An Optimal Stochastic Control Framework for Determining the Cost of Hedging of Variable Annuities

Peter Forsyth[*] Kenneth Vetzal[†]

February 25, 2014

Abstract

An implicit partial differential equation (PDE) method is used to determine the cost of hedging for a Guaranteed Lifelong Withdrawal Benefit (GLWB) variable annuity contract. In the basic setting, the underlying risky asset is assumed to evolve according to geometric Brownian motion, but this is generalized to the case of a Markov regime switching process. A similarity transformation is used to reduce a pricing problem with K regimes to the solution of K coupled one dimensional PDEs, resulting in a considerable gain in computational efficiency. The methodology developed is flexible in the sense that it can calculate the cost of hedging for a variety of different withdrawal strategies by investors. Cases considered here include both optimal withdrawal strategies (i.e. strategies which generate the highest possible cost of hedging for the insurer) and sub-optimal withdrawal strategies in which the policy holder's decisions depend on the moneyness of the embedded options. Numerical results are presented which demonstrate the sensitivity of the cost of hedging (given the withdrawal specification) to various economic and contractual assumptions.

Keywords: Optimal control, GLWB pricing, PDE approach, regime switching, no-arbitrage, withdrawal strategies

JEL Classification G22

Acknowledgements

This work was supported by Tata Consulting Services, the Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada.

1 Introduction

Over the past few decades there has been a general trend away from defined benefit pension plans. One result of this development has been an increased focus on financial contracts which are designed to assist investors with managing their pre-retirement savings and post-retirement spending plans. Variable annuities (VAs) are a prominent example. In contrast to traditional fixed annuities which provide a minimum specified rate of interest, VAs provide investors with additional flexibility in terms of how their contributions are invested (e.g. a choice among mutual funds). The term

[*]Cheriton School of Computer Science, University of Waterloo, Waterloo ON, Canada N2L 3G1, paforsyt@uwaterloo.ca

[†]School of Accounting and Finance, University of Waterloo, Waterloo ON, Canada N2L 3G1, kvetzal@uwaterloo.ca

"variable" refers to the fact that the returns can vary according to the investment choices made. In the U.S., VAs offer a tax-deferral advantage because taxes are not paid until income is withdrawn.[1] Although VAs have been offered for many years, the market for them exhibited dramatic growth beginning in the 1990s: according to Chopra et al. (2009), the annual rate of growth in the U.S. VA market during that decade was 21%, and the level of total assets reached almost USD 1 trillion by 2001. While the decline of traditional defined benefit pension plans was a contributing factor, another reason was that VA contracts began to incorporate several additional features which made them more attractive to investors. As noted by Bauer et al. (2008), these features can be divided into two broad types: guaranteed minimum death benefits (GMDBs) and guaranteed minimum living benefits (GMLBs). GMDBs provide a payout at least equal to the original amount invested (or this amount grossed up by a guaranteed minimum rate of return) if a policy holder dies. These provisions first became widely adopted in VA contracts in the 1990s. There are several types of GMLBs: guaranteed minimum accumulation benefits and guaranteed minimum income benefits both give investors a guaranteed asset level at some specified future time, the former in a lump sum amount and the latter in the form of an annuity. Guaranteed minimum withdrawal benefits (GMWBs) allow investors to withdraw funds each period (e.g. year) from their VA accounts up to specified limits, regardless of the investment performance of the accounts.[2] A variation of GMWBs known as guaranteed lifelong withdrawal benefits (GLWBs) permits such withdrawals as long as the investor remains alive. These various GMLB features were widely introduced to the U.S. market in the early 2000s. The size of assets in U.S. VA accounts grew to about USD 1.5 trillion by the end of 2007 (Chopra et al., 2009). In addition to the U.S., similar VA-type contracts have been marketed to investors in many other countries, including Japan, the U.K., Germany, Italy, France, and Canada. Further information about the historical development of the VA market and the types of contracts available can be found in Bauer et al. (2008) and Chopra et al. (2009).

The onset of the financial crisis in the latter half of 2007 resulted in dismal equity returns, sustained low interest rates, and high market volatility. Many of the options embedded into VA contracts became quite valuable, and insurers were faced with the prospect of having to make large payments in order to meet the terms of these written options. Some insurers were not effectively hedged, and large losses resulted (Kling et al., 2011). Subsequently, VA sales have generally declined, in part because insurers have reduced offerings, raised fees, and attempted to buy existing investors out of their contracts (Tracer and Pak, 2012). MetLife took a USD 1.6 billion impairment charge related to its annuity business in the third quarter of 2012 (Tracer and Pak, 2012), and some firms have tried to sell off their annuity business units (Nelson, 2013).[3]

An unfortunate aspect of this is that in principle it makes sense for relatively sophisticated financial institutions to offer risk management services for retail clients. However, this is subject to the caveat that the institutions themselves need to adopt effective hedging strategies to offset the risk exposures resulting from selling these types of contracts. Many insurers did attempt to hedge the risks. According to a report cited by Chopra et al. (2009), hedging programs saved the industry about $40 billion in September-October 2008, offsetting almost 90% of the industry's increase in liability valuations during that period. However, the hedging programs which were adopted were clearly not entirely successful and so there is definite scope for research as to how they might be improved.

[1] The deferral advantage is offset somewhat by having withdrawals taxed at ordinary income rates rather than capital gains rates.

[2] Investors can withdraw funds in excess of the specified limits, but are typically charged penalties to do so.

[3] Even so, as a reflection of earlier sales and recent stronger equity market performance, total assets in U.S. VA contracts reached an all-time high of around USD 1.7 trillion at the end of the first quarter of 2013 (Insured Retirement Institute, 2013).

This article contributes to the literature by developing a simple and computationally efficient framework to evaluate the cost of hedging these types of contracts. Of course, the cost of hedging depends on the issuer's hedging strategy, so it is worth discussing this in more detail here. As a point of comparison, consider the standard approach for valuing an American put option on some underlying asset. The fundamental idea is to determine the initial cost of a dynamic self-financing replicating portfolio which is designed to provide an amount at least equal to the payoff of the contract, on the assumption that the purchaser of the contract adopts an exercise strategy that maximizes the monetary value of the contract. If the purchaser follows any other exercise strategy, the contract writer will be left with a surplus. The initial cost of establishing this hedging strategy is the no-arbitrage price of the contract—if the contract were to trade for a different price than this portfolio, then arbitrage profits could be made in principle by exploiting this price difference. This is not a pure arbitrage argument, in the sense that it is subject to modelling assumptions about the value of the underlying asset and parameters such as volatility. The VA setting differs in two key respects. First, the option premium is not paid up as an up-front instalment, but rather is deducted over time as a proportional fee applied to the value of the assets in the investor's account. Second, it is important to allow for alternative possible assumptions regarding the investor's option exercise strategy. This is because an investor may follow what appears to be a sub-optimal strategy that does not maximize the monetary value of the embedded option. This could be for idiosyncratic reasons such as liquidity needs or tax circumstances.[4] We use the term "cost of hedging" to refer to the fair hedging fee to be deducted that finances a dynamic replicating portfolio for the options embedded in the contract under the assumption of a particular exercise strategy. The replicating portfolio is managed so as to provide sufficient funds to meet any future payouts that arise from writing the contracts, at least under the model considered. This is distinguished from the "no-arbitrage fee" by the possibility of alternative exercise behaviour. Our terminology is intended to remind the reader of this generalization: the no-arbitrage fee would be a special case under the assumption that the investor's strategy is to maximize the monetary value of the options embedded in the contract. Of course, as with the no-arbitrage value of standard option contracts, the cost of hedging for VA contracts is still subject to modelling assumptions about the value of the underlying asset over time. We also emphasize that the cost of hedging calculated under the assumption that investors act to maximize the value of the options that they hold does offer an important benchmark in that it is a worse case scenario for the contract writer—again, under the particular model assumed for the value of the underlying asset.

In this article, we will focus exclusively on GLWBs. Our approach can easily be adapted to the simpler case of GMWBs with a fixed maturity date, but recent concerns over longevity risk (i.e. retirees outliving their savings) imply that GLWBs may be of greater significance.[5] These contracts are typically initiated by making a single lump sum payment to an insurance company. This payment is then invested in risky assets, usually a mutual fund. The benefit base, or guarantee account balance, is initially set to the amount of the lump sum payment. The holder of the contract is entitled to withdraw a fixed fraction of the benefit base each period (e.g. year) for life, even if the actual investment in the risky asset declines to zero. Upon the death of the contract holder, his or her estate receives the remaining amount in the risky asset account. Typically, these contracts have ratchet provisions (a.k.a. "step-ups"), which periodically increase the benefit base if the risky asset investment has increased to a value larger than the guarantee account value. In addition,

[4]Prepayment options in mortgages offer a useful analogy. While these options may be exercised for the monetary advantage of being able to re-finance a home at a lower prevailing interest rate, they could also be exercised for a variety of other reasons such as a transfer of employment, a divorce, or a simple desire to move to a larger house.

[5]In addition, since GLWBs can last for much longer than GMWBs, it is more important to develop efficient valuation methods for GLWBs.

the benefit base may also be increased if the contract holder does not withdraw in a given year. This is known as a bonus or roll-up. Finally, the contract holder may withdraw more than the contractually specified amount, including complete surrender of the contact, upon payment of a penalty. Complete surrender here means that the contract holder withdraws the entire amount remaining in the investment account, and the contract terminates. In most cases, the penalty for full or partial surrender declines to zero after five to seven years. As noted above, investors are charged a proportional fee from their risky asset accounts to pay for these features.

The valuation and hedging of the various provisions embedded into VA contracts is a challenging exercise as the options involved are long term, path-dependent, and complex. Investors can be modelled as facing a non-trivial optimal stochastic impulse control problem when determining their withdrawal strategies in GMWB/GLWB contracts. In general, the prior literature on these contracts has taken one of two alternative approaches: (i) focusing primarily on the investor's optimization problem in the context of a relatively simple specification such as geometric Brownian motion (GBM) for the stochastic evolution of the underlying investment; or (ii) concentrating on a richer stochastic specification incorporating features such as random volatility and/or random interest rates, while assuming that the investor follows a simple pre-specified strategy, typically involving always withdrawing the contractually specified amount each period (i.e. the maximum withdrawal that can be made without paying a penalty), no matter what happens to the value of the investment account. As an early example of the former approach, Milevsky and Salisbury (2006) value GMWB contracts under GBM and two extreme cases of policy holder behaviour: withdrawal of the contractually specified amount at all times in all circumstances or maximizing the economic value of the embedded options. Numerical PDE techniques are used to solve the valuation problems. The fair hedging fee for the contract is shown to increase substantially if investors are assumed to act to maximize the value of their embedded options rather than to passively withdraw the contract amount. Dai et al. (2008) model the GMWB pricing problem as a singular stochastic control problem in the GBM setting and provide an efficient numerical PDE approach for solving the problem. Illustrative calculations show dramatic differences in the fair hedging fee as parameters such as the allowed withdrawal amount, the penalty for excess withdrawals, and volatility are changed. Chen et al. (2008) provide a detailed study of the effects of various parameters on the fair fees for hedging GMWBs, showing that in addition to assumed levels for volatility and the risk-free rate, the fact (often ignored in the literature) that the total fees charged to investors is split between fees made available for hedging purposes and fees paid for managing the underlying mutual funds can have large effects. Most of the results reported are for the case of GBM, but an extension to a jump-diffusion setting is also considered and shown to have a potentially significant impact on the fair hedging fee. In addition, Chen et al. (2008) also explore an alternative assumption about policy holder behaviour based on an idea put forth by Ho et al. (2005). Under this scenario, the contract holder is assumed to withdraw the contract amount unless the embedded options are sufficiently deep-into-the-money, in which case the assumption is that the holder will act "optimally" to capture the option value. Again, the implied fair hedging fees are quite sensitive to this behavioural specification. In Bauer et al. (2008), a very general framework is developed for pricing a wide variety of VA contracts. The numerical cases considered are all in the GBM context. In contrast to the papers cited above which rely on various types of numerical PDE approaches, Bauer et al. (2008) use Monte Carlo methods for some policy holder behaviour assumptions and a combination of Monte Carlo methods and a numerical integration technique to determine the optimal strategy. However, the approaches considered are quite inefficient compared to the PDE-based alternatives, at least in the simple GBM setting. The general framework of Bauer et al. is applied to the specific case of GLWBs by Holz et al. (2012). A variety of contractual features are considered, as well as some alternative assumptions about policy holder behaviour. The stochastic setting is GBM, and Monte Carlo methods are used. Again, the same general conclusion emerges as with several other studies: the fair hedging fee for the contract is extremely sensitive to assumptions about behaviour, the risk-free rate, volatility, and contractual provisions. Piscopo (2010) assumes GBM and uses a Monte Carlo method to estimate fair hedging fees for GLWBs for two cases of policy holder behaviour: always withdraw the contract amount each period, or do so except if the value of the account less any penalty is less than the present value of the future benefits, in which case completely withdraw all funds. Illustrative calculations show that the estimated fair hedging fee under the latter strategy is close to double that for the former. Piscopo and Haberman (2011) assume GBM and a base case strategy of always withdrawing the contract amount every period. Monte Carlo methods are used to show the effects of a variety of contractual provisions such as step-ups and roll-ups, assuming that the policy holder follows a strategy such as making no withdrawals for a specified number of years. In addition, an extension to stochastic mortality risk is considered. Yang and Dai (2013) develop a tree-based method for analyzing GMWB type contracts in the GBM context. Yang and Dai emphasize and demonstrate the importance of various contractual provisions, but they do not provide results for cases where investors are assumed to optimize discretionary withdrawals. Huang and Kwok (2013) carry out a theoretical analysis in the GBM setting of withdrawal strategies assuming the worst case for the contract provider (i.e. the policy which maximizes the value of the guarantee).

As noted above, the second general type of approach involves using more complex models for the value of the underlying fund, but specifying simpler strategic behaviour on the part of policy holders. For example, Shah and Bertsimas (2008) use Monte Carlo and numerical integration methods to estimate the fair hedging fees for GLWB contracts assuming that investors always withdraw the contractually specified amount from their accounts. Three specifications are considered: GBM, GBM with stochastic interest rates, and a generalization with both stochastic interest rates and stochastic volatility. Incorporating random interest rates and volatility results in somewhat higher fees compared to GBM.[6] Similarly, Kling et al. (2011) consider a stochastic volatility model, using Monte Carlo methods but assuming non-optimal behaviour by policy holders. Kling et al. conclude that while stochastic volatility does not matter too much for pricing GLWBs, it can have a significant effect on hedging strategies and risk exposures. Bacinello et al. (2011) consider a variety of VA embedded options in a setting with stochastic interest rates and stochastic volatility. Most of the results presented are for GMDBs or other specifications which do not provide for early withdrawals at the discretion of the policy holder. When considering contracts which do allow for these features, the assumptions regarding investor behaviour are basically the same as those of Piscopo (2010): always withdraw the contract amount, or completely surrender the policy. Monte Carlo methods are used to estimate hedging costs for contracts without discretionary withdrawal features and to estimate contract values given assumed fee levels otherwise. Standard conclusions apply regarding the sensitivity of fee levels (or contract values) to contractual specifications and financial market parameters. Peng et al. (2012) augment the standard GBM specification with stochastic interest rates. Under the assumption of deterministic withdrawals, they derive analytic upper and lower bounds for the fair values of GMWB contracts. Donelly et al. (2014) explore the valuation of guaranteed withdrawal benefits under stochastic interest rates and stochastic volatility. A PDE-based numerical scheme is used, and allowing for randomness of both interest rates and volatility is shown to have potentially large effects. However, the investor is simply assumed to always withdraw the contract amount.

[6] Shah and Bertsimas (2008) actually underestimate the fair hedging fees for all specifications because they assume that the fees are paid separately by investors rather than being deducted from the GLWB account itself. The reduction in the account value makes the embedded guarantee features on the original investment more valuable, implying higher hedging costs.

The departures from the simple GBM context cited thus far all involve having volatility and/or the risk-free rate follow a distinct diffusion process. A simpler alternative is to combine GBM with Markov regime switching. This results in a parsimonious representation that in principle can account for random changes in volatility and interest rates (Hardy, 2001). In the context of pricing standard options, such models are discussed in sources such as Naik (1993), Bollen (1998), Duan et al. (2002), Yuen and Yang (2009), and Shen et al. (2013). In the specific case of VAs and other equity-linked insurance contracts, regime switching models have been suggested by Siu (2005), Lin et al. (2009), Bélanger et al. (2009), Yuen and Yang (2010), Ngai and Sherris (2011), Jin et al. (2011), and Uzelac and Szimayer (2014). However, most of these papers do not consider any form of withdrawal benefits.[7]

The papers cited above illustrate the general tradeoff between the underlying stochastic model and the assumed strategies followed by investors. Making models more realistic by allowing departures from GBM such as stochastic volatility and/or interest rates is clearly desirable for these long term contracts. At the same time, it is also important to be able to consider a variety of possible assumptions about investor behaviour. The intrinsic difficulty of determining optimal behaviour in a model with stochastic volatility and interest rates has led the authors of previous papers to emphasize one or the other of these model features. In this paper, we use an implicit PDE approach to value GLWB guarantees. We initially consider the GBM setting, but then generalize to a Markov regime switching framework. This allows us to analyze a variety of assumptions regarding the withdrawal strategies of investors in a setting which permits a simple specification of stochastic volatility and interest rates. Our implicit method contrasts with lattice or tree-based methods (e.g. Yuen and Yang, 2010; Yang and Dai, 2013), which are essentially explicit difference schemes. Such explicit approaches are characterized by well-known time step size limitations due to stability considerations. These restrictions are particularly costly in the case of long term contracts such as GLWBs.

If we consider a modelling scenario with K regimes, then the use of a similarity transformation reduces the computational problem to solving a system of K coupled one-dimensional PDEs. This makes the computational cost of pricing GLWB contracts very modest, and far more efficient than Monte Carlo based alternatives. In order to determine the fair hedging fee, we use Newton iteration combined with a sequence of refined grids. In most cases, just a single Newton iteration is required on the finest grid.

From a financial perspective, option valuation in regime switching models is complicated by the fact that standard Black-Scholes arguments which rely on hedging written options using a replicating portfolio consisting of the underlying asset and a risk-free asset are inapplicable due to the extra risk associated with a potential change in regime—the market is incomplete. This implies that the martingale measure is not unique, and additional criteria are needed to pin down the measure to be used from pricing. This can be done in a variety of ways. One possibility is to use the Esscher transform, which can be justified on economic grounds in the context of a representative agent model with power utility. This was first suggested in the regime switching setting by Elliott et al. (2005), and applied to the context of VAs and other equity-linked insurance contracts in papers such as Siu (2005) and Lin et al. (2009). As an alternative, we consider an expanded set of hedging instruments. Examples could include other derivative contracts on the underlying or bonds of various maturities. As pointed out by Naik (1993) in a model with two regimes, one additional hedging instrument is needed beyond the underlying asset and the risk-free asset. More

[7] Bélanger et al. (2009) analyze GMDB contracts which permit partial early withdrawals assuming investors act so as to maximize the value of this option. Ngai and Sherris (2011) primarily focus on longevity risk, but do consider GLWBs in cases where policy holders are assumed to withdraw at the contract rate or to completely surrender their contracts.

generally, if there are K regimes, a total of $K + 1$ hedging instruments can be used to construct the replicating portfolio. If two of the instruments are the risk-free asset and the underlying asset, then $K - 1$ additional contracts such as traded options are required.

Based on this set of hedging instruments, we develop a general method for determining the cost of hedging GLWB contracts without making any specific assumptions about the withdrawal strategy of the contract holder. We use a dynamic programming approach (i.e. we solve the PDE backwards in time), which allows us to explore various withdrawal strategies. As illustrations, we consider two examples:

1. Taking the point of view of the worst case for the hedger, we assume that the contract holder follows an optimal withdrawal strategy. We use the term "optimal" here in the sense of the discussion above: the optimal withdrawal strategy maximizes the monetary value of the guarantee. We also allow the contract holder to surrender the contract when it is optimal to do so, again in the sense of maximizing the monetary gain from the contract.

2. We use the model discussed in Ho et al. (2005) which assumes that investors will withdraw at the contractually specified rate unless it is significantly advantageous for them to deviate from this strategy.[8]

We emphasize, however, that these two cases are merely illustrative: our method can value GLWB contracts under a wide variety of withdrawal strategies. Of course, given assumptions about parameter values (e.g. volatility), no alternative strategy can lead to a higher cost of hedging for the insurer than the "optimal" one. In this sense, the costs calculated here under the first scenario above are a worst-case upper bound for the insurer.

We use the terms "fair hedging fee" and "cost of hedging" interchangeably to refer to the fee which is required to maintain a replicating portfolio. A description of this replicating portfolio is given in the derivations of our valuation equations which are provided in Appendices A and B (see also Chen et al. 2008 and Bélanger et al. 2009).

The main contributions of this paper are as follows:

- We formulate the task of determining the worst case hedging cost as an optimal stochastic impulse control problem. In the context of a regime switching model, we derive a coupled system of PDEs and optimal control decisions across withdrawal dates that can be used in one of two ways. First, given an assumed fee, the solution of the system provides the value of the GLWB contract. Second, by numerically searching across alternative fees, the fair hedging fee can be determined as that which makes the initial value of the contract equal to the lump sum invested. Since the PDEs in the coupled system are one-dimensional, the model can be implemented in a way that is far more efficient than Monte Carlo based approaches that are common in the literature.

- We present numerical examples demonstrating the convergence of this method, and the sensitivity of the fair hedging fee to various modelling parameters.

- We consider two specific withdrawal assumptions: the worst case for the hedger (optimal withdrawal) and withdrawals depending on the moneyness of the guarantee (Ho et al., 2005; Knoller et al., 2013). However, we emphasize that our procedure can be adapted to other withdrawal specifications.

[8]Knoller et al. (2013) conduct an empirical investigation of the behaviour of Japanese VA policy holders. They find that the moneyness of the embedded options is the single most important factor in explaining exercise decisions by policy holders, lending support to the model of Ho et al. (2005).

- We consider the effect of misspecification risk for the case where the hedger incorrectly assumes that there is only a single regime. We assume that market prices for traded options and risk-free bonds are generated by a regime-switching model, and calibrate the parameters of the single regime model to match these prices. For some parameters, the implied hedging costs for the single regime model are reasonably close to those for the regime switching model, but this is not always true. Overall, we find that a single regime model cannot be assumed to consistently give an effective approximation.

Overall, the implicit PDE method developed here can be used to rapidly explore the effect of economic, contractual, and longevity assumptions on the fair hedging fee, under a variety of alternative withdrawal strategies for investors. The balance of the paper proceeds as follows. Section 2 develops the GLWB model in the GBM setting with just one regime. This model is extended to the regime switching context in Section 3. Various alternative policy holder withdrawal assumptions are discussed in Section 4. The numerical approach is described in Section 5. This is followed by Sections 6 and 7, which contain an extensive set of illustrative results and a brief concluding summary.

2 Formulation: Single Regime Case

We begin by considering the simplified case without regime switching. We assume that mortality risk is diversifiable across a large number of contract holders.[9] Let the mortality function $\mathcal{M}(t)$ be defined such that the fraction of the original owners of the GLWB contract who die in the interval $[t, t+dt]$ is $\mathcal{M}(t)dt$. The fraction of the original owners still alive at time t is denoted by $\mathcal{R}(t)$, with

$$\mathcal{R}(t) = 1 - \int_0^t \mathcal{M}(u)du. \tag{2.1}$$

Time t is measured in years from the contract inception date. Typically, mortality tables are given in terms of integer ages $\{0, 1, \dots\}$. Specifically, let x_0, y, and ω be integers with

$$\begin{aligned} x_0 &= \text{ insured's age at contract inception} \\ {}_y p_{x_0} &= \text{ probability that an } x_0 \text{ year old will survive the next } y \text{ years} \\ q_{x_0+y} &= \text{ probability for an } x_0+y \text{ year old to die in the next year} \\ \omega &= \text{ age beyond which survival is impossible.} \end{aligned} \tag{2.2}$$

This gives

$$\mathcal{M}(t) = {}_y p_{x_0} q_{x_0+y} \quad \text{where } t \in [y, y+1) \tag{2.3}$$

with $\mathcal{R}(t)$ given from equation (2.1). Note that $\mathcal{M}(t)$ is assumed constant for $t \in [y, y+1)$.

Let S be the amount in the investment account (i.e. mutual fund) of any holder of the GLWB contract still alive at time t. Let A be the guarantee account balance. We suppose that percentage fees based on the value of the investment account S are charged to the policy holder at the annual rate α_{tot} and withdrawn continuously from that account. These fees include mutual fund management fees α_m and a fee charged to fund the guarantee α_g, so that $\alpha_{tot} = \alpha_g + \alpha_m$. It is worth noting that while most existing contracts deduct fees as a fraction of the investment account, some insurance companies are now charging fees as a fraction of the guarantee account balance A or even

[9] In the case that this assumption is not justified, then the risk-neutral value of the contract can be adjusted using an actuarial premium principle (Gaillardetz and Lakhmiri, 2011).

$\max(S, A)$. Although these types of alternative fee structures can be easily incorporated into our general approach, we will consider only fees that are proportional to the value of the investment account S in the remainder of this article.

To determine the fair hedging fee for the GLWB contract, we use hedging arguments similar to Windcliff et al. (2001), Chen et al. (2008), and Bélanger et al. (2009). Note that we assume that α_m is given exogenously, so that "fair hedging fee" refers only to α_g. In Windcliff et al. (2001) and Bélanger et al. (2009), the value of the guarantee portion of the contract was determined. In this work, it is convenient to pose the problem in terms of the entire contract value, i.e. the total of the guarantee and the investor's investment account balance. However, we build on the work in Windcliff et al. (2001) and Bélanger et al. (2009) by first considering the guarantee portion, and then using an algebraic transformation to determine the value of the entire contract.

2.1 Impulse Controls

We suppose that there is a set of deterministic discrete impulse control times t_i , with $i = 0, \ldots, M$, which we label *event times*. Normally, event times are on either an annual or quarterly basis.

At these event times withdrawals, ratchets, and bonuses may occur. At any of these times $\{t_i\}$, the holder of the GLWB contract can give the system an impulse $\mathcal{C}_i$, moving the state variables (S_t, A_t) to the state $(S(\mathcal{C}_i), A(\mathcal{C}_i))$ and producing cash flows $f(\mathcal{C}_i, S_t, A_t, t_i)$.

The set of impulse controls for this problem is then the set

$$\mathcal{C} = \{\{t_0, \mathcal{C}_0\}, \{t_1, \mathcal{C}_1\}, \ldots, \{t_M, \mathcal{C}_M\}\} \ . \tag{2.4}$$

2.2 Evolution of Value Excluding Event Times

We first consider the evolution of the uncontrolled state variables between event times, that is for $t \in (t_i, t_{i+1})$, where t_i are the event times. Let the value of the guarantee portion of the contract be $\mathcal{U}(S, A, t)$. This guarantee portion is also known as the GLWB rider. Its value will incorporate the effects of mortality. Since for now there is just a single regime, assume that the value of the investment account follows the GBM process

$$dS = (\mu - \alpha_{tot})Sdt + \sigma SdZ, \tag{2.5}$$

where μ is the drift rate, σ is the volatility, and dZ is the increment of a Wiener process. We assume that the mutual fund in the investment account tracks an index $\hat{S}$ without any basis risk. The index follows

$$d\hat{S} = \mu \hat{S} dt + \sigma \hat{S} dZ. \tag{2.6}$$

We further assume that it is not possible for the insurance company to short the mutual fund S for fiduciary reasons (Windcliff et al., 2001).

As shown in Appendix A, $\mathcal{U}(S, A, t)$ satisfies a PDE of the form

$$\mathcal{U}_t + \frac{\sigma^2 S^2}{2}\mathcal{U}_{SS} + (r - \alpha_{tot})S\mathcal{U}_S - r\mathcal{U} - \mathcal{R}(t)\alpha_g S = 0, \tag{2.7}$$

where r is the risk-free interest rate and where the term $\mathcal{R}(t)\alpha_g S$ represents the stream of fees from the investors remaining in the guarantee at time t to the hedger. Equation (2.7) is identical to equation (5) in Bélanger et al. (2009), with the exception that we assume here that there is no GMDB. Although many contracts allow the holder to purchase both a *money back* GMDB provision and a GLWB feature, we focus exclusively on the GLWB rider in this article.

Consider time $T = \omega$, when there are no longer any policy holders left alive. From the point of view of the hedger of the GLWB rider, the value of the guarantee is zero at this time

$$\mathcal{U}(S, A, T) = 0. \tag{2.8}$$

We emphasize that equation (2.7) is valid only at times excluding event times. The complete problem also requires the addition of the control decisions across the event times, which will be discussed in Section 2.3. It is these control decisions that add A dependence to the contract.

The event times correspond to withdrawals made by the contract holder. If the contract holder never makes any withdrawals (except for death benefits), then solving equation (2.7) would lead to a negative contract value. This is easily understood: the contract holder is paying a fee and obtaining no financial benefit. Clearly, this would not be the optimal strategy for the contract holder.

Now, consider the value of the entire GLWB contract $\mathcal{V}(S, A, t)$

$$\mathcal{V}(S, A, t) = \mathcal{U}(S, A, t) + \mathcal{R}(t)S, \tag{2.9}$$

which includes the rider $\mathcal{U}$ and the amount in the investment accounts of those remaining alive. Note that only the investment account is affected by the survival probability, since mortality is already included in the PDE for $\mathcal{U}$. It is easily shown (see Appendix A) that

$$\mathcal{V}_t + \frac{\sigma^2 S^2}{2}\mathcal{V}_{SS} + (r - \alpha_{tot})S\mathcal{V}_S - r\mathcal{V} + \alpha_m \mathcal{R}(t)S + \mathcal{M}(t)S = 0. \tag{2.10}$$

It is convenient for computational purposes to express time in terms of "backward time" $\tau = T - t$, i.e. the time remaining until none of the original contract holders is left alive. Letting $V(S, A, \tau = T - t) = \mathcal{V}(S, A, t)$, equation (2.10) becomes

$$V_\tau = \frac{\sigma^2 S^2}{2} V_{SS} + (r - \alpha_{tot})SV_S - rV + \alpha_m \mathcal{R}(t)S + \mathcal{M}(t)S. \tag{2.11}$$

From equations (2.8) and (2.9) we have

$$V(S, A, \tau = 0) = \mathcal{R}(T)S = 0, \tag{2.12}$$

since $R(T = \omega) = 0$. We note that equation (2.11) is identical to equation (5) in Chen et al. (2008) if we set $\mathcal{R}(t) = 1$ and $\mathcal{M}(t) = 0$.

2.3 Impulse Controls at Event Times

Recall that we denote the contractually specified times where withdrawals, bonuses and ratchets occur as event times t_i. Since we will solve the PDE backwards in time, it is convenient to denote $\tau_i = T - t_i$, and let $\tau_i + \epsilon = \tau_i^+$, $\tau_i - \epsilon = \tau_i^-$, where $\epsilon > 0, \epsilon \ll 1$. However, many of the contractual features and the mortality tables are specified in terms of forward event times t_i, so we will use t_i as the argument for the contractual and actuarial parameters, while we use τ_i as the argument for the contract value $V(S, A, \tau)$. We now proceed to describe the control decisions for the various types of events.

We parameterize the contract holder's actions at event time t_i by a policy parameter $\gamma_i \in [0, 2]$. In the case of a ratchet, we denote the ratchet policy parameter by $\mathbb{R}_i \in \{0, 1\}$ where $\mathbb{R}_i = 0$

denotes no ratchet and $\mathbb{R}_i = 1$ denotes a ratchet event. The possible impulse controls $\mathcal{C}_i = (\mathbb{R}_i, \gamma_i)$ at t_i are then

$$\begin{aligned} \mathcal{C}_i &= (0, \gamma_i) \ ; \ \text{No ratchet} \\ \mathcal{C}_i &= (1, 0) \ ; \ \text{Ratchet} \ , \end{aligned} \tag{2.13}$$

where we do not permit a ratchet and withdrawal at the same time.

We can write the impulse control at event time t_i in the general form

$$V(\mathcal{C}_i, S, A, \tau_i^+) = V\left(S(\mathcal{C}_i), A(\mathcal{C}_i), \tau^-\right) + f(\mathcal{C}_i, S, A, t_i), \tag{2.14}$$

where $f(\mathcal{C}_i, S, A, t_i)$ is the cash flow from the event.

Ratchet Event ($\mathbb{R}_i = 1, \gamma_i = 0$). If the contract specifies a ratchet (step-up) feature, then the value of the guarantee account A is increased if the investment account has increased. The guarantee account A can never decrease, unless the contract is partially or fully surrendered. Using our general form (2.14), at a ratchet event time τ_i, we then have

$$\begin{aligned} \mathcal{C}_i &= (1, 0) \\ S(\mathcal{C}_i) &= S \\ A(\mathcal{C}_i) &= \max(S, A) \\ f(\mathcal{C}_i, S, A, t_i) &= 0 \ . \end{aligned} \tag{2.15}$$

General Withdrawal Event ($\mathbb{R}_i = 0$). The contract will typically specify a withdrawal rate G_r. Given a time interval of $t_i - t_{i-1}$ between withdrawals, the contract withdrawal amount at $t = t_i$ is $G_r(t_i - t_{i-1})A$. In the case of a withdrawal event $\mathcal{C}_i = (0, \gamma_i)$, where γ_i is the withdrawal policy of the contract holder.

At this point we do not make any particular assumptions about the withdrawal strategy of the policy holder. In general terms, the policy holder's actions at t_i can be represented by the policy parameter γ_i, where $0 \leq \gamma_i \leq 2$. Withdrawals of amounts less than or equal to the contract withdrawal amount $G_r(t_i - t_{i-1})A$ are represented by $\gamma_i \in [0, 1]$. Withdrawals in excess of the contract amount are indicated by $\gamma_i \in (1, 2]$, with $\gamma_i = 2$ corresponding to full surrender. We next consider different withdrawal events, represented by different values of γ_i.

Bonus Event ($\mathbb{R}_i = 0, \gamma_i = 0$). If the contract holder chooses not to withdraw at $t = t_i$, this is indicated by $\gamma_i = 0$. Let the bonus fraction be denoted by $B(t_i)$. If no bonus is possible at $t = t_i$, then $B(t_i) = 0$. To allow for a bonus, we have

$$\begin{aligned} \mathcal{C}_i &= (0, 0) \\ S(\mathcal{C}_i) &= S \\ A(\mathcal{C}_i) &= A(1 + B(t_i)) \\ f(\mathcal{C}_i, S, A, t_i) &= 0 \ . \end{aligned} \tag{2.16}$$

Withdrawal not Exceeding the Contract Amount ($\mathbb{R}_i = 0, \gamma_i \in (0, 1]$). The case where $\gamma_i < 1$ corresponds to a partial withdrawal of the contract amount, while $\gamma_i = 1$ implies a full withdrawal, i.e.

$$\begin{aligned} \mathcal{C}_i &= (0, \gamma_i) \ \ ; \ \ \gamma_i \in (0, 1] \\ S(\mathcal{C}_i) &= \max\left(S - \gamma_i G_r(t_i - t_{i-1})A, 0\right) \\ A(\mathcal{C}_i) &= A \\ f(\mathcal{C}_i, S, A, t_i) &= \mathcal{R}(t_i)\gamma_i G_r(t_i - t_{i-1})A \ . \end{aligned} \tag{2.17}$$

Note that withdrawals of funds up to and including the contract amount are allowed each period even if the amount in the investment account $S = 0$. In addition, the usual contract specification states that $A(\gamma_i) = A$ if withdrawal is not greater than the contract amount (i.e. the guarantee account value remains constant).

Partial or Full Surrender ($\mathbb{R}_i = 0, \gamma_i \in (1, 2]$). Next, consider the case of a withdrawal of an amount greater than the contract amount $G_r(t_i - t_{i-1})A$, i.e.

$$\text{Withdrawal amount} = G_r(t_i - t_{i-1})A + (\gamma_i - 1)S'\left(1 - \kappa(t_i)\right) \tag{2.18}$$

where $S' = \max\left(S - G_r(t_i - t_{i-1})A, 0\right)$ and $\kappa(t_i) \in [0, 1]$ is a penalty for withdrawal above the contract amount. In equation (2.18), $\gamma_i = 2$ represents complete surrender of the contract, while $1 < \gamma_i < 2$ represents partial surrender. Writing this in the general form (2.14), we have

$$\begin{aligned} \mathcal{C}_i &= (0, \gamma_i) \ \ ; \ \ \gamma_i \in (1, 2] \\ S(\mathcal{C}_i) &= S'(2 - \gamma_i) \\ A(\mathcal{C}_i) &= A(2 - \gamma_i) \\ f(\mathcal{C}_i, S, A, t_i) &= \mathcal{R}(t_i)\left(G_r(t_i - t_{i-1})A + (\gamma_i - 1)S'\left(1 - \kappa(t_i)\right)\right) . \end{aligned} \tag{2.19}$$

Contrast equation (2.19) with equation (2.17). In the case of withdrawal above the contract amount (i.e. partial or full surrender), the guarantee account value A is reduced proportionately for any withdrawal above the contract amount.

Death Benefit Payments. Our PDE (2.11) assumes that the amount remaining in the investment account is paid out immediately upon the death of the contract holder. However, in order to compare with some previous work, we also consider the possibility that death benefits are paid out only at event times t_i. In this case, if $t_{i-1} \le t \le t_i$ (i.e. $\tau_i \le \tau \le \tau_{i-1}$), then we solve the PDE

$$V_\tau = \frac{\sigma^2 S^2}{2} V_{SS} + (r - \alpha_{tot})SV_S - rV + \alpha_m \mathcal{R}(t_{i-1})S, \tag{2.20}$$

between event times. Death benefits are paid at t_i according to

$$V^{death}(S, A, \tau_i^+) = V(S, A, \tau_i^-) + \left(\mathcal{R}(t_{i-1}) - \mathcal{R}(t_i)\right) S. \tag{2.21}$$

The terminal condition at $T = \omega$ is modified in this case (recall that $\mathcal{R}(T = \omega) = 0$) to

$$V(S, A, 0) = \mathcal{R}(t_{N-1})S \tag{2.22}$$

where t_{N-1} is the penultimate event before the terminal event at T. Note that we will use equations (2.20-2.22) in the following only for a single numerical comparison with previous work. Generally, we will assume that death benefits are paid out continuously as in equations (2.11-2.12).

In practice, contracts typically specify that several events occur at the same contract times t_i. Mathematically, we can consider these events as occurring at times infinitesimally apart. A careful examination of the contract specification is usually required to determine the precise order of these events. We assume that the order of events occurring at an event time t_i is (in forward time):

1. Death benefit payments: equation (2.21).

2. Withdrawal, bonus, surrender: equation (2.14).

3. Ratchet: equation (2.15).

This order is typically seen in actual contracts. These events then occur in reverse order in backwards time.

2.4 Complete Pricing Problem

To summarize, in the case of a single regime the complete pricing problem consists of PDE (2.11) with initial condition (2.12), or PDE (2.20) and initial condition (2.22), which are valid for times excluding event times τ_i. As discussed above, we have the additional event conditions (2.15-2.19). If equation (2.20) is used, we have the additional event (2.21). Note that PDE (2.11) has no A dependence. However, $V = V(S, A, \tau)$ in general since the event conditions (2.15-2.19) generate A dependence at event times. Note that the solution to the complete pricing problem determines the value of the contract, given a specified hedging fee α_g. The fair hedging fee is the value of α_g such that the initial value of the contract equals the investor's initial contribution.

Given a withdrawal policy specified by the policy parameter γ_i, then the solution of the pricing problem is completely specified. We emphasize here that once the γ_i are given, then the cost of hedging can be determined. The choice of the model for γ_i is controversial, with various alternatives suggested in the literature. However, our pricing methodology isolates the choice for γ_i. In principle, any reasonable method can be used to determine γ_i and the remainder of the pricing method would remain the same. We will discuss some possibilities for selecting γ_i below in Section 4.

3 Extension to Regime Switching

We now extend the arguments above to the case where there are K possible regimes of the economy. As noted above in Section 1, this is a relatively simple way of incorporating uncertainty about interest rates and volatility into option valuation. It has been particularly popular in studying embedded options in insurance contracts since these contracts are typically quite long-term, and so it is harder to justify the assumption that such parameters will remain constant throughout the life of the contract.

The framework we adopt is basically the same as that originally proposed in the equity option valuation context by Naik (1993), and used more recently in sources such as Yuen and Yang (2009) and Shen et al. (2013). The set of possible regimes is $\mathcal{K} = \{1, 2, \ldots, K\}$. Let $\mathcal{K}_i$ be the set of states excluding state i, i.e. $\mathcal{K}_i = \mathcal{K} \setminus \{i\}$. The state of the economy is assumed to evolve according to a finite state continuous time observable Markov chain X on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is the real world probability measure. Following Elliott et al. (1995), the state space of X is identified with a finite set of unit basis vectors $\{e_1, \ldots, e_K\}$ where e_i is a $K \times 1$ vector with i-th component equal to unity and all other components zero. Let the rate matrix of the chain under $\mathbb{P}$ be A. Element (j, k) of this matrix is a constant transition intensity from state j to state k, denoted by $\lambda^{j \to k}$ for $j, k \in \mathcal{K}$. Note that

$$\begin{aligned} &\lambda^{j\to k} \geq 0 \text{ if } j \neq k \\ &\lambda^{j\to j} = -\sum_{k\in\mathcal{K}_j} \lambda^{j\to k}. \end{aligned}$$

The time index set of the model is $\mathcal{T} = [0, T < \infty]$. From the martingale representation theorem (Elliott et al., 1995),

$$X(t) = X(0) + \int_0^t A'X(s)ds + M(t), \quad t \in \mathcal{T},$$

where A' is the tranpose of A and $\{M(t)|t \in \mathcal{T}\}$ is an $(\mathbb{F}^X, \mathbb{P})$-martingale, with $\mathbb{F}^X = \left\{\mathcal{F}^X(t)|t \in \mathcal{T}\right\}$ being the filtration generated by X satisifying the usual conditions of being $\mathbb{P}$-complete and right-continuous. Denote t^- as $t - \epsilon$, $0 < \epsilon \ll 1$, and similarly $s^- = s - \epsilon$.

Following Elliott et al. (1995), let $\langle x, y \rangle$ denote the inner product $x'y$ of two column vectors x and y in $\mathbb{R}^K$ and define the martingale

$$\begin{aligned} m^{j\to k}(t) &\equiv \int_0^t \left\langle X(s^-), e_j \right\rangle e_k' dM(s) \\ &= \int_0^t \left\langle X(s^-), e_j \right\rangle e_k' dX(s) - \int_0^t \left\langle X(s^-), e_j \right\rangle e_k' A' X(s^-) ds \\ &= N^{j\to k}(t) - \lambda^{k\to j} \int_0^t \left\langle X(s^-), e_j \right\rangle ds \end{aligned}$$

where $N^{j\to k}(t)$ is the number of transitions from state j to state k up until time t. For each $k \in \mathcal{K}$, let $N^k(t)$ be the total number of transitions from other states into state k up to time t. Then

$$\begin{aligned} N^k(t) &= \sum_{j\in\mathcal{K}_k} N^{j\to k}(t) \\ &= \sum_{j\in\mathcal{K}_k} m^{j\to k}(t) + \sum_{j\in\mathcal{K}_k} \lambda^{k\to j} \int_0^t \left\langle X(s^-), e_j \right\rangle ds \end{aligned}$$

Denote the $(\mathbb{F}^X, \mathbb{P})$-martingale

$$\widetilde{N}^k(t) = \sum_{j\in\mathcal{K}_k} m^{j\to k}(t) = N^k(t) - \sum_{j\in\mathcal{K}_k} \int_0^t \left\langle X(s^-), e_j \right\rangle ds.$$

Letting

$$\lambda^k(t) = \sum_{j\in\mathcal{K}_k} \lambda^{k\to j} \left\langle X(t), e_j \right\rangle,$$

we then have

$$d\widetilde{N}^k(t) = dN^k(t) - \lambda^k(t^-)dt, \quad k \in \mathcal{K}.$$

The level of the instantaneous risk-free interest rate r is assumed to vary with the state of the economy. Let $\vec{r} = \left(r^1, r^2, \ldots, r^K\right)'$ be its possible values, so that $r(t) = \langle \vec{r}, X(t) \rangle$. Between event times, the value of the investor's account S is assumed to evolve according to GBM with coefficients that are similarly modulated by the Markov chain X. In particular, if $\vec{\nu} = \left(\nu^1, \nu^2, \ldots, \nu^K\right)'$ denotes the possible values for the expected growth rate (before percentage fees represented as in the single regime context by α_{tot}), the value of this parameter at time t is $\nu(t) = \langle \vec{\nu}, X(t) \rangle$. Similarly, $\vec{\sigma} = \left(\sigma^1, \sigma^2, \ldots, \sigma^K\right)'$ represents the regime-dependent values of the volatility term, which at time t is given by $\sigma(t) = \langle \vec{\sigma}, X(t) \rangle$. The Brownian motion term Z is assumed to be independent of X under $\mathbb{P}$. If there is a change in state, the level of the investor's account S is allowed to jump discretely but deterministically. The size of a jump depends on the state transition. In particular, let Ξ be a $K \times K$ matrix of parameters. Element (j, k) of this matrix, denoted by $\xi^{j\to k}$, determines the size of the jump associated with a transition from state j to state k, i.e. $S(t) = \xi^{j\to k} S(t^-)$ upon such a transition at time t. We restrict $\xi^{j\to j} = 1$ for each $j \in \mathcal{K}$, so that there are no jumps in S in the absence of a regime switch. Let $\xi(t)$ be the relevant row of Ξ given the state of the economy at time t, i.e. $\xi(t) = \Xi' X(t)$, and denote the elements of $\xi(t)$ by $\left(\xi^{X(t)\to 1}, \xi^{X(t)\to 2}, \ldots, \xi^{X(t)\to K}\right)$. Then the value of the investor's account between event times is modelled as evolving according to

$$
\begin{aligned}
dS(t) &= \left(\nu(t^-) - \alpha_{tot}\right) S(t^-)dt + \sigma(t^-)S(t^-)dZ(t) + \sum_{k\in\mathcal{K}} \left(\xi^{X(t^-)\to k} - 1\right) S(t^-)d\widetilde{N}^k(t) \\
&= \left(\nu(t^-) - \alpha_{tot} - \sum_{k\in\mathcal{K}} \left(\xi^{X(t^-)\to k} - 1\right) \lambda^k(t^-)\right) S(t^-)dt + \sigma(t^-)S(t^-)dZ(t) \\
&+ \sum_{k\in\mathcal{K}} \left(\xi^{X(t^-)\to k} - 1\right) S(t^-)dN^k(t) \\
&= \left(\mu(t^-) - \alpha_{tot}\right) S(t^-)dt + \sigma(t^-)S(t^-)dZ(t) + \sum_{k\in\mathcal{K}} \left(\xi^{X(t^-)\to k} - 1\right) S(t^-)dN^k(t)
\end{aligned} \tag{3.1}
$$

where $\mu(t^-) = \nu(t^-) - \sum_{k\in\mathcal{K}} \left(\xi^{X(t^-)\to k} - 1\right) \lambda^k(t^-)$. Note that the part of the instantaneous expected return that is due to jumps associated with all transitions out of the current regime is given by $\sum_{k\in\mathcal{K}} \left(\xi^{X(t^-)\to k} - 1\right) \lambda^k(t^-)S(t^-)dt$.

As noted above in Section 1, if we only use standard Black-Scholes arguments based on hedging with just the risk-free asset and the underlying asset, the market is incomplete. To address this, we extend the hedging argument using an expanded set of hedging instruments (details are given in Appendix B). To facilitate the related discussion, we rewrite equation (3.1) more compactly as

$$
dS = \left(\mu^j - \alpha_{tot}\right) Sdt + \sigma^j SdZ + \sum_{k=1}^{K} \left(\xi^{j\to k} - 1\right) SdX^{j\to k}, \quad j \in \mathcal{K} \tag{3.2}
$$

where μ^j and σ^j are the values of μ and σ in regime j and

$$
dX^{j\to k} = \begin{cases} 1 & \text{if there is a transition during } dt \text{ from regime } j \text{ to regime } k \\ 0 & \text{otherwise} \end{cases}.
$$

It is assumed that there can only be one transition during dt. Moreover,

$$
dX^{j\to k} = \begin{cases} 1 & \text{with probability } \lambda^{j\to k}dt + \delta^{j\to k} \\ 0 & \text{with probability } 1 - \lambda^{j\to k}dt - \delta^{j\to k} \end{cases}
$$

where $\delta^{j\to k} = 1$ if $j = k$ and is otherwise zero.

As in Section 2, the mutual fund in the investor's account tracks an index $\hat{S}$ which follows

$$
d\hat{S} = \mu^j \hat{S}dt + \sigma^j \hat{S}dZ + \sum_{k=1}^{K} \left(\xi^{j\to k} - 1\right) \hat{S}dX^{j\to k}, \quad j = 1, \ldots, K, \tag{3.3}
$$

where terms are defined analogously to those in equation (3.2).

Both equation (3.2) and equation (3.3) are specified under $\mathbb{P}$, the real world probability measure. Denote risk-neutral transition intensities by $\lambda_{\mathbb{Q}}^{j\to k}$, and define the quantity

$$
\rho^j = \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \left(\xi^{j\to k} - 1\right). \tag{3.4}
$$

Let $\mathcal{V}^j(S,t)$ denote the value of the contract in regime j.[10] It is shown in Appendix B that $\mathcal{V}^j$ can be determined by solving the coupled system of PDEs

$$\mathcal{V}_t^j + \frac{(\sigma^j)^2 S^2}{2}\mathcal{V}_{SS}^j + \left(r^j - \alpha_{tot} - \rho^j\right) S\mathcal{V}_S^j - r^j\mathcal{V}^j + \left[\alpha_m \mathcal{R}(t) + \mathcal{M}(t)\right] S + \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}(\mathcal{V}^k(\xi^{j\to k}S,t) - \mathcal{V}^j(S,t)) = 0 \ ,$$

for $j = 1,\ldots,K$. Rewriting the expression above in terms of backward time τ (i.e. using $V^j(S,\tau = T-t) = \mathcal{V}^j(S,t)$), we obtain

$$\begin{aligned} V_\tau^j &= \frac{(\sigma^j)^2 S^2}{2}V_{SS}^j + (r^j - \alpha_{tot} - \rho^j)SV_S^j - r^jV^j + \left[\alpha_m\mathcal{R}(t) + \mathcal{M}(t)\right] S \\ &+ \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}(V^k(\xi^{j\to k}S,\tau) - V^j(S,\tau)), \quad j = 1,\ldots,K. \end{aligned} \tag{3.5}$$

Note that equation (3.5) assumes that death benefits are paid continuously. In the case that death benefits are paid only at event times, then the generalization of equation (2.20) to the regime switching case is

$$\begin{aligned} V_\tau^j &= \frac{(\sigma^j)^2 S^2}{2}V_{SS}^j + (r^j - \alpha_{tot} - \rho^j)SV_S^j - r^jV^j + \alpha_m\mathcal{R}(t_{i-1})S \\ &+ \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}(V^k(\xi^{j\to k}S,\tau) - V^j(S,\tau)), \quad j = 1,\ldots,K. \end{aligned} \tag{3.6}$$

Without reference to market prices, the risk-neutral transition intensities $\lambda_{\mathbb{Q}}^{j\to k}$ are arbitrary non-negative functions, implying that the equivalent martingale pricing measure is not unique. However, the risk-neutral transition intensities are uniquely determined by the prices of the set of hedging instruments. In practice, the parameters in equation (3.5) can be calibrated to the observed prices of traded options. This means that the $\lambda_{\mathbb{Q}}^{j\to k}$ in equation (3.5) will correspond to those from the market's pricing measure.

Equation (3.5) holds between event times. All of the event conditions that were described above in Section 2.3 continue to hold in this regime switching context at the specified event times. In particular, the event conditions are simply applied to each individual regime.

4 Policy Holder Withdrawal

The pricing formulations outlined above take the policy holder's withdrawal strategy as given. In general terms, our intent above was to indicate how to calculate the fair value of the contract for a variety of potential withdrawal strategies. Specifying a particular withdrawal strategy amounts to specifying a model for γ_i as used above in Section 2.3. We now discuss some of the possibilities.

[10] In general, $\mathcal{V}^j$ will also depend on the guarantee account value A as well as the regime-dependent parameters r^j, μ^j, and σ^j, but this dependence is suppressed here for convenience.

4.1 Worst Case Hedging

If we take the position that the insurer should charge a price which ensures that no losses can occur, assuming that the claim is hedged, then the withdrawal strategy is assumed to be

$$\gamma_i = \arg\max_{\substack{\gamma \in [0,2] \\ \mathcal{C}=(0,\gamma)}} \left\{ V(S(\mathcal{C}), A(\gamma), \tau_i^-) + f(\mathcal{C}, S, A, t_i) \right\}$$

$$V(S, A, \tau_i^+) = V(S(\mathcal{C}_i), A(\mathcal{C}_i), \tau_i^-) + f(\mathcal{C}_i, S, A, t_i) \;\; ; \;\; \mathcal{C}_i = (0, \gamma_i) \; , \qquad (4.1)$$

with $S(\mathcal{C}_i), A(\mathcal{C}_i)$ and $f(\mathcal{C}_i, S, A, t_i)$ as given in Section 2.3. Assuming such a strategy by policy holders and hedging against it is obviously very conservative from the standpoint of the insurer, since it seeks to provide complete protection against policy holder withdrawal behaviour, given assumptions about parameter values such as volatilities. In other words, if investors follow this strategy, and if the insurer hedges continuously, the balance in the insurer's overall hedged portfolio will be zero. On the other hand, if investors deviate from this strategy, then the insurer's portfolio will have a positive balance.

4.2 Suboptimal Withdrawal

The withdrawal assumption underlying worst case hedging is often referred to as *optimal withdrawal*. This terminology is unfortunate, in that any withdrawal strategy different from strategy (4.1) is sub-optimal only in the sense that it does not maximize the cost of hedging. This may have little to do with any given policy holder's economic circumstances. Completely rational actions for a given policy holder may depart from the strategy (4.1). As noted by many authors, and particularly in Cramer et al. (2007), this is a controversial issue.

One possible approach that is quite simple is to assume that the contract holder will follow the default strategy of withdrawing at the contract rate at each event time t_i unless the extra value obtained by withdrawing optimally is greater than $\mathcal{F} G_r A(t_i - t_{i-1})$. In this case, $\mathcal{F} = 0$ corresponds to withdrawing optimally, while $\mathcal{F} = \infty$ corresponds to withdrawing at the contract rate. This approach was suggested in Ho et al. (2005) and Chen et al. (2008). It is worth noting that this approach is similar in spirit to a model proposed by Stanton (1995) in the context of mortgage-backed securities. In that model, mortgage holders can choose to refinance their mortgages for interest rate reasons or other exogenous factors. Mortgage holders face a transaction cost associated with prepayment (which could include both monetary fees and also non-monetary factors such as time and effort). Because of this transaction cost, mortgage holders do not prepay "optimally" (in the sense of maximizing the value of their financial option), even in the absence of exogenous factors.

4.3 Utility Models

A more complicated approach would be to specify a utility model to determine the withdrawal strategy of the policy holder. This would entail solving a system of PDEs, even in the single regime case. One PDE would be used to determine the withdrawal strategy (γ_i), based on maximizing the contract holder's utility. With this withdrawal strategy determined, the corresponding γ_i would be substituted into equation (2.14), and the contract value could then be determined by solving equation (2.7). Since the fee charged for hedging would influence the utility, these PDEs would be coupled. Of course, there are many possibilities here, with variations including the type of preferences, bequest motives, etc. As just one example, Moenig and Bauer (2011) consider a utility-based model for withdrawal behaviour. In addition, Moenig and Bauer (2011) also propose a model in which policy holders maximize (risk-neutral) after tax cash flows. To the extent that investors have to pay some form of tax on withdrawals, this is similar to the transaction cost notion noted above that was developed by Stanton (1995) in the context of mortgage-backed securities.

4.4 Summary of Withdrawal Models

A primary motivation in Moenig and Bauer (2011) is to explain the fact the observed fees charged by industry seem to be significantly lower than what would be suggested by "optimal" withdrawal assumptions. In this context it is useful to note that any strategy different from that in equation (2.14) cannot produce a larger fee, and will usually result in a smaller fee. Consequently, it is difficult to distinguish between various models of *sub-optimal* behaviour, simply because any such model will tend to produce a fee smaller than the worst case cost of hedging.

Hilpert et al. (2012) note that secondary markets for insurance products have been in place in many countries for some time, and appear to be growing. Financial third parties can potentially profit (through hedging strategies) from any financial instrument which is not priced using the worst case assumption in Section 4.1. As pointed out in Hilpert et al. (2012), this would lead to a general increase in fees charged by insurance companies for these products.

As noted above, Knoller et al. (2013) carry out an empirical study of policy holder behaviour in the Japanese variable annuity market. Their study shows that the moneyness of the guarantee has the greatest explanatory power for the rate at which policy holders surrender their policies. This supports the simple sub-optimal withdrawal model suggested in Section 4.2. In addition, Knoller et al. (2013) point out that several large Canadian insurers have recently suffered large losses related to increased lapse (surrender) rates, indicating that the fees being charged were insufficient to hedge worst case lapsation.

Consequently, in this paper we restrict attention to the worst case cost of hedging (Section 4.1) and the simple one parameter sub-optimal model (Section 4.2). We emphasize that once the strategy is specified (based on any reasonable model) the cost of hedging is determined from equations (2.11) and (2.14), in the single regime case. We defer investigation of other models such as utility-based approaches to future work, since in general an additional PDE must be solved to determine the policy holder strategy.

5 Numerical Method

We now describe several aspects of the numerical approach that we use to solve our valuation equations.

5.1 Localization

The PDE (3.5) is originally posed on the domain $(S, A, \tau) \in [0, \infty) \times [0, \infty) \times [0, T]$. For computational purposes, we need to truncate this domain to $(S, A, \tau) \in [0, S_{\max}] \times [0, A_{\max}] \times [0, T]$. Substituting $S = 0$ into equation (3.5), we obtain

$$V^j_\tau = -r^j V^j + \sum_{k \in \mathcal{K}_j} \lambda_{\mathbb{Q}}^{j \to k} \left(V^k(0, \tau) - V^j(0, \tau) \right). \tag{5.1}$$

Equation (5.1) serves as the boundary condition at $S = 0$. At $S = S_{\max}$, we impose the linearity condition

$$V^j_{SS} = 0; \quad S = S_{\max}. \tag{5.2}$$

This is an approximation. However, the error in regions of interest can be made small if $S_{\max}$ is sufficiently large. This will be verified in numerical tests.

No boundary condition is required at $A = 0$. We choose $A_{\max} = S_{\max}$, and impose an artificial cap on the contract of $A = A_{\max}$. This means that we replace equation (2.16) by

$$A(\mathcal{C}_i) = \min\left(A\left(1 + B(t_i)\right), A_{\max}\right) \tag{5.3}$$

The effect of this approximation can be made small by selecting $A_{\max} = S_{\max}$ sufficiently large. At $\tau = 0$, we have the obvious generalization of equation (2.12)

$$V^j(S, A, \tau = 0) = \mathcal{R}(T)S = 0. \tag{5.4}$$

5.2 Discretization

Between withdrawal times, we solve PDE (3.5) using second order (as much as possible) finite difference methods in the S direction, while still retaining the positive coefficient condition (Bélanger et al., 2009). Crank-Nicolson timestepping is used, with Rannacher smoothing (Rannacher, 1984). The discretized equations are solved at each timestep using a fixed point iteration scheme (Huang et al., 2011, 2012).

When determining the γ_i for worst-case hedging, we need to determine the withdrawal strategy (4.1). For a given level of PDE mesh refinement, we discretize the control $\gamma_i \in [0, 2]$. At each withdrawal date, the maxima are determined by a linear search. If data is needed at non-grid points, linear interpolation is used. The control grid discretization is reduced as we reduce the PDE mesh size, thus producing a convergent method. In fact, we observe that the worst case values for γ_i are always the discrete values $\{0, 1, 2\}$. In the case of pure GBM, it can be shown that the worst case controls are always $\gamma_i \in \{0, 1, 2\}$ (Azimzadeh, 2013). This appears to also be true for regime switching, although we have no proof of this. We emphasize that all our numerical results do not make this assumption.

5.3 Similarity Reduction

If $\lambda_{\mathbb{Q}}^{j\to k}$, $\xi^{j\to k}$, and σ^j are independent of S, then it is easy to verify that the solution $V^j(S, A, \tau)$ of PDE (3.5) with boundary conditions (5.4) and event conditions (2.15-2.19) has the property that

$$V^j(\eta S, \eta A, \tau) = \eta V^j(S, A, \tau) \tag{5.5}$$

for any scalar $\eta > 0$. Therefore, choosing $\eta = A^*/A$ we obtain

$$V^j(S, A, \tau) = \frac{A}{A^*} V^j\left(\frac{SA^*}{A}, A^*, \tau\right) \tag{5.6}$$

which means that we need only solve for a single representative value of $A = A^*$. This effectively reduces the system of coupled two dimensional PDEs to a system of coupled one dimensional PDEs, resulting in a large saving in computational cost. For a problem with K regimes, the entire pricing problem reduces to solution of K coupled one dimensional PDEs. We observe that the similarity reduction (5.5) was exploited in Shah and Bertsimas (2008). Also note that the similarity reduction holds for PDE (3.6) in the case that death benefits are only paid at year end.

Refinement	S, A nodes	Time steps
0	68	240
1	135	480
2	269	960
3	537	1920
4	1073	3840

Table 6.1: *Grid and timestep information for various levels of refinement. Normally, the similarity reduction is used so that there is one A node.*

5.4 Fair Hedging Fee

At $\tau = T$, the initial value of the guarantee level is set to the initial amount in the investment account $A_0 = S_0$. We can regard the solution as being parameterized by the rider fee α_g, i.e. $V(\alpha_g; S, A, \tau)$, so that the fair hedging fee (i.e. the cost of maintaining a replicating portfolio) is determined by solving the equation

$$V(\alpha_g; S = S_0, A = S_0, \tau = T) = S_0. \tag{5.7}$$

Equation (5.7) implies that since no up front fee is charged to enter into the contract, the fee α_g collected must be sufficient to cover the hedging costs.

We solve equation (5.7) by using Newton iteration, with tolerance

$$|\alpha_g^{k+1} - \alpha_g^k| < 10^{-8}, \tag{5.8}$$

with α_g^k being the k-th iterate. A sequence of grids is used, with the initial iterate for the finer grid being the converged solution from the coarse grid. Usually, only a single Newton iteration is required on the finest grid, which makes determination of the fair hedging fee very inexpensive.

6 Numerical Examples

6.1 Computational Parameters

In the localized domain $(S, A, \tau) \in [0, S_{\max}] \times [0, A_{\max}] \times [0, T]$, we set $S_{\max} = A_{\max} = 100S_0$, with $S_0 = 100$. Increasing $S_{\max}$ to $S_{\max} = 1000S_0$ resulted in no change to the solution to 10 digits. Since we use an unequally spaced grid, having a large $S_{\max}$ is computationally inexpensive.

We solve the PDE on a sequence of grids. At each refinement level, we insert a new fine grid node between each two coarse grid nodes and halve the timestep size. The grid and timestep information is shown in Table 6.1.

We use the DAV 2004R mortality table for a 65 year old German male (Base Table, first order) from Pasdika and Wolff (2005) to construct the mortality functions. For the convenience of the reader, this data is provided in Appendix C.

In the following, we will predominantly use the model which allows for continuously paid death benefits, i.e. equation (3.5). We use the model which allows for death benefits only paid at event times, equation (3.6), for a single validation case in Section 6.2.

6.2 Validation: Single Regime

In order to validate our basic numerical approach, we consider the special case where the contract holder withdraws deterministically at the contract rate at yearly intervals. Since there is no opti-

Parameter	Value
Volatility σ	.15
Interest rate r	.04
Expiry time T	57
Management fee α_m	0.0
Initial payment S_0	100
Mortality	DAV 2004R (65 year old male) (Pasdika and Wolff, 2005)
Mortality payments	At year end
Withdrawal rate G_r	.05
Bonus	No
Strategy	Deterministic withdrawal $G_r A(t_i - t_{i-1})$
Event times	yearly

TABLE 6.2: *Data used for validation example. Data from Holz et al. (2012), single regime.*

mal decision making in this problem, we can independently validate the solution. Once we have calculated the fair hedging fee α_g, we can verify that it is correct by using Monte Carlo simulation. We use the data suggested in Holz et al. (2012), and, in order to be comparable with that source, all death benefits are paid out at event times as described in Section 2.3. The stochastic process is a single regime GBM, with the parameters given in Table 6.2, along with contractual details.[11] We assume that the holder deposits the initial premium S_0 at $t = 0$, and begins withdrawing at $t = 1$ year.

Table 6.3 shows the results for the fair hedging fee computed using a sequence of refined grids, using the data in Table 6.2, assuming both no ratchet and an annual ratchet (2.15). Note that the estimated fair hedging fee is expressed in terms of basis points (bps), i.e. hundredths of a per cent. Table 6.4 shows the contract value at $(S, A, t) = (100, 100, 0)$, using α_g from the finest grid in Table 6.3. In Table 6.4, the ratio of successive changes in the value of the contract at $(S, A, t) = (100, 100, 0)$ is asymptotically approaching four, indicating quadratic convergence of the numerical method, as expected. The value converges to $V(100, 100, 0) = 100$ on the finest grid, consistent with Table 6.3. Both the similarity reduction (see Section 5.3) and the full two dimensional solution are shown for the ratchet case. Each of these methods appear to converge to the same value.

As noted above, there is no optimal decision making in this case as the policy holder simply withdraws at the contract rate, so we can use a straightforward Monte Carlo method to value these contracts. We use the α_g computed from the PDE method for both the ratchet and no-ratchet cases, and determine the value of the contract at $(S = A = S_0, t = 0)$ via a Monte Carlo simulation. If the fee α_g determined from the PDE method is correct, the value at $(S = A = S_0, t = 0)$ should converge to a value of $S_0 = 100$. Between event times, we use the exact solution for the GBM stochastic differential equation, and hence there is no timestepping error. Table 6.5 indicates that the fee obtained from the PDE solution does in fact appear to be correct.

Table 6.3 shows that the fair hedging fee reported for the same contracts in Holz et al. (2012) differs significantly from our results, especially if there is an annual ratchet feature. In particular, our estimated fee is about 7.5 basis points lower than that reported by Holz et al. (2012) without

[11] Note that the expiry time of $T = 57$ given in Table 6.2 (and in some subsequent tables) is based on the mortality table provided in Appendix C.

Refine level	Fair hedging fee α_g (bps)	
	No ratchet	Annual ratchet (2.15)
0	35.634697	64.645973
1	35.530197	64.834653
2	35.511140	64.898828
3	35.506538	64.915398
4	35.505335	64.919617
Result in Holz et al. (2012)	43	80

TABLE 6.3: *Results from validation test, data in Table 6.2. No ratchet and annual ratchet. Similarity reduction, single regime. The fee is shown in basis points (bps) (hundredths of a per cent).*

	Similarity reduction (5.3)				Full 2-d	
	No ratchet		Annual ratchet (2.15)		Annual ratchet (2.15)	
Refine level	Contract value	Ratio of changes	Contract value	Ratio of changes	Contract value	Ratio of Changes
0	100.012791		99.978287		100.048265	
1	100.002461		99.993258		100.010246	
2	100.000575	5.2	99.998351	3.0	100.003270	5.5
3	100.000119	4.2	99.999665	3.9	100.000804	2.8
4	100.000000	3.8	100.000000	3.8	100.000308	4.9

TABLE 6.4: *Results from validation test, data in Table 6.2, no ratchet case and annual ratchet. No ratchet case, α_g = 35.505335 bps. Ratchet case, α_g = 64.919617 bps. Contract value at $(S = A = S_0, t = 0)$, single regime.*

Number of simulations	No ratchet		Annual ratchet (2.15)	
	Contract value	Standard error	Contract value	Standard error
10^4	100.073	1.04	100.081	.90
10^5	99.899	.327	99.896	.28
10^6	99.970	.107	99.964	.090
10^7	99.998	.0327	99.994	.028

TABLE 6.5: *Monte Carlo validation. No ratchet case, α_g = 35.505335 bps. Ratchet case, α_g = 64.919617 bps. Contract value at $(S = A = S_0, t = 0)$, single regime.*

a ratchet, and about 15 basis points lower with an annual ratchet. One possibility is that we have misinterpreted some of the contractual specifications in Holz et al. (2012), leading to some subtle differences in the contracts that we are considering as compared to theirs, and these discrepancies result in different fees. Another potential explanation is that a Monte Carlo method was used to determine the fee by Holz et al. (2012). This may have introduced a significant error when calculating the fee unless a very large number of simulations was used.[12]

6.3 An Illustration of Complex Optimal Withdrawal Strategies

In situations where contract holders are assumed to behave optimally, it is interesting to note that their optimal withdrawal strategies can be quite complex.[13] As an example, it has been argued that the optimal strategy for the holder of a GLWB contract that does not include a ratchet or bonus provision must be to either withdraw at the contract rate or to fully surrender (Holz et al., 2012, p. 315). However, it is not clear whether this is still true if there are such features. To investigate this, we use the single regime data in Table 6.6. At any event time, the contract holder chooses the optimal strategy. Although our formulation allows any withdrawal amount in the range from no withdrawal to complete surrender, a few numerical tests indicated that the optimal strategy was either withdrawal at the contract rate, complete surrender, or not to withdraw at all.

The optimal withdrawal strategy varies over time, in part because the specified penalty for excess withdrawals declines. The particular case at $t = 0$ is shown as an illustrative example in Figure 6.1. For a fixed value of the guarantee level A, Figure 6.1 shows that it is optimal to withdraw at the contract rate if the investment account value is relatively low since the guarantee is in-the-money. On the other hand, if the investment account value is high, it is optimal to surrender the policy because the guarantee is out-of-the-money. In other words, the present value of withdrawing the entire balance of the investment account exceeds the value of either taking out the contractual amount and leaving the guarantee level unaffected or withdrawing nothing and having higher future guaranteed withdrawals due to the bonus feature. However, at intermediate values of the investment account it may be optimal to not withdraw at all (due to the bonus) or to withdraw at the contract rate. Note that the separators of the optimal strategy regions are straight lines passing through the origin. This is a consequence of the fact that the solution is homogeneous of degree one, as shown by equation (5.5).

6.4 Parameter Sensitivities: Single Regime

In Table 6.7 we specify the data for our base case. In all subsequent tests, we use level 3 grid refinement, which gives the fair fee for hedging correct to at least three digits.

We start by exploring the effects of some of the contract provisions. Table 6.8 shows results obtained by removing various contract features. It is interesting to observe that the bonus feature of the contract adds no value in our base case. In contrast, the surrender and ratchet features together account for about one half of the base case fair hedging fee.

We next consider the effects of the volatility parameter and the level of the risk-free interest rate. In agreement with many other studies cited above, Tables 6.9 and 6.10 show that the fair hedging fee is quite sensitive to these parameters. Due to the long term nature of the contract, it

[12] Holz et al. (2012) do not provide any information regarding the number of simulations used or the precision of their Monte Carlo estimates.

[13] We are using the word "optimal" here subject to the caveats mentioned above: this really means the strategy that generates the highest cost of hedging for the insurer, not the strategy that optimizes the particular economic circumstances of a given individual.

Parameter	Value
Volatility σ	.20
Interest rate r	.04
Penalty for excess withdrawal $\kappa(t)$	$0 \leq t \leq 1$: 3% , $1 < t \leq 2$: 2%, $2 < t \leq 3$: 1%, $3 < t < \infty$: 0%
Expiry time T	57
Management fee α_m	0.0
Fair hedging fee α_g	150 bps
Initial payment S_0	100
Mortality	DAV 2004R (65 year old male) (Pasdika and Wolff, 2005)
Mortality payments	Continuous
Withdrawal rate G_r	.05 annually
Bonus (no withdrawal)	.06 annually
Ratchet	Every three years
Strategy	Optimal
Event times	yearly

TABLE 6.6: *Data used for optimal strategy example.*

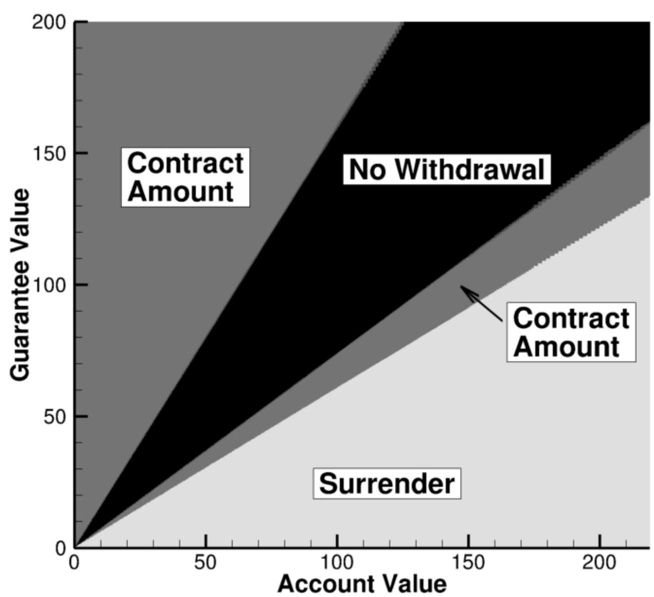

FIGURE 6.1: *Optimal withdrawal strategy at $t = 0$. Data in Table 6.6.*

Parameter	Value
Volatility σ	.15
Interest rate r	.04
Penalty for excess withdrawal $\kappa(t)$	$0 \leq t \leq 1$: 5%, $1 < t \leq 2$: 4%, $2 < t \leq 3$: 3%, $3 < t \leq 4$: 2%, $4 < t \leq 5$: 1%, $5 < t < \infty$: 0%
Expiry time T	57
Management fee α_m	0.0
Initial payment S_0	100
Mortality	DAV 2004R (65 year old male) (Pasdika and Wolff, 2005)
Mortality payments	Continuous
Withdrawal rate G_r	.05 annually
Bonus (no withdrawal)	.05 annually
Ratchet	Every three years
Strategy	optimal
Event times	yearly

Table 6.7: *Base case data.*

Case	Fair hedging fee (bps)
Base	70.7
No bonus	70.7
No surrender	52.4
No ratchet	63.1
No bonus, surrender, or ratchet	36.2

Table 6.8: *Effect of contractual provisions: fair hedging fee α_g for the data in Table 6.7, except as noted.*

Case	Fair hedging fee (bps)
Base ($\sigma = .15$)	70.7
$\sigma = .10$	27.4
$\sigma = .20$	132
$\sigma = .25$	209

TABLE 6.9: *Effect of volatility: fair hedging fee α_g for the data in Table 6.7, except as noted.*

Case	Fair hedging fee (bps)
Base ($r = .04$)	70.7
$r = .02$	242
$r = .06$	21.2

TABLE 6.10: *Effect of risk-free rate: fair hedging fee α_g for the data in Table 6.7, except as noted.*

may be particularly important to allow these parameters to be stochastic (again, as suggested in several papers cited above). We will address this below in a regime switching example.

GLWB riders are often marketed as an add-on to mutual funds managed by insurance companies. In many cases, these mutual funds already have fairly hefty management fees. Table 6.11 illustrates the effect of these management fees on the no-arbitrage guarantee fee. Consistent with similar results for GMWBs reported in Chen et al. (2008), the GLWB rider fee increases significantly as the underlying mutual fund management fee increases. This is easily understood. The guarantee applies initially to S_0, and never decreases unless excess withdrawals are made. The mutual fund management fees act as a drag on the investment account $S(t)$, increasing the value of the guarantee. This raises the following interesting observation. If an insurer wishes to provide its customers with the cheapest possible insurance (rather than collecting management fees for mutual funds), the best strategy would seem to be to provide a GLWB rider on an inexpensive exchange traded index fund, rather than a managed mutual fund.

As a final example for the single regime setting, we next explore the effect of sub-optimal withdrawal using the approach suggested in Ho et al. (2005) and Chen et al. (2008), and described in Section 4.2. We assume that the holder of the contract will withdraw at the contract rate at each event time t_i unless the extra value obtained by withdrawing optimally is greater than $\mathcal{F} G_r A(t_i - t_{i-1})$. In this case, $\mathcal{F} = 0$ corresponds to withdrawing optimally, while $\mathcal{F} = \infty$ corresponds to withdrawing at the contract rate. Table 6.12 shows that $\mathcal{F} = 0.1$ results in a fee very close to the optimal withdrawal assumption, while $\mathcal{F} = 1.0$ gives rise to a fee very close to that found

Case	Fair hedging fee (bps)
Base ($\alpha_m = 0$)	70.7
$\alpha_m = 50$ (bps)	84.7
$\alpha_m = 100$ (bps)	101
$\alpha_m = 150$ (bps)	119
$\alpha_m = 200$ (bps)	141

TABLE 6.11: *Effect of management fee: fair hedging fee α_g for the data in Table 6.7, except as noted.*

Case	Fair hedging fee (bps)
Base ($\alpha_m = 0$)	70.7
$\mathcal{F} = .05$	70.4
$\mathcal{F} = .10$	69.6
$\mathcal{F} = .5$	57.7
$\mathcal{F} = 1.0$	52.5
$\mathcal{F} = \infty$	52.4

TABLE 6.12: *Effect of sub-optimal withdrawal strategy: fair hedging fee α_g for the data in Table 6.7, except that the holder withdraws at the contract rate unless the extra value obtained by withdrawing optimally is larger than $\mathcal{F}G_rA(t_i - t_{i-1})$.*

for $\mathcal{F} = \infty$, i.e. always withdraw at the contract rate. If it is deemed desirable to value these contracts using sub-optimal behaviour, $\mathcal{F}$ could be estimated from empirical data on withdrawals. Alternatively, one could use contract pricing data to infer the degree of sub-optimality that is being implicitly assumed by the insurer.

6.5 Regime Switching

As a basic initial test, the numerical method used to solve the regime switching PDE (3.5) was applied to value plain vanilla options. The prices found were in close agreement with those from analytic solutions (where available) and Fourier timestepping methods (Jackson et al., 2008). We refer the reader to Sohrabi (2010) for the detailed validation.

We now consider the valuation of a GLWB under a regime switching model as described in Section 3. The base case contract parameters and regime switching data are given in Table 6.13. The regime switching parameters were obtained in O'Sullivan and Moloney (2010) by calibration to FTSE 100 options in January of 2007. We assume that any dividends paid out by the the underlying index are immediately reinvested in the index. For the purposes of computing the fair hedging fee, it is assumed that the process is in regime one at $t = 0$.

Our base case scenario assumes that the system is initially in regime one. Table 6.14 shows that if the system is initially in regime two (which is the more volatile regime), the fair hedging fee increases substantially.

Although the fitting exercise in O'Sullivan and Moloney (2010) used the same risk-free rate in both regimes, this is not necessary and probably not realistic. In fact, at least in the current aftermath of the financial crisis, it appears that we are in a regime characterized by low interest rates and high volatility. In view of the strong effect of interest rates on the value of the GLWB guarantee that was noted in Section 6.4, we will explore the sensitivity of the base case regime switching results to regime-dependent interest rates. Table 6.14 shows that adding regime dependent interest rates can dramatically increase the value of the guarantee. This is, of course, due to having long periods of low interest rates ($\simeq$ 19 years) interspersed with shorter periods ($\simeq$ 7.3 years) of high interest rates. We remind the reader here that the transition probabilities are risk-adjusted, so that the duration in each regime is under a risk-neutral setting that is obtained by calibration to market prices. In other words, these are not the same durations as for the objective probability measure.

Table 6.14 also shows the effect of increasing the volatilities in each regime, which, as one might expect, causes a large increase in the fair hedging fee. Finally, the table also contains a comparison between $\mathcal{F} = 0$ (optimal withdrawal) with $\mathcal{F} = \infty$ (withdrawal at the contract rate). For the case considered, optimal withdrawal strategies result in a fair hedging fee that is close to double the

Parameter	Value
Volatilities $\{\sigma_1, \ \ \sigma_2\}$	$\{.0832, \ \ .2141\}$
Interest rates $\{r_1, \ \ r_2\}$	$\{.0521, \ \ .0521\}$
Transition intensities $\{\lambda_{\mathbb{Q}}^{1\to2}, \ \ \lambda_{\mathbb{Q}}^{2\to1}\}$	$\{.0525, \ \ .1364\}$
Jump sizes $\{\xi^{1\to2}, \ \ \xi^{2\to1}\}$	$\{1.0, \ \ 1.0\}$
Initial regime ($t = 0$)	One
Penalty for excess withdrawal $\kappa(t)$	0.0
Expiry time T	57
Management fee α_m	0.01
Initial payment S_0	100
Mortality	DAV 2004R (65 year old male) (Pasdika and Wolff, 2005)
Mortality payments	Continuous
Withdrawal rate G_r	.05 annually
Bonus (no withdrawal)	.05 annually
Ratchet	Every three years
Strategy	Optimal
Event times	yearly

TABLE 6.13: *Data for regime switching example.*

corresponding fee when withdrawals occur at the contract rate.

While the results above show that regime switching can have a significant effect on the values of GLWB contracts, a question which naturally arises is whether a single regime model can provide an effective approximation to regime switching. For a stochastic process with K regimes, the $K \times K$ generator matrix Q of the risk-neutral Markov process is given by

$$[Q]_{ij} = \begin{cases} \lambda_{\mathbb{Q}}^{i\to j} & i \neq j \\ -\sum_{k\neq i} \lambda_{\mathbb{Q}}^{i\to k} & i = j \end{cases} \tag{6.1}$$

where $\lambda_{\mathbb{Q}}^{i\to j}$ are the risk-neutral transition intensities. Since these intensities are risk-neutral, we can consider that the market price of regime switch risk is embedded in these risk-neutral intensities.

Let the interest rate in regime j be r^j, and define the interest rate matrix R as the diagonal matrix with $R_{jj} = r^j$. The price of a zero coupon bond maturing in T years is given by

$$V = e^{(Q-R)\ T} \cdot \mathbf{1} \tag{6.2}$$

where $\mathbf{1} = [\mathbf{1}, \ldots, \mathbf{1}]'$, $e^{(Q-R)\ T}$ is the matrix exponential and $[V]_i$ is the zero coupon bond value assuming we are in regime i at $t = 0$. The effective single regime interest rate, assuming the process is in regime i at $t = 0$ is then given by

$$(r_{\text{eff}})_i = -\frac{\log(V_i)}{T}. \tag{6.3}$$

In order to determine the effective single regime volatility, we first price a T-year European call option (at the money) using a regime switching model. The effective single regime volatility is then the implied volatility which matches this price, assuming that the process is in regime i, and the effective interest rate is $(r_{\text{eff}})_i$. The results in Table 6.15 were determined using $T = 10$.

Case	Fair hedging fee (bps)
Base case (Table 6.13)	31.6
Initial regime ($t = 0$): Two	123
$r_1 = .04$, $r_2 = .06$	52.1
$r_1 = .03$, $r_2 = .07$	85.2
$r_1 = .02$, $r_2 = .08$	150
$\sigma_1 = .10$, $\sigma_2 = .20$	38.0
$\sigma_1 = .15$, $\sigma_2 = .25$	86.1
$\mathcal{F} = 0$ (Optimal withdrawal)	
$\sigma_1 = .15$, $\sigma_2 = .25$	114
$r_1 = .04$, $r_2 = .08$	
$\mathcal{F} = \infty$ (Withdrawal at contract rate)	
$\sigma_1 = .15$, $\sigma_2 = .25$	65.7
$r_1 = .04$, $r_2 = .08$	

TABLE 6.14: *Fair hedging fee α_g for the regime switching data in Table 6.13, except as noted.*

Some interesting patterns are revealed in this table. First, consider just the results for the regime switching model. Higher volatility in a regime tends to increase the fair hedging fee for that regime. A lower interest rate in a given regime also causes the fee to rise for that regime. These results are of course consistent with what was observed earlier in the single regime context (see Tables 6.9 and 6.10). However, at least for the parameter values used in Table 6.15, the volatility effect appears to be stronger. As a general rule, the initial regime with higher volatility results in a higher fair hedging fee than does the initial regime having a lower interest rate. The only exception to this is when the low (high) volatility regime has an interest rate of 2% (8%), and even then the fee is not that different (150 bps vs. 141 bps).

Turning to the single regime approximation to the regime switching model, it might seem that the fair hedging fees for the one-factor approximation in any given regime should lie between the fair hedging fees across the two regimes in the switching model. The simple intuition for this is that the single regime presumably acts as a blend of the two regimes. However, it turns out that while this is the general pattern, it is not always the case. It does occur when interest rates are constant across the regimes, but not always when they differ, particularly when there is a relatively big difference between the level of the interest rate under the two regimes. Overall, it would be difficult to draw the conclusion that the single regime model can consistently provide an effective approximation to the regime switching model. There are certainly parameter values for which this would be the case, but there are also situations where the approximation would be quite poor. In summary, the use of a single regime process with blended parameters to approximate a true regime switching process gives unreliable estimates for the fair hedging fee. The consequences of this could be significant for insurers offering these contracts. As an example, consider the case where $\sigma_1 = 8.32\%$, $\sigma_2 = 21.41\%$, $r_1 = 7\%$, and $r_2 = 3\%$. If the economy is initially in regime 1, the appropriate fee for hedging under the regime-switching model is about 37 basis points, 5 basis points higher than the estimated fee from single regime approximation. The insurer would be exposed to losses here since the fees charged would not be sufficient to completely hedge the risk exposure. While 5 basis points may not seem to be a large difference, its effects do add up over time since it applies every year and these are long-term contracts. More significantly, suppose that the economy is initially in regime 2. According to the single regime approximation, the fee charged should be 164

σ_1	σ_2	r_1	r_2	Initial Regime	Fair Hedging Fee (bps) Regime Switching	Single Regime Approximation
.0832	.2141	.0521	.0521	1	31.6	39.6
.0832	.2141	.0521	.0521	2	126	99.5
.0832	.2141	.04	.06	1	52.1	62.9
.0832	.2141	.04	.06	2	127	94.3
.0832	.2141	.06	.04	1	29.8	36.2
.0832	.2141	.06	.04	2	151	133
.0832	.2141	.03	.07	1	85.2	97.8
.0832	.2141	.03	.07	2	97.8	85.1
.0832	.2141	.07	.03	1	36.7	31.9
.0832	.2141	.07	.03	2	239	164
.0832	.2141	.02	.08	1	150	165
.0832	.2141	.02	.08	2	141	81.5
.0832	.2141	.08	.02	1	30.9	30.2
.0832	.2141	.08	.02	2	213	207
.15	.25	.0521	.0521	1	86.1	92.7
.15	.25	.0521	.0521	2	184	158
.10	.20	.0521	.0521	1	38	43
.10	.20	.0521	.0521	2	108	91.3

Table 6.15: *Fair hedging fee. Regime switching parameters as in Table 6.13, except as noted. Single regime approximation parameters computed as described in Section 6.5.*

basis points. This is dramatically lower (75 basis points) than the appropriate fee arising from the regime-switching model. Such a scenario would leave the insurer with a significantly underfunded hedging strategy. While this is merely illustrative, it does point to the potential benefits of adopting a regime-switching model, as well as the need for future detailed empirical research into parameter estimation for such models.

7 Conclusions

In this article, we have developed an implicit PDE method for valuing GLWB contracts assuming the underlying risky asset follows a Markov regime switching process. Assuming a process with K regimes, use of a similarity transformation reduces this problem to solving a system of K coupled one dimensional PDEs. The fair hedging fee (i.e. the cost of maintaining the replicating portfolio) is determined using a sequence of grids, coupled with Newton iteration. The entire procedure is computationally inexpensive. Since the valuation framework is developed independently of the withdrawal strategy, the methodology can easily accommodate a variety of alternative assumptions about policy holder behaviour.

The long term nature of these guarantees suggests that regime switching is a parsimonious model capable of modeling long term economic trends, having both stochastic volatilities and interest rates. Regime switching offers a much simpler approach compared to specifying volatility and/or interest rates as following separate diffusion processes. Another advantage of regime switching processes is that the model parameters are easy to interpret. A possible extension to the regime switching concept would be to include different mortality regimes, which could be used to model stochastic mortality improvement.

Our numerical tests indicate that regime dependent interest rate and volatility parameters have a large effect on the fair hedging fee, as does the assumption of optimal versus non-optimal policy holder withdrawal strategies. We conclude by pointing out that our method is quite flexible in that it can accommodate a wide variety of policy holder withdrawal strategies such as ones derived from utility-based models. We defer exploration of such models to future research.

Appendices

A Single Regime: Equation Between Event Times

This appendix provides a derivation of the PDE for the contract value between event times in the case of a single regime. The value of the underlying investment account S is assumed to evolve according to equation (2.5). The mutual fund in the investment account is assumed to track the index $\hat{S}$, which follows equation (2.6). Moreover, the fraction of the original owners of the contract who remain alive at time t is given by equation (2.1).

Suppose that the writer of the guarantee forms a portfolio Π which, in addition to being short the guarantee, is long x units of the index $\hat{S}$, i.e.

$$\Pi = -\mathcal{U}(S, A, t) + x\hat{S}. \tag{A.1}$$

Recalling that between event times $dA = 0$, then, by Itô's lemma,

$$d\Pi = -\left[\frac{\sigma^2 S^2}{2}\mathcal{U}_{SS} + (\mu - \alpha_{tot})\, S\mathcal{U}_S + \mathcal{U}_t\right] dt - \sigma S\mathcal{U}_S dZ + x\mu\hat{S}dt + x\sigma\hat{S}dZ + \mathcal{R}(t)\alpha_g S dt, \tag{A.2}$$

where the last term reflects fees collected from the fraction of the original holders of the contract who are still alive at time t that are used to fund the cost of hedging the guarantee. Setting $x = \mathcal{U}_S S/\hat{S}$ in equation (A.2) gives

$$d\Pi = -\left[\frac{\sigma^2 S^2}{2}\mathcal{U}_{SS} - \alpha_{tot} S\mathcal{U}_S + \mathcal{U}_t\right] dt + \mathcal{R}(t)\alpha_g S dt. \tag{A.3}$$

Since the portfolio is now (locally) riskless, it must earn the risk-free interest rate r. Setting $d\Pi = r\Pi dt$ results in

$$\begin{aligned} -\left[\frac{\sigma^2 S^2}{2}\mathcal{U}_{SS} - \alpha_{tot} S\mathcal{U}_S + \mathcal{U}_t\right] dt + \mathcal{R}(t)\alpha_g S dt &= r\left[\frac{S}{\hat{S}}\mathcal{U}_S\hat{S} - \mathcal{U}\right] \\ \Rightarrow \mathcal{U}_t + \frac{\sigma^2 S^2}{2}\mathcal{U}_{SS} + (r - \alpha_{tot})\, S\mathcal{U}_S + -r\mathcal{U} - \mathcal{R}(t)\alpha_g S &= 0. \end{aligned} \tag{A.4}$$

Now consider the value of the entire contract $\mathcal{V}(S, A, t)$. Again, recall that A does not change between event times. Then, $\mathcal{V}(S, A, t)$ is just the sum of the value of the guarantee and the amount in the investment accounts of the surviving contract holders, i.e.

$$\mathcal{V}(S, A, t) = \mathcal{U}(S, A, t) + \mathcal{R}(t)S. \tag{A.5}$$

Since $\mathcal{U}(S, A, t) = \mathcal{V}(S, A, t) - \mathcal{R}(t)S$,

$$\begin{aligned} \mathcal{U}_t &= \mathcal{V}_t - \mathcal{R}'(t)S = \mathcal{V}_t + \mathcal{M}(t)S \quad \text{(from equation (2.1))} \\ \mathcal{U}_S &= \mathcal{V}_S - \mathcal{R}(t) \\ \mathcal{U}_{SS} &= \mathcal{V}_{SS}. \end{aligned}$$

Substituting these expressions into (A.4) and simplifying using $\alpha_{tot} = \alpha_g + \alpha_m$ gives

$$\mathcal{V}_t + \frac{\sigma^2 S^2}{2}\mathcal{V}_{SS} + (r - \alpha_{tot})\, S\mathcal{V}_S - r\mathcal{V} + [\alpha_m \mathcal{R}(t) + \mathcal{M}(t)]\, S = 0. \tag{A.6}$$

B Multiple Regimes

This appendix provides a derivation of the set of PDEs governing the value of the contract between event times when there are K possible regimes. The material presented here draws heavily from Kennedy (2007). As in Appendix A, we start by considering the value of the guarantee portion of the contract, denoted by $\mathcal{U}^j$ in regime j. To simplify the notation, let θ^j be a vector of regime-dependent parameters, $\theta^j = \left(\mu^j, \sigma^j, r^j\right)$.

In general, $\mathcal{U}^j$ depends on S, the guarantee account value A, the parameter vector θ, and time t, i.e. $\mathcal{U}^j\left(S, A, \theta, t\right)$. However, since we assume that withdrawals can only happen at preset discrete times, A only changes at these times, and $dA = 0$ between event times. From Itô's lemma,

$$d\mathcal{U} = \bar{\mu}^j dt + \bar{\sigma}^j dZ + \sum_{k \in \mathcal{K}_j} \Delta\mathcal{U}^{j\to k} dX^{j\to k}, \tag{B.1}$$

where

$$\begin{aligned}
\bar{\mu}^j &= \mathcal{U}^j_t + \frac{(\sigma^j)^2 S^2}{2}\mathcal{U}^j_{SS} + \left(\mu^j - \alpha_{tot}\right) S\mathcal{U}^j_S,\\
\bar{\sigma}^j &= \sigma^j S\mathcal{U}^j_S,\\
\Delta\mathcal{U}^{j\to k} &= \mathcal{U}^k\left(\xi^{j\to k} S, \theta^k, t\right) - \mathcal{U}^j\left(S, \theta^j, t\right).
\end{aligned} \tag{B.2}$$

Now consider a set of K hedging instruments, each of which has a value that is dependent on $\hat{S}$, as well as the parameter vector θ and time. Denote the value of the n-th hedging instrument by $F_n(\hat{S}, \theta, t)$, $n = 1, \ldots, K$. As shown in Kennedy (2007), if these hedging instruments form a nonredundant set, then it is possible to construct a perfect hedge. Note that the hedging instruments can include $\hat{S}$ as well as any nonlinear securities such as traded option contracts on $\hat{S}$, not necessarily the change of state contracts suggested in Guo (2001). For example, F_n could be short term puts or calls with different strikes which are rolled over upon expiry. For the moment, we will not assume that $\hat{S}$ itself is one of these instruments. By Itô's lemma,

$$dF_n = \hat{\mu}^j_n dt + \hat{\sigma}^j_n dZ + \sum_{k \in \mathcal{K}_j} \Delta F^{j\to k}_n dX^{j\to k}, \tag{B.3}$$

where

$$\begin{aligned}
\hat{\mu}^j_n &= F^j_{n,t} + \frac{\sigma_j^2 \hat{S}^2}{2} F^j_{n,SS} + \mu^j \hat{S} F^j_{n,S}\\
\hat{\sigma}^j_n &= \sigma^j \hat{S} F^j_{n,S}\\
\Delta F^{j\to k}_n &= F_n(\xi^{j\to k}\hat{S}, \theta^k, t) - F_n(\hat{S}, \theta^j, t).
\end{aligned} \tag{B.4}$$

Note that $F^j_{n,t}$ and $F^j_{n,S}$ in the above denote partial derivatives of the n-th hedging instrument in regime j with respect to t and S respectively, while $F^j_{n,SS}$ is the second partial derivative of the n-th hedging instrument in regime j with respect to S.

A hedging portfolio Π is formed which is short the guarantee $\mathcal{U}$ and which has an amount w_n invested in the n-th hedging instrument, i.e.

$$\Pi = -\mathcal{U} + \sum_{n=1}^{K} w_n F_n^j. \tag{B.5}$$

The value of Π will evolve according to

$$\begin{aligned} d\Pi &= -d\mathcal{U} + \sum_{n=1}^{K} w_n dF_n^j + \mathcal{R}(t)\alpha_g S dt \\ &= \left[-\bar{\mu}^j + \sum_{n=1}^{K} w_n \hat{\mu}_n^j + \mathcal{R}(t)\alpha_g S \right] dt + \left[-\bar{\sigma}^j + \sum_{n=1}^{K} w_n \hat{\sigma}_n^j \right] dZ \\ &+ \sum_{k\in\mathcal{K}_j} \left[-\Delta\mathcal{U}^{j\to k} + \sum_{n=1}^{K} w_n \Delta F_n^{j\to k} \right] dX^{j\to k}. \end{aligned} \tag{B.6}$$

As in equation (A.2), the term involving $\mathcal{R}(t)\alpha_g S$ reflects fees paid to fund the cost of the guarantee from the fraction of policyholders who are still alive at time t. Equation (B.6) contains two types of risk: diffusion risk involving dZ and regime switching risk involving $dX^{j\to k}$. The diffusion risk can be hedged away by setting

$$\sum_{n=1}^{K} w_n \hat{\sigma}_n^j = \bar{\sigma}^j, \tag{B.7}$$

while the regime switching risk can be eliminated by setting

$$\sum_{n=1}^{K} w_n \Delta F_n^{j\to k} = \Delta\mathcal{U}^{j\to k}, \quad k = 1, \ldots, K, \quad k \neq j. \tag{B.8}$$

Assuming that (B.7)-(B.8) are satisfied, then Π is (locally) risk-free, and so no-arbitrage requires that $d\Pi = r^j \Pi dt$. This implies that

$$\begin{aligned} -\bar{\mu}^j + \sum_{n=1}^{K} w_n \hat{\mu}_n^j + \mathcal{R}(t)\alpha_g S &= r^j \left[-\mathcal{U}^j + \sum_{n=1}^{K} w_n F_n^j \right] \\ \Rightarrow \sum_{n=1}^{K} w_n \left(\hat{\mu}_n^j - r^j F_n^j \right) &= \bar{\mu}^j - r^j \mathcal{U}^j - \mathcal{R}(t)\alpha_g S. \end{aligned} \tag{B.9}$$

In matrix form, we can write out equations (B.7)-(B.9) as follows:

$$\begin{bmatrix} \hat{\sigma}_1^j & \hat{\sigma}_2^j & \ldots & \hat{\sigma}_K^j \\ \Delta F_1^{j\to 1} & \Delta F_2^{j\to 1} & \ldots & \Delta F_K^{j\to 1} \\ \Delta F_1^{j\to 2} & \Delta F_2^{j\to 2} & \ldots & \Delta F_K^{j\to 2} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta F_1^{j\to j-1} & \Delta F_2^{j\to j-1} & \ldots & \Delta F_K^{j\to j-1} \\ \Delta F_1^{j\to j+1} & \Delta F_2^{j\to j+1} & \ldots & \Delta F_K^{j\to j+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta F_1^{j\to K} & \Delta F_2^{j\to K} & \ldots & \Delta F_K^{j\to K} \\ \hat{\mu}_1^j - r^j F_1^j & \hat{\mu}_2^j - r^j F_2^j & \ldots & \hat{\mu}_K^j - r^j F_K^j \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ w_{K-1} \\ w_K \end{bmatrix} = \begin{bmatrix} \bar{\sigma}^j \\ \Delta\mathcal{U}^{j\to 1} \\ \Delta\mathcal{U}^{j\to 2} \\ \vdots \\ \Delta\mathcal{U}^{j\to j-1} \\ \Delta\mathcal{U}^{j\to j+1} \\ \vdots \\ \Delta\mathcal{U}^{j\to K} \\ \bar{\mu}^j - r^j \mathcal{U}^j - \mathcal{R}(t)\alpha_g S \end{bmatrix}. \tag{B.10}$$

Equation (B.10) is a system of $K+1$ equations in K unknowns, $w_1, w_2, \ldots, w_K$, so these equations must be linearly dependent. Denote the i-th equation in (B.10) by b_i, $i = 1, \ldots, K+1$, and consider the linear combination

$$\Lambda^j b_1 - \lambda_{\mathbb{Q}}^{j\to 1} b_2 - \lambda_{\mathbb{Q}}^{j\to 2} b_3 - \cdots - \lambda_{\mathbb{Q}}^{j\to K} b_K - b_{K+1}.$$

Each component of this combination must be zero for some choice of Λ^j, $\lambda^{j\to k}$, $k = 1, \ldots, K$, $k \neq j$. From the right hand side of (B.10), this implies that

$$\Lambda^j \bar{\sigma}^j - \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \Delta\mathcal{U}^{j\to k} = \bar{\mu}^j - r^j \mathcal{U}^j - \mathcal{R}(t)\alpha_g S.$$

Using (B.2), the expression above becomes

$$\Lambda^j \sigma^j S \mathcal{U}_S^j - \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \Delta\mathcal{U}^{j\to k} = \frac{(\sigma^j)^2 S^2}{2} \mathcal{U}_{SS}^j + \left(\mu^j - \alpha_{tot}\right) S \mathcal{U}_S^j + \mathcal{U}_t^j - r^j \mathcal{U}^j - \mathcal{R}(t)\alpha_g S,$$

which is equivalent to

$$\mathcal{U}_t^j + \frac{(\sigma^j)^2 S^2}{2} \mathcal{U}_{SS}^j + \left(\mu^j - \alpha_{tot} - \Lambda^j \sigma^j\right) S \mathcal{U}_S^j - r^j \mathcal{U}^j - \mathcal{R}(t)\alpha_g S + \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \Delta\mathcal{U}^{j\to k} = 0. \quad \text{(B.11)}$$

In equation (B.11), Λ^j is the market price of diffusion risk and the $\lambda_{\mathbb{Q}}^{j\to k}$ terms are risk-neutral transition intensities between regimes. Note that this equation holds regardless of whether or not $\hat{S}$ is one of the hedging instruments.

Now consider the first column of the left hand side of (B.10). The same linear combination as above implies that

$$\Lambda^j \hat{\sigma}_1^j - \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \Delta F_1^{j\to k} = \hat{\mu}_1^j - r^j F_1^j, \quad \text{(B.12)}$$

for the first hedging instrument. If this hedging instrument is in fact $\hat{S}$, then we can eliminate the Λ^j term, as follows. Using equation (B.4), and specifying $F_1 = \hat{S}$, we have

$$\begin{aligned} \hat{\mu}_1^j &= \mu^j \hat{S} \\ \hat{\sigma}_1^j &= \sigma^j \hat{S} \\ \Delta F_1^{j\to k} &= \left(\xi^{j\to k} - 1\right) \hat{S}. \end{aligned}$$

Substituting these expressions into (B.12) gives

$$\begin{aligned} \Lambda^j \sigma^j \hat{S} - \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \left(\xi^{j\to k} - 1\right) \hat{S} &= \left(\mu^j - r^j\right) \hat{S} \\ \Rightarrow \mu^j - \Lambda^j \sigma^j &= r^j - \rho^j \end{aligned} \quad \text{(B.13)}$$

where

$$\rho^j = \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k} \left(\xi^{j\to k} - 1\right).$$

In turn, substitution of (B.13) into (B.11) gives

$$\mathcal{U}_t^j + \frac{(\sigma^j)^2 S^2}{2}\mathcal{U}_{SS}^j + \left(r^j - \alpha_{tot} - \rho^j\right) S\mathcal{U}_S^j - r^j\mathcal{U}^j - \mathcal{R}(t)\alpha_g S + \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}\Delta\mathcal{U}^{j\to k} = 0. \tag{B.14}$$

As in the single regime case described in Appendix A, the value of the entire contract $\mathcal{V}(S, A, t) = \mathcal{U}(S, A, t) + \mathcal{R}(t)S$. This implies

$$\begin{aligned}
\mathcal{U}_t^j &= \mathcal{V}_t^j - \mathcal{R}'(t)S = \mathcal{V}_t^j + \mathcal{M}(t)S \\
\mathcal{U}_S^j &= \mathcal{V}_S^j - \mathcal{R}(t) \\
\mathcal{U}_{SS}^j &= \mathcal{V}_{SS}^j \\
\sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}\Delta\mathcal{U}^{j\to k} &= \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}\Delta\mathcal{V}^{j\to k} - \mathcal{R}(t)S \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}\left(\xi^{j\to k} - 1\right) \\
&= \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}\Delta\mathcal{V}^{j\to k} - \mathcal{R}(t)S\rho^j.
\end{aligned}$$

Substitution of these expressions into (B.14) gives

$$\mathcal{V}_t^j + \frac{(\sigma^j)^2 S^2}{2}\mathcal{V}_{SS}^j + \left(r^j - \alpha_{tot} - \rho^j\right) S\mathcal{V}_S^j - r^j\mathcal{V}^j + \left[\alpha_m\mathcal{R}(t) + \mathcal{M}(t)\right] S + \sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}\Delta\mathcal{V}^{j\to k} = 0. \tag{B.15}$$

Note that equation (B.15) holds between withdrawal times for $j = 1, 2, \ldots, K$, i.e. it is a coupled system of K one-dimensional PDEs. Since $\Delta\mathcal{V}^{j\to k} = \mathcal{V}^k(\xi^{j\to k}S, \theta^k, t) - \mathcal{V}^j(S, \theta^j, t)$, equation (B.15) can also be written in the form

$$\begin{aligned}
\mathcal{V}_t^j + \frac{(\sigma^j)^2 S^2}{2}\mathcal{V}_{SS}^j + \left(r^j - \alpha_{tot} - \rho^j\right) S\mathcal{V}_S^j - r^j\mathcal{V}^j + \left[\alpha_m\mathcal{R}(t) + \mathcal{M}(t)\right] S + \\
\sum_{k\in\mathcal{K}_j} \lambda_{\mathbb{Q}}^{j\to k}(\mathcal{V}^k(\xi^{j\to k}S, t) - \mathcal{V}^j(S, t)) = 0,
\end{aligned} \tag{B.16}$$

where the dependence on θ in the summation has been suppressed for brevity. Finally, note that we must have $\lambda_{\mathbb{Q}}^{j\to k} \geq 0$ in order to guarantee that a nonnegative payoff always produces an nonnegative contract value.

C Mortality Table: DAV 2000R

The mortality table in Pasdika and Wolff (2005) is specified in terms of q_x, which is the probability that a person aged x will die in the next year. We reproduce this data in Table C.1.

Age	q_x	Age	q_x	Age	q_x	Age	q_x	Age	q_x
65	0.008886	66	0.009938	67	0.011253	68	0.012687	69	0.014231
70	0.015887	71	0.017663	72	0.019598	73	0.021698	74	0.023990
75	0.026610	76	0.029533	77	0.032873	78	0.036696	79	0.041106
80	0.046239	81	0.052094	82	0.058742	83	0.066209	84	0.074583
85	0.083899	86	0.094103	87	0.105171	88	0.116929	89	0.129206
90	0.141850	91	0.154860	92	0.168157	93	0.181737	94	0.195567
95	0.209614	96	0.223854	97	0.238280	98	0.252858	99	0.267526
100	0.278816	101	0.293701	102	0.308850	103	0.324261	104	0.339936
105	0.355873	106	0.372069	107	0.388523	108	0.405229	109	0.422180
110	0.439368	111	0.456782	112	0.474411	113	0.492237	114	0.510241
115	0.528401	116	0.546689	117	0.565074	118	0.583517	119	0.601976
120	0.620400	121	1.0						

TABLE C.1: *DAV 2004R Mortality table, 65-year old German male (Pasdika and Wolff, 2005). q_x is the probability that a person aged x will die within the next year.*

References

Azimzadeh, P. (2013). Hedging costs for variable annuities. MMath thesis, Cheriton School of Computer Science, University of Waterloo.

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

Bauer, D., A. Kling, and J. Russ (2008). A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin* 38, 621–651.

Bélanger, A., P. Forsyth, and G. Labahn (2009). Valuing the guaranteed minimum death benefit clause with partial withdrawals. *Applied Mathematical Finance* 16, 451–496.

Bollen, N. P. B. (1998). Valuing options in regime-switching models. *Journal of Derivatives 6*(1), 38–49.

Chen, Z., K. Vetzal, and P. Forsyth (2008). The effect of modelling parameters on the value of GMWB guarantees. *Insurance: Mathematics and Economics* 43, 165–173.

Chopra, D., O. Erzan, G. de Gantès, L. Grepin, and C. Slawner (2009). Responding to the variable annuity crisis. McKinsey Working Papers on Risk Number 10, April.

Cramer, E., P. Matson, and L. Rubin (2007). Common practices relating to FASB statement 133, Accounting for Derivative Instruments and Hedging Activities as it Relates to Variable Annuities with Guaranteed Benefits. Practice Note, American Academy of Actuaries.

Dai, M., Y. K. Kwok, and J. Zong (2008). Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance* 18, 595–611.

Donelly, R., S. Jaimungal, and D. H. Rubisov (2014). Valuing guaranteed withdrawal benefits with stochastic interest rates and volatility. *Quantitative Finance* 14, 369–382.

Duan, J.-C., I. Popova, and P. Ritchken (2002). Option pricing under regime switching. *Quantitative Finance* 2, 116–132.

Elliott, R. J., L. Aggoun, and J. B. Moore (1995). *Hidden Markov Models: Estimation and Control.* Springer-Verlag, New York.

Elliott, R. J., L. Chan, and T. K. Siu (2005). Option pricing and Esscher transform under regime switching. *Annals of Finance* 1, 423–432.

Gaillardetz, P. and J. Lakhmiri (2011). A new premium principle for equity indexed annuities. *Journal of Risk and Insurance* 78, 245–265.

Guo, X. (2001). Information and option pricing. *Quantititive Finance* 1, 38–44.

Hardy, M. (2001). A regime switching model of long term stock returns. *North American Actuarial Journal* 5, 41–53.

Hilpert, C., J. Li, and A. Szimayer (2012). The effect of secondary markets on equity linked life insurance with surrender guarantees. Working paper, Bonn Graduate School of Economics.

Ho, T., S. B. Lee, and Y. Choi (2005). Practical considerations in managing variable annuities. Working paper, Thomas Ho Company.

Holz, D., A. Kling, and J. Russ (2012). GMWB for life: An analysis of lifelong withdrawal guarantees. *Zeitschrift für die gesamte Versicherungswissenschaft* 101, 305–325.

Huang, Y., P. Forsyth, and G. Labahn (2011). Methods for pricing American options under regime switching. *SIAM Journal on Scientific Computing* 33, 2144–2168.

Huang, Y., P. Forsyth, and G. Labahn (2012). Combined fixed point policy iteration for HJB equations in finance. *SIAM Journal on Numerical Analysis* 50, 1849–1860.

Huang, Y. T. and Y. K. Kwok (2013). Analysis of optimal dynamic withdrawal policies in withdrawal guarantees products. Working paper, Hong Kong University of Science and Technology.

Insured Retirement Institute (2013). Quarter-to-quarter variable annuity sales improve in Q1 2013. News release, June 5.

Jackson, K., S. Jaimungal, and V. Surkov (2008). Fourier space time-stepping for option pricing with Levy models. *Journal of Computational Finance* 12(2), 1–29.

Jin, Z., Y. Wang, and G. Yin (2011). Numerical solutions of quantile hedging for guaranteed minimum death benefits under a regime-switching jump-diffusion formulation. *Journal of Computational and Applied Mathematics* 235, 2842–2860.

Kennedy, J. (2007). *Hedging contingent claims in markets with jumps.* Ph. D. thesis, Department of Applied Mathematics, University of Waterloo.

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, 511–545.

Knoller, C., G. Kraut, and P. Schoenmaekers (2013). On the propensity to surrender a variable annuity contract: an empirical analysis of dynamic policyholder behavior. Working paper 7, Munich Risk and Insurance Center.

Lin, X., K. Tan, and H.Yang (2009). Pricing annuity guarantees under a regime switching model. *North American Actuarial Journal* 13(3), 316–338.

Milevsky, M. A. and T. S. Salisbury (2006). Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics* 38, 21–38.

Moenig, T. and D. Bauer (2011). Revisiting the risk-neutral approach to optimal policyholder behaviour: a study of withdrawal guarantees in variable annuities. Working paper, Georgia State University.

Naik, V. (1993). Option valuation and hedging strategies with jumps in the volatility of asset returns. *Journal of Finance* 48, 1969–1984.

Nelson, J. (2013). Setback for Sun Life. *The Globe and Mail*, June 22.

Ngai, A. and M. Sherris (2011). Longevity risk management for life and variable annuities: The effectiveness of static hedging using longevity bonds and derivatives. *Insurance: Mathematics and Economics* 49, 100–114.

O'Sullivan, C. and M. Moloney (2010). The variance gamma scaled self-decomposable process in actuarial modelling. Working paper, University College, Dublin, School of Business.

Pasdika, U. and J. Wolff (2005). Coping with longevity: The new German annuity valuation table DAV 2004R. The Living to 100 and Beyond Symposium, Orlando Florida.

Peng, J., K. S. Leung, and Y. K. Kwok (2012). Pricing guaranteed minimum withdrawal benefits under stochastic interest rates. *Quantitative Finance* 12, 933–941.

Piscopo, G. (2010). Withdrawal strategy for guaranteed lifelong withdrawal benefit option. *Perspectives of Innovations, Economics and Business* 5, 47–49.

Piscopo, G. and S. Haberman (2011). The valuation of guaranteed lifelong withdrawal benefit options in variable annuity contracts and the impact of mortality risk. *North American Actuarial Journal* 15(1), 59–76.

Rannacher, R. (1984). Finite element solution of diffusion problems with irregular data. *Numerische Mathematik* 43, 309–327.

Shah, P. and D. Bertsimas (2008). An analysis of the guaranteed withdrawal benefits for life option. Working paper, Sloan School of Management, MIT.

Shen, Y., K. Fan, and T. K. Siu (2013). Option valuation under a double regime switching model. *Journal of Futures Markets*, forthcoming.

Siu, T. K. (2005). Fair valuation of participating policies with surrender options and regime switching. *Insurance: Mathematics and Economics* 37, 533–552.

Sohrabi, M. (2010). Option pricing under regime switching: analyical, PDE and FFT methods. MMath thesis, School of Computer Science, University of Waterloo.

Stanton, R. (1995). Rational prepayment and the valuation of mortgage-backed securities. *Review of Financial Studies* 8, 677–807.

Tracer, Z. and S. Pak (2012). Variable-annuity sales fall most in 12 quarters. *Bloomberg News*, November 19.

Uzelac, F. and A. Szimayer (2014). Valuation of equity-linked life insurance contracts with surrender guarantees in a regime-switching rational expectation model. *Quantitative Finance* 14, 357–368.

Windcliff, H., P. Forsyth, and K. Vetzal (2001). Valuation of segregated funds: shout options with maturity extensions. *Insurance: Mathematics and Economics* 29, 1–21.

Yang, S. S. and T.-S. Dai (2013). A flexible tree for evaluating guaranteed minimum withdrawal benefits under deferred life annuity contracts with various provisions. *Insurance: Mathematics and Economics* 52, 232–242.

Yuen, F. and H. Yang (2010). Pricing Asian options and equity-indexed annuities with regime switching by the trinomial tree method. *North American Actuarial Journal* 14(2), 256–277.

Yuen, F. L. and H. Yang (2009). Option pricing in a jump-diffusion model with regime switching. *ASTIN Bulletin* 39, 515–539.