Fee Structure and Surrender Incentives in Variable Annuities

by

Anne MacKay

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2014

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Variable annuities (VAs) are investment products similar to mutual funds, but they also protect policyholders against poor market performance and other risks. They have become very popular in the past twenty years, and the guarantees they offer have grown increasingly complex. Variable annuities, also called segregated funds in Canada, can represent a challenge for insurers in terms of pricing, hedging and risk management. Simple financial guarantees expose the insurer to a variety of risks, ranging from poor market performance to changes in mortality rates and unexpected lapses.

Most guarantees included in VA contracts are financed by a fixed fee, paid regularly as a fixed percentage of the value of the VA account. This fee structure is not ideal from a risk management perspective since the resulting amount paid out of the fund increases as most guarantees lose their value. In fact, when the account value increases, most financial guarantees fall out of the money, while the fixed percentage fee rate causes the fee amount to grow.

The fixed fee rate can also become an incentive to surrender the variable annuity contract, since the policyholder pays more when the value of the guarantee is low. This incentive deserves our attention because unexpected surrenders have been shown to be an important component of the risk faced by insurers that sell variable annuities (see Kling, Ruez, and Russ (2014)). For this reason, it is important that the surrender behaviour be taken into account when developing a risk management strategy for variable annuity contracts. However, this behaviour can be hard to model.

In this thesis, we analyse the surrender incentive caused by the fixed percentage fee rate and explore different fee structures that reduce the incentive to optimally surrender variable annuity contracts. We introduce a "state-dependent" fee, paid only when the VA account value is below a certain threshold. Integral representations are presented for the price of different guarantees under the state-dependent fee structure, and partial differential equations are solved numerically to analyse the resulting impact on the surrender incentive. From a theoretical point of view, we study certain conditions that eliminate the incentive to surrender the VA contract optimally. We show that the fee structure can be modified to design contracts whose optimal hedging strategy is simpler and robust to different surrender behaviours.

The last part of this thesis analyzes a different problem. Group self-annuitization schemes are similar to life annuities, but part, or all, of the investment and longevity risk is borne by the annuitant through periodic adjustments to annuity payments. While they may decrease the price of the annuity, these adjustments increase the volatility of the payment patterns, making the product risky for the annuitant. In the last chapter of this thesis, we analyse optimal investment strategies in the presence of group self-annuitization schemes. We show that the optimal strategies obtained by maximizing the utility of the retiree's consumption may not be optimal when they are analysed using different metrics.

Acknowledgements

First, I would like to thank my advisors, Professors Carole Bernard and Mary Hardy, for the research ideas that form the backbone of this thesis. I am grateful for the guidance, the advice and the encouragements they have provided me over the past three years. I want to thank Carole for always being available, wherever she was in the world. Her enthusiasm and her energy were a great motivation, and without her I would not have been able to complete my degree so quickly. Mary's dedication to actuarial science is a constant source of inspiration. I am very thankful for all the help, advice and comments she has provided me over the past three years. These have been instrumental in the expansion of my research horizons and professional growth.

I am also grateful to Professor Phelim Boyle for giving me the opportunity to collaborate on a project that became a chapter of this thesis. I am very lucky to have had the chance to learn from him, through lectures, discussions, and while writing this thesis. Professor David Saunders has also been very helpful throughout this process. I especially appreciate his suggestions, which greatly improved my work. I want to thank him for his time over the past year.

I am also indebted to Professors Justin Wan and Pierre Devolder, who graciously accepted to be part of my examining committee. I want to thank them for their time and their helpful comments.

Professor Patrice Gaillardetz, my masters advisor, has had an important influence on my path as a graduate student. I want to thank him for his precious advice and guidance. I also want to acknowledge all the professors who have taught me since I started graduate school, as they all have, in a way, contributed to this thesis. In particular, I had very helpful discussions with Professors Adam Kolkiewicz and Bin Li, to whom I am very thankful. I also appreciate all the support and advice provided by Professor Jose Garrido.

All the comments and suggestions I have received when I presented my research at different conferences and seminars have also greatly improved the quality of this thesis. I am grateful for the comments provided by Dr. Jochen Russ and Dr. Daniel Bauer.

I am thankful to have had the chance to know the PhD students at Waterloo, who are supportive and inspiring. I also want to thank the administrative staff in statistics and actuarial science, especially Mary Lou Dufton, without whom the department would not be the same.

Finally, I want to thank my friends, who always believed in me. In particular, I am grateful to Mélina and Maciek for the advice and encouragement.

Thank you to my parents for giving me all the opportunities I have had, and for teaching me to always aim high. Thanks to the whole family for always being there, if not in person, at least on the phone. *Merci d'être là pour moi.* And thank you Chris for your unwavering support and for all the editing.

Table of Contents

List of Tables

List of Figures

Introduction

Overview of the thesis

This thesis is divided into two parts. The first part contains Chapters 1 to 5, and is concerned with the impact of different fee structures on the incentive to surrender variable annuity contracts. The second part is Chapter 6. In this chapter, we analyse the place of group self-annuitization schemes in the portfolio of a new retiree who seeks to maximize the utility of his consumption, and study the resulting payment patterns.

Fee structure and the surrender incentive in variable annuities

Introduction and motivation

Over the past 15 years, equity-linked insurance products have grown in popularity. By offering participation in market performance while protecting the initial investment, they are very attractive to many types of investors. While they used to be considered almost riskless, equity-linked products eventually proved to carry their share of risk, especially during the past financial crisis. This coincided with a rapid growth in the literature on the subject. Equity-linked insurance products are comprised of different types of contracts that differ in their features, but all offer financial guarantees that may expose their issuer to different types of risk.

In this thesis, we focus on variable annuities (VAs), which are also referred to as segregated funds in Canada. They are similar to mutual funds, but have a fixed term and guaranteed minimum payments at the time of death of the policyholder or at maturity. These guarantees, along with the tax advantages they bring, have made variable annuities very popular. However, they do present some challenges, in particular in terms of design, pricing, valuation and risk management; of course, each of these is intricately related to the others (Hardy (2003), Boyle and Hardy (2003), Palmer (2006), Coleman, Kim, Li, and Patron (2007)).

Nowadays, the guarantees that can be added to variable annuity contracts are numerous. The range of guaranteed minimum benefits is often referred to as "GMxBs" (Bauer, Kling, and Russ (2008)), to cover GMDBs (death benefits), GMMBs (maturity), GMIBs (income) and so on. The more complex guarantees evolved to distinguish the products from their competitors and to retain the policyholders. Additional guaranteed withdrawal riders can also be added to a typical variable annuity contract. In this thesis, we focus on GMMBs and apply some results to GMDBs. Note that GMMBs can also be referred to as guaranteed minimum accumulation benefits (GMABs), and the two terms will be used interchangeably throughout this thesis. Our results could eventually be extended to other types of guarantees.

In many cases, the financial guarantees embedded in VAs are analogous to financial options written on stock or indices. Techniques developed to price financial options have often been used to analyse the value of the guarantees embedded in variable annuities and other types of equity-linked insurance products. The first to do so were Boyle and Schwartz (1977), while Boyle and Hardy (1997) and Barbarin and Devolder (2005) compare and combine actuarial and financial pricing methods. There are however some differences between the way financial options and guarantees embedded in VAs are financed. In particular, financial options sold on the market are paid for upfront, whereas VA guarantees are usually funded via a fee paid out of the VA account, which is also the asset underlying the financial guarantee. The fee is typically set as a fixed percentage of the account value and is paid regularly throughout the contract. It is similar to the management fee paid out of a mutual fund to cover investment and other expenses. The fee rate charged on VA accounts is usually higher than in the mutual fund case, because it also covers the different financial guarantees.

The structure of the management fee in VAs creates a misalignment between the income and the cost of the option. When the fund value is high, large fees are received, but the option value is low because it has a small probability of being triggered at maturity. The opposite happens when the fund value is low. This discrepancy represents an incentive for the policyholder to surrender the contract when the guarantee is well out-of-the-money (see Bauer, Kling, and Russ (2008) and Milevsky and Salisbury (2001) for example). In fact, if the fund value is high enough that the guarantee has a very low probability of being in-the-money at maturity, then there is little point in continuing to pay for that guarantee. In that case, the policyholder should lapse and buy a new policy with the fund value. This new policy would be at-the-money for a similar cost. Although the maturity would be extended, it may still be an optimal strategy for the policyholder (see Moenig and Bauer (2012)).

To reduce the surrender incentive, most VA contracts include surrender charges during at least the early part of the contract duration. The surrender charge reduces the payoff received on surrender, so the policyholder does not receive the full value accumulated in the account. This surrender charge is also in place to recover the high expenses related to the sale of the VA contract. While this fee does give the policyholder an incentive to remain in the contract, there are many situations where it is optimal to surrender, even after taking the surrender charge into account.

Literature about the surrender option

Over the past 20 years, numerous papers have been concerned with pricing different equity-linked insurance contracts. In particular, many authors have analysed the impact of market assumptions on the price of equity-linked products. For example, Lin and Tan (2003) and Gaillardetz (2008) use stochastic interest rate models to price equity-indexed annuities, while Kling, Ruez, and Ruß (2011) study the impact of stochastic volatility on VAs. In this thesis, since we want to focus on the effect of the fee structure on the surrender incentive, we mostly consider a Black-Scholes market model.

The surrender problem has been treated in different ways in the literature. Nonetheless, all agree that unexpected lapses represent a significant risk for the insurer (see Kling, Ruez, and Russ (2014)). This is why policyholder behaviour needs to be accounted for when VA contracts are priced. Different assumptions can be used to model lapse behaviour, ranging from a simple deterministic lapse rates to more sophisticated models, like De Giovanni (2010)'s rational expectation and Li and Szimayer (2014)'s limited rationality. Under most of these assumptions, the policyholder is not able to assess the exact risk-neutral value of the contract. In addition, exogenous factors can affect her decision.

Another approach to modeling policyholder behaviour is to assume that the policyholder is perfectly rational and surrenders the contract as soon as it is optimal to do so from a financial perspective. Under this assumption, the surrender option can be viewed as an American option that can be exercised at any time before maturity (see Grosen and Jørgensen (2000)). The value obtained for the VA contract using this assumption represents an upper bound for its price, as it considers the worst-case scenario for the insurer. Even if the value thus obtained is not used as the final price, it sheds light on the intrinsic value of surrender option and on the risk it bears. Furthermore, while there are many other factors why policyholders lapse, Knoller, Kraut, and Schoenmaekers (2011) show that the moneyness of the embedded guarantee plays a role in surrender decisions. This is not dissimilar to surrendering optimally when the guarantee is out-of-the-money. They also find that financial literacy increases sensitivity towards the moneyness.

Pricing a VA contract assuming optimal surrender strategy can be justified if the insurer wants to cover the worst-case scenario. However, optimal surrenders are more complex to hedge and to manage. For this reason insurers can be tempted to ignore lapse risk or to make simplifying assumptions that do not reflect actual lapse behaviour. These flawed assumptions can significantly reduce the efficiency of a hedging strategy. For example, Kling, Ruez, and Russ (2014) show that hedging effectiveness can be threatened when lapse behaviour assumptions fail to predict actual surrenders. Thus, early surrenders are an important component of the risk faced by issuers of VA contracts.

Although most variable annuity contracts charge a constant fee as a percentage of the account value to cover embedded guarantees, many authors assume that these benefits are covered by the initial premium (see, for example, Grosen and Jørgensen (2002), Bacinello (2003a), Bacinello (2003b), Siu (2005), Bacinello, Biffis, and Millossovich (2009), Bacinello, Biffis, and Millossovich (2010), Bernard and Lemieux (2008)). However, as discussed earlier, the management fee has an impact on the surrender incentive and should be considered when the policyholder is assumed to lapse optimally. This is mentioned by Bauer, Kling, and Russ (2008) and Milevsky and Salisbury (2001). In particular, Milevsky and Salisbury (2001) argue that surrender charges are necessary to complete the market; they allow the insurer to fairly price the VA contract and hedge it appropriately.

Under the rationality assumption, the surrender option embedded in a VA contract can be analysed with tools developed for American options. A vast literature has been developed on this topic, so we will not try to cover it all here. American options can be priced in many different ways, each of which has its advantages and disadvantages. In particular, Kim and Yu (1996) use no-arbitrage arguments to derive an integral form for the value of the early exercise premium. In this thesis, we apply this technique to VA contracts to isolate the value of the surrender option and to understand the different factors affecting its value. This value can also be obtained through the method developed in Kim (1990). American options can also be priced using partial differential equations (PDEs), such as in Carr, Jarrow, and Myneni (1992). PDEs are also used in the context of equity-linked insurance products (for example in Dai, Kuen Kwok, and Zong (2008), Chen, Vetzal, and Forsyth (2008) and Belanger, Forsyth, and Labahn (2009)). In this thesis, we use them to assess the surrender incentive when the integral representation cannot be obtained.

Exploring new fee structures

Chapter 1 of this thesis presents the notation and concepts used throughout the first five chapters of the thesis. It reviews certain notions of risk-neutral pricing, fair pricing and optimal surrender. It also describes the VA contracts considered in this thesis.

In Chapter 2, we propose a technique to isolate the value of the surrender option in VA contracts with different types of accumulation benefits. Relying on the no-arbitrage arguments presented by Kim and Yu (1996), we develop an integral representation for the value added by the possibility to surrender a VA contract early. From this representation, it becomes obvious that optimal surrender incentives depend on the value of the fee paid when the account value is high above the guaranteed level. In other words, decreasing the fee paid when the maturity guarantee is out-of-the-money would reduce the surrender incentive.

With this result in mind, we introduce a new "state-dependent" fee structure in Chapter 3. Under this new setting, the fee is still paid as a constant percentage of the VA account, but only when the value of this account is below a certain threshold. Chapter 3 explores the fair fee for the accumulation benefit with this new fee structure. Using the appropriate change of measure and the necessary trivariate density derived in Karatzas and Shreve (1984), we obtain an integral representation for the value of the maturity benefit. This allows us to perform different sensitivities on the price of the contract, and to do a preliminary analysis of the surrender incentive.

Chapter 4 studies the effect of the state-dependent fee on the surrender incentive. Since the problem now includes optimal surrenders, it is no longer possible to obtain an analytic formula for the value of the contract. Instead, the price is obtained numerically by solving a partial differential equation using finite difference schemes. This also allows us to visualise the optimal surrender region for a VA with a simple GMAB at maturity. In particular, we show that the state-dependent fee combined with early surrender charges is effective in reducing the optimal surrender incentive. In this chapter, we also explain how to design a marketable contract for which the optimal behaviour is to keep it until maturity. In other words, we eliminate the surrender incentive completely, thus greatly reducing the complexity of the strategy required to hedge the optimal lapse behaviour. By analysing the hedging errors resulting from the application of such a strategy, we demonstrate that it is effective at mitigating lapse risk.

In Chapter 5, we modify the fee structure and study the case where the fee is paid as a fixed amount (instead of a fixed percentage of the account). This fee structure can be seen as a function of the account value. In particular, when the maturity guarantee is out-of-the-money, the fee rate paid by the policyholder is smaller than when the guarantee is in-the-money. We show that this fixed fee affects the shape of the optimal surrender region, also reducing the surrender incentive.

Optimal investment strategies at retirement in the presence of group self-annuitization schemes

While the first five chapters are concerned with a product that is typically used for pre-retirement savings, Chapter 6 explores post-retirement investment. Group-self annuitization schemes can be compared to life annuities with variable payments, which depend on the investment and mortality experience of the group. They are attractive to pension plan sponsors because they transfer investment and longevity risk to the retirees. For the same reason, they can result in very volatile payment patterns, which is particularly risky if they constitute the main source of income for retirees. In Chapter 6, we assume that a retiree seeks to maximize the expected utility of his consumption by investing in one or more of the following:

- A risk-free bank account
- A balanced fund
- A fixed life annuity
- A self-annuitization scheme.

Our results show that fixed life annuities still have a place in a retiree's portfolio, even when their price includes a margin for investment and longevity risk. Using a different metric, we also show that utility maximization does not necessarily yield the most appropriate investment strategies for retirees.

Chapter 1

Notation and setting

In this chapter, we introduce the market model used in the next four chapters. We also review some notions of financial and actuarial mathematics and define the main concepts discussed throughout this thesis.

1.1 Market model

We consider a variable annuity contract with maturity T and assume that its account tracks the value of an index $\{S_t\}_{0 \leqslant t \leqslant T}$ with real-world ($\mathbb{P}$-measure) dynamics

$$dS_t = S_t \left(\mu dt + \sigma dW_t^{\mathbb{P}} \right),$$

where $W_t^{\mathbb{P}}$ is a $\mathbb{P}$-Brownian motion. We work on a filtered probability space $(\Omega, \mathcal{F}_T, \mathbb{F}, \mathbb{P})$, where $\mathcal{F}_t = \sigma(\{W_s^{\mathbb{P}}\}_{0 \leqslant s \leqslant t})$ is the filtration induced by the Brownian motion, and $\mathbb{F} = \{\mathcal{F}_t\}_{0 \leqslant t \leqslant T}$.

For $0 \leqslant t \leqslant T$, we let F_t be the value of the VA account at time t and denote by C_t the total fees paid between times 0 and t. We assume that the fee paid at time t can be function of time and of the account value, so its dynamics are given by

$$dC_t = F_t \mathfrak{c}(t, F_t) dt,$$

where $C_0 = 0$ and $\mathfrak{c}(t, F_t)$ represents the fee rate. Since the management fee is paid out of the VA account, the process $\{F_t\}_{0 \leqslant t \leqslant T}$ follows

$$\begin{aligned} dF_t &= \frac{F_t}{S_t} dS_t - dC_t \\ &= F_t \left((\mu - \mathfrak{c}(t, F_t))dt + \sigma dW_t^{\mathbb{P}} \right). \end{aligned}$$

The different fee structures used in this thesis are as follows:

- In Chapter 2, we assume that the fee is paid continuously out of the fund at a rate c, so $\mathfrak{c}(t, F_t) = c$.
- In Chapter 3 and 4, we consider a fee that is paid only when the account value is below a level β. In this case, $\mathfrak{c}(t, F_t) = c\mathbb{1}_{\{F_t < \beta\}}$, where $\mathbb{1}_A$ is the indicator function of the set A.
- In Chapter 5, we explore a fee set as the sum of a deterministic amount p_t at time t and a fixed percentage c of the VA account F_t. Thus, $\mathfrak{c}(t, F_t) = c + \frac{p_t}{F_t}$.

Throughout this thesis, we consider that the assumptions of the Black-Scholes model (see Black and Scholes (1973)) hold. In particular, as demonstrated by Harrison and Pliska (1981), this means that there exists an unique equivalent risk-neutral measure $\mathbb{Q}$ under which discounted price processes are martingales. Under this measure, the dynamics of the VA account are given by

$$dF_t = F_t \left((r - \mathfrak{c}(t, F_t))dt + \sigma dW_t^{\mathbb{Q}} \right),$$

where $W_t^{\mathbb{Q}}$ is a $\mathbb{Q}$-Brownian motion. In the subsequent chapters, to simplify the notation, we will drop the superscript indicating under which measure the Brownian motion is defined whenever the context is clear.

1.2 Variable annuity contracts

1.2.1 Maturity benefit

In this thesis, we focus on a policy with a simple guarantee effective at maturity T. The payoff of the contract is the maximum between a pre-determined amount G and $\phi(T, F_\bullet)$ a function which may depend on the entire path of the fund value. In Chapter 2, we discuss payoffs that can be path-dependent. In particular, $F_\bullet$ can be the average of the fund value over time. In the subsequent chapters, we restrict ourselves to the case $\phi(T, F_\bullet) = F_T$. In all cases, the payoff of the VA at maturity is thus $\max(G, \phi(T, F_\bullet))$. We typically express the pre-determined amount G as the initial premium $P = F_0$ rolled-up to T at a conservative rate $g < r$, so that $G = F_0 e^{gT}$.

We denote the value of the maturity benefit at time t by $U(t, F_t)$ and define it as the risk-neutral expectation of the payoff discounted from T to t:

$$U(t, F_t) = E_{\mathbb{Q}}\left[e^{-r(T-t)} \max(G, \phi(T, F_\bullet))\right].$$

1.2.2 Optimal surrender

In Chapter 3, we focus on pricing the maturity benefit only. However, in the other chapters, we also consider the value added by the possibility of surrendering the contract before its maturity. This requires us to define the concept of optimal surrender and optimal surrender region, which is done here. These concepts, which will be reviewed in the next few chapters, are analogous to the ones used in the literature on American options (Wu and Fu (2003), for example).

We denote the price of the VA contract by $V(t, F_t)$. If this contract is surrendered at time $t \in [0, T)$, the policyholder receives

$$(1 - \kappa_t)\phi(t, F_\bullet),$$

where $\kappa_t \in [0, 1)$ is the surrender charge at t and $\phi(t, \cdot)$ is a function of the fund value or of its path to time t. Further assumptions will be made on the form of the surrender charge in the next chapters.

We also let τ be a stopping time with respect to $\mathcal{F}_t$ and denote by $\mathcal{T}_t$ the set of all stopping times τ greater than t and bounded by T. Denote the continuation value of the VA contract at time t by $V^*(t, F_t)$ and define it by

$$V^*(t, x) = \sup_{\tau \in \mathcal{T}_t} E\left[e^{-r(\tau - t)}\psi(\tau, F_\bullet) | F_t = x\right],$$

where $\psi(\tau, F_\bullet)$ denotes the benefit received if the contract is surrendered or expires at τ. That is,

$$\psi(t, x) = \begin{cases} (1 - \kappa_t)x, & \text{if } 0 \leqslant t < T \\ \max(x, G), & \text{if } t = T. \end{cases}$$

Heuristically, the continuation value is the discounted maximum value that the policyholder can expect to receive if she holds the contract at least one instant more. For each time $t \in (0, T)$, we can define the optimal surrender region $\mathcal{R}_t$ as

$$\mathcal{R}_t = \{F_t : V^*(t, F_t) \leqslant \psi(t, F_\bullet)\}. \tag{1.1}$$

We assume that the contract is surrendered as soon as the account value enters the optimal surrender region. Finally, the price of the contract is equal to the continuation value in the optimal surrender region. On the boundary of this region, the policyholder is indifferent between the continuation value $V^*(t, F_t)$ and the surrender value $\psi(t, F_t)$. Outside of this region, it is simply equal to the benefit received on surrender. Thus, for $t \in [0, T)$, we have

$$V(t, F_t) = \begin{cases} (1 - \kappa_t)\phi(\tau, F_\bullet), & \text{if } F_t \in \mathcal{R}_t, \\ V^*(t, F_t), & \text{otherwise.} \end{cases}$$

1.2.3 Fair fee

In Chapters 2 to 5, we are concerned with pricing VA contracts under different assumptions and fee structures. Here, we will consider that the fair fee rate is the smallest rate for which the initial premium is equal to the risk-neutral expectation of the VA payoff. Denoting the fair fee rate by c^*, it is the smallest rate that satisfies

$$P = F_0 = V^{(c^*)}(0, F_0), \tag{1.2}$$

where the superscript c^* represents the dependence of the value of the policy on the fee rate. We will usually drop this superscript, unless the fee rate used is not clear from the context. Note that Equation (1.2) also sheds light on another assumption used throughout this thesis — the initial premium P is equal to the initial fund value F_0. In other words, there are no upfront fees paid by the policyholder when buying the contract; the entire premium is deposited in the account. Unless otherwise stated, we consider that the policyholder does not make further deposits in the VA account.

1.3 Other assumptions

Throughout this thesis, we mostly consider VA contracts assuming that the policyholder is still alive at maturity. However, most policies offer additional guarantees if she passes away before maturity of the contract, and the maturity benefit is only paid if the policyholder is alive. For this reason, insurers need to account for mortality risk when pricing VAs. This is particularly true when insurers offer income guarantees, which can be valid as long as the policyholder is alive. In that case, longevity risk becomes an important part of the risk faced by the insurer, and modelling mortality improvements accurately is crucial. However, since our goal is to concentrate on the surrender incentive for products with fixed maturity, and because we want to isolate this incentive, we believe that our simplifying assumption is justified.

Since we are using the Black-Scholes model, we assume that the risk-free rate is deterministic and constant. Long-term financial guarantees like the ones embedded in VA contracts are sensitive to changes in the interest rate, so they should be priced and further studied using a model that allows for stochastic interest rates. Future extensions of our work should include analysis of our conclusions under stochastic interest rate models.

Chapter 2

Optimal surrender under constant fee structure

2.1 Introduction

This chapter is based on a paper that was written in collaboration with Dr. Carole Bernard and Max Muehlbeyer (from Ulm University), and that was published in *Insurance: Mathematics and Economics* (see Bernard, MacKay, and Muehlbeyer (2014)).

In this chapter, we investigate the optimal surrender strategy for a variable annuity contract with a minimum accumulation benefit, when the fee is paid as a constant percentage of the fund. We first consider a simple point-to-point guarantee and derive an integral representation for the continuation value of the contract, which can be solved to compute the optimal surrender boundary. To do so, we use no-arbitrage arguments presented, among others, by Kim and Yu (1996) and Carr, Jarrow, and Myneni (1992). This technique, originally designed for vanilla call options, can be extended to more complex path-dependent payoffs linked, for example, to the average fund value. Our objective is to illustrate a general technique to compute the optimal surrender strategy for a possibly path-dependent contract. This technique may help to understand the effect of complex path-dependent benefits on surrender incentives and could be useful to reduce the surrender option value by modifying the type of benefits offered and assess the riskiness of path-dependent benefits. The assessment of the value of the surrender option is also crucial to developing an upper bound for the price of a VA contract.

The chapter is organized as follows. In Section 2.2 we state the setting. The optimal surrender policy is derived in Section 2.3. Section 2.4 extends this method to path-dependent payoffs. In Section 2.5 we apply these results to numerical examples and analyse the sensitivity of the boundary with respect to a range of parameters. Section 2.6 concludes.

2.2 Setting

Consider a variable annuity contract with a guaranteed minimum accumulation benefit G at maturity T. This accumulation benefit is computed as $G = F_0 e^{gT}$ where g represents the guaranteed roll-up rate. Let F_t denote the underlying accumulated fund value of the variable annuity at time t. We assume that the insurance company charges a constant fee c for the guarantee, which is continuously withdrawn from the accumulated fund value F_t. Furthermore, we assume that the policyholder pays a single premium to initiate the contract.

The insurer then invests this premium in the fund or index that was chosen by the policyholder. We denote this underlying fund or index by S_t and assume that it follows a geometric Brownian motion. Therefore, its dynamics under the risk-neutral measure $\mathbb{Q}$ are given by

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{2.1}$$

where r is the risk-free interest rate, $\sigma > 0$ the constant volatility and W_t the Brownian motion. We denote by $\mathcal{F}_t$ the natural filtration associated with this Brownian motion. In this case, the stock price at time $u > t$ given the stock price at time t has a lognormal distribution and is explicitly given by

$$S_u = S_t e^{(r-\frac{\sigma^2}{2})(u-t)+\sigma(W_u - W_t)}$$

In this chapter, we are only concerned with pricing the surrender option and as such, we can treat the whole problem under the risk-neutral measure. This choice is also motivated by the use of no-arbitrage arguments in the derivation of the expression for the surrender option. It is based on the assumption that policyholders optimize over all possible surrender strategies and will choose to surrender optimally from a financial perspective.

The following results (2.2) and (2.3) will be useful to derive the results of this chapter. Since the insurance company continuously withdraws the fee from the fund value at a rate c, we have the following relationship between S_u and F_u at any time $u \in [0, T]$

$$F_u = e^{-cu} S_u = F_t e^{(r-c-\frac{\sigma^2}{2})(u-t)+\sigma(W_u - W_t)}. \tag{2.2}$$

Therefore, the conditional distribution of $F_u|F_t$ for $u > t$ is a lognormal distribution with log-scale parameter $\ln(F_t) + (r - c - \frac{\sigma^2}{2})(u - t)$ and shape parameter $\sigma^2(u - t)$. Hence, the risk-neutral transition density function of F_u at time $u > t$ given F_t is given by

$$f_{F_u}(x|F_t) = \frac{1}{\sqrt{2\pi\sigma^2(u-t)}x} e^{-\frac{[\ln(\frac{x}{F_t}) - (r-c-\frac{\sigma^2}{2})(u-t)]^2}{2\sigma^2(u-t)}}, \; x > 0. \tag{2.3}$$

Note that in this chapter we restrain ourselves to the case when the underlying follows a geometric Brownian motion, which presents a simple closed expression for its transition density. However, the method we present here can easily be extended to more general market models. We discuss this point briefly in the concluding remarks.

2.2.1 Fair Fee for the European Benefit

We assume initially that the VA cannot be surrendered early. Let c be the fee charged by the insurer between 0 and T. Note that the fund value at time T depends on this fee. We denote by F_T^c the value at T of the fund given that the fee charged during $[0, T]$ is equal to c and by $\phi(T, F_\bullet^c)$ the payoff at maturity T which may depend on the path of the fund denoted by $F_\bullet^c$. If the fee c is fair (for the European benefit), we denote it by c^* and it satisfies

$$F_0 = E[e^{-rT} \max(T, \phi(F_\bullet^{c^*}), G)], \tag{2.4}$$

where F_0 is the lump sum paid initially by the policyholder net of initial expenses and management fees. For $\phi(F_\bullet) = F_T$, and for other usual payoff functions $\phi(F_\bullet^{c^*})$, this fee c^* exists and is unique. To compute this fair fee, it is always possible to use Monte Carlo techniques. However when the distribution of $\phi(T, F_\bullet^c)$ is known, an analytical formula may be derived, which subsequently can be solved for c^*. For example when $\{X_t\}_{t\in[0,T]}$ is a Markov process with $X_T|X_t \sim \mathcal{LN}(M_t, \mathcal{V}_t)$ (a lognormal distribution with log-scale parameter M_t and shape parameter $\mathcal{V}_t$), then $E[\max(X_T, G)]$ can be computed as

$$E[\max(X_T, G)|\mathcal{F}_t] = e^{M_t + \frac{\mathcal{V}_t}{2}} \mathcal{N}\Big(\frac{-\ln(G) + M_t + \mathcal{V}_t}{\sqrt{\mathcal{V}_t}}\Big) + G\mathcal{N}\Big(\frac{\ln(G) - M_t}{\sqrt{\mathcal{V}_t}}\Big) \tag{2.5}$$

We omit the proof as it is a rather standard computation. The expression (2.5) can be used to compute the value of the maturity benefit of the VA in a Black Scholes setting when $\phi(T, F_\bullet^{c^*}) = F_T^{c^*}$, which is the simplest benefit: a GMAB on the terminal fund value payable at time T (Section 2.3). We can then solve for the fair fee in (2.4). It will also be applied when $\phi(T, F_\bullet^{c^*})$ is the geometric average of the fund value in Section 2.4.

2.2.2 Surrender Option

We now assume that the policyholder is allowed to surrender the policy at any time $t \in [0, T)$ for a surrender benefit equal to

$$(1 - \kappa_t)\phi(t, F^c_\bullet)$$

where κ_t is a penalty percentage charged for surrendering at time t. This is consistent with the modeling of surrender charges in Milevsky and Salisbury (2001). A standard penalty is typically decreasing over time. Examples of penalty functions are given in Palmer (2006).

In the absence of a surrender penalty ($\forall t, \kappa_t = 0$), we will see in the numerical analysis in Section 2.5 that the optimal surrender boundary is decreasing as a function of c. This result is intuitive: if the fee c charged on the fund is high, the policyholder has a larger incentive to surrender the contract when the guarantee is out of the money, because she is paying more for it.[1] This observation means that it may be difficult to pay for the surrender benefit by withdrawing a higher fixed percentage of the fund. Indeed if, for example, it is optimal to surrender when $F_t > 125$ when $c = 1\%$, then by charging $c = 2\%$ it might be optimal to surrender when $F_t > 100$. Increasing the fee c to take into account the surrender benefit increases the value of the surrender option. Alternatives include the possibility to charge for this benefit initially as a lump payment or to design a sufficiently high surrender penalty to decrease the incentive to surrender. This point is already present in the analysis of Milevsky and Salisbury (2001). It is clear that when κ_t is sufficiently high then it is never optimal to surrender at time t.

For simplicity, throughout the chapter, we assume that κ_t is exponentially decreasing and equal to $1 - \exp(-\kappa(T - t))$ so that the surrender benefit is equal to

$$e^{-\kappa(T-t)}\phi(t, F^c_\bullet), \tag{2.6}$$

for $\kappa < c$. For example when the surrender benefit at time t is $e^{-\kappa(T-t)}F^c_t$, then the inequality $\kappa < c$ ensures that it can be optimal to surrender the VA for a sufficiently high value of the fund F^c_t. The continuation value of the contract at time t is indeed always strictly greater than $F^c_t e^{-c(T-t)}$ because the policyholder will receive $\max(F^c_T, G)$ at time T and thus at least the fund F^c_T. At time t, the value of receiving F^c_T at time T is given by $E[F^c_T e^{-r(T-t)} | \mathcal{F}_t] = e^{-c(T-t)}F^c_t$. By assuming that $\kappa < c$, we ensure that for any fixed time

[1] In other words at a given time, the higher c, the larger the future fees to pay before the maturity, whereas the final benefit is decreasing in c, so the gap between the future benefit associated with the guarantee option and the future expected fees remaining to be paid increases and thus the incentive to surrender increases as well.

$t \in [0, T)$, there exists a fund value high enough that the surrender benefit is worth more than the maturity benefit so that surrendering the policy might become optimal.

2.3 Derivation of the optimal surrender boundary

This section presents the technique used to derive the optimal surrender boundary. As mentioned earlier it is sometimes optimal for the policyholder to surrender the contract before the maturity T because the fee c is charged as a percentage of the fund value. Thus, assuming the fund value is sufficiently high, the fee paid for the guarantee would exceed the actual value of the guarantee. This mismatch leads to an optimal early surrender of the variable annuity.

Consider the variable annuity contract from Section 2.2 with a payoff of $\max(F_T, G)$ at maturity T. Here we assume that c is given and thus omit the superscript c in the notation for the value of the fund at time t. If the contract is surrendered early, at time $t < T$, the policyholder receives the accumulated fund value F_t reduced by the surrender penalty, so that the surrender benefit is given by $e^{-\kappa(T-t)}F_t$ (particular case of (2.6)). Let B_t denote the value of the optimal surrender boundary at time t, i.e. if the fund value crosses this value from below, it is optimal for the policyholder to lapse the contract and receive the amount B_t.[2]

In order to derive the value of the surrender option and the optimal surrender boundary we use the same technique as Kim and Yu (1996) and Carr, Jarrow, and Myneni (1992). We first seek to calculate the continuation value of the VA contract, which is defined in Chapter 1. It represents the value of the policy given that it is kept at least one instant longer, and it is the value used to define the boundaries of the optimal surrender region. Outside of the optimal surrender region, the continuation value of the VA contract is equal to its price. Throughout this chapter, we use "price" and "continuation" value interchangeably.

To calculate the continuation value at time t, denoted by $V^*(t, F_t)$, we decompose it into a European part and a surrender option. To understand the intuition behind this approach, consider the following trading strategy which "converts" the full contract value into the corresponding value of the maturity benefit and the surrender option. We know that the price of the VA at time $t < T$ along the surrender boundary is equal to $e^{-\kappa(T-t)}F_t$, the value of the surrender benefit at t. This comes from the definition of the optimal

[2]Here we assume that the optimal surrender region is of the form $\{F_t > B_t\}$, in other words the optimal surrender behaviour is based on a threshold strategy where optimal surrender is driven by the value of the underlying fund crossing a barrier. This assumption is discussed and justified in Appendix 2.A.

surrender boundary given in Chapter 1. Moreover, $B_0 > F_0$ because otherwise it would not be optimal for the policyholder to buy the VA at time 0 for a price F_0. We neglect all transaction costs.

Assume that the policyholder has bought the VA at time $t = 0$. Now whenever the fund value crosses the optimal surrender boundary from below, she exercises the option and surrenders the contract. And whenever the fund value crosses the boundary from above, she buys back the VA contract (given that the boundary is exactly equal to the value of the VA by definition). Any profits resulting from this trading strategy constitute the value added by the possibility of lapsing the contract before maturity — the surrender option. So assume that at time t the fund value F_t crosses the optimal surrender boundary from below. The policyholder surrenders the contract and receives $e^{-\kappa(T-t)}F_t = e^{-\kappa(T-t)}e^{-ct}S_t$ which she instantaneously invests in the stock S_t. However, since S_t is not subject to the guarantee fee c, S_t outperforms F_t. Therefore, in the case that the fund value crosses the surrender boundary from above, say at time $u > t$, the value of the contract on the boundary is $e^{-\kappa(T-u)}F_u$, the policyholder only needs to pay $e^{-\kappa(T-u)}F_u$ to re-enter, that is $e^{-\kappa(T-u)}e^{-cu}S_u = e^{-\kappa T}e^{-(c-\kappa)u}S_u < e^{-\kappa(T-t)}e^{-ct}S_u$ (because $c - \kappa > 0$). The profit from this strategy is the value of the surrender option. A formal derivation is given in the proof of Proposition 2.3.1 below.

Proposition 2.3.1. *The benefit associated with the exercise of the surrender option between* $[t, t+dt]$ *is equal to* $h(t) = e^{-\kappa(T-t)}(c-\kappa)F_t dt + g(dt)$*, where* $g(dt)$ *is* $o(dt)$ *as* $dt \to 0$.[3]

Proof. Assume the VA contract is surrendered at time t. Then the policyholder receives an amount of $e^{-\kappa(T-t)}F_t = e^{-\kappa(T-t)}e^{-ct}S_t$, which is invested in the index S_t. In order to buy it back at time $t + dt > t$, she only needs $e^{-\kappa(T-(t+dt))}F_{t+dt} = e^{-\kappa T}e^{-(c-\kappa)(t+dt)}S_{t+dt}$. Therefore, consider the following decomposition of the amount received at time t:

$$\begin{aligned} e^{-\kappa(T-t)}e^{-ct}S_t &= e^{-\kappa T}e^{-(c-\kappa)(t+dt)}S_t + e^{-\kappa T}S_t\big(e^{-(c-\kappa)t} - e^{-(c-\kappa)(t+dt)}\big) \\ &= e^{-\kappa(T-(t+dt))}e^{-c(t+dt)}S_t + e^{-\kappa(T-t)}e^{-ct}S_t(1 - e^{-(c-\kappa)dt}) \end{aligned} \tag{2.7}$$

The first addend is the amount invested in the asset S_t that is needed to re-enter the contract at time $t + dt$ (in other words, it is the no-arbitrage price of $e^{-\kappa(T-(t+dt))}e^{-c(t+dt)}S_{t+dt}$ paid at time $t + dt$). The second addend is the amount that needs to be siphoned off and is invested in the risk-free asset. This decomposition is going to be the key step in generalizing this proof to more general benefits (see Section 2.4 for an example of path-dependent benefit).

[3] A function $f(x)$ is $o(g(x))$ when $x \to 0$ if $\lim_{x\to 0} \frac{f(x)}{g(x)} = 0$.

Now we can look at what happens to this portfolio after we perform the time step from t to $t + dt$. We use the first order approximation to approximate $e^{-(c-\kappa)dt}$ and e^{rdt}. Then the right hand side of (2.7) becomes

$$\begin{aligned}
e^{-\kappa T}e^{-(c-\kappa)(t+dt)}S_{t+dt} + e^{-\kappa T}e^{-(c-\kappa)t}S_t e^{rdt}(1 - e^{-(c-\kappa)dt}) \\
&= e^{-\kappa T}e^{-(c-\kappa)(t+dt)}S_{t+dt} + e^{-\kappa T}e^{-(c-\kappa)t}S_t(1 + rdt)(c-\kappa)dt + o(dt) \\
&= e^{-\kappa T}e^{-(c-\kappa)(t+dt)}S_{t+dt} + e^{-\kappa T}e^{-(c-\kappa)t}S_t(c-\kappa)dt + o(dt) \\
&= e^{-\kappa(T-(t+dt))}F_{t+dt} + e^{-\kappa(T-t)}(c-\kappa)F_t dt + o(dt)
\end{aligned}$$

The first part of the expression is the cost of buying back the variable annuity. Then the policyholder is left with the benefit of surrender of $h(t) := e^{-\kappa(T-t)}(c-\kappa)F_t + g(dt)$, where $g(dt)$ is $o(dt)$. $\square$

Using Proposition 2.3.1 and the trading strategy explained above we are now able to derive a pricing formula for the variable annuity contract with a surrender benefit similarly to Kim and Yu (1996).

Theorem 1. *Let $V^*(t, F_t)$ denote the continuation value at time t of the variable annuity with guarantee G at maturity and a surrender benefit equal to the accumulated fund value with some penalty $\kappa > 0$, $e^{-\kappa(T-t)}F_t$. Then $V^*(t, F_t)$ can be decomposed into a corresponding European part $U(t, F_t)$ and a surrender option $e(t, F_t)$*

$$V^*(t, F_t) = U(t, F_t) + e(t, F_t), \tag{2.8}$$

where

$$\begin{cases} U(t, F_t) = e^{-c(T-t)}F_t\mathcal{N}\big(d_1(F_t, G, T, t)\big) + e^{-r(T-t)}G\mathcal{N}\big(d_2(F_t, G, T, t)\big), \\ e(t, F_t) = e^{-\kappa T}(c-\kappa)F_t e^{ct}\int_t^T e^{-(c-\kappa)u}\mathcal{N}\big(d_1(F_t, B_u, u, t)\big)du, \end{cases} \tag{2.9}$$

and $\mathcal{N}(x)$ is the standard normal distribution function with d_1 and d_2 defined as

$$\begin{cases} d_1(x, y, T, t) := \frac{\ln(\frac{x}{y})+(r-c+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \\ \\ d_2(x, y, T, t) := \sigma\sqrt{T-t} - d_1(x, y, T, t). \end{cases} \tag{2.10}$$

Proof. First we prove the formula for the European part $U(t, F_t)$ of the VA. Since $F_T|F_t \sim \mathcal{LN}(\ln(F_t) + (r - c - \frac{\sigma^2}{2})(T-t), \sigma^2(T-t))$, we can use (2.5) to calculate the

European part of the VA. Define d_1 and d_2 as in (2.10). Then it follows that

$$\begin{aligned} U(t,F_t) &= e^{-r(T-t)}\Bigg[F_t e^{(r-c)(T-t)}\mathcal{N}\Big(\frac{-\ln(G)+\ln(F_t)+(r-c+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\Big) \\ &\quad +G\mathcal{N}\Big(\frac{\ln(G)-\ln(F_t)-(r-c-\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\Big)\Bigg], \end{aligned}$$

and we find the first part of (2.9). Secondly, we prove the formula for the surrender option $e(t,F_t)$. Define $\widetilde{\mu}(x) := \ln F_t) + (r-c-\frac{\sigma^2}{2})(x-t)$ and $\widetilde{\sigma}^2(x) := \sigma^2(x-t)$. From Proposition 2.3.1, the benefits on surrender amount to $e^{-\kappa(T-u)}(c-\kappa)F_u du + g(du)$ for each period of time $[u, u+du]$ during which the account value F_u is above the optimal surrender boundary B_u. We divide the remaining duration of the contract, $T-t$, into N intervals of length $du = \frac{T-t}{N}$ and denote $u_k = t + k\ du$, for $k = 1, 2, \ldots, N$. Therefore, the surrender option at $t < T$ can be approximated by

$$\sum_{k=1}^{N} e^{-r(u_k-t)} \int_{B_{u_k}}^{\infty} (e^{-\kappa(T-u_k)}(c-\kappa)x\ du + g(du)) f_{F_{u_k}}(x|F_t)dx.$$

To get the desired result in continuous time, we take the limit as $N \to \infty$ to obtain the following integral form.

$$\begin{aligned} e(t,F_t) &= \int_t^T e^{-r(u-t)} \int_{B_u}^{\infty} e^{-\kappa(T-u)}(c-\kappa)x f_{F_u}(x|F_t)dxdu \\ &\overset{(2.3)}{=} (c-\kappa)\int_t^T e^{-\kappa(T-u)}e^{-r(u-t)}\int_{B_u}^{\infty} x\frac{1}{\sqrt{2\pi\sigma^2(u-t)}\,x}e^{-\frac{[\ln(\frac{x}{F_t})-(r-c-\frac{\sigma^2}{2})(u-t)]^2}{2\sigma^2(u-t)}}dxdu \\ &= (c-\kappa)\int_t^T e^{-\kappa(T-u)}e^{-r(u-t)}\int_{B_u}^{\infty}\frac{1}{\sqrt{2\pi\widetilde{\sigma}^2(u)}}e^{-\frac{[\ln(x)-\widetilde{\mu}(u)]^2}{2\widetilde{\sigma}^2(u)}}dxdu \\ &\overset{y=\ln(x)}{=} (c-\kappa)\int_t^T e^{-\kappa(T-u)}e^{-r(u-t)}\int_{\ln(B_u)}^{\infty}\frac{1}{\sqrt{2\pi\widetilde{\sigma}^2(u)}}e^{-\frac{[y-\widetilde{\mu}(u)]^2}{2\widetilde{\sigma}^2(u)}}e^y dydu \\ &= (c-\kappa)\int_t^T e^{-\kappa(T-u)}e^{-r(u-t)}\int_{\ln(B_u)}^{\infty}\frac{1}{\sqrt{2\pi\widetilde{\sigma}^2(u)}}e^{-\frac{[y-(\widetilde{\mu}(u)+\widetilde{\sigma}^2(u))]^2}{2\widetilde{\sigma}^2(u)}}\underbrace{e^{\widetilde{\mu}(u)+\frac{\widetilde{\sigma}^2(u)}{2}}}_{=F_t e^{(r-c)(u-t)}}dydu \end{aligned}$$

$$
\begin{aligned}
&= (c-\kappa)F_t \int_t^T e^{-\kappa(T-u)} e^{-c(u-t)} \left[1 - \mathcal{N}\left(\frac{\ln(B_u) - (\widetilde{\mu}(u) + \widetilde{\sigma}^2(u))}{\widetilde{\sigma}(u)}\right)\right] du \\
&= (c-\kappa)F_t \int_t^T e^{-\kappa(T-u)} e^{-c(u-t)} \mathcal{N}\left(\frac{\ln(\frac{F_t}{B_u}) + (r - c + \frac{\sigma^2}{2})(u-t)}{\sigma\sqrt{u-t}}\right) du \\
&= (c-\kappa)F_t \int_t^T e^{-\kappa(T-u)} e^{-c(u-t)} \mathcal{N}\big(d_1(F_t, B_u, u, t)\big) du.
\end{aligned}
$$

The expression for the surrender option in (2.9) follows. □

Theorem 1 provides a way to calculate the price of a VA with surrender benefit. However, since the surrender option depends on the optimal surrender boundary B_t, one needs to compute it first. In the following we derive the optimal surrender boundary condition analogously to Kim and Yu (1996).

First, note that at maturity $B_T = G$. We also know that along the surrender boundary we have

$$V^*(t, B_t) = e^{-\kappa(T-t)} B_t.$$

Thus, by formula (2.8) and (2.9) we have

$$
\begin{aligned}
B_t &= e^{\kappa(T-t)} \left(U(t, F_t) + e(t, F_t)\right) \\
&= e^{\kappa(T-t)} (e^{-c(T-t)} B_t e^{\kappa(T-t)} \mathcal{N}(d_1(B_t e^{\kappa(T-t)}, G, T, t)) + e^{-r(T-t)} G \mathcal{N}(d_2(B_t e^{\kappa(T-t)}, G, T, t)) \\
&\quad + (c-\kappa) B_t e^{(c-\kappa)t} \int_t^T e^{-(c-\kappa)u} \mathcal{N}\left(d_1\left(B_t e^{\kappa(T-t)}, B_u, u, t\right)\right) du). \qquad (2.11)
\end{aligned}
$$

This integral equation can be used to compute the optimal surrender boundary B_t. Observe, however, that in order to calculate B_t the optimal surrender boundary for future times must be known. Since it holds that $B_T = G$ at expiration, we work backwards to recursively recover the optimal surrender boundary. Because (2.11) does not have an analytic solution, numerical integration schemes must be used. Practically this can be done by dividing the interval $[0, T]$ into n equidistant subintervals $0 = t_0 < t_1 < ... < t_n = T$ where times t_i, $i = 0, .., n$, represent the only possible early surrender times. Define $g(u) := e^{-(c-\kappa)u} \mathcal{N}(d_1(B_t e^{\kappa(T-t)}, B_u, u, t))$. Then, the integral in (2.11) is approximated by

$$I(k) = \frac{T}{n} \sum_{i=1}^{k-1} g(t_{n-i}), \qquad k = 1, .., n. \qquad (2.12)$$

Note, that at time t_{n-1} the early surrender premium $I(1)$ is equal to zero because there is no possibility for the policyholder to surrender the option in the last interval. Therefore, the premium has to be zero.

Proposition 2.3.2 (Derivation of the optimal surrender boundary)**.** *The following backward procedure generates an approximation to the surrender boundary.*

- $B_{t_n} = B_T = G$.
- *Recursively, for $k = 1..n$, compute $I(k)$ in* (2.12) *to approximate the integral part of* (2.11) *and solve the following equation for the only unknown $B_{t_{n-k}}$*

$$
\begin{aligned}
B_{t_{n-k}} &= e^{-c(T-t_{n-k})} B_{t_{n-k}} e^{\kappa(T-t_{n-k})} \mathcal{N}(d_1(B_{t_{n-k}} e^{\kappa(T-t_{n-k})}, G, T, t_{n-k})) \\
&\quad + e^{-r(T-t_{n-k})} G \mathcal{N}(d_2(B_{t_{n-k}} e^{\kappa(T-t_{n-k})}, G, T, t_{n-k})) + (c-\kappa) B_{t_{n-k}} e^{(c-\kappa)t_{n-k}} I(k).
\end{aligned}
$$

The method described in this section can be extended to any path-independent payoff for which $\phi(F^c_\bullet(T), T) = \ell(F^c_T, T)$ for some function $\ell(\cdot)$. In Section 2.4 we illustrate how to derive the optimal surrender boundary when $\phi(F^c_\bullet(T), T)$ is path-dependent, that is it depends on the path $\{F_t\}_{t\in[0,T]}$.

2.3.1 Alternative derivation of the optimal surrender boundary

We now present a second way to derive the optimal surrender boundary. Here, we will assume that there is no surrender charge (i.e. $\kappa = 0$), but the method still works as long as κ is such that the optimal surrender strategy is of the threshold type. The method presented here is based on the method used in Kim (1990) to price American calls on dividend-paying stocks. The main idea behind the technique is to assume that the contract can only be surrendered at a finite number of times, and to calculate the value of the contract at each of those surrender times. The appropriate limit is then taken to retrieve the integral representation presented in Theorem 1.

Proposition 2.3.3 gives the continuation value of the contract when it can only be surrendered at a finite number of points in time.

Proposition 2.3.3. *Assume that the surrender option can only be exercised at a finite number of points in time and denote these times by t_k, $k = 0, 1, ..., n-1, n$, where $t_0 = 0$, $t_n = T$ and $t_k - t_{k-1} = \Delta t = \frac{T}{n}$. For simplicity of notation, write $F_{t_k} = F_k$ and $B_{t_k} = B_k$, and denote by $\upsilon(t_{n-m}, F_{n-m}; B_{n-m+1})$ the continuation value of the VA contract at time t_{n-m}, assuming that it can only be surrendered at a finite number of points between 0 and T. Then, $\upsilon(t_{n-m}, F_{n-m}; B_{n-m+1})$ is given by*

$$
\begin{aligned}
&\upsilon(t_{n-m}, F_{n-m}; B_{n-m+1}) = U(t_{n-m}, F_{n-m}) \\
&\quad + \sum_{k=1}^{m-1} e^{-(m-k)r\Delta t} E[(F_{n-k} - E[e^{-r\Delta t} F_{n-k+1} | \mathcal{F}_{n-k}]) \mathbb{1}_{\{F_{n-k} \geqslant B_{n-k}\}} | \mathcal{F}_{n-m}] - \hbar(\Delta t),
\end{aligned}
\tag{2.13}
$$

where $\hbar(\Delta t)$ *is* $\mathcal{O}(\Delta t)^4$, *and with the optimal exercise boundary* B_{n-k} *defined implicitly and recursively by*

$$\begin{aligned}B_{n-k} &= U(t_{n-k}, B_{n-k}) \\ &+ \sum_{j=1}^{k-1} e^{-(k-j)r\Delta t} E[(F_{n-j} - E[e^{-r\Delta t} F_{n-j+1} | \mathcal{F}_{n-j}]) \mathbb{1}_{\{F_{n-j} \geqslant B_{n-j}\}} | F_{n-k} = B_{n-k}] - \hbar(\Delta t). \quad (2.14)\end{aligned}$$

Proof. To obtain the value of the contract, we use backward induction. At maturity, the option is automatically exercised when $F_T < G$. Thus, we let $B_n = G$. We can then move back one period and consider the price of the VA contract one period before expiry. To clarify the notation, we write $\upsilon(t_k, F_k; B_{k+1})$ to denote the continuation value of the VA contract given the optimal surrender boundary one period later. At time t_{n-1}, since the surrender option cannot be exercised between time t_{n-1} and T, we have

$$\begin{aligned}\upsilon(t_{n-1}, F_{n-1}; B_n) &= E[e^{-r\Delta t} \max(F_n, G) | \mathcal{F}_{n-1}] \\ &= U(t_{n-1}, F_{n-1}).\end{aligned}$$

We denote the optimal boundary at time t_{n-1} by B_{n-1} and define it implicitly by

$$B_{n-1} = V^*(t_{n-1}, B_{n-1}; B_n).$$

This means that when the fund value at time t_{n-1} is above B_{n-1}, it is optimal to surrender the contract.

We now move back to time t_{n-2} and calculate $\upsilon(t_{n-2}, F_{n-2}; B_{n-1})$ as the risk-neutral expectation of the value of the contract at t_{n-1}. We know that the surrender option will be exercised if the fund value is greater than the optimal surrender boundary. Thus, when F_{n-1} is above B_{n-1}, the value of the contract at t_{n-1} is simply the fund value. For F_{n-1} below the optimal boundary B_{n-1}, the value of the contract is the value of the live contract calculated in the previous step.

[4] A function $f(x)$ is $\mathcal{O}(g(x))$ when $x \to 0$ if and only if there exist positive numbers K and δ such that $|f(x)| \geqslant K|g(x)|$ for $|x| < \delta$.

Then, we have

$$\begin{aligned}\upsilon(t_{n-2}, F_{n-2}, B_{n-1}) &= E[e^{-r\Delta t}\upsilon(t_{n-1}, F_{n-1}; B_n)\mathbb{1}_{\{F_{n-1}<B_{n-1}\}}|\mathcal{F}_{n-2}] \\ &\quad + E[e^{r\Delta t}F_{n-1}\mathbb{1}_{\{F_{n-1}\geqslant B_{n-1}\}}|\mathcal{F}_{n-2}] \\ &= U(t_{n-2}, F_{n-2}) - E[e^{-r\Delta t}\mathbb{1}_{\{F_{n-1}\geqslant B_{n-1}\}}E[e^{r\Delta t}\max(F_n, G)|\mathcal{F}_{n-1}]|\mathcal{F}_{n-2}] \\ &\quad + E[e^{r\Delta t}F_{n-1}\mathbb{1}_{\{F_{n-1}\geqslant B_{n-1}\}}|\mathcal{F}_{n-2}] \\ &= U(t_{n-2}, F_{n-2}) + E[e^{-r\Delta t}(F_{n-1} - E[e^{-r\Delta t}F_n|\mathcal{F}_{n-1}])\mathbb{1}_{F_{n-1}>B_{n-1}}|\mathcal{F}_{n-2}] \\ &\quad - E[e^{-r\Delta t}\mathbb{1}_{\{F_{n-1}>B_{n-1}\}}E[(G - F_n)\mathbb{1}_{\{F_n<G\}}|\mathcal{F}_{n-1}]|\mathcal{F}_{n-2}] \qquad (2.15)\end{aligned}$$

Kim (1990) shows that the last term of (2.15) is $\mathcal{O}(\Delta t)$. In fact, it is the price of a put option issued at t_{n-1}, with maturity T and strike price K, discounted back to time t_{n-2}, conditional on F_{n-1} begin greater than B_{n-1}. We show that as Δt goes to 0, this integral also goes to 0. Intuitively, if the option is out-of-the-money a very short time before expiry, then its value will be very small because the probability that it becomes in-the-money before maturity is very low. A more rigourous proof is given in the Appendix of Kim (1990). Then, the price of the contract at t_{n-1} is given by

$$\begin{aligned}\upsilon(t_{n-2}, F_{n-1}, B_{n-1}) &= U(t_{n-2}, F_{n-2}) + E[e^{-r\Delta t}(F_{n-1} - E[e^{-r\Delta t}F_n|\mathcal{F}_{n-1}])\mathbb{1}_{\{F_{n-1}\geqslant B_{n-1}\}}|\mathcal{F}_{n-2}] \\ &\quad + \hbar_{n-2}(\Delta t), \qquad (2.16)\end{aligned}$$

where $\hbar_{n-m}(\Delta t)$ is a function that contains the terms of order Δt or higher at time $n - m$.

To complete the induction, we show that the formula holds for a general step $n - m$. We suppose that

$$\begin{aligned}&\upsilon(t_{n-m}, F_{n-m}; B_{n-m+1}) = U(t_{n-m}, F_{n-m}) \\ &+ \sum_{k=1}^{m-1} e^{-(m-k)r\Delta t}E[(F_{n-k} - E[e^{-r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m}] + \hbar_{n-m}(\Delta t). \qquad (2.17)\end{aligned}$$

Then, the optimal exercise boundary B_{n-m} is defined implicitly by

$$\begin{aligned}&B_{n-m} = U(t_{n-m}, B_{n-m}) \\ &+ \sum_{k=1}^{m-1} e^{-(m-k)r\Delta t}E[(F_{n-k} - E[e^{-r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|F_{n-m} = B_{n-m}] \\ &+ \hbar_{n-m}(\Delta t). \qquad (2.18)\end{aligned}$$

As in the previous step, we move back one period and calculate $\upsilon(t_{n-m-1}, F_{n-m-1}; B_{n-m})$ as the risk-neutral expectations of the value of the live contract at t_{n-m}. We can show that

$$\begin{aligned}&\upsilon(t_{n-m-1}, F_{n-m-1}; B_{n-m}) = U(t_{n-m-1}F_{n-m-1})\\&\quad+\sum_{k=1}^{m} e^{((m+1)-k)r\Delta t}E[(F_{n-k} - E[e^{-r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m-1}] + \hbar_{n-m}(\Delta t).\end{aligned}$$

This step completes the proof; details can be found in Appendix 2.B. □

Proposition 2.3.3 gives us a recursive equation for the value of the live VA contract when the surrender option can only be exercised at discrete times. The next step is to take the limit when $\Delta t \to 0$ to obtain the integral representation presented in Theorem 1. Since the fee is paid at a constant rate (that is, $c(t, F_t) = cF_t$), then $F_{n-m+1} = F_{n-m}e^{\left(r-c-\frac{\sigma^2}{2}\right)\Delta t+\sigma W_{\Delta t}}$ and $E[e^{-r\Delta t}F_{n-m+1}|\mathcal{F}_{n-m}] = F_{n-m}e^{-c\Delta t}$. So (2.13) from Proposition 2.3.3 becomes

$$\begin{aligned}&\upsilon(t_{n-m}, F_{n-m}; B_{n-m+1}) = U(t_{n-m}, F_{n-m})\\&\quad+\sum_{k=1}^{m-1} e^{-(m-k)r\Delta t}E[F_{n-k}(1-e^{-c\Delta t})\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m}] + \hbar(\Delta t).\end{aligned} \tag{2.19}$$

Now since $1 - e^{-c\Delta t} = c\Delta t + \mathcal{O}(\Delta t)$, we can re-write $\upsilon(t_{n-m}, F_{n-m}, ; B_{n-m+1})$ as

$$\begin{aligned}&\upsilon(t_{n-m}, F_{n-m}, ; B_{n-m+1}) = U(t_{n-m}, F_{n-m})\\&\quad+\sum_{k=1}^{m-1} e^{(m-k)r\Delta t}\int_{B_{n-k}}^{\infty} c\Delta t F_{n-k}p(F_{n-k}, (m-k)\Delta t; F_{n-m})dF_{n-k} + \hbar(\Delta t).\end{aligned} \tag{2.20}$$

Taking the limit as $n \to \infty$ (so that $\Delta t \to 0$) and defining $F_t = F_{n-m}$, we obtain

$$\upsilon(t, F_t; B_\bullet) = V^*(t, F_t) \tag{2.21}$$

$$= U(t, F_t) + \int_0^{T-t} e^{-r(T-t-s)}\int_{B(T-t-s)}^{\infty} cF_{T-t-s}p(F_{T-t-s}, T-t; F)dF_{T-t-s}ds$$

$$= U(t, F_t) + cFe^{-ct}\int_t^T Fce^{-cu}\mathcal{N}(d_1(F_t, B_u, u, t))ds, \tag{2.22}$$

where $d_1(x, y, T, t)$ is defined as in Theorem 1. Thus, we recover the result of Theorem 1 when $\kappa = 0$. □

2.4 Path-dependent payoff

In this section, we consider a path-dependent design for the payoff of the variable annuity. The example that we study is based on the payoff $\phi(F^c_\bullet) = \max(G, Y_T)$. In other words, the policyholder receives the maximum of the geometric average of the fund value at time T, Y_T, and the guarantee G at time T. The geometric average Y_t is defined as

$$Y_t = \exp\left(\frac{1}{t}\int_0^t \ln F_s ds\right). \tag{2.23}$$

Our goal is twofold. First we illustrate a general method to derive the optimal surrender strategy when there are path-dependent benefits. Second, we want to understand the impact of Asian benefits on the surrender incentive in VAs.

We need a few preliminary results. Defining the geometric average of the index as

$$\widetilde{Y}_t = \exp\left(\frac{1}{t}\int_0^t \ln S_s ds\right) \tag{2.24}$$

gives us the following relation between Y_t and $\widetilde{Y}_t$ at any time t

$$Y_t = e^{-\frac{ct}{2}}\widetilde{Y}_t. \tag{2.25}$$

An important difference with the setting of Section 2.3 is that this payoff is path-dependent since Y_t includes all values of F_s for times $s \in [0, t]$. We assume that the surrender benefit at time t is now also path-dependent and equal to

$$e^{-\kappa(T-t)}Y_t, \tag{2.26}$$

where κ is sufficiently small so that it can still be optimal to surrender the policy. In particular, throughout the section, we have the following assumption.

Assumption 2.4.1. *The parameters r, c and κ are such that*

- $\kappa < \frac{r+c+\frac{\sigma^2}{6}}{2}$, *and*
- $c < r - \frac{\sigma^2}{6}$.

Note that this assumption is not very restrictive. In fact, with a fee rate that would fail to meet the second criterion of Assumption 2.4.1, the policy would hardly be marketable.

In the same setting as described in Section 2.2, the conditional distribution of $\widetilde{Y}_u|(\widetilde{Y}_t, S_t)$ follows a lognormal distribution

$$\widetilde{Y}_u|(\widetilde{Y}_t, S_t) \sim \mathcal{LN}\left(\frac{t}{u}\ln \widetilde{Y}_t + \frac{u-t}{u}\ln S_t + \frac{r-\frac{\sigma^2}{2}}{2u}(u-t)^2\ ,\ \ \frac{\sigma^2}{3u^2}(u-t)^3\right).$$

This result is known and can be found for example in Hansen and Jørgensen (2000). Using the relationships (2.25) and (2.2), it is easy to show from the previous result on S_t and $\widetilde{Y}_t$ that

$$Y_u|(Y_t, F_t) \sim \mathcal{LN}\left(\frac{t}{u}\ln Y_t + \frac{u-t}{u}\ln F_t + \frac{r-c-\frac{\sigma^2}{2}}{2u}(u-t)^2\ ,\ \ \frac{\sigma^2}{3u^2}(u-t)^3\right). \tag{2.27}$$

Therefore, the conditional distribution function of Y_u given (Y_t, F_t) is known, similarly to the conditional distribution of $F_u|F_t$ in (2.3) which was key in the derivation of the early surrender premium for path-dependent benefits.

Using a similar trading strategy as in Section 2.3, we compute the early surrender premium of the variable annuity with Asian benefits and are able to prove the following proposition.

Proposition 2.4.1. *The benefit associated with the exercise of the surrender option between* $[t, t+dt]$ *is equal to*

$$h(t, Y_t, F_t) = e^{-\kappa(T-t)}Y_t\left(r - \kappa + \frac{1}{t}\ln\left(\frac{Y_t}{F_t}\right)\right)dt + g^g(dt),$$

when at time t*, it is optimal to surrender with* (Y_t, F_t)*, and where* $g^g(dt)$ *is* $o(dt)$ *as* $dt \to 0$*.*

Proof. The proof is in the same spirit as the proof of Proposition 2.3.1 for path-independent benefits. At the optimal boundary, the value of the VA is exactly equal to the surrender benefit (2.26), therefore

$$V^*(t, F_t, Y_t) = e^{-\kappa(T-t)}Y_t$$

At time $t + dt$, the value of the contract at the surrender boundary is

$$V^*(t+dt, F_{t+dt}, Y_{t+dt}) = e^{-\kappa(T-t-dt)}Y_{t+dt}$$

Assume that the VA is surrendered at time t, then the policyholder receives $e^{-\kappa(T-t)}Y_t$, we now have to compute how much is gained by staying out of the contract between t and $t+dt$. The main difficulty is to find a trading strategy at time t to ensure that we are able to re-enter the contract at $t+dt$ and to measure the profit from this strategy needed in the calculation of the surrender option.

Let us compute at time t the no-arbitrage value of $e^{-\kappa(T-t-dt)}Y_{t+dt}$. To do so, consider $u > t$ and compute first

$$E[e^{-r(u-t)}Y_u|\mathcal{F}_t] = e^{-r(u-t)}\exp\left(\frac{t}{u}\ln Y_t + \frac{u-t}{u}\ln F_t + \frac{r-c-\frac{\sigma^2}{2}}{2u}(u-t)^2 + \frac{\sigma^2(u-t)^3}{6u^2}\right)$$

using the conditional distribution of $Y_u|(Y_t, F_t)$. For $u = t + dt$, we find that

$$E[e^{-rdt}Y_{t+dt}|\mathcal{F}_t] = e^{-rdt}\exp\left(\frac{t}{t+dt}\ln Y_t + \frac{dt}{t+dt}\ln F_t + \frac{r-c-\frac{\sigma^2}{2}}{2(t+dt)}dt^2 + \frac{\sigma^2 dt^3}{6(t+dt)^2}\right)$$

$$\to E[e^{-rdt}e^{-\kappa(T-t-dt)}Y_{t+dt}|\mathcal{F}_t] = e^{-\kappa(T-t-dt)}e^{-rdt}\exp\left((1-\frac{dt}{t})\ln Y_t + \frac{dt}{t}\ln F_t + o(dt)\right),$$

which can be further simplified into

$$E[e^{-rdt}e^{-\kappa(T-t-dt)}Y_{t+dt}|\mathcal{F}_t] = e^{-\kappa(T-t)}Y_t - e^{-\kappa(T-t)}Y_t\left(r-\kappa+\frac{1}{t}\ln\left(\frac{Y_t}{F_t}\right)\right)dt + o(dt).$$

At time t, the policyholder receives $e^{-\kappa(T-t)}Y_t$. Note the following decomposition,

$$e^{-\kappa(T-t)}Y_t = e^{-\kappa(T-t-dt)}E[e^{-rdt}Y_{t+dt}|\mathcal{F}_t] + e^{-\kappa(T-t)}Y_t\left(r-\kappa+\frac{1}{t}\ln\left(\frac{Y_t}{F_t}\right)\right)dt + o(dt)$$

One can invest $Y_t e^{-\kappa(T-t)} - Y_t e^{-\kappa(T-t)}\left(r-\kappa+\frac{1}{t}\ln\left(\frac{Y_t}{F_t}\right)\right)dt$ at time t in the delta hedging strategy that generates $e^{-\kappa(T-t)}Y_{t+dt}$ at time $t+dt$. The remainder is left in a bank account at time t, so that the surrender option between t and $t+dt$ can be computed as $h(t, Y_t, F_t)$ in Proposition 2.4.1. □

Note that it seems that the surrender option can be negative. This is actually not the case, as if it is optimal to surrender at time t, then one cannot get more value by waiting for another dt, therefore

$$Y_t e^{-\kappa(T-t)} \geqslant E[e^{-rdt}Y_{t+dt}e^{-\kappa(T-t-dt)}|\mathcal{F}_t],$$

and thus $h(t, Y_y, F_t) \geqslant 0$ at any time t when it is optimal to surrender with (Y_t, F_t).

Proposition 2.4.2. *Let F_t denote the fund value process given in* (2.2) *and Y_t the geometric average based on F_t given in* (2.23). *Then, for $u > t$,*

$$Y_u|(Y_t, F_t, F_u = f) \sim \mathcal{LN}\left(M_f, \widehat{\mathcal{V}}_{u,t}\right), \tag{2.28}$$

where

$$\begin{cases} M_f := M_{Y_u|Y_t,F_t,F_u=f} = \frac{t}{u}\ln Y_t + \frac{1}{2}\frac{u-t}{u}\ln F_t + \frac{u-t}{2u}\ln f, \\ \widehat{\mathcal{V}}_{u,t} := \mathcal{V}_{Y_u|Y_t,F_t,F_u=f} = \frac{\sigma^2}{12u^2}(u-t)^3. \end{cases}$$

Proof. Conditionally on (Y_t, F_t), we have that $(\ln(Y_u), \ln(F_u))$ is a bivariate normal distribution. Thus $\ln(Y_u)|(\ln(F_u), F_t, Y_t)$ is normally distributed with mean $M_{Y_u|Y_t,F_t,F_u}$ and variance $\mathcal{V}_{Y_u|Y_t,F_t,F_u}$. To compute the conditional moments of $X|Y$ where $X = \ln Y_u|F_t, Y_t$ and $Y = \ln F_u|F_t, Y_t$ for $u > t$ we first compute

$$\begin{cases} \mathbb{E}[X] = \mathbb{E}[\ln Y_u|F_t, Y_t] = \frac{t}{u}\ln Y_t + \frac{u-t}{u}\ln F_t + \frac{r-c-\frac{\sigma^2}{2}}{2u}(u-t)^2 \\ \mathbb{E}[Y] = \mathbb{E}[\ln F_u|F_t, Y_t] = \mathbb{E}[\ln F_u|F_t] = \ln(F_t) + (r - c - \frac{\sigma^2}{2})(u-t) \\ \mathrm{Var}[X] = \mathrm{Var}[\ln Y_u|F_t, Y_t] = \frac{\sigma^2}{3u^2}(u-t)^3 \\ \mathrm{Var}[Y] = \mathrm{Var}[\ln F_u|F_t, Y_t] = \mathrm{Var}[\ln F_u|F_t, Y_t] = \sigma^2(u-t) \\ \mathrm{cov}[X,Y] = \mathrm{cov}[\ln F_u, \ln Y_u|F_t, Y_t] = \frac{\sigma^2}{2}\frac{(u-t)^2}{u} \\ \mathrm{corr}[X,Y] = \frac{\sqrt{3}}{2} \end{cases} \tag{2.29}$$

using (2.3) and (2.27) for the conditional means and variances. The only missing element is the covariance. From (2.23), recall that $Y_u = Y_t^{\frac{t}{u}} e^{\frac{1}{u}\int_t^u \ln(F_s)ds}$. It is thus clear that

$$\mathrm{cov}[X,Y] = \mathrm{cov}\left[\ln F_u, \frac{t}{u}\ln Y_t + \frac{1}{u}\int_t^u \ln F_s ds|F_t, Y_t\right] = \mathrm{cov}\left[\ln F_u, \frac{1}{u}\int_t^u \ln F_s ds|F_t, Y_t\right]$$

Using the linearity of the covariance

$$\mathrm{cov}[X,Y] = \frac{1}{u}\int_t^u \mathrm{cov}\left[\ln F_u, \ln F_s|F_t, Y_t\right] ds$$

where we are left with the computation of $\mathrm{cov}\left[\ln F_u, \ln F_s|F_t, Y_t\right]$ for $t \leqslant s \leqslant u$. It is clear that

$$\begin{aligned} \mathrm{cov}\left[\ln F_u, \ln F_s|F_t, Y_t\right] &= \sigma^2 \mathrm{cov}\left[(W_u - W_t), (W_s - W_t)|F_t, Y_t\right] \\ &= \sigma^2 \mathrm{cov}\left[W_{u-t}, W_{s-t}|F_0, Y_0\right] \\ &= \sigma^2 \min(u-t, s-t). \end{aligned}$$

Integrating over s gives the desired result. Then using the inputs in (2.29) and the well-known conditional moments of a bivariate normal distribution

$$\begin{aligned} M_{Y_u|Y_t,F_t,F_u} &= \mathbb{E}(X) + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}(Y - \mathbb{E}(Y)) \\ \mathcal{V}_{Y_u|Y_t,F_t,F_u} &= (1-\rho^2)\,\mathrm{var}(X), \end{aligned}$$

where $\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$. The claim follows and we have that $Y_u|Y_t, F_t, F_u = f$ is distributed according to a LogNormal distribution with these moments. □

We can now state a result for the surrender option with Asian benefits, similar to the one derived in Section 2.3.

Theorem 2. *Let $V^{*g}(t, F_t, Y_t)$ denote the continuation value at time t of the variable annuity with guarantee G at maturity and a surrender benefit equal to the accumulated geometric average $e^{-\kappa(T-t)}Y_t$, and suppose that Assumption 2.4.1 holds. Then $V^{*g}(t, F_t, Y_t)$ can be decomposed into a corresponding "European" part, $U^g(Y_t, F_t, t)$, and a surrender option, $e^g(Y_t, F_t, t)$, that is*

$$V^{*g}(t, F_t, Y_t) = U^g(Y_t, F_t, t) + e^g(Y_t, F_t, t), \tag{2.30}$$

where

$$\begin{aligned} U^g(Y_t, F_t, t) &= e^{-r(T-t)} e^{M_t^g + \frac{\mathcal{V}_t^g}{2}} \mathcal{N}\Big(\frac{-\ln(G) + M_t^g + \mathcal{V}_t^g}{\sqrt{\mathcal{V}_t^g}}\Big) \\ &\quad + e^{-r(T-t)} G \mathcal{N}\Big(\frac{\ln(G) - M_t^g}{\sqrt{\mathcal{V}_t^g}}\Big), \end{aligned} \tag{2.31}$$

$$e^g(Y_t, F_t, t) = e^{-\kappa T} e^{rt} \int_t^T e^{u(\kappa - r)} e^{\frac{\widehat{\mathcal{V}}_{u,t}}{2}} Y_t^{\frac{t}{u}} F_t^{\frac{u-t}{2u}} E\left[k(u, F_u, t)\right] du \tag{2.32}$$

where M_t^g and $\mathcal{V}_t^g$ are the conditional moments of $\ln(Y_T)|(Y_t, F_t)$ *given in* (2.27)*, F_u is LogNormal with density $f_{F_u}^{(c)}(f|F_t)$ in* (2.3) *and where*

$$k(u, f, t) = f^{\frac{u-t}{2u}} \left(\mathcal{N}\left(\frac{H_u(B_u(f), f)}{\sqrt{\widehat{\mathcal{V}}_{u,t}}} \right) \left(\frac{H_u(f, f)}{u} + r - \kappa \right) + \frac{\sqrt{\widehat{\mathcal{V}}_{u,t}}}{u\sqrt{2\pi}} \, e^{-\frac{1}{2}\frac{H_u(B_u(f), f)^2}{\widehat{\mathcal{V}}_{u,t}}} \right)$$

with $H_u(x, f) = M_f + \widehat{\mathcal{V}}_{u,t} - \ln(x)$ and where M_f and $\widehat{\mathcal{V}}_{u,t}$ are the conditional moments of $Y_u|(Y_t, F_t, F_u = f)$ for $u > t$ given in Proposition 2.4.2.

Remark 2.4.1. *As in Theorem 1, the value of the surrender option also depends on the fee rate c, although in a more indirect way. The dependence on c comes from the density $f_{F_u}^{(c)}(f|F_t)$ of F_u, which is a function of c. In this theorem, we slightly modify the notation and add the subscript $^{(c)}$ to highlight the link between the fee rate and the value of the surrender option.*

Proof. First we prove the formula for the European part $U^g(Y_t, F_t, t)$ of the VA. Since $Y_T|(Y_t, F_t) \sim \mathcal{LN}(M_t^g, \mathcal{V}_t^g)$, we can use (2.5) to calculate the European part of the VA. $U^g(Y_t, F_t, t)$ in (2.30) follows immediately. Secondly, we prove the formula for the surrender option $e^g(Y_t, F_t, t)$. Performing a similar substitution as in the derivation of the surrender option in Section 2.3, and taking a similar limit as in the proof of Theorem 1 to obtain the integral from a sum, we get

$$e^g(Y_t,F_t,t) = \int_t^T e^{-r(u-t)} \int_0^\infty \int_{B_u(f)}^\infty h(u,y,f) f_{Y_u}(y|Y_t,F_t,F_u=f)dy f_{F_u}(f|F_t,Y_t)df du \tag{2.33}$$

with $h(u,y,f)$ given in Proposition 2.4.1 and where the rationale is to derive the optimal boundary $B_u(f)$ for Y_u at time u given $F_u = f$. Indeed the optimal surrender policy at time u now depends on Y_u and F_u. We first condition on F_u and assume that $F_u = f$ is given. We then look for the critical level for Y_u which triggers the optimal surrender of the policy. The surrender region is of the form $Y_u > B_u(f)$.[5]

To compute $e^g(Y_t,F_t,t)$ note that $f_{F_u}(f|Y_t,F_t) = f_{F_u}(f|F_t)$ is known in (2.3), and that the distribution of $f_{Y_u}(y|Y_t,F_t,F_u=f)$ is given in Proposition 2.4.2. Let us thus simplify the surrender option (2.33) as

$$\int_t^T e^{-r(u-t)} e^{-\kappa(T-u)} \int_0^\infty \int_{B_u(f)}^\infty y\left(r-\kappa+\frac{1}{u}\ln\left(\frac{y}{f}\right)\right) f_{Y_u}(y|Y_t,F_t,F_u=f)dy f_{F_u}(f|F_t)df du$$

and thus

$$e^g(Y_t,F_t,t) = e^{-\kappa T}e^{rt}\int_t^T e^{u(\kappa-r)}\int_0^\infty \left[\left(r-\kappa-\frac{\ln(f)}{u}\right)\mathbb{E}_1+\frac{1}{u}E_2\right] f_{F_u}(f|F_t)df du$$

where $E_1 := E\left[\mathbb{1}_{Y>B_u(f)}Y\right]$ and $E_2 := \mathbb{E}\left[\mathbb{1}_{Y>B_u(f)}Y\ln(Y)\right]$, and where Y is lognormal with log moments M_f and $\widehat{\mathcal{V}}_{u,t}$ (mean and variance of $\ln(Y_u)|Y_t,F_t,F_u=f$ calculated in Proposition 2.4.2). It is then easy to prove that

$$\begin{cases} E_1 = \mathcal{N}\left(\frac{M_f+\widehat{\mathcal{V}}_{u,t}-\ln(B_u(f))}{\sqrt{\widehat{\mathcal{V}}_{u,t}}}\right)e^{M_f+\frac{\widehat{\mathcal{V}}_{u,t}}{2}} \\ E_2 = \sqrt{\widehat{\mathcal{V}}_{u,t}}B_u(f)^{1+\frac{M_f}{\widehat{\mathcal{V}}_{u,t}}}\frac{e^{-\frac{1}{2}\frac{M_f^2+(\ln(B_u(f)))^2}{\widehat{\mathcal{V}}_{u,t}}}}{\sqrt{2\pi}}+(M_f+\widehat{\mathcal{V}}_{u,t})E_1 \end{cases}$$

[5] See Appendix 2.C.

This observation allows us to further simplify the surrender option to

$$e^g(Y_t,F_t,t) = e^{-\kappa T}e^{rt}\int_t^T e^{u(\kappa-r)}\int_0^\infty \left[\frac{\sqrt{\widehat{\mathcal{V}}_{u,t}}}{u\sqrt{2\pi}}\; B_u(f)^{1+\frac{M_f}{\widehat{\mathcal{V}}_{u,t}}}\; e^{-\frac{1}{2}\frac{M_f^2+(\ln(B_u(f)))^2}{\widehat{\mathcal{V}}_{u,t}}}\right.$$
$$\left. +\mathcal{N}\left(\frac{M_f+\widehat{\mathcal{V}}_{u,t}-\ln(B_u(f))}{\sqrt{\widehat{\mathcal{V}}_{u,t}}}\right)e^{M_f+\frac{\widehat{\mathcal{V}}_{u,t}}{2}}\left(\frac{M_f+\widehat{\mathcal{V}}_{u,t}-\ln(f)}{u}+(r-\kappa)\right)\right] f_{F_u}(f|F_t)df\,du$$

Replacing $B_u(f)$ by $\exp(\ln(B_u(f)))$, noting that $\widehat{\mathcal{V}_{u,t}}$ does not depend on f, and denoting by $H_u(x,f) := M_f + \widehat{\mathcal{V}}_{u,t} - \ln(x)$, this expression further simplifies to

$$e^g(Y_t,F_t,t) = e^{-\kappa T}e^{rt}\int_t^T e^{u(\kappa-r)}e^{\frac{\widehat{\mathcal{V}}_{u,t}}{2}}\int_0^\infty \left[\frac{\sqrt{\widehat{\mathcal{V}}_{u,t}}}{u\sqrt{2\pi}}\; e^{-\frac{1}{2}\frac{H_u(B_u(f),f)^2}{\widehat{\mathcal{V}}_{u,t}}}\right.$$
$$\left. +\mathcal{N}\left(\frac{H_u(B_u(f),f)}{\sqrt{\widehat{\mathcal{V}}_{u,t}}}\right)\left(\frac{H_u(f,f)}{u}+r-\kappa\right)\right] e^{M_f} f_{F_u}(f|F_t)df\,du$$

then

$$e^g(Y_t,F_t,t) = e^{-\kappa T}e^{rt}\int_t^T e^{u(\kappa-r)}e^{\frac{\widehat{\mathcal{V}}_{u,t}}{2}}Y_t^{\frac{t}{u}}F_t^{\frac{u-t}{2u}}E\left[k(u,F_u,t)\right]du$$

where $k(u,f,t) = f^{\frac{u-t}{2u}}\left(\mathcal{N}\left(\frac{H_u(B_u(f),f)}{\sqrt{\widehat{\mathcal{V}}_{u,t}}}\right)\left(\frac{H_u(f,f)}{u}+r-\kappa\right)+\frac{\sqrt{\widehat{\mathcal{V}}_{u,t}}}{u\sqrt{2\pi}}\; e^{-\frac{1}{2}\frac{H_u(B_u(f),f)^2}{\widehat{\mathcal{V}}_{u,t}}}\right)$ and F_u is a LogNormal variable with density $f_{F_u}(f|F_t)$. □

Theorem 2 provides a formula for the price of a VA with Asian benefits including a surrender option. However, since the surrender option depends on the optimal surrender boundary $B_t(f)$ it is not an explicit formula that can be implemented directly. One first needs to compute this boundary in analogy to Kim and Yu (1996). Note that the value of $B_T(F_T)$ at maturity is known and equal to $B_T(F_T) = G$. The procedure is then similar to the one-dimensional case except that one has a double integral to compute.

To make the problem more tractable and reduce the number of equations to solve, we make the following assumption on the shape of the barrier. The benefit of this assumption appears clearly in Proposition 2.4.3 below, which describes the algorithm for the optimal surrender boundary.

Assumption 2.4.2. *Assume that the boundary $B_u(f)$ is given by the following form*

$$B_u(F_u) = \max(Ge^{-r(T-u)}, a_u + b_uF_u) \tag{2.34}$$

At any time before maturity, it is never optimal to surrender unless the immediate payoff is at least equal to the discounted value of the minimum terminal payoff G, as this is the minimum amount guaranteed at time T. We also know that along the surrender boundary it holds that

$$V^*(t, F_t, B_t(F_t)) = e^{-\kappa(T-t)} B_t(F_t) = \max(Ge^{-r(T-t)}, a_t + b_t F_t).$$

Thus, by formula (2.30)

$$B_t(F_t) = e^{\kappa(T-t)} \left(U^g(\max(Ge^{-r(T-t)}, a_t + b_t F_t), F_t, t) + e^g(\max(Ge^{-r(T-t)}, a_t + b_t F_t), F_t, t) \right).$$

This is an integral equation for the optimal surrender boundary because of the form of $e^g(\cdot, \cdot, \cdot)$ in (2.32). Observe, however, that in order to compute $\max(Ge^{-r(T-t)}, a_t + b_t F_t)$ at time t, the optimal surrender boundary for future times must be known. Since $B_T = G$ ($b_T = 0$) at expiration, we work backwards through time to recover the optimal surrender boundary recursively. Because formula (2.32) does not have an analytic solution, numerical integration schemes must be used. Practically this is done by dividing the interval $[0, T]$ into n equidistant subintervals $0 = t_0 < t_1 < ... < t_n = T$, where the times t_i, $i = 0, .., n$, represent the possible surrender times. Define $g(u) := e^{-\kappa T} e^{rt} e^{u(\kappa - r)} e^{\frac{\hat{\mathcal{V}}_{u,t}}{2}} Y_t^{\frac{t}{u}} F_t^{\frac{u-t}{2u}} E\left[k(u, F_u, t)\right]$. Then, the integral in (2.32) is approximated by

$$I(k) = \frac{T}{n} \sum_{i=1}^{k-1} g(t_{n-i}), \;\; k = 1, .., n. \tag{2.35}$$

Note, that at time t_{n-1}, $I(1) = 0$.

Proposition 2.4.3 (Derivation of the optimal surrender boundary)**.** *The following backward procedure generates the approximate surrender boundary.*

- $B_{t_n} = B_T = G$, $b_T = 0$.
- *Recursively, for* $k = 1..n$*:*
  - *For* m *values of* $F_{t_{n-k}}$*, compute the optimal boundary* $B_{t_{n-k}}(F_{t_{n-k}})$ *using* (2.35) *and solving*

    $$B_{t_{n-k}}(F_{t_{n-k}}) = e^{\kappa(T-t)} \left(U^g(B_{t_{n-k}}(F_{t_{n-k}}), F_{t_{n-k}}, t_{n-k}) + I(k) \right).$$

  - *Out of the* m *values obtained, use those above* $Ge^{-r(T-t)}$ *to perform a linear regression and obtain* $a_{t_{n-k}}$ *and* $b_{t_{n-k}}$*.*

A numerical illustration is given in the next section. Note that the technique described in this section will apply for other types of path-dependent benefits. The derivation holds when at any time u, conditional on the value of the underlying fund $F_u = f$ at time u, the optimal strategy is driven by checking whether some other quantity (here the geometric average) is above a level $B_u(f)$ (in other words the optimal strategy is a threshold strategy conditionally on the fund value at time u). Finally, note that the approximation (2.34) significantly simplifies the implementation as it locally approximates the surrender boundary with a piecewise linear function. From our numerical experiments we found that this is a satisfactory approximation.

2.5 Numerical Examples

This section presents some numerical examples to illustrate the techniques presented in Sections 2.3 and 2.4 respectively.

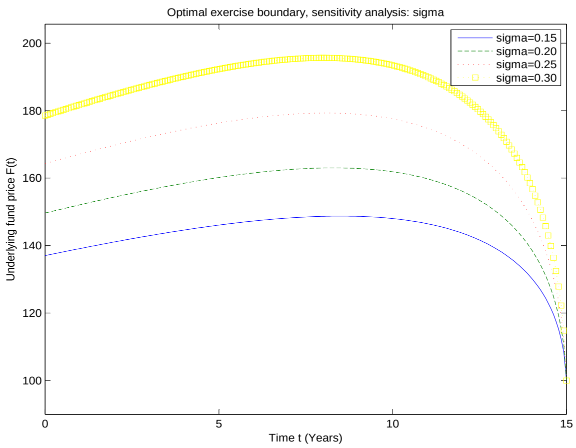
2.5.1 Optimal Boundary for the VA studied in Section 2.3

We perform a sensitivity analysis to further shed light on some properties of the surrender boundary derived in Section 2.3. Unless stated otherwise, we assume that $\kappa = 0$, $r = 0.03$, $\sigma = 0.2$ and $T = 15$ (years). The guaranteed amount G is equal to 100. The fair fee is $c^* = 0.91\%$, ignoring the surrender benefit.

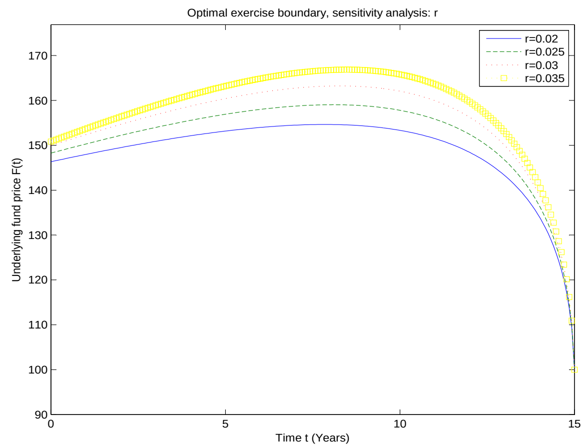
Figure 6.4 shows optimal surrender boundaries for the set of parameters given above when varying one parameter at a time. There are a few things to be noticed. First, as discussed earlier, the time zero value of the boundary is greater than the fund value at time 0 and the value at maturity T is equal to the guarantee G. Secondly, the graph of the surrender boundary is generally non-monotonic. The curve slowly increases to its maximum and then declines rapidly to G.

In the following we examine the sensitivity of the optimal surrender boundary with respect to the parameters σ, r, c, T, G and κ. Panel A of Figure 6.4 illustrates the sensitivity with respect to the volatility σ. We compute the surrender boundary for values of $\sigma = 15\%, 20\%, 25\%$ and 30%. We observe that as volatility increases the optimal surrender boundary gets pushed further up. With a high volatility, the policyholder would surrender the contract at higher values of the underlying fund than if she had invested in a fund with a lower volatility. Intuitively this result can be explained by the fact that the fund fluctuates more heavily if the volatility is higher. Therefore, the maturity benefit is more valuable.

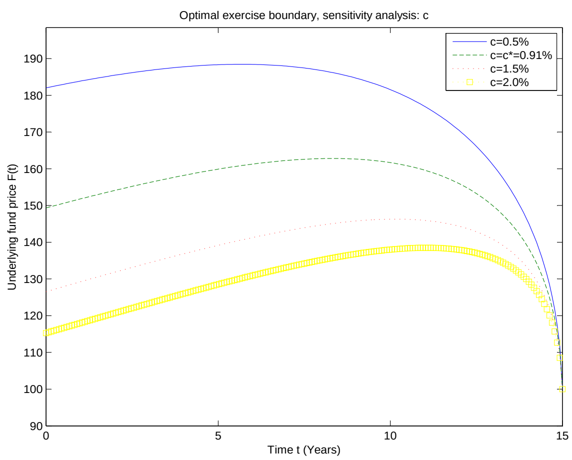
Panel B of Figure 6.4 displays the sensitivity with respect to the risk-free interest rate r. We vary the interest rate between 2% and 3.5% and compute the optimal surrender boundary. Similarly to the sensitivity with respect to the volatility, we observe that the optimal surrender boundary is higher for higher interest rates.



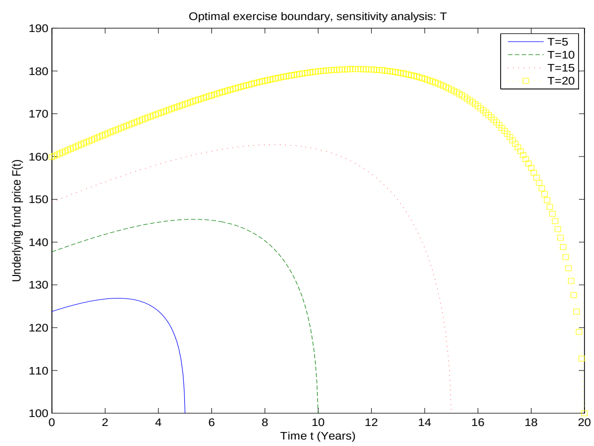
Panel A: Sensitivity to σ



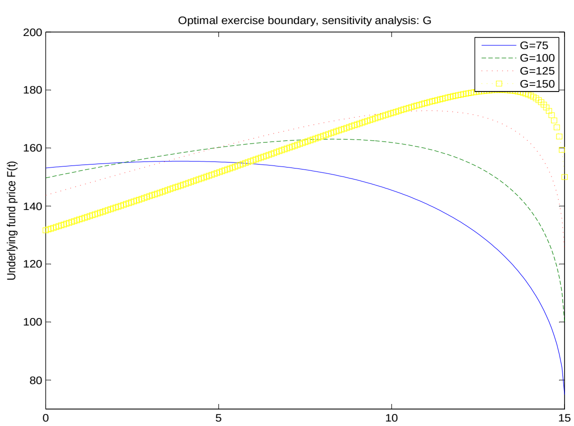
Panel B: Sensitivity to r



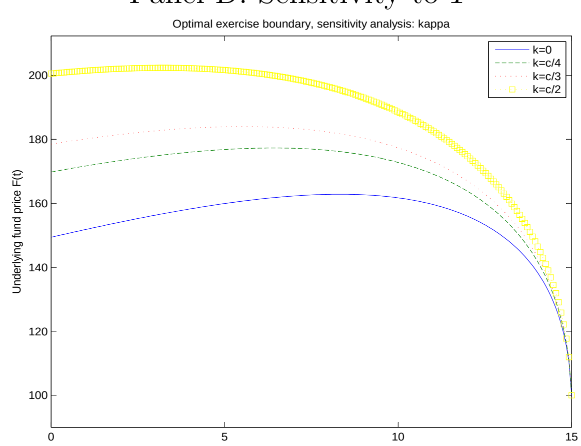
Panel C: Sensitivity to c



Panel D: Sensitivity to T



Panel E: Sensitivity to G



Panel F: Sensitivity to κ

Figure 2.1: Sensitivity analysis: The fee rates are computed to make the contract with the European benefit fair in all panels except in Panel C in which the sensitivity to the fee rate c is studied.

In Panel C of Figure 6.4 we show the sensitivity of the optimal surrender boundary of the sensitivity analysis with respect to the fee c. Since insurance companies do not always charge the fair fee, it is interesting to investigate what happens if the fee is somewhat higher or lower. In our case, the fee takes values from 0.5% to 2.0%. The figure shows that with a higher fee the optimal surrender boundary is lower. This is intuitive since with a higher fee, the policyholder has to pay more for the guarantee. Thus, the mismatch between the premium for the guarantee and its value is even greater resulting in earlier surrender times. This also increases the value of the surrender option, showing that increasing c is not a good way to pay for surrender benefits (see also Milevsky and Salisbury (2001)). We also observe that the optimal surrender boundary is very sensitive to changes in the fee. From an initial optimal surrender value of 150 at time zero for the fair fee, the optimal surrender value drops to about 115 for a fee of 2.0%. Likewise if the fee is reduced to 0.5% the optimal surrender value increases to just above 180.

Panel D of Figure 6.4 shows the sensitivity with respect to the maturity T. It illustrates that with increasing maturity the optimal surrender boundary increases as well. Considering a short time to maturity the fund value is less likely to reach high values. It is also known that the price of plain vanilla options are negatively correlated with the time to maturity, i.e. it loses value the closer it gets to maturity. Therefore, if we decrease the maturity T the option is worth less and should thus be surrendered at a lower fund level.

We analyse the sensitivity of the surrender boundary with respect to the guarantee G in Panel E of Figure 6.4. For $G = 75, 100, 125$ and 150 we compute the optimal surrender boundary B_t. The graphs look quite different from the ones above. We observe that the higher the guarantee the lower is the initial value of the surrender boundary. However at the same time the slope is higher for graphs with a higher guarantee. This effect can be explained by considering the fees c^* displayed in the table below. The fee for a contract with a guarantee of 150 is about 15 times greater than the fee of a contract with a guarantee of 75. So on the one hand the policyholder has a high guaranteed return at maturity. But on the other hand she has to pay a high fee for it. For this reason, it is better for the policyholder to surrender the contract earlier than if she had a lower guarantee implying a lower fee.

G	75	100	125	150
c^*	0.35%	0.91%	2.02%	5.28%

Lastly, Panel F of Figure 6.4 represents the effect of κ. The optimal boundary quickly moves up when κ increases: the surrender incentive is much lower because of the surrender penalty. In practice κ can be chosen high enough to have very low surrender incentives. Throughout our study we assumed $\kappa = 0$ to find the maximum risk for companies if they do not charge for the surrender option.

2.5.2 Optimal Boundary for the VA studied in Section 2.4

We illustrate the shape of the optimal boundary for a VA with path-dependent payoff in Figure 2.2. Since the optimal boundary depends on time t and on the value of the fund F_t, the optimal surrender boundary throughout the life of the contract must thus be represented by a surface (Figure 2.2). It is also possible to fix a value F_t and obtain a curve which shows the evolution of the boundary through time (as it is in Figure 2.3). We consider a 10-year contract with payoff $\max(Y_T, G)$ as defined in Section 2.4. Here, we assume that the guaranteed roll-up rate is 0.025 so that $G = F_0 e^{0.025T}$. We also assume that there is no surrender charge. Neglecting the surrender benefit, we use the fair fee $c^* = 0.0197$. Market assumptions are as in the previous section.

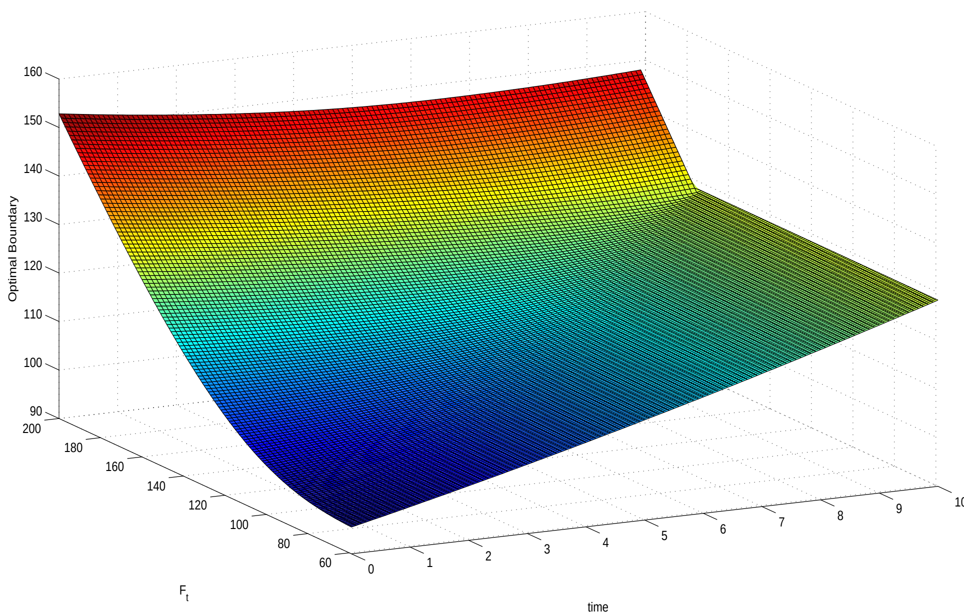


Figure 2.2: Optimal surrender surface for a 10-year geometric average VA with $G = F_0 e^{0.025\times 10}$ and $\kappa = 0$.

For high values of F_T, the boundary drops at maturity, because for any value $Y_T > G$, the option is exercised. However, before maturity, it is not necessarily optimal to surrender because the average of the fund might still increase. This is especially the case when $F_t > Y_t$. For low values of F_t, the boundary is close to $Ge^{-r(T-t)}$, the discounted value of G. When the fund value is low, it drags the average down and decreases the probability that the average at maturity is above the guarantee. Thus, for low values of F_t, it may be optimal to lapse the contract and cash in the gains earlier. In general, for a fixed time t, this causes the boundary to increase with F_t. This behaviour is more noticeable at the beginning of the contract since there is more time for the average to increase. The optimal surrender boundaries are relatively low, because the average is a lot less volatile than the fund. For this reason, it is often optimal to surrender early, even when the fund value is high, because the expected increase in the average is less than the risk-free rate. Thus, it would be optimal to withdraw the amount of the average and invest it at the risk-free rate. In fact, when Y_t increases past $Ge^{-r(T-t)}$, the value of the option drops quickly because of the low volatility of the average. This indicates that Asian-type maturity benefits tend to increase the value of the surrender option.

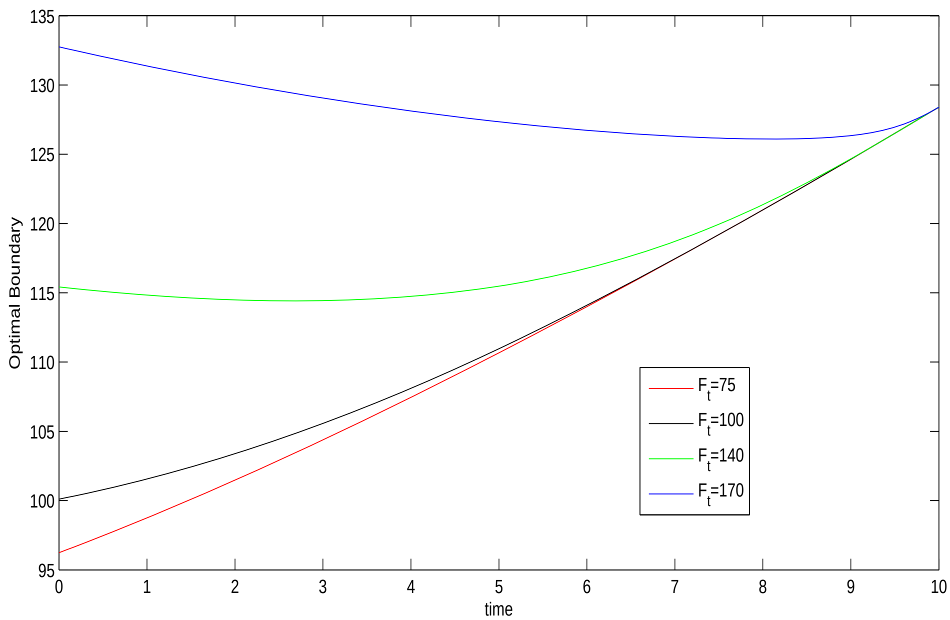


Figure 2.3: Optimal surrender boundary for a 10-year geometric average VA as a function of time for different values of F_t with $G = F_0e^{0.025\times 10}$ and $\kappa = 0$.

2.6 Concluding Remarks

In this chapter, we presented a method that allows us to derive a formula for the continuation value of a VA contract. We do so by decomposing the value into a corresponding European part and a surrender option. The latter can be expressed as an integral that depends on the optimal boundary at future times. Thus, the price of the contract and the surrender boundary must be solved iteratively through backward recursion. We implemented the formulae thus obtained and performed some numerical examples. They revealed that when the maturity benefit depends only on the terminal fund value, the optimal surrender boundary is a non-monotonic function which increases at first and then decreases to finally reach the guaranteed amount at maturity. By performing sensitivity analysis we found that with increasing volatility, interest rate, surrender charge and maturity the optimal surrender boundary is pushed up. If we increase the guarantee, however, we find a lower boundary at the beginning, but, due to a higher slope, the boundary takes higher values as maturity is approached before dropping back to the guaranteed level. This effect is explained by the higher fair fees for contracts with a high guarantee.

Our method is general enough to be used when the benefits are path-dependent. We considered the geometric average of the fund as an example of such a payoff. Analogously to the first case, we derived a pricing formula and an integral equation for the optimal surrender boundary which depends on the geometric average as well as on the fund value itself. We found that the low volatility of the geometric average decreases the value of the guarantee and increases surrender incentives.

In this chapter, we assumed that the underlying follows a geometric Brownian motion. Although this model is too simple to fit actual market data, it is sufficient to shed some light on the different factors influencing the optimal surrender boundary. Since the transition density of the underlying asset is known explicitly, we are able to obtain integral representations for the value of the surrender option. Our method can easily be extended to other market models as long as the model guarantees the existence of a portfolio that replicates the fund value using traded assets. In the case when the transition density is not known in explicit form, the method can still be used, without deriving an analytical form for the integrand but approximating it by Monte Carlo techniques for instance. Thus, our method can be extended to obtain the surrender boundary under more realistic market models.

The results obtained in this chapter, especially those pertaining to the maturity benefit that depends only on the terminal fund value, give an insight into the factors that influence the value of the option to surrender the variable annuity contract. In particular, the fee rate paid when the account value is high contributes to surrender incentives, both by influencing the optimal surrender boundary and by affecting the value of the surrender option. In the next chapters, we will explore different fee structures designed to reduce the surrender incentive by decreasing the fee rate paid when the fund value is high.

Appendix

2.A Optimal Surrender Region for GMAB

In this appendix, we discuss the shape of the optimal surrender region for a simple guaranteed minimum accumulation benefit. Throughout Chapter 2, we assume that the optimal strategy is of the threshold type, that is, for every time $t \in [0, T)$, the optimal surrender region has the form $[a_t, \infty)$, with $0 \leqslant a_t < \infty$. In order for our method to be valid, this condition must hold. Since the difference between the value of the maturity benefit and the account value decreases monotonically in F_t, it is somewhat intuitive that the strategy should be of the threshold type. However, proving it rigorously is not straight-forward.

For most American-type options studied in the past 40 years, the optimal exercise stratgy is of the threshold type. In other words, the optimal exercise region for any $t \in [0, T)$ is of the form $[a_t, \infty)$ or $[0, b_t]$, with $0 \leqslant a_t < \infty$ and $0 \leqslant b_t < \infty$. However, after Dayanik and Karatzas (2003) gave examples of options with optimal surrender regions that were not of the threshold type, more literature on the shape of this region emerged[6]. This has given rise to literature studying the conditions that American-type options must satisfy in order for their optimal exercise strategy to be of the threshold type. In particular, Villeneuve (2007) studies sufficient conditions for threshold strategies, while Strulovici and Szydlowski (2012) analyse the existence of optimal strategies. These papers focus on time-homogenous payoffs — payoffs that only depend on the value of the underlying. However, in our case, the payoff of the variable annuity with surrender option is inhomogeneous in time, since the financial guarantee is only applied at maturity. For this reason, we need to use the results obtained by Jönsson, Kukush, and Silvestrov (2005a) and Jönsson, Kukush, and Silvestrov (2005b) for time inhomogeneous payoffs in a discrete setting. By showing that our payoff function satisfies certain conditions, we can confirm that the optimal strategy, in discrete time, is of the threshold type. We can then use Amin and Khanna (1994)'s convergence results to confirm that the results also hold in continuous time. In this appendix, we

[6] In Chapter 4 of this thesis, we also present some variable annuity designs that have non-threshold optimal surrender strategies.

show how the results of Jönsson, Kukush, and Silvestrov (2005b) and Amin and Khanna (1994) can be applied to our problem.

A rigourous proof of the threshold strategy is out of the scope of this Appendix. To justify our assumption that the optimal strategy is to always surrender the contract when the account value is above a certain boundary, we show that in discrete time, our problem satisfies the conditions of Theorem 1 of Jönsson, Kukush, and Silvestrov (2005b), which confirms the one-threshold structure of the optimal stopping domain.

Since Jönsson, Kukush, and Silvestrov (2005b)'s results hold in discrete time, we must modify our original problem and set it in discrete time. Assume that the timeline from 0 to T is divided into n intervals of size $\Delta t = \frac{T}{n}$, with $t_0 = 0$, $t_1 = \Delta t$, $\ldots$, $t_n = T$, and that the VA contract can only be surrendered or exercised at times $t_0, t_1, \ldots, t_{n-1}, t_n$. Further assume that the underlying index S_t follows a binomial model, such that it approximates our Black-Scholes setting and converges to it in the limit. Following Cox, Ross, and Rubinstein (1979), we let the value of the index at time t_k, S_k, be given by $S_k = A(S_{k-1}, Y_k)$, for $k = 1, 2, \ldots, n$. Y_k is a random variable taking value $u = e^{\sigma \Delta t}$ with probability p and $d = e^{-\sigma \Delta t}$ with probability $1-p$, where $p = \frac{e^{r\Delta t} - d}{u - d}$. We also let $A(x, y) = xy$, so that $E[e^{-r\Delta t} A(S_{k-1}, Y_k) | S_{k-1}] = S_{k-1}$. Since the variable annuity fee is paid at a constant rate c, we denote the value of the account at time t_k by F_k and let $F_k = e^{-ct_k} S_k$. As in the continuous time case, we have $S_0 = F_0 = P$.

For each time t_k, the payoff function of the contract is given by

$$\psi_k(x) = \begin{cases} xe^{-\kappa(T-t_k)-ct}, & \text{if } k = 0, 1, \ldots, n-1. \\ \max(xe^{-cT}, G), & \text{if } k = n. \end{cases} \tag{2.36}$$

Note that the payoff functions are expressed here in terms of the index value.

To apply Theorem 1 of Jönsson, Kukush, and Silvestrov (2005b), we must check assumptions $\tilde{\mathbf{A}}\mathbf{2}$, $\tilde{\mathbf{G}}\mathbf{1}$, $\tilde{\mathbf{B}}\mathbf{1}$, $\tilde{\mathbf{E}}\mathbf{1}$ and $\tilde{\mathbf{E}}\mathbf{2}$, and assumption **A1** from Jönsson, Kukush, and Silvestrov (2005a). Assumption **A1** ensures that if $S_i = 0$, then $S_j = 0$ for all $i < j \leqslant n$. Assumption $\tilde{\mathbf{A}}\mathbf{2}$ requires that $A(x, y)$ be non-decreasing and convex in x, which is satisfied in our model. Assumption $\tilde{\mathbf{E}}\mathbf{2}$ is satisfied if for any $0 < d < \infty$, $\lim_{x\to\infty} P\left[A(x, Y_{k+1}) < d\right] = 0$. These three assumptions are easily satisfied by the binomial model. To satisfy condition $\tilde{\mathbf{G}}\mathbf{1}$, we must show that each function $\psi_k(x)$, $k = 0, 1, \ldots, n$ is non-decreasing and convex. For $k = 0, 1, ..., n-1$, the function is linear and increasing in x, so the condition is satisfied. It is also easy to show that $\psi_n(x)$ is non-decreasing and convex. Condition $\tilde{\mathbf{E}}\mathbf{1}$ states that for $k = 0, 1, \ldots, n-1$, we must have

$$\liminf_{x\to\infty} \frac{e^{-r\Delta t} E[\psi_{k+1}(S_{t_{k+1}}) | S_{t_k} = x]}{\psi_k(x)} < 1. \tag{2.37}$$

For $k = 0, 1, \ldots, n-2$, $E[\psi_{k+1}(S_{t_{k+1}}) | S_{t_k} = x] = xe^{-\kappa(T-t_{k+1})-c(t_k+\Delta t)}$ and $\psi_k(x) = xe^{-\kappa(T-t_k)-ct_k}$, so $\frac{e^{-r\Delta t} E[\psi_{k+1}(S_{t_{k+1}}) | S_{t_k}=x]}{\psi_k(x)} = e^{-(c-\kappa)\Delta t}$, which is less than 1 under our assumption that $\kappa < c$. For

$k = n-1$, $E[\psi_n(S_{t_n})|S_{t_{n-1}} = x] = xe^{-cT} + P(xe^{-ct_{n-1}}, \Delta t)$, where $P(x, \Delta t)$ is the of a European put option on the fund with $F_t = x$, with time to maturity Δt and strike G. So

$$\frac{e^{-r\Delta t} E[\psi_n(S_{t_{n-1}})|S_{t_n} = x]}{\psi_{n-1}(x)} = e^{-(c-\kappa)\Delta t} + \frac{e^{\kappa \Delta t} P(xe^{-ct_{n-1}\Delta t})}{xe^{-ct_{n-1}}}. \tag{2.38}$$

Since the second term of goes to 0 as $x \to \infty$, condition $\tilde{\mathbf{G}}\mathbf{1}$ holds for all $k = 0, 1, \ldots, n$.

Finally, to be able to conclude that Theorem 1 of Jönsson, Kukush, and Silvestrov (2005b) holds, we must check that

$$\psi_k'(x) \geqslant e^{-r\Delta t} E\left[\psi_{k+1}'(A(x, Y_{k+1}))\ A'(x, Y_{k+1})\right], \tag{2.39}$$

for all $k = 0, 1, \ldots, n-1$, where $\psi_k'(x)$ and $A'(x, Y_{k+1})$ are the right-derivative of the corresponding functions. First, note that $A'(x, Y_{k+1}) = Y_{k+1}$, and that for $k = 0, 1, \ldots, n-1$, $\psi_k'(x) = e^{-\kappa(T)-(c-\kappa)t_k}$. We also have $\psi_n'(x) = e^{-cT}\mathbb{1}_{\{x \geqslant Ge^{cT}\}}$. Thus, for $k = 0, 1, \ldots, n-2$, we have

$$\begin{aligned} e^{-r\Delta t} E\left[\psi_{k+1}'(A(x, Y_{k+1}))\ A'(x, Y_{k+1})\right] &= e^{-\kappa T-(c-\kappa)t_{k+1}} \\ &= e^{-\kappa T-(c-\kappa)t_k-(c-\kappa)\Delta t} \\ &\leqslant e^{-\kappa T-(c-\kappa)t_k} = \psi_k'(x). \end{aligned} \tag{2.40}$$

To show that (2.39) holds for $k = n-1$, observe that

$$\begin{aligned} e^{-r\Delta t} E\left[\psi_n'(A(x, Y_{k+1}))\ A'(x, Y_{k+1})\right] &\leqslant e^{-r\Delta t} E\left[e^{-cT} Y_n\right] \\ &= e^{-cT} \\ &\leqslant e^{-cT+(c-\kappa)\Delta t} = \psi_{n-1}'(x). \end{aligned} \tag{2.41}$$

Note that these conditions hold only when $\kappa < c$.

Theorem 1 from Jönsson, Kukush, and Silvestrov (2005b) states that when conditions $\tilde{\mathbf{A}}\mathbf{2}$, $\tilde{\mathbf{G}}\mathbf{1}$, $\tilde{\mathbf{B}}\mathbf{1}$, $\tilde{\mathbf{E}}\mathbf{1}$ and $\tilde{\mathbf{E}}\mathbf{2}$ hold, then for any time t_k, $k = 0, 1, \ldots, n$, the optimal stopping region at time t_k has the form $[a_k, \infty)$, with $0 \leqslant d_k < \infty$.

Having checked all the necessary conditions for Theorem 1 from Jönsson, Kukush, and Silvestrov (2005b) to hold, we can use it to confirm that for any time t_k, $k = 0, 1, ..., n$, the optimal surrender strategy at time t_k is a threshold strategy. The idea behind the proof of this theorem is to show that the optimal surrender strategy at time $n-1$ is a threshold strategy. This is done using conditions $\tilde{\mathbf{B}}\mathbf{1}$, $\tilde{\mathbf{E}}\mathbf{1}$ and $\tilde{\mathbf{G}}\mathbf{1}$. These conditions are then used to show recursively that a threshold strategy is also optimal at time $n-2, n-3, \ldots, 1, 0$.

We must then verify that this result also is also true in continuous time. While Cox, Ross, and Rubinstein (1979) shows that the convergence holds for European-type contingent claims, the convergence is harder to prove for options that can be surrendered at any time before maturity, because they involve an optimal control problem. Amin and Khanna (1994) show that under certain regularity conditions in the financial model, if the payoff function of the American option is uniformly integrable[7], convergence of the price and the optimal strategy holds. For a binomial model, Amin and Khanna (1994) show that $\sup_k E\left[\sup_{k<j\leqslant n} S_{t_j}^{\gamma} | S_{t_k}\right] < \infty$ for every $\gamma > 0$, so convergence holds for any payoff that is bounded above by a polynomial function of the stock price. Since our payoff function at most the sum of the account value and an European put on the account value (which has a bounded payoff), it is uniformly integrable in the binomial model, and thus convergence holds.

2.B Last step of the proof of Proposition 2.3.3

In this last step, we suppose that

$$\begin{aligned} &\upsilon(t_{n-m}, F_{n-m}; B_{n-m+1}) = U(t_{n-m}, F_{n-m}) \\ &\quad + \sum_{k=1}^{m-1} e^{-(m-k)r\Delta t} E[(F_{n-k} - E[e^{-r\Delta t} F_{n-k+1} | \mathcal{F}_{n-k}]) \mathbb{1}_{\{F_{n-k} \geqslant B_{n-k}\}} | \mathcal{F}_{n-m}] - \mathcal{O}(\Delta t). \end{aligned} \tag{2.42}$$

We move back one period and calculate $\upsilon(t_{n-m-1}, F_{n-m-1}; B_{n-m})$ as the risk-neutral expectations of the value of the live contract at t_{n-m}:

$$\begin{aligned} &\upsilon(t_{n-m-1} F_{n-m-1}; B_{n-m}) = \\ &\quad E[e^{-r\Delta t} \upsilon(t_{n-m}, F_{n-m}; B_{n-m+1}) \mathbb{1}_{\{F_{n-m} < B_{n-m}\}} | \mathcal{F}_{n-m+1}] \\ &\qquad + E[e^{-r\Delta t} F_{n-m} \mathbb{1}_{\{\{F_{n-m} \geqslant B_{n-m}\}\}} | \mathcal{F}_{n-m+1}] \\ &\quad = U(t_{n-m-1}, F_{n-m-1}) \\ &\qquad + \sum_{k=1}^{m-1} e^{-(m-k+1)r\Delta t} E[(F_{n-k} - E[e^{-r\Delta t} F_{n-k+1} | \mathcal{F}_{n-k}]) \mathbb{1}_{\{F_{n-k} \geqslant B_{n-k}\}} | \mathcal{F}_{n-m+1}] \\ &\qquad + E[e^{-r\Delta t} (F_{n-m} - U(t_{n-m}, F_{n-m})) \mathbb{1}_{\{F_{n-m} \geqslant B_{n-m}\}} | \mathcal{F}_{n-m-1}] - \mathcal{O}(\Delta t) \\ &\qquad - E\Big[\, e^{-r\Delta t} \mathbb{1}_{\{F_{n-m} \geqslant B_{n-m}\}} \\ &\qquad \sum_{k=1}^{m-1} E[e^{-(m-k)r\Delta t} (F_{n-k} - E[e^{-r\Delta t} F_{n-k+1} | \mathcal{F}_{n-k}]) \mathbb{1}_{\{F_{n-k} \geqslant B_{n-k}\}} | \mathcal{F}_{n-m}] | \mathcal{F}_{n-m+1} \,\Big]. \end{aligned} \tag{2.43}$$

[7]The payoff function $\psi_k(x)$ is uniformly integrable if for some $\delta > 1$, $\sup_k E\left[\sup_{k<j\leqslant n} \|\psi_j(S_{t_j})\|^{\delta} | S_{t_k}\right] < \infty$.

Note that the last term can be written as

$$
\begin{aligned}
&- E\left[e^{-r\Delta t}\left\{E[e^{-r\Delta t}F_{n-m+1}|\mathcal{F}_{n-m}] - e^{-r\Delta t}E[F_{n-m+1}\mathbb{1}_{\{F_{n-m+1}<B_{n-m+1}\}}|\mathcal{F}_{n-m}]\right.\right.\\
&\quad - E[E^{-r\Delta t}E[e^{-r\Delta t}F_{n-m+2}|\mathcal{F}_{n-m+1}]\mathbb{1}_{\{F_{n-m+1}\geqslant B_{n-m+1}\}}|\mathcal{F}_{n-m}]\\
&\quad \left.\left. + \sum_{k=1}^{m-2} e^{-(m-k)r\Delta t}E[(F_{n-k} - E[e^{r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m}]\ \right\}|\mathcal{F}_{n-m+1}\right]. \qquad (2.44)
\end{aligned}
$$

Using (2.44) in (2.43) and re-arranging, we obtain

$$
\begin{aligned}
&\upsilon(t_{n-m-1}, F_{n-m-1}; B_{n-m})\\
&\quad = U(t_{n-m-1}F_{n-m-1})\\
&\quad + \sum_{k=1}^{m} e^{((m+1)-k)r\Delta t}E[(F_{n-k} - E[e^{-r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m-1}] - \mathcal{O}(\Delta t)\\
&\quad - E\left[e^{-r\Delta t}\mathbb{1}_{\{F_{n-m}\geqslant B_{n-m}\}}\left\{U(t_{n-m}, F_{n-m}) - e^{-r\Delta t}E[F_{n-m+1}|\mathcal{F}_{n-m}]\right.\right.\\
&\quad + e^{-r\Delta t}E[(F_{n-m+1} - E[e^{-r\Delta t}F_{n-m+2}|\mathcal{F}_{n-m+1}])\mathbb{1}_{\{F_{n-m+1}\geqslant B_{n-m+1}\}}|\mathcal{F}_{n-m}]\\
&\quad \left.\left.\sum_{k=1}^{m-2} e^{-(m-k)r\Delta t}E[(F_{n-k} - E[e^{-r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m}]\right\}|\mathcal{F}_{n-m-1}\right]. \qquad (2.45)
\end{aligned}
$$

We will now show that the last expectation in (2.45), denoted by L_{m+1} is of order $\mathcal{O}(\Delta t)$ or higher. First, we know from (2.18) that for $F_{n-m} > B_{n-m}$,

$$
\begin{aligned}
F_{n-m} &> U(t_{n-m}, F_{n-m})\\
&+ \sum_{k=1}^{m-1} e^{-(m-k)r\Delta t}E[(F_{n-k} - E[e^{-r\Delta t}F_{n-k+1}|\mathcal{F}_{n-k}])\mathbb{1}_{\{F_{n-k}\geqslant B_{n-k}\}}|\mathcal{F}_{n-m}] - \mathcal{O}(\Delta t),
\end{aligned}
$$

which gives us the following upper bound for L_{n-m-1}:

$$L_{n-m-1} < E[e^{-r\Delta t}\mathbb{1}_{\{F_{n-m}\geqslant B_{n-m}\}}\left(F_{n-m} - E[e^{-r\Delta t}F_{n-m+1}|\mathcal{F}_{n-m}]\right)|\mathcal{F}_{n-m-1}].$$

Before we go further, we need to examine $E[e^{-r\Delta t}F_t|\mathcal{F}_s]$, $0 \leqslant s < t \leqslant T$, more closely. Observe that

$$
\begin{aligned}
F_t &= F_s \exp\left\{\left(r - \frac{\sigma^2}{2}\right) - \int_0^{t-s} c(F_{s+u})du + \sigma(W_t - W_s)\right\}\\
&\geqslant F_s \exp\left\{\left(r - \bar{c} - \frac{\sigma^2}{2}\right) + \sigma(W_t - W_s)\right\}, \qquad (2.46)
\end{aligned}
$$

since we assumed that $c(t, F_t) \leqslant \bar{c}$ for $0 \leqslant t \leqslant T$. Then we have a new upper bound for L_{n-m-1}:

$$L_{n-m-1} < E[e^{-r\Delta t} \mathbb{1}_{\{F_{n-m} \geqslant B_{n-m}\}} F_{n-m} \left(1 - e^{-c\Delta t}\right) |\mathcal{F}_{n-m-1}].$$

This allows us to evaluate the expectation and obtain

$$\begin{aligned} & E[e^{-r\Delta t} \mathbb{1}_{\{F_{n-m} \geqslant B_{n-m}\}} F_{n-m} \left(1 - e^{-c\Delta t}\right) |\mathcal{F}_{n-m-1}] = \\ & (1 - e^{-c\Delta t}) F_{n-m-1} N(d_1(F_{n-m-1}, \Delta t; B_{n-m})), \end{aligned} \tag{2.47}$$

where $d_1(F_{n-m-1}, \Delta t; B_{n-m}) = \frac{\ln\left(\frac{F_{n-m-1}}{B_{n-m}}\right) + \left(r - c - \frac{\sigma^2}{2}\right)\Delta t}{\sigma\sqrt{\Delta t}}$ and $\mathcal{N}(\cdot)$ is the cumulative distribution of a standard normal random variable. To see that (2.47) is of order $\mathcal{O}(\Delta t)$ or higher, we observe that $(1 - e^{-c\Delta t})$ is of order $\mathcal{O}(\Delta t)$ and we show that

$$\lim_{\Delta t \to 0} \mathcal{N}(d_1(F_{n-m-1}, \Delta t; B_{n-m})) = 0.$$

Here, F_{n-m-1} is strictly less than B_{n-m-1} since the option is still "alive" one period later (which means that $F_{n-m-1} < B_{n-m-1}$). Then, for any $\varepsilon > 0$, there exists $\Delta t^\star$ small enough such that for any $\Delta t < \Delta t^\star$, $|B_{n-m-1} - B_{n-m}| < \varepsilon$. Now let $\varepsilon' = B_{n-m} - F_{n-m-1}$. Since $F_{n-m-1} < B_{n-m-1}$, $\varepsilon' > 0$. By the continuity of the optimal exercise boundary $B(\cdot)$, there exists Δt small enough so that $B_{n-m} > B_{n-m-1} - \varepsilon' = F_{n-m-1}$. Hence, as $\Delta t \to 0$, $\ln\left(\frac{F_{n-m-1}}{B_{n-m}}\right)$ becomes negative and $d_1 \to -\infty$. Thus, $\lim_{\Delta t \to 0} \mathcal{N}(d_1(F_{n-m-1}, \Delta t; B_{n-m})) = 0$. Then, we can say that L_{n-m-1} is of order $\mathcal{O}(\Delta t)$ or higher. □

2.C Optimal Surrender Region with Asian Benefits

We prove here that the optimal surrender strategy for the path-dependent payoff introduced in Section 2.4 is also a threshold strategy. That is, we show that when the surrender charge is of the form $\kappa_t = 1 - e^{-\kappa(T-t)}$, $\kappa < c$, and satisfies the conditions stated at the beginning of Section 2.4, then for any time t before maturity and any value F_t, there exists a geometric average Y_t^* above which the value of the contract is less than the surrender benefit available immediately. Here, we will not refer to the results used in Appendix 2.A because we are now facing a path-dependent problem. Instead, this proof is inspired by Section 3 of Wu and Fu (2003). We let τ be a stopping time with respect to $\mathcal{F}_t$ and denote by $\mathcal{T}_t$ the set of all stopping times τ greater than t and bounded by T. We express the value at time t of the variable annuity contract $V^*(t, x)$ by

$$V^{*g}(t, f, x) = \sup_{\tau \in \mathcal{T}_t} E\left[e^{-r(\tau - t)} \psi(\tau, Y_\tau) | F_t = f, Y_t = x\right],$$

and

$$\psi(t,x) = \begin{cases} e^{-\kappa(T-t)}x, & \text{if } 0 \leqslant t < T \\ \max(x,G), & \text{if } t = T. \end{cases} \tag{2.48}$$

We also define the optimal surrender region at time t, denoted $\mathcal{R}_t(F_t)$, by

$$\mathcal{R}_t(F_t) = \left\{ Y_t : \sup_{\tau \in \mathcal{T}_t} E\left[e^{-r(\tau-t)}\psi(\tau, Y_\tau) | \mathcal{F}_t \right] \leqslant \psi(t, Y_t) \right\}. \tag{2.49}$$

We can also rewrite $\mathcal{R}_t(F_t)$ as

$$\mathcal{R}_t(F_t) = \left\{ Y_t : \frac{V^{*g}(t, F_t, Y_t)}{\psi(t, Y_t)} \leqslant 1 \right\}.$$

We analyse the function $\gamma^g(t, F_t, x) \equiv \frac{V^{*g}(t,F_t,x)}{\psi(t,x)}$ and obtain Lemma 4.B.2.

Lemma 2.C.1. *Let* $\gamma^g(t,f,x) = \frac{V^{*g}(t,f,x)}{\psi(t,x)}$ *for* $t \in [0,T]$. *Then,*

- *For* $t = T$, $\gamma^g(T, f, x) = 1$.
- *For* $t \in [\,0, T\,)$, $\gamma(t, f, x)$ *is non-increasing in* x.

Proof. At $t = T$, we have that $\frac{V^{*g}(T,F_T,Y_T)}{\psi(T,Y_T)} = \frac{\psi(T,Y_T)}{\psi(T,Y_T)} = 1$.

For what follows, we use the fact that $Y_u | F_t, Y_t$ has the same distribution as $Y_t^{\frac{t}{u}} F_t^{\frac{u-t}{u}} e^{\mu(t,u)+\sigma(t,u)Z}$, where Z is a standard normal random variable, $\mu(t,u) = \frac{r-c-\frac{\sigma^2}{2}}{2u}(u-t)^2$ and $\sigma^2(t,u) = \frac{\sigma^2}{3u^2}(u-t)^3$. For $t \in [\,0, T\,)$, note that $\gamma^g(t, f, x)$ can be rewritten as

$$\begin{aligned} \gamma^g(t,f,x) = \frac{V^{*g}(t,f,x)}{\psi(t,x)} &= \frac{\sup_{\tau \in \mathcal{T}_t} E\left[e^{-r(\tau-t)}\psi(\tau, Y_\tau) | F_t = f, Y_t = x \right]}{e^{-\kappa(T-t)}x} \\ &= \frac{\sup_{\tau \in \mathcal{T}_t} E\left[e^{-r(\tau-t)} e^{-\kappa(T-\tau)} Y_\tau + e^{-r(T-t)}(G - Y_T)^+ \mathbb{1}_{\{\tau=T\}} | F_t = f, Y_t = x \right]}{x e^{-\kappa(T-t)}} \end{aligned}$$

where $X_{t,u} = \mu(t,u) + \sigma(t,u)Z$. Note that because Brownian motion increments are independent, $\gamma^g(t, f, x)$ only depends on $\{F\}_{0 \leqslant s < t}$ through Y_t, the value of the average at time t. In addition, since x is a positive real number and $E\left[e^{-r(\tau-t)}\psi(\tau, Y_\tau) | F_t = f, Y_t = x \right]$ is finite, we can take $\frac{1}{xe^{-\kappa(T-t)}}$ inside the supremum.

Now for any $\tau \in \mathcal{T}_t$ and $\varepsilon > 0$, we have

$$\begin{aligned}
&E\left[e^{-(r-\kappa)(\tau-t)}\left(\frac{Y_t}{F_t}\right)^{\frac{t}{\tau}-1} e^{X_{t,\tau}}|Y_t = x, F_t = f\right] \\
&= E\left[e^{-(r-\kappa)(\tau-t)}\left(\frac{x}{f}\right)^{\frac{t}{\tau}-1} e^{X_{t,\tau}}|F_t = f\right] \\
&\geqslant E\left[e^{-(r-\kappa)(\tau-t)}\left(\frac{x+\varepsilon}{f}\right)^{\frac{t}{\tau}-1} e^{X_{t,\tau}}|F_t = f\right] \\
&= E\left[e^{-(r-\kappa)(\tau-t)}\left(\frac{Y_t}{F_t}\right)^{\frac{t}{\tau}-1} e^{X_{t,\tau}}|Y_t = x+\varepsilon, F_t = f\right],
\end{aligned} \tag{2.50}$$

since $\frac{t}{\tau} - 1 \leqslant 0$, and

$$\begin{aligned}
&E\left[e^{-r(T-t)}\left(\frac{G}{Y_t e^{-\kappa(T-t)}} - \left(\frac{Y_t}{F_t}\right)^{\frac{t}{\tau}-1} e^{X_{t,T}}\right)^+ \mathbb{1}_{\{\tau=T\}}|Y_t = x, F_t = f\right] \\
&= E\left[e^{-r(T-t)}\left(\frac{G}{x e^{-\kappa(T-t)}} - \left(\frac{x}{f}\right)^{\frac{t}{\tau}-1} e^{X_{t,T}}\right)^+ \mathbb{1}_{\{\tau=T\}}|F_t = f\right] \\
&\geqslant E\left[e^{-r(T-t)}\left(\frac{G}{(x+\varepsilon) e^{-\kappa(T-t)}} - \left(\frac{x}{f}\right)^{\frac{t}{\tau}-1} e^{X_{t,T}}\right)^+ \mathbb{1}_{\{\tau=T\}}|F_t = f\right] \\
&\geqslant E\left[e^{-r(T-t)}\left(\frac{G}{(x+\varepsilon) e^{-\kappa(T-t)}} - \left(\frac{x+\varepsilon}{f}\right)^{\frac{t}{\tau}-1} e^{X_{t,T}}\right)^+ \mathbb{1}_{\{\tau=T\}}|F_t = f\right] \\
&= E\left[e^{-r(T-t)}\left(\frac{G}{Y_t e^{-\kappa(T-t)}} - \left(\frac{Y_t}{F_t}\right)^{\frac{t}{\tau}-1} e^{X_{t,T}}\right)^+ \mathbb{1}_{\{\tau=T\}}|Y_t = x+\varepsilon, F_t = f\right].
\end{aligned} \tag{2.51}$$

To obtain the third and fourth line, note that for $a > a' \geqslant 0$ and $0 \leqslant b < b'$, $(a-b)^+ \geqslant (a'-b)^+$ and $(a-b)^+ \geqslant (a-b')^+$. Using (2.50) and (2.51) together, we obtain

$$\frac{E\left[e^{-r(\tau-t)}\psi(\tau, Y_\tau)|F_t = f, Y_t = x\right]}{\psi(t,x)} \geqslant \frac{E\left[e^{-r(\tau-t)}\psi(\tau, Y_\tau)|F_t = f, Y_t = x+\varepsilon\right]}{\psi(t,x+\varepsilon)} \tag{2.52}$$

for any $\tau \in \mathcal{T}_t$ and $\varepsilon > 0$. Taking the supremum over all $\tau \in \mathcal{T}_t$ on both sides, we obtain the desired result. $\square$

The result presented in Lemma 4.B.2 allows us to say that if we can find Y_t^* such that $\gamma^g(t, F_t, Y_t^*) = 1$, then for any $Y_t \geqslant Y_t^*$, $\gamma(t, F_t, Y_t) \leqslant 1$. Thus, the optimal surrender region $\mathcal{R}_t(F_t)$ has the form $[\ Y_t^*, \infty\ )$.

To complete the proof that the optimal surrender strategy is of the threshold type, we need to show that for any t, F_t, $0 \leqslant t < T$, $F_t > 0$, there exists a value Y_t^* such that $V^{*g}(t, F_t, Y_t) \leqslant \psi(t, Y_t)$. This is shown in Theorem 2.C.1.

Theorem 2.C.1. *The optimal exercise strategy for the path-dependent surrender option is to surrender the contract when* $Y_t \geqslant B_t(f)$, *with*

$$B_t(f) = \inf\{x : V^{*g}(t, f, x) \leqslant \psi(t, x)\},$$

for $t \in [\ 0, T\ )$, $f > 0$. $B_t(f) < \infty$ *for all* $t \in [0, T], f > 0$ *if the surrender charges are of the form* $\kappa_t = 1 - \exp(-\kappa(T - t))$ *and satisfy* $\kappa < \frac{r + c + \frac{\sigma^2}{6}}{2}$, *and* $c < r - \frac{\sigma^2}{6}$.

Proof. We show that for any $t \in [\ 0, T\ )$, $f > 0$ it is possible to find x such that $V^{*g}(t, f, x) \leqslant \psi(t, x)$. Note that for $t \in [\ 0, T\ )$, $\psi(t, x) = xe^{-\kappa(T-t)}$. Thus, we need to show that it is possible to find x such that

$$V^{*g}(t, f, x) \leqslant xe^{-\kappa(T-t)}.$$

First, fix $t \in [\ 0, T\ )$ and observe that for any stopping time $\tau \in \mathcal{T}_t$, we have

$$\begin{aligned}
&E[e^{-r(\tau-t)}\psi(\tau, Y_\tau)|Y_t = x, F_t = f] \\
&\quad = E[e^{-r(\tau-t)}Y_\tau e^{-\kappa(T-\tau)}|Y_t = x, F_t = f] + E[e^{-r(T-t)}(G - Y_T)^+ \mathbb{1}_{\{\tau=T\}}|Y_t = x, F_t = f] \\
&\quad \leqslant E[e^{-r(\tau-t)}Y_\tau e^{-\kappa(T-\tau)}|Y_t = x, F_t = f] + E[e^{-r(T-t)}(G - Y_T)^+|Y_t = x, F_t = f]
\end{aligned}$$

The second term of the equation is simply the price of a geometric Asian put option with strike G. This term goes to 0 as $x \to \infty$ (see for example Kemna and Vorst (1990)). Now by the same reasoning as in the proof of Theorem 6.1, it suffices to show that there exists x^* such that

$$E[e^{-r(\tau-t)}Y_\tau e^{-\kappa(T-\tau)}|Y_t = x^*, F_t = f] < x^* e^{-\kappa(T-t)}.$$

Then, by Lemma 4.B.2, the inequality will hold for any $x > x^*$. For a fixed $t \in [0, T)$, $f \in (0, \infty)$, this can be done by taking any $x > f$. Let $f < x^* < \infty$. Then, by first conditioning on the stopping time τ, we have

$$
\begin{aligned}
&E[e^{-r(\tau-t)}Y_\tau e^{-\kappa(T-\tau)}|Y_t = x^*, F_t = f] \\
&= E[e^{-\kappa(T-\tau)}e^{-r(\tau-t)+\frac{t}{\tau}\ln x^* + \frac{\tau-t}{\tau}\ln f + \frac{r-c-0.5\sigma^2}{2\tau}(\tau-t)^2 + \frac{\sigma^2(\tau-t)^3}{6\tau^2}}] \\
&< E[x^* e^{-\kappa(T-\tau)}e^{-r(\tau-t)+\frac{r-c-0.5\sigma^2}{2\tau}(\tau-t)^2 + \frac{\sigma^2(\tau-t)^3}{6\tau^2}}] \\
&< E[x^* e^{-\kappa(T-\tau)}e^{-r(\tau-t)+\frac{r-c-\sigma^2/6}{2\tau}(\tau-t)^2}] \\
&< E[x^* e^{-\kappa(T-\tau)}e^{-\frac{1}{2}(r+c+\frac{\sigma^2}{6})(\tau-t)}] \\
&< E[x^* e^{-\kappa(T-\tau)}e^{\kappa(\tau-t)}] \\
&= x^* e^{-\kappa(T-t)}
\end{aligned}
$$

To get the fourth and the fifth line, we use the assumption $c < r - \frac{\sigma^2}{6}$ and the fact that $\tau > \tau - t$. By taking the supremum over all stopping times, this allows us to conclude that under our assumptions, it is always possible to find an average fund value x such that

$$V^{*g}(t, f, x) \leqslant xe^{-\kappa(T-t)}.$$

□

Chapter 3

State-dependent fees for variable annuity guarantees

3.1 Introduction

This chapter is based on a paper that was written in collaboration with Dr. Carole Bernard and Dr. Mary Hardy, and that is forthcoming in *ASTIN Bulletin* (see Bernard, Hardy, and MacKay (2014)).

Variable annuity guarantees are typically funded by a fixed fee rate. This fee structure is unsatisfactory from a risk management perspective because there is a misalignment of the fee income and the option cost. When markets fall, the option value is high, but the fee income is reduced. When markets rise, the fee income increases, but there is negligible guarantee liability. As explained in the previous chapter, a fixed fee rate represents an incentive for the policyholder to surrender his contract when the account value is well above the guarantee.

In this chapter we investigate a dynamic fee structure for GMMBs and GMDBs, where the fee rate depends in some way on the evolving embedded option value. Specifically, we develop pricing formulas for a contract for which the fee (applied as a proportion of the policyholder's fund) is only payable when the fund value is below the guaranteed benefit – that is, when the embedded option is in-the-money. A similar formula was derived for equity options in Karatzas and Shreve (1984). However, our formula is more general, and we present details that will make it easier to use for the reader. We also extend the results to allow for a threshold for fee payments that is higher than the guaranteed amount, so that the fee begins to be paid when the fund moves close to being in-the-money. For convenience, we refer to this fee structure as 'state-dependent', where the states being considered are simply (1) the policyholder's fund F_t lies above a given barrier β, and (2) F_t lies below the barrier, β. The motivation for exploring this fee structure was a VA-type contract issued by Prudential UK, whose "flexible investment plan" is an equity-linked policy with an optional GMDB, for which the fee is paid only if the fund value is below the guaranteed level. See Prudential(UK) (2012), page 11, for details. Thus, this is a practical design proposition, at least for the GMDB. The advantages for an optional GMDB rider are clear; once the guarantee is no longer valuable, the policyholder pays the same fees as those who did not select the GMDB rider. Additionally, the alignment of income and liability value is better managed. One major disadvantage however is that by paying for the option only when the option is in the money, the fee rates at those times will be high, especially for GMMBs, to the extent that the contract will no longer appear competitive with other investment alternatives, and which could also exacerbate the option liability through fee drag. We discuss this further in Section 3.4.

Although the policyholder's option to surrender is impacted by the fee structure, in this chapter we derive and evaluate fair fees assuming, initially, no policy surrenders. It would be straightforward to adapt the formulae to allow for deterministic surrenders. In Section 3.5 we consider further the impact of the fee structure on the surrender incentive. Many recent works on variable annuities have incorporated optimal surrenders in the pricing analysis (see, for example, Bacinello, Millossovich, Olivieri, and Pitacco (2011) and Belanger, Forsyth, and Labahn (2009)). The complexity added by the state-dependent fee makes this more complex in this case than in the flat fee context, and a full analysis will be performed in Chapter 4. In Section 3.5 we present a preliminary analysis of the surrender incentive, to support the intuition that the state-dependent fee structure reduces the incentive to surrender for policies that are substantially out-of-the-money.

The rest of this chapter is organized as follows. In Section 3.2 we present the model (which follows the standard Black-Scholes assumptions) and the pricing of a general maturity benefit given that the fee is only paid when the fund value is below some critical level. In Section 3.3 we apply the results to the simple guaranteed minimum maturity and death benefits embedded in a VA contract. Some numerical illustrations can be found in Section 3.4. A first analysis of the surrender incentive is presented in Section 3.5. Section 3.6 studies the robustness of the results to the Black-Scholes model assumption. Section 3.7 concludes.

3.2 Pricing with state-dependent fee rates

3.2.1 Notation

Consider a simple variable annuity contract and denote by F_t the underlying fund value at time t. Assume that there is a single premium P paid at time 0 by the policyholder. Further assume that the premium is fully invested in an equity index S_t which follows a geometric Brownian motion under the risk-neutral measure $\mathbb{Q}$[1]:

$$dS_t = S_t \left(r \ dt + \sigma dW_t \right).$$

Here r denotes the continuously compounded risk-free rate and σ is the volatility. Furthermore suppose that management fees are paid out of the fund at a constant percentage c whenever the value of the fund, F_t, lies below a given level β. Let C_t be the total fees paid at time t; its dynamics are given by

$$dC_t = cF_t \mathbb{1}_{\{F_t < \beta\}} \ dt,$$

where $\mathbb{1}_A$ is the indicator function of the set A. The dynamics of the fund are thus given by

$$\frac{dF_t}{F_t} = \frac{dS_t}{S_t} - c\mathbb{1}_{\{F_t < \beta\}} \ dt, \quad 0 \leqslant t \leqslant T, \tag{3.1}$$

with $F_0 = P - e$, where P is the initial premium and e is the fixed expense charge deducted by the insurer at inception. Throughout this chapter, we assume $e = 0$. Using Ito's lemma to compute $d \ln(F_t)$, and integrating, we obtain the following representation of (3.1),

$$F_t = F_0 \exp \left\{ \left(r - \frac{\sigma^2}{2} \right) t - c \int_0^t \mathbb{1}_{\{F_s < \beta\}} ds + \sigma W_t \right\}. \tag{3.2}$$

3.2.2 Pricing the VA including guarantees with state-dependent fee rates

We first consider a simple maturity benefit guarantee, with term T-years, and with a payout that depends only on F_T. Let $g(T, F_T)$ be the total payoff of the VA – that is, the policyholder's fund plus any additional payments arising from the embedded option. For example $g(T, F_T) = \max(G, F_T)$ where G is the guaranteed payout at time T. Our first objective in this chapter is to compute the following value

$$U_{0,F_0} := E_{\mathbb{Q}}[e^{-rT} g(T, F_T)],$$

which corresponds to the market value (or no-arbitrage price) of the VA at time 0 for its payoff $g(T, F_T)$ paid at time T. Note that in this chapter, we slightly modify the notation for the price of the maturity benefit to highlight its dependence on other factors such as the fair fee and the level under which this fee is paid. As we discuss above, for simplicity, we ignore lapses and surrenders, and value the contract as a definite term investment.

As the fee is only paid when the fund value F_t is below the level β, we introduce the occupation

[1]This chapter is only concerned with the risk-neutral pricing of the maturity benefit. For this reason, we only work under the risk-neutral measure $\mathbb{Q}$.

time

$$\Gamma^{-}_{t,F}(\beta) = \int_0^t \mathbb{1}_{\{F_s<\beta\}} ds, \tag{3.3}$$

which corresponds to the total time in the interval $(0, t)$ for which the fund value lies below the fee barrier β.

Proposition 3.2.1 (No-Arbitrage Price)**.** *Let* $\theta_t := -a + \frac{c}{\sigma}\mathbb{1}_{\{F_t<\beta\}}$ *where* $a = \frac{1}{\sigma}\left(r - \frac{\sigma^2}{2}\right)$ *and let* $K = \frac{1}{\sigma}\ln\left(\frac{\beta}{F_0}\right)$. *Define a probability* $\widetilde{\mathbb{Q}}$ *by its Radon-Nikodym derivative with respect to the risk-neutral probability*[2] $\mathbb{Q}$

$$\left(\frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}\right)_T = \exp\left\{\int_0^T \theta_s dW_s - \frac{1}{2}\int_0^T \theta_s^2 ds\right\}. \tag{3.4}$$

Then,

$$U_{0,F_0} = e^{-rT-\frac{a^2}{2}T} E_{\widetilde{\mathbb{Q}}}\left[\exp\left\{a\widetilde{W}_T + b\Gamma^{-}_{T,\widetilde{W}}(K) - \frac{c}{\sigma}\int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}} d\widetilde{W}_s\right\} g\left(T, F_0 e^{\sigma\widetilde{W}_T}\right)\right], \tag{3.5}$$

where $b = \frac{c}{\sigma^2}\left(r - \frac{\sigma^2}{2} - \frac{c}{2}\right)$ *and* $\widetilde{W}_t = W_t - \int_0^t \theta_s ds$ *is a* $\widetilde{\mathbb{Q}}-$*Brownian motion, and* $\Gamma^{-}_{T,\widetilde{W}}(K)$ *is defined similarly as in* (3.3) *and represents the occupation time of* $\widetilde{W}$ *below* K.

Proof. This proof is based on Section 5 of Karatzas and Shreve (1984). Recall that W_t is a standard Brownian motion under the risk-neutral probability measure $\mathbb{Q}$. By Girsanov's Theorem it is clear that $\widetilde{W}_t = W_t - \int_0^t \theta_s ds = W_t + at - \frac{c}{\sigma}\int_0^t \mathbb{1}_{\{F_s<\beta\}} ds$ is a standard Brownian motion under the probability measure $\widetilde{\mathbb{Q}}$ defined in (3.4). Furthermore, observe that $F_t = F_0 e^{\sigma\widetilde{W}_t}$, for $0 \leqslant t \leqslant T$, so that the occupation time of the fund below the level β can be rewritten using the Brownian motion $\widetilde{W}$ under the new probability measure $\widetilde{\mathbb{Q}}$

$$\{F_t < \beta\} = \left\{\widetilde{W}_t < K\right\}, \tag{3.6}$$

where $K := \frac{1}{\sigma}\ln\left(\frac{\beta}{F_0}\right)$. In other words, the occupation time $\Gamma^{-}_{t,F}(\beta)$ (given in (3.3)), is also $\Gamma^{-}_{t,\widetilde{W}}(K) = \int_0^t \mathbb{1}_{\{\widetilde{W}_s<K\}} ds$ (using the equivalence in (3.6)). Note that we can express the Radon-

[2] This means that $\widetilde{\mathbb{Q}}(A) = E\left[\frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}\mathbb{1}_A\right]$ for any measurable set A.

Nikodym derivative $\left(\frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}\right)_T$ as

$$\begin{aligned}\left(\frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}\right)_T &= \exp\left\{\int_0^T \theta_s(d\widetilde{W}_s + \theta_s ds) - \frac{1}{2}\int_0^T \theta_s^2 ds\right\} \\ &= \exp\left\{\int_0^T \theta_s d\widetilde{W}_s + \frac{1}{2}\int_0^T \theta_s^2 ds\right\} \\ &= \exp\left\{-a\widetilde{W}_T + \frac{c}{\sigma}\int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}} d\widetilde{W}_s + \frac{a^2}{2}T - b\Gamma^-_{T,\widetilde{W}}(K)\right\} \end{aligned} \tag{3.7}$$

where

$$b = \frac{c}{\sigma^2}\left(r - \frac{\sigma^2}{2} - \frac{c}{2}\right).$$

To get the last equality, (3.7), we used the fact that

$$\theta_s^2 = a^2 - \frac{2c}{\sigma^2}\left(r - \frac{\sigma^2}{2} - \frac{c}{2}\right)\mathbb{1}_{\{\widetilde{W}_s<K\}}.$$

This makes it possible to price a claim $g(T, F_T)$ on the fund at maturity, under the measure $\widetilde{\mathbb{Q}}$:

$$\begin{aligned} U_{0,F_0} = E_{\mathbb{Q}}\left[e^{-rT}g(T,F_T)\right] &= E_{\widetilde{\mathbb{Q}}}\left[\left(\frac{d\mathbb{Q}}{d\widetilde{\mathbb{Q}}}\right)_T e^{-rT}g(T,F_T)\right] \\ &= E_{\widetilde{\mathbb{Q}}}\left[e^{-rT+a\widetilde{W}_T-\frac{c}{\sigma}\int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}}d\widetilde{W}_s-\frac{a^2}{2}T+b\Gamma^-_{T,\widetilde{W}}(K)} g\left(T, F_0 e^{\sigma\widetilde{W}_T}\right)\right], \end{aligned}$$

which ends the proof of Proposition 3.2.1. □

From Proposition 3.2.1, it is clear that we need the joint distribution of $(\widetilde{W}_t, \int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}}d\widetilde{W}_s, \Gamma^-_{t,\widetilde{W}}(K))$ under $\widetilde{\mathbb{Q}}$ to evaluate the initial premium U_{0,F_0}. Using the Tanaka formula for the Brownian local time, $L_{t,\widetilde{W}}(K)$ (see Proposition 6.8 of Karatzas and Shreve (1991)), we can write

$$\int_0^t \mathbb{1}_{\{\widetilde{W}_s<K\}}d\widetilde{W}_s = L_{t,\widetilde{W}}(K) - (\widetilde{W}_t - K)^- + (0-K)^-, \tag{3.8}$$

where we recall that $x^- = \max(-x, 0)$ so that $(0-K)^- = \max(K, 0)$. Thus, to price the Variable Annuity contract, we compute the expectation under $\widetilde{\mathbb{Q}}$ using the trivariate density[3] $\widetilde{\mathbb{Q}}\left(\widetilde{W}_T \in dx, L_{T,\widetilde{W}} \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right)$, which we recall in the following proposition.

[3]Note that the terminology "density" is used but it could be mixed, as there could be some probability mass on $\{L_{T,\widetilde{W}}(K) = 0, \Gamma^-_{T,\widetilde{W}}(K) = T\}$ or on $\{L_{T,\widetilde{W}}(K) = 0, \Gamma^-_{T,\widetilde{W}}(K) = 0\}$.

Proposition 3.2.2 (Distribution of $\left(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K)\right)$ under $\widetilde{\mathbb{Q}}$)**.** *Denote by* $f(x,y,z) = \widetilde{\mathbb{Q}}\left(\widetilde{W}_T \in dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right)$ *and by* $h(s,x) := \frac{|x|}{\sqrt{2\pi s^3}} \exp\left\{-\frac{x^2}{2s}\right\}$.

For $K \geqslant 0$*, the joint density of* $\left(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K)\right)$ *is given by*

$$f(x,y,z) = \begin{cases} 2h(z, y+K)\ h(T-z, y-K+x), & x > K,\ y > 0,\ 0 < z < T \\ 2h(T-z, y)\ h(z, y+2K-x), & x < K,\ y > 0,\ 0 < z < T \\ \frac{1}{\sqrt{2\pi T}}\left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}}\right), & x < K,\ y = 0,\ z = T. \end{cases}$$

For $K < 0$*, the joint density of* $\left(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K)\right)$ *is given by*

$$f(x,y,z) = \begin{cases} 2h(z, y)\ h(T-z, y+x-2K), & x > K, y > 0, 0 < z < T \\ 2h(T-z, y-K)\ h(z, y-x+K), & x < K, y > 0, 0 < z < T \\ \frac{1}{\sqrt{2\pi T}}\left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}}\right), & x > K, y = 0, z = 0. \end{cases}$$

Proof. Section 3 of Karatzas and Shreve (1984) gives the trivariate density of $(Z_T,\ L_{T,Z}(0),\ \Gamma^+_{T,Z}(0))$ for a Brownian motion Z with given initial value Z_0. To make use of this result, we express $L_{T,\widetilde{W}}(K)$ and $\Gamma^-_{T,\widetilde{W}}(K)$ in terms of $(Z_T, L_{T,Z}(0), \Gamma^+_{T,Z}(0))$ where Z is a Brownian motion such that $Z_0 = K$. A detailed proof of Proposition 3.2.2 can be found in Appendix 3.A. □

From the expression of the price in Proposition 3.2.1, and from the Tanaka formula in (3.8), we observe that the trivariate density given in Proposition 3.2.2 can be used to price any European VA. The following proposition summarizes this result.

Proposition 3.2.3 (No-arbitrage price)**.** *The no-arbitrage price at time* 0 *of a VA contract with payoff* $g(T, F_T)$ *can be computed as*

$$U_{0,F_0} = e^{-rT-\frac{a^2}{2}T} E_{\widetilde{\mathbb{Q}}}\left[e^{a\widetilde{W}_T + b\Gamma^-_{T,\widetilde{W}}(K)\ -\frac{cL_{T,\widetilde{W}}(K)}{\sigma} + \frac{c}{\sigma}\left((\widetilde{W}_T - K)^- - \max(K,0)\right)} g\left(T, F_0 e^{\sigma\widetilde{W}_T}\right)\right]$$

using the trivariate density of $\left(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K)\right)$ *given in Proposition 3.2.2.*

3.3 Examples

In this section we apply Proposition 3.2.3 to price a VA with level GMMB, and a VA with a level GMDB, with state-dependent fees payable when the policyholder's fund falls below some pre-specified level.

3.3.1 State-dependent fee rates for a VA with GMMB

The next proposition gives the price of a variable annuity with a guaranteed minimum maturity benefit (GMMB). It pays out the maximum between the value of the fund at maturity, F_T, and a guaranteed amount G at a given maturity date T. The payoff is then

$$g(T, F_T) := \max(G, F_T) = F_T \mathbb{1}_{\{G \leqslant F_T\}} + G \mathbb{1}_{\{F_T < G\}}. \tag{3.9}$$

Let $U^M_{0,F_0}(T, \beta, G, c)$ be the no-arbitrage price of this VA at inception assuming that a fee c is taken from the fund continuously, whenever the fund value drops below the level β.

Proposition 3.3.1 (No-arbitrage price for GMMB with guarantee G)**.** *The initial price $U^M_{0,F_0}(T, \beta, G, c)$ of the GMMB contract is given by*

$$U^M_{0,F_0}(T, \beta, G, c) = e^{-rT - \frac{a^2}{2}T - \frac{c}{\sigma}\max(0,K)}(F_0 A_T + G D_T), \tag{3.10}$$

where

$$\begin{cases} A_T = \int\limits_{H}^{+\infty} \int\limits_{0}^{+\infty} \int\limits_{0}^{T} e^{(\sigma+a)x + bz - \frac{c}{\sigma}y + \frac{c}{\sigma}(x-K)^-} f(x,y,z) dz\, dy\, dx \\ \\ D_T = \int\limits_{-\infty}^{H} \int\limits_{0}^{+\infty} \int\limits_{0}^{T} e^{ax + bz - \frac{c}{\sigma}y + \frac{c}{\sigma}(x-K)^-} f(x,y,z) dz\, dy\, dx \end{cases}$$

with $H := \frac{1}{\sigma}\ln\left(\frac{G}{F_0}\right)$. Recall that $K := \frac{1}{\sigma}\ln\left(\frac{\beta}{F_0}\right)$ where β is the level that triggers the payment of the continuous fee c.

Note that the maturity benefit is usually conditional on the survival of the policyholder. In Proposition 3.3.1, we ignore mortality risk. In order to add mortality risk, it would suffice to multiply the value of the benefit by the probability that the policyholder survives to maturity.

Proof. Using Proposition 3.2.3 and the expression of the payoff $g(T, F_T)$ given in (3.9), the result follows. □

We now replace the density f by its expression given in Proposition 3.2.2 and we can simplify the expressions for A_T and D_T. To do so we distinguish four cases. Details are given in Appendix 3.B.

The fair fee rate c^* is then computed such that, for a given premium P, the VA value is equal to the premium paid, that is

$$P = U^M_{0,F_0}(T, \beta, G, c^*). \tag{3.11}$$

We ignore here other types of expenses and assume there are no other guarantees attached. Note that the fair fee is the amount that needs to be paid out of the fund in order for the insurer to build the replicating portfolio.

Proposition 3.3.1 gives integral expressions for the no-arbitrage price of the GMMB. In the special case when the fee is paid only when the option is in-the-money, in other words when $\beta = G$, the price of the GMMB is given in the next proposition.

Proposition 3.3.2 (No-arbitrage price for GMMB with guarantee G when $\beta = G$)**.** *When the fee is only paid when the option is in-the-money, that is $\beta = G$, the initial price $U^M_{0,F_0}(T,\beta,G,c)$ of the GMMB contract is given by*

$$U^M_{0,F_0}(T,\beta,G,c) = e^{-rT-\frac{a^2}{2}T-\frac{c}{\sigma}\max(0,K)}(F_0C_1 + G(C_2 + C_3)),$$

where

$$C_3 = \left(\frac{\beta}{F_0}\right)^{\frac{c}{\sigma^2}} e^{\left(r-\frac{\sigma^2}{2}\right)^2 \frac{T}{2\sigma^2}} \left(N\left(\frac{\ln\left(\frac{\beta}{F_0}\right)-\eta T}{\sigma\sqrt{T}}\right) - \left(\frac{\beta}{F_0}\right)^{\frac{2\eta}{\sigma^2}} N\left(\frac{\ln\left(\frac{F_0}{\beta}\right)-\eta T}{\sigma\sqrt{T}}\right)\right)$$

with $\eta = r - \frac{\sigma^2}{2} - c$ and $N(\cdot)$ stands for the cdf of the standard normal distribution. For $K \geqslant 0$ (that is $G = \beta \geqslant F_0$),

$$\begin{cases} C_1 = \int\limits_K^\infty \int\limits_0^\infty \int\limits_0^T \frac{(y+K)(y-K+x)}{\pi\sqrt{z^3(T-z)^3}} e^{(a+\sigma)x+bz-\frac{cy}{\sigma}-\frac{(y+K)^2}{2z}-\frac{(y-K+x)^2}{2(T-z)}} dz\ dy\ dx \\ C_2 = \int\limits_{-\infty}^K \int\limits_0^\infty \int\limits_0^T \frac{y(y+2K-x)}{\pi\sqrt{z^3(T-z)^3}} e^{\left(a-\frac{c}{\sigma}\right)x+bz-\frac{cy}{\sigma}+\frac{cK}{\sigma}-\frac{y^2}{2(T-z)}-\frac{(y+2K-x)^2}{2z}} dz\ dy\ dx, \end{cases}$$

and for $K < 0$ (that is $G = \beta < F_0$)

$$\begin{cases} C_1 = \int\limits_K^\infty \int\limits_0^\infty \int\limits_0^T \frac{y|y+x-2K|}{\pi\sqrt{z^3(T-z)^3}} e^{(a+\sigma)x+bz-\frac{cy}{\sigma}-\frac{y^2}{2z}-\frac{(y-2K+x)^2}{2(T-z)}} dz\ dy\ dx \\ C_2 = \int\limits_{-\infty}^K \int\limits_0^\infty \int\limits_0^T \frac{|y-K|(y-x+K)}{\pi\sqrt{z^3(T-z)^3}} e^{\left(a-\frac{c}{\sigma}\right)x+bz-\frac{cy}{\sigma}+\frac{cK}{\sigma}-\frac{(y-K)^2}{2(T-z)}-\frac{(y+K-x)^2}{2z}} dz\ dy\ dx. \end{cases}$$

Details can be found in Appendix 3.B in the case when $H = K$.

3.3.2 State-dependent fee rates for a VA with GMDB

In this section, we price a guaranteed minimum death benefit rider, which guarantees a minimum amount if the policyholder dies before maturity of the contract. Let us first introduce some notation. Let τ_x be the random variable representing the future lifetime of a policyholder aged x and let $\llcorner\tau\lrcorner_x$ denote his curtate future lifetime[4]. Let also ${}_tp_x = P(\tau_x > t)$ be the probability that a policyholder aged x survives t years and $q_{x+t} = P(t < \tau_x \leqslant t+1 | \tau_x > t)$ be the probability

[4]Here, $\llcorner\cdot\lrcorner$ denotes the floor function.

that a policyholder aged $x+t$ dies during year t. We assume that the GMDB is paid at the end of the year of death, only if death occurs strictly before the maturity T of the contract. The fair price at time 0 of the GMDB is given by

$$E_{\mathbb{Q}}\left[e^{-r(\llcorner\tau\lrcorner_x+1)}\max(G_{\llcorner\tau\lrcorner_x+1},F_{\llcorner\tau\lrcorner_x+1})\mathbb{1}_{\{(\llcorner\tau\lrcorner_x+1)<T\}}+e^{-rT}F_T\mathbb{1}_{\{(\llcorner\tau\lrcorner_x+1)\geqslant T\}}\right],$$

where G_t is the guarantee paid at time t. Assume that mortality rates are deterministic and that death benefits are paid at the end of the year of death. Under those assumptions, the no-arbitrage price of the payoff of the variable annuity with a GMDB rider, denoted by $U^D_{0,F_0}(T,\beta,c)$, can be computed using the price for a GMMB obtained in the previous section. It can be seen as a weighted sum of GMMB prices. See, for example, Dickson, Hardy, and Waters (2009) for details.

Proposition 3.3.3 (Fair price for GMDB)**.** *Under the assumption of deterministic mortality rates, the fair price of a GMDB, $U^D_{0,F_0}(T,\beta,c)$ is given by*

$$\begin{aligned}U^D_{0,F_0}(T,\beta,c)&=E_{\mathbb{Q}}\left[\sum_{t=0}^{T}{}_tp_x\,q_{x+t}\,e^{-rt}\max(G_t,F_t)+{}_Tp_x\,e^{-rT}F_T\right]\\&=\sum_{t=0}^{T}{}_tp_x\,q_{x+t}U^M_{0,F_0}(t,\beta,G_t,c)+\ {}_Tp_x\,E_{\mathbb{Q}}\left[e^{-rT}F_T\right],\end{aligned}$$

where $U^M_{0,F_0}(t,\beta,G_t,c)$ is the market value of a variable annuity of maturity t with a GMMB rider with guaranteed amount G_t. We also have that

$$E_{\mathbb{Q}}\left[e^{-rT}F_T\right]=F_0e^{-rT-\frac{a^2}{2}T}A^D_T,$$

with

$$A^D_D=\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^T e^{(a+\sigma)x+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}\left((x-K)^--\max(x,0)\right)}f(x,y,z)dz\ dy\ dx,$$

where $f(x,y,z)$ is defined as in Proposition 3.2.2.

Proof. This result is immediate using (3.5) and Proposition 3.2.1. □

The fair fee rate for the GMDB benefit is such that the initial premium is equal to the expected value under $\mathbb{Q}$ of the discounted payoff. That is, for a given premium P, and assuming the GMDB is the only guarantee, c^* satisfies

$$P=U^D(T,\beta,c^*). \tag{3.12}$$

3.4 Numerical Results

In this section, we compare the state-dependent fee rates for VAs with guarantees, for a range of parameters and settings. Unless otherwise indicated, we let $r = 0.03$ and $\sigma = 0.2$; the contract term is 10 years. We assume that the initial premium is $P = 100$, with no initial fixed expense, so that $e = 0$. The policyholder is assumed to be aged 50, and mortality is assumed to follow the Gompertz model, with force of mortality $\mu_y = 0.00002\, e^{0.1008y}$.

Let g denote the guaranteed roll-up rate, which relates the guarantee G applying at T, and the initial premium P, as $G = P\, e^{gT}$. When $g = 0$, the guarantee is a 'return-of-premium' guarantee, so that $G = P$. For now, we assume that the fee is paid only below the guaranteed level ($\beta = G$).

In Figure 3.1, we show the market value of the VA with a GMMB as a function of the fee rate c. The point on the x-axis at which the curve of the no-arbitrage price of the payout crosses the line $y = 100$ corresponds to the fair fee rate (as computed in (3.11)). We show different curves for $g = 0\%$, $g = 1\%$ and $g = 2\%$. As expected, the fair fee increases as the guaranteed payoff increases, reaching almost 0.1 when the guaranteed roll-up rate g is 2%.

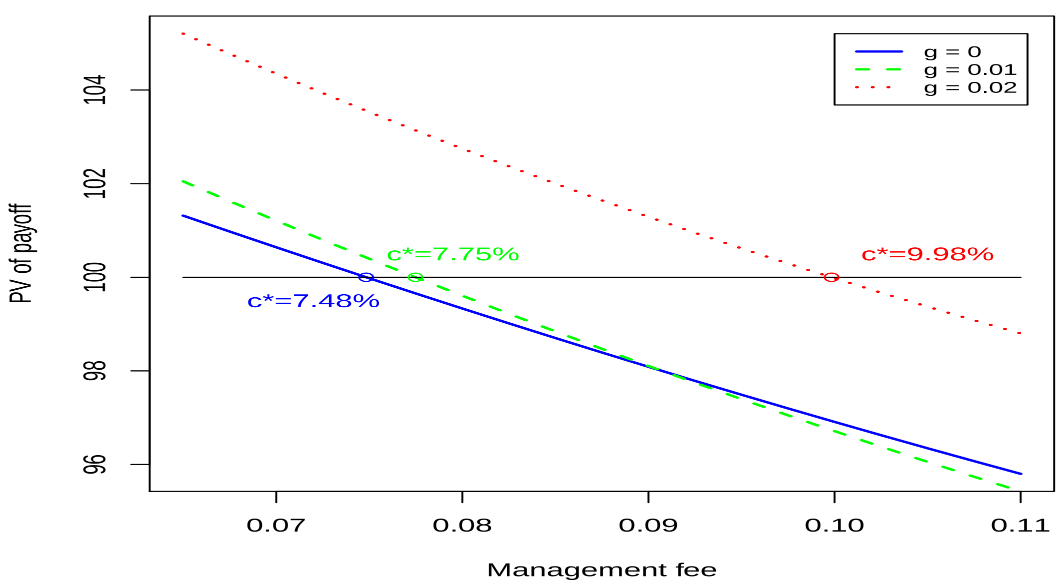


Figure 3.1: No arbitrage price of a VA with a 10-year GMMB as a function of c, with \$100 initial premium and guaranteed roll-up rates $g = 0\%$, 1% and 2%.

Figure 3.1 also illustrates, though, that the relationship between the guarantee level and fee rate is not straightforward. We note that the curves for $g = 0\%$ and $g = 1\%$ cross at around $c = 9\%$. At first this observation may seem counter-intuitive; for a conventional, static fee structure, a lower guarantee would lead to a lower fee rate. However, in our case, as the fee is paid only when the option is in the money, the effect of the guarantee roll-up rate g on the state-dependent fee rate c is not so clear. A higher guarantee generates a higher occupation time, so that the cost is spread over a longer payment period, which in some cases will lead to a lower fee rate.

Figure 3.2 illustrates the sensitivity of the dynamic fair fee rate with respect to the volatility σ and the term T. We find that the fair fee rate increases with σ and decreases with T. In relative terms, the maturity of the GMMB does not seem to affect the sensitivity of the fair fee to changes in σ.

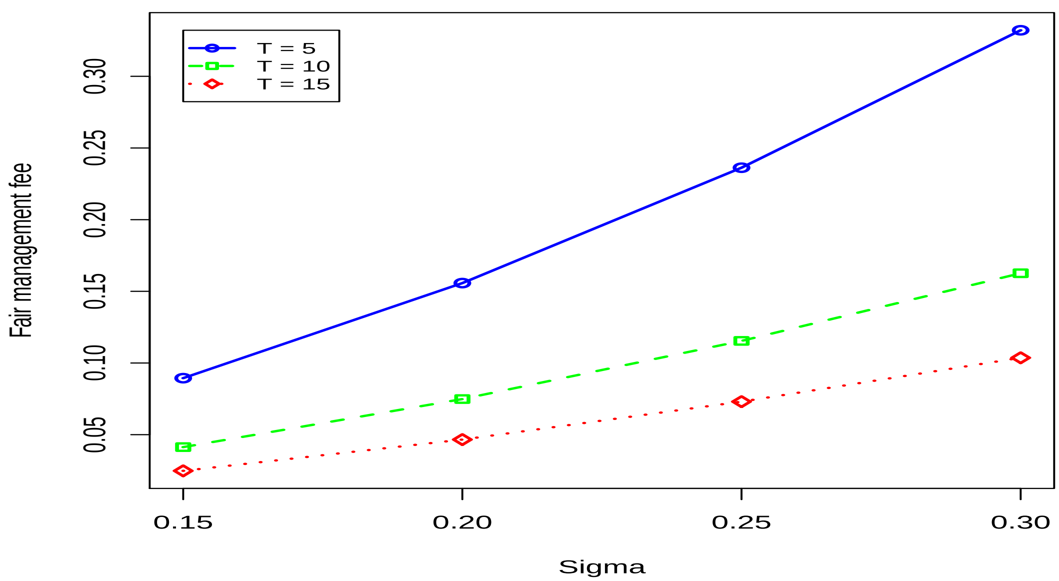


Figure 3.2: Sensitivity of fair fee rates for GMMB with respect to volatility (from $\sigma = 15\%$ to $\sigma = 30\%$) and contract term, $T = 5$, 10 or 15 years for a contract with $g = 0$.

Table 3.1 considers two contracts; a GMMB and a GMDB, both with fees paid only when the guarantee is in the money. For comparison purposes, results pertaining to a GMMB with a continuous fee rate are also illustrated. As expected, the fair fee rate for the GMMB is much higher when it is paid only when the guarantee is in the money. The difference is much more significant for shorter maturities. The GMDB results in a lower fair fee rate since the benefit is paid only if the policyholder dies during the life of the contract. In this case, since the policyholder is 50 years old, the probability of having to pay the benefit is quite low. Observe that the GMMB fair rate (as computed in (3.11)) is decreasing with respect to T whereas the GMDB fair rate (as computed in (3.12)) is increasing with respect to T. This is explained by the fact that the payoff of the GMDB benefit is a weighted sum of GMMB payoffs with different maturities (see Proposition 3.3.3), with weights linked to the probability that the policyholder dies during a given year. By increasing the maturity of the contract, we extend the period during which the death benefit can be paid, thus increasing the probability that it is paid at any given point before maturity. In addition, since the mortality rates increase with time, the fair fee rate is calculated assuming that less premium will be collected in later years since some policyholders will have died. For this reason, the fee rate needs to be higher for longer maturities.

Table 3.1: Fair fee rates (%) for the GMMB and GMDB with respect to maturity T when $\beta = G$. "(s-d fee)" refers to state-dependent fees and "(cst fee)" refers to a constant fee.

Type of Contract	$T = 5$	$T = 7$	$T = 10$	$T = 12$	$T = 15$
GMMB (s-d fee)	15.58	11.01	7.48	6.08	4.66
GMMB (cst fee)	3.53	2.43	1.58	1.24	0.91
GMDB (s-d fee)	0.1	0.12	0.17	0.21	0.27
GMDB (cst fee)	0.04	0.04	0.06	0.06	0.08

Table 3.2 illustrates the sensitivity of the state-dependent fee to the volatility assumption, compared to that of the static fee. Although the state-dependent fair fee is consistently higher, it is not significantly more sensitive to a small increase in volatility. In fact, when the volatility increases from 0.15 to 0.2, the static fee increases by about 84% while the state-dependent fee is 81% higher. However, to keep up with larger increases in volatility, the state-dependent fee rises faster. When the volatility doubles from 0.15 to 0.3, the state-dependent fee is multiplied by 3.94 while its constant counterpart is only 3.74 times higher. Thus, as σ increases, the sensitivity of the state-dependent fee to the volatility also increases, compared to the static fee.

For a simple return-of-premium GMMB, we see in Figure 3.1 that the fair fee under the dynamic fee structure is around 7.5% of the fund. This is a very high rate, and is not practical in marketing terms. This result though, is not that surprising. The static fee under the same assumptions is around 2%. The result shows that 2% of the fund paid throughout the policy term has the same value as 7.5% of the fund paid only when the fund is low. Thus, the state-dependent fee is paid for a shorter time, and is paid on a smaller fund.

If we accept that such a high fee is unlikely to be feasible, we can move the payment barrier higher, so that the option being funded still has significant value, even though it is out-of-the-

Table 3.2: Fair fee rates (%) for the 10-year GMMB with respect to volatility σ when $\beta = G$. "S-D" refers to state-dependent fees and "CST" refers to a constant fee.

Type of Fee	$\sigma = 0.15$	$\sigma = 0.2$	$\sigma = 0.25$	$\sigma = 0.3$
S-D	4.13	7.48	11.54	16.26
CST	0.86	1.58	2.38	3.22

money. That is, the policyholder would pay for the option when it is close to the money, and cease paying when it moved further away from the money. This adjustment retains the advantage of the differential fee structure with respect to disintermediation risk when the guarantee is far out-of-the-money, but, by increasing the occupation time, offers a more reasonable rate for policies which are close to the money.

Suppose that the dynamic fee is paid when the fund value is below $\beta = (1 + \lambda)G$, for some $\lambda \geqslant 0$. Figure 3.3 illustrates the fair fee rate for different values of λ. We see a dramatic effect on the fee rate for relatively low values of λ. Increasing the barrier from G to $1.2G$ causes the fair fee rate to decrease from 7.48% to 3.77%, which may improve the marketability of the contract significantly. Moreover, fixing the barrier at $1.34G$ or higher causes the fair fee to drop below 3.00%, which brings the risk-neutral drift of the fund value process back above 0. Asymptotically, the state-dependent fee rate of the GMMB converges to the static fee rate which is not state-dependent. Increasing the payment barrier decreases the fair fee rate, but it also adds an incentive to surrender. In fact, if the fee is paid while the guarantee is out-of-the-money, it may be optimal for the policyholder to surrender when the fund value is close to the guaranteed amount. This might happen when the expected value of the discounted future fees is more than the value of the financial guarantee. However, even with an increased payment barrier, the incentive to surrender for very high fund values is still eliminated, thus decreasing the value of the surrender option. We discuss this further in Section 3.5.

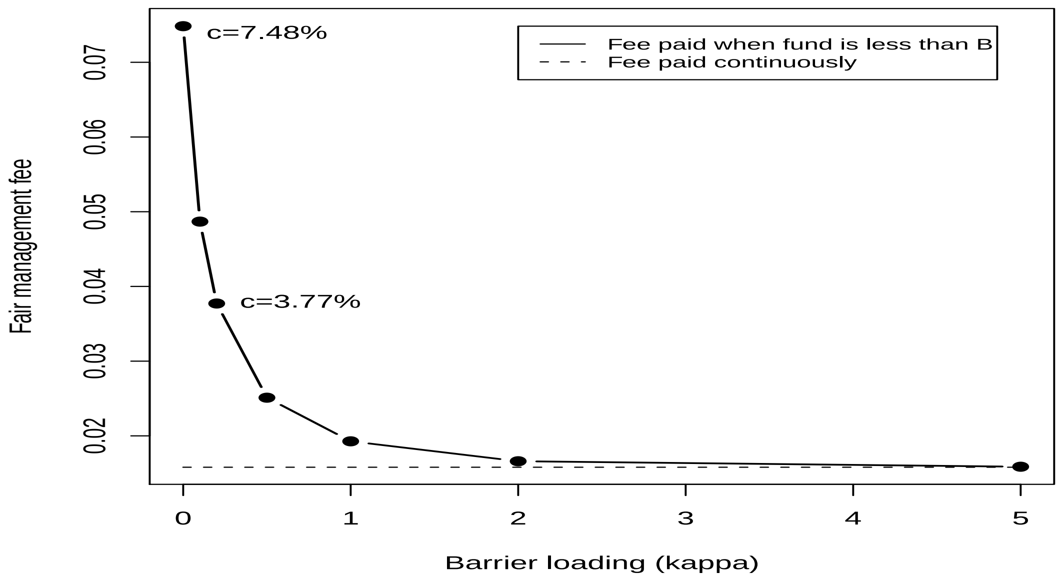


Figure 3.3: State-dependent fee rates for GMMB, for λ=0 to 5, $r = 0.03$, $T = 10$

3.5 Analysis of the Surrender Incentive

One of the motivations of state-dependent fees is to reduce the incentive to surrender the VA contract when the fund value is high, that is when the guarantee is out-of-the-money. It is clear that if the policyholder pays a constant fee, at some point the fund value could become so high that the option is essentially valueless, while the option fees have considerable drag on the fund accumulation. It would be rational to surrender, in order to stop paying the fees. However, if the fee is state-dependent, when the fund value is high the policyholder no longer pays fees, and will gain no benefit from surrendering.

In order to demonstrate this numerically, we consider the relationship between the value of the financial guarantee at maturity and the value of the future fees. At some time t before the maturity date T, the policyholder may surrender, with a payoff of F_t (we assume no surrender charge). If the policyholder holds the contract until maturity, she will receive the fund at T plus the additional guarantee payout (if any) – that is, she will receive $F_T + (G - F_T)^+$.

We will show, in Proposition 3.5.1, that it is optimal to retain the contract at time t if at that time the value of the future fees is less than the value of the maturity guarantee, which is the value of the put option payout $(G - F_T)^+$. However, the reverse is not generally true. That is, it may not be optimal to surrender the contract even when the value of the future fees is greater than the value of the maturity put option. The reason is that the surrender option is American-style, and if the value of the fees is not too much greater than the option value, it may be optimal to postpone the decision to surrender.

The option to postpone the surrender decision when fees are state-dependent will be examined in Chapter 4. However, a comparison of the value of the maturity put option benefit with the value of the future fees does allow some comparison of surrender incentives. We will demonstrate through an example how the state-dependent fee reduces the surrender incentive for out-of-the-money options, compared with a constant fee.

Proposition 3.5.1 (Sufficient conditions to retain the contract at time t)**.** *Consider a GMMB contract that can be surrendered and denote the present value of the maturity benefit at time t by $P(t, F_t)$. A first sufficient condition to retain the contract at time t is given by*

$$P(t, F_t) := E_t[e^{-r(T-t)} \max(F_T, G)] \geqslant F_t. \tag{3.13}$$

Let $m(t, F_t)$ denote the value at t of the future fees, and let $p(t, F_t)$ denote the value at t of the maturity put option, so that

$$m(t, F_t) := F_t - E_t[e^{-r(T-t)} F_T] \quad \text{and } p(t, F_t) := E_t\left[e^{-r(T-t)}(G - F_T)^+\right] \tag{3.14}$$

Then a second sufficient condition to retain the contract at time t is given by

$$p(t, F_t) \geqslant m(t, F_t). \tag{3.15}$$

Proof. Fix $t \in [0, T)$ and denote the value at time t of the full contract, including the surrender option, by $V(t, F_t)$. The surrender region is $\mathcal{A}$, defined as the region of F_t where $F_t > V(t, F_t)$ (as the surrender benefit is equal to F_t).

The full contract value can never be less than the present value of the maturity benefit, so $V(t, F_t) \geqslant P(t, F_t)$.

Suppose also that (3.13) holds, so that $P(t, F_t) \geqslant F_t$; then

$$V(t, F_t) \geqslant P(t, F_t) \geqslant F_t \Rightarrow F_t \notin \mathcal{A}$$

that is, F_t is outside the surrender region when $P(t, F_t) \geqslant F_t$. Condition (3.13) is proved.

Condition (3.15) follows from (3.13), as follows

$$
\begin{aligned}
P(t, F_t) &= E_t[e^{-r(T-t)}F_T] + E_t\left[e^{-r(T-t)}(G - F_T)^+\right] \\
&= F_t - m(t, F_t) + p(t, F_t) \\
&\Rightarrow P(t, F_t) - F_t = p(t, F_t) - m(t, F_t)
\end{aligned}
$$

then

$$
P(t, F_t) > F_t \Leftrightarrow p(t, F_t) > m(t, F_t)
$$

and condition (3.15) is proved. □

To illustrate the effect of the fee structure on the incentive to surrender, we consider a 5-year variable annuity contract. All other parameters are the same as in Section 3.4. We assume that the fee rate is set at the fair rate to fund the maturity benefit without allowing for surrenders.

3.5.1 Constant Fee

We first consider the case where a level fee is paid continuously, regardless of the value of the fund. Figure 3.4 shows the difference between the value of the financial guarantee and the expected value of the future fees.

As expected, at any time, there is a fund value above which the future fees are worth more than the financial guarantee at maturity. This can be clearly seen in the panels in Figure 3.4 which give snapshots at $t = 1, 2$ and 4. The curves all cross at a single point. Below this point, the option value is greater than the value of the future fees, and the policyholder should remain in the contract. Above the cross over point, there is another threshold, above which it would be optimal to surrender the policy. Between the cross over point and the surrender threshold, it is optimal to postpone the surrender decision. It is well documented that when the fee is constant, the optimal surrender strategy is to lapse the contract whenever the fund value is above a certain level (see, for example Chapter 2 and Milevsky and Salisbury (2001)). In this example the threshold is 125.2 at $t = 1$, 126.4 at $t = 2$ and 123.7 at $t = 4$ (these numbers were calculated following the algorithm in Chapter 2). That is, for example, at $t = 1$, the policyholder should surrender if the fund value is greater than 125.2, which has a probability of around 20%.

3.5.2 State-Dependent Fee

Next, we consider a contract with a state-dependent fee. We first assume that the payment barrier is $\beta = G = F_0$. In this case the state-dependent fee is very high. The difference between the value of the financial guarantee and the expected value of the future fees is shown in Figure

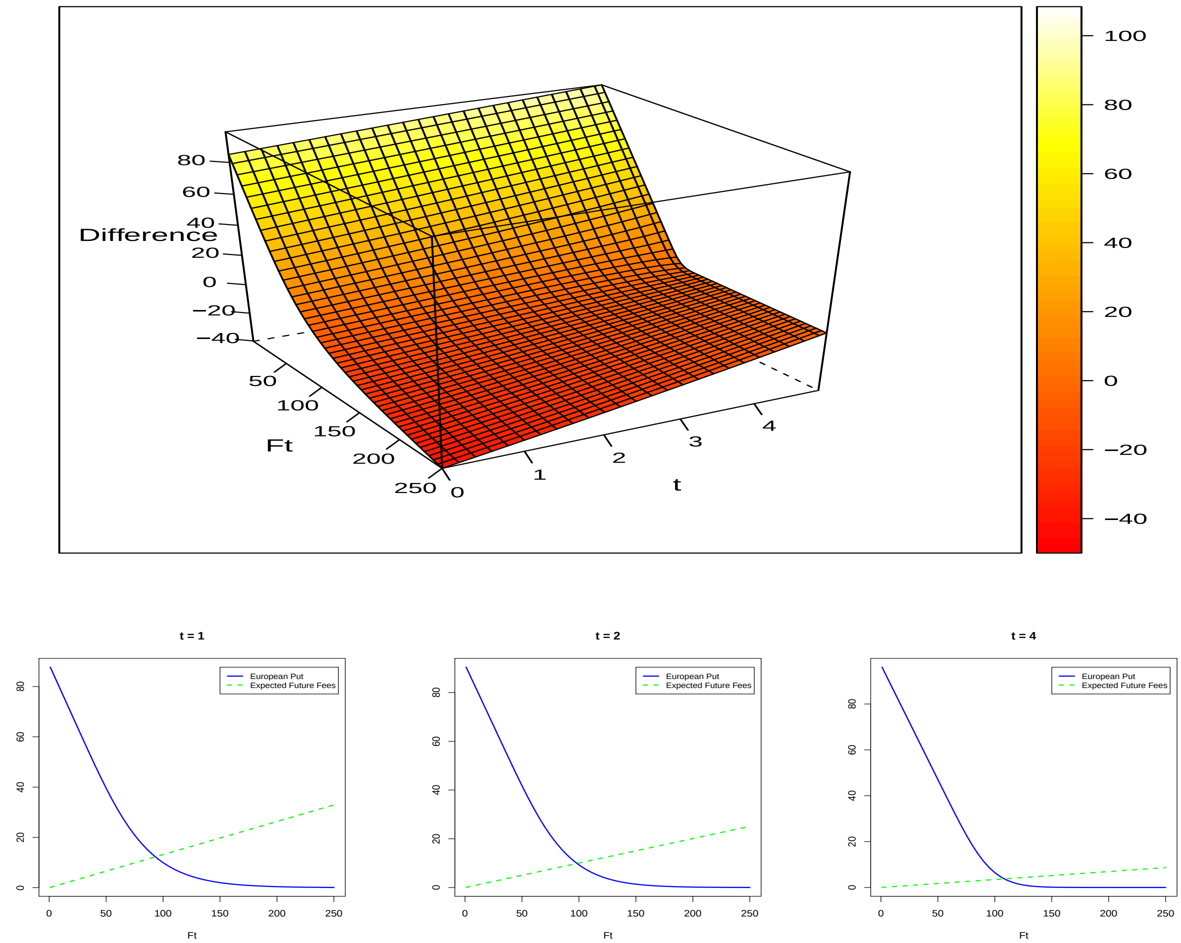


Figure 3.4: Difference at time t (when the fund value is F_t) between the value of the financial guarantee at maturity T and the expected value at t of the discounted future fees for a 5-year GMMB contract with $G = F_0 = 100$ and $c = 3.53\%$ paid continuously. Snapshots at $t = 1$, $t = 2$ and $t = 4$.

3.5. The smaller panels show the relationship between the future fee value and the option value at $t = 1, 2$ and 4, in terms of F_t.

One major difference in comparison to the constant fee case is that the surface does not drop below 0. That means that in this example, there is no fund value at any date where the value of the future fees exceeds the value of the option. Hence, following Proposition 3.5.1, it is never optimal to surrender this contract.

To complete this analysis, we consider a similar contract, but with an increased payment barrier β. We let $\beta = 1.4G$, again assuming $G = F_0$. In this case, the difference between the value of the financial guarantee and the expected future fee can become negative, but returns to 0 for high fund values (see Figure 3.6).

However, when the payment barrier is increased, there is an interval around the payment barrier where it may become optimal to surrender the contract since the policyholder can expect to pay more than he will receive (see Figure 3.6). This indicates that the surrender region when the fee is state-dependent has a different form than when the fee is constant; the optimal surrender strategy is no longer based on a simple threshold; there may be a surrender corridor, where surrender is optimal for F_t in the corridor, and not for higher or lower values of F_t. However, it is apparent that the area where there is no surrender incentive, because the option value is greater than or equal to the future fee value, is much larger than for the constant fee case.

We have demonstrated through an example that the optimal surrender area for the constant fee contract is much larger than for the state-dependent fee contract. However, we have not included a full analysis of optimal surrenders, nor have we priced the contracts assuming optimal (or deterministic) surrenders. As mentioned before, the small panels of Figure 3.6 indicate that the optimal surrender strategy is not of the threshold type. This makes a full analysis of the optimal surrender strategy more complex than for the traditional constant fee[5].

3.6 Model Risk

In the previous sections we have assumed that fund values may be modelled by a geometric Brownian motion, which allows us to obtain solutions in integral form. However, it is well-known that this model does not provide a good fit to the empirical distribution of stock returns over longer terms. Our results in the previous section may thus be viewed as approximations for real market values. In order to test their sensitivity to model risk, we find the fair fee rate when the stock returns are assumed to follow a regime-switching log-normal (RSLN) model. Hardy (2001) shows that this model provides a better fit for the distribution of long-term stock returns, allowing for heavier tails. It also reproduces the volatility clustering observed in empirical data.

[5]Because of the shape of the surrender region, we cannot use the integral representation techniques of Kim and Yu (1996), and least squares Monte Carlo techniques (used, for example, by Bacinello, Millossovich, Olivieri, and Pitacco (2011)) break down due to the fact that the difference between the surrender benefit and the continuation option is so small for much of the range of F_t, that numerical errors are too significant. The problem is manageable using partial differential equations, as in Dai, Kuen Kwok, and Zong (2008), Chen, Vetzal, and Forsyth (2008) and Belanger, Forsyth, and Labahn (2009), but the technical challenges involved take such analysis beyond the scope of this chapter.

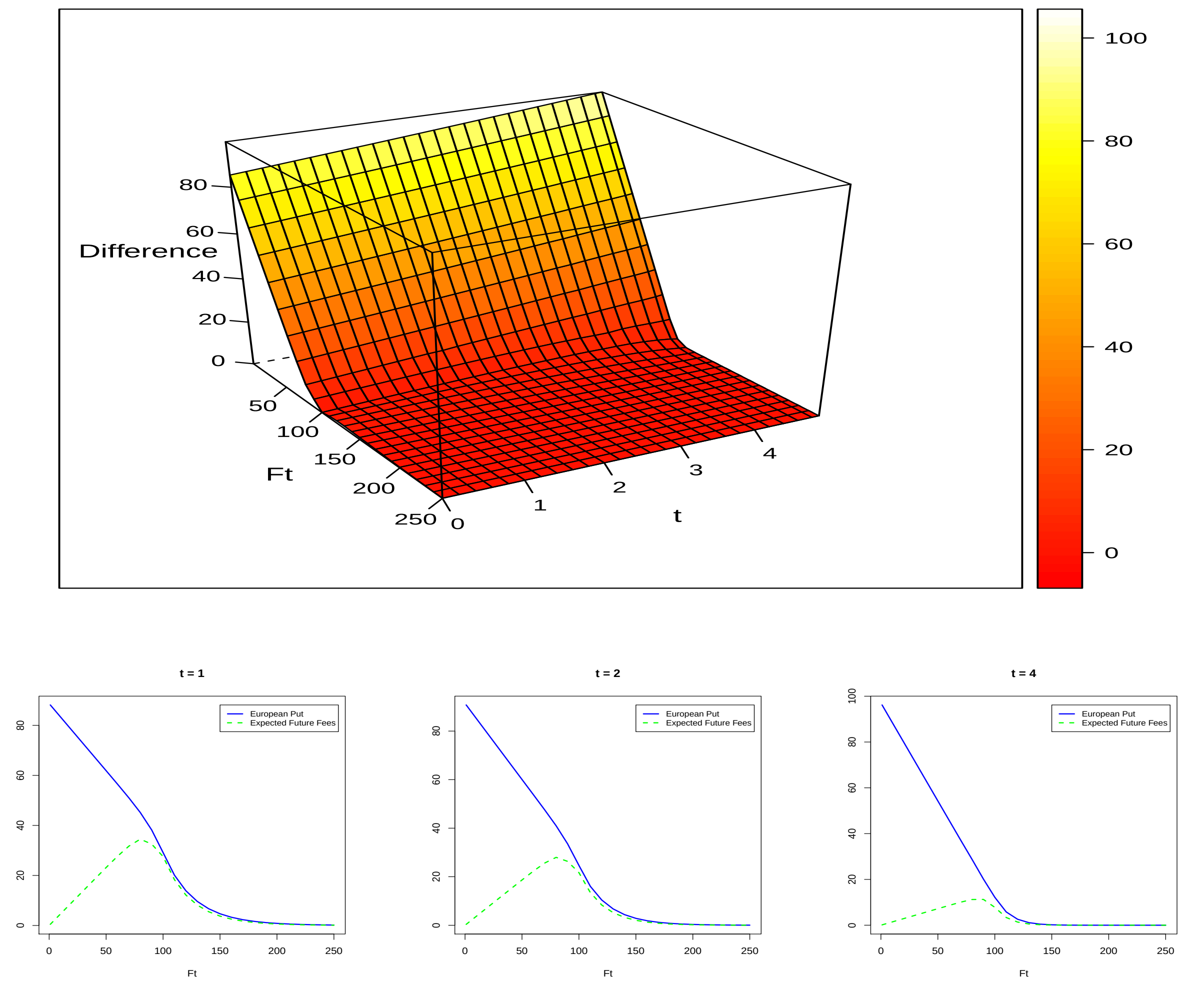


Figure 3.5: Difference at time t (when the fund value is F_t) between the value of the financial guarantee at maturity T and the expected value at t of the discounted future fees, for a 5-year GMMB contract with $\beta = G$, $G = F_0 = 100$ and $c = 15.58\%$ paid when $F_t < \beta$. Snapshots at $t = 1$, $t = 2$ and $t = 4$.

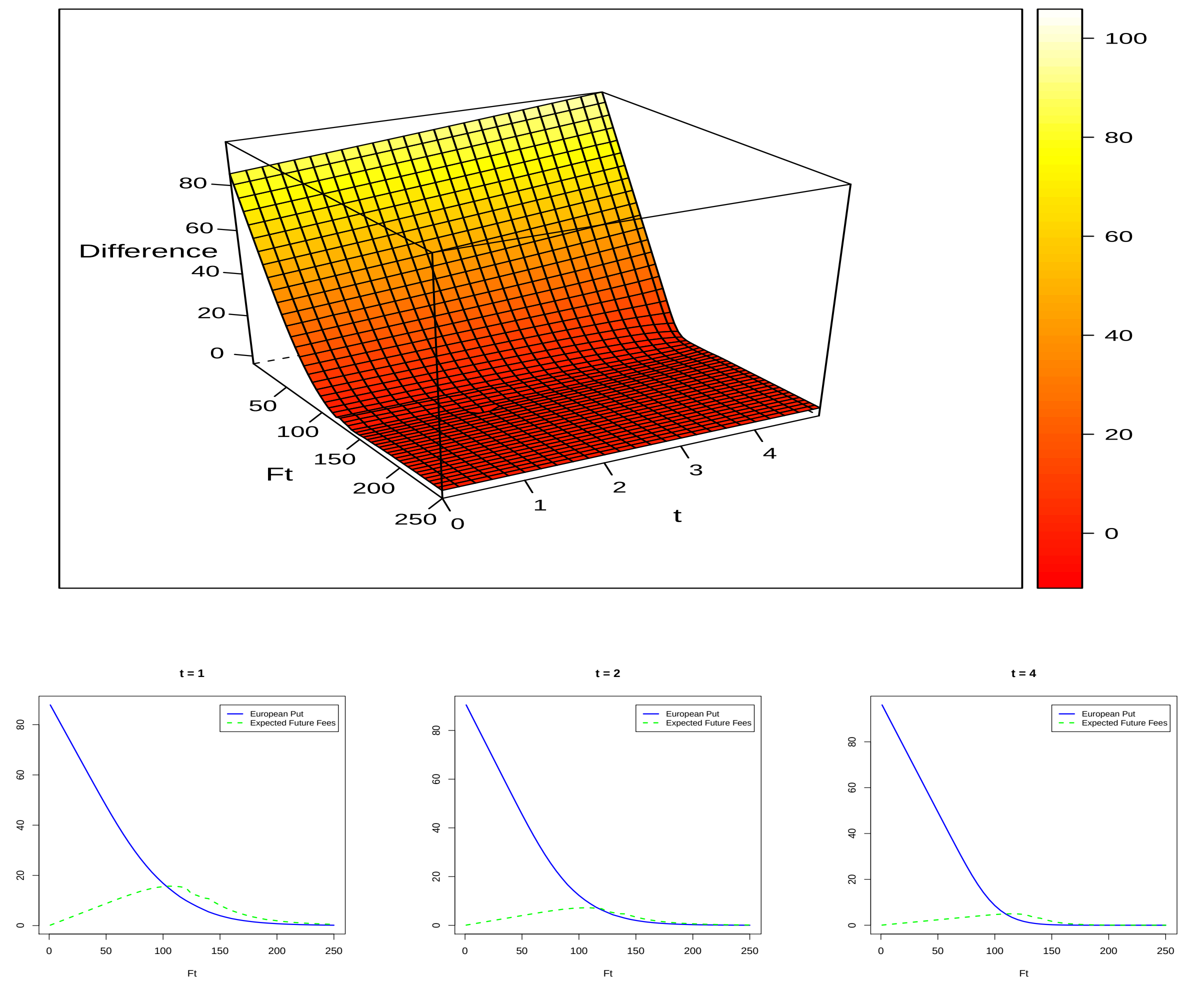


Figure 3.6: Difference at time t (when the fund value if F_t) between the value of the financial guarantee at maturity T and the expected value at t of the discounted future fees for a 5-year GMMB contract with $\beta = 1.4G$, $G = F_0 = 100$ and $c = 4.84\%$ paid when $F_t < \beta$. Snapshots at $t = 1$, $t = 2$ and $t = 4$.

The RSLN model is based on an underlying state variable, the value of which is governed by a transition matrix P (of size 2×2 if there are 2 possible regimes). The elements of this matrix, denoted $p_{i,j}$ represent the probability that the state variable moves to state j given that it is currently in state i. Between each transition, stock returns follow a log-normal distribution whose parameters are determined by the regime indicated by the state variable. Hardy (2001) presents a two-regime model in which transitions occur monthly. These characteristics are very suitable for our purposes. In fact, while we assumed earlier that management fees were withdrawn in a continuous manner, many insurance companies collect them monthly. This also has an impact on the fair fee rate that should be charged.

In this section, in addition to testing the sensitivity of the fair fee rate to model changes, we also analyze the impact of discrete fee collection. When fees are withdrawn in discrete time, getting an analytical expression for the present value of the guarantee payoff is not always possible. For this reason, we use Monte Carlo simulations to obtain the results given in this section. We consider GMMB guarantees with maturities ranging from 5 to 15 years. To quantify the model impact and fee discretization separately, we find the fair fee rates under three models. The first case is identical to the setting presented in the previous sections; the value of the fund follows a geometric Brownian motion (GBM) and fees are paid continuously. In the second case, the return on the fund is also assumed to be log-normal, but the fees are paid monthly. The last case assumes that fund returns follow a regime-switching log-normal model and that the fees are paid monthly. We use a risk-free rate of 0.03. For the GBM, we use $\sigma = 0.14029$ to match Hardy (2001). For the RSLN, we used the parameters given in Table 3.3, which are taken from Hardy (2001). The volatility parameters σ_1 and σ_2 are expressed per month.

Table 3.3: Regime-switching log-normal parameters used for Monte Carlo simulations

Parameter	σ_1	σ_2	$p_{1,2}$	$p_{2,1}$
Value	0.035	0.0748	0.0398	0.3798

Table 3.4 presents the fair fee rates obtained using 5×10^6 simulations. The column RSLN corresponds to the fair rates computed by Monte Carlo simulations when the underlying index price follows a regime switching model and fees are paid monthly when the fund is in-the-money, assuming a return-of-premium guarantee. The column GBM-D corresponds to the fair rates computed by Monte Carlo when the underlying index price evolves as in the Black-Scholes model and fees are paid monthly when the fund is below the return-of-premium guarantee. Finally GBM-C corresponds to the fair rates computed using Proposition 3.3.2 and solving (3.11) when the fee is paid continuously, conditional on the fund value lying below the return-of-premium guarantee.

We observe from Table 3.4 that the RSLN fees are close to the Black-Scholes fees, given

Table 3.4: Fair fee rates (%) in the Regime Switching model (RSLN) and in the Black-Scholes model when fees are paid monthly (GBM-D) and when fees are paid continuously (GBM-C)

T	RSLN	GBM-D	GBM-C
5	7.18	7.27	7.82
10	3.43	3.44	3.57
15	2.07	2.06	2.11

monthly payments, which indicates that the results appear to be fairly robust with respect to model risk, particularly with respect to fat tails. A possible explanation for this may be the fact that when the fund does not perform well, the guarantee is more likely to be in the money and the fee is collected. If the guarantee is in the money more often, the fee is also paid for a longer period of time. Thus, although fatter tails may lead to potentially higher benefits to pay, they can drag the fund down and cause the fee to be paid more often. In this case, a slightly lower fee rate can still be sufficient to cover the guarantee. We also note that the Black-Scholes model with continuous fees generates slightly higher fee rates than for the monthly fees, which arises from the difference between the occupation time for the discrete case and the continuous case, with lower expected occupation time where the process is continuous.

3.7 Concluding Remarks

This chapter finds the fair fee rate for European-type guaranteed benefits in variable annuities when the fee payment is contingent on the position of the value of the underlying fund relative to some critical level β.

This is a first step towards a dynamic state-dependent charging structure, where the fee rate depends on the fund value and the dynamic value of the embedded guarantees. We have considered pricing here, but we recognize that the fee structure could also have an important impact on hedging performance and on the surrender rates. In future research we will investigate valuation and risk management allowing for dynamic, state-dependent surrenders, and also consider the impact of a non-level fee threshold.

Appendix

3.A Proof of Proposition 3.2.2

Proof. We first prove the case where $K \geqslant 0$. Observe that

$$\begin{aligned}
&\widetilde{\mathbb{Q}}\left(\widetilde{W}_T \in dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right) \\
&\quad = \widetilde{\mathbb{Q}}\left(-\widetilde{W}_T \in -dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right) \\
&\quad = \widetilde{\mathbb{Q}}\left(Z_T \in K - dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right),
\end{aligned}$$

where $Z_t = K - \widetilde{W}_t$ is a $\widetilde{\mathbb{Q}}$-Brownian motion starting at K.
Next, we need to express $L_{T,\widetilde{W}}(K)$ and $\Gamma^-_{T,\widetilde{W}}(K)$ in terms of $L_{T,Z}(0)$ and $\Gamma^+_{T,Z}(0)$ (where $\Gamma^+_{T,Z}(0)$ denotes the occupation time of Z above 0). Then, it will be possible to use the results of Karatzas and Shreve (1984) to obtain the desired distribution. Note that

$$\Gamma^-_{T,\widetilde{W}}(K) = \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} ds = \int_0^T \mathbb{1}_{\{Z_s \geqslant 0\}} ds = \Gamma^+_{T,Z}(0). \tag{3.16}$$

Using the Tanaka formula, we also have

$$
\begin{aligned}
L_{T,\widetilde{W}}(K) &= (\widetilde{W}_T - K)^- - (0-K)^- + \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} d\widetilde{W}_s \\
&= (-Z_T)^- - K - \int_0^T \mathbb{1}_{\{Z_s \geqslant 0\}} dZ_s \\
&= (-Z_T)^- - K - \int_0^T \left(1 - \mathbb{1}_{\{Z_s<0\}}\right) dZ_s \\
&= (-Z_T)^- - K - Z_T + K + \int_0^T \mathbb{1}_{\{Z_s<0\}} dZ_s \\
&= (-Z_T)^+ + \int_0^T \mathbb{1}_{\{Z_s<0\}} dZ_s, \qquad (3.17)
\end{aligned}
$$

where we use $\int_0^T dZ_s = Z_T - Z_0 = Z_T - K$ to obtain the second to last equation. The Tanaka formula also allows us to express $L_{T,Z}(0)$ as

$$
L_{T,Z}(0) = Z_T^- - K^- + \int_0^T \mathbb{1}_{\{Z_s<0\}} dZ_s, \qquad (3.18)
$$

where $K^- = 0$ as $K \geqslant 0$. Re-arranging (3.17) using (3.18), we obtain

$$
L_{T,\widetilde{W}}(K) = L_{T,Z}(0) + (-Z_T)^+ - Z_T^-.
$$

Since $(-Z_T)^+ - Z_T^- = 0$, we get

$$
L_{T,\widetilde{W}}(K) = L_{T,Z}(0). \qquad (3.19)
$$

Then using (3.16) and (3.19), we can write

$$
\begin{aligned}
&\widetilde{\mathbb{Q}}\left(Z_T \in K - dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right) \\
&\quad = \widetilde{\mathbb{Q}}\left(Z_T \in K - dx, L_{T,Z}(0) \in dy, \Gamma^+_{T,Z}(0) \in dz\right). \qquad (3.20)
\end{aligned}
$$

Since Z_t is a Brownian motion starting at K, we can use the trivariate density of (Z_T, $L_{T,Z}(0)$, $\Gamma^+_{T,Z}(0)$) given in Section 4 of Karatzas and Shreve (1984) and get the desired result.

Similarly, for $K < 0$, we can observe that

$$\begin{aligned}
&\widetilde{\mathbb{Q}}\left(\widetilde{W}_T \in dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right) \\
&\quad = \widetilde{\mathbb{Q}}\left(\widetilde{W}_T - K \in dx - K, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right) \\
&\quad = \widetilde{\mathbb{Q}}\left(H_T \in dx - K, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right),
\end{aligned}$$

where $H_t = \widetilde{W}_t - K$ be a $\widetilde{\mathbb{Q}}$-Brownian motion starting at $-K$. We now express $L_{T,\widetilde{W}}(K)$ and $\Gamma^-_{T,\widetilde{W}}(K)$ in terms of $L_{T,H}(0)$ and $\Gamma^+_{T,H}(0)$. We have

$$\Gamma^-_{T,\widetilde{W}}(K) = \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} ds = \int_0^T \mathbb{1}_{\{H_s \leqslant 0\}} ds = \int_0^T 1 - \mathbb{1}_{\{H_s \geqslant 0\}} ds = T - \Gamma^+_{T,H}(0).$$

We also have

$$\begin{aligned}
L_{T,\widetilde{W}}(K) &= (\widetilde{W}_T - K)^- - (0 - K)^- + \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} d\widetilde{W}_s \\
&= H_T^- + \int_0^T \mathbb{1}_{\{H_s \leqslant 0\}} dH_s \\
&= L_{T,H}(0).
\end{aligned} \tag{3.21}$$

Then we can write

$$\begin{aligned}
&\widetilde{\mathbb{Q}}\left(H_T \in dx - K, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\right) \\
&= \widetilde{\mathbb{Q}}\left(H_T \in dx - K, L_{T,H}(0) \in dy, \Gamma^+_{T,H}(0) \in T - dz\right).
\end{aligned} \tag{3.22}$$

Since H_t is a Brownian motion starting at $-K$ (remember that $K < 0$), the result follows from Section 4 of Karatzas and Shreve (1984). □

3.B Details for the GMMB price

Note that A_T and D_T in the expression (3.10) in Proposition 3.3.1 depend on the trivariate density established in Proposition 3.2.2, which depends on K and β. We now discuss all possible cases needed to implement this formula.

Assume $K \geqslant 0$ and $\beta \geqslant G$.

In this case $K \geqslant H$ and we have $A_T = A_T^{(1)} + A_T^{(2)} + A_T^{(3)}$ where

$$\begin{cases} A_T^{(1)} = \int\limits_H^K \int\limits_0^{+\infty} \int\limits_0^T e^{(\sigma+a)x+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z,y)\; h(z,y+2K-x)dzdydx \\ A_T^{(2)} = \int\limits_K^{+\infty} \int\limits_0^{+\infty} \int\limits_0^T e^{(\sigma+a)x+bz-\frac{c}{\sigma}y}\, 2h(z,y+K)\; h(T-z,y-K+x)dzdydx \\ A_T^{(3)} = \frac{e^{bT+\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int\limits_H^K e^{\left(a+\sigma-\frac{c}{\sigma}\right)x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

We also have $D_T = D_T^{(1)} + D_T^{(2)}$ where

$$\begin{cases} D_T^{(1)} = \int\limits_{-\infty}^H \int\limits_0^{+\infty} \int\limits_0^T e^{ax+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z,y)\; h(z,y+2K-x)dzdydx \\ D_T^{(2)} = \frac{e^{bT+\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int\limits_{-\infty}^H e^{\left(a-\frac{c}{\sigma}\right)x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

Assume $K < 0$ and $\beta \geqslant G$.

In this case $K \geqslant H$ and we have $A_T = A_T^{(4)} + A_T^{(5)} + A_T^{(6)}$ where

$$\begin{cases} A_T^{(4)} = \int\limits_H^K \int\limits_0^{+\infty} \int\limits_0^T e^{(\sigma+a)x+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z,y-K)\; h(z,y-x+K)dzdydx \\ A_T^{(5)} = \int\limits_K^{+\infty} \int\limits_0^{+\infty} \int\limits_0^T e^{(\sigma+a)x+bz-\frac{c}{\sigma}y}\, 2h(z,y)\; h(T-z,y+x-2K)dzdydx \\ A_T^{(6)} = \frac{e^{\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int\limits_H^K e^{\left(a+\sigma-\frac{c}{\sigma}\right)x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

We also have $D_T = D_T^{(3)} + D_T^{(4)}$ where

$$\begin{cases} D_T^{(3)} = \int\limits_{-\infty}^H \int\limits_0^{+\infty} \int\limits_0^T e^{ax+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z,y-K)\; h(z,y-x+K)dzdydx \\ D_T^{(4)} = \frac{e^{\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int\limits_{-\infty}^H e^{\left(a-\frac{c}{\sigma}\right)x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

Assume $K \geqslant 0$ and $\beta < G$.

In this case $K < H$ and we have

$$A_T = \int_H^{+\infty} \int_0^{+\infty} \int_0^T e^{(\sigma+a)x+bz-\frac{c}{\sigma}y}\, 2h(z, y+K)\ h(T-z, y-K+x) dz dy dx.$$

We also have $D_T = \bar{D}_T^{(1)} + \bar{D}_T^{(2)} + \bar{D}_T^{(3)}$ where

$$\begin{cases} \bar{D}_T^{(1)} = \int_{-\infty}^{K} \int_0^{+\infty} \int_0^T e^{ax+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z, y)\ h(z, y+2K-x) dz dy dx \\ \bar{D}_T^{(2)} = \int_K^{H} \int_0^{+\infty} \int_0^T e^{ax+bz-\frac{c}{\sigma}y}\, 2h(z, y+K)\ h(T-z, y-K+x) dz dy dx \\ \bar{D}_T^{(3)} = \frac{e^{bT+\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_{-\infty}^{K} e^{\left(a-\frac{c}{\sigma}\right)x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

Assume $K < 0$ and $\beta < G$.

In this case $K < H$ and we have

$$A_T = \int_H^{+\infty} \int_0^{+\infty} \int_0^T e^{(\sigma+a)x+bz-\frac{c}{\sigma}y}\, 2h(z, y)\ h(T-z, y+x-2K) dz dy dx$$

We also have that $D_T = \bar{D}_T^{(4)} + \bar{D}_T^{(5)} + \bar{D}_T^{(6)}$ where

$$\begin{cases} \bar{D}_T^{(4)} = \int_{-\infty}^{K} \int_0^{+\infty} \int_0^T e^{ax+bz-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z, y-K)\ h(z, y-x+K) dz dy dx \\ \bar{D}_T^{(5)} = \int_K^{H} \int_0^{+\infty} \int_0^T e^{ax+bz-\frac{c}{\sigma}y}\, 2h(z, y)\ h(T-z, y+x-2K) dz dy dx \\ \bar{D}_T^{(6)} = \frac{e^{\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_{-\infty}^{K} e^{\left(a-\frac{c}{\sigma}\right)x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

To compute $A_T^{(3)}$, we observe that for any $\alpha \in \mathbb{R}$, we have the following identity

$$\frac{1}{\sqrt{2\pi T}} \int_H^K e^{\alpha x} \left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}} \right) dx =$$
$$e^{\frac{\alpha^2 T}{2}} \left[N\left(\frac{\alpha T - H}{\sqrt{T}}\right) - N\left(\frac{\alpha T - K}{\sqrt{T}}\right) - e^{2\alpha K} N\left(\frac{2K - H + \alpha T}{\sqrt{T}}\right) + e^{2\alpha K} N\left(\frac{K + \alpha T}{\sqrt{T}}\right) \right].$$

After replacing $\alpha = a + \sigma - \frac{c}{\sigma} = \frac{r-c}{\sigma} + \frac{\sigma}{2}$, and simplify, we obtain

$$A_T^{(3)} = \left(\frac{\beta}{F_0}\right)^{\frac{c}{\sigma^2}} e^{\frac{T}{2\sigma^2}(\gamma^2 - c(c-2r+\sigma^2))} \left[N\left(\frac{\ln\left(\frac{F_0}{G}\right) + \gamma T}{\sigma\sqrt{T}}\right) - N\left(\frac{\ln\left(\frac{F_0}{\beta}\right) + \gamma T}{\sigma\sqrt{T}}\right)\right.$$
$$\left. - \left(\frac{\beta}{F_0}\right)^{\frac{2\gamma}{\sigma^2}} N\left(\frac{\ln\left(\frac{\beta^2}{F_0 G}\right) + \gamma T}{\sigma\sqrt{T}}\right) + \left(\frac{\beta}{F_0}\right)^{\frac{2\gamma}{\sigma^2}} N\left(\frac{\ln\left(\frac{\beta}{F_0}\right) + \gamma T}{\sigma\sqrt{T}}\right)\right],$$

where $\gamma = r + \frac{\sigma^2}{2} - c$.

Define $\mathcal{I}(\alpha_1, \alpha_2)$ as follows

$$\mathcal{I}(\alpha_1, \alpha_2) := \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{\alpha_2} e^{\alpha_1 x} \left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}}\right) dx.$$

We use the following identity to get closed-form expressions for $\bar{D}_T^{(3)}$ and $D_T^{(2)}$:

$$\mathcal{I}(\alpha_1, \alpha_2) = e^{\frac{\alpha_1^2 T}{2}} \left[N\left(\frac{\alpha_2 - \alpha_1 T}{\sqrt{T}}\right) - e^{2\alpha_1 K} N\left(\frac{\alpha_2 - \alpha_1 T - 2K}{\sqrt{T}}\right)\right].$$

It is clear that $\bar{D}_T^{(3)} = e^{bT} \left(\frac{\beta}{F_0}\right)^{\frac{c}{\sigma^2}} \mathcal{I}\left(a - \frac{c}{\sigma}, K\right)$ and $D_T^{(2)} = e^{bT} \left(\frac{\beta}{F_0}\right)^{\frac{c}{\sigma^2}} \mathcal{I}\left(a - \frac{c}{\sigma}, H\right)$. After simplifications, we obtain $\bar{D}_T^{(3)}$

$$\bar{D}_T^{(3)} = \left(\frac{\beta}{F_0}\right)^{\frac{c}{\sigma^2}} e^{\left(r - \frac{\sigma^2}{2}\right)^2 \frac{T}{2\sigma^2}} \left(N\left(\frac{\ln\left(\frac{\beta}{F_0}\right) - \eta T}{\sigma\sqrt{T}}\right) - \left(\frac{\beta}{F_0}\right)^{\frac{2\eta}{\sigma^2}} N\left(\frac{\ln\left(\frac{F_0}{\beta}\right) - \eta T}{\sigma\sqrt{T}}\right)\right),$$

where $\eta = r - \frac{\sigma^2}{2} - c$. Note that $D_T^{(2)}$ can be obtained similarly as the only difference is that H replaces K. Thus we only need to replace β by G in the above expression to obtain $D_T^{(2)}$. □

Chapter 4

Optimal surrender under the state-dependent fee structure

4.1 Introduction

This chapter is based on a paper that was written in collaboration with Dr. Maciej Augustyniak (Université de Montréal), Dr. Carole Bernard and Dr. Mary Hardy.

This chapter aims to provide some insight on a very practical question: How can an insurer take advantage of product design to mitigate lapse risk, and to simplify risk management (hedging) in VAs? To answer this question, we examine the interplay between the fee structure of a VA with a guaranteed minimum accumulation benefit (GMAB) and the schedule of surrender charges. We propose to reduce the surrender incentive using product design to construct a contract which will be rarely optimal to lapse, if ever, while still being marketable. By achieving this, we greatly simplify the strategy that hedges optimal policyholder behaviour, since the hedge can be established as if no surrenders were allowed. Most importantly, it eliminates the need to model surrender behaviour for pricing and hedging purposes, thus reducing the risk of having an inappropriate lapse model. In such a product design, lapse assumptions mainly impact the profitability analysis of the product, and have little influence on the hedging strategy.

When the fee is paid as a fixed percentage of the fund, we demonstrate that there exists a model-free minimal surrender charge function which results in lapsation being sub-optimal during the whole length of the contract, and derive it in explicit form. However, these surrender penalties generally lead to a product design which is not marketable. For this reason, we consider the state-dependent fee structure presented in Chapter 3, where the fee is paid only when the account value is below a certain threshold. We analyse the optimal surrender behaviour under such a fee structure in the presence of surrender charges. We show how to solve for the minimal surrender charge function which results in lapsation being sub-optimal during the whole length of the contract. We explore different product designs that are able to eliminate the surrender incentive while keeping the contract marketable and attractive to policyholders. We find that by combining a state-dependent fee with surrender charges, it is possible design a contract that can be hedged and managed reasonably well, while remaining attractive to policyholders (with relatively low fees). In particular, when the surrender incentive is eliminated, the hedging strategy is much simpler since it is reduced to replicating the maturity benefit. Through the analysis of hedging errors, we also show that such a hedging strategy performs well under optimal and sub-optimal lapse behaviour, making the state-dependent fee an attractive design from a risk-management perspective.

Section 4.2 of this chapter introduces the model and the contract to price. In Section 4.3, we derive examples of optimal surrender regions. Section 4.4 presents the theoretical results useful to design a contract that eliminates the surrender incentive. In Section 4.5, we give an example of such a contract and analyse the effectiveness of a dynamic hedge under different surrender behaviours. Section 4.6 concludes.

4.2 Pricing the GMMB

4.2.1 Market and Notation

We consider a VA contract with maturity T and underlying account value at time t denoted by F_t, $t \in [0, T]$. Suppose that the initial premium F_0 is fully invested in an index whose value process $\{S_t\}_{0 \leqslant t \leqslant T}$ has real-world ($\mathbb{P}$-measure) dynamics

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t^{\mathbb{P}},$$

where $W_t^{\mathbb{P}}$ is a $\mathbb{P}$-Brownian motion.[1] Suppose also that the usual assumptions of the Black-Scholes model are satisfied. Then, the market is complete and there exists a unique risk-neutral measure $\mathbb{Q}$ under which the index S_t follows a geometric Brownian motion with drift equal to the risk-free rate r. Then we have

$$\frac{dS_t}{S_t} = r dt + \sigma dW_t^{\mathbb{Q}}.$$

[1]We work on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leqslant t \leqslant T}, \mathbb{P})$ where $(\Omega, \mathcal{F})$ is a measurable space, $\{\mathcal{F}_t\}_{0 \leqslant t \leqslant T}$ is the natural filtration generated by the Brownian motion (with $\mathcal{F}_t = \sigma(\{W_s\}_{0 \leqslant s \leqslant t})$) and $\mathbb{P}$ is the real-world measure. We assume that the probability space is complete ($\mathcal{F}_0$ contains the $\mathbb{P}$-null sets) and right-continuous.

We consider the state-dependent fee struture explored in Chapter 3, under which the insurer only charges the fee when the account value is below a given level β. This level is set as a multiple of the guaranteed amount and is defined as

$$\beta = (1+\lambda)G,$$

where $\lambda \geqslant 0$ is the fee payment barrier loading. In this setting, the $\mathbb{P}$-dynamics of the account value are given by

$$\frac{dF_t^{(\beta)}}{F_t^{(\beta)}} = (\mu - c\mathbb{1}_{\{F_t^{(\beta)}<\beta\}})dt + \sigma dW_t,$$

where $\mathbb{1}_A$ is the indicator function of the set A and where the superscript β indicates the dependence of the account value on the fee barrier. Without loss of generality, we assume that $F_0 = S_0$. Chapter 3 provides integral representations for the prices of guaranteed minimum accumulation and death benefits (GMAB and GMDB) when surrenders are not allowed. These integral representations allow us to solve for the fair value of c numerically when the fee is state-dependent. When the fee is paid regardless of the account value, or $\beta = \infty$, these integral representations simplify to expressions very similar to the Black-Scholes formula. In that case, the account value at time t is denoted by $F_t^{(\infty)}$.

4.2.2 Pricing VAs in the presence of a state-dependent fee and surrender charges

We focus on a T-year VA contract with a GMAB, having payoff $\max(G, F_T^{(\beta)})$ at maturity. The symbol G denotes a pre-determined guaranteed amount equal to

$$G = e^{gT}F_0^{(\beta)},$$

where $0 \leqslant g < r$ is the guaranteed roll-up rate. If the policyholder surrenders the contract at any time $0 < t < T$, she receives $(1-\kappa_t)F_t^{(\beta)}$: the account value diminished by the surrender charge $\kappa_t F_t^{(\beta)}$, where $0 \leqslant \kappa_t < 1$. Typically, κ_t is a decreasing function of time to discourage policyholders to lapse in the first years of the contract. In fact, early surrenders affect insurers more significantly because VA contracts have acquisition expenses that are expected to be recouped during the first years of the contract. For this reason, insurers generally charge high surrender penalties during the first few years of the contract. Since the contract cannot be surrendered at maturity, we define $\kappa_T = 0$. We will make further assumptions on the form of κ_t in the numerical examples.

We let $V(t, F_t^{(\beta)})$ denote the value of the contract at time t, $0 \leqslant t \leqslant T$. Since the VA contract can be surrendered at any time before maturity, its pricing becomes an optimal stopping problem. To define this problem, we must first introduce further notation. Denote by $\mathcal{T}_t$ the set of all stopping times τ greater than or equal to t and bounded by T. Then, define the *continuation value* of the VA contract with surrender as

$$V^*(t, F_t^{(\beta)}) = \sup_{\tau \in \mathcal{T}_t} E_{\mathbb{Q}}[e^{-r(\tau - t)}\psi(\tau, F_\tau^{(\beta)})|\mathcal{F}_t],$$

where,

$$\psi(t, F_t^{(\beta)}) = \begin{cases} (1-\kappa_t)F_t^{(\beta)}, & \text{if } t \in (0, T), \\ \max(G, F_T^{(\beta)}), & \text{if } t = T, \end{cases}$$

is the payoff of the contract at surrender or at maturity.

Let $\mathcal{R}_t$ be the optimal surrender region at time $t \in [0, T]$ and define it by

$$\mathcal{R}_t = \{F_t^{(\beta)} < \infty : \psi(t, F_t^{(\beta)}) \geqslant V^*(t, F_t^{(\beta)})\}.$$

Thus, the optimal surrender region is defined as the fund values for which the surrender benefit is worth at least as much as the VA contract when the policyholder continues to hold on to it for at least a small amount of time. The complement of $\mathcal{R}_t$, denoted by $\mathcal{C}_t$ will be referred to as the continuation region. When the VA fee is paid regardless of the account value ($\beta = \infty$), the surrender region at time t, if it exists, is of threshold type, i.e., $\mathcal{R}_t = \{F_t^{(\infty)} \geqslant B_t\}$, with or without surrender penalties (see Chapter 2). B_t represents the fund threshold which induces a rational policyholder to surrender his VA contract at time t. Section 4.4 analyzes the surrender region for the case of a state-dependent fee and shows that it is not necessarily of threshold-type.

Finally, we can define the price of a VA contract with GMAB and surrender option as

$$V(t, F_t^{(\beta)}) = \begin{cases} V^*(t, F_t^{(\beta)}), & \text{if } F_t^{(\beta)} \in \mathcal{C}_t, \\ \psi(t, F_t^{(\beta)}), & \text{if } F_t^{(\beta)} \in \mathcal{R}_t. \end{cases}$$

Since we are working in the Black-Scholes framework, under the usual no-arbitrage assumptions $V(t, F_t^{(\beta)})$ must satisfy the partial differential equation (PDE) in the continuation region $\mathcal{C}_t$,

$$\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial {F^{(\beta)}}_t^2}{F^{(\beta)}}_t^2\sigma^2 + \frac{\partial V}{\partial F_t^{(\beta)}}F_t^{(\beta)}(r - c\mathbb{1}_{\{F_t^{(\beta)} < \beta\}}) - rV = 0, \tag{4.1}$$

for $0 \leqslant t \leqslant T$ (and $F_t^{(\beta)} \in \mathcal{C}_t$). In the optimal surrender region $\mathcal{R}_t$, we have

$$V(t, F_t^{(\beta)}) = \psi(t, F_t^{(\beta)}), \tag{4.2}$$

for $0 \leqslant t \leqslant T$ (and $F_t^{(\beta)} \in \mathcal{R}_t$).[2] The derivation of (4.1) is given in Appendix 4.A. The solution

[2]For more details on this characterization, see for example Björk (2004).

also needs to satisfy the following boundary conditions:

$$V(T, F_T^{(\beta)}) = \max(G, F_T^{(\beta)}),$$
$$\lim_{F_t^{(\beta)} \to 0} V(t, F_t^{(\beta)}) = V(t, 0) = Ge^{-r(T-t)}.$$

The first boundary condition reflects the payoff of the VA at maturity. The second condition comes from the fact that when the account value approaches 0, only the maturity guarantee is valuable. To solve the PDE in (4.1), we also need to specify an upper boundary. However, the behaviour of the contract price for high account values is different whether the fee is constant or state-dependent, and it is generally not possible to specify this boundary exactly for a finite value of F_t.

In the constant fee case, when the optimal strategy is to lapse whenever the account value is above a certain threshold, we can specify an exact upper boundary because the price of the contract corresponds to the surrender benefit when the fund value is sufficiently high. Such a situation can occur if surrender charges are low enough. However, Section 4.4.2 shows that there exists a minimal surrender charge function such that the optimal strategy is to never surrender the contract until maturity. When this happens, the asymptotic behaviour of the contract price is the same as if only the maturity benefit was considered, so that (see Section 4.4.2)[3]

$$\lim_{F_t^{(\infty)} \to \infty} \frac{V(t, F_t^{(\infty)})}{F_t^{(\infty)}} = e^{-c(T-t)}.$$

Of course, there is no need to use numerical methods to solve the PDE in this case, because in the absence of surrenders, the contract price has a closed-form expression based on the Black-Scholes formula.

With a state-dependent fee ($\beta < \infty$), the following asymptotic behaviour holds regardless of the form assumed for the surrender charge function:

$$\lim_{F_t^{(\beta)} \to \infty} \frac{V(t, F_t^{(\beta)})}{F_t^{(\beta)}} = 1. \tag{4.3}$$

A proof of this assertion is given in Appendix 4.B. This result allows us to use $F_t^{(\beta)}$ as an upper

[3]Intuitively, this is due to the fact that the maturity benefit is worth close to nothing at very high fund values $F_t^{(\infty)}$, which implies that we can write:

$$E_{\mathbb{Q}}[e^{-r(T-t)} \max(G, F_T^{(\infty)}) | \mathcal{F}_t] \approx E_{\mathbb{Q}}[e^{-r(T-t)} F_T^{(\infty)} | \mathcal{F}_t] = F_t^{(\infty)} e^{-c(T-t)}.$$

boundary for $V(t, F_t^{(\beta)})$ when solving the PDE in (4.1) numerically. Intuitively, this limiting behaviour stems from the fact that when the account value is very high, the maturity benefit is worth close to nothing, and the policyholder does not expect to pay any more fees. Thus, the value of the contract can be estimated by

$$E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}|\mathcal{F}_t] \approx E_{\mathbb{Q}}[e^{-r(T-t)}F_t^{(\beta)}\frac{S_T}{S_t}|\mathcal{F}_t] = F_t^{(\beta)}, \quad \text{when } F_t^{(\beta)} \gg \beta.$$

4.3 Numerical Examples

In this section, we solve the PDE presented in equation (4.1) of Section 4.2 under different fee structures, and observe the impact of the fee structure on the shape of the optimal surrender region. Understanding the interplay between the fee structure and the surrender incentive is the first step towards designing a contract that eliminates the surrender incentive while offering reasonable fee rates and surrender charges. Moreover, having a better understanding of the surrender incentive allows the insurer to establish more effective risk-management strategies.

4.3.1 Solving the PDE numerically

To solve (4.1), we use finite difference methods. The equation is first expressed in terms of $x_t = \ln F_t$ and discretized over a rectangular grid representing the truncated, discretized domain of (t, x_t). The upper truncation point of the grid depends on the contract whose price we are solving for. For example, when an optimal surrender boundary exists for all $t \in [0, T]$, it is not necessary to consider values that are above this boundary, because the price of the contract is known exactly in this region (and is equal to the value of the surrender benefit). In more general cases, the grid in the x_t dimension must be large enough that the asymptotic results derived in Section 4.2 can be used reliably to approximate the contract price at the highest fund values in the grid. For small account values at time t, the contract price is well approximated by $Ge^{-r(T-t)}$ and we do not need to include fund values which are very close to zero. The maximum value of the grid, $x^{\max}$, is high enough that it has a very small probability of being reached by the process x_t. When the contract is always optimally surrendered at values below $e^{x^{\max}}$, we use a lower maximal value to decrease computational time. We use an explicit method with space steps dx and time steps $dt = \frac{(dx/\sigma)^2}{3}$ to ensure stability of our numerical scheme (see, for example, Racicot and Théoret (2006)). Implicit methods were also explored for validation purposes and to examine stability, but the explicit scheme was selected and was implemented in C++ to improve the speed of calculation. Central differences were used to approximate the first order term. Again, other methods were explored. In particular, we also used forward differences to make sure that all the coefficients were positive (for more details, see Chapter 9 of Duffy (2006)), but the precision of the results obtained using central differences was very similar.

4.3.2 Numerical Results

In this section, we consider a 10-year VA contract guaranteeing an amount of $G = F_0^{(\beta)} = 100$ at maturity (in other words, the guaranteed roll-up rate of the GMAB is $g = 0$). The market parameters were fitted to a data set of weekly percentage log-returns on the S&P500 from October 28, 1987 to October 31, 2012, from which we obtained the parameters $\mu = 0.07$ and $\sigma = 0.165$. We further assume $r = 0.03$.

The grid that we use spans from $\ln 20$ to $\ln 400$. Note that $E_{\mathbb{P}}[S_T|S_0 = 100] = 201.38$ and $\sqrt{Var_{\mathbb{P}}[S_T|S_0 = 100]} = 112.65$, so our grid covers the most likely paths of $F_t^{(\beta)}$ since $F_t^{(\beta)} \leqslant S_t$ for any $\beta \geqslant 0$ and $t \in [0, T]$. The space steps we use have length $dx = 0.0005$.

Fair fee

We define the fair fee for a variable annuity contract as the fee rate c^* satisfying

$$F_0^{(\beta)} = V(0, F_0^{(\beta)}; c^*), \tag{4.4}$$

where $V(0, F_0^{(\beta)}; c^*)$ is the price of the contract calculated using the fee c^*. Table 6.1 presents the fair fees obtained for different payment barrier loadings λ and surrender charge functions κ_t, assuming either optimal policyholder behaviour or that the contract is held to maturity. In the latter case, the price of the contract is simply the expectation of the discounted maturity benefit. In Chapter 3, we derive an integral formula for this price, which can also be obtained by solving an equation similar to (4.1).

We consider values of $\lambda = 0.2, 0.5$ and ∞, corresponding to threshold levels of $\beta = 120$, 150, and ∞, where $\beta = (1 + \lambda)G$. $\lambda = \infty$ corresponds to the case where the fee is paid continuously, regardless of the value of the account ($\beta = \infty$). Note that Chapter 3 studied the case where $\lambda = 0$ without any surrender charges. This design leads to a very high fair fee which may not always be marketable. In this section, we focus on more realistic contract designs, and also incorporate surrender charges. We consider two decreasing surrender charge functions, in addition to the case $\kappa_t = 0$. First, we use the function, $\kappa_t = 1 - e^{-\kappa(T-t)}$, studied in Chapter 2. Second, we consider a "vanishing" surrender charge function, $\kappa_t = \kappa(1 - t/T)^3$. This function mimics surrender penalties found on the market, which are typically high in the first years of the contract, and drop rapidly to make the VA a more liquid investment.

The resulting fair fees are presented in Table 6.1. In general, the fair fee decreases when the payment barrier loading λ increases, because the fee is expected to be paid for a longer period of time. The fair fee also decreases with increasing surrender charges, as these charges discourage lapsation and represent an additional revenue for the insurer. Further analysis of the fair fee is conducted in the next part of this section to discuss the optimal surrender region under constant and state-dependent fees.

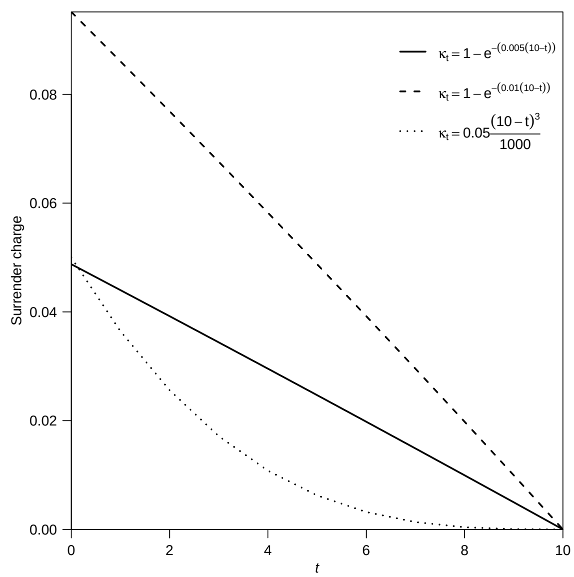


Figure 4.1: Evolution of surrender charge κ_t from $t = 0$ to $t = 10$ for different surrender charge functions.

λ	β	No Surrender	Optimal Surrender			
			$\kappa_t = 0$	$\kappa_t = 1 - e^{-0.005(T-t)}$	$\kappa_t = 1 - e^{-0.01(T-t)}$	$\kappa_t = 0.05(1 - t/T)^3$
0.2	120	0.02359	0.03473	0.02364	0.02361	0.02371
0.5	150	0.01550	0.03473	0.01585	0.01557	0.01763
∞	∞	0.01062	0.03473	0.01394	0.01075	0.01697

Table 4.1: Fair fee for different VA contracts with $T = 10$, $r = 0.03$, and $\sigma = 0.165$.

Optimal surrender region

No surrender charge ($\kappa_t = 0$)

When there is no penalty on early surrender, the analysis of the optimal surrender region presented in this section reveals that the state-dependent fee structure may not be sufficient to decrease the surrender incentive. This is first hinted at in Table 4.1, where the fair fee was found to be the same for all three values of λ studied when $\kappa_t = 0$. Such a result suggests that a policyholder behaving optimally is not able to profit from the state-dependent fee. For example, it could be rational to lapse before reaching the fee barrier threshold. Figure 4.2 shows that this is exactly what occurs, i.e., the optimal surrender boundaries for all three values of λ are identical and lie below 120. Therefore, a policyholder behaving optimally will never reach the fee barrier threshold of $\beta = 120$ or 150, and from his perspective, product designs with $\beta = 120$, 150, or ∞ are equivalent. This explains why the fair fee remains the same for these three designs.

In the absence of surrender charges, the minimal account value at which the policyholder should lapse the contract just after inception is equal to the initial premium ($F_0^{(\beta)} = 100$), when the fair fee is calculated assuming optimal surrenders. In fact, the fair fee is calculated such that $V(0, F_0^{(\beta)}) = F_0^{(\beta)}$, which coincides with the definition of the optimal surrender boundary at time 0 (assuming no surrender charges). This is clearly illustrated in Figure 4.2, where the optimal surrender boundary crosses 100 at $t = 0$. It is this phenomenon that Milevsky and Salisbury (2001) refer to when they argue that without surrender charges, the VA contract is not marketable. Next, we explore such charges as a way to decrease the surrender incentive.

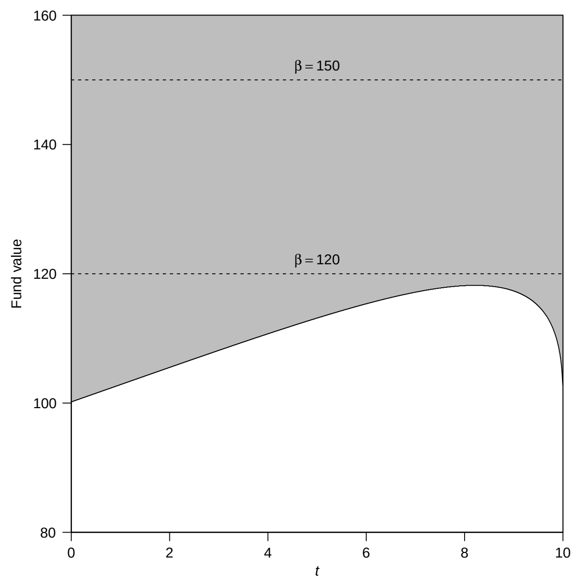


Figure 4.2: Optimal surrender region for $\lambda = 0.2$, 0.5 or ∞ and $\kappa_t = 0$, priced assuming optimal surrenders ($c = 0.03473$).

Adding surrender charges

The addition of surrender charges to a VA policy with any kind of fee structure (constant or state-dependent) reduces the incentive to lapse because it decreases the surrender benefit. We first revisit the surrender charge function, $\kappa_t = 1 - e^{\kappa(T-t)}$, considered in Section 4.3.2. This results in a surrender penalty starting approximately at 10κ at time 0 and decreasing almost linearly to 0 at time 10. Figure 4.3 presents the optimal surrender regions for $\kappa = 0.005$ and 0.01, when $\lambda = 0.5$. Interestingly, it is only optimal to lapse the contract closer to maturity, and we

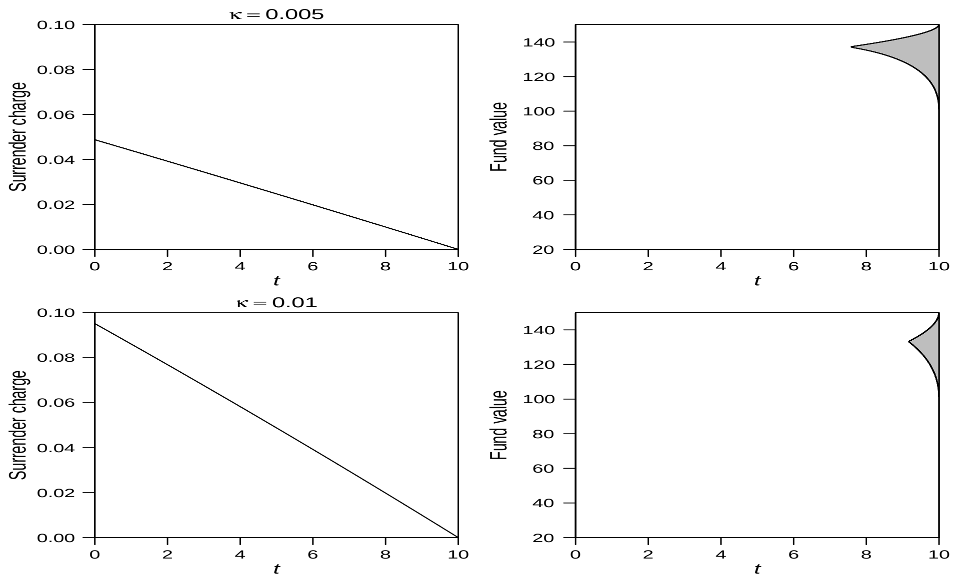


Figure 4.3: *Left column*: Surrender charge as a function of time when $\kappa_t = 1 - e^{-\kappa(10-t)}$. *Right column*: Optimal surrender region for $\lambda = 0.5$, fairly priced assuming optimal surrenders (see Table 4.1.

observe that the size of the optimal surrender region diminishes when the surrender charge rate κ increases. In other words, the presence of surrender penalties encourages the policyholder to wait for the account value to grow above the fee barrier threshold.

Figure 4.4 shows the optimal surrender region when $\kappa_t = \kappa(1 - t/T)^3$. Compared to the previous case, this function leads to lower penalties for almost all $t \in [0, T)$, even when the parameters are chosen such that the charge at $t = 0$ is the same under both functions. For this reason, the surrender incentive is higher for the "vanishing" surrender charge. This is also reflected in the fair fee, which is higher for all three levels of the payment barrier considered (see Table 4.1). Nonetheless, Figure 4.4 shows that by combining a state-dependent fee with surrender charges, we are again able to eliminate the surrender incentive at the beginning of the contract, and when the account value is close to and above the payment barrier. In the next section, we will formalize this result theoretically and demonstrate that the introduction of surrender charges eliminates the surrender incentive above the payment barrier. We will also show how to obtain the minimal surrender charge function which does not give rise to a surrender incentive during the whole length of the contract.

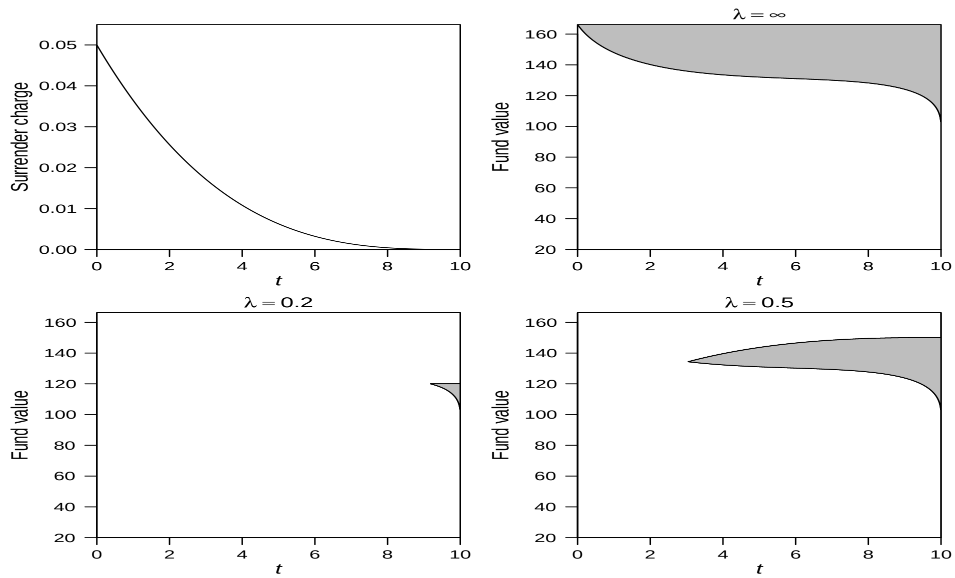


Figure 4.4: *Upper left*: Surrender charge as a function of time when $\kappa_t = 0.05(1 - t/10)^3$. *Upper right and bottom row:* Optimal surrender region when $\lambda = 0.2,\ 0.5,\ \infty$, fairly priced assuming optimal surrenders (see Table 4.1).

4.4 Theoretical analysis of the surrender incentive

In the previous section, we analysed the optimal surrender region resulting from different product designs. In particular, we showed that a combination of surrender charges and state-dependent fees can eliminate the surrender incentive for high account values. In this section, we formalize some of the results that were observed in Section 4.3. These results are then used in Section 4.5 to design a contract that completely eliminates the surrender incentive.

4.4.1 Surrender incentive for large account values when $\beta < \infty$

We first show that when the fee is state-dependent (that is, when $\beta < \infty$) and the account value is above the fee barrier, the contract is always worth at least as much as the surrender benefit. Thus, the policyholder never has a clear incentive to lapse when the account value is above the payment barrier. This result is formalized in the following proposition.

Proposition 4.4.1. *Let $F_t^{(\beta)}$ and κ_t be defined as in Section 4.2 and let $\beta < \infty$. Then, for any $t \in [0, T]$ and $F_t^{(\beta)} > \beta$,*

$$V^*(t, F_t^{(\beta)}) \geqslant F_t^{(\beta)}. \tag{4.5}$$

If $\kappa_t > 0$ at time t, the inequality in (4.5) is strict.

Proof. Suppose that there are no surrender charges, i.e., $\kappa_t = 0$, $\forall t$, that the fee is only paid below β, and that we are at time t with $F_t^{(\beta)} \geqslant \beta$. Consider the stopping time,

$$\tau_\beta = \inf\left\{t < u < T : F_u^{(\beta)} < \beta\right\},$$

with the convention that $\tau_\beta = T$, if the barrier β is never reached. Then, we can write,

$$\begin{aligned}
V^*(t, F_t^{(\beta)}) &= \sup_{\tau \in \mathcal{T}_t} E_{\mathbb{Q}}[e^{-r(\tau-t)}\psi(\tau, F_\tau^{(\beta)})|\mathcal{F}_t] \\
&\geqslant E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}\psi(\tau_\beta, F_{\tau_\beta}^{(\beta)})|\mathcal{F}_t] \\
&= E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}F_{\tau_\beta}^{(\beta)}\mathbb{1}_{\{\tau_\beta\in(t,T)\}}|\mathcal{F}_t] + E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}F_{\tau_\beta}^{(\beta)}\mathbb{1}_{\{\tau_\beta=T\}}|\mathcal{F}_t] \\
&= \beta E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}\mathbb{1}_{\{\tau_\beta\in(t,T)\}}|\mathcal{F}_t] + E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{\tau_\beta=T\}}|\mathcal{F}_t],
\end{aligned}$$

where the first term is the payoff of a down rebate option which pays β if the fund $F_t^{(\beta)}$ reaches β before maturity T and zero otherwise, and the second term is the payoff of a down-and-out European call option with zero strike which pays $F_T^{(\beta)}$ at maturity T, provided that $F_u^{(\beta)} \geqslant \beta$, for $t \leqslant u \leqslant T$. Since the combined payoff of these two options can be replicated by holding the fund $F_t^{(\beta)}$ and selling it as soon as $F_t^{(\beta)} = \beta$, simple no-arbitrage arguments imply that the total price of these two options is exactly $F_t^{(\beta)}$, which gives,

$$V^*(t, F_t^{(\beta)}) \geqslant F_t^{(\beta)}.$$

Now consider the case where the surrender charge function κ_u, $t \leqslant u \leqslant T$, is a decreasing function of u, and is strictly positive at t:

$$\begin{aligned}
V^*(t, F_t^{(\beta)}) &= \sup_{\tau \in \mathcal{T}_t} E_{\mathbb{Q}}[e^{-r(\tau-t)}\psi(\tau, F_\tau^{(\beta)})|\mathcal{F}_t] \\
&\geqslant E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}\psi(\tau_\beta, F_{\tau_\beta}^{(\beta)})|\mathcal{F}_t] \\
&= E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}F_{\tau_\beta}^{(\beta)}(1-\kappa_{\tau_\beta})\mathbb{1}_{\{\tau_\beta\in(t,T)\}}|\mathcal{F}_t] + E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}F_{\tau_\beta}^{(\beta)}\mathbb{1}_{\{\tau_\beta=T\}}|\mathcal{F}_t] \\
&= \beta E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}(1-\kappa_{\tau_\beta})\mathbb{1}_{\{\tau_\beta\in(t,T)\}}|\mathcal{F}_t] + E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{\tau_\beta=T\}}|\mathcal{F}_t] \\
&> (1-\kappa_t)\left\{\beta E_{\mathbb{Q}}[e^{-r(\tau_\beta-t)}\mathbb{1}_{\{\tau_\beta\in(t,T)\}}|\mathcal{F}_t] + E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{\tau_\beta=T\}}|\mathcal{F}_t]\right\} \qquad (4.6) \\
&= (1-\kappa_t)F_t^{(\beta)},
\end{aligned}$$

as the term inside the braces in equation (4.6) was shown to be exactly $F_t^{(\beta)}$. This result implies that in the presence of surrender charges, it is never optimal to surrender the variable annuity contract when the fund is above or equal to the fee threshold barrier β. □

Because of the way the optimal surrender region was defined in Section 4.2.2 (the policyholder is assumed to lapse when the contract is worth at least as much as the surrender benefit), Proposition 4.4.1 does not necessarily characterize the region above the payment barrier. For example, in Figure 4.2, it is optimal to surrender the contract when $F_t^{(\beta)} > \beta$, while Figures 4.3 and 4.4 show continuation regions above β. Note that for the state-dependent fee designs studied in Figure 4.2, the policyholder is actually indifferent to lapse, but is assumed to surrender because of the way we defined the surrender region. This is because the contract value and the surrender benefit are equal above β for this specific case.

4.4.2 Minimal surrender charge to eliminate the surrender incentive

In the remainder of this section, we seek to design a contract that does not have an optimal surrender region. In other words, the optimal behaviour when buying such a contract is to hold it until maturity. To design this policy, we look for a sufficiently large surrender charge such that $\psi(t, F_t^{(\beta)}) - V^*(t, F_t^{(\beta)}) \leqslant 0$, for any account value $F_t^{(\beta)}$, and any time $t \in [0, T)$. We first study the constant fee case, where a general result can be obtained in closed-form. We then discuss contracts with state-dependent fees.

Constant fee ($\beta = \infty$)

When a constant fee is charged, and there are no surrender charges, there exists an optimal lapsation boundary above which a rational investor should not hold on to the VA contract. Assuming once again that $F_0^{(\infty)} = G = 100$, $T = 10$, $r = 0.03$, and $\sigma = 0.165$, Figure 4.5 illustrates three such boundaries[4], each of which is associated with a given level of c (displayed on the curve). Since a lower fee c lessens the incentive to surrender, this boundary shifts upwards when c decreases.

When the insurer does not charge a penalty for early surrender, the fee income represents his only revenue. This income compensates him for both the guarantee offered and early surrender risk (the risk of not being able to collect future fees on the account value). To fully mitigate lapse risk, he must charge a fee for which the value of the guarantee offered assuming optimal policyholder behaviour equals the amount invested. Under the assumptions stated above, this fair fee corresponds to 3.50%, which is very high. In fact, if the insurer were certain that the policyholder would keep her variable annuity until maturity, then he would be able to offer her

[4] These boundaries can be obtained using the method described in Chapter 2, or by solving the PDE in equation (4.1) numerically. See Section 4.3 for more details.

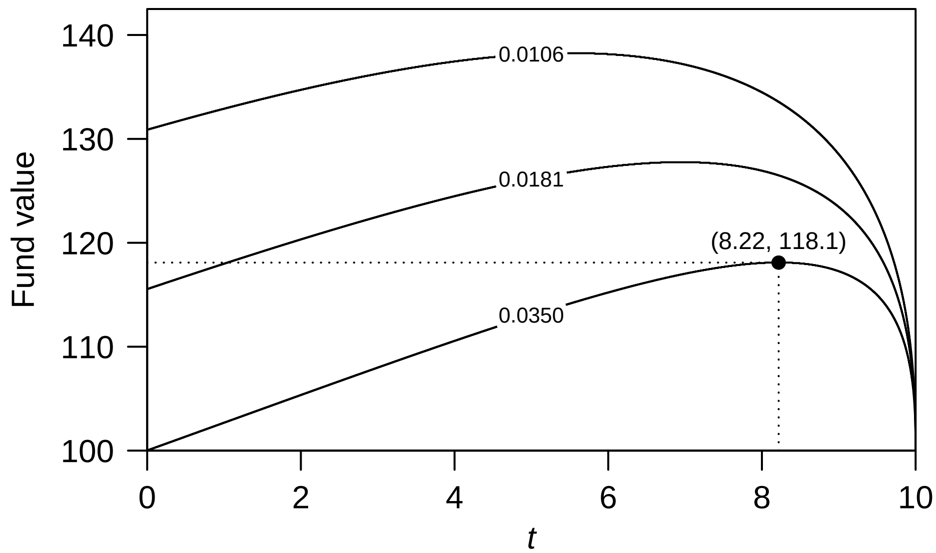


Figure 4.5: Optimal surrender boundary when there are no surrender charges. Each of the three curves is associated with a fee c. The fair value of c under optimal behaviour is 3.5%. The fair value of c when the policyholder is not allowed to lapse is 1.06%. The fair value of c assuming the policyholder lapses suboptimally as soon as the fund reaches 150 is 1.81%.

a fee of only 1.06%. One way to decrease the fair fee while still fully mitigating lapse risk is to introduce surrender penalties in the product design. These penalties represent an additional income for the insurer, enabling him to reduce the constant fee charge, and also work as a disincentive to lapse. Proposition 4.4.2 states the minimal value of κ_u at each time $u \in [t, T)$, so that it is never optimal for the policyholder to surrender her contract from time t until maturity.

Proposition 4.4.2. *Let $F_t^{(\beta)}$ and κ_t be defined as in Section 4.2 and let $\beta = \infty$. If the insurer wants to charge the minimal value of κ_u at each time $u \in [t, T)$, so that an optimal lapsation boundary from $u = t$ to T does not exist, then he must set:*

$$\kappa_u = 1 - e^{-c(T-t)}, \quad t \leqslant u < T.$$

Proof. First, suppose that the surrender charge κ_u, for $t \leqslant u < T$, is sufficiently high for there to be no optimal lapsation boundary for $t \leqslant u < T$. This situation is possible because we can consider the extreme case where $\kappa_u = 1$, for $t \leqslant u < T$. Then, the value of the contract at time u must simply be the risk-neutral discounted expectation of the payoff at maturity, and be greater or equal to the surrender benefit:

$$V^*(u, F_u^{(\beta)}) = E_{\mathbb{Q}}[e^{-r(T-u)}\max(F_T^{(\beta)}, G)|\mathcal{F}_u] \geqslant F_u^{(\beta)}(1-\kappa_u), \quad \forall F_u^{(\beta)} \geqslant 0.$$

The previous inequality can be rewritten as:

$$\kappa_u \geqslant 1 - \frac{E_{\mathbb{Q}}[e^{-r(T-u)}\max(F_T^{(\beta)}, G)|\mathcal{F}_u]}{F_u^{(\beta)}}, \quad \forall F_u^{(\beta)} \geqslant 0.$$

Therefore, the minimal surrender penalty that can be charged at time u while the inequality above is satisfied corresponds to:

$$\kappa_u^\star = \max\left(1 - \inf_{F_u^{(\beta)} \geqslant 0}\left\{\frac{E_{\mathbb{Q}}[e^{-r(T-u)}\max(F_T^{(\beta)}, G)|\mathcal{F}_u]}{F_u^{(\beta)}}\right\}, 0\right).$$

Since

$$\begin{aligned}\frac{E_{\mathbb{Q}}[e^{-r(T-u)}\max(F_T^{(\beta)}, G)|\mathcal{F}_u]}{F_u^{(\beta)}} &= \frac{F_u^{(\beta)}e^{-c(T-u)} + E_{\mathbb{Q}}[e^{-r(T-u)}\max(G - F_T^{(\beta)}, 0)|\mathcal{F}_u]}{F_u^{(\beta)}} \\ &> e^{-c(T-u)}, \quad \forall F_u \geqslant 0,\end{aligned} \tag{4.7}$$

and,

$$\lim_{F_u^{(\beta)} \to \infty} \frac{E_{\mathbb{Q}}[e^{-r(T-u)}\max(F_T^{(\beta)}, G)|\mathcal{F}_u]}{F_u^{(\beta)}} = e^{-c(T-u)},$$

then, we must have,

$$\kappa_u^\star = 1 - e^{-c(T-u)}, \quad t \leqslant u < T. \tag{4.8}$$

□

Remark 4.4.1. *Note that the result given by Proposition 4.4.2 is very general because it is essentially model-free, and holds for any arbitrage-free complete market model, and not just for the Black-Scholes model. It shows that if the surrender charge function is chosen according to (4.8), for $t \leqslant u < T$, then it will not be optimal to surrender the contract. However, the condition $\kappa_t \geqslant \kappa_t^\star$ is also sufficient to guarantee that it is not optimal to lapse at time t, regardless of the form assumed for κ_u, $t < u < T$. To see why, observe that,*

$$\frac{V^*(t, F_t^{(\beta)})}{F_t^{(\beta)}(1-\kappa_t)} \geqslant \frac{E_{\mathbb{Q}}[e^{-r(T-t)}\max(F_T^{(\beta)}, G)|\mathcal{F}_t]}{F_t^{(\beta)}(1-\kappa_t)} > \frac{e^{-c(T-t)}}{1-\kappa_t}, \quad \forall F_t^{(\beta)} \geqslant 0.$$

It is clear that whenever $\kappa_t \geqslant \kappa_t^\star$, the continuation value, $V^*(t, F_t^{(\beta)})$, must be greater than the surrender value, $F_t^{(\beta)}(1-\kappa_t)$, for any values of $F_t^{(\beta)}$ at time t, making surrender sub-optimal.

Note that the converse of this result does not necessarily hold, i.e., the condition $\kappa_t \geqslant \kappa_t^\star$ is not necessary when surrender is not optimal at time t, $\forall F_t^{(\beta)} \geqslant 0$. In other words, there may be a value of $\kappa_t \in (0, \kappa_t^\star)$ which makes lapsation not optimal at time t. In fact, $\kappa_t^\star$ is a strict lower bound for the surrender charge at time t if and only if it is never optimal to surrender the contract after t.

State-dependent fee ($\beta < \infty$)

We have seen in the previous section that to fully mitigate lapse risk in the presence of a constant fee, the insurer can use various product designs, the two extremes being: (I) $c = 3.5\%$ and no surrender charges, and (II) $c = 1.06\%$ and a schedule of surrender charges corresponding to $\kappa_t = 1 - e^{-0.0106(T-t)}$. Both of these designs may be difficult to market in practice, because the first has a high fee and the second heavily penalizes the policyholder in the event of lapse. Therefore, we may opt for a product design such as the one with $\kappa = 0.005$ in Figure 4.6. However, this design may be difficult to hedge because it gives rise to an optimal lapsation boundary, and the insurer must therefore hedge an American option to fully mitigate lapse risk. Note that for scenario (II), when $c = \kappa = 1.06\%$, the insurer is required to hedge a European option as this product design makes surrender sub-optimal at any given time. We therefore ask the following question: Can we design a product that incorporates a reasonable fee and surrender charges, and that gives no incentive to lapse for a rational policyholder? If we consider a state-dependent fee, the answer to this question is yes. Figure 4.7 illustrates the optimal lapsation boundary when the fee barrier threshold is $\beta = 110$ (left) or $\beta > 118.1$ (right), and there are no surrender charges.

First, note that when $\beta > 118.1$, the optimal lapsation boundary is always under β, and, therefore, a rational policyholder will always surrender her product before reaching the no-fee region. If the insurer wants to fully mitigate lapse risk, then for pricing and hedging purposes this design is equivalent to having a constant fee and no surrender charges ($c = 3.5\%$, see Figure 4.5). Nevertheless, in practice, the state-dependent fee design with $\beta > 118.1$ may give a false incentive for sub-optimal policyholders to hold on to the product. For this reason, it may be argued that the state-dependent fee design is preferable. When $\beta = 110$, the state-dependent fee design does not give rise to an optimal lapsation boundary between $t = 2.0$ to 9.9, and is therefore able to make lapse sub-optimal in that range with only a slight fee increase of 0.08% with respect to the design with a constant fee. However, this fee is too high to be marketable in practice, and we are still confronted with an American option for risk management purposes. This suggests that combining a state-dependent fee with surrender charges can lead to a very interesting design.

In the state-dependent fee case, obtaining a simple closed-form expression for the minimal surrender charge that eliminates the surrender incentive is not possible, for different reasons. First, although there exists an integral representation for the value of the maturity benefit (see, Chapter 3), or for the discounted expectation of the account value at future times, they are generally complex and depend on the current account value $F_t^{(\beta)}$ in more than one way. In addition, the

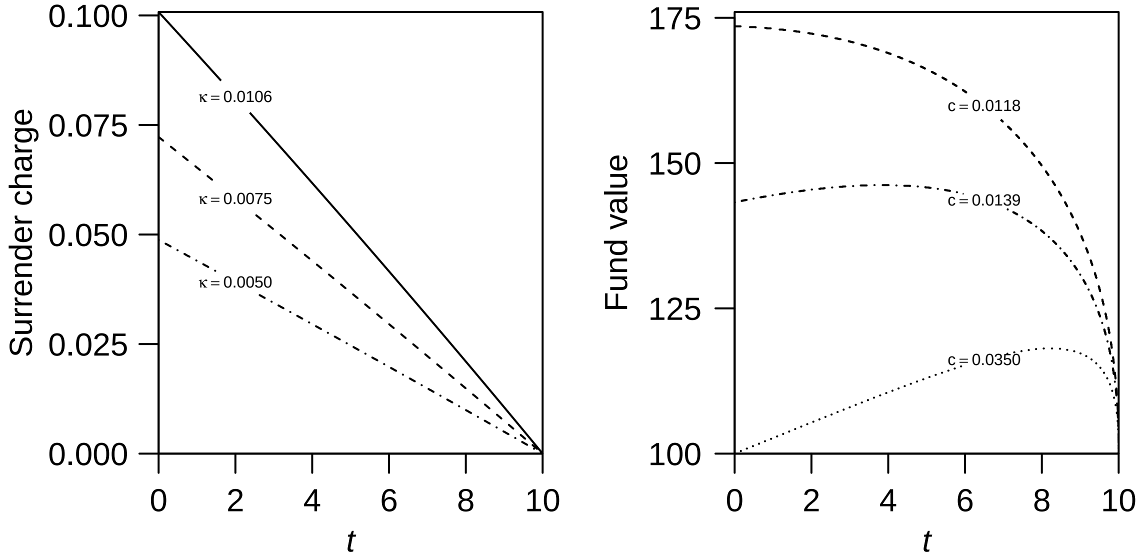


Figure 4.6: *Left:* Surrender charges when $\kappa_t = 1 - e^{-\kappa(10-t)}$ for $\kappa = 0.005,\ 0.0075, 0.0106$. *Right:* Optimal surrender boundaries for a constant fair fee associated with the surrender charge functions presented on the left.

function $E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(\beta)}, G)|\mathcal{F}_t]/F_t^{(\beta)}$ is generally not monotone in F_t because the expected future fees are not monotone in F_t, due to their state-dependent nature. Nevertheless, it is still possible to solve for the minimal surrender charge that eliminates the surrender incentive, but we must do so numerically. The procedure is outlined in the following steps:

1. Find the fair fee assuming that the contract is always held until maturity. This can be done numerically, for example by using finite differences, or by using the formula given in Chapter 3.

2. For each $t \in [0, T)$, numerically obtain the account value $F_t^\star$ at which the ratio of the maturity benefit to the fund value is the smallest:

$$F_t^\star = \arg\inf_{F_t^{(\beta)} \geqslant 0} \left\{ \frac{E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(\beta)}, G)|\mathcal{F}_t]}{F_t^{(\beta)}} \right\}.$$

   Note that in this step, we can assume $V^*(t, F_t^{(\beta)}) = E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(\beta)}, G)|\mathcal{F}_t]$, since this must hold $\forall t$ when κ_t is set so that surrenders are never optimal during the entire length of the contract.

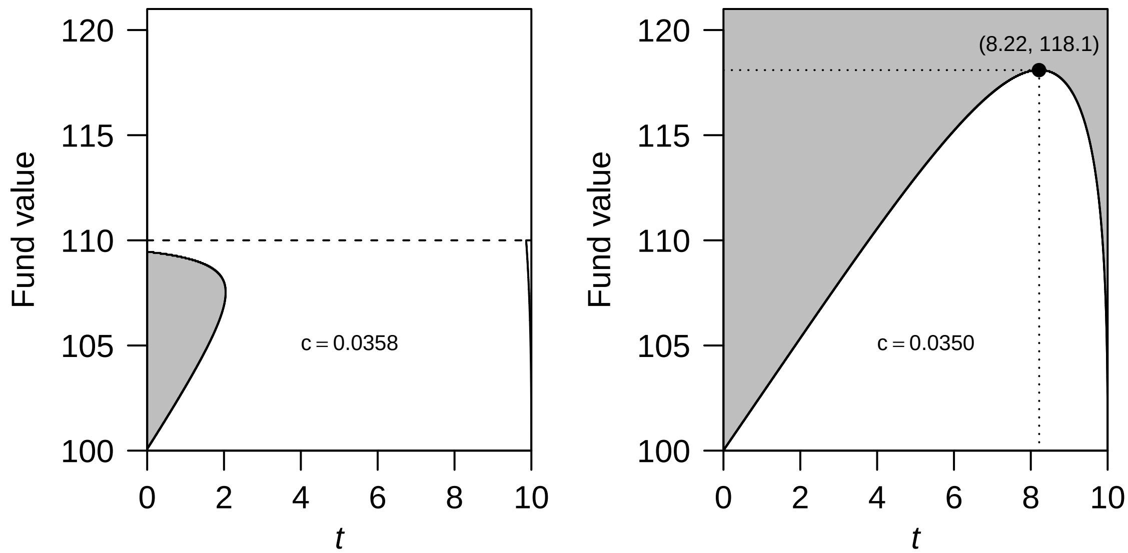


Figure 4.7: *Left:* Optimal surrender boundary for a state-dependent fee paid only under $\beta = 110$, and assuming fair pricing ($c = 3.58\%$) and no surrender charges. *Right:* Optimal surrender boundary for a state-dependent fee paid only under any choice of threshold $\beta > 118.1$, and assuming fair pricing ($c = 3.50\%$) and no surrender charges.

3. Set

$$\kappa_t^\star = \max\left(1 - \frac{E_{\mathbb{Q}}[e^{-r(T-t)}\max(F_T^{(\beta)}, G)|F_t^{(\beta)} = F_t^\star]}{F_t^\star}, 0\right),$$

where $\kappa_t^\star$ is the minimal surrender charge eliminating the surrender incentive.

As an example, we revisit a design presented in Section 4.3. We assume $\beta = 150$ and $F_0^{(150)} = G = 100$ because this design results in a realistic fee rate. The market parameters are the same as in Section 4.3. Using the procedure described above, we obtain the surrender charge function illustrated on the left of Figure 4.8. In this case, the fair fee is $c = 1.55\%$. The values $F_t^\star$ used in the calculation of the surrender charge are given on the right. Observe that exactly on this boundary, we have $V(t, F_t^\star) = (1-\kappa_t)F_t^\star$, and that surrender is never optimal on either side of the boundary. Then, with this surrender charge function, the policyholder never has a clear incentive to surrender (that is, the contract value is always worth at least as much as the surrender benefit). Another way to see this is to say that whenever the policyholder lapses, she does so sub-optimally and this brings additional revenue to the insurer.

The surrender charge structure obtained is satisfying because it is not too high, starting below 3.5% and decreasing to 0 at maturity. This is significantly lower than the minimal surrender charge required to eliminate the surrender incentive when $\beta = \infty$. As explained in Section 4.4.2, such a penalty would start at over 10% at inception and would not drop below 5% until the fifth year of the contract. Thus, the state-dependent fee allows us to eliminate optimal surrender incentives using much lower surrender charges.

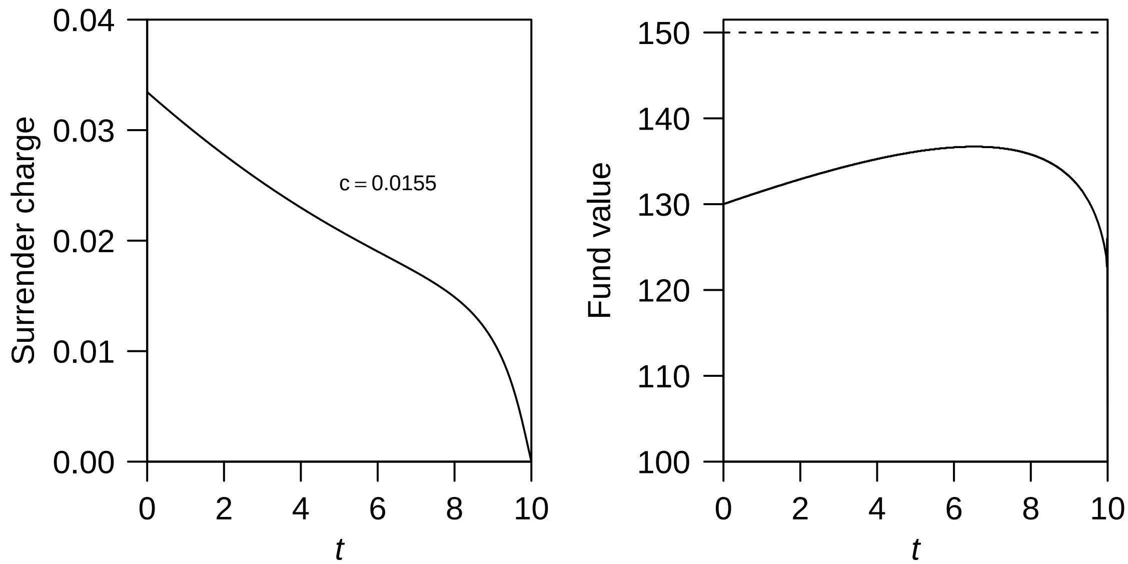


Figure 4.8: *Left*: Minimal surrender charge function not giving rise to an optimal surrender boundary when $c = 0.0155$ and $\beta = 150$. *Right*: Values of $F_t^{(150)}$ at which the infima of the function $E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(150)}, G)|\mathcal{F}_t]/F_t^{(150)}$ were computed.

4.5 Dynamic hedging

This section illustrates why eliminating the surrender incentive in the VA product design can simplify the insurer's hedging strategy and make it more effective. Before presenting our results on dynamic hedging, we review some concepts with respect to hedging VAs, and explain how we calculate the insurer's hedged loss.

4.5.1 Calculation of the net hedged loss at maturity

Assume that we have a path of stock values, $\{S_t\}_{0 \leqslant t \leqslant T}$, and corresponding account values, $\{F_t^{(\beta)}\}_{0 \leqslant t \leqslant T}$, sampled at discrete time intervals h, where, for example, $h = 1/52$ implies weekly observations. We define the net hedged loss at maturity as $L - H$, where,

$$\begin{aligned}L &= \text{Net unhedged loss at maturity,}\\ H &= \text{Cumulative mark-to-market gain on the hedge.}\end{aligned}$$

When the insurer does not use a hedging strategy, his net loss at maturity is L. When he employs a hedging strategy, his net loss is $L - H$. The losses are *net* because they take into account the fee income and surrender charges received by the insurer.

If the policyholder does not surrender her contract, the net unhedged loss at maturity T is

$$\begin{aligned}L &= \text{payoff to the policyholder} \; - \; \text{accumulated value of fees}\\ &= \max(0, G - F_T^{(\beta)}) - \sum_{i=0}^{T/h-1} F_{ih}^{(\beta)}(1 - e^{-ch})e^{r(T-ih)}\mathbb{1}_{\{F_{ih}^{(\beta)} < \beta\}}.\end{aligned}$$

In the event of surrender at time $t = \tau$, the net unhedged loss at maturity T is

$$\begin{aligned}L &= -(\text{accumulated value of fees and surrender charges})\\ &= -\sum_{i=0}^{\tau/h-1} F_{ih}^{(\beta)}(1 - e^{-ch})e^{r(T-ih)}\mathbb{1}_{\{F_{ih}^{(\beta)} < \beta\}} - F_\tau^{(\beta)}\kappa_\tau e^{r(T-\tau)}.\end{aligned}$$

To calculate the net hedged loss at maturity, the cumulative mark-to-market gain on the hedge must be subtracted from the net unhedged loss. Assuming that the hedging strategy consists of a delta hedge, the mark-to-market gain at time $t + h$ of the hedge established at time t is

$$\Delta_t(S_{t+h} - S_t e^{rh}),$$

where Δ_t is the *delta* used in the hedge, and will be defined in Section 4.5.2. The cumulative mark-to-market gain on the hedge corresponds to the accumulated value of these gains to maturity:

$$H = \sum_{i=0}^{\tau/h-1} \Delta_{ih}(S_{(i+1)h} - S_{ih}e^{rh})e^{r(T-(i+1)h)},$$

where τ represents the time at which the hedging strategy is stopped (surrender or maturity). Finally, the net hedged loss at maturity is simply $L - H$.

4.5.2 Calculation of Δ_t

When constructing a hedging strategy, the insurer must first specify the assumptions that will be used for hedging, as well as the objective function that needs to be hedged. For example, suppose that $\beta = \infty$, and that the insurer wants to set-up a delta hedge of his net liability assuming no surrenders occur. In this context, the net liability of the insurer towards the policyholder at time t, denoted by Ψ_t, corresponds to the fair value of the maturity benefit minus the account value:

$$\begin{aligned}\Psi_t &= E_{\mathbb{Q}}[e^{-r(T-t)}\max(F_T^{(\infty)}, G)|\mathcal{F}_t] - F_t^{(\infty)} \\ &= E_{\mathbb{Q}}[e^{-r(T-t)}\max(G - F_t^{(\infty)}, 0)|\mathcal{F}_t] + F_t^{(\infty)}e^{-c(T-t)} - F_t^{(\infty)} \\ &= E_{\mathbb{Q}}[e^{-r(T-t)}\max(G - F_t^{(\infty)}, 0)|\mathcal{F}_t] - F_t^{(\infty)}(1 - e^{-c(T-t)}).\end{aligned} \tag{4.9}$$

Equation (4.9) offers an alternative interpretation of the net liability, as the value of the underlying European put option minus the fair value of the fees that will be collected by the insurer until maturity.[5] For this particular case, the delta of the net liability is available in closed form as,

$$\begin{aligned}\frac{\partial}{\partial S_t}\Psi_t &= \frac{\partial \Psi_t}{\partial F_t^{(\infty)}}\frac{\partial F_t^{(\infty)}}{\partial S_t} \\ &= [-e^{-c(T-t)}N(-d_1) - (1 - e^{-c(T-t)})]e^{-ct} \\ &= -e^{-cT}N(-d_1) - (e^{-ct} - e^{-cT}),\end{aligned}$$

where $N(\cdot)$ denotes the standard normal cumulative distribution function, and

$$d_1 = \frac{\log(F_t^{(\infty)}/G) + (r - c + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}.$$

In more general situations, where we want to hedge assuming optimal policyholder behaviour, we cannot obtain the delta analytically. However, we can write,

$$\Psi_t = V(t, F_t^{(\beta)}) - F_t^{(\beta)},$$

where $V(t, F_t^{(\beta)})$ represents the fair value of the VA contract with surrender option, as defined in

[5]The fair value of the fees that will be collected by the insurer after time t corresponds to $F_t^{(\infty)}(1 - e^{-c(T-t)})$. To see why, we can interpret the fee c as a dividend rate. The fair value of dividends to be received between times t and T is the difference between the fund value at time t ($F_t^{(\infty)}$) and the prepaid forward price at t for a claim paying $F_T^{(\infty)}$ at time T ($F_t^{(\infty)}e^{-c(T-t)}$).

equation (4.2.2). The delta is then obtained with,

$$
\begin{aligned}
\frac{\partial}{\partial S_t}\Psi_t &= \frac{\partial \Psi_t}{\partial F_t^{(\beta)}}\frac{\partial F_t^{(\beta)}}{\partial S_t} \\
&= \left[\frac{\partial V(t, F_t^{(\beta)})}{\partial F_t^{(\beta)}} - 1\right]\frac{\partial F_t^{(\beta)}}{\partial S_t},
\end{aligned}
\tag{4.10}
$$

where, $\partial V(t, F_t^{(\beta)})/\partial F_t^{(\beta)}$ must be estimated numerically based on a finite difference grid (as in Section 4.3) and,

$$
\frac{\partial F_t^{(\beta)}}{\partial S_t} = \frac{F_t^{(\beta)}}{S_t}. \tag{4.11}
$$

This is implicitly obtained in the derivation of equation (4.1) in Appendix 4.B, since the principle behind this derivation is a portfolio that hedges the delta of the variable annuity contract.

Remark 4.5.1. *When the insurer prices the VA contract and hedges its delta assuming optimal policyholder behaviour, his hedging strategy is a theoretical self-financing super-hedge. In other words, the hedge will always yield enough money for the insurer to help him cover the payoff of the VA as well as the surrender benefit. If the policyholder adopts a sub-optimal behaviour, then the insurer will also be able to derive a gain from the hedge. Unfortunately, these statements are only valid under the rather stringent assumptions of the Black-Scholes model. In practice, even if the insurer implements the optimal hedge, the presence of both discretization and model errors will cause the hedging strategy to lose its self-financing property and expose the insurer to a possible loss.*

4.5.3 Modeling policyholder behaviour

Given that the insurer establishes his hedging strategy assuming a particular form of policyholder behaviour, it is important to verify that the effectiveness of this strategy is robust to a wide range of dynamic lapsation behaviour observed in practice. For example, there is empirical evidence (e.g., Knoller, Kraut, and Schoenmaekers, 2013; Milliman, 2011) that the moneyness of the guarantee is a key driver of lapse behaviour among policyholders. The Canadian Institute of Actuaries (2002) and the American Academy of Actuaries (2005) both recommended to model surrenders by varying the lapse rate according to the moneyness of the guarantee. Based on a report from the Society of Actuaries (2012), approximately 60% of insurers follow this practice.

Therefore, we consider the following stopping time to model different forms of policyholder behaviour in our analysis of hedging effectiveness:

$$\tau_M = \inf_{0<t<T} \left\{ \frac{F_t^{(\beta)}(1-\kappa_t)}{G} \geqslant M_t \right\}, \tag{4.12}$$

where $F_t^{(\beta)}(1-\kappa_t)/G$ denotes what we call the moneyness ratio at time t, and M_t is a moneyness threshold, which when reached induces surrender. If the moneyness threshold is never attained, then we set $\tau_M = T$. When $M_t = \infty$, $\forall t$, then $\tau_M = T$ a.s., and this strategy falls back to keeping the contract until maturity. Moreover, since we can rewrite the condition $F_t^{(\beta)}(1-\kappa_t)/G \geqslant M_t$ as $F_t^{(\beta)} \geqslant M_t G/(1-\kappa_t)$, this stopping time encompasses all strategies of threshold-type, and, therefore all optimal strategies for the case $\beta = \infty$. For instance, if we choose $M_t = M_t^{\text{opt}}$, such that $M_t^{\text{opt}} G/(1-\kappa_t)$ matches the optimal lapsation fund threshold for $0 \leqslant t \leqslant T$, then this stopping time is the optimal one. Besides allowing us to consider two extreme cases of lapse modeling, i.e, no surrender and optimal behaviour, the stopping time in (4.12) can also help us define realistic sub-optimal lapse assumptions. For example, if M_t is constant $\forall t$, say $M_t = 1.5$, then a policyholder adopting such a strategy will surrender her contract when the surrender benefit, $F_t^{(\beta)}(1-\kappa_t)$, is at least 50% larger than the guarantee. The rationale behind this type of surrender behaviour is to avoid paying fees when the guarantee has little value. We will consider such surrender strategies based on a fixed moneyness ratio in our hedging analysis.

4.5.4 Results

To illustrate why eliminating the surrender incentive in the VA product design can simplify the insurer's hedging strategy and make it more effective, we revisit the example analyzed in Section 4.4. This example incorporates a state-dependent fee paid only when the account value is below $\beta = 150$, and the following assumptions: $F_0^{(150)} = G = 100$, $T = 10$, $r = 0.03$, and $\sigma = 0.165$. We assume that the surrender charge function is the minimal one which makes lapsation sub-optimal during the whole length of the contract, see Figure 4.8. The fair value of c for this particular case is 0.0155.

To highlight the importance of product design on risk management, we contrast this example to a typical constant fee product design ($\beta = \infty$) with the exact same schedule of surrender charges. It turns out that when pricing under optimal policyholder behaviour, the fair value of c is once again 0.0155, and the optimal lapsation boundary corresponds to the curve on the right of Figure 4.8. This result, which may seem surprising at first, has an intuitive explanation which we detail in the next paragraph.

First, note that for a given surrender charge function, and assuming that the policyholder lapses optimally, the state-dependent fair fee ($\beta < \infty$) is always at least as much as the constant fair fee ($\beta = \infty$). This is due to the fact that under the state-dependent fee design, the fee might be paid over a period of time shorter than in the constant fee case. Consequently, for a given lapse assumption, the state-dependent fee is an upper bound for the fair fee when $\beta = \infty$. In our specific example, this implies that the fair fee when $\beta = \infty$ is at most 0.0155. To see why 0.0155 is also a lower bound for the constant fair fee, consider a policyholder who lapses as soon as the account value hits the curve on the right of Figure 4.8. At this exact moment, we know that the surrender benefit is exactly equal to the value of the VA contract in the state-dependent fee case ($\beta = 150$). This is simply because the (minimal) surrender charge schedule was established so that the following condition is satisfied along the curve on the right of Figure 4.8:

$$E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(150)}, G)|\mathcal{F}_t] = (1-\kappa_t)F_t^{(150)}. \tag{4.13}$$

This strategy (holding a constant-fee contract and surrendering as soon as the account value hits the curve on the right of Figure 4.8) can be replicated by holding the state-dependent fee contract with $\beta = 150$ and $c = 0.0155$, and surrendering it as soon as $(1-\kappa_t)F_t^{(150)} = E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(150)}, G)|\mathcal{F}_t]$. In both cases, the surrender boundary would be the same, because it was defined through Equation (4.13). Since that surrender boundary is always under $\beta = 150$, the policyholder will pay fees continuously until surrender or maturity in both contracts. We know that the state-dependent fee contract is priced fairly at $c = 0.0155$. Thus, since under this particular surrender strategy the policyholder receives the same payoff from holding the constant fee or the state-dependent fee contract, they should both have the same price of 0.0155. This implies that $c = 0.0155$ must be a lower bound for the fair fee when $\beta = \infty$, as it is the fair c under *one* possible surrender strategy. Finally, the arguments presented in this paragraph imply also that (i) the fair fee must be exactly 0.0155 because it is bounded above and below by this value, and (ii) the curve on the right of Figure 4.8 must be the optimal lapsation boundary in the constant fee case ($\beta = \infty$).[6]

In summary, we consider two product designs which are priced assuming optimal policyholder behaviour. The first one is a constant fee design ($\beta = \infty$), $c = 0.0155$ and the surrender charge schedule given in Figure 4.8. This contract is fairly priced, as explained above. The optimal hedging strategy for this design is to hedge assuming the lapsation boundary is given by the curve on the right-hand side of Figure 4.8. The second design has the same surrender charge schedule and the same fair fee rate, but this fee is now paid only when the account value is below $\beta = 150$. The optimal hedging strategy for this design is to hedge assuming the policyholder will hold on to her contract until maturity.

Table 4.2 shows the statistics of the insurer's net delta hedging loss at maturity $(H - L)$ for the first product design with $\beta = \infty$ based on 500,000 stock paths projected on a weekly

[6]Note that this result about the equivalence of the fair fee when $\beta < \infty$ and $\beta = \infty$ will hold when (i) the surrender charge function is chosen as the minimal one making lapsation sub-optimal for the case $\beta < \infty$, and (ii) the value of F_t at the infimum of the function $E_{\mathbb{Q}}[e^{-r(T-t)} \max(F_T^{(\beta)}, G)|\mathcal{F}_t]/F_t^{(\beta)}$ is below β, for $t \leqslant 0 \leqslant T$.

frequency ($h = 1/52$) over $T = 10$ years with the Black-Scholes model and a real-world drift parameter of $\mu = 0.07$. The hedging portfolio is rebalanced weekly, and is established assuming either optimal behaviour (Opt) or no surrenders (NS). We also consider five possible types of surrender behaviours based on the stopping time in (4.12): $M_t = M_t^{\text{opt}}$, 1.3, 1.5, 1.7, and ∞ (see Section 4.5.3 for more details).

Table 4.2: Statistics of the insurer's net hedging loss

$\beta = \infty$										
Behaviour	$M_t = M_t^{\text{opt}}$		$M_t = 1.3$		$M_t = 1.5$		$M_t = 1.7$		$M_t = \infty$	
Hedge	Opt	NS	Opt	NS	Opt	NS	Opt	NS	Opt	NS
Mean	0.0	2.5	0.0	2.5	−1.0	2.9	−2.7	2.2	−10.4	−4.1
StDev	0.7	4.1	0.7	4.3	1.3	5.8	2.5	6.6	9.6	0.7
95% CTE	1.6	7.7	1.6	8.5	1.5	12.4	1.4	14.9	1.4	−2.5
99% VaR	1.9	8.0	1.9	8.8	1.8	12.9	1.8	15.7	1.8	−2.3

$\beta = 150$					
Behaviour	$M_t = M_t^{\text{opt}}$	$M_t = 1.3$	$M_t = 1.5$	$M_t = 1.7$	$M_t = \infty$
Hedge	NS	NS	NS	NS	NS
Mean	0.0	0.0	−1.1	−1.9	0.0
StDev	0.7	0.7	1.1	1.8	1.0
95% CTE	1.6	1.6	1.6	1.8	2.1
99% VaR	1.9	1.9	2.0	2.2	2.4

First, observe that on average the optimal hedging strategy never results in a loss, regardless of the policyholder behaviour. This is consistent with a super-hedge, but note that the insurer is still exposed to hedging risk as the 95% CTE is close to a loss of 1.5 for all scenarios. Nonetheless, hedging assuming optimal policyholder behaviour gives good results because it corresponds to hedging the worst-case scenario. However, given that the insurer sells many different VA products, implementing this optimal hedge for each product can be very impractical, if even possible. For this reason, insurers generally implement a simplified hedging strategy, such as a delta hedge that neglects the probability of surrenders. Unfortunately, the results in Table 4.2 show that this simplification significantly impairs hedging effectiveness when the policyholder can surrender her contract before maturity and is therefore not a viable solution for the insurer.

We now turn our attention to the second product design with $\beta = 150$. The second part of Table 4.2 shows the statistics of the insurer's net delta hedging loss at maturity $(H - L)$ for this product design based on the same 500,000 simulated weekly stock paths. We again analyse the same five surrender behaviours as in the first part of Table 4.2, but now consider only a delta hedge of the maturity benefit (no surrenders), as this strategy is also optimal for this design.

We observe that hedging effectiveness for the scenarios with $M_t = M_t^{\text{opt}}$ and $M_t = 1.3$, in the second part Table 4.2 is comparable with what was obtained in the first part, for the product design with $\beta = \infty$. However, the risk measures for the net hedging loss are a bit higher for the other scenarios. This increase is due to the fact that for the product design with $\beta = 150$, the insurer does not receive any fee income when the account value is above 150, but he is still exposed to hedging errors.

From a risk management standpoint, a product design which does not give rise to a surrender incentive seems preferable. First, the variable annuity product can be hedged conservatively assuming no surrenders which simplifies the construction of the hedging portfolio. Second, the hedging strategy can be implemented in a uniform manner across the portfolio of VAs because the optimal lapsation boundary does not have to be taken into consideration for each of the different product designs. Third, early surrenders can only be sub-optimal and generate additional revenue for the insurer. This additional revenue can compensate the insurer for the liquidity strain that arises with early surrenders, or for the need to adjust his hedging portfolio after a policyholder has lapsed.

4.6 Concluding Remarks

This chapter explores the surrender incentive resulting from a state-dependent fee structure. It shows that a combination of surrender charges and state-dependent fee structure can significantly reduce the incentive to surrender a variable annuity contract optimally, especially above the payment barrier. This chapter also explains how to obtain the minimal surrender charge to ensure that surrender incentives are eliminated. Under a state-dependent fee structure, such charges are shown to be much lower than when the fee is paid continuously. Finally, it is shown that when the surrender incentive is eliminated, hedging the maturity benefit is sufficient to protect the insurer against optimal and suboptimal lapse behaviour.

Further research should aim to test the robustness of the product design and dynamic hedging strategy under different market models. The product design could also be extended to variable annuity contracts offering other types of financial guarantees.

Appendix

4.A Proof of Equation (4.1)

In this section, we detail the derivation of (4.1). This derivation is very close to the derivation of the Black-Scholes equation for options on a dividend-paying stock. Since the derivation holds for any $\beta \in (0, \infty]$, in this appendix we omit the superscript (β) and refer to the account value by F_t for ease of exposition. First, we remind the reader of the dynamics of the VA account F_t:

$$dF_t = F_t((\mu - c\mathbb{1}_{\{F_t<\beta\}})dt + \sigma dW_t), \qquad 0 \leqslant t \leqslant T, \tag{4.14}$$

which is based on the value of the index S_t. The dynamics of S_t are given by

$$dS_t = S_t(\mu dt + \sigma dW_t), \qquad 0 \leqslant t \leqslant T. \tag{4.15}$$

We also consider the price of the VA contract, $V(t, F_t)$, which depends on time t and account value F_t. Thus, by Itô's lemma, we can write

$$dV(t, F_t) = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial F_t}dF_t + \frac{1}{2}\frac{\partial^2 V}{\partial F_t^2}\sigma^2 F_t^2 dt, \qquad 0 \leqslant t \leqslant T. \tag{4.16}$$

Now consider a portfolio composed of a long position in the VA contract and a short position in the index S_t. We denote the value of the portfolio at time t by Π_t and define it by

$$\Pi_t = V(t, F_t) - \Delta_t S_t, \tag{4.17}$$

where Δ_t is the number of shares of the index S_t in the portfolio at time t. Using (4.16), (4.14) and (4.15), and since the portfolio is assumed to be self financing, we have

$$\begin{aligned} d\Pi_t &= dV - \Delta_t dS_t \\ &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial F_t}dF_t + \frac{1}{2}\frac{\partial^2 V}{\partial F_t^2}\sigma^2 F_t^2 dt - \Delta_t S_t(\mu dt + \sigma dW_t) \\ &= \left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial F_t}F_t(\mu - c\mathbb{1}_{\{F_t<\beta\}}) + \frac{1}{2}\frac{\partial^2 V}{\partial F_t^2}\sigma^2 F_t^2 - \Delta_t S_t \mu\right)dt + \left(\frac{\partial V}{\partial F_t}F_t\sigma - \Delta_t S_t \sigma\right)dW_t. \end{aligned}$$

To make the portfolio risk-free, we need to eliminate the term in dW_t. This is done by setting $\Delta_t = \frac{\partial V}{\partial F_t}\frac{F_t}{S_t}$. Using this hedging ratio, we get

$$d\Pi_t = \left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial F_t}F_t(\mu - c\mathbb{1}_{\{F_t<\beta\}}) + \frac{1}{2}\frac{\partial^2 V}{\partial F_t^2}\sigma^2 F_t^2 - \Delta_t S_t \mu\right)dt. \tag{4.18}$$

This portfolio is thus risk-free. By no-arbitrage arguments, the return of this portfolio must be the risk-free rate. In other words, we must have

$$d\Pi_t = r\Pi_t dt = r(V(t, F_t) - \Delta_t S_t)dt. \tag{4.19}$$

Since (4.18) and (4.19) are equal, and since $\Delta_t = \frac{\partial V}{\partial F_t}\frac{F_t}{S_t}$ we have

$$\begin{aligned} &\frac{\partial V}{\partial t} - \frac{\partial V}{\partial F_t}F_t c\mathbb{1}_{\{F_t<\beta\}} + \frac{1}{2}\frac{\partial^2 V}{\partial F_t^2}\sigma^2 F_t^2 - rV(t, F_t) + r\frac{\partial V}{\partial F_t}F_t = 0 \\ &\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial F_t^2}\sigma^2 F_t^2 + \frac{\partial V}{\partial F_t}F_t(r - c\mathbb{1}_{\{F_t<\beta\}}) - rV(t, F_t) = 0. \end{aligned}$$

4.B Proof of Equation (4.3)

In this section, we prove (4.3) from Section 4.2. More precisely, we show that

$$\lim_{x\to\infty}\frac{V(t,x)}{x} = 1, \tag{4.20}$$

where $V(t, x)$ is the price of the VA contract with maturity payoff $\max(G, F_T^{(\beta)})$, $\beta < \infty$, and surrender option.

We first present two lemmas that will be used to prove (4.3) in Proposition 4.B.1.

Lemma 4.B.1. *Let $F_t^{(\beta)}$, $t \leqslant 0 \leqslant T$ be as defined in Section 4.2. Then,*

$$\lim_{x\to\infty} \frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}|F_t^{(\beta)} = x]}{x} = 1. \tag{4.21}$$

Proof. Let $m_F(t,u) = \inf_{t\leqslant s\leqslant u} F_s^{(\beta)}$ and $m_S(t,u) = \inf_{t\leqslant s\leqslant u} S_s$ be the minimum values attained by the account and the index, respectively, between times t and u. Then, (4.21) can be re-written as

$$\frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}|F_t^{(\beta)} = x]}{x} = \\ \frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{m_F(t,T)>\beta\}}|F_t^{(\beta)} = x]}{x} + \frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{m_F(t,T)\leqslant\beta\}}|F_t^{(\beta)} = x]}{x}. \tag{4.22}$$

To prove that $\lim_{x\to\infty}\frac{V(t,x)}{x} = 1$, we show that the first term of (4.22) goes to 1 as $x \to \infty$, and then show that the second term goes to 0 as $x \to \infty$.

To do so, let $C_t = e^{-c\int_0^t \mathbb{1}_{\{F_s^{(\beta)}<\beta\}}ds}$ and note that C_t is $\mathcal{F}_t$-measurable. Observe that if $F_t^{(\beta)} = C_t S_t > \beta$, then

$$F_u^{(\beta)}\mathbb{1}_{\{m_F(t,u)>\beta\}} = C_t S_u \mathbb{1}_{\{m_S(t,u)>\frac{\beta}{C_t}\}}, \text{ a.s. for } t < u \leqslant T, \tag{4.23}$$

since the fee is not paid as long as the account value is above β. It follows that

$$E_{\mathbb{Q}}[e^{-r(u-t)}F_u^{(\beta)}\mathbb{1}_{\{m_F(t,u)>\beta\}}|F_t^{(\beta)} = x] = C_t E_{\mathbb{Q}}\left[e^{-r(u-t)}S_u\mathbb{1}_{\{m_S(t,u)>\frac{\beta}{C_t}\}} \mid S_t = \frac{x}{C_t}\right]. \tag{4.24}$$

The expectation on the right-hand side of (4.24) is the price of a down-and-out call option with strike 0 and barrier $\frac{\beta}{C_t}$. Under the Black-Scholes model, the price of this option has a closed-form solution (see, for example, Chapter 18 of Björk (2004)). So we can write

$$C_t E_{\mathbb{Q}}\left[e^{-r(u-t)}S_u\mathbb{1}_{\{m_S(t,u)>\frac{\beta}{C_t}\}}|S_t = \frac{x}{C_t}\right] = \\ x\mathcal{N}\left(\frac{\ln\frac{x}{\beta} + \left(r + \frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right) - \beta\left(\frac{\beta}{x}\right)^{\frac{2r}{\sigma^2}}\mathcal{N}\left(\frac{\ln\frac{\beta}{x} + \left(r + \frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right).$$

Thus, we have

$$\frac{E_{\mathbb{Q}}[e^{-r(u-t)}F_u^{(\beta)}\mathbb{1}_{\{m_F(t,u)>\beta\}}|F_t^{(\beta)}=x]}{x}$$
$$= \frac{\frac{x}{C_t}\mathcal{N}\left(\frac{\ln\frac{x}{\beta}+\left(r+\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right)}{x} - \frac{\beta\left(\frac{\beta}{x}\right)^{\frac{2r}{\sigma^2}}\mathcal{N}\left(\frac{\ln\left(\frac{\beta}{x}\right)+\left(r+\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right)}{x}$$
$$= \frac{1}{C_t}\mathcal{N}\left(\frac{\ln\left(\frac{x}{\beta}\right)+\left(r+\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right) - \left(\frac{\beta}{x}\right)^{\frac{2r}{\sigma^2}+1}\mathcal{N}\left(\frac{\ln\left(\frac{\beta}{x}\right)+\left(r+\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right).$$

The result follows since $\lim_{y\to\infty}\mathcal{N}(y)=1$ and $\lim_{y\to-\infty}\mathcal{N}(y)=0$.

To show that the second term of (4.22) vanishes for large values of x, we first note that

$$\frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{m_F(t,T)\leqslant\beta\}}|F_t^{(\beta)}=x]}{x} \leqslant \frac{E_{\mathbb{Q}}[e^{-r(T-t)}S_T\mathbb{1}_{\{m_S(t,T)\leqslant\frac{\beta}{C_t}\}}|S_t=\frac{x}{C_t}]}{x}, \tag{4.25}$$

since for any $0\leqslant t\leqslant T$, $F_t=S_tC_t\leqslant S_t$, a.s. The right-hand side of (4.25) is the price of of down-and-in call option with strike 0 and barrier $\frac{\beta}{C_t}$. The price option also has a closed-form solution (again, see Chapter 18 of Björk (2004)), which allows us to write

$$\frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{m_F(t,T)\leqslant\beta\}}|F_t^{(\beta)}=x]}{x} \leqslant$$
$$\frac{1}{C_t}\left\{\mathcal{N}\left(\frac{\ln\frac{\beta}{x}-(\widetilde{r}+\sigma^2)(u-t)}{\sigma\sqrt{u-t}}\right)+\left(\frac{\beta}{x}\right)^{\frac{2\widetilde{r}}{\sigma^2}+2}\mathcal{N}\left(\frac{\ln\frac{\beta}{x}+(\widetilde{r}+\sigma^2)(u-t)}{\sigma\sqrt{u-t}}\right)\right\}.$$

Since $\lim_{y\to-\infty}\mathcal{N}(y)=0$, $\lim_{x\to\infty}\frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}\mathbb{1}_{\{m_F(t,T)\leqslant\beta\}}|F_t^{(\beta)}=x]}{x}=0$. □

Lemma 4.B.2. *Let $F_t^{(\beta)}$, $t\leqslant 0\leqslant T$ be as defined in Section 4.2. Then,*

$$\lim_{x\to\infty}\frac{x+E_{\mathbb{Q}}[e^{-r(T-t)}(G-F_T^{(\beta)})^+|F_t^{(\beta)}=x]}{x}=1,$$

where $(G-F_T^{(\beta)})^+=\max(G-F_T^{(\beta)},0)$.

Proof. Denote by $p_{t,S_t}(T,G,\delta)$ the price at time t of a European put option with strike G and maturity T, on a stock S_t paying dividends at a continuous rate δ. Using

$$\frac{S_T e^{-c(T-t)}}{S_t} < \frac{F_T^{(\beta)}}{F_t^{(\beta)}} < \frac{S_T}{S_t}, \qquad \text{a.s.},$$

it is easy to show that

$$p_{t,x}(T,G,0) \leqslant E_{\mathbb{Q}}[e^{-r(u-t)}(G-F_u^{(\beta)})^+|F_t^{(\beta)}=x] \leqslant p_{t,x}(T,G,c).$$

Since $\forall \delta \geqslant 0$, $\lim_{x\to\infty} p_{t,x}(T,G,\delta)=0$, the desired result follows from

$$\lim_{x\to\infty} \frac{E_{\mathbb{Q}}[e^{-r(T-t)}(G-F_T)^+|F_t^{(\beta)}=x]}{x} = 0.$$

Using Lemmas 4.B.1 and 4.B.2, we can now prove the main result of this section.

Proposition 4.B.1. *Let $V(t,F_t^{(\beta)})$ be as defined in Section 4.2. Then,*

$$\lim_{x\to\infty} \frac{V(t,x)}{x} = 1. \tag{4.26}$$

Proof. To prove Proposition 4.B.1, we first show

$$E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}|\mathcal{F}_t] \leqslant V(t,F_t^{(\beta)}) \leqslant F_t + E_{\mathbb{Q}}[e^{-r(T-t)}(G-F_T^{(\beta)})^+|\mathcal{F}_t]. \tag{4.27}$$

The first inequality stems from the fact that the price of the contract with surrender option, $V(t,F_t^{(\beta)})$ is worth at least as much as the present value of the maturity benefit, which is itself at least equal to the expectation of the account value at maturity. To show the second inequality, recall that the payoff of the contract is either $(1-\kappa_u)F_u^{(\beta)}$ if the contract is surrendered at time $u<T$, or $F_T^{(\beta)}+(G-F_T^{(\beta)})^+$ at time T if the contract is kept until then. Notice also that the present value of the surrender benefit is at most $F_t^{(\beta)}$ since for any $u<t<T$,

$$E_{\mathbb{Q}}[e^{-r(u-t)}(1-\kappa_u)F_u^{(\beta)}|\mathcal{F}_t] \leqslant E_{\mathbb{Q}}[e^{-r(u-t)}F_u^{(\beta)}|\mathcal{F}_t] \leqslant F_t^{(\beta)}. \tag{4.28}$$

Thus, the value of the variable annuity contract is bounded above by the sum of the expected value of the two possible payoffs, and it follows that

$$V(t,F_t^{(\beta)}) \leqslant F_t + E_{\mathbb{Q}}[e^{-r(T-t)}(G-F_T^{(\beta)})^+|\mathcal{F}_t].$$

From (4.27), it follows that

$$\frac{E_{\mathbb{Q}}[e^{-r(T-t)}F_T^{(\beta)}|F_t^{(\beta)}=x]}{x} \leqslant \frac{V(t,x)}{x} \leqslant \frac{F_t^{(\beta)}+E_{\mathbb{Q}}[e^{-r(T-t)}(G-F_T^{(\beta)})^{+}|F_t^{(\beta)}=x]}{x}. \tag{4.29}$$

To complete the proof of Proposition 4.B.1, it suffices to take the limit of (4.29) as $x \to \infty$. The result follows from Lemma 4.B.1 and Lemma 4.B.2, since the first and the third terms of (4.29) both go to 1 in the limit. □

Chapter 5

Optimal surrender under deterministic fee structure

5.1 Introduction

This chapter is based on a chapter that was written in collaboration with Dr. Carole Bernard and that was submitted to the book *Innovations in Risk Management*.

In this chapter, we consider a VA with guaranteed minimum accumulation benefit (GMAB) and the option to surrender. We propose to change the fee structure so that a fixed percentage c is paid from the fund while an additional fee is paid via regular installments of a deterministic amount p_t at time t, throughout the term of the contract. Another interpretation would be to say that it is a “state-dependent percentage fee” computed as

$$c(t, F_t) = c + \frac{p_t}{F_t}$$

which denotes the percentage of the fund taken to pay for the options. This is equivalent to saying that $c(t, F_t)F_t = cF_t + p_t$ is the fee paid at time t.

This fee structure can be seen as a compromise between a constant fee rate (as in Chapter 2) and a state-dependent fee paid only under a certain threshold (as in Chapters 3 and 4). The deterministic fee expressed as a fixed amount is more in line with the cost of the option; as a percentage of the account, the fee rate decreases with the value of the financial guarantee. This results in lower fee rates (as a percentage of the account) when the account value is high, when compared to the constant fee. As demonstrated in Chapter 2, the surrender incentive depends on the fee paid when the account value is large. This is why the deterministic fee structure can reduce surrender incentives. An advantage of this fee structure over the state-dependent one presented in Chapter 3 is that it is easy to explain to a less sophisticated policyholder. In addition, since the fee can be expressed as a continuous function of the account value, pricing the VA is mathematically simpler than when the fee is a discontinuous function of the account value F_t.

We want to investigate the impact of such fee structure on the value of the surrender option. In Section 5.2, we describe the model and the VA contract. Section 5.3 introduces a theoretical result and discusses the valuation of the surrender option. Numerical examples are presented in Section 5.4 and Section 5.5 concludes.

5.2 Assumptions and Model

Consider a market with a risk-free asset yielding a constant risk-free rate r and an index $\{S_t\}_{0\leqslant t\leqslant T}$ evolving as in the Black-Scholes model so that

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t,$$

under the risk-neutral measure $\mathbb{Q}$, where $\sigma > 0$ is the constant instantaneous volatility of the index. Let $\mathcal{F}_t$ be the natural filtration associated with the Brownian motion W_t.

5.2.1 Variable Annuity

We consider a VA contract with an underlying fund fully invested in the index S_t. At time t, we assume that the fee paid is the sum of a constant percentage $c \geqslant 0$ of the account value and a deterministic amount p_t. Setting $p_t = 0$, we will find back results commonly used in the literature with the fee being only paid as a percentage of the fund.

We further assume that the investment of the policyholder at time 0 is $P = F_0$, and that regular additional premiums a_t are invested at time t. Additional contributions are common in variable annuities, but they are regularly neglected and most academic research focuses on the single premium case as it is simpler.

When additional contributions can be made to the account, VAs are called Flexible Premiums Variable Annuities (FPVAs). Chi and Lin (2012) provide examples of such VAs where the policyholder is given the choice between a single premium and a periodic monthly payment in addition to some initial lump sum. Analytical formulae for the value of such contracts can be found in Costabile (2013) and Huerlimann (2010). In the first part of this chapter, we show how flexible premium payments influence the surrender value.

We assume that all premiums invested at 0 and at later times t are invested in the fund. All fees (percentage or fixed fees) are paid from the fund. We need to model the dynamics of the fund. Our approach is inspired by Chi and Lin (2012) who study flexible premiums paid over time. For the sake of simplicity, we assume that all cash flows happen in continuous time, so that a fixed payment of A at time 1 (say, end of the year) is similar to a payment made continuously over the interval $[0, 1]$. Due to the presence of a risk-free rate r, an amount paid at time T equal to A is equivalent to an instantaneous contribution of $a_t\ dt$ at any time $t \in (0, 1]$ so that the annual amount paid per year is $A = \int_0^1 a_t e^{r(1-t)} dt$. By abuse of notation, if a_t is constant over the year, we will write that a_t is the annual rate of contribution per year (although there is no compounding effect).

Precisely the dynamics of the fund can be written as follows

$$dF_t = (r - c)F_t dt + \sigma F_t dW_t + a_t dt - p_t dt$$

with $F_0 = V_0$, and where F_t denotes the value of the fund at time t, a_t is the annual amount of contributions (or equivalently, the fee paid at the end of each year), c is the annual rate of fees and p_t is the annual amount of fee to pay for the options. Similarly to Chi and Lin (2012), it is straightforward to show that

$$F_t = F_0 e^{(r-c-\frac{\sigma^2}{2})t+\sigma W_t} + \int_0^t (a_s - p_s) e^{(r-c-\frac{\sigma^2}{2})(t-s)+\sigma(W_t - W_s)} ds,\ t \geqslant 0,$$

that is

$$F_t = S_t e^{-ct} + \int_0^t (a_s - p_s) e^{-c(t-s)} \frac{S_t}{S_s} ds. \tag{5.1}$$

To simplify the notation, we will write

$$F_t = S_t e^{-ct} + \int_0^t b_s e^{-c(t-s)} \frac{S_t}{S_s} ds, \tag{5.2}$$

where $b_s = a_s - p_s$ can take values in $\mathbb{R}$. While in the case of regular contributions, b_s is typically positive, it can also be negative, for example in the single premium case, or if the regular premiums are very low. We will split b_s into contributions a_s and deterministic fee p_s when it is needed for the interpretation of the results.

This formulation can be seen as an extension of the case studied in Chi and Lin (2012) who assume a constant contribution parameter $a_t = a$ for all t and no periodic fees ($p_t = 0$). As can be seen from (5.2), the account value becomes path-dependent and involves a continuous arithmetic average.

To simplify the notation, we assume without loss of generality that $F_0 = S_0 = V_0$.

5.2.2 Benefits

We assume that there is a guaranteed minimum accumulation rate $g < r$ on all the contributions of the policyholder until time T so that the accumulated guaranteed benefit G_t at time t has dynamics

$$dG_t = gG_t dt + a_t dt$$

where $G_0 = V_0$ at time 0. Thus, at time t the guaranteed amount G_t can be expressed as

$$G_t = V_0 e^{gt} + \int_0^t a_s e^{g(t-s)} ds.$$

When the contribution rate is constant ($a_t = a$), the guaranteed value can be simplified to

$$G_t = V_0 e^{gt} + a \left(\frac{e^{gt} - 1}{g} \mathbb{1}_{\{g>0\}} + t \mathbb{1}_{\{g=0\}} \right).$$

Chi and Lin (2012) develop techniques to price and hedge the guarantee at time t. Using their numerical approach it is possible to estimate the fair fee for the European VA (Proposition 3 in their paper).

As in Bernard, MacKay, and Muehlbeyer (2014) and Milevsky and Salisbury (2001), we assume that the policyholder has the option to surrender the policy at any time t and to receive a surrender benefit at surrender time equal to

$$(1 - \kappa_t) F_t$$

where κ_t is a penalty percentage charged for surrendering at time t. As presented for instance in Bernard and Lemieux (2008), Milevsky and Salisbury (2001) or Palmer (2006), a standard surrender penalty is a non-increasing function of time. Typical VAs sold in the US have a surrender charge period. A typical example is New York Life's Premier Variable Annuity (New York Life (2014)), for which the surrender charge starts at 8% in the first contract year, decreases by 1% per year to reach 2% in year 7. From year 8 on, there is no penalty on surrender. In another example, "the surrender charge is 7% during the first Contract Year and decreases by 1% each subsequent Contract Year. No surrender charge is deducted for surrenders occurring in Contract Years 8 and later" (Thrivent Financial (2014)). In general, the maximum surrender charge will be around 8% of the amount withdrawn based on the annuity. The percentage of the surrender charge varies and generally decreases during the surrender charge period.

5.3 Valuation of the surrender option

In this section, we discuss the valuation of the full variable annuity contract, with maturity benefit and surrender option.[1] We first present a sufficient condition to eliminate the possibility of optimal surrender. We then explain how we evaluate the surrender option using partial differential equations (PDEs). We consider a variable annuity contract with maturity benefit only, which can be surrendered. We choose to ignore the death benefits that are typically added to that type of contract since our goal is to analyze the effect of the fee structure on the value of the surrender option.

5.3.1 Theoretical Result on Optimal Surrender Behaviour

According to (5.2) the account value F_t can be written as follows at time t

$$F_t = e^{-ct}S_t + \int_0^t b_s e^{-c(t-s)}\frac{S_t}{S_s}ds, \ t \geqslant 0,$$

and at time $t + dt$, it is equal to

$$F_{t+dt} = e^{-c(t+dt)}S_{t+dt} + \int_0^{t+dt} b_s e^{-c(t+dt-s)}\frac{S_{t+dt}}{S_s}ds.$$

Proposition 5.3.1 (Sufficient condition for no surrender)**.** *For a fixed time $t \in [0, T]$, a sufficient condition to* **not** *surrender at time t is given by*

$$(\kappa_t' + (1-\kappa_t)c)F_t < b_t(1-\kappa_t), \tag{5.3}$$

where $\kappa_t' = \frac{\partial \kappa_t}{\partial t}$. Here are some special cases of interest:

- *When $a_t = p_t = 0$ (no periodic investment, no periodic fee) and $\kappa_t = 1 - e^{-\kappa(T-t)}$ (situation considered in Chapter 2) then $b_t = 0$ and (5.3) becomes*

$$\kappa > c.$$

- *When $a_t = 0$ (no periodic investment, i.e. a single lump sum paid at time 0), then $b_t = -p_t \leqslant 0$. Assume that $p_t > 0$ so that $b_t < 0$ thus*

[1] In this chapter, we quantify the additional value added by the possibility for the policyholder to surrender his policy. We call it the surrender option, as in Milevsky and Salisbury (2001). It is not a guarantee that can be added to the variable annuity, but rather a real option created by the fact that the contract can be surrendered.

- *If* $\kappa_t' + (1-\kappa_t)c > 0$ *(for example if* κ *is constant), then the condition can never be satisfied and no conclusion can be drawn.*
- *If* $\kappa_t' + (1-\kappa_t)c < 0$ *then it is not optimal to surrender when*

$$F_t > \frac{-p_t(1-\kappa_t)}{\kappa_t' + (1-\kappa_t)c}.$$

- *When* $\kappa_t = \kappa$ *and* $b_t = b$ *are constant over time, the condition* (5.3) *can be rewritten as*

$$F_t < \frac{b(1-\kappa)}{c(1-\kappa)} = \frac{b}{c}.$$

Remark 5.3.1. *Proposition 5.3.1 shows that in the absence of periodic fees and investment, an insurer can easily ensure that it is never optimal to exercise by choosing a surrender penalty of* $1 - e^{-\kappa t}$ *with a penalty parameter* κ *higher than the percentage fee* c*. This is in line with our result from Chapter 4. Proposition 5.3.1 shows that it is also possible when there are periodic fees and investment opportunities and the conditions are more complicated.*

Remark 5.3.2. *Equation* (5.3) *of Proposition 5.3.1 has an intuitive interpretation. In fact, it can be re-written as*

$$(\kappa_t' + (1-\kappa_t)c)F_t + p_t(1-\kappa_t) < a_t(1-\kappa_t),$$

where the left-hand side is approximately the gain made from being out of the contract from t *to* $t+dt$*, and right-hand side approximates the increase in the account value between times* t *and* $t+dt$*. Thus, it is not optimal to surrender if the gain from surrender is less than the increase in the account value.*

Proof. Consider a time t at which it is optimal to surrender. This implies that for any time interval of length $dt > 0$, it is better to surrender at time t than to wait until time $t+dt$. In other words, the surrender benefit at time t must be at least equal to the expected discounted value of the contract at time $t+dt$, and in particular larger than the surrender benefit at time $t+dt$. Thus

$$(1-\kappa_t)F_t \geqslant E[e^{-rdt}(1-\kappa_{t+dt})F_{t+dt}\ |\mathcal{F}_t]$$

Using the martingale property for the discounted stock price S_t and the independence of increments for the Brownian motion, we know that $E[S_{t+dt}e^{-rdt}] = S_t$ and $E\left[\frac{S_{t+dt}}{S_t}\middle|\mathcal{F}_t\right] = E\left[\frac{S_{t+dt}}{S_t}\right] =$

e^{rdt} thus

$$
\begin{aligned}
E[e^{-rdt}F_{t+dt}|\mathcal{F}_t] &= e^{-c(t+dt)}S_t + \int_0^t b_s e^{-c(t+dt-s)}\frac{S_t}{S_s}ds + \int_t^{t+dt} b_s e^{-c(t+dt-s)}e^{-rdt}E\left[\frac{S_{t+dt}}{S_s}\right]ds, \\
&= e^{-c(t+dt)}S_t + \int_0^t b_s e^{-c(t+dt-s)}\frac{S_t}{S_s}ds + \int_t^{t+dt} b_s e^{-c(t+dt-s)}ds, \\
&= e^{-cdt}F_t + e^{-cdt}\int_t^{t+dt} b_s e^{-c(t-s)}ds. \qquad (5.4)
\end{aligned}
$$

Thus

$$
(1-\kappa_t)F_t \geqslant (1-\kappa_{t+dt})\left(e^{-cdt}F_t + e^{-cdt}\int_t^{t+dt} b_s e^{-c(t-s)}ds\right)
$$

We then use $\kappa_{t+dt} = \kappa_t + \kappa_t' dt + o(dt)$, $e^{-cdt} = 1 - cdt + o(dt)$ and $\int_t^{t+dt} b_s e^{-c(t-s)}ds = b_t dt + o(dt)$ to obtain

$$
(1-\kappa_t)F_t \geqslant (1-\kappa_t-\kappa_t' dt)\left((1-cdt)F_t + (1-cdt)b_t dt\right) + j(dt)
$$

which can be further simplified into

$$
(\kappa_t' + (1-\kappa_t)c)F_t dt \geqslant b_t(1-\kappa_t)dt + j(dt). \qquad (5.5)
$$

where the function $j(dt)$ is $o(dt)$. Since this holds for any $dt > 0$, we can divide (5.5) by dt and take the limit as $dt \to 0$. Then, we get that if it is optimal to surrender the contract at time t,

$$
(\kappa_t' + (1-\kappa_t)c)F_t \geqslant b_t(1-\kappa_t).
$$

It follows that if $(\kappa_t' + (1-\kappa_t)c)F_t < b_t(1-\kappa_t)$, it is not optimal to surrender the contract at t.□

5.3.2 Valuation of the surrender option using PDEs

To evaluate the surrender option, we compare the value of the variable annuity contract with and without surrender. These contracts can be compared to American and European options, respectively, since the guarantee in the former can be exercised at any time before maturity while the latter is only triggered when the contract expires.

From now on, we assume that the deterministic fee p_t is constant over time, so that $p_t = p$ for any time t. We also assume that the policyholder makes no contribution after the initial premium ($a_t = 0$ for any t), and denote $G_T = G$ to simplify the notation.

We denote by $U(t, F_t)$ and $V(t, F_t)$ the value of the contract without and with surrender option, respectively. The value of the contract without the surrender option is simply the risk-neutral expectation, conditional on the filtration up to time t, of the payoff at maturity.

$$U(t, F_t) = E[e^{-r(T-t)} \max(G, F_T)|\mathcal{F}_t] \tag{5.6}$$

It is well-known[2] that the value of a European contingent claim on the fund value F_t follows the following PDE:

$$\frac{\partial U}{\partial t} + \frac{1}{2}\frac{\partial^2 U}{\partial F_t^2} F_t^2 \sigma^2 + \frac{\partial U}{\partial F_t}(F_t(r-c) - p) - rU = 0. \tag{5.7}$$

Note that(5.7) is very similar to the Black-Scholes equation for a contingent claim on a stock that pays dividends (here, the constant fee c represents the dividends), with the addition of the term $\frac{\partial U}{\partial F_t} p$ resulting from the presence of a deterministic fee. Since it represents the contract described in Section 5.2, (5.7) is subject to the following conditions:

$$U(T, F_T) = \max(G, F_T)$$
$$\lim_{F_t \to 0} U(t, F_t) = G e^{-r(T-t)}.$$

The last condition results from the fact that when the fund value is very low, the guarantee is certain to be triggered. When $F_t \to \infty$, the problem is unbounded. However, we have the following asymptotic behaviour:

$$\lim_{F_t \to \infty} \frac{U(t, F_t)}{E_t[F_T e^{-r(T-t)}]} = 1, \tag{5.8}$$

which stems from the fact that when the fund value is very high, the value of the guarantee approaches 0. We will use this asymptotic result to solve the PDE numerically, when truncating the grid of values for F_t. The expected value in (5.8) is easily calculated and is given in the proof of Proposition 5.3.1.

To express the value of the variable annuity contract with the surrender option, we must introduce further notation. We denote by $\mathcal{T}_t$ the set of all stopping times τ greater than t and bounded by T. Then we can express the continuation value of the VA contract with surrender as

$$V^*(t, F_t) = \sup_{\tau \in \mathcal{T}_t} E[e^{-r(\tau - t)} \psi(\tau, F_\tau)],$$

where

$$\psi(t, x) = \begin{cases} (1-\kappa_t)x, & \text{if } t \in (0, T) \\ \max(G, x), & \text{if } t = T \end{cases}$$

[2]See, for example, Björk (2004), Section 7.3. The derivation of this PDE is similar to the one used in Chapter 4.

is the payoff of the contract at surrender or maturity. Finally, we let $\mathcal{R}_t$ be the optimal surrender region at time $t \in [0, T]$ and define it by

$$\mathcal{R}_t = \{F_t : V^*(t, F_t) \leqslant \psi(t, F_t)\}.$$

In other words, the optimal surrender region is given by the fund values for which the surrender benefit is worth at least as much as the VA contract if the policyholder continues to hold it for at least a small amount of time. The complement of the optimal surrender region $\mathcal{R}_t$ will be referred to as the continuation region. We also define B_t, the optimal surrender boundary at time t, by

$$B_t = \inf_{F_t \in [0,\infty)} \{F_t \in \mathcal{R}_t\}.$$

As is the case for the American put option[3], the VA contract with surrender option gives rise to a free boundary problem. In the continuation region, $V^*(t, F_t)$ follows (5.7), the same equation as for the contract without surrender option. However, in the optimal surrender region, the value of the contract with surrender is the value of the surrender benefit:

$$V^*(t, F_t) = \psi(t, F_t), \qquad t \in [0, T], F_t \in \mathcal{R}_t. \tag{5.9}$$

For the contract with surrender, the PDE to solve is thus subject to the following conditions:

$$\begin{aligned}
&V^*(T, F_T) = \max(G, F_T) \\
&\lim_{F_t \to 0} V^*(t, F_t) = Ge^{-r(T-t)} \\
&\lim_{F_t \to B_t} V^*(t, F_t) = \psi(t, B_t). \\
&\lim_{F_t \to B_t} \frac{\partial}{\partial F_t} V^*(t, F_t) = 1 - \kappa_t.
\end{aligned}$$

For any time $t \in [0, T]$, the value of the VA with surrender is given by $V(t, F_t) = \max(V^*(t, F_t), \psi(t, F_t))$. This free boundary problem is solved in Section 5.4 using numerical methods.

5.4 Numerical Example

To price the VA using a PDE approach, we modify (5.7) to express it in terms of $x_t = \ln F_t$. We discretize the resulting equation over a rectangular grid with time steps $dt = 0.0001$ ($dt = 0.0002$ for $T = 15$) and $dx = \sigma\sqrt{3dt}$,[4] from 0 to T in t and from 0 to $\ln 450$ in x. We use an explicit scheme with central difference in x and in x^2.

[3]See, for example, Carr, Jarrow, and Myneni (1992)

[4]As suggested in Racicot and Théoret (2006).

5.4.1 Numerical Results

We now consider variable annuities with the maturity benefit described in Section 5.2. We assume that the initial premium $P_0 = 100$, that there is no periodic premium ($a_s = a = 0$), that the deterministic fee is constant ($p_t = p$) and that the guaranteed roll-up rate is $g = 0$. We further assume $r = 0.03$, $\sigma = 0.2$ and that the surrender charge, if any, is of the form $\kappa_t = 1 - e^{\kappa(T-t)}$.

For a 10-year variable annuity contract with and without surrender charge, the results are presented in Table 5.1. In each case, the fee levels c and p are chosen such that the maturity benefit is fully covered.[5] It is interesting to note that as a percentage of the initial premium, the fair fee when it is paid as a fixed amount is higher than the fair constant percentage fee. This is due to the fact that when the fee is a fixed amount, it represents a lower percentage of the fund when the fund value is high. The insurer thus receives less than when the fee is a fixed percentage. When the fund value is low, the fixed amount fee represents a larger proportion of the fund compared to the constant percentage fee. This higher percentage drags the fund value down and increases the option value. This can explain the difference between the fair fixed percentage and fixed amount fees.

Fee		Surrender Option	
c	p	$\kappa = 0$	$\kappa = 0.005$
0.0000	2.0321	3.07	1.02
0.0050	1.3875	3.50	1.46
0.0100	0.7443	3.92	1.89
0.0158	0.0000	4.43	2.39

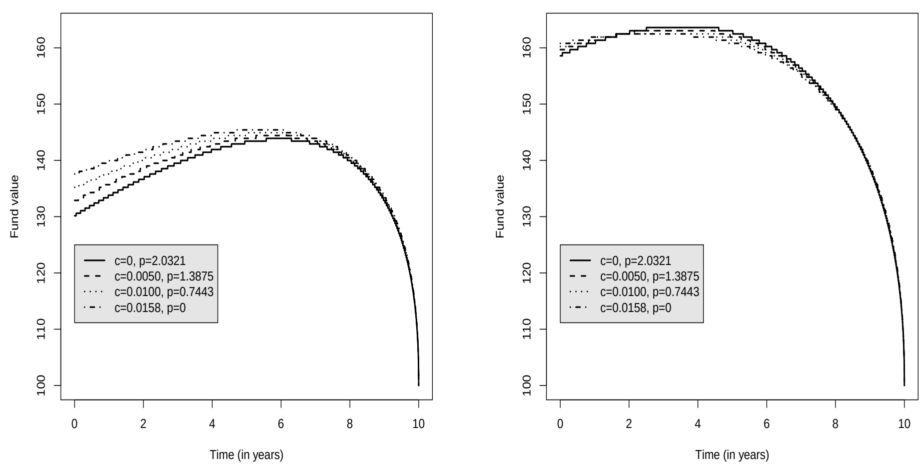
Table 5.1: Value of the surrender option for a 10-year variable annuity contract for different fee structures.

In Table 5.1, we present the values of the surrender option for different fee structures. In particular, we study the case where the fee is set only as a fixed percentage, and the one where the fee is only paid as a deterministic amount. Fee structures combining both types of fees are also analysed. The results in Table 5.1 show that when the fee is set as a fixed amount, the value of the surrender option is lower than when the fee is expressed as a percentage of the fund. When a mix of both types of fees is applied, the value of the surrender option decreases as the fee set as a percentage of the fund decreases. When the fee is set as a fixed amount, a lower percentage of the fund is paid out when the fund value is high. Consequently, the fee paid by the policyholder is lower when the value of the guarantee is low. This explains why the value of the surrender option decreases when the fee is paid as a fixed amount. This can be observed

[5] That is, $P_0 = U(t, F_t)$.

both with and without surrender charges. However, surrender charges decrease the value of the surrender option, as expected. The effect of using a fee set as a fixed amount, instead of a fixed percentage, is even more noticeable when there is a surrender charge.

Figure 5.1 shows the optimal surrender boundaries for the fee structures presented in Table 5.1. As expected, the optimal boundaries are higher when there is a surrender charge. Those charges are put in place in part to discourage policyholders from surrendering early. The boundaries are also less sensitive to the fee structure when there is a surrender charge. In fact, when there is a surrender charge, setting the fee as a fixed amount leads to a higher optimal boundary during most of the contract. This highlights the advantage of the fixed amount fee structure combined with surrender charges. Without those charges, the fixed fee amount could lead to more surrenders.



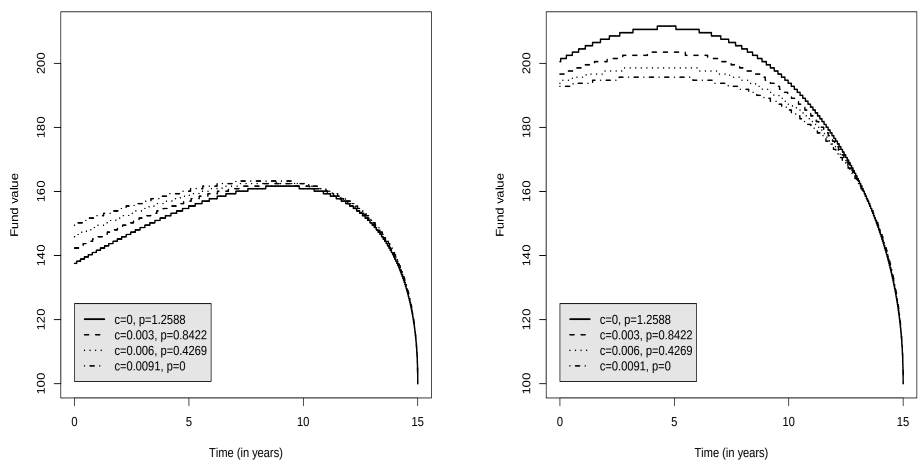
$\kappa = 0$ $\kappa = 0.005$

Figure 5.1: Optimal surrender boundary when $T = 10$.

Table 5.2 shows the effect of the fee structure on the surrender options for 5-year and 15-year contracts. For the 15-year contract, we lowered the surrender charge parameter to $\kappa = 0.004$ to ensure that the optimal surrender boundary is always finite. For both maturities, setting the fee as a fixed amount instead of a fixed percentage has a significant effect on the value of the surrender option. This effect is amplified for longer maturities. As for the 10-year contract, combining the fixed amount fee with a surrender charge further reduces the value of the surrender option, especially when $T = 15$. The optimal surrender boundaries for different fee structures when $T = 15$ are presented in Figure 5.2. For longer maturities such as this one, the combination of surrender charges and deterministic fee increases the surrender boundary more significantly.

$T=5$				$T=15$			
Fee		Surrender Option		Fee		Surrender Option	
c	p	$\kappa=0$	$\kappa=0.005$	c	p	$\kappa=0$	$\kappa=0.004$
0.0000	4.1500	3.09	2.09	0.0000	1.2588	2.76	0.23
0.0100	2.9714	3.32	2.33	0.0030	0.8422	3.30	0.77
0.0200	1.7955	3.56	2.57	0.0060	0.4269	3.84	0.84
0.0353	0.0000	3.92	2.94	0.0091	0.0000	4.40	1.86

Table 5.2: Value of the surrender option for 5-year and 15-year variable annuity contracts for different fee structures.



$\kappa=0$ $\kappa=0.004$

Figure 5.2: Optimal surrender boundary when $T=15$.

5.5 Concluding Remarks

In this chapter, we introduced the notion of a deterministic fee as a new type of state-dependent fee. Combined with surrender charges, it can reduce the surrender incentive. Further investigation into the form of the surrender charge could lead to designs eliminating the optimal surrender incentive. Nonetheless, this type of fee structure is an interesting alternative to the state-dependent fee introduced in Chapter 3, since it is easier to explain to the policyholder. In addition, it does not present a discontinuity, which simplifies the pricing of liabilities.

We also explored a sufficient condition that allows to eliminate the possibility of optimal surrender for variable annuity contracts with fairly general fee structures. This result extends the one presented in Chapter 2. It could be used to explore fee structures that eliminate the surrender incentive.

Future work should focus on more general payouts in more general market models, and include death benefits. These considerations would introduce our results to a more realistic setting, which would probably increase their attractiveness to the insurance industry.

Chapter 6

Group Self-Annuitization Schemes: How optimal are the 'optimal strategies'?

6.1 Introduction and Motivation

Group self-annuitization (GSA) schemes allow individuals with retirement funds to pool their assets with other similar individuals, with a view to providing income through retirement. By pooling funds, the members benefit from risk sharing. Each year end (say), the income of the surviving members is adjusted to reflect the investment experience of the pooled fund, the mortality experience of the annuitants or possibly both.

With the decline of defined benefit plans, increased attention on longevity risk, and (perceived) high cost of annuities purchased through insurance companies, GSA schemes have been attracting attention from researchers and from the pensions industry. Some defined contribution (DC) pension plan sponsors are using GSAs to offer retirees the benefits of pooling (albeit to a limited extent), without requiring sponsors to retain the investment, mortality and longevity risks associated with offering fixed annuities. For individual retirees the benefits of pooling may be available at substantially less cost than through the fixed annuity market. Furthermore, if investments perform above expectations, and longevity is adequately anticipated, then the extra return in a GSA scheme is returned to the participants, whereas for a fixed annuity (offered by a sponsor or purchased through the annuity market), any excess investment income would not increase benefits (at least, not directly). This upside opportunity may be an attraction for participants, and it has been suggested (for example, by Maurer, Mitchell, Rogalla, and Kartashov

(2013)) that GSAs could increase annuitization of retirement benefits, which is generally assumed to be below optimal levels (as proposed by, for example, Yagi and Nishigaki (1993)).

Previous research on GSA schemes covers a range of different designs. Piggott, Valdez, and Detzel (2005) derive closed-form expressions for the benefits coming from annuity pools, both closed and open to new participants. Using these results, Qiao and Sherris (2013) use simulations to highlight the risks of GSAs, especially when the group is closed, for retirees reaching high ages. They argue that there needs to be solidarity between younger and older pensioners to reduce longevity risk, and that the fund needs to be open to new pensioners to avoid volatility in payments and declining income at high ages. Nonetheless, Stamos (2008) demonstrates that pooled annuity funds can protect against longevity risk, even when the pool is small. He shows that a utility maximizing retiree will often favor pooled annuities when there is a charge for transferring mortality risk to an insurer; van de Ven and Weale (2009) obtain similar results. Donnelly, Guillén, and Nielsen (2013) also find that participants are willing to absorb the mortality risk in a pooled arrangement. They show that the expected return to the annuitants is higher and that the expected lifetime utility is also increased compared with a pooled fund where only the investment risk is pooled. However, they do not compare the value of a pooled fund compared with a regular fixed annuity product.

Independently, Sabin (2010) develops a similar product called *tontine annuity* and gives the condition under which the pool remains fair for all participants.

In this chapter, we assess the value of a GSA-type annuity within a retiree's portfolio. We note that variants of these schemes are offered by some North American employers, and are likely to become more popular with plan sponsors who are de-risking their pension benefits. For example, it is a feature of the University of British Columbia (UBC) pension plan[1]. Under the UBC version, the yearly amount of the annuity is computed based on an assumed mortality table and an assumed interest rate, which can be selected by the participant to be 4% or 7% per year. The group of retirees share the investment risk and the mortality experience. Every year the annuity payments are recomputed on the same valuation basis (4% or 7%) given the funds available, which depend on the investment return on the fund, the mortality experience and the cash paid out as annuity payments during the year.

Intuitively, this arrangement seems somewhat risky for the retiree, unless she has significant other stable income. The graph in Figure 6.1 plots the possible annual values of this type of annuity for 100 simulation paths when the retirement fund follows a geometric Brownian motion with parameters $\mu = 0.0583$ and $\sigma = 0.2$, and the plan assumed interest rate is $i = 0.04$. We see that for a small number of paths the annuitant does extremely well, but for a relatively large number of paths she fares poorly. We question the appropriateness of this profile for most retirees. The UBC plan results available for the period 1996 to 2013 show that the retirees selecting the

[1]See UBC Faculty Pension Plan (2013).

GSA option have had a volatile ride – their annuities fell by almost 20% in a single year during the financial crisis[2].

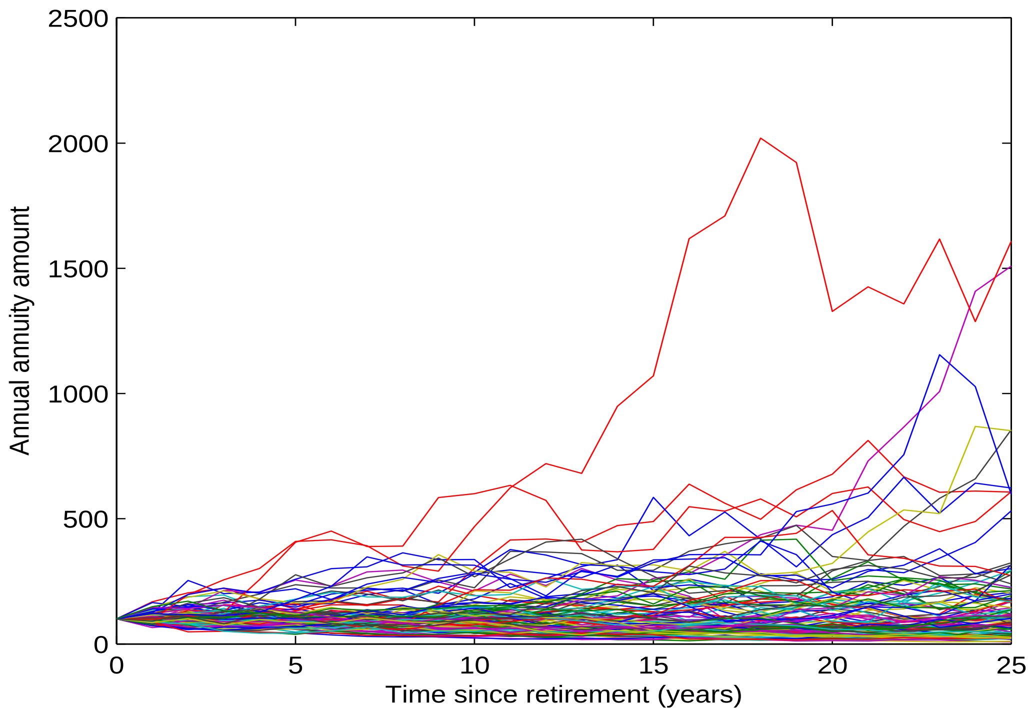


Figure 6.1: 100 simulated paths of the GSA. Initial annuity value is 100 pa, $\mu = 0.0583$, $\sigma = 0.2$, $i = 0.04$, mortality follows RP-2000 table (75% male), group of 100 retirees

However, some of the recent literature has proposed that GSAs have an important role in retirement portfolios. Maurer, Mitchell, Rogalla, and Kartashov (2013) consider the impact of stochastic mortality and investment risk together. If the insurer retains the mortality risk they label the annuity as non-participating and assume the insurer uses a quantile approach to charge for the mortality risk. The authors load the premium so that the solvency probability given a stochastic mortality model is very high. The so-called participating contracts pass the systematic mortality risk to the annuitants. The authors find, in the context of their optimal allocation approach, that the participating contracts are preferred by consumers, relative to the

[2]In 2009, the payments in the UBC plan were reduced by 17.45% for the 4% option, and by 19.8% for the 7% option.

non-participating contracts. This is consistent with similar analysis in Donnelly, Guillén, and Nielsen (2013).

Most authors who investigate retirement asset allocation seek to optimize the utility of the retiree's consumption. Horneff, Maurer, and Rogalla (2010) study optimal portfolio decisions with deferred annuities in a life-cycle and find that a significant portion of the wealth goes to buying deferred annuities. Hanewald, Piggott, and Sherris (2013) use simulations to analyse different portfolios that include life, deferred and inflation-indexed annuities, group self-annuitization and self-annuitization. They split mortality risk into idiosyncratic and systematic and consider fee loadings. When loadings are present, GSAs are preferred to other types of annuities, even inflation-indexed ones. However, the GSA group is assumed to be sufficiently large that all idiosyncratic mortality risk is diversified away, which is not assumed in our work.

In this chapter we use dynamic programming to obtain the optimal investment and consumption strategy for a retiree who has access to both a GSA-scheme (very similar to the one offered by the UBC pension plan) and a fixed whole-life annuity. To reflect the language in the UBC plan, we use the term *Variable Payment Life Annuity* for the GSA scheme. We also assume that the retiree can self-annuitize and we analyze the resulting income and annuity payments. We find that a utility maximizing retiree will invest a significant part of her wealth in the life annuity, even in the presence of a fee loading, where the objective is to maximize lifetime utility of consumption, assuming CRRA (constant relative risk averse) utility.

While most authors consider systematic mortality and longevity risk, in this chapter we limit ourselves to studying the effect of idiosyncratic mortality risk. This work is a first step towards a full-blown analysis of the VPLA, which should include longevity risk since it is an important factor in the development of group self-annuitization schemes. In addition, adding longevity risk would increase the riskiness of the product, and would further confirm our conclusions about the riskiness of the VPLA.

The CRRA utility optimization approach is used by many authors, including Stamos (2008), Maurer, Mitchell, Rogalla, and Kartashov (2013), Horneff, Maurer, and Rogalla (2010) and Donnelly, Guillén, and Nielsen (2013). However, it may not adequately model the financial risk management challenge faced by retirees. We extend our analysis to explore different criteria, in particular, the risk of a 50% drop in retirement income. We show that strategies which emphasize downside protection lead to payment patterns that may be more appropriate or realistic for retirees than the CRRA utility maximization approach.

Section 2 presents the GSA scheme. In Section 3, we describe the optimization problem. Numerical results for different sets of assumptions are presented in Section 4. Some assumptions are relaxed in Section 5 and Section 6 concludes.

6.2 Variable Payment Life Annuities

In this section, we introduce the Variable Payment Life Annuity (VPLA) product in more detail. This term is not widely mentioned in the literature, but it is the one used in the UBC plan, which is part of the motivation for this chapter. This product is very similar to the GSA described in Piggott, Valdez, and Detzel (2005). A VPLA is a life annuity with payments that vary depending on the performance of the fund relative to a fixed *assumed rate* i. In this work, we assume that it is a type of annuity offered to members of a pension plan. Thus, the evolution of the annuity payments depends on the performance of the pension fund, or of the sub-fund allocated to the VPLA.

We first consider the situation where N members of a pension plan retire at the same time. Assume there are no other retirees. Each retiree decides to allocate part of their wealth at retirement to a VPLA. Let (x_n) denote the n^{th} member of the group, for $n = 1, 2, \ldots, N$, and let $V_0(x_n)$ denote (x_n)'s initial assets allocated to the VPLA. Then, at time 0, the n^{th} retiree will receive an amount

$$L_0^V(x_n) = \frac{V_0(x_n)}{\ddot{a}_{x_n}}, \tag{6.1}$$

where $\ddot{a}_{x_n}$ is the present value of a life annuity-due to (x_n). One year later, the amount received by the retiree is denoted by $L_1^V(x_n) = L_0^V(x_n)(1 + j_0)$, where $1 + j_0$ is the adjustment factor for the first year. More generally, if (x_n) is still alive at time t,

$$L_t^V(x_n) = L_{t-1}^V(x_n)(1 + j_{t-1}) = L_0^V(x_n) \prod_{k=0}^{t-1}(1 + j_k). \tag{6.2}$$

The adjustment factor may take into account only the investment experience of the fund, or the investment and mortality experience. Note that the adjustment factor for a given year is the same for all the retirees in the group. In the next subsections, we explain how the adjustment factor is obtained in each case.

6.2.1 Adjustment factor without mortality

In this case, the only variable that contributes to the adjustment factor is the performance of the pension fund. There is no mortality risk pooling and each retiree can be considered individually. The adjustment factor from year t to $t + 1$ is given by

$$1 + j_t = \frac{1 + R_t^V}{1 + i},$$

where R_t^V is the return on the pension fund during year t. Thus, if the actual return is less than the assumed rate, the payment during year $t+1$ is less than the previous one.

6.2.2 Adjustment factor with mortality

When the adjustment factor takes mortality into account, the mortality experience of the whole retiree group affects the benefits of the survivors.

We begin by deriving the first adjustment factor $(1 + j_0)$. The total value of the pension fund at time 0 is

$$F_0 = \sum_{n=1}^{N} V_0(x_n) = \sum_{n=1}^{N} L_0^V(x_n)\ddot{a}_{x_n}.$$

Note this is the value of the fund before the annuity payments. The value of the fund after the annuity payments are made is

$$F_{0+} = \sum_{n=1}^{N}(V_0(x_n) - L_0^V(x_n)) = \sum_{n=1}^{N} L_0^V(x_n)\, a_{x_n}, \tag{6.3}$$

where $a_{x_n} = \ddot{a}_{x_n} - 1$. F_{0+} is the amount available for investment in the first year. The fund earns a return of R_0^V during the first year, so the value of the fund at $t = 1$, before paying the annuities is

$$F_1 = \left(\sum_{n=1}^{N} L_0^V(x_n)\ a_{x_n}\right)\ (1 + R_0^V), \tag{6.4}$$

For each (x_n), define the survival indicator function at t as

$$I_t(x_n) = \begin{cases} 1 & \text{if } (x_n) \text{ is alive at } t \\ 0 & \text{if } (x_n) \text{ dies before } t \end{cases} \tag{6.5}$$

We use conservation of value to obtain the payments at time one. The annuity payment to (x_n) at $t = 1$ is given by

$$L_1^V(x_n) = (1 + j_0)\ L_0^V(x_n)\ I_1(x_n),$$

where $1 + j_0$ is the adjustment factor during the first year. The fund at $t = 1$ must be sufficient to pay the adjusted annuities of the survivors, which gives the prospective equation

$$F_1 = \sum_{n=1}^{N} (1 + j_0)\ I_1(x_n)\ L_0^V(x_n)\ \ddot{a}_{x_n+1}$$

and this must be equal to the right hand side of equation (6.4), which is the retrospective fund value. Solving for j_0 gives

$$1 + j_0 = \frac{\left(\sum_{n=1}^{N} \; L_0^V(x_n) \; a_{x_n}\right) \; (1 + R_0^V)}{\sum_{n=1}^{N} \; I_1(x_n) \; L_0^V(x_n) \; \ddot{a}_{x_n+1}}.$$

Notice that we can write F_1 as

$$F_1 = \sum_{n=1}^{N} \; L_1^V(x_n) \; \ddot{a}_{x_n+1},$$

where some of the $L_1^V(x_n)$'s may be zero. The value of the fund at $t = 1$ after paying the survivors is

$$F_{1+} = \sum_{n=1}^{N} L_1^V(x_n) \, \ddot{a}_{x_n+1} - \sum_{n=1}^{N} L_1^V(x_n) = \sum_{n=1}^{N} L_1^V(x_n) \, a_{x_n+1}.$$

For subsequent years, we have a fund F_t at the start of the $(t+1)^{\text{th}}$ year. Proceeding as above, before the annuity payments at t we have

$$F_t = \sum_{n=1}^{N} \; L_t^V(x_n) \; \ddot{a}_{x_n+t}$$

and after paying the annuities at t we have

$$F_{t+} = \sum_{n=1}^{N} \; L_t^V(x_n) \; a_{x_n+t}$$

During the period $(t, t+1)$ the return on the fund is R_t^V. Hence the value of the fund at time $t+1$ before the annuity payments is

$$F_{t+1} = F_{t+} \, (1 + R_t^V) = \left(\sum_{n=1}^{N} \; L_t^V(x_n) \; a_{x_n+t}\right) \; (1 + R_t^V).$$

This must be sufficient to fund the future annuity payments starting from $t+1$, with value

$$F_{t+1} = \sum_{n=1}^{N} L_t^V(x_n) \, I_{t+1}(x_n) \, (1 + j_t) \, \ddot{a}_{x_n+t+1}. \tag{6.6}$$

We obtain j_t by equating the two expressions for F_{t+1}, giving

$$(1+j_t) \;= \frac{\left(\sum_{n=1}^{N} \; L_t^V(x_n)\; a_{x_n+t}\right)(1+R_t^V)}{\sum_{n=1}^{N} \;\; L_t^V(x_n)\; I_{t+1}(x_n)\;\; \ddot{a}_{x_n+t+1}}. \tag{6.7}$$

Notice that if we assume that everybody retires at the same age, x, say, then (6.7) simplifies to

$$\begin{aligned}(1+j_t) &= \frac{\sum_{n=1}^{N} \; L_t^V(x_n)}{\sum_{n=1}^{N} L_t^V(x_n)\; I_{t+1}(x_n)} \; \frac{a_{x+t}\,(1+R_t^V)}{\ddot{a}_{x+t+1}} \\ &= \left(\frac{\sum_{n=1}^{N} \; L_t^V(x_n)}{\sum_{n=1}^{N} L_t^V(x_n)\; I_{t+1}(x_n)}\right) p_{x+t} \left(\frac{1+R_t^V}{1+i}\right) \\ \Rightarrow (1+j_t) &= \frac{p_{x+t}}{p^*_{x+t}} \; \frac{1+R_t^V}{1+i}\end{aligned} \tag{6.8}$$

where p_{x+t} and i are the assumptions used for $\ddot{a}_x + t$, and p^*_{x+t} is the experienced survival rate, weighted by the annuity values. That is,

$$p^*_{x+t} = \frac{\sum_{n=1}^{N} L_t^V(x_n)\; I_{t+1}(x_n)}{\sum_{n=1}^{N} L_t^V(x_n)} \tag{6.9}$$

This form shows the two components of the adjustment factor. The first is linked to actual mortality and is greater than 1 if more people than expected die, weighted by the annuity payments. The second is greater than 1 if the actual return R_t^V is higher than the return assumed by the plan, i. Our Equation (6.8) is identical to Equation (4) from Piggott, Valdez, and Detzel (2005).

In this section, we have assumed that the group of retirees is closed to new entrants. While this assumption may be realistic in some cases, for example, if the administrators separate groups or cohorts or retirees into different funds with different adjustment factors, the VPLA plan that we discuss here would typically be open to new retirees. Pooling the younger and older cohorts reduces the volatility of the payments by increasing the diversification (see Qiao and Sherris (2013)). However, it would add to the complexity of our optimization problem and make it infeasible to solve using dynamic programming. So we first consider a closed, heterogeneous group. It is important to keep in mind that this leads to more volatile VPLA payments, especially in older ages.

6.3 The Optimization Problem

Now that we have introduced the VPLA product, we incorporate it in an optimization problem that models the decisions an individual must make through their retirement.

The main goal of this section is to use the standard construct of the annuitization literature (see, for example, Hainaut and Devolder (2006),Stamos (2008), Maurer, Mitchell, Rogalla, and Kartashov (2013), Horneff, Maurer, and Rogalla (2010), Donnelly, Guillén, and Nielsen (2013), and Milevsky and Young (2007)). Under this standard approach we assume that the consumption of each individual is entirely flexible and is one of the controls under a dynamic optimization (that is, an optimization involving payments and decisions at different times, which depend on the evolving underlying processes). The other control is the proportion of wealth invested in different asset types. The optimal values for the control variables are determined by maximizing the total expected utility of the consumption process, discounted by the retiree's subjective discount rate. The usual assumption, which we also adopt, is a utility function that has constant relative risk aversion (CRRA). CRRA utility uses a parameter of relative risk aversion $\gamma > 0$, and takes the form

$$U(c) = \begin{cases} \frac{c^{1-\gamma}}{1-\gamma} & \text{for } \gamma \neq 1, \\ \log(c) & \text{for } \gamma = 1. \end{cases}$$

CRRA utility is chosen partly for its tractability. However, it may not be the best choice for the annuitization problem. CRRA implies that utility depends on proportional changes in wealth, not on absolute values. Since (again, following the usual practice) we assume all individuals in the group have the same risk aversion parameter γ, we are assuming that an individual with a starting pension of \$20,000 has the same aversion to a 10% drop in income as an individual with a starting pension of \$200,000.

Although these assumptions may be questioned, we continue with them as a benchmark. It allows us to compare our results with other researchers' findings. It also gives us a strategy that we can test under more realistic constraints.

6.3.1 The model, assumptions and notation

We assume a retirement age of 65 and set the time of retirement at $t = 0$. We also assume all lives expire by age 120, which means we will project for 54 years.

A new retiree has wealth A_0 to divide between four assets: the money market, a balanced fund, a life annuity with fixed payments and a VPLA. After this initial decision, the retiree rebalances her non-annuitized wealth between the saving bonds and the balanced fund at the beginning of each year.

The proportions of initial wealth invested in balanced fund, the fixed annuity and the variable annuity are denoted by ω_B, ω_F and ω_V, respectively. The remaining wealth is invested in the money market.

Denote the non-annuitized (or liquid) wealth of the retiree at time t by W_t.

We slightly modify and extend the notation from the previous section and define L^F to be the annual income from the fixed annuity, and L_t^V to be the retiree's income at time t from the VPLA. The total annuity income at t is

$$L_t = L_t^V + L^F.$$

At $t = 0$, we obtain L^F and L_0^V by dividing the amount invested in each annuity by $\ddot{a}_{65}^F$ and $\ddot{a}_{65}^V$, respectively. We allow for different assumptions to be used when determining the initial payments for the fixed and variable annuities because the issuer of the fixed annuity typically incorporates margins for the retained investment and longevity risk. We assume that this margin is incorporated in the interest rate assumption for $\ddot{a}_{65}^F$. We denote this adjusted interest rate by i^F and define it by

$$i^F = i(1 - \lambda),$$

where i is the interest rate used for the VPLA annuity factor, $\ddot{a}_{65}^V$, and $\lambda \in (\,0, 1\,]$ is the interest margin parameter.

Thus, starting with an accumulated amount at retirement A_0, the annuity payments and liquid wealth at time 0, after investment decisions are made, are given by

$$\begin{aligned} L^F &= \frac{\omega_F A_0}{\ddot{a}_{65}^F} \\ L_0^V &= \frac{\omega_V A_0}{\ddot{a}_{65}^V} \\ L_0 &= L^F + L^V \\ W_0 &= A_0(1 - \omega_F - \omega_V) + L_0, \end{aligned}$$

At times $t = 1, 2, ..., 54$, the only investment decision that the retiree must make is how to divide her non-annuitized wealth, W_t, between the money market and the balanced fund. We denote by ω_t the proportion of the wealth invested in the balanced fund at t. Let $\boldsymbol{\omega}$ denote the set of portfolio control variables, $\{\omega_B, \omega_V, \omega_F, \omega_1, \omega_2, \ldots, \omega_T\}$.

Investments in the money market are assumed to earn the risk-free rate, r, and the return on the balanced fund in $(t, t + 1)$ is denoted R_t^B. Hence, the return on the non-annuitized wealth during the year starting at time t, denoted R_t, is given by

$$R_t = r + \omega_t(R_t^B - r).$$

After one period, the total liquid wealth, W_1, and the annuity income, L_1, are given by

$$\begin{aligned} L_1^V &= L_0^V(1 + j_0) \\ L_1 &= L_1^V + L^F \\ W_1 &= (W_0 - C_0)(1 + R_0) + L_1, \end{aligned}$$

where C_0 is the amount consumed at time 0 and $(1 + j_t)$ is the adjustment factor derived in the previous section.

To determine the adjustment factors we assume that a proportion α_V of the VPLA fund is invested in the risky asset, while the rest is in the risk-free asset. For $t = 0, 1, 2, ..., T - 1$, the total wealth W_t and the annuity income L_t evolve according to the following equations, where N_t denotes the number of survivors at t.

$$\begin{aligned} L_{t+1}^V &= \begin{cases} L_t^V(1 + j_t) & \text{if } N_t \in [2, 3, ..., N_0] \\ L_t^V \frac{1+R_t^V}{1+i} & \text{if } N_t = 1 \end{cases} \\ L_{t+1} &= L_{t+1}^V + L^F \\ W_{t+1} &= (W_t - C_t)(1 + R_t) + L_{t+1}, \end{aligned}$$

where R_t^V is the return on the VPLA fund from time t to $t + 1$. We assume that N_t follows a binomial with N_{t-1} trials and probability of success p_{x+t-1}, and that when there is only one retiree left in the plan, the payments are no longer adjusted to mortality experience. This is in line with Donnelly, Guillén, and Nielsen (2013). It keeps the payments from declining too much if there is only one person left in the plan, but will distort results at very high ages.

6.3.2 Solving the Optimization Problem

We assume a constant time-preference discount factor, β, for the retiree. The objective then is to find the control vectors $\boldsymbol{\omega}$, and $\boldsymbol{C}$, such that

$$E\left[\sum_{t=0}^{54} \beta^t U(C_t)\right] \tag{6.10}$$

is maximized. We assume that the retiree's utility is given by

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma}, \qquad \gamma > 1 \tag{6.11}$$

Since the utility function is time separable,[3] the optimization problem can be represented recursively by the following equations

$$H(t, W_t, L_t, N_t) = \max_{\omega_t, C_t} \{U(C_t) + E_t [\beta H(t+1, W_{t+1}, L_{t+1}, N_{t+1})]\} \tag{6.12}$$

$$\text{for } \; t = 1, 2, ..., T-1$$

$$H(0, W_0, L_0, N) = \max_{\omega_B, \omega_F, \omega_V, C_0} \{U(C_0) + E_0 [\beta H(1, W_1, L_1, N_1)]\} \quad \text{for } t = 0 \tag{6.13}$$

where $E_t[\cdot]$ is the expectation conditional on the information up to time t. This optimization can be solved recursively starting from the last period using dynamic programming. We describe this process in Appendix A. More general detail about the methodology is given in, for example, Pennacchi (2008).

6.4 Results of the Optimization Problem

In this section we show the optimal investment and consumption strategy during retirement under the maximized expected CRRA utility, using the optimization approach described above. Once the optimal controls are obtained, we use Monte Carlo methods to simulate the resulting consumption and annuity payments, assuming a retiree follows the optimal strategy. We analyze the distribution of the payments at different ages after retirement. The maximized utility approach combines all future payments into a single expected present value of future consumption, weighted by the utility function. It is interesting to look at the potential income paths generated by the optimal strategy. This allows us to get a better feel for the risks associated with the VPLA, even when utility is optimized.

As a simple measure of the adequacy of the resulting income flow to the retiree, we measure the probability that the annual consumption level drops below a certain percentage of the original level at retirement. This is important because we assume that this is the retiree's only source of revenue, and if she is no longer able to consume at the original level, her lifestyle may be affected. In this analysis, we set the "poverty threshold" at 50% of the initial consumption. That is, whenever the consumption in a given year is less than 50% of the consumption in the first year, we consider that the retiree has hit the poverty level.

[3] A utility function is time separable if the current utility only depends on current consumption and not on past consumption or expected future consumption.

Parameter	Value
Age at retirement x	65
Annuitization interest rate i	0.03
Risky asset lognormal distribution μ	0.04078
Risky asset lognormal distribution σ	0.18703
Risk free rate r	0.02
Percentage of VPLA fund invested in risky asset α_V	0.4
Time-preference discount factor β	0.96
Risk aversion parameter γ	5
Size of retiree group N_0	100
Mortality table	RP-2000 combined healthy

Table 6.1: Parameters used to obtain numerical results for the optimization problem

We first solve the problem with a set of assumptions similar to the one presented in Maurer, Mitchell, Rogalla, and Kartashov (2013), and then modify key assumptions to understand how they affect investment and consumption decisions.

6.4.1 Assumptions and Parameters

The parameters used in this section are presented in Table 6.1. Most of our parameters come from Maurer, Mitchell, Rogalla, and Kartashov (2013), as that enables us to benchmark our results against theirs. The parameters that differ are explained in this section. Note that the risk aversion parameter is higher than many researchers use – a more common value would be around 2 (see, for example Maier and Rüger (2010))[4]. Using $\gamma = 5$ indicates that we are assuming a very strong aversion to risk.

We also have to make an assumption about the investment choices in the VPLA fund. In Maurer, Mitchell, Rogalla, and Kartashov (2013), the retiree has control over the fund composition. However, we assume that the fund is managed by the plan sponsor and that the investment proportions are fixed. The parameter α_V was chosen to reflect the average results from Maurer, Mitchell, Rogalla, and Kartashov (2013), giving an expected return on the VPLA fund of 3.6%, which is slightly higher than the risk free rate used to determine fixed annuity prices. This assumption will be tested in Section 6.4.3.

[4]This is a significant difference. For example, an individual who risks losing 80% of their wealth, with a probability of 1%, would pay a premium of 40% of their wealth for full insurance with $\gamma = 5$, but only 4% of their wealth with $\gamma = 2$.

Additional parameters, such as the fee load applied to fixed annuities, do not appear in Table 6.1 as we will consider a range of values in our analysis.

6.4.2 Numerical Results of Utility Maximization

In this section, we present the results of the utility maximization problem introduced in the previous section. Once the optimal controls are obtained, we use 100,000 Monte Carlo simulations to investigate the optimal strategies. The balanced fund return and the number of survivor processes are each simulated 100,000 times. The first sample path for the returns is used with the first sample path for the survivors, and so on until the last one. Market and mortality risks are assumed to be independent.

We obtain our first results under the assumption that $\lambda = 0$, which means that the same annuitization rate and mortality tables are used to price both the fixed and variable annuities. Under this assumption, it is optimal to invest all the wealth at retirement in the fixed annuity (see Table 6.2 for a summary of the results). This result is intuitive – for the same price, a risk-averse retiree prefers an annual fixed payment to an uncertain one, even if both payments have similar expected values.

As explained previously, a plan sponsor would typically charge more to retain investment and longevity risk; if the plan sponsor does not offer a fixed annuity, the retiree would have to purchase the annuity from an insurer, and the margins for risk and profit would generally lead to substantially higher annuity prices compared with the VPLA rate. For this reason, we also performed the optimization problem for different values of the interest margin parameter λ. The optimal investment choices at retirement are summarized in Table 6.2. We describe the key results with these parameters.

- It is always optimal (using these assumptions) to invest all of the retirement funds in a combination of the fixed and variable annuity; that is, $\omega_B = 0$, $\omega_V + \omega_F = 1$.
- It is always optimal to consume the full annuity payment each year – that is, $C_t = L_t$ for all t.
- As the cost of the fixed annuity increases, the retiree optimizes the utility of her consumption by investing a greater part of her initial wealth in the VPLA. However, even when the interest margin parameter is very high ($\lambda = 0.5$ and $\lambda = 0.6$), the retiree should still invest over 20% of her wealth in the fixed annuity. This allows her to have a minimum guaranteed annual payment.
- When the interest margin parameter λ is greater than 0.2, the minimum guaranteed payment is less than half of the initial payment. In that case, there is a positive probability

that, throughout the life of the retiree, the annual payment drops below 50% of the initial level.

- Even when the fee load is high, the initial payments are quite similar. The distribution of the payments throughout retirement is however very different under the different values for λ.

λ	0	0.1	0.2	0.3	0.4	0.5	0.6
i^F	0.0300	0.0270	0.0240	0.0210	0.018	0.0150	0.012
$\ddot{a}_{65}^F$	13.9301	14.3064	14.7006	15.1136	15.5466	16.0010	16.4780
ω_V	0	0.12	0.39	0.60	0.71	0.77	0.77
ω_F	1	0.88	0.61	0.40	0.29	0.23	0.23
L_0	7.18	7.00	6.94	6.95	6.96	6.96	6.93
L_0^V	0	0.76	2.82	4.36	5.10	5.52	5.56
L_0^F	7.18	6.24	4.12	2.59	1.86	1.44	1.37

Table 6.2: Optimal investment at retirement for different interest margins λ, with associated initial annuity payments, as % of wealth at retirement.

We study in more detail the optimal strategy when $\lambda = 0.2$, since this is a plausible interest spread parameter. This leads to an annuity factor $\ddot{a}_{65}^F$ of 1.055 times the factor used to price the VPLA[5]. Using Monte Carlo simulation, we analyse the distribution of the annual payment throughout retirement, conditional on the retiree being alive at each age. The distribution illustrated in Figure 6.2 is summarized in Table 6.8 of Appendix B (column PS1). We see that for $\lambda = 0.2$, the optimal strategy is to invest a little over 60% of initial wealth in the fixed annuity. The resulting annuity has a fixed component of over 50%, meaning that the annual payment can never be less than half of the initial payment. This ensures a level of financial security for the retiree during a period of her life when she is particularly vulnerable. While the average and the median annual payment slowly decrease throughout retirement, there is also a possibility that mortality experience and investment returns could be favorable for the retiree, leading to increased payments. Nonetheless, by age 95, in over half of the cases, the annual payment will have dropped to below 91% of the first payment. In this case, since a significant part of the wealth remains in the fixed annuity, the retiree is always sure to receive a payment that is at least 59% of her first payment. Figure 6.2 shows a drop in the mean and the median average payment just after year 30. This is due to the fact that our distribution is conditional on the retiree still being alive at that age. Keeping one retiree alive drops the actual mortality experience below the mortality rate assumed by the plan, thus affecting the payments negatively. When all retirees in the group have died except for one, payments are only affected by investment returns.

[5]Note that this is lower than the factor of 1.1 assumed by Milevsky (2001).

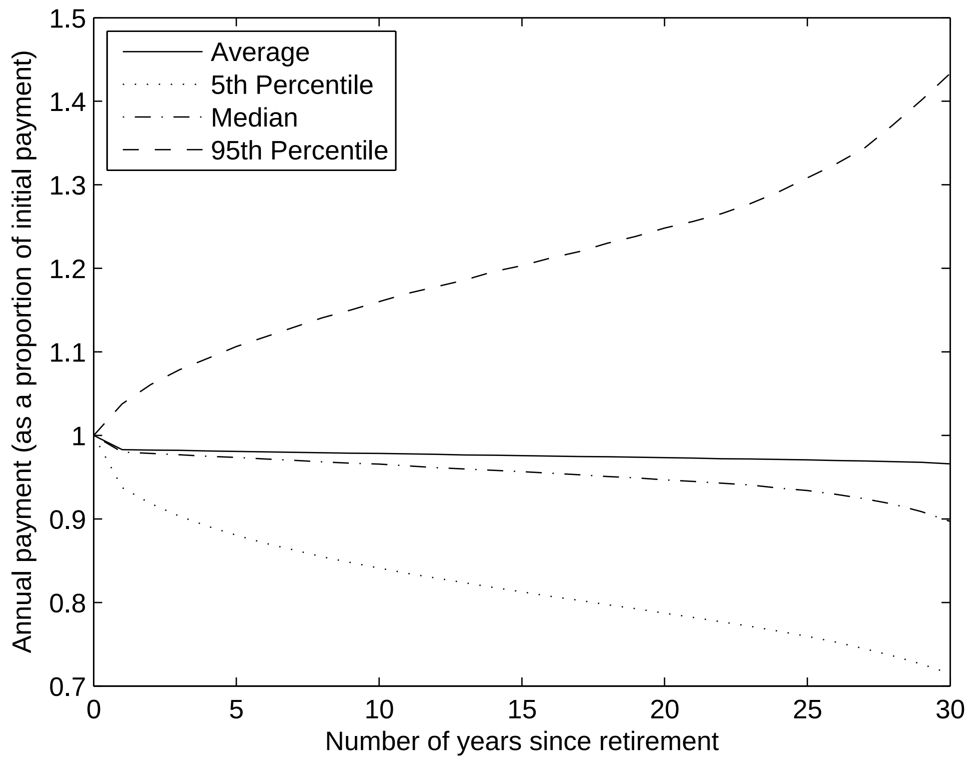


Figure 6.2: Distribution of the annual payment during retirement as a percentage of initial wealth, conditional on survival; $\lambda = 0.2$.

Next, we analyse the distribution of the annual payments when $\lambda = 0.3$. This means that the fixed annuity factor $\ddot{a}_{65}^{F}$ is 8.5% higher than the VPLA annuity factor, which is still a plausible fee load. Under these assumptions, the optimal strategy at retirement is to invest 40% of the wealth in the fixed annuity and 60% in the VPLA, and nothing in the balanced fund or money market. In this case the fixed annuity component contributes 37% of the first payment, so there is a possibility that the total payment could drop below half of the first payment, if the VPLA falls by more than around 65%.

Column PS2 of Table 6.8 summarizes the distribution presented in Figure 6.3. When $\lambda = 0.3$, a larger part of the wealth is invested in the VPLA, so the range of possible payments, especially at more advanced ages, is wider. The median also drops faster, but the average remains around the same when λ increases from 0.2 to 0.3. We can conclude that the investment strategy that maximizes utility leads the same level of consumption, *on average*, even when the price of the fixed annuity increases. However, the variability of the annual payments is increased, which significantly increases the risk that the retiree hits the poverty threshold during retirement. These probabilities are presented in column PS2 of Table 6.9. When $\lambda = 0.3$ and the retiree maximizes the utility of her consumption, she is not at risk before age 90, but she has a 5.32% probability of reaching the poverty threshold between ages 90 and 100. This appears unsatisfactory, since it coincides with a period of higher expected consumption needs (medical expenses, for example) and the nonagenarian is unlikely to have the opportunity to return to the workforce to supplement her income.

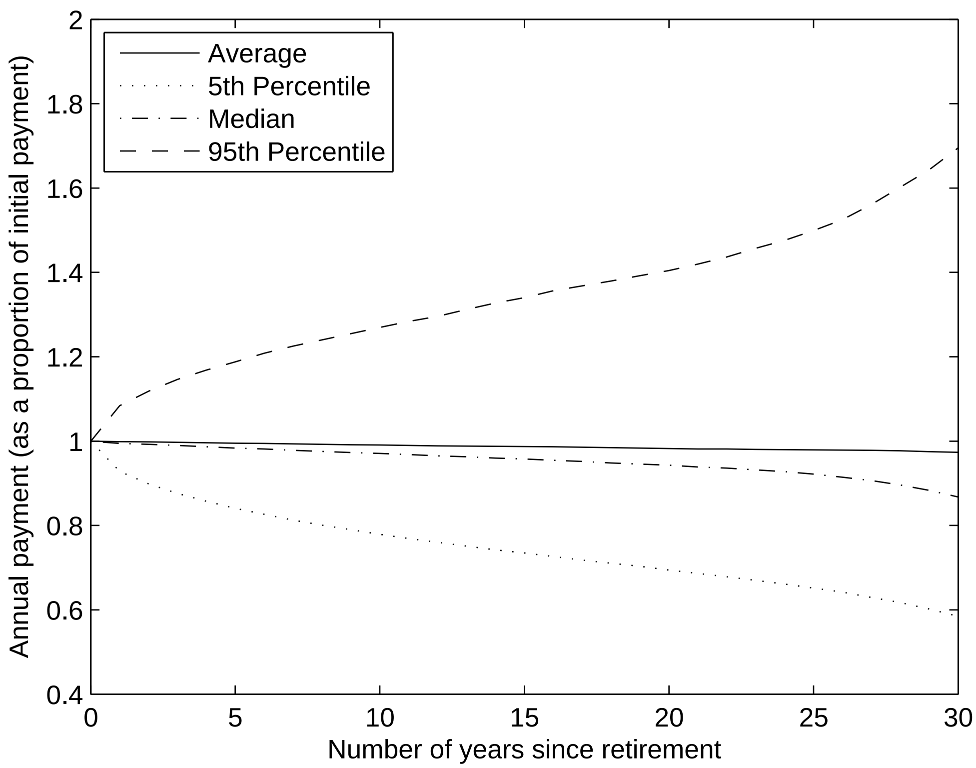


Figure 6.3: Distribution of the annual payment during retirement as a percentage of initial wealth, $\lambda = 0.3$.

To conclude this analysis, we keep $\lambda = 0.3$ and set ω_V and ω_F so that the probability of hitting the poverty level during retirement remains at 0. Thus, we set the fixed annuity part of the first payment to be worth half of the total amount. By letting

$$\frac{\omega_F}{a_{65}^F} = 0.5\left(\frac{\omega_V}{a_{65}^V} + \frac{\omega_F}{a_{65}^F}\right),$$

we get $\omega_F = 0.57$ and $\omega_V = 0.43$, which results in an initial payment of 6.88% of the initial wealth, a lower initial payment than when the retiree maximizes the utility of her consumption. However, since the fixed annuity makes up half of the initial payment, the retiree can never reach the poverty level. The distribution of the payments during retirement is summarized in column PS3 of Table 6.8. While the average payment is slightly lower when poverty level is avoided, this new investment strategy leads to higher median payments at ages 85 and 95. At all ages, the 95^{th} percentile is higher with the revised strategy, and this difference increases with time.

This result is interesting because it highlights the fact that simple utility maximizing strategies may lead to investment strategies that are too risky for retirees and that do not take their particular needs into account. By commuting the income stream into a single present value, upside opportunities can balance downside risk (though not symmetrically using risk averse utility). Thus, for example, using the parameters of this section, but with a CRRA parameter of $\gamma = 2$, a 65-year old life is deemed to be indifferent between consuming \$20,000 per year for life, and consuming \$31,000 per year for 20 years, followed by only \$5,000 per year after age 85.

Rather than focus on the expected utility, we focus here on maintaining a minimum level of income for life. This strategy might better reflect the risk preferences of retirees than the CRRA utility does. We note that behavioral science shows that most investors fear a decrease in income, which is reflected in the concept of *habit formation* (see MacDonald, Jones, Morrison, Brown, and Hardy (2013) and Pollak (1970)), and a strategy which offers lower initial payments, but with less chance of catastrophic reduction in income would be preferred over higher initial payments and/or higher upside potential, but incorporating the severe downside risk.

6.4.3 Exploring different assumptions

VPLA only

In this section, we explore the case where the retiree cannot invest in a fixed annuity. That is, if she wants to annuitize her wealth, she can only invest in the VPLA. Again, we use dynamic programming to obtain the optimal controls.

For this section, we use different assumptions for the annuitization rate i to reflect the terms offered in the UBC pension plan. We start by assuming $i = 4\%$. We find that the optimal investment and consumption strategy is to invest the full amount at retirement in the VPLA and to consume the total payment every year. The distribution of the annual payments is illustrated in Figure 6.4 and is summarized in column PS4 of Table 6.8. The range of possible payments in this case is a lot wider, and the average payment increases during the first 20 years of retirement. However, if the retiree survives to age 100, there is a 36.6% probability that her annual payment will eventually be less than half of the original payment, which is significant. Investing the full amount available at retirement in a VPLA and relying on this investment to provide a retiree's main, and often only, source of income is extremely risky and does not seem appropriate.

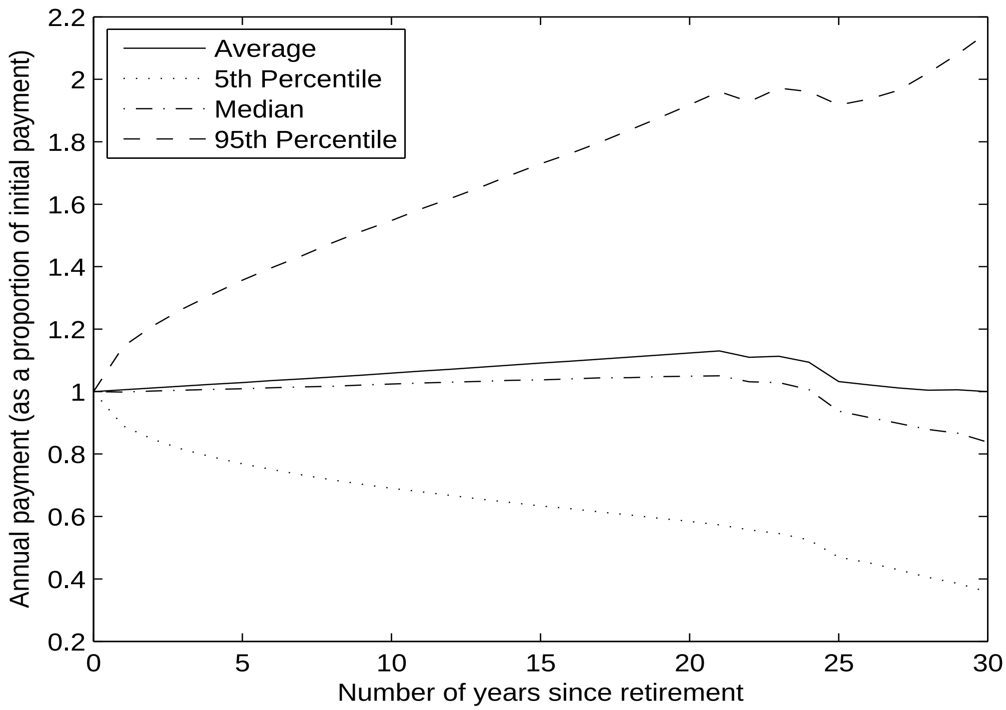


Figure 6.4: Distribution of the annual payment during retirement as a percentage of initial wealth when $\omega_F = 0$.

Pension plans may offer a VPLA option priced assuming a much higher interest rate. For example, a retiree in the UBC plan can choose a VPLA priced with a rate of 7%. Under this assumption, the initial annuity payment will be a lot higher than when $i = 4\%$. However, unless market performance meets the assumption, the annual payment will decrease steadily. This only appears to be an interesting option for someone with a significantly shorter life expectancy than the rest of the group. Here, we study the optimal investment and consumption strategies when $i = 7\%$ (and using the same mortality assumptions, with no adjustment for adverse selection). In this case, we assume that the VPLA fund is fully invested in the risky asset ($\alpha_V = 1$) to increase the average return. The resulting consumption strategy is interesting: in the first years, when annuity payments are high, the retiree should not consume all the payment. She should save it to complement her VPLA income, which will most likely decrease later in her retirement. Column PS4 of Table 6.9 summarizes the distribution of the consumption at different ages. The optimal strategy allows the mean consumption to increase until age 95. However, the range of possible scenarios is very wide, which translates into a risky strategy. In the worst 5% of the cases, the annual consumption drops to 60% of the initial consumption only 10 years after retirement, and fails to recover in subsequent years. Figure 6.5 illustrates the evolution of the non-annuitized (liquid) wealth during retirement. When market performance and mortality experience is favorable, she is able to re-build the wealth she had at retirement. However, in the worst cases the saved annuity payments do not protect her from decreasing consumption, potentially to less that one-third of the initial consumption (and unfavorable tax treatment could make the situation even worse).

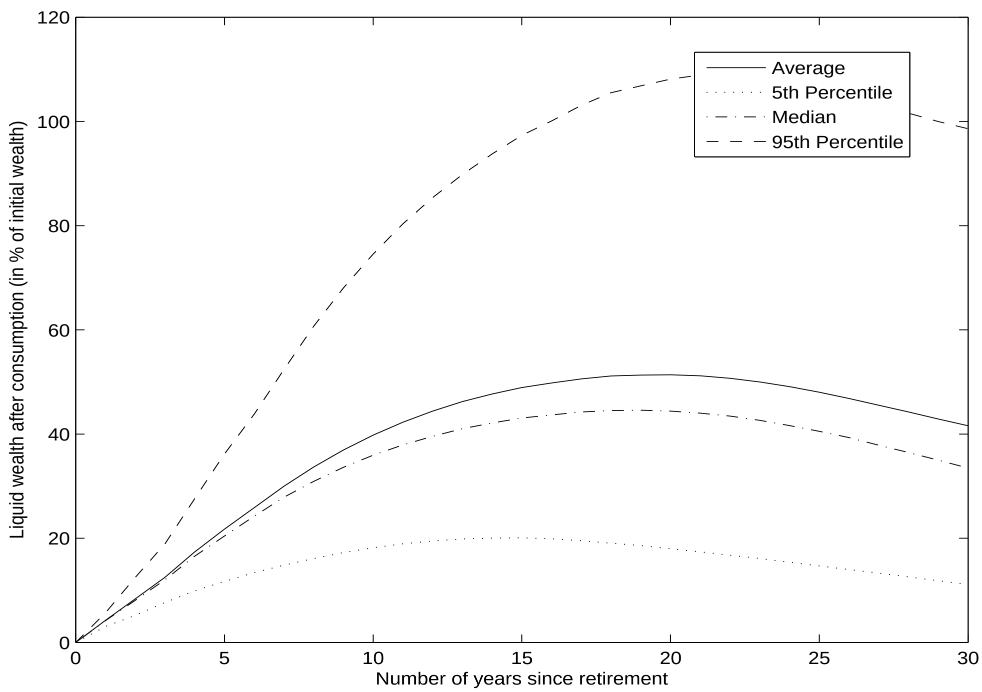


Figure 6.5: Distribution of the annual liquid wealth during retirement as a percentage of initial wealth when $\omega_F = 0$, $i = 0.07$.

Changing the risk aversion parameter

Here we consider the optimal investment choices for a retiree with a lower risk aversion parameter ($\gamma = 2$). As noted above, Maurer, Mitchell, Rogalla, and Kartashov (2013) use $\gamma = 5$, but according to Maier and Rüger (2010) this is on the high side of empirical estimates. Harrison, List, and Towe (2007) use a more middle of the road assumption that $\gamma = 2$. The results in column PS6 of Table 6.8 show that when there is no additional fee, the fixed annuity is still preferred. However, as soon as the fee load λ is above 0.2, the retiree maximizes the utility of her consumption by investing all of her accumulated wealth in the VPLA. This strategy results in the payment distribution summarized in colum PS6 of Table 6.8.

While this strategy maximizes the utility of the retiree's consumption, it results in very risky payment patterns, given that we are assuming that this is the retiree's only source of income. There is a significant probability that the annual payments become insufficient to meet her needs. In fact, the probability that the annual payment decreases below 50% of the initial payment before age 100 is almost 40%. With an interest margin of $\lambda = 0.2$ on the fixed annuity, she could buy a fixed annuity that pays her 6.896% of her initial wealth each year, which represents 96% of the initial VPLA payment. There is a very high probability (94.22%) that the VPLA payment drops below that amount before the retiree reaches age 100. Thus, the higher initial payment and the possibility of increased income provided by the VPLA may not be sufficient to justify investing in the VPLA, even when the utility is maximized.

λ	0	0.05	0.1	0.15	0.2	0.25
i^F	0.0300	0.0285	0.0270	0.0255	0.0240	0.0225
$\ddot{a}^F_{65}$	13.9301	14.1161	14.3064	14.5012	14.7006	14.9046
ω_V	0	0	0.27	0.66	1	1
ω_F	1	1	0.73	0.34	0	0
L_0	7.18	7.08	5.10	2.34	0	0
L^V_0	0	0	1.94	4.74	7.18	7.18
L^F	7.18	7.08	7.04	7.08	7.18	7.18

Table 6.3: Optimal investment with associated initial annuity payments (as % of wealth at retirement), at retirement for different interest margins λ, when $\gamma = 2$.

Changing group size

In the results above we have assumed a group size of 100. This group size is reasonable if the retirees are separated into cohorts. Better longevity risk pooling can however be accomplished when the retiree group is larger. In this section, we consider $N_0 = 200$ and observe the effect on the optimal consumption levels when $\lambda = 0.2$ and $\lambda = 0.3$. We return to the original assumption with respect to the risk aversion parameter ($\gamma = 5$).

When $\lambda = 0.2$, the retiree should invest 55% of her wealth in the fixed annuity and the rest in the VPLA (compared to 61% when $N_0 = 100$), confirming that a larger retiree group makes the VPLA more attractive. Column PS7 of Table 6.8 gives an insight into the distribution of the annual payment at ages 75, 85 and 95. This distribution is similar to the one obtained with the same assumptions for a group size of 100. The mean and the median are slightly higher at all three ages. The difference between the 5th and the 95th percentiles, smaller in the case of the larger group, shows that there is a better longevity risk pooling in this case. The optimal strategy is thus less risky, even if a larger percentage of the initial wealth is invested in the VPLA.

If the fee load is increased, at $\lambda = 0.3$, the retiree maximizes her utility by investing 35% of her wealth in the fixed annuity (compared to 40% when $N_0 = 100$). Column PS8 of Table 6.8 summarizes the distribution of the payments and clearly shows that the optimal investment strategy is less risky when the retiree group is larger. However, the 95^{th} percentile does not increase as much over time, thus decreasing the probability of very high payments. Column PS8 of Table 6.9 confirms that a larger group better protects the retiree against the probability of hitting the poverty threshold during retirement, consistent with the results of Hanewald, Piggott, and Sherris (2013).

Changing market assumptions

Throughout this analysis, we have assumed that 40% of the VPLA fund was invested in the risky asset. Now we change this assumption and choose α_V so that the expected return on the VPLA fund is equal to the annuitization rate i, which gives us $\alpha_V = 0.25$. The VPLA fund returns are less volatile than under the previous assumption, so the payment pattern is less risky. However, retirees do not benefit as much from market performance, although they still benefit from mortality risk pooling.

The optimal investment choices when $\alpha_V = 0.25$ are summarized in Table 6.4. When $\lambda = 0.2$, it is optimal for a retiree to invest over 81% of her portfolio in the fixed annuity. This is a much higher proportion than when $\alpha_V = 0.4$. Since the expected value of the investment portion of the VPLA adjustment factor, $((1 + R^V)/(1 + i))$, is equal to 1.0, the retiree does not have an expectation of investment gains in the VPLA. In addition, the mortality part of the adjustment factor makes the payments more volatile. Thus, when the fee loading is not too high, investing in the VPLA is not worth it. The payment pattern resulting from the optimal investment strategy when $\lambda = 0.2$ is given in column PS9 of Table 6.8. The payment distribution is a lot less risky, but there is less upside potential. Furthermore, the initial payment of 6.87% of the retirement wealth is lower for lower values of α_V. This shows that if a plan sponsor wants to offer this sort of GSA scheme, the expected return on the pooled fund should be higher than the annuity interest rate if they want the VPLA to be more attractive than the fixed annuity. Of course, if the interest margin is higher on the fixed annuity, the VPLA is attractive even without any incentive from the equity market, since it offers a certain protection against longevity risk at no additional cost, compared with self-annuitization.

λ	0	0.05	0.1	0.2	0.25	0.3
i^F	0.0300	0.0285	0.0270	0.0240	0.0225	0.021
$\ddot{a}_{65}^F$	13.9301	14.1161	14.3064	14.7006	14.9046	15.1136
ω_V	0	0	0	0.19	0.39	0.58
ω_F	1	1	1	0.81	0.61	0.42
L_0	7.18	7.08	6.99	6.87	6.89	6.94
L_0^V	0	0	0	1.36	2.80	4.16
L^F	7.18	7.08	6.99	5.51	4.09	2.78

Table 6.4: Optimal investment at retirement, with associated initial consumption in % of fund, for different interest margins λ when $\omega_V = 0.25$.

6.5 Optimal Strategies in an Open Group of Retirees

So far, we have assumed that the VPLA payments are based on a closed group of retirees. In the present section, we relax this assumption. Using an open retiree group should decrease the volatility of payments and eliminate the downward trend at very high ages, as noted in Qiao and Sherris (2013).

We perform this analysis using Monte Carlo simulations and the adjustment factor given by Equation (6.8)[6]. We assume that when the retiree enters the group at age 65 (we call this time 0), there are 82 retirees in the group, distributed as outlined in the first two columns of Table 6.5. The starting amount for a retiree aged x, $x > 65$ is given by

$$L_0(x) = L_0^V(65)\left(\frac{E[1+R_t]}{1+i}\right)^{x-65}. \tag{6.14}$$

In other words, we assume that at time 0, a retiree receives the average amount she would have received if she started in the plan at age 65 with the same amount and investment strategy as

[6] The adjustment factor we use differs from the one suggested in Piggott, Valdez, and Detzel (2005) and fails to satisfy the four criteria presented in their work.

the new retirees at time 0, and if mortality experience followed the assumptions. Column 3 of Table 6.5 shows an example of the starting amounts when the assumptions follow parameter set 1 (PS1) from Table 6.7. We assume that each year 3 new retirees join the group at age 65, with the same initial amount and the same investment strategies as the previous retirees. The other assumptions are the same as in the previous section (outlined in Table 6.1). In the Monte Carlo simulation, each participant is considered individually. Each year, each participant stays alive with the appropriate probability from the mortality table RP-2000 combined healthy. Each retiree's pension amount is tracked and is used in the calculation of the adjustment factor.

Age (x)	Number	$L_{x,0}$ (under PS2)
65	20	7.18
70	19	7.39
75	16	7.61
80	13	7.83
85	9	8.06
90	4	8.30
95	1	8.55

Table 6.5: Composition of retiree group at $t = 0$ when $\omega_V = 0.39$.

To compare with the closed group, we consider the case where $\lambda = 0.2$ (so that $\ddot{a}^F_{65} = 1.055\,\ddot{a}^V_{65}$) and apply the optimal investment strategy derived for the closed group (i.e. we use PS2 from Table 6.7). Under the same assumptions, we also increase the proportion of the initial wealth invested in the VPLA. Results are summarized in Table 6.6.

By comparing columns 1 and 2 of Table 6.6, we find that for the same investment strategy ($\omega_V = 0.39$), the mean and the median of the total payment are higher when the group is open than when it is closed. They also increase as the retiree ages. New, younger retirees entering the group make the payments more stable and keep them from decreasing through time. The open group also increases the 5^{th} and 95^{th} percentiles, thus improving the financial security of the annuitant.

The last column of Table 6.6 displays the statistics of the distribution of payments when all the wealth is invested in the VPLA. In this case, the resulting payment pattern is less risky than when the group is closed, confirming the role of the new entrants in stabilizing the annuity payments. Columns 2, 3 and 4 show the distribution of the payments resulting from other investment strategies when the wealth is split between the fixed annuity and the VPLA. While there is added risk to investing a larger proportion of wealth in the VPLA, the resulting payments are not as volatile as when the group is closed.

Figures 6.6 and 6.7 illustrate the distribution of payments when $\omega_V = 0.39$ and $\omega_V = 1$, respectively.

ω_V	0.39	0.39	0.5	0.6	0.7	1
ω_F	0.61	0.61	0.5	0.4	0.3	0
Group Type	Open	Closed	Open	Open	Open	Open
l_0	6.9492	6.9492	6.9906	7.0282	7.0658	7.1787
Mean consumption, % of L_0						
Age 75	102.44	99.41	103.07	103.69	104.30	106.09
Age 85	104.90	98.88	706.18	107.65	108.72	112.12
Age 95	107.20	98.13	108.92	110.97	113.13	117.83
Median						
Age 75	100.96	98.12	101.12	101.46	101.67	102.36
Age 85	101.86	96.18	102.38	102.96	103.43	104.88
Age 95	102.69	91.13	103.14	104.03	105.16	107.34
5th Percentile						
Age 75	87.34	85.45	83.85	80.67	77.74	68.57
Age 85	83.56	79.97	79.11	74.96	71.09	59.40
Age 95	80.71	72.61	75.60	70.76	66.61	54.07
95th Percentile						
Age 75	122.54	117.86	128.63	134.28	139.79	155.96
Age 85	136.52	126.81	146.24	156.36	164.94	189.81
Age 95	148.63	145.62	162.77	174.81	188.01	221.55

Table 6.6: Statistics of the distribution of the annual payment as a percentage of the initial consumption for different investment strategies.

6.6 Conclusion

In this chapter, we used dynamic programming to obtain the optimal investment and consumption strategy for a retiree whose choices at retirement include a GSA scheme, a fixed annuity and self-annuitization.

We found that a mix of the GSA scheme and the fixed annuity are preferred over self-anuitization. We tested the sensitivity of the investment choices to changes in the different parameters. One of our main conclusions is that in most cases, a reasonably risk-averse retiree should invest part of her portfolio in a fixed annuity. We also confirm that the fees charged by insurance companies to take on longevity and investment risks make GSA schemes more attractive. We demonstrate that risk is better pooled in a bigger retiree group, and that opening the group leads to less volatile payment patterns.

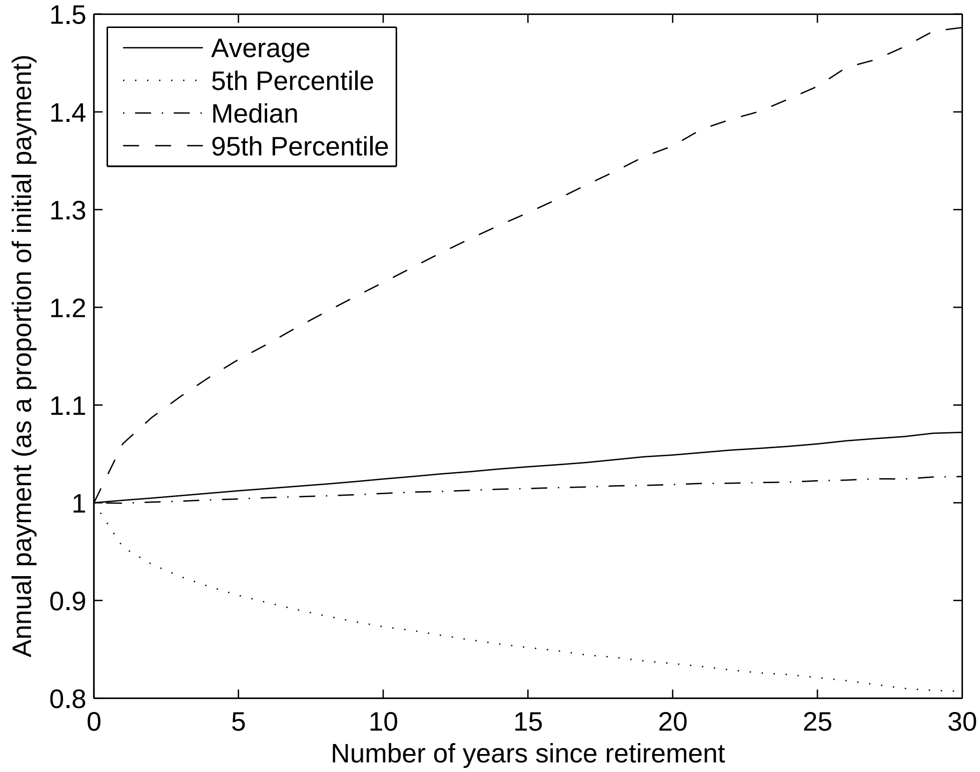


Figure 6.6: Distribution of the annual payment during retirement as a percentage of initial wealth when $\omega_V = 0.39$ and the retiree group is *open*.

We show that utility maximization can lead to suboptimal strategies from a cashflow perspective, and that other criteria, such as keeping the income above a certain threshold throughout retirement, should also be used to assess investment strategies.

The present analysis does not take systematic mortality risk into account. However, uncertainty about future mortality is one of the main reasons why insurance companies charge a premium for fixed annuities. Adding mortality improvements to our model without changing the design of the product would have caused payments to be riskier and to decrease over longer horizons. Nonetheless, even without systematic mortality risk, we show that fixed annuities still have a place in a retirement portfolio. Some GSA schemes have payments that are adjusted to changes in mortality assumptions. Future work could aim to analyse optimal retirement strategies in the presence of such products, using models that include mortality improvements. Such

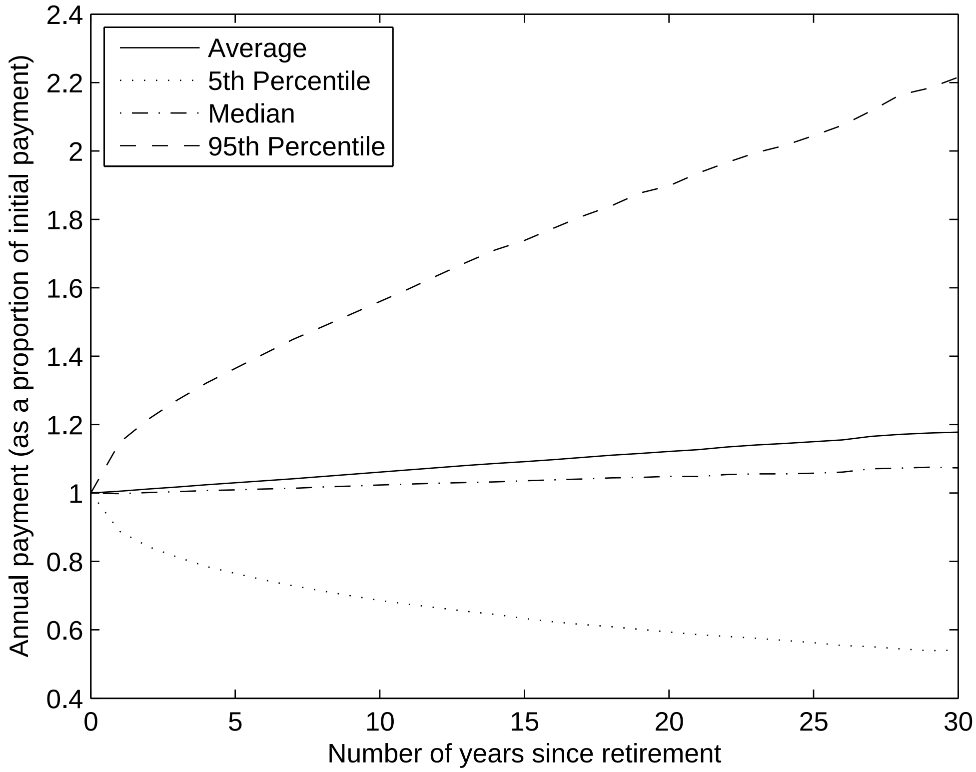


Figure 6.7: Distribution of the annual payment during retirement as a percentage of initial wealth when $\omega_V = 1$ and the retiree group is *open*.

analysis could also be performed on larger groups.

Appendix

6.A Solving the optimization problem using dynamic programming

In this section, we explain in greater detail how to solve the optimization problem through dynamic programming. Our optimization problem has four state variables: W_t, L_t^V, L^F and N_t. However, to illustrate the method, we assume only one state variable here, omitting the annuity payments L^V and L^F. The technique presented can easily be extended to higher dimensions.

Suppose that $W_t = (W_{t-1} - C_{t-1})((1+r) + \omega_{t-1}(R_{t-1}^B - r)))$ and consider the objective function

$$H(0, W_0) = \max_{\omega_0, C_0} \{U(C_0) + E_0 \left[\beta H(1, W_1)\right]\}, \tag{6.15}$$

where ω and C are the controls we want to solve for. More generally, let

$$H(t, W_t) = \max_{\omega_t, C_t} \{U(C_t) + E_t \left[\beta H(t+1, W_{t+1)}\right]\}. \tag{6.16}$$

Notice that the function $H(t, W_t)$ is always the maximized future discounted expected utility. We assume that the utility function is time-separable. In other words, the optimal consumption at a given time is independent of past consumption except through the process W_t. This allows us to treat each period, recursively, from end to start. At a given time t, in the one-variable problem, the optimal controls are only dependent on W_t. In other words, we can construct a set of values for W_t, together with the optimal values for the control variables given W_t.

However, we cannot entirely solve the problem at each t since we do not know the value of the function $H(t+1, W_{t+1})$. Generally, it is only possible to write this function in analytical form at time T, that is, $H(T, W_T) = 0$ for any W_T, as it is assumed all lives have died by T, and there is no bequest motive. Since this is the only known value for the derived utility function, we begin our optimization from the second-to-last period $T-1$. At that time, given our assumption that no lives survive to T, it is optimal to consume all remaining wealth, giving

$$H(T-1, W_{T-1}) = \max_{\omega_{T-1}, C_{T-1}} \{U(C_{T-1}) + E_{T-1}[H(T, W_T)]\} = U(W_{T-1}) \tag{6.17}$$

Since we cannot know the value of the wealth process at time $T-1$ (it will depend on the optimal controls during periods 1 to $T-1$), we calculate and store the derived utility function $H(T-1, W_{T-1})$ for a range of different values of W_{T-1}. These are chosen to represent the range of feasible values for W_{T-1}. Then, we move back one period and, again, for a range of values of W_{T-2}, solve for the optimal controls ω_{T-2} and C_{T-2} that will maximize

$$U(C_{T-2}) + E_{T-2}\left[\beta\, H(T-1, W_{T-1})\right] = U(C_{T-2}) + \beta\, E[U(W_{T-1})]$$

However, this time, we only know $H(T-1, W_{T-1})$ for the selected discrete values that we used for W_{T-1}. Our candidate controls ω_{T-2} and C_{T-2} will most likely not return one of the values W_{T-1} for which we have calculated $H(T-1, W_{T-1})$. Thus, we have to interpolate from the values we know to approximate the derived utility function for any value W_{T-1}. This will allow us to obtain the optimal controls at time $T-2$. The same procedure is repeated until we find the controls at time 0.

Here is the algorithm that is followed to obtain the optimal controls for a problem with T periods, using n discrete values for each W_t.

1. Build a grid of values of W_t at which the derived utility function will be calculated. This grid will have n rows and T columns. Each column represents a vector of possible wealths at a given time.

2. Build another grid of the same size to store the values of $H(t, W_t)$. Fill the last column with zeros, since we assume no bequest function.

3. Build two other grids of the same size to store the optimal values of ω and C_t at each time, for different wealths.

4. For each column $t = T-1$ to 1, apply the following to each element $i = 1$ to n of the column:

   (a) Given wealth W_t^i, find the optimal controls ω_t^i and C_t^i. Note that the function to optimize will use interpolation to calculate the value of the derived utility function one period later.

   (b) Store the optimal controls and the derived utility in the corresponding grid.

5. Now the grids are filled out and the first period needs to be solved.

6. Given wealth W_0, find the optimal controls ω_0 and C_0. Again, the function to optimize will use interpolation to calculate the value of the derived utility one period later.

To apply this method to our optimization problem, we need to extend it to the case where there are three state variables. Hence, instead of having a vector of values W_t and its associated vector J_t at each time t, we have a four-dimensional array with values W_t, L_t^V, L^F and N_t at each time t (denote by n_W, n_{LV}, n_{LFA} and n_N the number of values of W_t, L_t^V, L_t^F and N_t that are considered, respectively). The interpolation that needs to be performed to solve the problem at each data point is thus 3-dimensional. This method extends quite easily to multiple dimensions. However, the number of data points at which the derived utility function must be calculated is multiplied ($n_W \times n_{LV} \times n_{LFA} \times n_N$ instead of n), and the interpolation can become computationally burdensome.

6.A.1 Simplifying the optimization problem by normalizing

The normalization described in this section was inspired by Hubener, Maurer, and Rogalla (2013).

The optimization results shown in Section 6.4 were simplified from the Bellman equation, (6.12) by normalizing with respect to W_t, which reduced the number of state variables from four to three. That is, instead of working with the variables L_t^V, L^F, and C_t, we use the normalized variables

$$l_t^V = \frac{L_t^V}{W_t}, \quad l_t^F = \frac{L^F}{W_t}, \text{ and } \quad c_t = \frac{C_t}{W_t}. \tag{6.18}$$

This simplification is possible because the optimization problem is homothetic in wealth. This effectively means that the level of wealth does not impact the utility maximizing strategy, so that working with l_t^F, l_t^V and c_t gives the same results as working with W_t, L_t^V, L^F and C_t for any W_t. Intuitively, this seems reasonable given the nature of CRRA utility. However, we demonstrate more formally here.

We need to show that

$$H(t, W_t, L_t^V, L_t^F, N_t) = W_t^{1-\gamma} h(t, l_t^V, l_t^F, N_t) \tag{6.19}$$

for some function $h(.)$.

We show this by backwards induction. To make the proof easier to read, we will omit the arguments in H and h other than the time variable.

At T, $H(T) = 0$, so the result holds trivially. In the penultimate period, we have

$$\begin{aligned} H(T-1) &= U(W_{T-1}) \\ &= \frac{W_{T-1}^{1-\gamma}}{1-\gamma} \\ &= W_{T-1}^{1-\gamma}\, h(T-1), \end{aligned}$$

where $h(T-1) = 1/1-\gamma$.

Now, assume that for some $t+1$, $1 \leqslant t+1 \leqslant T$ the result in equation (6.19) holds. Then consider the function at t.

$$\begin{aligned} &H(t) = \max_{\omega_t, C_t} \{U(C_t) + E_t\,[\beta\, H(t+1)]\} \\ &\text{and } U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma} = \frac{(W_t\, c_t)^{1-\gamma}}{1-\gamma} = W_t^{1-\gamma}\, U(c_t) \\ &\Rightarrow H(t) = \max_{\omega_t, c_t} \left\{W_t^{1-\gamma} U(c_t) + E_t\left[\beta\, W_{t+1}^{1-\gamma} h(t+1)\right]\right\} \text{ (using the inductive hypothesis).} \\ &\text{Now } W_{t+1} = (W_t - C_t)\,(1+R_t) + I_x(t+1)\left(L_{t+1}^V + L^F\right) \end{aligned}$$

where $I_x(t+1)$ is the survival indicator function for (x) at $t+1$. Recall that

$$L_{t+1}^V = L_t^V\,(1+j_t). \tag{6.20}$$

Then

$$\begin{aligned} &W_{t+1} = W_t\left((1-c_t)(1+R_t) + I_x(t+1)\left(l_t^V\,(1+j_t) + l_t^F\right)\right) \\ &\Rightarrow H(t) = \max_{\omega_t, c_t} \Big\{W_t^{1-\gamma} U(c_t) \\ &\qquad\qquad + E_t\left[\beta\, W_t^{1-\gamma}\left((1-c_t)(1+R_t) + I_x(t+1)\left(l_t^V\,(1+j_t) + l_t^F\right)\right)^{1-\gamma}\ h(t+1)\right]\Big\} \\ &\Rightarrow H(t) = W_t^{1-\gamma}\, h(t) \quad \text{where} \\ &h(t) = \max_{\omega_t, c_t} \left\{U(c_t) + E_t\left[\beta\ \left((1-c_t)(1+R_t) + I_x(t+1)\left(l_t^V\,(1+j_t) + l_t^F\right)\right)^{1-\gamma}\ h(t+1)\right]\right\} \end{aligned}$$

Note that, as required, $h(t)$ is a function of l_t^V, l_t^f, but not of W_t, L_t^V or L^F.

Using this normalization we can perform the optimization problem using dynamic programming with three-dimensional grids. We normalize further by using

$$l_t = l_t^V + l_t^F \quad \text{and} \quad \rho_t = \frac{l_t^V}{l_t}.$$

This way, both continuous state variables exist on the interval $[0, 1]$ and all combinations are attainable.

To accelerate the computation, we use the method described in Section 5.1 of Carroll (2011) to calculate the expectation of functions of lognormal random variables. In this method, the lognormal distribution is discretized and the integral is approximated by a sum, in which each term represents an interval of equal probability.

To obtain the results presented in Section 6.4, we discretized the space of state variables (l_t, p_t, N_t) over a grid of size $6 \times 16 \times 9$. Larger grid sizes were explored, with similar results.

6.B Parameter sets and associated numerical results

Parameters	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9
λ	0.2	0.3	0.3	-	-	0.2	0.2	0.3	0.2
ω_V	0.39	0.6	0.43	1	1	1	0.45	0.65	0.19
ω_F	0.61	0.4	0.57	0	0	0	0.55	0.35	0.81
α_V	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.25
i	0.03	0.03	0.03	0.04	0.07	0.03	0.03	0.03	0.03
γ	5	5	5	5	5	2	5	5	5
N_0	100	100	100	100	100	100	200	200	100

Table 6.7: Definition of parameter sets (PS).

Parameter Set	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9
l_0	6.9492	6.9538	6.8583	8.4804	11.1479	7.1787	6.9718	6.9820	6.8739
Mean consumption, % of L_0									
Age 75	99.41	99.08	99.70	96.58	70.51	98.72	99.27	99.07	99.12
Age 85	98.89	98.23	99.01	102.49	96.13	97.46	98.63	98.17	98.29
Age 95	98.13	97.34	98.26	91.20	93.26	96.17	98.07	97.27	97.50
Median									
Age 75	98.12	97.09	98.07	93.41	60.35	95.47	97.79	96.87	98.83
Age 85	96.18	94.29	95.75	95.65	74.50	91.29	95.84	94.10	97.68
Age 95	91.13	86.74	89.95	76.43	64.72	83.52	92.28	88.79	94.95
5th Percentile									
Age 75	85.45	77.93	82.59	62.96	25.66	64.79	83.88	76.67	94.71
Age 85	79.97	69.44	75.79	53.31	23.38	51.94	78.03	68.15	92.04
Age 95	72.61	58.46	66.71	32.69	17.23	37.77	71.50	58.58	87.69
95th Percentile									
Age 75	117.86	126.96	122.52	141.17	149.36	143.61	119.70	128.78	104.54
Age 85	126.81	140.46	133.33	174.98	240.60	163.84	128.83	141.87	106.60
Age 95	145.62	169.53	156.68	196.13	259.71	197.17	143.48	163.68	115.19

Table 6.8: Statistics of the distribution of the annual payment as a percentage of the initial payment for different sets of parameters.

$Pr(L_t < 0.5\ L_0)$	PS2	PS4	PS8
Before age 80	0	0.0070	0
Before age 90	0	0.0550	0
Before age 100	0.0532	0.3660	0.0504
Before age 120	0.3562	0.5633	0.4381

Table 6.9: Probabilities of hitting the poverty level before certain ages for different sets of parameters.

References

American Academy of Actuaries (2005): "Recommended Approach for Setting Regulatory Risk-Based Capital Requirements for Variable Annuities and Similar Products," .

Amin, K., and A. Khanna (1994): "Convergence of American option values from discrete- to continuous-time financial models," *Mathematical Finance*, 4(4), 289–304.

Bacinello, A. R. (2003a): "Fair valuation of a guaranteed life insurance participating contract embedding a surrender option," *Journal of Risk and Insurance*, 70(3), 461–487.

——— (2003b): "Pricing guaranteed life insurance participating policies with annual premiums and surrender option," *North American Actuarial Journal*, 7(3), 1–17.

Bacinello, A. R., E. Biffis, and P. Millossovich (2009): "Pricing life insurance contracts with early exercise features," *Journal of Computational and Applied Mathematics*, 233(1), 27–35.

——— (2010): "Regression-based algorithms for life insurance contracts with surrender guarantees," *Quantitative Finance*, 10(9), 1077–1090.

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

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

Bauer, D., A. Kling, and J. Russ (2008): "A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities," *ASTIN Bulletin*, (38), 621–651.

Bauer, D., A. Kling, and J. Russ (2008): "A universal pricing framework for guaranteed minimum benefits in variable annuities," *ASTIN Bulletin*, 38(2), 621.

Belanger, A., P. Forsyth, and G. Labahn (2009): "Valuing the guaranteed minimum death benefit clause with partial withdrawals," *Applied Mathematical Finance*, 16(6), 451–496.

Bernard, C., M. R. Hardy, and A. MacKay (2014): "State-Dependent Fees for Variable Annuity Guarantees," *ASTIN, forthcoming.*

Bernard, C., and C. Lemieux (2008): "Fast simulation of equity-linked life insurance contracts with a surrender option," in *Proceedings of the 40th Conference on Winter Simulation*, pp. 444–452. Winter Simulation Conference.

Bernard, C., A. MacKay, and M. Muehlbeyer (2014): "Optimal surrender policy for variable annuity guarantees," *Insurance: Mathematics and Economics*, 55, 116–128.

Björk, T. (2004): *Arbitrage theory in continuous time*. Oxford university press.

Black, F., and M. Scholes (1973): "The pricing of options and corporate liabilities," *The journal of political economy*, pp. 637–654.

Boyle, P. P., and M. R. Hardy (1997): "Reserving for maturity guarantees: Two approaches," *Insurance: Mathematics and Economics*, 21(2), 113–127.

Boyle, P. P., and M. R. Hardy (2003): "Guaranteed Annuity Options," *ASTIN Bulletin*, 33(2), 125–152.

Boyle, P. P., and E. S. Schwartz (1977): "Equilibrium Prices of Guarantees under Equity-Linked Contracts," *Journal of Risk and Insurance*, 44, 639–660.

Canadian Institute of Actuaries (2002): "Report of the CIA Task Force on Segregated Fund Investment Guarantees," .

Carr, P., R. Jarrow, and R. Myneni (1992): "Alternative characterizations of American put options," *Mathematical Finance*, 2(2), 87–106.

Carroll, C. D. (2011): "Solution methods for microeconomic dynamic stochastic optimization problems," *http://www. econ. jhu. edu/people/ccarroll/solvingmicrodsops. pdf. (accessed October 26, 2012).*

Chen, Z., K. Vetzal, and P. A. Forsyth (2008): "The effect of modelling parameters on the value of GMWB guarantees," *Insurance: Mathematics and Economics*, 43(1), 165–173.

Chi, Y., and X. S. Lin (2012): "Are Flexible Premium Variable Annuities Under-Priced?," *ASTIN Bulletin*, 42(2), 559.

Coleman, T., Y. Kim, Y. Li, and M. Patron (2007): "Robustly Hedging Variable Annuities With Guarantees Under Jump and Volatility Risks," *Journal of Risk and Insurance*, 74(2), 347–376.

Costabile, M. (2013): “Analytical valuation of periodical premiums for equity-linked policies with minimum guarantee,” *Insurance: Mathematics and Economics*, 53(3), 597–600.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979): “Option pricing: A simplified approach,” *Journal of Financial Economics*, 7(3), 229–263.

Dai, M., Y. Kuen Kwok, and J. Zong (2008): “Guaranteed minimum withdrawal benefit in variable annuities,” *Mathematical Finance*, 18(4), 595–611.

Dayanik, S., and I. Karatzas (2003): “On the optimal stopping problem for one-dimensional diffusions,” *Stochastic Processes and their Applications*, 107(2), 173–212.

De Giovanni, D. (2010): “Lapse rate modeling: a rational expectation approach,” *Scandinavian Actuarial Journal*, 2010(1), 56–67.

Dickson, D. C., M. R. Hardy, and H. Waters (2009): *Actuarial Mathematics for Life Contingent Risks*. Cambridge University Press., Cambridge, England.

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

Duffy, D. J. (2006): *Finite Difference methods in financial engineering: a Partial Differential Equation approach*. John Wiley & Sons.

Gaillardetz, P. (2008): “Valuation of life insurance products under stochastic interest rates,” *Insurance: Mathematics and Economics*, 42(1), 212–226.

Grosen, A., and P. L. Jørgensen (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.

——— (2002): “Life insurance liabilities at market value: an analysis of insolvency risk, bonus policy, and regulatory intervention rules in a barrier option framework,” *Journal of Risk and Insurance*, 69(1), 63–91.

Hainaut, D., and P. Devolder (2006): “Life annuitization: Why and how much?,” *ASTIN Bulletin*, 36(2), 629.

Hanewald, K., J. Piggott, and M. Sherris (2013): “Individual post-retirement longevity risk management under systematic mortality risk,” *Insurance: Mathematics and Economics*, 52, 87–97.

Hansen, A., and P. Jørgensen (2000): “Analytical valuation of American-style Asian options,” *Management Science*, 46(8), 1116–1136.

Hardy, M. R. (2001): "A Regime-Switching Model of Long-Term Stock Returns," *North American Actuarial Journal*, 5(2), 41–53.

——— (2003): *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Harrison, G. W., J. A. List, and C. Towe (2007): "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica*, 75(2), 433–458.

Harrison, J. M., and S. R. Pliska (1981): "Martingales and stochastic integrals in the theory of continuous trading," *Stochastic processes and their applications*, 11(3), 215–260.

Horneff, W., R. Maurer, and R. Rogalla (2010): "Dynamic portfolio choice with deferred annuities," *Journal of Banking & Finance*, 34(11), 2652–2664.

Hubener, A., R. Maurer, and R. Rogalla (2013): "Optimal Portfolio Choice with Annuities and Life Insurance for Retired Couples," *Review of Finance*, p. rfs046.

Huerlimann, W. (2010): "Analytical pricing of the unit-linked endowment with guarantees and periodic premiums," *ASTIN Bulletin*, 40(2), 631.

Jönsson, H., A. Kukush, and D. Silvestrov (2005a): "Threshold structure of optimal stopping strategies for American type option. I," *Theory of Probability and Mathematical Statistics*, 71, 93–103.

Jönsson, H., A. Kukush, and D. Silvestrov (2005b): "Threshold structure of optimal stopping strategies for American type option. II," *Teor. Ĭmovīr. Mat. Stat.*, (72), 42–53.

Karatzas, I., and S. E. Shreve (1984): "Trivariate Density of Brownian Motion, Its Local and Occupation Times, with Application to Stochastic Control," *Annals of Probability*, 12(3), 819–828.

——— (1991): *Brownian Motion and Stochastic Calculus*. Springer, New York, 2nd edn.

Kemna, A., and A. Vorst (1990): "A pricing method for options based on average asset values," *Journal of Banking & Finance*, 14(1), 113 – 129.

Kim, I. J. (1990): "The Analytic Valuation of American Options," *The Review of Financial Studies*, 3(4), 547–572.

Kim, I. J., and G. G. Yu (1996): "An alternative approach to the valuation of American options and applications," *Review of Derivatives Research*, 1(1), 61–85.

Kling, A., F. Ruez, and J. Russ (2011): “The impact of stochastic volatility on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities,” *ASTIN Bulletin*, 41(02), 511–545.

Kling, A., F. Ruez, and J. Russ (2014): “The impact of policyholder behavior on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities,” Discussion paper, forthcoming in European Actuarial Journal.

Knoller, C., G. Kraut, and P. Schoenmaekers (2011): “On the Propensity to Surrender a Variable Annuity Contract,” Discussion paper, Working Paper, Ludwig Maximilian Universitaet Munich.

——— (2013): “On the Propensity to Surrender a Variable Annuity Contract-An Empirical Analysis of Dynamic Policyholder Behavior,” *Munich Risk and Insurance Center Working Paper*, (7).

Li, J., and A. Szimayer (2014): “The effect of policyholders’ rationality on unit-linked life insurance contracts with surrender guarantees,” *Quantitative Finance*, 14(2), 327–342.

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

MacDonald, B.-J., B. Jones, R. J. Morrison, R. L. Brown, and M. R. Hardy (2013): “Research and Reality: A Literature Review on Drawing Down Retirement Financial Savings,” *North American Actuarial Journal*, 17(3).

Maier, J., and M. Rüger (2010): “Measuring risk aversion model-independently,” *Discussion Papers in Economics*, 2010-33.

Maurer, R., O. S. Mitchell, R. Rogalla, and V. Kartashov (2013): “Lifecycle Portfolio Choice With Systematic Longevity Risk and Variable Investment Linked Deferred Annuities,” *Journal of Risk and Insurance*, 80(3), 649 – 676.

Milevsky, M. A. (2001): “Optimal annuitization policies: Analysis of the options,” *North American Actuarial Journal*, 5(1), 57–69.

Milevsky, M. A., and T. S. Salisbury (2001): “The real option to lapse a variable annuity: Can surrender charges complete the market,” in *Conference Proceedings of the 11th Annual International AFIR Colloquium*, p. 537.

Milevsky, M. A., and V. R. Young (2007): “Annuitization and asset allocation,” *Journal of Economic Dynamics and Control*, 31(9), 3138–3177.

Milliman (2011): “Variable annuity dynamic lapse study: A data mining approach,” Research report, Milliman.

Moenig, T., and D. Bauer (2012): “Revisiting the Risk-Neutral Approach to Optimal Policyholder Behavior: A Study of Withdrawal Guarantees in Variable Annuities,” Working Paper, Georgia State University.

New York Life (2014): “Premier Variable Annuity,” www.newyorklife.com/nyl-internet/file-types/NYL-Premier-VA-Fact-Sheet-No-CA-NY.pdf.

Palmer, B. (2006): “Equity-Indexed Annuities: Fundamental Concepts and Issues,” *Working Paper*.

Pennacchi, G. G. (2008): *Theory of asset pricing*. Pearson/Addison-Wesley.

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

Pollak, R. A. (1970): “Habit formation and dynamic demand functions,” *The Journal of Political Economy*, 78(4), 745–763.

Prudential(UK) (2012): “Key Features of the Flexible Investment Plan (No Initial Charge and Initial Charge option) - Additional Investments,” www.pru.co.uk/pdf/FIPK10080.pdf; Retrieved February 27, 2013.

Qiao, C., and M. Sherris (2013): “Managing Systematic Mortality Risk With Group Self-Pooling and Annuitization Schemes,” *Journal of Risk and Insurance*, 80(4), 949–974.

Racicot, F.-E., and R. Théoret (2006): *Finance computationnelle et gestion des risques*. Presses de l’Universite du Quebec.

Sabin, M. (2010): “Fair tontine annuity,” *Available at SSRN 1579932*.

Siu, T. K. (2005): “Fair valuation of participating policies with surrender options and regime switching,” *Insurance: Mathematics and Economics*, 37(3), 533–552.

Society of Actuaries (2012): “Policyholder Behavior in the Tail: Variable Annuity Guaranteed Benefits—2011 Survey Results,” Research report, Society of Actuaries.

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

Strulovici, B., and M. Szydlowski (2012): “On the smoothness of value functions and the existence of optimal strategies,” *Available at SSRN 1996808*.

Thrivent Financial (2014): "Flexible Premium Deferred Variable Annuity," www.thrivent.com/insurance/autodeploy/Thrivent_VA_Prospectus.pdf.

UBC Faculty Pension Plan (2013): *The Variable Payment Life Annuity Explained*, http://www.pensions.ubc.ca/faculty/publications/brochures/annuities.pdf. (accessed on January 18, 2014).

van de Ven, J., and M. Weale (2009): "Mortality Risk and Pricing of Annuities," Discussion paper, National Institute of Economic and Social Research.

Villeneuve, S. (2007): "On threshold strategies and the smooth-fit principle for optimal stopping problems," *Journal of Applied Probability*, pp. 181–198.

Wu, R., and M. C. Fu (2003): "Optimal exercise policies and simulation-based valuation for American-Asian options," *Operations Research*, 51(1), 52–66.

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