials of higher degrees tend to explode beyond the fitting range" (Krah et al., 2020). The aim of this article is to combine the advantages of LSMC and NN, and work around reservations against them. In the case of deep learning and neural networks, such reservations exist regarding the model interpretability ("black-box" techniques). For this reason, we follow a slightly different path compared to, for example, Castellani et al. (2021). We represent least-squares Monte-Carlo as a small-sized neural network architecture with a squared activation function. In other words, we represent LSMC as a special case of NN. This allows us to start with a neural network architecture with a similar performance to LSMC. From this link between LSMC and NN, we have two advantages: (1) Small neural networks are easy to handle and do not suffer from the curse of dimensionality or the choice of basis functions in LSMC. In NN, the complexity of model calibration depends on the number of layers and units per layer rather than the number of risk factors. (2) We can use the network structure to slightly extend LSMC, adding more neurons or layers to the network structure. With this approach, we can easily trade off between the interpretability of LSMC and the fitting capability of NN.

In our numerical analysis, we restrict ourselves to a specific model framework, a Li-Lee model for stochastic mortality and a two-factor Vasicek model for stochastic interest rates, see Günther and Hieber (2024) for more background and further references on this model. It is straightforward to extend these results to more complex model frameworks. We compare our neural network algorithm to the benchmark LSMC approach (see, for example, Bacinello et al., 2021). We consider different loss functions and show that it is promising to replace the least-squared error with the more general Huber loss function. We show that neural networks achieve similar performance as LSMC techniques while not suffering from the curse of dimensionality. More complex models with many risk factors could therefore be handled more efficiently by neural networks. These can be seen as an alternative to LSMC techniques, providing a more or less automatic and simultaneous selection of interaction terms and model calibration.

The article is organised as follows: Section 4.2 introduces the theoretical model framework and the neural network architecture to approximate the value of annuity factors. Section 4.2 also shows how the neural networks relate to least-squares Monte-Carlo algorithms as introduced in Bacinello et al. (2021). Section 4.3 shows how the neural networks are calibrated. Section 4.4 presents an in-depth numerical analysis, discussing the choice of hyper-parameters for the deep learning approach and comparing the results to benchmark techniques.

4.2. Model framework and neural network architecture

We denote time by $t > 0$ and consider a cohort of individuals aged x at time $t = 0$, retiring in $T \in \mathbb{N}$ years from now. To keep things simple, we consider two processes (interest rate risk by the short rate $\{r_t\}_{t\geq 0}$ and mortality risk by the central death rate $\{m_{x+t,t}\}_{t\geq 0}$) on the probability space $(\Omega, \mathcal{G}, \mathbb{P})$ enlarged with a filtration $\{\mathcal{G}_t\}_{t\geq 0}$ accounting for mortality and financial risks, i.e. $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$. The sequence $\{\mathcal{F}_t\}_{t\geq 0}$ filters the financial information $\{r_t\}_{t\geq 0}$ and the sequence $\{\mathcal{H}_t\}_{t\geq 0}$ describes the mortality uncertainty $\{m_{x+t,t}\}_{t\geq 0}$.

Since we are using a cohort approach, we consider future improvements in the central death rate after the retirement time T and are interested in the realisations of $m_{x+t,T+t}$, $t = 0, 1, \ldots$. This is opposed to a period approach where we would only consider the realised rates at the time of retirement T, i.e. $m_{x+T,T}$ for different ages $x \in \mathbb{N}$. For an extensive comparison of biometric indices under period- and cohort-based approaches see Bacinello et al. (2024). In our cohort-based approach, the future annuity factor at the retirement time T is a conditional expectation given the information available up to the evaluation time T. If the p risk factors $\boldsymbol{Z_t} = (Z_1(t), Z_2(t)\ldots, Z_p(t)), t \geq 0$, are Markov processes, the problem then translates into

$$
\begin{aligned}
a_x(T, \boldsymbol{z}) &= \mathbb{E}\left[\sum_{t=1}^{\omega-x} \mathrm{e}^{-\int_T^{T+t} r_s\,\mathrm{d}s} \mathrm{e}^{-\sum\limits_{s=0}^{t-1} m_{x+T+s,T+s}} \,\Big|\, \boldsymbol{Z_T} = \boldsymbol{z}\right] \\
&\approx \mathbb{E}\left[\sum_{t=1}^{\omega-x} \mathrm{e}^{-\sum\limits_{s=0}^{t-1} r_{T+s} + m_{x+T+s,T+s}} \,\Big|\, \boldsymbol{Z_T} = \boldsymbol{z}\right], \qquad (59)
\end{aligned}
$$

where we assumed a maximal age ω of the individuals and $z = (z_1, z_2, \ldots, z_p) \in \mathbb{R}^p$. Closed-form solutions to equation (59) are rare to obtain in practice, unless the stochastic processes $\{m_{x,t}\}_{t\geq 0}$ and $\{r_t\}_{t\geq 0}$ are independent and of the affine form (Biffis (2005), Luciano and Vigna (2008)).
Note that the realisation $\boldsymbol{z}$ of $\boldsymbol{Z_T}$ is not known at $t = 0$ and is thus a stochastic variable itself. A naive approach to determine risk measures of the future annuity factors $a_x(T, \boldsymbol{z})$ would require expensive nested Monte-Carlo simulations. In the following, we therefore want to provide efficient algorithms to compute $a_x(T,\ \boldsymbol{z})$ as a function of the p risk factors $Z_1, \ldots, Z_p$ that determine the interest rate r_t and the mortality rate $m_{x,t}$. Benchmark algorithms are least squares Monte-Carlo techniques where the conditional expectation in (59) is obtained by comonotonic approximations or Taylor series expansions (Denuit, 2008; Hoedemakers et al., 2005; Dowd et al., 2011; Liu, 2013; Bacinello et al., 2021) rather than nested simulations.

4.2.1. Approximation error.

We want to find an approximating function of $a_x(T,\ \boldsymbol{z})$ using deep learning. Let us first introduce the approximating model $e(\boldsymbol{z}) = e(z_1, \ldots, z_p)$ and define the distance to the true value $Y = a_x(T, \boldsymbol{z})$ by an error function L_δ. In this article, we use the so-called Huber-loss function, which is a generalisation of least-squared error. It is defined as:

$$L_\delta\big(Y,\, e(\boldsymbol{z})\big) = \begin{cases} \frac{1}{2}\big(Y - e(\boldsymbol{z})\big)^2\,, & \text{for } \big|Y - e(\boldsymbol{z})\big| \le \delta\,, \\ \delta\big(\big|Y - e(\boldsymbol{z})\big| - \frac{1}{2}\delta\big)\,, & \text{otherwise.} \end{cases}$$

The limit $\delta \to \infty$ corresponds to the special case of the least-squared error function $L_\infty\big(Y,\, e(\boldsymbol{z})\big) = \frac{1}{2}\big(Y - e(\boldsymbol{z})\big)^2$. For $\delta < \infty$, the Huber loss reduces the squared effect of large deviations $\big|Y - e(\boldsymbol{z})\big|$ and gives more flexibility in the error evaluation. We then determine the model that minimises the loss, that is:

$$\operatorname{argmin}_{e(\boldsymbol{z})} \mathbb{E}\big[L_\delta(Y,\, e(\boldsymbol{z}))\big]\,. \tag{60}$$

To apply our algorithms, we simulate $n \in \mathbb{N}$ scenarios of the tuple $\big(Y^{(c)}, \boldsymbol{z}^{(c)}\big)$, $c = 1, \ldots, n$. This can be used to determine a training sample estimate of (60) as:

$$\mathbb{E}\big[L_\delta\big(Y,\, e(\boldsymbol{z})\big)\big] \approx \frac{1}{n}\sum_{c=1}^{n} L_\delta\big(Y^{(c)},\, e(\boldsymbol{z})\big)\,. \tag{61}$$

4.2.2. Neural network architecture. We consider a neural network architecture with $J \in \mathbb{N}$ layers and $N^{(j)} \in \mathbb{N}$ neurons per layer, for $j = 1, 2, \ldots, J$. Every neuron $k = 1, 2, \ldots, N^{(j)}$ in a layer j is activated by an activation function of the type $\phi_k^{(j)}(x)$. The first layer is defined as:

$$y_k^{(1)} = \phi_k^{(1)}\left(\beta_{0,k}^{(1)} + \beta_{1,k}^{(1)} z_1 + \ldots \beta_{p,k}^{(1)} z_p\right)\,, \tag{62}$$

where we define weights $\beta_{1,k}^{(1)} \in \mathbb{R}$ and intercepts $\beta_{0,k}^{(1)} \in \mathbb{R}$ for $k = 1, 2, \ldots, N^{(1)}$. The activation function $\phi_k^{(1)}$ needs to be chosen appropriately. The outputs $y_k^{(1)}$ are used as inputs for the second layer. In the $j - th$ layer, for $j = 2, \ldots, J - 1$, we have

$$y_k^{(j)} = \phi_k^{(j)}\left(\beta_{0,k}^{(j)} + \sum_{l=1}^{N^{(j-1)}} \beta_{l,k}^{(j)} \cdot y_l^{(j-1)}\right)\,, \tag{63}$$

where, in each layer j, we have weights and intercepts $\beta_{l,k}^{(j)}$, $l = 1, 2, \ldots, N^{(j-1)}$, $k = 1, 2, \ldots, N^{(j)}$. The last layer then provides the final output of the neural network:

$$e(\boldsymbol{z}) = y_1^{(J)} = \phi_1^{(J)}\left(\beta_{0,1}^{(J)} + \sum_{l=1}^{N^{(J)}} \beta_{l,1}^{(J)} \cdot y_l^{(J-1)}\right)\,. \tag{64}$$

4.2.3. Least squares Monte-Carlo as special case. Least-squares Monte-Carlo techniques were originally introduced in Finance by Longstaff and Schwartz (2001) to price American options. The technique involves numerical approximations of functional payoffs through affine combinations of basis functions. Common choices in the literature include *Hermite*, *Laguerre* or *Chebychev* polynomials, however, in many applications, expansions of up to 2nd order polynomials are already able to catch much of the variation in the data (see Longstaff and Schwartz, 2001; Cairns, 2011). In our example, an approximation by M basis functions $e_l\big(\boldsymbol{z}\big)$, $l = 0, 1, \ldots, M - 1$ would lead to the model:

$$e(\boldsymbol{z}) = \sum_{l=0}^{M-1} \beta_l \cdot e_l\big(\boldsymbol{z}\big)\,. \tag{65}$$

Least-squares Monte-Carlo was employed in lattice structures to approximate the continuation value of the American option under a discrete-time framework. Therefore, the basis representation (65) was obtained by regressing over the state variable(s) and basis functions. LSMC is used to compute capital requirements and balance sheet projections in life insurance where computations are otherwise based on a nested simulation approach, see Bauer et al. (2010).

A two-dimensional neural network ($J = 2$) can be parameterised to resemble the basis functions of a least-squares Monte-Carlo algorithm. To illustrate this, Example 4.2.1 gives the polynomial basis functions used by Bacinello et al. (2021) and embeds this as a neural network appropriately choosing activation functions. Note that this example can be generalised to more risk factors and higher degrees of the basis functions.

Example 4.2.1 (Least-squares Monte-Carlo, 3rd order expansion). Let our model comprise two risk factors $\boldsymbol{Z} = (Z_1, Z_2)$, where Z_1 describes the state of the interest rate process r_T and Z_2 describes the state of the central death rate $m_{x+T,T}$.
Consider the expansion (65) and the $M = 7$ basis functions

$$\begin{aligned}
&e_1(Z_1, Z_2) = Z_1\,, && e_2(Z_1, Z_2) = Z_2\,, && e_3(Z_1, Z_2) = Z_1\,,\\
&e_4(Z_1, Z_2) = Z_2^2\,, && e_5(Z_1, Z_2) = Z_1^3\,, && e_6(Z_1, Z_2) = Z_2^3\,,\\
&e_7(Z_1, Z_2) = Z_1 \cdot Z_2\,.
\end{aligned}$$

We can choose weights and activation functions in the first layer of the neural network to represent this with $N^{(1)} = 6$ neurons. In (62), this means that we choose $\phi_1^{(1)}(z) = \phi_2^{(1)}(z) = z$, $\phi_k^{(1)}(z) = z^2$ for $k = 3, 4, 7, 8$ and $\phi_k^{(1)}(z) = z^3$ for $k = 5, 6$. The weights $\beta_{i,k}^{(1)}$, $i = 0, 1, 2$, are then chosen appropriately such that:

$$\begin{aligned}
&y_k^{(1)} = \phi_k^{(1)}(Z_k)\,, && k = 1,\ 2,\\
&y_k^{(1)} = \phi_k^{(1)}(Z_{k-2})\,, && k = 3,\ 4,\\
&y_k^{(1)} = \phi_k^{(1)}(Z_{k-4})\,, && k = 5,\ 6,\\
&y_7^{(1)} - y_8^{(1)} = \phi_7^{(1)}((Z_1 + Z_2)/2) - \phi_8^{(1)}((Z_1 - Z_2)/2) = Z_1 \cdot Z_2\,.
\end{aligned}$$

With this, the output (64) of the neural network with one hidden layer can be chosen to be identical to the least-squares basis function expansion (65).

4.3. Calibrating the neural network

The calibration of the neural network architecture is carried out based on the tuple $\left(Y^{(c)}, \boldsymbol{z}^{(c)}\right)$, for $c = 1, \ldots, n$, as introduced in Section 4.2.1. In our case, these were obtained by simulation from some economic scenario generator.

Algorithm 4.3.1 (Calibrating the neural network).
The calibration then works in the following steps:

- (i) *Determine the outer scenarios* $\boldsymbol{z}^{(c)} = z_1^{(c)}, \ldots, z_p^{(c)}$ *for* $c = 1, \ldots, n$.
- (ii) *For each outer scenario* $c = 1, \ldots, n$, *compute inner scenarios to obtain an estimate for* $Y^{(c)}$ *as described in Section 4.2.1.*
- (iii) *Specify the neural network architecture, that is the number of layers* J, *the number of neurons* $N^{(j)}$ *in each layer and the activation functions.*
- (iv) *Use the back-propagation algorithm and the objective function* (60) *to calculate all weights* $\beta_{l,k}^{(j)}$ *of the neural network.*

As activation functions we will use the *Square* activation function

$$\mathrm{Square}(z) = z^2\,,$$

and the *Exponential Linear Units* ($a \in \mathbb{R}^+$) activation function

$$\mathrm{Elu}(z) = \begin{cases} z\,, & \text{if } z \geq 0\,,\\ a\,(\mathrm{e}^z - 1)\,, & \text{if } z < 0. \end{cases}$$

For more details on possible activation functions and hyperparameter selection, please refer to the appendices of the paper.

4.4. Numerical results and comparison

In this section, we analyse the performance of the neural network architecture to approximate annuity factors, discuss the choice of hyper-parameters and compare to least-squares Monte-Carlo.

4.4.1. Exemplary financial and actuarial model. We use an exemplary financial and actuarial model that is used for solvency assessment and regulation in central Europe, see Günther and Hieber (2024) for a discussion and background of this model. Mortality is modelled by a Li-Lee multi-population model Li and Lee (2005):

$$\ln(m_{x,t}) = \ln(m^{\mathrm{g}}_{x,t}) + \ln(m^{\mathrm{c}}_{x,t})\,, \tag{66}$$

where $m^{\mathrm{g}}_{x,t}$ describes the mortality of a larger general population and $m^{\mathrm{c}}_{x,t}$ defines the country-specific deviation from the general trend. In particular, we have

$$\begin{aligned} \ln(m^{\mathrm{g}}_{x,t}) &= A_x + B_x K_t\,,\\ \ln(m^{\mathrm{c}}_{x,t}) &= a_x + b_x k_t\,, \end{aligned}$$

where $a_x, A_x \in \mathbb{R}$ describe the age-specific pattern of mortality, $k_t, K_t \in \mathbb{R}$ represents the time trend of death rates, $b_x, B_x \in \mathbb{R}$ measures the loading to the particular age groups when the mortality indices k_t, K_t change. The mortality improvement factor K_t evolves as a Brownian motion with drift, while k_t is assumed to behave as an auto-regressive process of order 1. Thus, we assume

$$\begin{aligned} K_t &= K_{t-1} + \gamma + \epsilon_t\,,\\ k_t &= c + d \cdot k_{t-1} + \delta_t\,, \end{aligned}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2_\epsilon)$ and $\delta_t \sim \mathcal{N}(0, \sigma^2_\delta)$ are i.i.d. error terms, while $\gamma, c, d \in \mathbb{R}$ define the two auto-regressive processes.

The model is calibrated on data from North-Western European countries with a specific focus on the deviation of the Swiss subpopulation. The data is available on the Human Mortality Database via https://www.mortality.org. On the financial side, we assume a two-factor Vasicek model for the interest rate:

$$\begin{aligned} r_t &= \psi_t + x_t + y_t\,,\\ \mathrm{d}x_t &= -\alpha\, x_t\,\mathrm{d}t + \nu\,\mathrm{d}W^{(1)}_t\,,\\ \mathrm{d}y_t &= -\beta\, y_t\,\mathrm{d}t + \eta\Big(\rho\,\mathrm{d}W^{(1)}_t + \sqrt{1-\rho^2}\,\mathrm{d}W^{(2)}_t\Big)\,,\\ \psi_t &= f^M(0,t) + \frac{\nu^2}{2}\cdot g_a(t)^2 + \frac{\eta^2}{2}\cdot g_b(t)^2 + \rho\nu\eta\cdot g_a(t)\cdot g_b(t)\,, \end{aligned} \tag{67}$$

where x and y describe short- and long-term deviations from the deterministic function ψ_t which contains the initial forward-yield curve $f^M(0,t)$. The parameters $\alpha > 0, \beta > 0$ describe the mean reversion speed of the two processes x and y, while $\nu > 0, \eta > 0$ define their volatilities. The Brownian motions $W_t^{(1)}$ and $W_t^{(2)}$ are independent and the correlation between the two processes x and y is given by ρ. The initial forward yield curve follows a Nelson-Siegel-Svensson yield curve

$$f^M(0,t) = \beta_0 + \beta_1 \frac{\theta_1}{t}\left(1 - \mathrm{e}^{-\frac{t}{\theta_1}}\right) + \beta_2 \frac{\theta_1}{t}\left(1 - \mathrm{e}^{-\frac{t}{\theta_1}} \frac{t+\theta_1}{\theta_1}\right) + \beta_3 \frac{\theta_2}{t}\left(1 - \mathrm{e}^{-\frac{t}{\theta_1}} \frac{t+\theta_2}{\theta_2}\right).$$

and its parameters are taken from estimations by the German federal bank from November 2023. The remaining parameters are adapted from Graf et al. (2021) and Günther and Hieber (2024), see also the 4.C.

To summarise, the Markov process $\boldsymbol{Z_t}$ describing the central death rate as well as the interest rate process is composed of the four risk factors K_t, k_t, x_t and y_t, that is

$$z_t = (Z_1(t), Z_2(t), Z_3(t), Z_4(t) = (K_t, k_t, x_t, y_t)\,.$$

Note that these risk factors are uniquely determined by the innovations $\epsilon_t, \delta_t, W_t^{(1)}$ and $W_t^{(2)}$. In practice, we replace the risk factors by their corresponding normally-distributed innovations, because most training algorithms for neural networks are optimised for standardised input data.

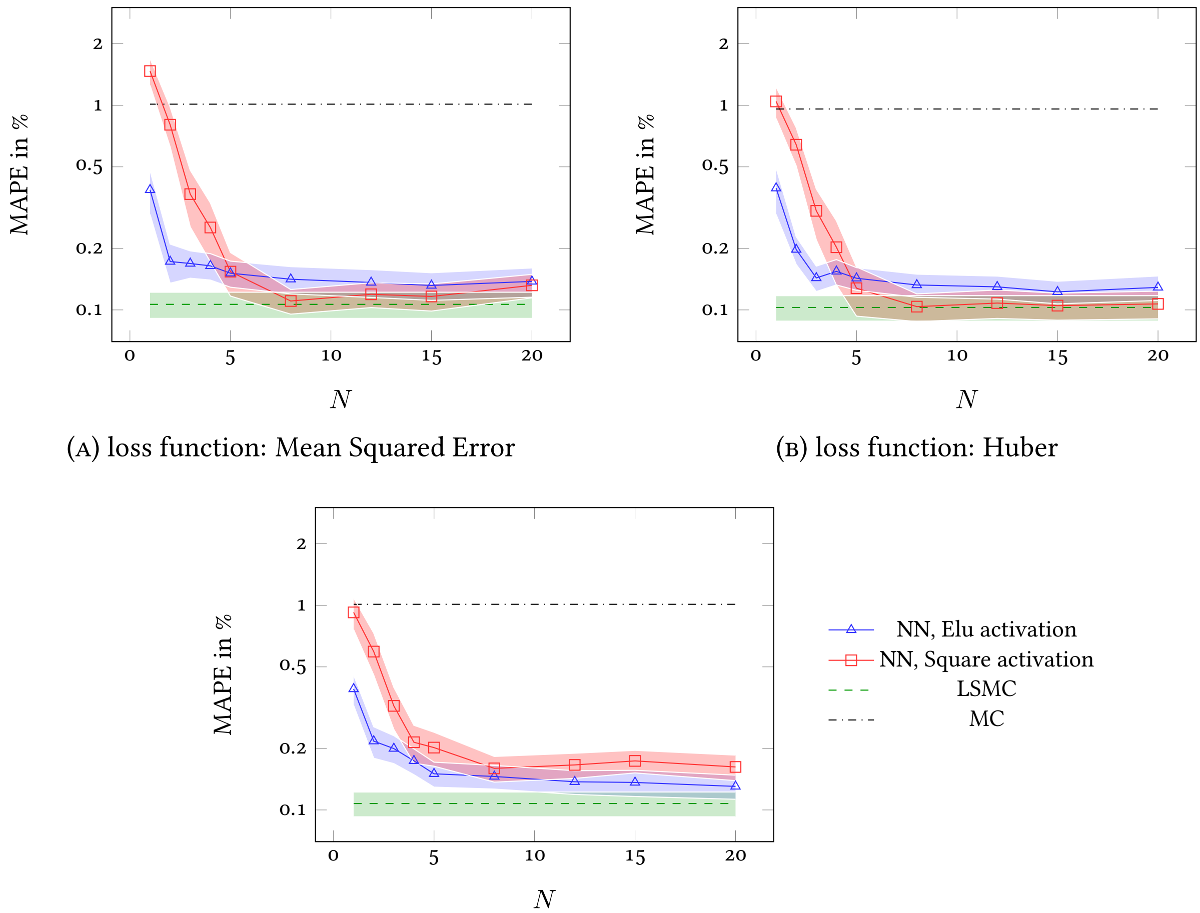
4.4.2. Numerical Results. We compare the results from the neural network (NN) method as described in Algorithm 4.3.1 with those from the LSMC approach. To compute the error of both techniques, we compute a benchmark approximation based on a nested simulation with $100\,000$ outer and $100\,000$ inner simulations.

We run the NN and the LSMC algorithm 100 times to get a reliable estimate and confidence intervals of the average error for both methods. For each run i, we employ a new nested simulation with $1\,000$ outer and 10 inner scenarios, that serve as the input for both algorithms. After fitting the two models, we can rapidly compute the estimated results for a large number of scenarios. To this end, we generate a new set of $1\,000\,000$ outer scenarios $\boldsymbol{z}^{(c)}$ and compute their corresponding estimated annuity values $\hat{Y}^{(c)}$, $c = 1, \ldots, 1\,000\,000$.
We measure the performance of both techniques on the 90^{th} percentile $q_{90\%}^{(i),\text{LSMC/NN}}$ of the annuity values $\hat{Y}^{(c)}$ at $t = 1$. Finally, we compute the *Mean Absolute Percentage Error* (MAPE) over all 100 runs:

$$\text{MAPE}^{\text{LSMC/NN}} = \frac{1}{100} \sum_{i=1}^{100} \left| \frac{q_{90\%}^{(i),\text{LSMC/NN}} - q_{90\%}^{\text{true}}}{q_{90\%}^{\text{true}}} \right|.$$

For the LSMC algorithm, we use polynomials up to degree $d = 2$ as a basis. This is in line with the work by Bacinello et al. (2021), who show that polynomials of higher degrees do not result in an improvement in the accuracy for a problem of this complexity.

We set up the neural network with one hidden layer, in which we use either an exponential linear unit (ELU) or a square as an activation function. The single neuron in the output layer is not transformed by any activation function. Fine-tuning the remaining hyper-parameters of the neural network and the backpropagation algorithm is a delicate task. After several initial experiments, we decided to use a decaying learning rate which is initialised at 0.003 and run the backpropagation algorithm for $E = 100$ epochs at a batch size of $B = 128$. We use the packages `keras` and `tensorflow` for the implementation.



(A) loss function: Mean Squared Error

(B) loss function: Huber

(C) loss function: Mean Absolute Percentage Error

FIGURE 1. Relative error of the MC simulations and the LSMC and Neural Network algorithms as a function of the number of neurons N in the hidden layer for different activation functions and loss functions. Corresponding 95% confidence intervals in shaded colours.

Figure 1 compares the MAPE of the neural network algorithm with that of the LSMC for different loss functions. After 100 epochs and for all loss functions, we can see that with an increasing number of neurons, the error decreases from between 0.4% and 2% for just one neuron in the hidden layer to below 0.2% for 20 neurons. For networks with fewer than five neurons in the hidden layer, the Elu activation function outperforms the Square activation function. On the other hand, for networks with a larger number of neurons, this advantage is less pronounced and even reversed for the Mean Squared Error and Huber loss functions. The strongest results are obtained when combining the quadratic activation function in the hidden layer with a Huber loss function. Here, the confidence intervals of the NN approach and the LSMC method overlap after 8 neurons in the hidden layer.

However, the neural networks are still slightly outperformed by the LSMC approach. It achieves an average MAPE of 0.10%, while the NN approach with 20 neurons and a quadratic activation function remains at 0.11% MAPE. This can be explained by several factors. First, 1 000 input samples and 4 explanatory variables are very small numbers for neural networks. Their strength might come into play for more complex problems with a multitude of variables and possible interaction effects. Secondly, using only 10 inner scenarios leaves us with very noisy estimations of the annuity factors in each outer scenario. Thus, the LSMC approach is at an advantage, because it imposes a well-defined structure on the estimation problem. In contrast to the NN approach, the training for the LSMC only requires a matrix inversion which is computationally fast for a small number of risk factors. Thus, the LSMC algorithm only took an average of 2.3 seconds of computation time, while the neural network algorithm terminated after an average of 21.5 seconds. The number of neurons in the hidden layer does not considerably affect the computation time. The Monte-Carlo simulations used for both algorithms took an average of 0.2 seconds while being around 10 times less accurate.

However, neural networks come very close to the accuracy of the LSMC algorithm and are not subject to the curse of dimensionality. As illustrated in Figure 2, the time spent on the neural network algorithm does not rely significantly on the number of risk factors. The LSMC algorithm, on the other hand, takes around 10 times as long when increasing the number of risk factors from 1 to 4. The number of basis functions m for polynomials of p risk factors up to degree 2 is

$$m = \binom{p+2}{2} = \frac{p^2+3p+2}{2}.$$

Efficient algorithms for inverting matrices of dimension $m \times m$ as described in Pan and Reif (1985) are based on the Newton method which relies on matrix multiplications. Thus, the complexity of a matrix inversion is at least $\mathcal{O}(m^2) = \mathcal{O}(p^4)$. Adding more risk factors, we expect to see a break-even point beyond which it is numerically more expensive to use the LSMC algorithm. To this end, we can interpret a neural network that uses the Mean Squared Error as its loss function and a square activation function as an alternative way to simultaneously calibrate and select the interaction terms of an LSMC approach. Note that computing the annuity factors for n cohorts would require us to apply the neural network algorithm or the LSMC algorithm n times respectively. However, we can reuse the simulated realisations of the random variables ϵ_t, δ_t, $W_t^{(1)}$ and $W_t^{(2)}$ for the Monte-Carlo simulation in both algorithms.

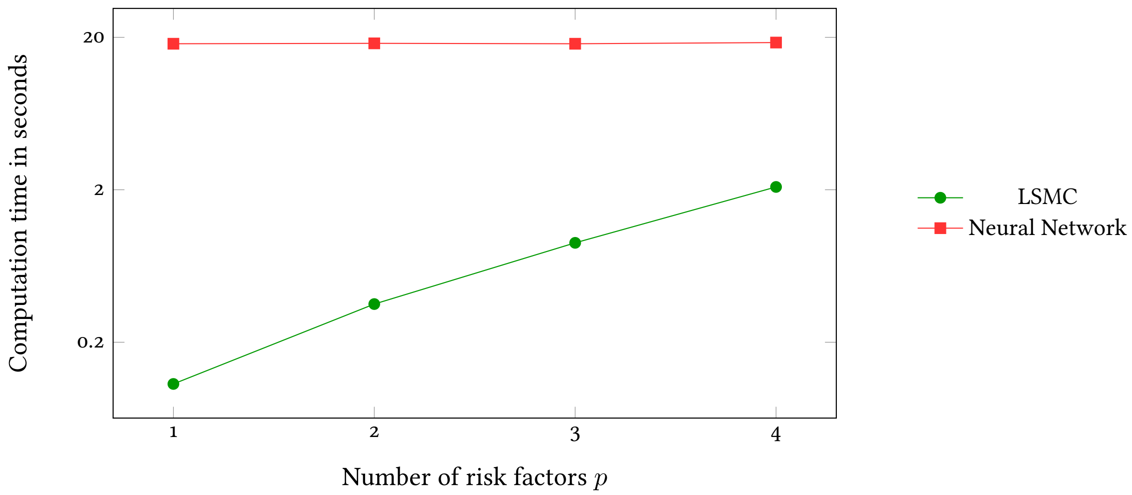


FIGURE 2. The computation time in seconds spent on running the LSMC algorithm and the neural network algorithm for 100 epochs in relation to the number of risk factors p.

4.5. Conclusion

In this article, we investigate deep learning techniques for annuity factor estimation. When the annuitisation of capital is due at a future time point, closed-form analytical solutions are generally not available. We compare feed-forward neural networks to the least-squares Monte-Carlo technique.
We obtain the following results: Surprisingly, neural networks cannot outperform the accuracy of the LSMC algorithm. However, the combination of the squared activation function and Huber loss function can already help small network sizes replicate the LSMC technique. This connection between LSMC techniques and neural networks (as presented in Example 4.2.1) is a promising approach to profit at the same time from the interpretability of the LSMC technique and the almost automatic calibration of the neural network architecture. This calibration avoids the tedious choice of interaction terms for LSMC techniques and does not suffer from the curse of dimensionality if many risk factors are needed.

4.A. Choice of activation functions

Step 3 in the algorithm leaves a lot of freedom on how to specify the neural network architecture.
The annuity factors (59) we want to predict are defined on the positive half-plane $[0, \infty)$. We want to avoid activation functions whose range is bounded from above, such as sigmoid or tanh. Therefore, we consider:

- *Rectified Linear Units*
$$\mathrm{ReLU}(z) = \max\{0,\ z\}\,,$$
- *Gaussian Error Linear Units*
$$\mathrm{Gelu}(z) = z\frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right]\,,$$
- *Exponential Linear Units ($a \in \mathbb{R}^+$)*
$$\mathrm{Elu}(z) = \begin{cases} z\,, & \text{if } z \geq 0\,,\\ a\,(e^z - 1)\,, & \text{if } z < 0\,,\end{cases}$$
- *Scaled Exponential Linear Units ($\tau,\ a \in \mathbb{R}^+$)*
$$\mathrm{Selu}(z) = \begin{cases} \tau \cdot z\,, & \text{if } z > 0\\ \tau \cdot \alpha\,(e^z - 1)\ , & \text{if } z \leq 0\,,\end{cases}$$
- *Softplus*
$$\mathrm{Softp}(z) = \ln(1 + e^z)\,,$$
- *Softsign*
$$\mathrm{Softs}(z) = \left(\frac{z}{|z| + 1}\right)\,,$$
- *Swish*
$$\mathrm{Swish}(z) = \frac{z}{1 + e^{-z}}\,,$$
- *Square*
$$\mathrm{Square}(z) = z^2\,,$$
- *Cubic*
$$\mathrm{Cubic}(z) = z^3\,.$$

4.B. Choice of other hyper-parameters

Apart from the choice of activation function, the neural network architecture allows for more flexibility by a variety of hyper-parameters:

- The network *depth*, that is the number of layers J. Since Cybenko (1989), it is acknowledged that any continuous function can be approximated by a single-layer neural network. Explicit solutions for the construction of the network architecture in terms of depth and hyperparameter selection to achieve a target degree of precision are hard to obtain in practice and only approximation bounds are available (Yarotsky, 2017).
- The layer dimension, $N^{(j)}$. The drawback of improvements based on the layers' dimension is that computational time increases in the number of neurons since those represent additional units to be optimised during the learning phase.
- *Pre-processing* and *regularisation*. It is possible to envisage specific layers which are aimed at flattening the data input, to shuffle and/or induce sparsity by regularisation or dropout techniques, to improve the standardisation efficiency by batch or layer-tailored solutions (Bengio, 2012; Goodfellow et al., 2016).
- The *optimisation routine*. Repeated local optimisation over small batches is generally preferred to a globally implemented routine. Batch size B and epochs E, i.e. the sub-sample sizes over which to optimise and the number of iterations to be performed while implementing the routine are discretionary interventions on the learning routine proficiency. If the representative capability is expected to improve over the number of training epochs up to a convergence rate, for the batch size a trade-off between accuracy and over-fitting generally holds.

4.C. Choice of parameters for mortality and interest rates

The parameters of the financial market are described in Table 1.

TABLE 1. Choice of parameters for the financial market.

Nelsson-Siegel-Svensson yield curve	β_0	0.0154617
	β_1	0.0247537
	β_2	−0.2622674
	β_3	0.2604199
	τ_1	0.0480792
	τ_2	0.0583035
mean reversion speed of x_t	a	0.3912
mean reversion speed of y_t	b	0.0785
volatility of x_t	ν	0.0201
volatility of y_t	η	0.0135
correlation between x_t and y_t	ρ	−0.6450

The general population in the Li-Lee model has been from the countries of Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Luxembourg, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom. Figure 3 presents the calibrated parameters. Future developments of the mortality improvements K_t of the North-Western European population are modelled as

$$K_t = K_{t-1} + \gamma + \epsilon_t \,,$$

with drift $\gamma = -2.0014$ and $\epsilon_t \sim \mathcal{N}(0, 2.4180)$. The future development of the parameter k_t for the Swiss deviation is given by

$$k_t = c + d \cdot k_{t-1} + \delta_t \,,$$

where $c = -0.0620$, $d = 0.9825$ and $\delta_t \sim \mathcal{N}(0, 1.2720)$.

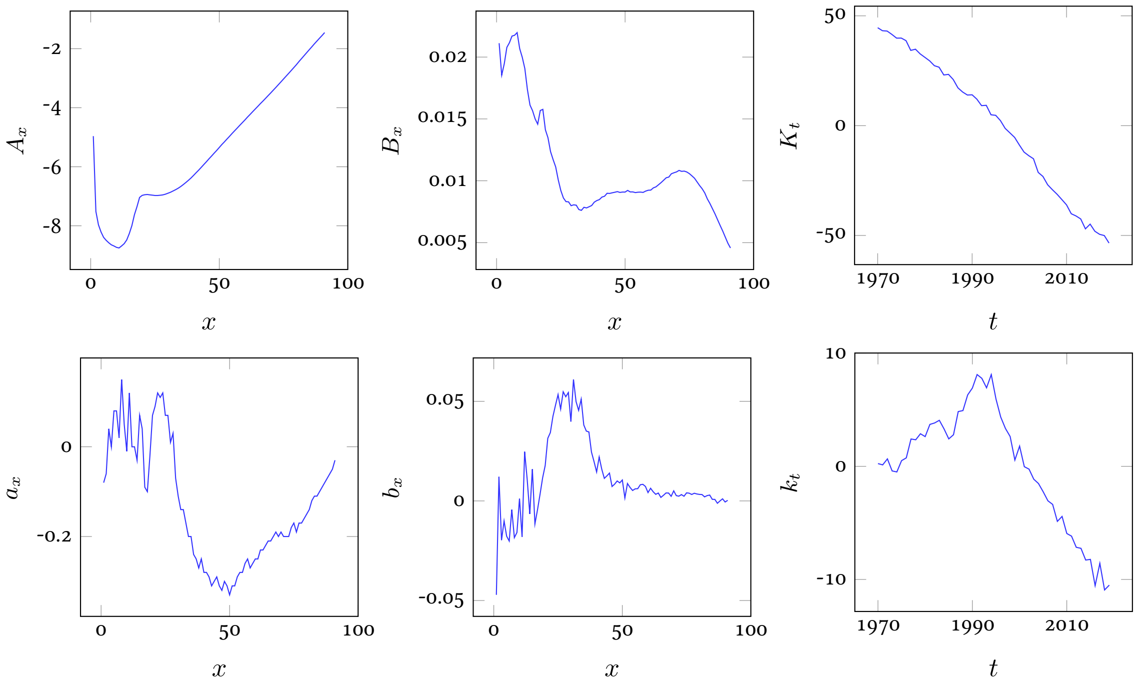


FIGURE 3. *Li-Lee* parameters A_x, B_x, K_t for the North-Western European population and a_x, b_x, k_t for the Swiss deviation.

CHAPTER 5

Modern pension design by refundable income tontines

This chapter is based on the following article[1]:

S. Günther, P. Hieber: Modern pension design by refundable income tontines. *Preprint*.

Abstract. Life and pension annuities often contain a "safety net" in case of an early death in the retirement phase. Within a specified guarantee period, parts of the paid premium are refunded as a bequest. Such a guarantee may increase the attractiveness of retirement products, helping to reduce concerns from (risk-averse) policyholders as well as regulators. We show how to include such a guarantee into a *tontine fund* (also named pooled annuity, collective defined contribution or group-self annuitisation scheme), without leaving any risks with the insurance provider. This product is compared to refundable income annuities, where payments and bequests are fully guaranteed to policyholders. Given the reluctance of states, employers and pension funds to provide long-term guarantees, a pooled annuity fund with premium-refund guarantee may be a way to complement existing state pensions by an additional funded layer.

5.1. Introduction

In many industrialised countries, first-pillar retirement income is about to decline, leading to an increased risk of old-age poverty. To keep sufficient funds also in case of a long life to 110 or 120 years of age, we need to complement the first-pillar retirement income by additional, funded retirement schemes. In this article, we suggest a pooled annuity fund with an optional premium refund guarantee. In contrast to an annuity, a pooled annuity fund shares risks within a pool of annuitants. While retirement income is not guaranteed but volatile, this avoids the need for risk capital and the resulting expenses for reporting and capital provision. Expected payments can thus be 5%-15% higher than standard annuities. In addition, such a fund ensures higher transparency and lower administration costs. The option of a guarantee of contributions in case of an early death in retirement aims to reduce policyholder's concerns regarding annuitisation and investments in retirement products. Such a product may be introduced as an additional obligatory but funded layer in the first or second pillar retirement income. It may also help to set up a voluntary second or third-pillar retirement plan.

The inspiration for this article came from the design of the QSuper annuitisation fund[2] in Australia – a fund with a volume of more than $100 billion operating as a pooled annuity fund. If total payments until death are less than the original contributions made, the difference is paid as a death benefit. This ensures that an investor in QSuper received at least a zero return on total investment. We modify this contract feature and allow that the policyholder

[1]We acknowledge funding from the IAA Life Section by the "Life Section Research Grant" assigned to the project "Modern survivor funds".

[2]After a merger in 2022, the fund is operating under different names.

can initially decide on the percentage $\alpha \in [0\%, 100\%]$ of contributions that are guaranteed by the fund. For $\alpha = 0\%$, this corresponds to a standard pooled annuity fund without premium refund, for $\alpha = 100\%$ this closely resembles the QSuper fund with a premium-refund guarantee. In analogy to a refundable income annuity in the US (see, for example, Milevsky and Salisbury, 2022, for more details), the product may also be named a refundable income tontine[3]. In contrast to our product, payments and premium refunds in a refundable income annuity are guaranteed by an insurance provider (for a cost).

There are several recent articles that combine pooled annuities, tontines and group-self annuitisation schemes (which we will all use as synonyms in this article) with a bequest. While such a bequest is not ideal from an actuary's perspective as it lowers lifetime retirement income, it may be necessary to improve the attractiveness for risk-averse retirees. From a risk management perspective, this benefits also from a natural hedge, pooling life insurance with lifelong retirement income (see Chen and Rach, 2021, for a detailed discussion and references).

Let us give an overview of approaches in the literature: One common idea is to assume a "perfect pool" (Bernhardt and Donnelly, 2019), that is to consider mortality risk as fully diversified in a large pool. Sharing mortality risk then simply leads to an additional return, equal to the individual's mortality rate, on top of the financial return. While this approach simplifies reality, the resulting cash flows are easy to interpret and implement. Further, accepting the "perfect pool" assumption, individual accounts are actuarially fair at all times. Adding a bequest can simply be carried out by combining the pool with a second product, guaranteed by an insurance provider. The case of a deterministic bequest, as a function of time, is considered in Chen and Rach (2021). Another option is to treat the bequest stochastically, via bequest accounts. Milevsky and Salisbury (2025) introduce a Ricatti tontine that returns the initial contributions on average. The motivation for this design are regulatory concerns, implemented in the GuardPath modern tontine product in the US. Bernhardt and Donnelly (2019) and Ng and Nguyen (2024) assume that a fixed percentage of funds is always kept in a bequest account that is – in case of death – returned to the policyholder. Similar to the product discussed in this article, this interpolates between a tontine (if there is no bequest account) and an individual bank account (if all money is in the bequest account).

Different from the purpose of our article and also from the previously mentioned articles, several authors introduce (small) bequests to assure the actuarial fairness of the products (e.g., Donnelly et al., 2014; Denuit, 2019; Hieber and Lucas, 2022) or to compensate an administrator (e.g., Dhaene and Milevsky, 2024). Such death benefits are small (1%-5% of total premium) and do not aim to make the product more appealing to customers.

A crucial concept of tontines is that the systematic mortality risk is not taken over by an insurance company but by the tontine pool. Chen et al. (2019) and Chen and Rach (2021) consider this aspect in their work, while Blake and Burrows (2001) suggest an alternative approach to hedge systematic mortality risk through survivor bonds. Kurtbegu (2018) investigates a similar approach to replicate the longevity risk through the financial market. If the systematic mortality risk cannot be eliminated, it may become necessary to reduce or increase

[3]A literature stream on modern tontines started around 20 years ago, see Denuit et al. (2024) for a recent overview. Milevsky and Salisbury (2015) introduced a "natural tontine" for a closed and homogeneous pool. Sabin (2010) worked out conditions to set up fair distribution plans (without death benefits) among policyholders with different death probabilities and initial investments. Several further articles consider the design and fairness of heterogeneous tontine portfolios (e.g., Donnelly et al., 2013; Denuit et al., 2022; Bernard et al., 2024; Kabuche et al., 2024; Forsyth et al., 2024)).

the consumption levels of retirees. Piggott et al. (2005) account for this by introducing adjustment factors in group-annuitisation schemes.

In this article, we introduce a pooled annuity fund with a premium refund guarantee. The pool has a finite size and pool members may vary in their survival probabilities and contributed amount. The cashflows are defined in discrete time, following a similar setting as in Hieber and Lucas (2022). The actuarial fairness of the scheme allows new members to join at all times, avoiding the disadvantage of closed pools whose size is declining over time. As in Bernhardt and Donnelly (2019) and Ng and Nguyen (2024), funds are split between individual tontine accounts and bequest accounts. The tontine accounts are "at risk" and benefit from mortality credits from deceased pool members while the bequest accounts are individual and paid as bequest in case of death. Contrasting to Bernhardt and Donnelly (2019) and Ng and Nguyen (2024), the bequest account decreases by the annual payout until it is fully depleted. We want to compensate for an early death during the retirement phase but we will not pay any bequest at old age. The aim of this product is to share the risk of getting old (say above the age of 80) and provide sufficient funds to those that are lucky to live until ages 100 to 120. Secondly, we want to keep the product a mutual insurance product where risks are shared within a pool and no insurance provider takes risks. This is in contrast to deterministic death benefits as introduced by Chen and Rach (2021). Extending the approach in Piggott et al. (2005), we model a systematic longevity trend using a Lee-Carter mortality model. Conditional on this general trend, pool members' remaining lifetimes are independent, leading to a component of diversifiable mortality risk that can be fully removed by increasing pool size. We allow the pool to adapt its payoff according to changes in the systematic mortality trend. Such a policy avoids situations where new pool members are significantly advantaged or disadvantaged relative to the existing pool. We use these annual update factors to quantify the systematic mortality risk.

This analysis suggests that members of a tontine pool receive higher benefits than policyholders of an annuity who are subject to risk premia. This can already be achieved in small pools admitting only 100 new members annually. Guarantee levels of up to 100% are affordable at a markup of between 5 to 15% relative to a tontine without guarantee.

The results in this article have links to inter-generational risk sharing in collective defined contribution (CDC) plans. An early work by Allen and Gale (1997) investigates how external shocks can be smoothed by exchanging risks between different generations. Gollier (2008) and Cui et al. (2010) show that CDC plans are welfare-improving and Chen et al. (2023) determine the optimal investment strategy in a CDC plan with multiple overlapping generations.

The article is organised as follows: In Section 5.2, we introduce refundable income annuities and our new product refundable income tontines (also named pooled annuities with premium-refund guarantee). We distinguish two product designs that differ in the way the premium refund option is defined. In Section 5.3, we show how the benefits are adapted following changes in the systematic mortality trend. In Section 5.4, calibrating realistic mortality risk charges in Switzerland, we compare refundable tontines to annuities with premium refund and calculate the cost of the premium refund option.

5.2. Refundable retirement products

We consider an individual who contributes a single premium P_0 at the beginning of the retirement phase to receive an (average) life-long payment of $b \in \mathbb{R}$, paid at the end of each year. Payments cease in case of death. In case of an early death, the individual specifies a guarantee fraction $\alpha \in [0, 1]$. If total payments until death are below $\alpha \cdot P_0$, an additional bequest is paid from the contract. In the following, we introduce two products: A refundable income annuity where payments are guaranteed by the insurance provider (see Section 5.2.1) and a refundable income tontine whose average payments equal the refundable income annuity, see Section 5.2.2. In contrast to the refundable income annuity, risks in this product are shared within a pool of policyholders. There is no insurance provider guaranteeing payments or bequests.

Let us first introduce some basic notation. In the following, we consider a pool of n individuals. Individual j, for $j = 1, 2, \ldots, n$, has initial age x_j. We denote future time (in years) by $t = 1, 2, 3, \ldots$. We consider the following index sets:

- $\mathcal{L}_t$: If $j \in \mathcal{L}_t$, then individual j is alive at time t.
- $\mathcal{D}_t$: If $j \in \mathcal{D}_t$, then individual j died in the t-th year, that is in $(t-1,\, t]$.

Given this, the t-year survival probability of individual j is ${}_tp_{x_j} = \mathbb{E}_0[\mathbb{1}_{\{j \in \mathcal{L}_t\}}]$ and the annual death probability: $q_{x_j+t} = \mathbb{E}_{t-1}[\mathbb{1}_{\{j \in \mathcal{D}_t\}}]$. Throughout the article, $\mathbb{E}_t[\bullet] = \mathbb{E}[\bullet \mid \mathcal{F}_t]$ is a conditional expectation and $\mathcal{F}_t$ is the natural filtration with respect to $\mathcal{L}_t$.

In this article, we focus on longevity risk and ignore financial risks. We assume that a constant annual financial return of δ is credited to the financial accounts. We later comment on extensions to stochastic financial risks.

5.2.1. Refundable income annuities. In the case of a refundable income annuity, an insurance provider guarantees exactly a payment of b, paid at the end of each year. The bequest is immediately obtained: In case of death in year $\tau \in \{1, 2, \ldots\}$, that is in $(\tau - 1, \tau]$, the bequest is $d_j(\tau,\, P_0) = \max(\alpha P_0 - b \cdot (\tau - 1),\, 0)$. If the contract premium is a single premium paid at time $t = 0$, it is a fair premium if it satisfies the following equation:

$$P_0 = \mathbb{E}_0\left[\sum_{t=1}^{\omega - x_j} \mathrm{e}^{-\delta t} \left(\mathbb{1}_{\{j \in \mathcal{L}_t\}} \cdot b + \mathbb{1}_{\{j \in \mathcal{D}_t\}} \cdot d_j(t,\, P_0)\right)\right]$$

$$= \sum_{t=1}^{\omega - x_j} \mathrm{e}^{-\delta t}\, {}_tp_{x_j} \cdot b + \sum_{t=1}^{\omega - x_j + 1} \mathrm{e}^{-\delta t}\, {}_{t-1}p_{x_j} \cdot q_{x_j+t-1} \cdot \max\left(\alpha P_0 - b \cdot (t-1),\, 0\right), \tag{68}$$

where ω is a predetermined maximal age (for example $\omega = 126$). Note that equation (68) contains the premium P_0 on both sides of the equation and has to be solved for a fair premium P_0 numerically. For non-negative interest rates $\delta \geq 0$ and guarantee levels $\alpha \in [0, 1]$ the fixed-point problem has a unique solution. The special case $\alpha = 0$ leads to a regular annuity without a money-back guarantee.

5.2.2. Refundable income tontines. In the case of a refundable income tontine, we assume that the pool of $n \in \mathbb{N}$ individuals shares longevity risks. To manage the refundable income tontine, each individual's funds are split into two accounts:

- The **guarantee account** g_j: This account belongs to each individual and is paid out in case of death. At any time t, the guarantee account $g_j(t, P_0)$ is sufficient to cover the guaranteed bequest. For each member j, it is initialised as $g_j(0, P_0) = \alpha \cdot P_0$ and reduced annually until it is fully depleted.
- The **tontine account** c_j: This account is the "money at risk" that is shared within the pool of individuals. It is lost upon death but is also eligible to receive mortality credits if other pool members die. Its initial account value for individual j is $c_j(0) = P_0 - g_j(0, P_0) = (1 - \alpha) \cdot P_0$.

At the end of every year $t = 1, 2, \ldots$, the interest gained by both accounts gets attributed to the tontine account. Thus, if time t^- refers to the time right before the end of the year, we have:

$$g_j(t^-, P_0) = g_j(t-1, P_0) \tag{69}$$

$$c_j(t^-) = \mathrm{e}^\delta \cdot c_j(t-1) + (\mathrm{e}^\delta - 1) \cdot g_j(t-1, P_0)\,. \tag{70}$$

At the end of the year, payments to survivors as well as bequests are paid out. Apart from this, the tontine accounts c_i (70) of deceased pool members $i \in \mathcal{D}_t$ are redistributed to the tontine accounts c_j of survivors $j \in \mathcal{L}_{t-1}$. We define the pool's total mortality credits as:

$$X(t^-) := \sum_{j \in \mathcal{D}_t} c_j(t^-)\,.$$

We share this amount among all members by assigning them an individual mortality credit $\beta_j(t)$. We need to ensure that $\sum_{j=1}^{n} \beta_j(t) = X(t^-)$ and that we only pay to individuals $j \in \mathcal{L}_{t-1}$. An actuarially fair redistribution (see Theorem 5.2.1) also needs to satisfy that $\mathbb{E}_{t-1}[\beta_j(t)] = q_{x_j+t-1} \cdot c_j(t^-)$, that is that the mortality credits are proportional to the risk taken (indicated by the death probability q_{x_j+t-1}) and the amount invested (indicated by the account value $c_j(t^-)$). For more details and a discussion of different distribution schemes, see for example Denuit and Robert (2021) or Definition 2.1 in Hieber and Lucas (2022). In this article, we want to stick to the simplest distribution rule, a linear risk-sharing rule where

$$\beta_j(t) = \frac{q_{x_j+t-1} \cdot c_j(t^-)}{\sum\limits_{i \in \mathcal{L}_{t-1}} q_{x_i+t-1} \cdot c_i(t^-)} \cdot X(t^-)\,. \tag{71}$$

In addition to the uncertain mortality credits $\beta_j(t)$, the survivors $j \in \mathcal{L}_t$ withdraw a predetermined amount $s_j(t, P_0)$ from their two accounts. If an individual j dies, the amount $\beta_j(t)$ is instead paid to the insurance provider to cover costs related to the pool administration, see also Dhaene and Milevsky (2024). Altogether, this defines the total annual benefit $W_j(t)$ to individual $j \in \mathcal{L}_{t-1}$ as:

$$W_j(t) = \begin{cases} s_j(t, P_0) + \beta_j(t)\,, & \text{if } j \text{ survives time } t, \text{ i.e. } j \in \mathcal{L}_t \\ g_j(t, P_0)\,, & \text{if } j \text{ dies in } (t-1,\, t], \text{ i.e. } j \in \mathcal{D}_t \end{cases}. \tag{72}$$

Having added the financial return to the accounts in (70), we also need to define how the payments, bequests and administration charges affect the individual's guarantee account g_j and tontine account c_j. We assume that each year, the guarantee account is reduced by an amount $\Delta g_j(t, P_0)$, until it is fully depleted. As mortality credits are immediately paid out, the accounts of individual $j \in \mathcal{L}_t$ evolve as:

$$g_j(t, P_0) = g_j(t^-, P_0) - \Delta g_j(t, P_0)\,, \tag{73}$$

$$c_j(t) = c_j(t^-) + \Delta g_j(t, P_0) - s_j(t, P_0)\,. \tag{74}$$

In the following, we introduce possibilities on how to choose the amount $\Delta g_j(t)$. The first one pays the same premium refund as a refundable income annuity, that is we have that the guarantee account value is $g_j(t^-, P_0) = d_j(t, P_0) = \max(\alpha P_0 - b \cdot (t-1), 0)$ for all $t \in \{1, 2, \ldots\}$, see Section 5.2.2.1. Note that this means that in adverse scenarios with fewer deaths than expected, total payments until death may be lower than αP_0. We want to avoid this case in the second choice of $\Delta g_j(t)$. Here, one ensures that at least αP_0 is paid out, see Section 5.2.2.2.

5.2.2.1. *Deterministic death benefit.* If we want to ensure that $g_j(t^-, P_0) = d_j(t, P_0) = \max(\alpha P_0 - b \cdot (t-1),\, 0)$ for all $t \in \{1, 2, \ldots\}$, we need to choose $g_j(0, P_0) = \alpha P_0$ and:

$$\Delta g_j(t, P_0) = d_j(t, P_0) - d_j(t+1, P_0)\,.$$

The advantage of this choice is that it is deterministic and does not depend on mortality risk, pool size, or the death probability of other members in the pool. As long as the guarantee account g_j is not exhausted, it is reduced by an amount of $\Delta g_j(t, P_0)$. The tontine account c_j covers the fixed withdrawal $s_j(t, P_0)$ and increases by the amount $\Delta g_j(t, P_0)$, see (74).
Figure 1 illustrates the development of the total funds for a tontine member entering the pool at age 65 and emphasises the split between guarantee account and tontine account.

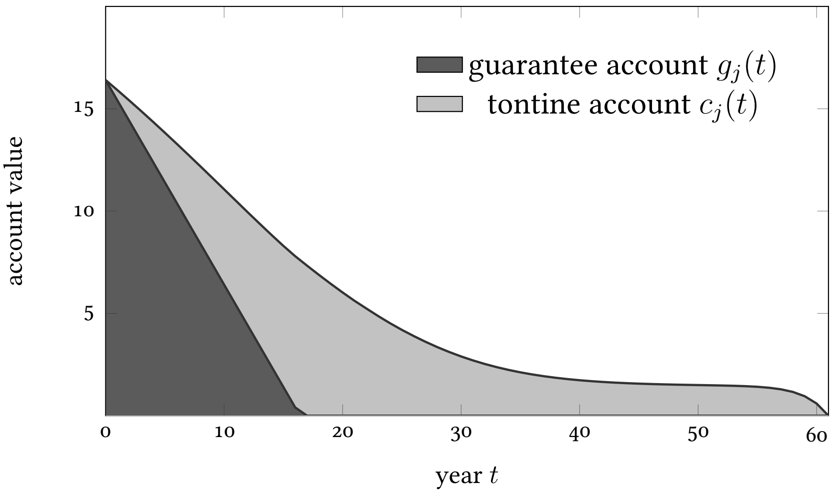


FIGURE 1. Decomposition of the total funds at time t into tontine account $c_j(t)$ and guarantee account $g_j(t)$ for guarantee level $\alpha = 100\%$.

For each individual j, the withdrawal amount is now chosen such that withdrawal plus mortality credits are (on average) equal to the desired annual benefit $b = 1$. A policyholder that survived until year $t-1$ expects to receive:

$$\begin{aligned}
\mathbb{E}_{t-1}\big[W_j(t)\,|\,j \in \mathcal{L}_t\big] &= s_j(t, P_0) + \mathbb{E}_{t-1}\big[\beta_j(t)\,|\,j \in \mathcal{L}_t\big] \\
&= s_j(t, P_0) + \frac{q_{x_j+t-1} \cdot c_j(t^-)}{\sum\limits_{i \in \mathcal{L}_{t-1}} q_{x_i+t-1} \cdot c_i(t^-)} \sum_{i \in \mathcal{L}_{t-1},\, i \neq j} q_{x_i+t-1} \cdot c_i(t^-) \\
&= s_j(t, P_0) + q_{x_j+t-1} \cdot c_j(t^-) \cdot \left(1 - \frac{q_{x_j+t-1} \cdot c_j(t^-)}{\sum\limits_{i \in \mathcal{L}_{t-1}} q_{x_i+t-1} \cdot c_i(t^-)}\right) \\
&\approx s_j(t, P_0) + q_{x_j+t-1} \cdot c_j(t^-)\,, \qquad (75)
\end{aligned}$$

using the linear risk-sharing rule (71). The last approximation "≈" is the key to keep this system analytically tractable. For a homogeneous pool and a pool size of $n = 1\,000$, this inaccuracy amounts to $1/n = 0.1\%$ of the mortality credits. In extreme cases, for example, a heterogeneous pool with a pool size of $n = 3$, this approximation may not be useful, see also Dhaene and Milevsky (2024) or Remark 5.2.4 below. However, we believe that for any reasonable pension pool, the effect of the approximation will be moderate.
Individual j aims for an annual benefit of b at the end of each year. To determine the fixed

annual withdrawals $s_j(t, P_0)$, we want to ensure that the policyholder's expected benefit is equal to b, i.e.

$$\mathbb{E}_{t-1}\big[W_j(t)\,|\,j \in \mathcal{L}_t\big] = b\,. \tag{76}$$

Combining this with equations (70) and (75), we can determine $s_j(t)$ iteratively, starting with $t_{\omega,j} = \omega - x_j$ (see also Hieber and Lucas, 2022). First, recall that $g_j(t^-, P_0) = d_j(t, P_0) = \max(\alpha P_0 - t + 1,\, 0)$ and $\Delta g_j(t, P_0) = d_j(t, P_0) - d_j(t+1, P_0)$ for all $t \in \{1, 2, \ldots, t_{\omega,j} - 1\}$. Using (75), the last withdrawal at terminal age ω is:

$$s_j(t_{\omega,j}, P_0) = \frac{b}{1 + q_{\omega-1}} = c_j(t^-_{\omega,j})\,, \quad \text{and} \quad c_j(t_{\omega,j}) = 0\,. \tag{77}$$

Going backwards in time for $t = t_{\omega,j} - 1,\ t_{\omega,j} - 2,\ \ldots, 1$, we obtain iteratively:

$$c_j(t) = \mathrm{e}^{-\delta} c_j(t + 1^-) - (1 - \mathrm{e}^{-\delta}) \cdot g_j(t, P_0)\,, \tag{78}$$

$$c_j(t^-) = \frac{c_j(t) - \Delta g_j(t, P_0) + b}{1 + q_{x_j+t-1}}\,, \tag{79}$$

$$s_j(t, P_0) = b - q_{x_j+t-1} \cdot c_j(t^-)\,. \tag{80}$$

This iteration can be carried out for each individual in the pool separately. We obtain a deterministic sequence $s_j(1, P_0)$, $s_j(2, P_0)$, …, $s_j(t_{\omega,j}, P_0)$ that defines the deterministic part of the benefit (72) in the case of a deterministic death benefit. Note that the annual withdrawals $s_j(t, P_0)$ depend on the expected benefits b and the death benefits $d_j(t, P_0)$, which again depend on P_0. Thus, we determine the fair premium P_0 of a tontine as the solution to the fixed point problem (68).
Theorem 5.2.1 shows that the introduced refundable income tontine is actuarially fair at all times, that is that the sum of the two account values $c_j(t) + g_j(t, P_0)$ is equal to the expected discounted value of future payments plus administration charges. Recall that the administration charges were equal to the mortality credits $\beta_j(t)$ of a deceased person.

THEOREM 5.2.1 (Actuarial fairness). *At any time $t = 0, 1, 2, \ldots, \omega - x_j$, the combined account value $g_j(t, P_0) + c_j(t)$ is actuarially fair. That is, it satisfies condition*

$$c_j(t) + g_j(t, P_0) = \underbrace{\mathbb{E}_t\left[\sum_{s=t+1}^{\omega-x_j} \mathrm{e}^{-\delta(s-t)} \cdot W_j(s)\right]}_{\textit{present value of benefits}} + \underbrace{\mathbb{E}_t\left[\sum_{s=t+1}^{\omega-x_j} \mathrm{e}^{-\delta(s-t)} \cdot \beta_j(s) \cdot \mathbb{1}_{\{j \in \mathcal{D}_s\}}\right]}_{\textit{present value of administration charges}}\,. \tag{81}$$

PROOF. See Appendix 5.A.1. □

The special case $\alpha = 0$ leads to a (non-refundable) tontine, a modern tontine as introduced in Hieber and Lucas (2022). Let us illustrate this section and Theorem 5.2.1 by Example 5.2.2.

Example 5.2.2 (Numerical example). Let us consider an $x_1 = 65$ year-old male Swiss individual, whose mortality table is obtained from the Human Mortality Database on Swiss data (see the numerical section for more details). The financial return is set to $\delta = 0.03$ and he aims for a lifelong benefit of $b = 1$ with a guarantee of $\alpha = 100\%$ on the paid single premium P_0. We set $d_j(t, P_0) = \max(\alpha P_0 - t + 1,\, 0)$ and $\Delta g_j(t, P_0) = d_j(t, P_0) - d_j(t + 1, P_0)$. The iteration (77)–(80) returns this individual's annual withdrawal schedule $s_j(1, P_0)$, $s_j(2, P_0)$, …, $s_j(t_{\omega,j}, P_0)$ as a function of the initial contribution P_0. We solve the fixed-point iteration

(68) to obtain the initial premium $P_0 = 16.4035$. Given that the initial split on the two accounts is $c_j(0) + g_j(0) = (1-\alpha)P_0 + \alpha P_0 = P_0 = 16.4035$, we obtain a decomposition of this premium in the present value of benefits plus the present value of administration charges. The administration charges can be approximated by:

$$\mathbb{E}\left[\sum_{t=1}^{\omega-x_j} \mathrm{e}^{-\delta t} \cdot \beta_j(t) \cdot \mathbb{1}_{\{j \in \mathcal{D}_t\}}\right] \approx \sum_{t=1}^{\omega-x_j} \mathrm{e}^{-\delta t} \cdot \big(1 - s_j(t, P_0)\big) \cdot {}_{t-1}p_{x_j} \cdot q_{x_j+t-1} \approx 0.2051\,.$$

Relative to the total premium, these administration charges amount to 1.25% of the premium. Figure 2 illustrates the withdrawal schedule $s_j(t, P_0)$ and the expected administration charges in this example. Initially, the death probabilities q_{x_j+t-1} grow faster than the tontine account $c_j(t)$ shrinks. Thus, the expected mortality credits $\mathbb{E}_{t-1}[\beta_j(t)] \approx q_{x_j+t-1} \cdot c_j(t)$ increase until the trend is reversed at the age of around 115 years. Note that the withdrawal schedule and expected benefit scale with the investment. If the male individual invests $100\,000$ in the tontine fund at age $x_1 = 65$, this person receives an average annual lifelong benefit of $100\,000/P_0 \approx 6\,096$ by withdrawing $6\,096 \cdot s_j(t)$ in year t. Note this is equal to buying $6\,096$ "shares" of a refundable income tontine, where each share pays (on average) 1 for life. The withdrawals $s_j(t, P_0)$ then also scale by the number of shares.

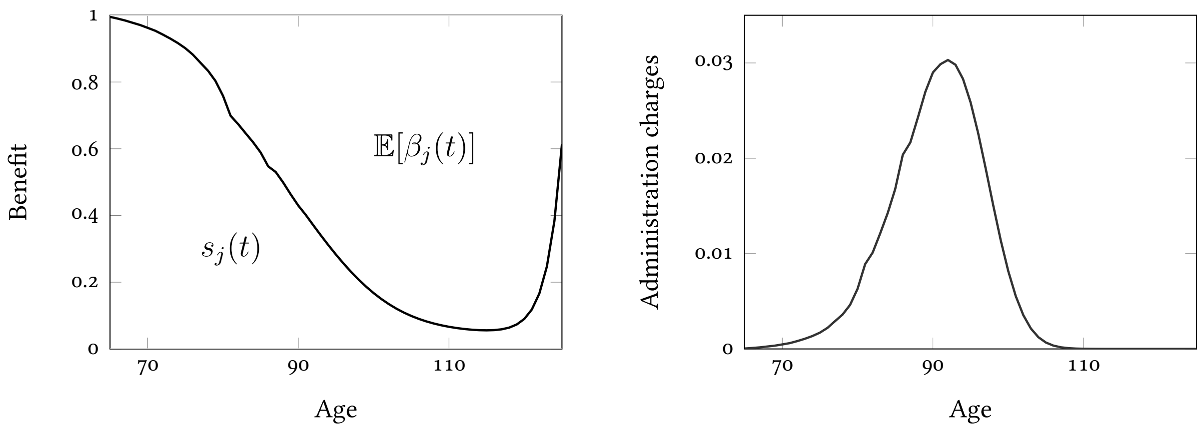


FIGURE 2. Left Panel: Separation of expected annual benefit into fixed withdrawals $s_j(t, P_0)$ and expected mortality credits $\mathbb{E}[\beta_j(t)]$ for contract specifications in Example 5.2.2. Right Panel: Expected administration charges over the contract duration.

Theorem 5.2.3 shows that if we do not consider systematic mortality risk and pool member's remaining lifetimes are independent, the benefit of a refundable income tontine approaches that of the refundable income annuity if the pool size approaches infinity ("perfect pool").

THEOREM 5.2.3 (Convergence for independent lifetimes).
Assume that all pool members have independent remaining lifetimes. Let all account values be such that $c_j(t^-) \in [c_{\min}, c_{\max}]$, *where* $0 < c_{\min} < c_{\max}$.

As the pool size n *increases, the benefit of a refundable income tontine almost surely converges to that of a refundable income annuity.*

PROOF. See Appendix 5.A.2. $\square$

Note that the condition $c_j(t^-) \in [c_{\min}, c_{\max}]$ ensures that the share of capital held by any pool member becomes negligible as $n \to \infty$. It can be enforced by setting arbitrary upper and lower bounds for the initial capital P_0 invested by any pool member.

Remark 5.2.4 (Modifications and generalisations). *To not overload notation, we wrote this section in terms of a constant benefit b for life and chose a fixed death benefit function $d_j(t, P_0)$. It is straightforward to extend these results to different death benefit functions $d_j(t, P_0)$ or individually increasing or decreasing benefits $b_j(t)$, see also Hieber and Lucas (2022). Increasing average benefits may be desirable for a risk-averse policyholder, see for example, Chen et al. (2019) for an illustration using power utility preferences.*

5.2.2.2. ***Stochastic death benefit.*** Now, we want to consider a death benefit that guarantees that the total payout of the tontine is at least equal to a portion α of the initial investment P_0. An individual $j \in \mathcal{D}_t$, that is a person that survived until time $t-1$ but died in $(t-1, t]$ receives a death benefit of:

$$g_j(t^-, P_0) = \max\left(\alpha P_0 - \sum_{s=1}^{t-1} W_j(s), 0\right) = \max\left(\alpha P_0 - \sum_{s=1}^{t-1} s_j(s, P_0) + \beta_j(s), 0\right), \tag{82}$$

ensuring that total payments sum up to at least αP_0. The death benefit (82) is stochastic, leading to a higher complexity in the product's analysis. Computations in this case have to mostly rely on simulation.

As long as the guarantee account is not exhausted ($g_j(t^-, P_0) > 0$), it is reduced by the total payment $\Delta d_j(t, P_0) = \min\big(g_j(t^-, P_0),\, W_j(t)\big)$, while the mortality credits $\beta_j(t)$ are credited to the tontine account. After the guaranteed amount has been paid out to the policyholder, only the tontine account remains active. In particular, adjustments for an individual $j \in \mathcal{L}_t$ follow (73)–(74).

This again ensures that at every time step t, the combined account values are reduced by $s_j(t, P_0)$. The guarantee account g_j is fully depleted as soon as total payments exceed αP_0.

In contrast to Section 5.2.2.1, both accounts c_j and g_j are stochastic as they depend also on the mortality credits $\beta_j(t)$. For the c_j-account, we obtain the following expression:

$$c_j(t) = \underbrace{P_0 \cdot e^{\delta t} - \sum_{s=1}^{t} s_j(s, P_0) \cdot e^{\delta(t-s)}}_{\text{total funds for policyholder } j} - \underbrace{\max\left(\alpha P_0 - \sum_{s=1}^{t-1} s_j(s, P_0) + \beta_j(s), 0\right)}_{\text{value of the guarantee account } g_j}. \tag{83}$$

To apply (76) and ensure that expected benefits at all times equal b, we need to compute the expected mortality credits and thus the expected account value $c_j(t)$. Using the risk-sharing rule (71), this means:

$$\mathbb{E}_{t-1}\big[\beta_j(t)\big] = q_{x_j+t-1} \cdot e^{\delta} \cdot \mathbb{E}_{t-1}\big[c_j(t-1)\big]\,.$$

Solving a backwards iteration similar to Section 5.2.2.1 poses a significant challenge as $\mathbb{E}_{t-1}\big[c_j(t-1)\big]$ can no longer be determined analytically, but depends also on other pool members and their individual contribution. However, especially for large pool sizes, the total payout $\sum_{s=1}^{t-1} s_j(s, P_0) + \beta_j(s)$ will be close to $b \cdot (t-1)$, ensuring that the death benefit (82) is close to the death benefit $\max(\alpha P_0 - b \cdot (t+1),\, 0)$ from Section 5.2.2.1. To avoid these numerical challenges, we suggest to use the same withdrawals $s_j(1, P_0)$, $s_j(2, P_0)$, ..., $s_j(t_{\omega,j}, P_0)$ that result from the iteration (77)–(80). Still, given this withdrawal schedule, the two types of guarantee in Section 5.2.2.1 and this section are different. In the numerical section, we try to quantify this difference.

5.3. Systematic mortality risk and regular update of the mortality table

To adequately analyse the policyholder's risks of a tontine, it is necessary to take into account the uncertainty in the mortality tables, that is that average death counts $\mathbb{E}[\mathbb{1}_{\{j\in\mathcal{D}_t\}}]$ do not agree with the probabilities q_{x_j+t}. Therefore, we model the evolution of future mortality rates $m_{x,t}$ for an x-year old in calendar year t stochastically. The procedures described in this section could be applied to any stochastic mortality model. In this article, however, we choose the Lee-Carter model as introduced in Lee and Carter (1992), i.e.

$$\ln(m_{x,t}) = a_x + b_x \cdot \kappa_t \, ,$$

where a_x describes the average mortality level for individuals of age x, κ_t denotes future mortality improvements in year t and b_x defines the relevance of the calendar year effect κ_t for different ages x. The mortality improvement factor κ_t evolves as a Brownian motion with drift

$$\kappa_{t+1} = \kappa_t + \theta + \xi_t \, , \tag{84}$$

where $\theta \in \mathbb{R}$ is the drift term and $\xi_t \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed (i.i.d.) normal random variables with variance $\sigma^2 > 0$.

Algorithm 1 Updating the mortality table

Input: κ_0, a_x, b_x, θ, σ^2.
- **for** $u = 1, 2, \ldots, \omega - x_j$ **do**
  - Draw $\xi_{u-1} \sim \mathcal{N}(0, \sigma^2)$
  - Update $\kappa_u = \kappa_{u-1} + \theta + \xi_{u-1}$
  - **if** $x \leq 90$ **then**
    - Calculate $q^{(u)}_{x_j+t-1} = 1 - \mathrm{e}^{-a_x - b_x\cdot(\kappa_u+\theta\cdot(t-u))}$.
  - **else**
    - Extrapolate $q^{(u)}_{x_j+t-1}$ using Kannisto's method (see Remark 5.3.1).
  - **end if**
- **end for**

Based on the Lee-Carter mortality rates, we compute the death probabilities q_{x_j+t-1}, used in Sections 5.2.2.1 and 5.2.2.2 as:

$$q_{x_j+t-1} = 1 - \mathbb{E}\left[\mathrm{e}^{-m_{x_j+t-1,t-1}}\right] \approx 1 - \mathrm{e}^{-a_x - b_x\cdot(\kappa_0+\theta\cdot t)} \, .$$

There is no closed-form solution to compute the above expectation. However, approximating the expected value by its median proves to be a sufficiently close approximation in practice. In year u after the initiation of the contract, the realisation of the mortality improvement factor κ_t leads to a new best estimate for the cohort life tables. For each future year $u = 1, 2, 3, \ldots$, we compute:

$$q^{(u)}_{x_j+t-1} \approx 1 - \mathbb{E}_u\left[\mathrm{e}^{-m_{x_j+t-1,t-1}}\right] \approx 1 - \mathrm{e}^{-a_x - b_x\cdot(\kappa_u+\theta\cdot(t-u))} \, . \tag{85}$$

We assume that there is agreement in the tontine fund to update the underlying mortality table in the pool each year $u = 1, 2, \ldots$. This update has the advantage that the mortality risk sharing within the pool always follows up-to-date mortality tables, providing fair conditions also for new entrants. Algorithm 1 gives the details for this update. In practice, it is possible to just recalibrate the Lee-Carter model each year to also obtain the death probability $q^{(u)}_{x_j+t-1}$.

Remark 5.3.1 (Extrapolation for high ages). *For high ages $x > 90$, available mortality observations are limited. Therefore, the estimation of Lee-Carter parameters is more volatile. That is why we may also rely on a method, introduced by Kannisto (1994), who observed that for a fixed year t, the logit of the mortality rates $m_{x,t}$ becomes linear in the age for the highest ages, i.e.*

$$\ln\left(\frac{m_{x,t}}{1-m_{x,t}}\right) \approx c_t \cdot x + d_t\,.$$

For each calendar year t, we use this relationship to extrapolate the observed mortality rates from ages up to 90 years to higher ages. For $x > 90$, we get:

$$\widehat{m}_{x,t} = \frac{\mathrm{e}^{c_t \cdot x + d_t}}{1+\mathrm{e}^{c_t \cdot x + d_t}}\,. \tag{86}$$

Empirically, the extrapolation works best when the linear model is fitted to the observations of age groups $x \in [80, 90]$.

Algorithm 2 Updating the withdrawal schedule in year u

Input: **benefit level *b***, u, $q^{(u)}_{x_j+t-1}$, $c_j(u)$, $g_j(u)$, $q_\omega = 1$, P_0, α.

Set $\Delta g_j(u+t) = \max(\alpha P_0 - t - u,\, 0) - \max(\alpha P_0 - t - u + 1,\, 0)$.
Set time to terminal age $t^{(u)}_{\omega,j} = \omega - x_j - u$.
$c_j(u + t^{(u)}_{\omega,j}) = s_j(u + t^{(u)}_{\omega,j}) = \frac{b}{1+q^{(u)}_{\omega-1}}$
for $t = t^{(u)}_{\omega,j} - 1, t^{(u)}_{\omega,j} - 2, \ldots, 1$ **do**
$c_j(u+t) = \mathrm{e}^{-\delta} c_j\big(u+t+1\big) - (1-\mathrm{e}^{-\delta}) \cdot g_j(u+t)$
$c_j(u+t) = \big(c_j(t) - \Delta g_j(u+t) + b\big) \,/\, (1 + q^{(u)}_{x_j+u+t-1})$
$s_j(u+t) = b - q^{(u)}_{x_j+u+t-1} \cdot c_j(u+t)$
end for
return updated withdrawal schedule after year u: $s_j(u+1),\ s_j(u+2), \ldots,\ s_j(\omega - x_j)$.

Practically, this means that each year, we slightly update the withdrawal schedule, that is the split between account withdrawals $s_j(t, P_0)$ and expected mortality credits. We adjust them in a way to retain a constant expected cash flow until death. Figure 3 illustrates how the expected payout would be shifted with and without an update of the withdrawal schedule after a mortality shock. Algorithm 2 shows how this is done in year u, following the iteration (77)–(80).

Using Algorithms 1 and 2, Algorithm 3 updates the withdrawal schedule and the average annual payment b. At time $t = 0$, this is slightly different. Here, benefits are fixed at at some level b. Given that $c_j(0) + g_j(0, P_0) = P_0$, the root search in Algorithm 3 is then with respect to the premium P_0 (and not the level of benefits b), see also Example 5.2.2. For updates at times $u = 1, 2, \ldots$, this first step fixes the death benefits $g_j(t^-, P_0) = \max(\alpha P_0 - t + 1, 0)$ for all $t = 0, 1, 2, \ldots$.

Equation (70) helps to get $c_j(1)$ from $c_j(0)$ (and at any later time $c_j(u+1)$ from $c_j(u)$). Given an update year u, $c_j(u)$ and $g_j(u, P_0)$, we can run Algorithm 3 to find the adjusted benefit b^* and the updated withdrawal schedule. We use that $c_j(u) + g_j(u, P_0)$ is equal to the present value of all future withdrawals $s_j(u+1, P_0)$, $s_j(u+2, P_0)$, …, $s_j(\omega - x_j, P_0)$. Given

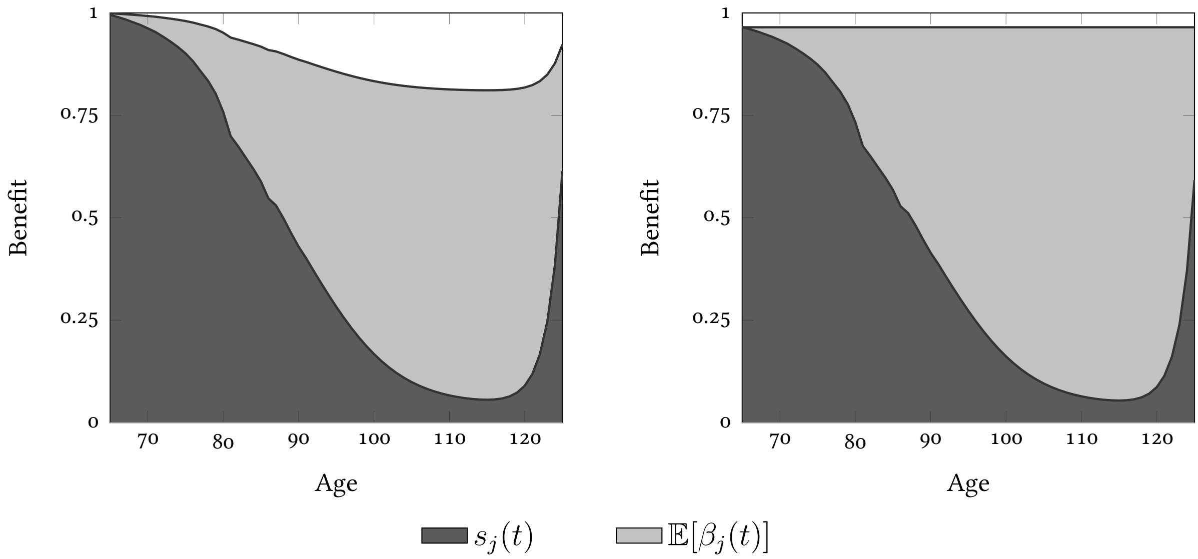


FIGURE 3. Expected future payments for a policyholder with guarantee level $\alpha = 100\%$ after a mortality shock with (right) and without (left) update of the withdrawal schedule. The mortality shock is assumed to lead to new death probabilities $\widetilde{q}_{x_j+t} = 0.8 \cdot q_{x_j+t}$.

the fixed guarantee account values $g_j(t, P_0)$, the annual withdrawals only depend on b. Thus, similarly to (**??**), we get the condition

$$g_j(u, P_0) + c_j(u) = \sum_{t=1}^{\omega - x_j - u} \mathrm{e}^{-\delta \cdot t} s_j(u+t, P_0)\,,$$

which we solve for $b = b^*$. Using this procedure, the tontine remains actuarially fair and it is possible to admit new members into the pool at the end of every year.

Algorithm 3 Annual updating step

Input: u, $c_j(u)$, $g_j(u)$, P_0.

Use Algorithm 1 to get $q^{(u)}_{x_j+t-1}$.

function PRESENT_VALUE(b)

Use Algorithm 2 to get updated withdrawal $s_j(u+1),\ s_j(u+2), \ldots,\ s_j(\omega - x_j)$ for benefit level b.

return present value at time u: $\sum_{t=1}^{\omega - x_j - u} \mathrm{e}^{-\delta t} s_j(u+t)$

end function

$b^* :=$ root of $g_j(u) + c_j(u)-$ PRESENT_VALUE(b)

return b^* and updated withdrawal $s_j(u+1),\ s_j(u+2), \ldots,\ s_j(\omega - x_j)$ for benefit level b^*.

Algorithm 3 has some similarity to group-self annuitisation schemes where also new benefits b^* are determined each year adjusting for the evolution in mortality, see, for example, (1) in Qiao and Sherris (2012). Considering this, it will be straightforward to suggest a similar updating scheme based on interest rate risk or other financial risk factors.

In Theorem 5.2.3, we have seen that for a large pool size, the refundable income annuity agrees with the refundable income tontine. This is no longer the case if there is a systematic mortality component as then longevity risk cannot be fully eliminated. Nevertheless, we can show that on average, the refundable income tontine is equal to the benefit of a refundable income annuity, see Theorem 5.3.2.

THEOREM 5.3.2 (Convergence for dependent lifetimes). *Assume that all pool members have dependent remaining lifetimes. Let all account values be such that* $c_j(t^-) \in [c_{\min}, c_{\max}]$, *where* $0 < c_{\min} < c_{\max}$.
As the pool size n *increases, the expected benefit of a refundable income tontine converges to that of a refundable income annuity.*

PROOF. See Appendix 5.A.2. □

At the end of this section, we want to provide the link between a modern tontine without premium refund ($\alpha = 0$) and group-self-annuitisation (GSA) schemes as introduced in Piggott et al. (2005). Ignoring financial risks, benefits b in a GSA are (annually) adjusted by mortality experience adjustment factors (MEA_u) and an annuity factor adjustment (AFA_u). In our notation, the two factors are given by:

$$\text{MEA}_u = \frac{p^*}{p} = \frac{\sum\limits_{j \in \mathcal{L}_{u-1}} c_j(u^-)}{\sum\limits_{j \in \mathcal{L}_u} \frac{1}{1-q_{x_j+u-1}} \cdot c_j(u^-)}, \qquad \text{AFA}_u = \frac{b^*}{b} = \frac{\sum\limits_{t=1}^{\omega - x_j - u} \mathrm{e}^{-\delta \cdot t} \cdot s_j^{\text{old}}(u+t, P_0)}{\sum\limits_{t=1}^{\omega - x_j - u} \mathrm{e}^{-\delta \cdot t} \cdot s_j^{\text{new}}(u+t, P_0)},$$

where b^*/b is the quotient between the old and new benefit payment at time $t = u$ following Algorithm 3. Further, p^*/p denotes the risk quotient of contribution-weighted actual number of deaths in $(u-1, u]$ vs. the contribution-weighted one-year survival probability at time $u-1$. While MEA_u adjusts for the difference between actual and expected deaths in $(u-1, u]$, AFA_u adjusts for systematic changes due to a variation in the mortality table. The major difference between the GSA scheme and the tontine product we introduce lies in the way mortality credits are redistributed within the pool. A GSA scheme assumes that all pool members receive the same return MEA_u on their funds for deaths in $(u-1, u]$, our mortality credit distribution takes into account the different risk level of participants (i.e. differs by age), see also the proof on actuarial fairness in Theorem 5.2.1.

5.4. Numerical example

In this section, we will provide a numerical analysis of refundable income tontine funds. We estimate the fair costs of the premium refund guarantee as a function of the guarantee level α and major parameters (for example the interest rate level δ and the initial age x_j). Then, we look at payout fluctuations, the effect of systematic and diversifiable mortality risk and a comparison to refundable income annuities.

5.4.1. Valuation. We fit the Lee-Carter mortality model to data from Switzerland available on Human Mortality Database (2022) from calendar years 1970 to 2022 for men aged 65 to 90. The exact parameters are given in Appendix 5.B. We extrapolate the death probabilities for ages 91 to 126 by the Kannisto method, see Remark 3.
The fair premium of a tontine fund with a guarantee significantly depends on the return level δ and the age x_j at which an individual j enters the tontine fund. Figure 4 shows the relative cost, i.e. the additional amount (in %) one has to pay for the refund guarantee as a function

of the guarantee level α. We see that a 65-year-old entering a tontine pays less than a 15% markup for a full guarantee ($\alpha = 100\%$) if the assumed return level is at least 2%.

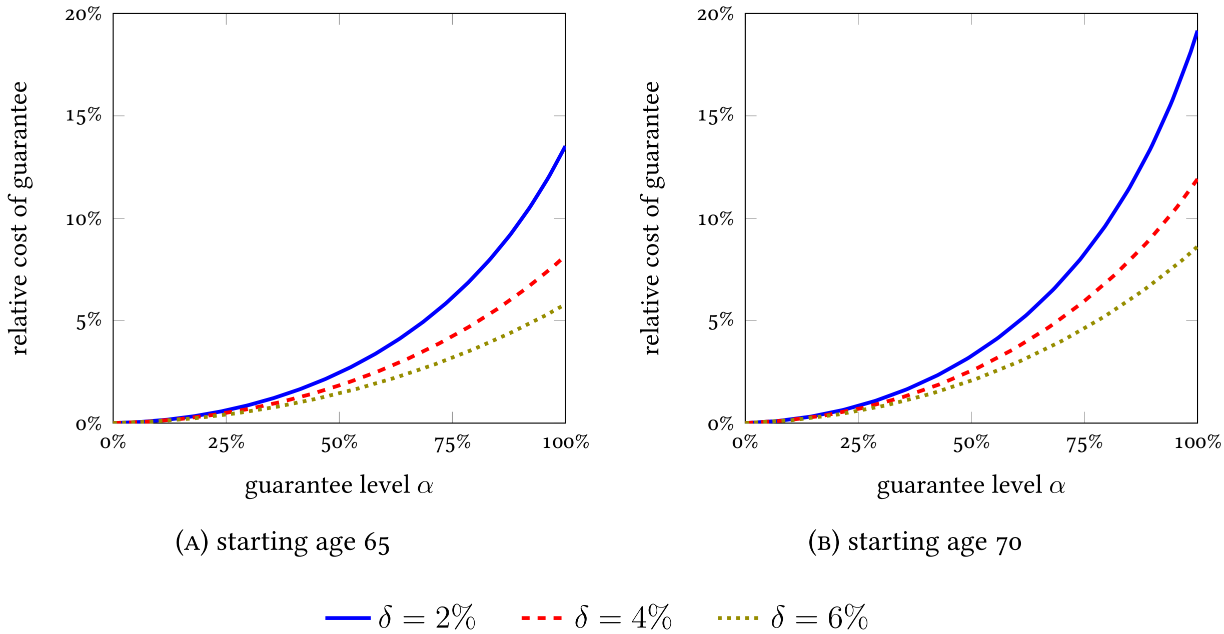


FIGURE 4. Cost of the guarantee level α for different ages at contract initiation. Values are given as a relative markup in comparison with a tontine without guarantee, i.e. $\alpha = 0\%$.

5.4.2. Simulating a heterogeneous tontine fund.

5.4.2.1. *Setup and procedure.* We carry out 10 000 simulation runs as described in Algorithm 4. We generate a tontine pool that, every year, admits $n = 100$ new members aged $x = 65$ years into the pool. Each new member chooses contract specifications according to the available parameters in Table 1.

TABLE 1. Available parameters for the specification of a tontine contract.

Attribute		Values
initial capital	P_0	100 000, 200 000, 300 000, 400 000, 500 000
guarantee fraction	α	0%, 25%, 50%, 75%, 100%
age	x	65

At the beginning of each year, we allow an update on the mortality table used for the allocation of mortality credits for the existing pool of N members. This update is optional and may follow a change in the Lee-Carter mortality trend, see for example Algorithm 1. In case of a mortality improvement in the pool, this update ensures that the withdrawal schedule is adjusted to keep a constant lifelong payment (on average), compare Figure 3, left hand side (without update) to the right hand side (with update). Secondly, if the mortality table is regularly updated, new members join to actuarially fair conditions that reflect current expectations on longevity. At the end of the year u, we record the benefit to all survivors and update all accounts $c_j(u)$ and $g_j(u, P_0)$.

Algorithm 4 Simulate a portfolio

Input: Lee-Carter parameters a_x, b_x, κ_t, θ and σ^2, number of newly admitted members per year n.

Initialise pool size $N = 0$
for $u = 1, 2, \ldots$ **do**
  Optional: Update the mortality table for mortality credit distribution.
  Optional: Update withdrawal scheme according to Algorithm 2.
  for $j = N+1, N+2, \ldots, N+n$ **do** ▷ **Add n new members to the pool**
    Randomly select an initial capital P_j and guarantee level α_j according to Table 1.
    Get expected benefit b and withdrawal schedule $s_j(u+t)$ as described in Example 5.2.2.
    Initialise the accounts $g_j(0) = \alpha_j \cdot P_j$ and $c_j(0) = (1-\alpha) \cdot P_j$.
  end for
  Get survivors from independent Bernoulli variables with survival probabilities $p^{(u)}_{x_j+u}$
  Remove deceased members from pool and update pool size N
  Update account values g_j and c_j according to (70), (73) and (74)
  Record mortality credits β_j as defined in linear risk-sharing rule (71)
end for

In order to provide a fair comparison between a refundable income annuity and a refundable income tontine, we first apply the same administration charges. As discussed in Example 5.2.2, the administration charges for the tontine are taken from those mortality credits that would have been assigned to recently deceased members. These are stochastic and amount to around 1.25% on average. For the refundable income annuity, we use the same level of administration charges. To add also a fair risk charge for an annuity, we consider the life table GRM95 that is used in Switzerland to price annuities. This life table was introduced to fit the mortality of the cohort that turned 65 in 2003. Comparing the annuity values corresponding to that life table with those derived from the Lee-Carter model implies a risk loading of around 5.25%. In total, for a refundable income annuity, we assume that an insurance company charges 6.5% in administration charges and risk premia.

5.4.2.2. *Results.* Figure 5 shows the average pool size N per year since the tontine was set up. It increases sharply for the first 30 years until the number of deaths almost equals the number $n = 100$ of new members. From then on, the pool only grows due to mortality improvements. We focus on the cohort that enters the tontine in year 61. By that time, the tontine is fully grown, i.e. the oldest members have reached the maximal age $\omega = 126$ years (if they have survived to that age).

For the following numerical study, we update the life table each year according to Algorithm 1. In the top left panel of Figure 6, we see the mean annual benefit of a tontine W_j and its 95% confidence interval compared to that of an annuity. The annual benefit remains at 100%

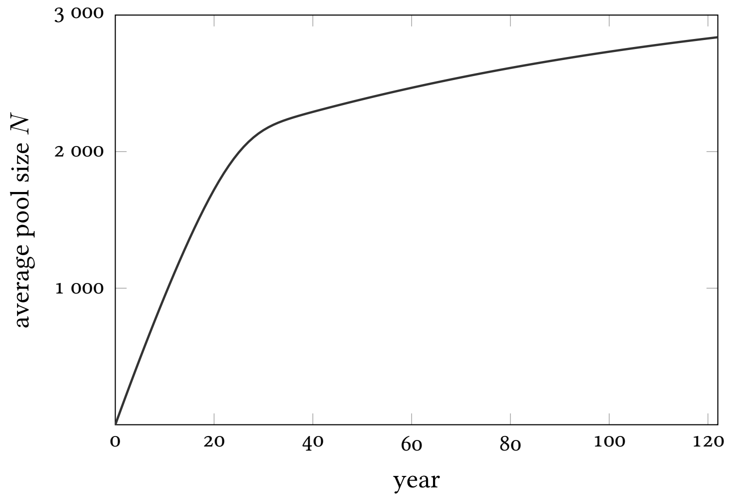


FIGURE 5. Average pool size N as a function of the number of years since the tontine was set up.

on average as proven in Theorem 5.3.2. However, it fluctuates between $+20\%$ and -20% of the initial expected payout. Although the average benefit is higher than that of an annuity (due to the risk premia of 5.25%), there are years in which there is a considerable probability of receiving less money than from an annuity. These years are mostly between the ages of 90 and 125. This is because, for higher ages, the annual benefit relies more on the mortality credits $\beta_j(t)$ compared to the fixed withdrawals $s_j(t, P_0)$, as shown in the left panel of Figure 2. Thus, if the pool experiences more or fewer deaths than expected, this has a stronger effect on older members who rely more on mortality credits.

However, the cumulative benefit does not only consider the number of deaths in one year but smoothes over all previous years. Additionally, it benefits from the less volatile first 15-20 years, during which the annual benefit is mostly made up of the fixed withdrawals $s_j(t)$. Thus, we notice that the variance is reduced considerably in the bottom left panel of Figure 6. The cumulative benefit varies by less than 5% compared to the expected value at contract initiation. We can conclude that with 95% probability a policyholder would receive more total payments from a tontine than from an annuity.

In short, the difference between the annual number of deaths within the portfolio and its expected value explains the variability of the benefit. This difference is determined by two effects. On the one hand, it depends on the diversification of the portfolio. A larger pool size will make use of the law of large numbers as shown in Theorem 5.2.3 and reduce the variance in the benefit. We can empirically show this by increasing the number of people we admit to the tontine each year to $n = 1\,000$, see the top right panel in Figure 6. For this pool size, the benefit now varies by no more than $\pm 10\%$ of the initial expected value.

On the other hand, the variability of the benefit is determined by the systematic mortality risk that is described by the Lee-Carter model. If, for example, mortality rates improve more than expected, the members of the pool tend to live longer and thus there are less mortality credits to distribute. We take this effect into account when updating the withdrawal schedule in Algorithm 3. There, we calculate a new expected benefit b^* based on the state of the mortality improvement factor κ_t. The right image in Figure 6 shows 95% confidence intervals for these updated expected benefits throughout the policyholder's lifetime. These quantify the systematic mortality risk and describe the expected benefit of an infinitely large portfolio. All variation around the values b^* could, in theory, be eliminated by increasing the pool size.

Again, we see that with a probability of more than 95%, the expected annual benefit remains higher than that of an annuity.

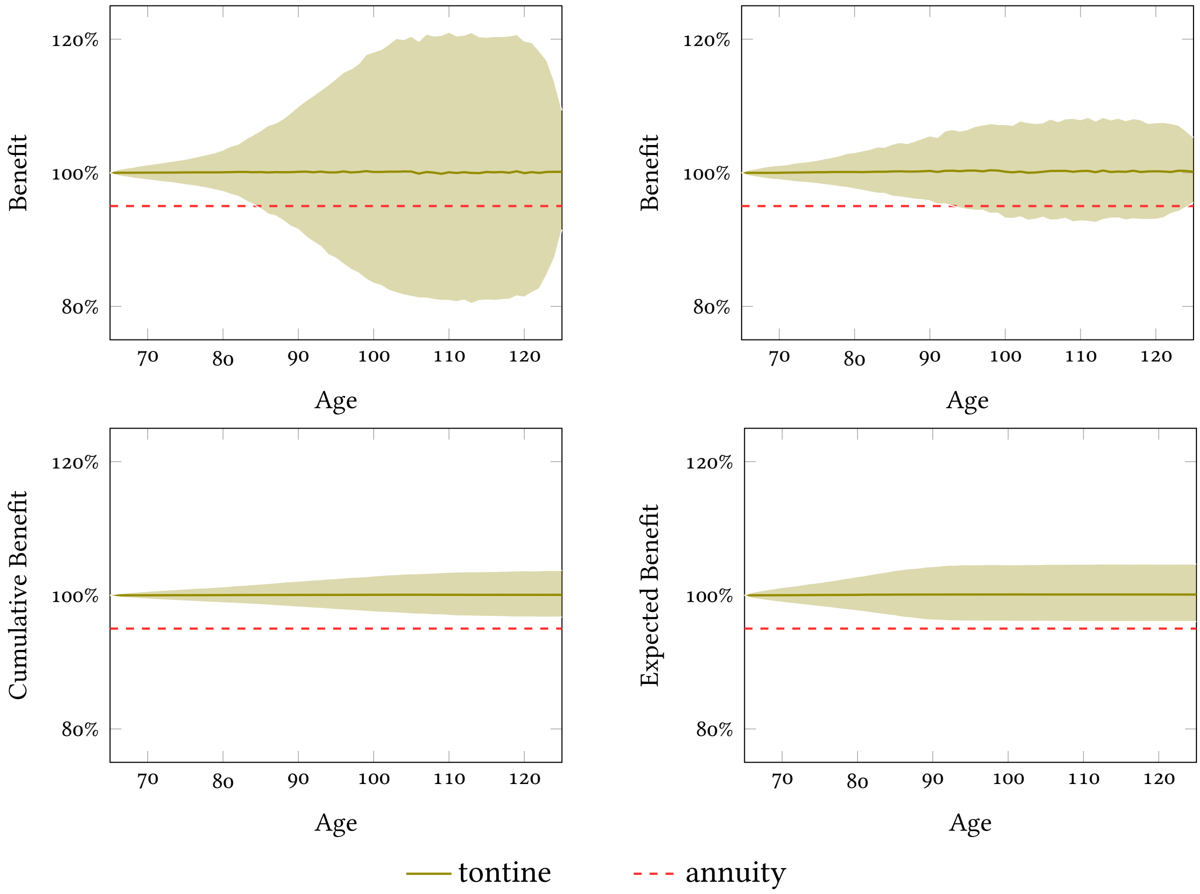


FIGURE 6. Top: Mean annual benefit of a tontine member, with guarantee level $\alpha = 100\%$ and initial investment $P_0 = 100\,000$ in a tontine pool accepting $n = 100$ (left) and $n = 1\,000$ (right) members per year. Bottom Left: Mean cumulative benefit of the same individual up to every age. Bottom Right: Mean expected benefits b^* for the same person computed in the update step in Algorithm 3.
All values are given as a percentage of the expected (cumulative) benefit at contract initiation. The results are based on $10\,000$ simulations of Algorithm 4 and the 95% confidence intervals are shaded in green.

5.5. Conclusion

We introduced a refundable income tontine that repays a fraction of the premium in case of an early death in retirement. If the benefit until death was lower than that fraction, the policyholder's heirs receive the remainder as a bequest. The product is the counterpart to a refundable income annuity and does not require any risk capital by an insurance company. In theory, by splitting the total funds into a guarantee and a tontine account, we can incorporate any deterministic death benefit.
We model the systematic mortality risk of the portfolio by a Lee-Carter model. We explain how to account for deviations from the expected mortality improvements by adjusting the future annual withdrawals based on an updated life table. These updated expected withdrawals

allow us to quantify the systematic mortality risk.

We compare refundable income tontines to refundable income annuities by simulating a heterogeneous portfolio. It takes larger pool sizes to receive a higher annual benefit from a tontine with high confidence. However, even small tontines admitting 100 new members annually into the pool can – with 95% confidence – provide the policyholder with a higher cumulative benefit than an annuity with realistic risk loadings.

5.A. Proofs of Theorems

5.A.1. Proof of Theorem 5.2.1. We prove the result by backwards induction. After $t = \omega - x_j$ years, policyholder j reaches the maximum possible age. Thus, there are no further withdrawals $s_j(t+1, P_0), s_j(t+2, P_0), \dots$ or death benefits to be paid. We arrive at $g_j(\omega - x_j, P_0) = c_j(\omega - x_j) = 0$. In the following, we will use that

$$\begin{aligned}
\mathbb{E}_{t-1}\left[W_j(t) + \beta_j(t) \cdot \mathbb{1}_{\{j \in \mathcal{D}_t\}}\right] &= \mathbb{E}_{t-1}\big[(s_j(t, P_0) + \beta_j(t)) \cdot \mathbb{1}_{\{j \in \mathcal{L}_t\}} \\
&\qquad + (\beta_j(t) + g_j(t^-, P_0)) \cdot \mathbb{1}_{\{j \in \mathcal{D}_t\}}\big] \\
&= \mathbb{E}_{t-1}\left[\beta_j(t) + s_j(t, P_0) \cdot \mathbb{1}_{\{j \in \mathcal{L}_t\}} + g_j(t^-, P_0) \cdot \mathbb{1}_{\{j \in \mathcal{D}_t\}}\right] \\
&= c_j(t^-) \cdot q_{x_j+t-1} + s_j(t, P_0) \cdot p_{x_j+t-1} + g_j(t^-, P_0) \cdot q_{x_j+t-1} \\
&= \big(c_j(t^-) + g_j(t^-, P_0)\big) \cdot q_{x_j+t-1} + s_j(t, P_0) \cdot p_{x_j+t-1}
\end{aligned}$$

and that for $j \in \mathcal{L}_t$, we get

$$c_j(t) + g_j(t, P_0) + s_j(t, P_0) = c_j(t^-) + g_j(t^-, P_0)\,,$$

which follows directly from (73) and (74).
Now, assume that (81) holds for t. Then we find for an individual $j \in \mathcal{L}_{t-1}$:

$$\begin{aligned}
&\mathbb{E}_{t-1}\left[\sum_{s=t}^{\omega-x_j} \mathrm{e}^{-\delta(s-t+1)} \cdot W_j(s)\right] + \mathbb{E}_{t-1}\left[\sum_{s=t}^{\omega-x_j} \mathrm{e}^{-\delta(s-t+1)} \cdot \beta_j(s) \cdot \mathbb{1}_{\{j \in \mathcal{D}_s\}}\right] \\
&= \mathrm{e}^{-\delta}\Bigg(\mathbb{E}_{t-1}\left[W_j(t) + \beta_j(t) \cdot \mathbb{1}_{\{j \in \mathcal{D}_t\}}\right] + \\
&\qquad p_{x_j+t-1} \cdot \mathbb{E}_t\left[\sum_{s=t+1}^{\omega-x_j} \mathrm{e}^{-\delta(s-t)} \cdot \big(W_j(s) + \beta_j(s) \cdot \mathbb{1}_{\{j \in \mathcal{D}_s\}}\big)\right]\Bigg) \\
&= \mathrm{e}^{-\delta}\Big(\big(c_j(t^-) + g_j(t^-, P_0)\big) \cdot q_{x_j+t-1} + p_{x_j+t-1} \cdot \big(s_j(t, P_0) + c_j(t) + g_j(t, P_0)\big)\Big) \\
&= \mathrm{e}^{-\delta}\Big(\big(c_j(t^-) + g_j(t^-, P_0)\big) \cdot q_{x_j+t-1} + p_{x_j+t-1} \cdot \big(c_j(t^-) + g_j(t^-, P_0)\big)\Big) \\
&= \mathrm{e}^{-\delta}\big(c_j(t^-) + g_j(t^-, P_0)\big) = c_j(t-1) + g_j(t-1, P_0)\,.
\end{aligned}$$

Thus, (81) also holds for $t - 1$.

5.A.2. Proof of Theorems 5.2.3 and 5.3.2. In case of death, both products have a certain benefit of $g_j(t, P_0)$. Thus, we need to consider the benefit in case of survival, which is 1 for the refundable income annuity. For the refundable income tontine and an expected benefit of $b_j(t) = 1$, we use (75) to determine the annual withdrawals as $s_j(t, P_0) = 1 - q_{x_j+t-1} \cdot c_j(t^-)$. We will show that, given that policyholder j survived until time t, we have

$$\beta_j(t) \cdot \mathbb{1}_{j \in \mathcal{L}_t} \xrightarrow{a.s.} (1 - s_j(t, P_0)) \cdot \mathbb{1}_{j \in \mathcal{L}_t} \quad \text{and} \quad \mathbb{E}[\beta_j(t) \,|\, j \in \mathcal{L}_t] = 1 - s_j(t, P_0)$$

for Theorems 5.2.3 and 5.3.2 respectively.

Proof of Theorem 5.2.3. We can rewrite $\beta_j(t)$ as

$$\beta_j(t) = q_{x_j+t-1} \cdot c_j(t^-) \cdot \frac{X(t^-)}{\mathbb{E}[X(t^-)]}\,,$$

where $X(t^-) = \sum_{i=1}^n \mathbb{1}_{i\in\mathcal{D}_t} c_i(t^-) =: \sum_{i=1}^n Y_i$.
The sequence Y_i satisfies Kolmogorov's criterion

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\operatorname{Var}\left(Y_i\right)}{i^2} = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{c_i(t^-)^2 \cdot q_{x_j+t-1} \cdot p_{x_j+t-1}}{i^2} < \infty\,,$$

because the tontine account values $c_i(t)$ are bounded. Therefore, by (Feller, 1968, Section X.7) the strong law of large numbers applies to the sequence Y_i and thus

$$\frac{X(t^-)}{\mathbb{E}[X(t^-)]} \xrightarrow[n\to\infty]{a.s.} 1\,.$$

Now, note that $X(t^-)\mathbb{1}_{j\in\mathcal{L}_t} = (X(t^-) - c_j(t^-) \cdot \mathbb{1}_{j\in\mathcal{L}_t}$. Hence, we find

$$\beta_j(t) \cdot \mathbb{1}_{j\in\mathcal{L}_t} = q_{x_j+t-1} \cdot c_j(t^-) \cdot \frac{X(t^-) - c_j(t^-)}{\mathbb{E}[X(t^-)]} \cdot \mathbb{1}_{j\in\mathcal{L}_t} \xrightarrow[n\to\infty]{a.s.} q_{x_j+t-1} \cdot c_j(t^-) \cdot \mathbb{1}_{j\in\mathcal{L}_t}\,.$$

$\square$

PROOF OF THEOREM 5.3.2. We have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[\beta_j(t) \,|\, j \in \mathcal{L}_t] &= \lim_{n\to\infty} \mathbb{E}\big[\mathbb{E}_{t-1}[\beta_j(t) \,|\, j \in \mathcal{L}_t]\big] \\
&= \mathbb{E}\left[\lim_{n\to\infty} q^{(t-1)}_{x_j+t-1} \cdot c_j(t^-) \cdot \left(1 - \frac{q^{(t-1)}_{x_j+t-1} \cdot c_j(t^-)}{\sum\limits_{i\in\mathcal{L}_{t-1}} q^{(t-1)}_{x_i+t-1} \cdot c_i(t^-)}\right)\right] \\
&= \mathbb{E}\left[q^{(t-1)}_{x_j+t-1} \cdot c_j(t^-)\right] = q_{x_j+t-1} \cdot c_j(t^-)\,.
\end{aligned}$$

Thus, in case of survival, we get $\mathbb{E}[W_j(t)] = s_j(t, P_0) + \mathbb{E}[\beta_j(t)] = 1$. $\square$

5.B. Lee-Carter parameters

We have calibrated the Lee-Carter model to data from the male Swiss population available on the Human Mortality Database (2022). We have only considered ages 65 to 90 and calendar years 1970 to 2022. Table 2 shows the values found for the age-dependent parameters a_x and b_x.

The remaining parameters are $\kappa_{2022} = -12.055910$, $\theta = -0.500177$ and $\sigma^2 = 1.014913$.

TABLE 2. Age-specific parameters a_x and b_x for the Lee-Carter model.

age x	a_x	b_x
65	−4.099196	0.046197
66	−4.005667	0.047028
67	−3.906698	0.048066
68	−3.818800	0.046572
69	−3.723774	0.048055
70	−3.636814	0.046763
71	−3.543625	0.047193
72	−3.441165	0.046121
73	−3.335674	0.046120
74	−3.231963	0.046192
75	−3.146928	0.044918
76	−3.033881	0.044482
77	−2.936804	0.042358
78	−2.832407	0.042294
79	−2.719201	0.041592
80	−2.600541	0.039383
81	−2.481278	0.036427
82	−2.375659	0.036259
83	−2.268151	0.033470
84	−2.157913	0.031149
85	−2.049979	0.028539
86	−1.937620	0.025265
87	−1.829462	0.024600
88	−1.722802	0.022409
89	−1.618813	0.019969
90	−1.498895	0.018580

Bibliography

Allen, F. and Gale, D. (1997). Financial markets, intermediaries, and intertemporal smoothing. Journal of Political Economy, 105(3):523–546.

Alonso-García, J., Wood, O., and Ziveyi, J. (2017). Pricing and hedging guaranteed minimum withdrawal benefits under a general Lévy framework using the COS method. Quantitative Finance, 18(6):1049–1075.

Bacinello, A. R. (2001). Fair pricing of life insurance participating policies with a minimum interest rate guaranteed. ASTIN Bulletin, 31(2):275–297.

Bacinello, A. R., Millossovich, P., Olivieri, A., and Pitacco, E. (2011). Variable annuities: A unifying valuation approach. Insurance: Mathematics and Economics, 49(3):285–297.

Bacinello, A. R., Millossovich, P., and Viviano, F. (2021). An efficient Monte Carlo based approach for the simulation of future annuity values. DEAMS research paper series.

Bacinello, A. R., Millossovich, P., and Viviano, F. (2024). An iterative least-squares Monte Carlo approach for the simulation of cohort based biometric indices. European Actuarial Journal.

Barbarin, J. and Devolder, P. (2005). Risk measure and fair valuation of an investment guarantee in life insurance. Insurance: Mathematics and Economics, 37(2):297–323.

Barigou, K. and Delong, Ł. (2022). Pricing equity-linked life insurance contracts with multiple risk factors by neural networks. Journal of Computational and Applied Mathematics, 404:113922.

Bauer, D., Bergmann, D., and Reuss, A. (2010). Solvency II and nested simulations – a least-squares Monte Carlo approach. In Proceedings of the 2010 ICA congress.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural Networks: Tricks of the Trade, pages 437–478. Springer Berlin Heidelberg.

Bernard, C., Courtois, O. L., and Quittard-Pinon, F. (2005). Market value of life insurance contracts under stochastic interest rates and default risk. Insurance: Mathematics and Economics, 36(3):499–516.

Bernard, C., Feliciangeli, M., and Vanduffel, S. (2024). Can an actuarially unfair tontine be optimal? The Geneva Risk and Insurance Review, pages 1–33.

Bernard, C. and Li, W. V. (2013). Pricing and hedging of cliquet options and locally capped contracts. SIAM Journal on Financial Mathematics, 4(1):353–371.

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

Berninger, C. and Pfeiffer, J. (2021). The Gauss2++ model: A comparison of different measure change specifications for a consistent risk neutral and real world calibration. European Actuarial Journal, 11(2):677–705.

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. Insurance: Mathematics and Economics, 37(3):443–468.

Biffis, E. and Millossovich, P. (2006). The fair value of guaranteed annuity options. Scandinavian Actuarial Journal, 2006(1):23–41.

Blake, D. and Burrows, W. (2001). Survivor bonds: Helping to hedge mortality risk. The Journal of Risk and Insurance, 68(2):339.

Boyer, M. M. and Stentoft, L. (2013). If we can simulate it, we can insure it: An application to longevity risk management. Insurance: Mathematics and Economics, 52(1):35–45.

Boyer, M. M. and Stentoft, L. (2017). Yes we can (price derivatives on survivor indices). Risk Management and Insurance Review, 20(1):37–62.

Brigo, D. and Mercurio, F. (Berlin: 2006). Interest Rate Models — Theory and Practice. Springer.

Briys, E. and Varenne, F. D. (1994). Life insurance in a contingent claim framework: Pricing and regulatory implications. The Geneva Papers on Risk and Insurance Theory, 19(1):53–72.

Bundesbank (2023). Tägliche Zinsstruktur für börsennotierte Bundeswertpapiere. https://www.bundesbank.de/de/statistiken/geld-und-kapitalmaerkte/zinssaetze-und-renditen/taegliche-zinsstruktur-fuer-boersennotierte-bundeswertpapiere-650724. Accessed: 11.10.2023.

Cairns, A. J. (2011). Modelling and management of longevity risk: Approximations to survivor functions and dynamic hedging. Insurance: Mathematics and Economics, 49(3):438–453.

Carr, P. and Madan, D. (1999). Option valuation using the fast Fourier transform. Journal of Computational Finance, 2(4):61–73.

Castellani, G., Fiore, U., Marino, Z., Passalacqua, L., Perla, F., Scognamiglio, S., and Zanetti, P. (2021). Machine learning techniques in nested stochastic simulations for life insurance. Applied Stochastic Models in Business and Industry, 37(2):159–181.

Chen, A. and Hieber, P. (2016). Optimal asset allocation in life insurance: The impact of regulation. ASTIN Bulletin, 46(3):605–626.

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

Chen, A., Kanagawa, M., and Zhang, F. (2023). Intergenerational risk sharing in a defined contribution pension system: Analysis with Bayesian optimization. ASTIN Bulletin, 53(3):515–544.

Chen, A. and Rach, M. (2021). Bequest-embedded annuities and tontines. Asia-Pacific Journal of Risk and Insurance, 16(1):1–46.

Cheridito, P., Ery, J., and Wüthrich, M. V. (2020). Assessing asset-liability risk with Neural Networks. Risks, 8(1):16.

Chourdakis, K. (2004). Non-affine option pricing. The Journal of Derivatives, 11(3):10–25.

Cotticelli, S. and Savelli, N. (2023). Capital requirement modeling for market and non-life premium risk in a dynamic insurance portfolio. Annals of Actuarial Science, 18(1):1–32.

Cui, J., De Jong, F., and Ponds, E. (2010). Intergenerational risk sharing within funded pension schemes. Journal of Pension Economics and Finance, 10(1):1–29.

Cui, Z., Kirkby, J. L., and Nguyen, D. (2017). Equity-linked annuity pricing with cliquet-style guarantees in regime-switching and stochastic volatility models with jumps. Insurance: Mathematics and Economics, 74:46–62.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems, 2(4):303–314.

Deelstra, G. and Hieber, P. (2023). Randomization and the valuation of guaranteed minimum death benefits. European Journal of Operational Research, 309(2):1218–1236.

Deelstra, G. and Rayée, G. (2013). Pricing variable annuity guarantees in a local volatility framework. Insurance: Mathematics and Economics, 53(3):650–663.

Denuit, M. (2008). Comonotonic approximations to quantiles of life annuity conditional expected present value. Insurance: Mathematics and Economics, 42(2):831–838.

Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. ASTIN Bulletin, 49(03):591–617.

Denuit, M., Dhaene, J., Feng, R., Hieber, P., and Robert, C. Y. (2024). Decentralized insurance: On the popularity of tontines and peer-to-peer (P2P) insurance schemes. Annals of Actuarial Science, 18(2):237–241.

Denuit, M., Hieber, P., and Robert, C. Y. (2022). Mortality credits within large survivor funds. ASTIN Bulletin, 52(3):813–834.

Denuit, M. and Robert, C. Y. (2021). From risk sharing to pure premium for a large number of heterogeneous losses. Insurance: Mathematics and Economics, 96:116–126.

Dhaene, J. and Milevsky, M. A. (2024). Egalitarian pooling and sharing of longevity risk a.k.a. can an administrator help skin the tontine cat? Insurance: Mathematics and Economics, 119:238–250.

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.

Dowd, K., Blake, D., and Cairns, A. J. G. (2011). A computationally efficient algorithm for estimating the distribution of future annuity values under interest-rate and longevity risks. North American Actuarial Journal, 15(2):237–247.

Eaton, M. L. (2007). Multivariate statistics: a vector space approach. Beachwood, Ohio: Institute of Mathematical Statistics.

Fan, K., Shen, Y., Siu, T. K., and Wang, R. (2015). Pricing annuity guarantees under a double regime-switching model. Insurance: Mathematics and Economics, 62:62–78.

Feller, W. (1968). An introduction to probability theory and its applications, volume 1 of Wiley series in probability and mathematical statistics. Wiley, 3 edition.

Feng, B. M., Li, J. S.-H., and Zhou, K. Q. (2022). Green nested simulation via likelihood ratio: Applications to longevity risk management. Insurance: Mathematics and Economics, 106:285–301.

Fernandez-Arjona, L. (2021). A neural network model for solvency calculations in life insurance. Annals of Actuarial Science, 15(2):259–275.

Floryszczak, A., Le Courtois, O., and Majri, M. (2016). Inside the Solvency 2 black box: Net asset values and Solvency capital requirements with a least-squares Monte-Carlo approach. Insurance: Mathematics and Economics, 71:15–26.

Forsyth, P. A., Vetzal, K. R., and Westmacott, G. (2024). Optimal performance of a tontine overlay subject to withdrawal constraints. ASTIN Bulletin, 54(1):94–128.

Gollier, C. (2008). Intergenerational risk-sharing and risk-taking of a pension fund. Journal of Public Economics, 92(5–6):1463–1485.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning. MIT Press.

Graf, S., Kling, A., and Ruß, J. (2011). Risk analysis and valuation of life insurance contracts: Combining actuarial and financial approaches. Insurance: Mathematics and Economics, 49(1):115–125.

Graf, S., Kling, A., and Ruß, J. (2021). Auswirkungen unterschiedlich hoher garantierter Rentenfaktoren bei aufgeschobenen Rentenversicherungen. https://www.ifa-ulm.de/studie-rentenfaktoren.pdf. Accessed: 11.10.2023.

Graf, S. and Korn, R. (2020). A guide to Monte Carlo simulation concepts for assessment of risk-return profiles for regulatory purposes. European Actuarial Journal, 10(2):273–293.

Grosen, A. and Jørgensen, P. L. (2000). Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies. Insurance: Mathematics and Economics, 26(1):37–57.

Günther, S. and Hieber, P. (2024). Analyzing the interest rate risk of equity-indexed annuities via scenario matrices. Insurance: Mathematics and Economics, 114:15–28.

Günther, S. and Hieber, P. (2024). Efficient simulation and valuation of equity-indexed annuities under a two-factor G2++ model. European Actuarial Journal, 14(3):905–928.

Ha, H. and Bauer, D. (2022). A least-squares Monte Carlo approach to the estimation of enterprise risk. Finance and Stochastics, 26(3):417–459.

Hainaut, D. and Akbaraly, A. (2023). Risk management with local least squares Monte Carlo. ASTIN Bulletin, 53(3):489–514.

Hieber, P. (2017). Cliquet-style return guarantees in a regime switching Lévy model. Insurance: Mathematics and Economics, 72:138–147.

Hieber, P. and Lucas, N. (2022). Modern life-care tontines. ASTIN Bulletin, 52(2):563–589.

Hieber, P., Natolski, J., and Werner, R. (2019). Fair valuation of cliquet-style return guarantees in (homogeneous and) heterogeneous life insurance portfolios. Scandinavian Actuarial Journal, 2019(6):478–507.

Hoedemakers, T., Darkiewicz, G., and Goovaerts, M. (2005). Approximations for life annuity contracts in a stochastic financial environment. Insurance: Mathematics and Economics, 37(2):239–269.

Human Mortality Database (2022).

Ignatieva, K., Song, A., and Ziveyi, J. (2016). Pricing and hedging of guaranteed minimum benefits under regime-switching and stochastic mortality. Insurance: Mathematics and Economics, 70:286–300.

Jonen, C., Meyhöfer, T., and Nikolić, Z. (2023). Neural networks meet least squares Monte Carlo at internal model data. European Actuarial Journal, 13(1):388–425.

Kabuche, D., Sherris, M., Villegas, A. M., and Ziveyi, J. (2024). Pooling functional disability and mortality in long-term care insurance and care annuities: A matrix approach for multi-state pools. Insurance: Mathematics and Economics, 116:165–188.

Kannisto, V. (1994). Development of oldest-old mortality, 1950-1990, volume 1 of Monographs on population aging. Odense University Press.

Kassberger, S., Kiesel, R., and Liebmann, T. (2008). Fair valuation of insurance contracts under Lévy process specifications. Insurance: Mathematics and Economics, 42(1):419–433.

Kijima, M. and Wong, T. (2007). Pricing of ratchet equity-indexed annuities under stochastic interest rates. Insurance: Mathematics and Economics, 41(3):317–338.

Kirkby, J. L. (2015). Efficient option pricing by frame duality with the fast Fourier transform. SIAM Journal on Financial Mathematics, 6(1):713–747.

Kirkby, J. L. (2023). Hybrid equity swap, cap, and floor pricing under stochastic interest by Markov chain approximation. European Journal of Operational Research, 305(2):961–978.

Korn, R., Temoçin, B. Z., and Wenzel, J. (2017). Applications of the central limit theorem for pricing cliquet-style options. European Actuarial Journal, 7(2):465–480.

Krah, A.-S., Nikolić, Z., and Korn, R. (2020). Least-squares Monte Carlo for proxy modeling in life insurance: Neural networks. Risks, 8(4):116.

Kurtbegu, E. (2018). Replicating intergenerational longevity risk sharing in collective defined contribution pension plans using financial markets. Insurance: Mathematics and Economics, 78:286–300.

Kushner, H. J. and Dupuis, P. (2001). Numerical Methods for Stochastic Control Problems in Continuous Time. New York: Springer.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U.S. mortality. Journal of the American Statistical Association, 87(419):659–671.

Li, J. S.-H., Li, J., Balasooriya, U., and Zhou, K. Q. (2019). Constructing out-of-the-money longevity hedges using parametric mortality indexes. North American Actuarial Journal, 25(1):S341–S372.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography, 42(3):575–594.

Li, Y. and Forsyth, P. A. (2019). A data-driven neural network approach to optimal asset allocation for target based defined contribution pension plans. Insurance: Mathematics and Economics, 86:189–204.

Lin, X. S. and Tan, K. S. (2003). Valuation of equity-indexed annuities under stochastic interest rates. North American Actuarial Journal, 7(4):72–91.

Liu, X. (2013). Annuity uncertainty with stochastic mortality and interest rates. North American Actuarial Journal, 17(2):136–152.

Lo, C. C. and Skindilias, K. (2014). An improved Markov chain approximation methodology: Derivatives pricing and model calibration. International Journal of Theoretical and Applied Finance, 17(07).

Longstaff, F. A. and Schwartz, E. S. (2001). Valuing American options by simulation: A simple least-squares approach. Review of Financial Studies, 14(1):113–147.

Luciano, E. and Vigna, E. (2008). Mortality risk via affine stochastic intensities: Calibration and empirical relevance. Belgian Actuarial Bulletin, 8:5–16.

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. (2022). Refundable income annuities: Feasibility of money-back guarantees. Insurance: Mathematics and Economics, 105:175–193.

Milevsky, M. A. and Salisbury, T. S. (2025). The Riccati tontine: How to satisfy regulators on average. The Geneva Risk and Insurance Review, 50(1):72–102.

Miltersen, K. R. and Persson, S. (2003). Guaranteed investment contracts: Distributed and undistributed excess return. Scandinavian Actuarial Journal, 2003(4):257–279.

Miltersen, K. R. and Persson, S.-A. (1999). Pricing rate of return guarantees in a Heath–Jarrow–Morton framework. Insurance: Mathematics and Economics, 25(3):307–325.

Ng, T. W. and Nguyen, T. (2024). Individual survivor fund account: The impact of bequest motives on tontine participation. SSRN.

Pan, V. and Reif, J. (1985). Efficient parallel solution of linear systems. In Proceedings of the seventeenth annual ACM symposium on Theory of computing - STOC '85, STOC '85, pages 143–152. ACM Press.

Persson, S.-A. and Aase, K. K. (1997). Valuation of the minimum guaranteed return embedded in life insurance products. The Journal of Risk and Insurance, 64(4):599.

Piccioni, M. (1987). Convergence of implicit discretization schemes for linear differential equations with application to filtering. In Lecture Notes in Mathematics, pages 208–229. Springer Berlin Heidelberg.

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.

Qiao, C. and Sherris, M. (2012). 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. SSRN Electronic Journal.

Shreve, S. (2010). Stochastic Calculus for Finance II. New York: Springer.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992 - 1994. National Bureau of Economic Research.

Swiss National Bank (2025). Offizielle Zinssätze. https://data.snb.ch/de/topics/snb/cube/snboffzisa. Accessed: 29.03.2025.

Yarotsky, D. (2017). Error bounds for approximations with deep ReLU networks. Neural Networks, 94:103–114.

Zaglauer, K. and Bauer, D. (2008). Risk-neutral valuation of participating life insurance contracts in a stochastic interest rate environment. Insurance: Mathematics and Economics, 43:29–40.