Revisiting the Risk-Neutral Approach to Optimal Policyholder Behavior: A Study of Withdrawal Guarantees in Variable Annuities*

Working Paper

Thorsten Moenig
Department of Risk Management and Insurance, Georgia State University
35 Broad Street, 11th Floor; Atlanta, GA 30303; USA
Email: thorsten@gsu.edu

Daniel Bauer
Department of Risk Management and Insurance, Georgia State University
35 Broad Street, 11th Floor; Atlanta, GA 30303; USA
Email: dbauer@gsu.edu

October 2011

Abstract

Policyholder exercise behavior presents an important risk factor for life insurance companies. Yet, most approaches presented in the academic literature – building on value maximizing strategies akin to the valuation of American options – do not square well with observed prices and exercise patterns.

Following a recent strand of literature, in order to gain insights on what drives policyholder behavior, this paper develops a life-cycle model for variable annuities (VA) with withdrawal guarantees. However, in contrast to these earlier contributions, we explicitly allow for outside savings and investments, which considerably affects the results. Specifically, we find that withdrawal patterns after all are primarily motivated by value maximization – but with the important asterisk that the value maximization should be taken out from the policyholders' perspective accounting for individual tax benefits.

To this effect, we develop and apply a risk-neutral valuation methodology that takes these different tax structures into consideration. The results are in line with corresponding findings from the life cycle model as well as prevalent market rates for the considered withdrawal guarantee. Also, our findings endorse the application of simple reduced-form exercise rules based on the "moneyness" of the guarantee that are slowly being adopted in the insurance industry.

Keywords: Variable Annuities, Guaranteed Minimum Benefits, Optimal Policyholder Behavior, Lifecycle Theory, Risk-Neutral Valuation with Taxation.

*An earlier version of the paper was entitled "Policyholder Exercise Behavior for Variable Annuities including Guaranteed Minimum Withdrawal Benefits". The authors are thankful for helpful comments from participants at the 2011 AFIR & ASTIN Colloquia, the 2011 ARIA Annual Meeting, the 46th Actuarial Research Conference as well as from seminar participants at Georgia State University. In particular, we are indebted to Glenn Harrison, Ajay Subramanian and Eric Ulm for valuable input. Financial support from the Society of Actuaries (CAE Grant) is also gratefully acknowledged. All remaining errors are ours.

1 Introduction

Policyholder behavior is an important risk factor for life insurance companies offering contracts that include exercise-dependent features, but so far it is little understood. Specifically, analyses of optimal policyholder behavior uncovered in the actuarial literature – building on the theory for evaluating American and Bermudan options – commonly yield exercise patterns and prices that are far from observations in practice.[1] A recent strand of literature believes to have identified the problem in the incompleteness of the insurance market.[2] More precisely, the argument is that in contrast to financial derivatives, policyholders may not have the possibility to sell (or repurchase) their contract at its risk-neutral continuation value so that exercising may be advisable – and rational – even if risk-neutral valuation theory does not suggest so. As a solution, these papers suggest to analyze exercise behavior in life-cycle utility optimization models where the decision to exercise is embedded in the overall portfolio problem of an individual or a household, although the associated complexity naturally necessitates profound simplifications.

In this paper, we follow this strand of literature in that we also develop a life-cycle utility model for a poster child of exercise-dependent options in life insurance, namely a variable annuity (VA) contract including a Guaranteed Minimum Withdrawal Benefit (GMWB) rider. However, compared to earlier work, we explicitly account for outside savings and allocation options. While of course this addition increases the complexity of the optimization problem, it affects the results considerably. We find that most risk allocations occur outside of the VA and that changes in the policyholder's wealth level, preferences, or other behavioral aspects have little effect on the optimal withdrawal behavior. In contrast, the exercise behavior appears to be primarily motivated by value maximization, however with the important wrinkle that taxation rules considerably affect this value.

To further analyze this assertion and as an important methodological contribution of the paper, we develop a valuation mechanism in the presence of different investment opportunities with differing tax treatments. The key idea is that if the pre-tax investment market for underlying investments such as stocks and bonds is complete, it is possible to replicate any given post-tax cash flow with a pre-tax cash flow of these underlying investments – irrespective of the tax treatment for the securities leading to the former cash flow. We show that when taking taxation into account via the proposed mechanism, a value-maximizing approach yields withdrawal patterns and pricing results that are close to the results from the life cycle model. Hence, our results can be interpreted as a vindication of the risk-neutral valuation approach – associated with all its benefits such as independence of preferences, wealth, or consumption decisions – although it is to be taken out from the perspective of the policyholder rather than the insurance company so that personal tax considerations matter.

As already indicated, we focus our attention on policyholder exercise behavior for VA contracts with GMWBs. Here, a VA essentially is a unit-linked, tax-deferred savings plan potentially entailing guaranteed payment levels, for instance upon death (Guaranteed Minimum Death Benefit, GMDB) or survival until expiration (Guaranteed Minimum Living Benefits, GMLB). A GMWB, on the other hand, provides the

[1] See, among others, Bauer et al. (2008), Grosen and Jørgensen (2000), Milevsky and Posner (2001), Milevsky and Salisbury (2006), Ulm (2006), or Zaglauer and Bauer (2008).

[2] See e.g. Gao and Ulm (2011), Knoller et al. (2011), and Steinorth and Mitchell (2011).

policyholder with the right but not the obligation to withdraw the initial investment over a certain period of time, irrespective of investment performance, as long as annual withdrawals do not exceed a pre-specified amount. To finance these guarantees, most commonly insurers deduct an option fee at a constant rate from the policyholder's account value.

In 2010, U.S. individual VA sales totaled over $140 billion, increasing the combined net assets of VAs to a record $1.5 trillion, whereby most of them are enhanced by one or even multiple guaranteed benefits. These figures indicate the importance for insurers to understand how policyholders may utilize these embedded options, especially because changes in economic or regulatory conditions have on occasion caused dramatic shifts in policyholder behavior that have caught the industry off-guard.[3] However, to date most liability models fail to capture this risk factor in an adequate fashion. In particular, companies usually rely on historic exercise probabilities or static exercise rules, although some insurers indicate they use simple dynamic assumptions in their C3 Phase II calculations (cf. Society of Actuaries (2009)).

The prevalent assumption for evaluating GMWBs in the actuarial literature is that policyholders may exercise optimally with respect to the value of the contract consistent with arbitrage pricing theory (see, among others, Milevsky and Salisbury (2006), Bauer et al. (2008), Chen and Forsyth (2008), or Dai et al. (2008)).[4] Specifically, the value is characterized by an optimal control problem identifying the supremum of the risk-neutral contract value over all admissible withdrawal strategies. While such an approach may be justified in that it – in principle – identifies the unique supervaluation and superhedging strategy robust to any policyholder behavior (cf. Bauer et al. (2010)), the resulting "fair" guarantee fees considerably exceed the levels encountered in practice. For example, Milevsky and Salisbury (2006) calculate the no-arbitrage hedging cost of a GMWB to range from 73 to 160 basis points, depending on parameter assumptions, although typically insurers charge only about 30 to 45 bps. While from the authors' perspective these observed differences between theory and practice are a result of "suboptimal" policyholder behavior, these deviations can also be attributed to the policyholder foregoing certain privileges and protection when making a withdrawal, even in the case of rational decision making. Most notably, tax benefits of VAs are a major reason for their popularity, so that it is proximate to assume that taxation also factors into the policyholder's decision-making process. Furthermore, in contrast to financial derivatives, policyholders generally are not able to sell their policy at its risk-neutral value, which may also affect withdrawal behavior.

To analyze whether or not there are rational reasons for the observed behavior – akin to related recent literature (cf. Gao and Ulm (2011) or Steinorth and Mitchell (2011)) – we introduce a structural model that explicitly considers the problem of decision making under uncertainty faced by the holder of a VA policy. More specifically, the policyholder's state-contingent decision process is modeled using a lifetime utility model of consumption and bequests, where we allow for stochasticity in both the financial market

[3] For instance, rising interest rates in the 1970s led to the so-called *disintermediation* process, which caused substantial increases in surrenders and policy loans in the whole life market (cf. Black and Skipper (2000), p. 111). Similarly, in 2000, the UK-based mutual life insurer Equitable Life – the world's oldest life insurance company – was closed to new business due to problems arising from a misjudgment of policyholder behavior with respect to exercising guaranteed annuity options within individual pension policies (cf. Boyle and Hardy (2003)). More recently, the U.S. insurer The Hartford had to accept TARP money, after losing "$2.75 billion in 2008, hurt by investment losses and the cost of guarantees it provided to holders of variable annuities."

[4] A few alternatives have also been put forward. For instance, Stanton (1995) proposes a rational expectations model with heterogeneous transaction costs in the case of prepayment options within mortgages, and De Giovanni (2010) develops a model for surrender options in life insurance contracts, which also allows for irrational in addition to rational exercises.

and individual lifetime. However, in contrast to previous contributions, we explicitly allow for an outside investment option and we include appropriate investment tax treatments. We parametrize the model based on reasonable assumptions about policyholder characteristics, the financial market, etc., and solve the decision making problem numerically using a recursive dynamic programming approach.

Based on the model, we are able to identify a variety of aspects that factor into the policyholder's decision process. First and foremost, withdrawals are infrequent and are optimal mainly upon poor market performance or – to be more precise – when the VA account has fallen below the tax base. For instance, in our benchmark case the policyholder will make one or more withdrawals prior to maturity less than one fourth of the time, and the probability that he will withdraw the full initial investment is less than 5%. These findings are in stark contrast to the results based on arbitrage pricing theory, which find that withdrawing at least the guaranteed amount is optimal in most circumstances (cf. Milevsky and Salisbury (2006)). In particular, our results indicate that the assumed guarantee fee of 50 basis points appears to sufficiently provide for the considered return-of-investment GMWB. Moreover, our results prove fairly insensitive to changes in individual and behavioral parameters such as wealth, income and the level of risk aversion. The differences are small but systematic in a way that is consistent with the market incompleteness resulting from an absence of life-contingent securities – other than the VA – within our model. Therefore, our results suggest that policyholder behavior is primarily driven by value maximization when taking the preferred tax treatment of VAs into account. In particular, taxation not only seems to be a major reason why people purchase VAs, but appears to also incentivize them not to withdraw prematurely.

To further elaborate on this observation, we devise a risk-neutral valuation mechanism in the presence of different investment opportunities with differing tax treatments. Relying on this mechanism, we implement an alternative approach to uncover the optimal withdrawal behavior with regards to maximizing the value of all payoffs akin to standard arbitrage pricing methods. As predicted, the numerical results of the value-maximizing strategy turn out to be similar to those of the – considerably more complex – utility-based model.[5]

On a practical note, our results endorse the use of simple dynamic exercise rules based on the "moneyness" of the guarantee, which are slowly adopted by some life insurance companies (cf. Society of Actuaries (2009)). While this result is in line with the empirical findings from Knoller et al. (2011), we note that the coherence of this rule in our setting is not primarily due to the "moneyness" factoring into the policyholder's decision process, but it is a consequence of the similarities between tax and benefits base.

The remainder of the paper is structured as follows: In the subsequent section, we introduce a life-time utility model for VAs. Section 3 is dedicated to the implementation of the model in a Black-Scholes framework and to discussing computational details. Section 4 details the numerical results of the life-cycle model. In Section 5, we develop a risk-neutral valuation approach with taxation and apply it to our valuation problem. This is followed by a discussion of implications for insurance practice in Section 6. And finally, Section 7 concludes and briefly discusses possible extensions.

[5]This result shows some resemblance to the findings of Carpenter (1998) in the context of employee stock options. In particular, her investigations suggest that a value-maximizing strategy (plus a fixed-probability exogenous exercise state) explains exercise behavior just as well as a complex utility-based model.

2 A Lifetime Utility Model for Variable Annuities with GMWBs

There exists a large variety of VA products available in the U.S. The policies differ by how the premiums are collected; policyholder investment opportunities, including whether the policyholder can reallocate funds after underwriting; and guarantee specifics, for instance what type of guarantees are included, how the guarantees are designed and how they are paid for, etc. For a detailed description of VAs and the guarantees available in the market, we refer to Bauer et al. (2008).

This section develops a lifetime utility model of VAs including (at least) a simple return-of-investment GMWB option, with stochasticity in policyholder lifetime and asset returns. The policyholder's state-contingent decision process entails annual choices over withdrawals from the VA account, consumption, and asset allocation in an outside portfolio.

In contrast to mutual funds, VAs grow tax deferred, which presents the primary reason for their popularity among individuals who exceed the limits of their qualified retirement plans. For instance, Milevsky and Panyagometh (2001) argue that variable annuities outperform mutual funds for investments longer than ten years, even when the option to harvest losses is taken into account for the mutual fund. Since the preferred tax treatment may also affect policyholder exercise behavior, we briefly describe current U.S. taxation policies on variable annuities and the way these are captured in our model in Section 2.4.

2.1 Description of the Variable Annuity Policy

We consider an x-year old individual who has just (time $t = 0$) purchased a VA with finite integer maturity T against a single up-front premium P_0. We assume that all cash flows as well as all relevant decisions come into effect at policy anniversary dates, $t = 1, \ldots, T$. In particular, the insurer will return the policyholder's concurrent account value – or some guaranteed amount, if eligible – at the end of the policyholder's year of death or at maturity, whichever comes first. In addition, the contract contains a GMWB option, which grants the policyholder the right but not the obligation to withdraw the initial investment P_0 free of charge and independent of investment performance, as long as annual withdrawals do not exceed the guaranteed annual amount g_t^W. Withdrawals in excess of either g_t^W or the remaining aggregate withdrawal guarantee – denoted by G_t^W – carry a (partial) surrender charge of $s_t \geq 0$ as a percentage of the excess withdrawal amount. We model a "generic" contract that may also contain a GMDB or other GMLB options. In that case, we denote by G_t^D, G_t^I, and G_t^A the guaranteed minimum death, income, and accumulation benefit, respectively.[6] For simplicity of exposition and without much loss of generality, we assume that all included guarantees are return-of-investment options. Thus all involved guarantee accounts have an identical *benefits base*

$$G_t \equiv G_t^W = G_t^D = G_t^A = G_t^I.$$

If an option is not included, we simply set the corresponding guaranteed benefit to zero. Hence this model allows us to include a variety of guarantees, without having to increase the state space, which makes the

[6] A Guaranteed Minimum Accumulation Benefit (GMAB) guarantees a minimal (lump-sum) payout at maturity of the contract, provided that the policyholder is still alive. Under the same conditions, a Guaranteed Minimum Income Benefit (GMIB) guarantees a minimal annuity payout.

problem computationally feasible.[7] However, other contract designs could be easily incorporated at the cost of a larger state space. We refer to Bauer et al. (2008) for details.

While for return-of-investment guarantees, the initial benefits base is $G^{\cdot}_0 = P_0$, this equality will no longer be satisfied after funds have been withdrawn from the account. More precisely, following Bauer et al. (2008), we model the adjustments of the benefits base in case of a withdrawal prior to maturity based on the following assumptions: If the withdrawal does not exceed the guaranteed annual amount g_t^W, the benefits base will simply be reduced by the withdrawal amount. Otherwise, the benefits base will be the lesser of that amount and a so-called pro rata adjustment. Hence,

$$G^{\cdot}_{t+1} = \begin{cases} (G^{\cdot}_t - w)^+ & : w \leq g_t^W \\ \left(\min\left\{ G^{\cdot}_t - w \, , \, G^{\cdot}_t \cdot \frac{X_t^+}{X_t^-} \right\} \right)^+ & : w > g_t^W, \end{cases} \tag{1}$$

where w is the withdrawal amount, $X_t^{-/+}$ denote the VA account values immediately before and after the withdrawal is made, respectively, and $(a)^+ \equiv \max\{a, 0\}$. To finance the guarantees, the insurer continuously deducts an option fee at constant rate $\phi \geq 0$ from the policyholder's account value.

With regards to the investment strategy for the VA, we assume the policyholder chooses an allocation at inception of the contract, and that it remains fixed subsequently. This is not unusual in the presence of a GMWB option since otherwise the policyholder may have an incentive to shift to the most risky investment strategy in order to maximize the value of the guarantee.

2.2 Policyholder Preferences

The policyholder gains utility from consumption, while alive, and from bequesting his savings upon his death (if death occurs prior to retirement). We assume time-separable preferences with an individual discount factor β, and utility functions $u_C(\cdot)$ and $u_B(\cdot)$ for consumption and bequests, respectively.

The policyholder is endowed with an initial wealth W_0, of which he invests P_0 in the VA. The remainder is placed in an "outside account". We suppose there exist d identical investment opportunities inside and outside the VA, the main difference being that adjustments to the investment allocations in the outside portfolio can be made every year at the policy anniversary. We denote the time-t value of the VA account by X_t and the time-t value of the outside account by A_t, where the corresponding investment allocations are specified by the d-dimensional vectors ν^X and ν_t for the VA and outside account, respectively. In either case, short sales are not allowed, so that we require

$$\nu_t, \nu^X \geq 0, \text{ and } \sum_i \nu_t(i) = \sum_i \nu^X(i) = 1. \tag{2}$$

For the values at policy anniversaries $t \in \{0, 1, \ldots, T\}$, we add superscript $^-$ to denote the level of an account (state variable) at the beginning of a period, just prior to the policyholder's decision, and superscript $^+$ to indicate its value immediately afterwards. Note that guarantee accounts do not change *between* periods, i.e. between $(t)^+$ and $(t+1)^-$, $t = 0, 1, \ldots, T-1$, but only through withdrawals at policy anniversary dates,

[7] In particular, we can also analyze contracts that do not contain a GMWB option at all.

so that no superscripts are necessary here. The policyholder receives annual (exogenous) net income I_t, which for the purpose of this paper is deterministic and paid in a single installment at each policy anniversary. Upon observing his current level of wealth, i.e. his outside account value A_t^-, the state of his VA account X_t^-, the level of his guarantees $G_t^\cdot$, and his VA tax basis H_t (see below), the policyholder chooses how much to withdraw from the VA account, how much to consume, and how to allocate his outside investments in the upcoming policy year.

Appendix A.1 describes the assumed timeline of events leading up to and following the policyholder's decision each period.

2.3 Mortality

Relying on standard actuarial notation, we denote by ${}_tq_x$ the probability that (x) dies within t years, and by ${}_tp_x \equiv 1 - {}_tq_x$ the corresponding probability of survival. In particular, we express the one-year death and survival probabilities by q_x and p_x, respectively. Consequently, the probability that (x) dies in the interval $(t, t+1]$ is given by ${}_tp_x \cdot q_{x+t}$.

Upon the policyholder's death, we assume that his bequest amount is converted to a risk-free perpetuity (reflecting that upon the beneficiary's death, remaining funds will be passed on to his own beneficiaries, and so on) at the risk-free rate, which for simplicity is assumed to be constant and denoted by r. Note that all previous earnings on the VA will be taxed as ordinary income at that point. We assume the beneficiaries have the same preferences as the policyholder, and that his bequest motive is B. That is, if he leaves bequest amount x (net of taxes), the bequest utility is given by:

$$\frac{1}{1-\beta} \cdot B \cdot u_C([1-e^{-r}] \cdot x).$$

2.4 Tax Treatment of Variable Annuities

We model taxation of income and investment returns based on concurrent U.S. regulation, albeit with a few necessary simplifications. More precisely, we assume that all investments into the VA are post-tax and non-qualified. As such, taxes will only be due on future investment gains, not the initial investment (principal) itself.

Investments inside a VA grow tax deferred. In other words, the policyholder will not be taxed on any earnings until he starts to make withdrawals from his account. However, all earnings from a VA will eventually be taxed as ordinary *income*. More precisely, withdrawals are taxed on a last-in first-out basis, meaning that earnings are withdrawn *before* the principal. Specifically, early withdrawals after an investment gain are subject to income taxes. Only if the account value lies below the tax base will withdrawals be tax free. In addition, withdrawals prior to the age of $59\frac{1}{2}$ are subject to an early withdrawal tax of s^g (typically 10%). At maturity, denoting the concurrent VA account value by X_T and the tax base by H_T, if the policyholder chooses the account value to be paid out as a lump-sum, he is required to pay taxes on the (remaining) VA earnings

$$\max\{X_T - H_T, 0\}$$

immediately (if applicable, we substitute G_T^A for X_T). However, if the policyholder chooses to annuitize his account value – e.g. in level annual installments, as we assume in this paper – his annual tax-free amount is his current tax base H_T divided by his life expectancy e_{x+T}, as computed from the appropriate actuarial table. In other words, the policyholder will need to declare any annuity payments from this VA in excess of H_T/e_{x+T} per year as ordinary income (see IRS (2003)).

For the initial tax base, we obviously have $H_1 = P_0$. The subsequent evolution of the tax base depends on both the evolution of the account value and withdrawals, where H_t essentially denotes the part of the account value that is left from the original principal. More precisely, the tax base remains unaffected by withdrawals smaller or equal to $X_t^- - H_t$, i.e. those withdrawals that are fully taxed (because they come from earnings), whereas tax-free withdrawals reduce the tax base dollar for dollar. Hence, formally we have

$$H_{t+1} = H_t - \left(w_t - \left(X_t^- - H_t\right)^+\right)^+. \tag{3}$$

In contrast, returns from a mutual fund are not tax deferrable. While in practice parts of these returns are ordinary dividends and thus taxed as income, others are long term capital gains and subject to the (lower) capital gains tax rate. We simplify taxation of mutual fund earnings to be at a constant annual rate, denoted by κ, which for future reference we call the *capital gains tax*, although it may be chosen a little higher than the actual tax on capital gains to reflect income from dividends or coupon payments, which are taxable at a higher rate. The income tax rate is also assumed to be constant over taxable money and time, at rate τ.[8]

2.5 Policyholder Optimization During the Lifetime of the Contract

The setup, as described in this section, requires four state variables: A_t^-, the value of the outside account just before the t-th policy anniversary date; X_t^-, the value of the VA account just before the t-th policy anniversary date; $G_t^{\cdot}$, the value of the benefits base (and thus all guarantee accounts) in period t; and H_t, the tax base in period t. At the t-th policy anniversary, given withdrawal of w_t, we define next-period benefits base and tax base by equations (1) and (3), respectively.

2.5.1 Transition from $(t)^-$ to $(t)^+$

Upon withdrawal of w_t, consumption C_t, and new outside portfolio allocation level ν_t, we update our state variables as follows:

$$\begin{aligned} X_t^+ &= \left(X_t^- - w_t\right)^+, \text{and} \\ A_t^+ &= A_t^- + I_t + w_t - C_t - \text{fee}_I - \text{fee}_G - \text{taxes}, \end{aligned} \tag{4}$$

where

$$\text{fee}_I = s \cdot \max\left\{w_t - \min(g_t^W, G_t^W), 0\right\}$$

[8] We believe this to be a reasonable simplification as holders of variable annuities are typically relatively wealthy, so that brackets over which the applicable marginal income tax rate is constant are fairly large. Moreover, we want to avoid withdrawal behavior being affected unpredictably by "fragile" tax advantages.

denotes the excess withdrawal fees the policyholder pays to the insurer,

$$\text{fee}_G = s^g \cdot (w_t - \text{fee}_I) \cdot \mathbf{1}_{\{x+t<59.5\}}$$

are the early withdrawal penalty fees the government collects on withdrawals prior to age 59.5, and

$$\text{taxes} = \tau \cdot \min\{w_t - \text{fee}_I - \text{fee}_G, (X_t^- - H_t)^+\}$$

are the (income) taxes the policyholder pays upon withdrawing w_t.

In our basic model, we update the guaranteed withdrawal account by (1). If the contract specifies guarantees to evolve differently (e.g. step-up or ratchet-type guarantees), the updating function must be modified accordingly. In that case we may also need to carry along an additional (binary) state variable to keep track of whether the policyholder has previously made a withdrawal. We refer to Bauer et al. (2008) for details.

2.5.2 Transition from $(t)^+$ to $(t+1)^-$

In our model, the only state variables changing stochastically between $(t)^+$ and $(t+1)^-$ are the account values inside and outside of the VA, both driven by the evolution of the financial assets, which are described by the vector-valued stochastic process, $(S_t)_{t\geq 0}$.[9] Similarly, the (row) vector ν_t captures the fraction of outside wealth A_t^+ the policyholder wants to invest in each asset. Taking into account the tax treatments as described in section 2.4 we can update the account values as follows:

$$\begin{aligned} A_{t+1}^- &= A_t^+ \cdot \left[\nu_t \cdot \tfrac{S_{t+1}}{S_t} - \kappa \cdot \left(\nu_t \cdot \tfrac{S_{t+1}}{S_t} - 1\right)^+\right], \text{ and} \\ X_{t+1}^- &= X_t^+ \cdot e^{-\phi} \cdot \left[\nu^X \cdot \tfrac{S_{t+1}}{S_t}\right], \end{aligned} \tag{5}$$

where $\frac{S_{t+1}}{S_t}$ denotes the component-wise quotient.

2.5.3 Bellman Equation

Denoting the policyholder's time-t value function by $V_t^- : \mathbb{R}^4 \to \mathbb{R}$, $y_t \equiv (A_t^-, X_t^-, G_t^-, H_t) \mapsto V_t^-(y_t)$, where we call y_t the vector of state variables, we can describe his optimization problem at each policy anniversary date recursively by

$$V_t^-(y_t) = \max_{C_t, w_t, \nu_t} u_C(C_t) + e^{-\beta} \cdot \mathbb{E}_t \left[q_{x+t} \cdot u_B(b_{t+1}|S_{t+1}) + p_{x+t} \cdot V_{t+1}^-(y_{t+1}|S_{t+1})\right], \tag{6}$$

subject to (1), (2), (3), (4), (5), the bequest amount

$$b_{t+1} = A_{t+1}^- + b_X - \tau \cdot (b_X - H_t, 0),$$

[9] As usual in this context, underlying our consideration is a complete filtered probability space $(\Omega, F, \mathbb{P}, \mathbf{F} = (F_t)_{t\geq 0})$, where $\mathbf{F}$ satisfies the usual conditions and $\mathbb{P}$ denotes the "physical" probability measure.

where $b_X = \max\left\{X_{t+1}^-, G_{t+1}^D\right\}$, and the choice variable constraints

$$\begin{aligned} 0 &\le C_t \le \quad A_t^- + I_t + w_t - \text{fee}_I - \text{fee}_G - \text{taxes}, \quad \text{and} \\ 0 &\le w_t \le \quad \max\left\{ X_t^- \ , \ \min\left\{ g_t^W \ , \ G_t^{\cdot} \right\} \right\}. \end{aligned}$$

2.6 Policyholder Behavior upon Maturity of the Variable Annuity

If the policyholder is alive when the Variable Annuity matures at time T, we assume that he retires immediately and no longer receives any outside income. He will live off his concurrent savings, which consist of the time-T value of his outside portfolio, plus the maximum of his VA account value and any remaining GMLB benefits. More precisely, we assume he uses these savings to purchase a single-premium whole life annuity, and that he no longer has a bequest motive; his consumption preferences, on the other hand, are the same as before.

We model the taxation of annuities following our discussion in Section 2.4. The outside account value A_T^- is already net of taxes, thus only future earnings (i.e. interest) need to be taxed. Therefore, A_T^- acts as the tax base for the whole life annuity. The outside account can thus be converted into net annuity payments of

$$c_A \equiv \frac{A_T^-}{e_{x+T}} + (1-\tau)\cdot\left(\frac{A_T^-}{\ddot{a}_{x+T}} - \frac{A_T^-}{e_{x+T}}\right) = \tau\cdot\frac{A_T^-}{e_{x+T}} + (1-\tau)\cdot\frac{A_T^-}{\ddot{a}_{x+T}} \tag{7}$$

at the beginning of every year as long as the policyholder is alive. Here, $\ddot{a}_{x+T}$ denotes the actuarial present value of an annuity due paying 1 at the beginning of each year while $(x+T)$ is alive, and e_{x+T} denotes the policyholder's complete life expectancy at maturity of the VA as used to determine tax treatment upon annuitization of the VA payout.

At maturity, the policyholder can withdraw the remainder of the account value (or some guaranteed level, if applicable) from the VA. That is:

$$w_T = \max\left\{X_T^-, \max\left[G_T^A, \min\left(G_T^W, g_T^W\right)\right]\right\}. \tag{8}$$

This results in life-long annual payments of

$$\begin{aligned} c_X \quad &\equiv \quad \min\left\{\tfrac{w_T}{\ddot{a}_{x+T}}, \tau\cdot\tfrac{H_T}{e_{x+T}} + (1-\tau)\cdot\tfrac{w_T}{\ddot{a}_{x+T}}\right\} \\ &= \quad \tfrac{w_T}{\ddot{a}_{x+T}} - \tau\cdot\max\left\{\tfrac{w_T}{\ddot{a}_{x+T}} - \tfrac{H_T}{e_{x+T}}, 0\right\}. \end{aligned} \tag{9}$$

If a GMIB is included in the contract, the policyholder can also choose to annuitize the guaranteed amount G_T^I at a guaranteed annuity factor $\ddot{a}_{x+T}^{guar}$, and thus receive annual payouts

$$c_I \equiv \frac{G_T^I}{\ddot{a}_{x+T}^{guar}} - \tau\cdot\max\left\{\frac{G_T^I}{\ddot{a}_{x+T}^{guar}} - \frac{H_T}{e_{x+T}} \ , \ 0\right\} \tag{10}$$

Overall, the policyholder can therefore consume $c_A + \max\{c_X, c_I\}$ every year during his retirement. The time-T expected lifetime utility for the policyholder is thus

$$V_T^-(A_T^-, X_T^-, G_T^{\cdot}, H_T) = \sum_{t=0}^{\omega-x-T} \exp(-\beta t) \cdot {}_t p_x \cdot u_C(c_A + \max\{c_X, c_I\}), \tag{11}$$

subject to equations (7) to (10).

3 Implementation in a Black-Scholes Framework

For our implementation, we consider two investment possibilities only, namely a risky asset $(S_t)_{t\geq 0}$ and a risk-free asset $(B_t)_{t\geq 0}$. More specifically, akin to the well-known Black-Scholes-Merton model, we assume that the risky asset evolves according to the Stochastic Differential Equation (SDE)

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dZ_t\,, \;\; S_0 > 0,$$

where $\mu, \sigma > 0$, and $(Z_t)_{t>0}$ is a standard Brownian motion, while the risk-free asset (savings account) follows

$$\frac{dB_t}{B_t} = r\; dt\,, \;\; B_0 = 1 \quad \Rightarrow \quad B_t = \exp(r\, t).$$

In this setting, optimization problem (6) takes the form

$$V_t^-(y_t) = \max_{C_t, w_t, \nu_t} u_C(C_t) + e^{-\beta} \int_{-\infty}^{\infty} \psi(\gamma) \left[q_{x+t} \cdot u_B(b_{t+1}|S'(\gamma)) + p_{x+t} \cdot V_{t+1}^-(y_{t+1}|S'(\gamma)) \right] d\gamma, \tag{12}$$

where $\psi(\gamma) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\gamma^2}{2})$ is the standard normal probability density function, and $S'(\gamma) = S_t \cdot e^{\sigma\gamma + \mu - \frac{1}{2}\sigma^2}$ is the annual gross return of the risky asset, subject to various constraints (see Appendix A.2 for a detailed list).

In the remainder of this section, we present a recursive dynamic programming approach for the solution. In particular, we address practical implementation problems arising from the complexity associated with the high dimensionality of the state space.

3.1 Estimation Algorithm

The key idea underlying our algorithm is a discretization of the state space at policy anniversaries. More specifically, our approach to derive the optimal consumption, allocation, and – particularly – withdrawal policies consists of the following steps.

Algorithm 3.1.

(I) Discretize the four-dimensional state space consisting of the values for $A, X, G^{\cdot}$, and H appropriately to create a grid.

(II) For $t = T$: for all grid points $(A, X, G^{\cdot}, H)$, compute $V_T^-(A, X, G^{\cdot}, H)$ via Equation (11).

(III) For $t = T-1, T-2, \ldots, 1$*:*

(1) Given V_{t+1}^{-}*, calculate* $V_t^{-}(A, X, G^{\cdot}, H)$ *recursively for each* $(A, X, G^{\cdot}, H)$ *on the grid via the (approximated) solution to Equation* (12)*.*

(2) Store the optimal state-contingent withdrawal, consumption, and allocation choices for further analyses.

(IV) For $t = 0$*: For the given starting values* $A_0 = W_0 - P_0$*,* $X_0 = P_0$*,* $G_0^{\cdot} = G_1^{\cdot} = P_0$ *and* $H_0 = H_1 = P_0$*, compute* $V_0^{-}(W_0 - P_0, P_0, P_0, P_0)$ *recursively from equation* (12)*.*

Storing the optimal choices in step $(III.2)$ not only allows us to analyze to what extent a representative policyholder makes use of the withdrawal guarantee, which is the primary focus of our paper, but we may also determine the time zero value of all collected fees and payouts to the policyholder or his beneficiaries via their expected present values under the risk-neutral measure $\mathbb{Q}$.[10] In particular, by comparing these values we can make an inference whether or not the contracted fee percentage within our representable contract sufficiently provides for the offered guarantees in the absence of other costs. However, before discussing our results in Section 4, the remainder of this section provides the necessary details about the implementation of the steps in Algorithm 3.1 as well as the choice of the underlying parameters.

3.2 Evaluation of the Integral Equation (12)

Since within step (III) of Algorithm 3.1 the (nominal) value function at time $t+1$ is only given on a discrete grid, it is clearly not possible to directly evaluate the integral in Equation (12). We consider two different approaches for its approximation by discretizing the underlying return space.

Since the integral entails the standard normal density function, one prevalent approach is to rely on a Gauss-Hermite Quadrature. However, to ascertain the accuracy of our approximation, we additionally consider a second approach. Note that our integral equation is of the form

$$K \equiv \int_{-\infty}^{\infty} \phi(u) F(\lambda(u))\, du, \tag{13}$$

where $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp(\frac{1}{2}u^2)$ is the standard normal density function, $\lambda(u) = \exp(\sigma u + \mu - \frac{1}{2}\sigma^2)$ corresponds to the annual stock return S_{t+1}/S_t, and

$$F(x) \equiv q_{x+t} \cdot u_B \left(b_{t+1} \middle| \frac{S_{t+1}}{S_t} = x \right) + p_{x+t} \cdot V_{t+1}^{-} \left(y_{t+1} \middle| \frac{S_{t+1}}{S_t} = x \right).$$

Dividing the return space $(-\infty, \infty)$ into $M > 0$ subintervals $[u_k, u_{k+1})$, for $k = 0, 1, \ldots, M-1$, where we set $-\infty = u_0 < u_1 < \ldots < u_{M-1} < u_M = \infty$, a consistent approximation of the integral (13) is given

[10] By the *fundamental theorem of asset pricing* the existence of the risk-neutral measure is essentially equivalent with the absence of arbitrage in the market. As is common in this context, here we choose the product measure of the (unique) risk-neutral measure for the (complete) financial market and the physical measure for life-contingent events.

by

$$K \approx \sum_{k=0}^{M-1} \Phi(u_{k+1}) \cdot [a_k - a_{k+1}] + \exp(\mu) \cdot \Phi(u_{k+1} - \sigma) \cdot [b_k - b_{k+1}]. \tag{14}$$

Here, $\Phi(.)$ is the standard normal cdf, $a_k \equiv \frac{x_{k+1} \cdot \psi_k - x_k \cdot \psi_{k+1}}{x_{k+1} - x_k}$, $b_k \equiv \frac{\psi_{k+1} - \psi_k}{x_{k+1} - x_k}$ for $k = 0, \ldots, M-1$, $a_M = b_M \equiv 0$, $x_k = \lambda(u_k)$ represent the gross returns, and $\psi_k \equiv F(x_k)$ are the function values evaluated at returns x_k (see Appendix B.1 for a derivation of (14)). With this approach, we have the discretion to choose the number $(M-1)$ and location (x_k) of all nodes, providing more flexibility than the Gauss-Hermite Quadrature method. It is important to note, however, that the values ψ_k cannot be calculated directly, but need to be derived from the value function grid at time $t+1$, where we rely on multilinear interpolation when necessary. We find very similar results for both approaches and therefore only present estimation results based on the approximation via Equation (13).

3.3 Monte-Carlo Simulations to Quantify Optimal Behavior

Using Algorithm 3.1, we can determine the policyholder's optimal decision variables for all time/state combinations. To aggregate and better compare results, and to analyze pricing implications, we perform Monte Carlo simulations. More precisely, we simulate 5 million paths over stock movements and individual mortality. Thus, based on optimal choices of withdrawal, consumption and investment, we can compute the evolution of state variables as well as a variety of withdrawal measures for each path. Tables 3 and 4 in the Results section 4 show the corresponding statistics for different parameter assumptions. Note that the first section of each table is based on paths generated under the risk-neutral measure $\mathbb{Q}$ (see Footnote 10). While we later argue that $\mathbb{Q}$ is not appropriate to value contingent claims from the *policyholder's* perspective due to tax considerations (see Section 5), an *insurer* replicating its liabilities does not pay taxes on the respective earnings, so that a direct valuation under $\mathbb{Q}$ is appropriate.

3.4 Parameter Assumptions

For our numerical analysis, we consider a male policyholder who purchases a 15-year VA with a return-of-investment GMWB at age 55.[11] Fee and guarantee structures are typical for contracts offered in practice. We further assume that the policyholder maximizes his expected lifetime utility over consumption and bequests, and that he exhibits CRRA preferences, that is

$$u_C(x) = \frac{x^{1-\gamma}}{1-\gamma}.$$

Assumptions about contract specifications and policyholder characteristics in the benchmark case are displayed in Table 1.

[11] We assume that his mortality follows the *2007 Period Life Table for the Social Security Area Population* for the United States (`http://www.ssa.gov/oact/STATS/table4c6.html`).

Parameter Assumptions

	Parameter	Value	Source
Age at inception	x	55	
Time to maturity (years)	T	15	
VA principal	P_0	$100,000$	
Guarantees included		GMWB	
Guarantee fee	ϕ	50 bps	
Annual guaranteed amount	g^W	$7,000$	
Excess withdrawal fee	$s_1, \ldots, s_8$	$8\%, 7\%, \ldots, 1\%$	
	$s_t, t \geq 9$	0%	
Early withdrawal tax	s^g	10%	U.S. tax policy
Life expectancy at maturity (years)	e_{x+T}	12.6	IRS (2003)
Income tax rate	τ	25%	
Capital gains tax rate	κ	15%	
Risk-free rate	r	5%	3-Month Treasury CMR, 1982-2010
Mean return on asset	μ	10%	S&P 500, 1982-2010
Volatility	σ	17%	
Initial wealth	W_0	$500,000$	U.S. Census data*
Income	I_t	$40,000$	
Risk aversion	γ	2	Nishiyama and Smetters (2005)
Subjective discount rate	β	0.968	
Bequest motive	B	1	

Table 1: Parameter Choices for benchmark case.

* In 2007, the median net worth of a U.S. household where the head is age 55 to 64 was roughly 250,000. Median annual (gross) income is around 57,000 for the same category. Our assumptions are based on anecdotal evidence that holders of VA policies are generally wealthier than average. In addition, our results indicate that the choices of wealth, income, etc. do not have a considerable effect on withdrawal behavior.

Account Value Grids

	Mean	95%-ile	99%-ile	99.9%-ile	99.99%-ile	max. grid points
X_T^-	$361,000$	$933,000$	$1,477,000$	$2,579,000$	$3,970,000$	$6,000,000$
A_T^-	$709,000$	$1,111,000$	$1,339,000$	$1,637,000$	$1,981,000$	$7,800,000$
$A_T^- + X_T^-$	$1,070,000$	$2,025,000$	$2,791,000$	$4,103,000$	$5,758,000$	n/a

Table 2: Distribution of terminal VA and outside account values (based on MC simulation, cf. 3.3), and choice of maximum grid point.

3.5 State Variable Grids

As discussed above, the choice-dependent state variables in our life-cycle model are A_t^-, X_t^-, $G_t^{\cdot}$, and H_t. The guarantee account $G_t^{\cdot}$ and the tax base H_t are bounded from above by their starting value $G_0^{\cdot} = P_0$ and $H_0 = P_0$, respectively. For both accounts, we divide the interval $[0, P_0]$ into 16 grid points, including the boundaries. Now note that the tax base can never fall below the benefits base: Both start off at the same level, namely the principal, and both are only affected by withdrawals. The benefits base, however, is reduced by *at least* the withdrawal amount (and possibly more if the withdrawal amount exceeds the annual guaranteed amount); the tax base, on the other hand, is reduced *at most* by the withdrawal amount (namely if withdrawals come from the principal, not earnings). Therefore, we only need to consider state vectors for which $H_t \geq G_t^{\cdot}$.

Since policyholder preferences are assumed to exhibit decreasing absolute risk aversion (cf. Section 3.4), and in the interest of keeping grid sizes manageable and the implementation computationally feasible, we choose grids for VA and outside account that are increasing in distance between grid points. More specifically, for the VA account X_t^-, we divide the interval from 0 to 6 million into 64 grid points, whereas for the outside account A_t^- we use 49 grid points and a range from 0 to 7.8 million. As displayed in Table 2, these values are well above the 99.99th percentile of account values as determined by simulation (see Section 3.3).[12]

4 Results I: Withdrawal behavior in the Life-Cycle Model

One of the primary objectives of this paper is to determine whether it is optimal for policyholders to withdraw prematurely from their VA. We commence by analyzing withdrawal patterns and incentives for the benchmark case parameters (cf. Table 1). Subsequently, we discuss how differences in underlying parame-

[12] We choose the grid for the outside account based on the percentiles of the combined terminal account values, $A_T^- + X_T^-$, in order to be able to accurately value the policyholder's lifetime utility even if he chooses to fully surrender his VA account.

ters affect the optimal withdrawal behavior.

4.1 Optimal Withdrawal Behavior in the Benchmark Case

Our key observation is that in the presence of taxation early withdrawals are an exception rather than the norm. More specifically, for the benchmark case parameters roughly 76% of all possible scenarios entail no withdrawals until maturity. And in only about 5% of all cases will our representative policyholder withdraw his entire guaranteed amount (cf. Table 3, Column [1]).

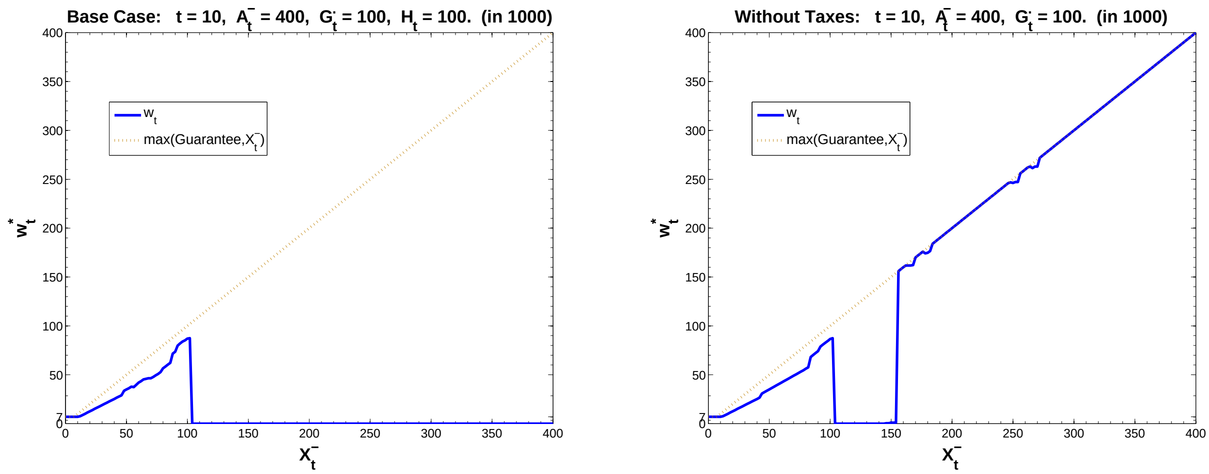
Figure 1 depicts optimal withdrawals, w_t, for our utility-based model as a function of the VA account value X_t^-, in the presence and in the absence of tax considerations, whereby we also include the maximal possible withdrawal amount as a reference. We find that withdrawal patterns are very similar for account values below the benefits base $G_t^{\cdot}$ – which coincides with the tax base H_t. Here, the policyholder withdraws the majority of his account since withdrawals are neither taxed nor subject to fees. However, the optimal strategies in the two cases differ fundamentally when the account is above the benefits and tax base: While we observe no out-of-the-money withdrawals *with* taxes, in the absence of tax considerations the policyholder surrenders his contract if the VA account exceeds approximately 150,000.[13] The intuition for this observation is that the benefit of deferred taxation outweighs the guarantee fees, whereas – when withdrawals are not taxed – the benefits of downside protection does not compensate for the incurred fees. For account values close to but above the benefits base, however, the latter comparison is inverted, leading to no withdrawals even in the case without taxation.

The findings in the absence of taxation are consistent with results from the existing literature that analyzes optimal withdrawal behavior based on arbitrage pricing theory (see e.g. Chen et al. (2008)). These studies also derive fair guarantee fees that are significantly above concurrent market rates. In contrast, our analysis – if we include taxation – suggests that an annual guarantee fee of $\phi = 50$ bps seems sufficient to cover the expected costs of the guarantee. More specifically, we find that the risk-neutral actuarial present value at time 0 of the collected guarantee fees is $5,971$ (plus an additional 30 in excess withdrawal charges), which far exceeds the risk-neutral value of payments attributable to the GMWB of $1,480$.

Figure 2 displays the optimal withdrawal behavior as a function of the VA account value X_t^- at different points in time and for differing but fixed levels of the benefits base $G_t^{\cdot}$ and the tax base H_t. The outside account A_t^- is identical over all panels, but sensitivities are analyzed in Section 4.2 below.

During the first four contract years, the policyholder has not reached age 59.5, and therefore all withdrawals are subject to a 10% early withdrawal tax. Withdrawals are still profitable for low account values as the policyholder may not be able to withdraw the guaranteed amount otherwise. However, this becomes less likely as the account value increases. For instance, as demonstrated by Figure 2(a) for the case of $t = 4$, there are no withdrawals beyond an account value of about $60,000$. Moreover, we do not observe excess

[13] It is worth noting that we would also observe positive withdrawals when the VA makes up the vast majority of the policyholder's total wealth, due to an overexposure to equity risk. More precisely, since he cannot change the allocation inside the VA, he withdraws from the VA to place the funds (after possible fee and tax payments) in the risk-free outside account. For even larger values of the VA account, the policyholder may also want to consume beyond the limits of his outside wealth, leading to withdrawals for the purpose of consumption smoothing. However, such scenarios are extremely unlikely (we observed no such case in 5 million simulations of the benchmark case) and do not have a sizable impact on the value of the guarantee. Hence, we will not delve into this issue any further.



(a) Base Case: $\tau = 25\%$.

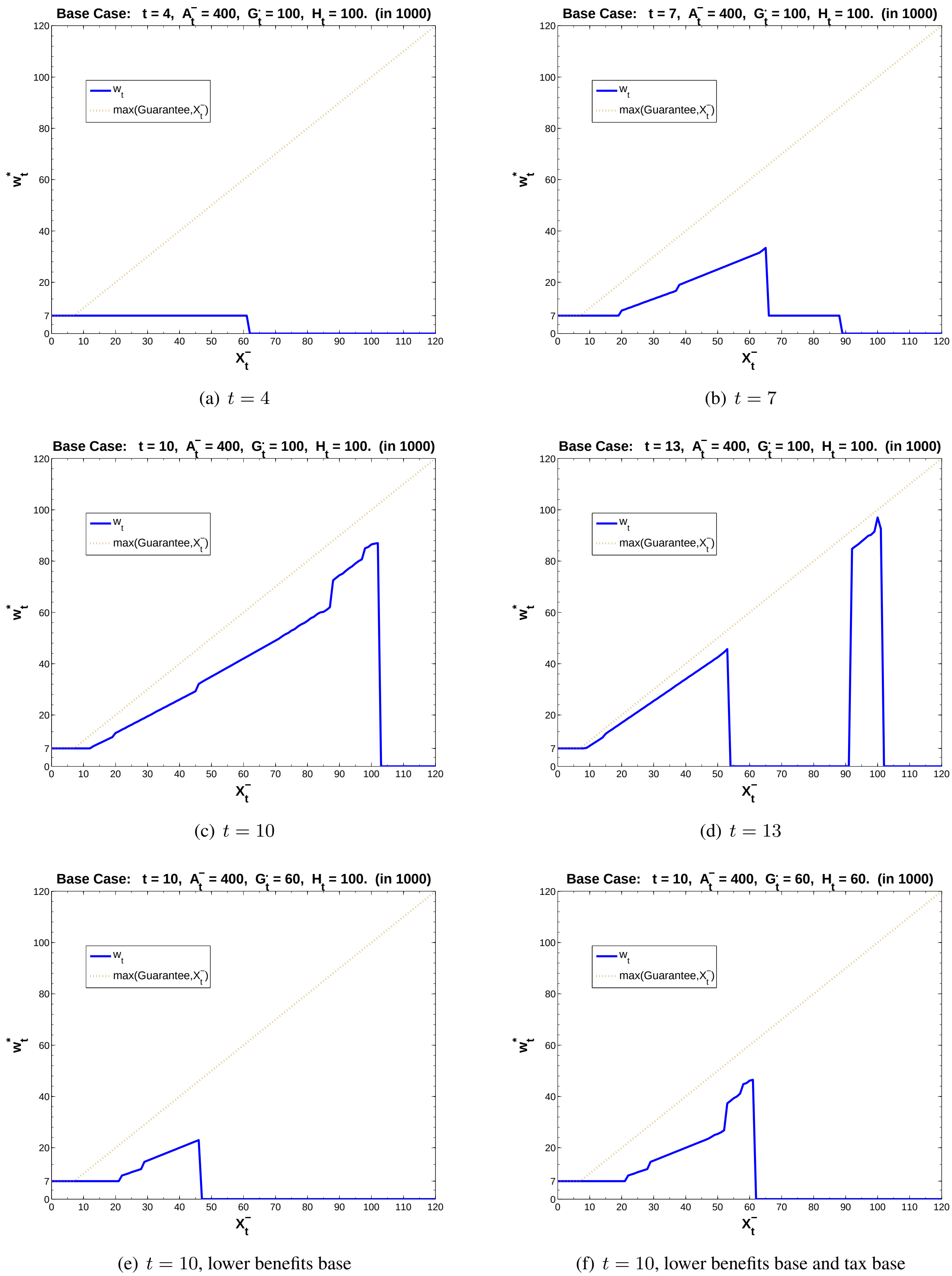
(b) No Taxes: $\tau = \kappa = 0\%$.

Figure 1: Withdrawal behavior with and without taxes as a function of the VA account X_t^-.

withdrawals in this case, which we attribute to the 15% charge ($10\% + 5\%$ excess withdrawal fee) on all excess withdrawals.

After his 60th birthday, the policyholder can withdraw $7,000$ annually free of charge, and he will do so whenever the VA is below the tax base, as evidenced by Figure 2(b) for $t = 7$. In addition, we observe excess withdrawals, despite a 2% excess withdrawal fee. The intuition is that this is the policyholder's best chance to access as much of the aggregate guarantee as possible: He withdraws the amount that reduces his benefits base to a level that leaves roughly the guaranteed amount of $7,000$ for each of the remaining withdrawal dates. In other words, he withdraws as much as possible without jeopardizing the future payouts from his GMWB rider. As the VA account increases, the optimal withdrawal amount increases as well and so do the associated excess withdrawal costs. Yet, beyond a certain amount, the benefit of the maximal guarantee will no longer compensate for the excessive withdrawal fee, so that the policyholder will prefer to withdraw the guaranteed amount only. When the excess withdrawal fee vanishes, however, excess withdrawals are optimal up to the full benefits base as evidenced by Figure 2(c) (time $t = 10$). Moreover, the optimal withdrawal curve becomes steeper as time progresses since there are fewer periods – and hence a smaller aggregate guaranteed amount – remaining.

This pattern of excess withdrawals continues as we approach the end of the contract term. However, during the final years before maturity, we observe zero withdrawals for low, but not too low, account values relative to the tax and benefits base (cf. Figure 2(d)). This can be once again explained by tax benefits: Since these account values are considerably below the tax base, any (likely) return over the remaining contract years will be tax free – unlike investments in the outside account; hence, even if the guarantee is worthless and cannot be brought in the money unless incurring considerable withdrawal penalties, paying the fee inside the tax-sheltered VA account is optimal. Keeping these effects in mind, it is then also not surprising that this "gap" widens as we get closer to maturity.



(a) $t = 4$

(b) $t = 7$

(c) $t = 10$

(d) $t = 13$

(e) $t = 10$, lower benefits base

(f) $t = 10$, lower benefits base and tax base

Figure 2: Withdrawal behavior in the Base Case as a function of the VA account X_t^-.

This is also the motivation that withdrawals vanish beyond a certain point when the tax base exceeds the benefits base, as shown in Figure 2(e). The ability to save sheltered of taxes yields no withdrawals beyond about $45,000$, whereas below that amount bringing the guarantee into the money pays off. However, if we also decrease the tax base so that it is again on par with the benefits base (Figure 2(f)), we uncover a similar pattern as before (Figure 2(c)). In particular, the relative size of the outside account value A_t^- appears to have little effect on optimal withdrawal patterns.

4.2 Sensitivities to Key Unobservables

A primary concern when implementing a utility-based model in practice is the choice of parameter assumptions, particularly those the insurer has little or no information about. In our case, these key unobservables include the level of initial wealth W_0, the policyholder's annual labor income I_t, his level of risk aversion γ, and his bequest motive B.[14] In that regard, our findings provide some encouraging evidence: As Table 3 shows, variations in these characteristics have relatively little effect on aggregate withdrawal statistics. Nonetheless, the effects of these deviations from the benchmark case are quite systematic and in some cases counterintuitive at first sight: For instance, while withdrawals appear to increase with the level of initial wealth, we observe the opposite effect when increasing annual income. Furthermore, it may appear that a more risk averse policyholder will make greater use of his guarantees.[15] Again, this is not what we observe.

Instead, our findings reflect the insight that in addition to taxation, in-the-moneyness, and fee structure, none of which change when varying the preference parameters, withdrawals are also affected by the policyholder's bequest motive. Here, it is important to realize that while the outside account "pays" irrespective of the policyholder's life status, the guarantee is only material while he is alive. Conversely, the annual income stream essentially is a life annuity, so that income and guarantee can be viewed as substitute goods. Moreover, a risk averse policyholder with a positive bequest motive will optimally allocate a certain proportion of his current wealth in annuities while the remainder should also serve as bequest protection. And under decreasing absolute risk aversion, the optimal *amount* of annuities is increasing in the wealth level.

Hence, if the policyholder starts off with a higher degree of wealth (Columns [2] vs. [1], [3] vs. [1] and [5] vs. [4]), he will ideally allocate a larger absolute amount to annuities. Since income remains fixed, his demand for the guarantee increases, and the policyholder has a greater incentive to move the guarantee (further) into the money, thus withdrawing more frequently and making more use of his guarantee.

As income increases (Columns [4] vs. [1] and [5] vs. [3]), on the other hand, ceteris paribus the demand for the guarantee goes down since the two are substitutes. The policyholder thus is willing to give up parts

[14] More generally, one might question the fundamental assumption of our expected utility framework. While a detailed analysis with respect to preferences or other assumptions is beyond the scope of this paper, it is worth noting that – say – more complex utility structures reflecting non time-separable preferences (see e.g. Epstein and Zin (1989)) and similar features can be easily implemented in our model. Furthermore, we highlight in the conclusions that we see evidence that such modifications will have little effect on our results.

[15] Consider e.g. a simplified version of the scenario in Figure 2(a): The policyholder faces the decision whether to withdraw the guaranteed amount of $7,000$. Withdrawing the money results in a *certain* payout of $6,300$ (i.e. $7,000$ minus 10% early withdrawal tax); otherwise he faces a lottery with a random payout of either $7,000$ or 0, depending on whether the portfolio will increase above the guarantee by maturity of the policy. This argument may suggest that a more risk averse policyholder would be drawn towards the certain payout, that is towards making the withdrawal.

Sensitivities to Key Unobservables

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
	BM	$W_0 = 250K$	$W_0 = 1M$	Inc = $70K$	$W_0 = 1M$, Inc = $70K$	$\gamma = 2.88$	$B = 0.2$
$\mathbb{E}^{\mathbb{Q}}$[Fees] [a]	$5,971$	$5,974$	$5,958$	$5,967$	$5,950$	$5,955$	$6,024$
$\mathbb{E}^{\mathbb{Q}}$[Excess-Fee]	30	33	30	34	33	35	48
$\mathbb{E}^{\mathbb{Q}}$[GMWB]	$1,480$	$1,423$	$1,534$	$1,421$	$1,516$	$1,280$	$1,765$
$\mathbb{E}$[agg. w/d] [b]	$13,449$	$13,431$	$13,533$	$13,727$	$13,852$	$14,144$	$11,693$
$\mathbb{E}$[excess w/d]	$8,555$	$8,636$	$8,463$	$9,058$	$8,915$	$9,692$	$7,005$
$\mathbb{E}$[w/d , $t \leq 4$]	151	51	263	52	218	51	264
$\mathbb{E}$[w/d , $5 \leq t \leq 8$]	$5,263$	$5,476$	$5,288$	$5,483$	$5,502$	$5,727$	$3,770$
$\mathbb{E}$[w/d , $t \geq 9$]	$8,035$	$7,903$	$7,982$	$8,193$	$8,132$	$8,366$	$7,659$
$\mathbb{E}[G_T^{\cdot}]$ [c]	$85,961$	$85,943$	$85,892$	$85,763$	$85,645$	$85,369$	$87,499$
$\mathbb{P}(G_T^{\cdot} == 0)$	4.9%	5.0%	4.9%	5.8%	5.8%	6.0%	3.1%
$\mathbb{P}(G_T^{\cdot} < P_0)$	23.6%	23.9%	23.6%	23.5%	23.6%	22.5%	19.4%
$\mathbb{E}[H_T]$	$86,865$	$86,896$	$86,775$	$86,686$	$86,523$	$86,298$	$88,525$

Table 3: Withdrawal statistics in the benchmark case (Column [1]), and for policyholders with different (unobservable) characteristics.

[a] All under the risk-neutral measure $\mathbb{Q}$: Actuarial present value (APV) of fees collected by the insurer; APV of excessive withdrawal fees charged to policyholder; APV of payouts made to policyholder only due to GMWB (i.e. when $X_t^- = 0$).
[b] Mean aggregate withdrawal amount (pre-maturity); excess withdrawal amount; withdrawals subject to early w/d tax; withdrawals subject to excess w/d fees but no early w/d tax; "free" withdrawals. (Note: values are added up under the physical measure $\mathbb{P}$ and without accounting for the time value of money.)
[c] Average level of benefits base at maturity; probability that the policyholder uses full guarantee; probability that at least one withdrawal is made; Average tax base at maturity.

of his guarantee in exchange for non-life contingent funds, and he can do so by not withdrawing and thereby letting the guarantee move out of the money. Therefore, the value of the GMWB decreases.

Similarly, a more risk averse policyholder (Column [6][16] vs. [1]) with a positive bequest motive is also more averse to mortality risk, and thus has an increased preference for a safe asset – in the form of the outside account – over a risky asset in the form of the annuity. Thus, ceteris paribus, the demand for the guarantee declines, and so will the incentives to withdraw and hence the value of the GMWB. Finally, a lower bequest motive (Column [7] vs. [1]) reduces the demand for payments in death states, thus raising the relative demand for the guarantee, and resulting in increased withdrawals.

Figure 3 further affirms this intuition by depicting withdrawals as a function of the VA account value X_t^- at time $t = 4$ (cf. Figure 2(a)). As discussed above, the withdrawal decision here is based on a trade-off between increasing the value of the guarantee and the cost of withdrawing. In line with the portrayed intuition, we observe slight changes in response to changes in the preference parameters. More precisely, the "safe asset" characteristic of the outside account yields increases in the withdrawal area for larger values of the outside account (Figure 3(b)) and for a lower bequest motive (Figure 3(e)). Conversely, the range of withdrawing at the guaranteed level shrinks for policyholders with a larger income (Figure 3(c)) or a higher level of risk aversion (Figure 3(d)).

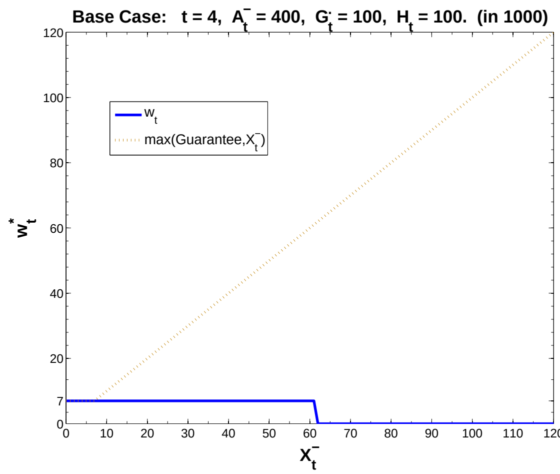
Hence, these deviations can be ultimately attributed to a lack of market completeness regarding bequest protection. In particular, the sensitivities might be less pronounced – or not even existent – if the policyholder had access to life insurance. This indicates that the sensitivity may be even less significant in practice, where agents have access to a large menu of life- and morbidity-contingent securities.

All in all, the insignificance of the sensitivity of optimal withdrawal behavior – and, thus, of resulting financial statistics – points towards value-maximization as being the key driver. In particular, it does not seem to be the market incompleteness that is responsible for the divergence of actual observations and results from the actuarial literature, which rely on value-maximization approaches. Rather, it appears to be a matter of perspective: While there, the calculations were carried out from the company's position, the focus should be on the policyholder's point of view. In particular, it appears imperative to take investment taxation rules into consideration. This idea is developed in the following section.

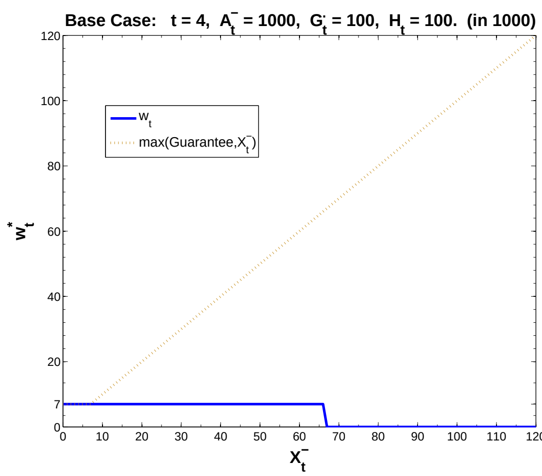
5 Risk Neutral Valuation from the Policyholder's Perspective

One of the primary results from the previous section is that optimal policyholder behavior in the life cycle model seems to be mainly driven by value maximization. This raises the interesting possibility that we can after all meaningfully analyze optimal policyholder behavior by a risk-neutral valuation approach – associated with all its benefits such as independence of preferences, wealth, and, in particular, outside allocation and consumption decisions. Specifically, the reason for the meager performance of risk-neutral valuation approaches in explaining observed prices and exercise patterns so far seems to be the disregard of important factors affecting the policyholder's decision, especially tax considerations, rather than a fundamental

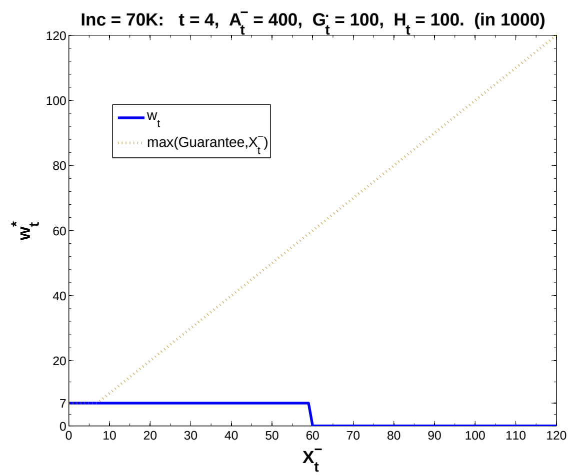
[16]The parameter value $\gamma = 2.88$ was chosen in order to attain a Merton Ratio (see Merton (1969)) of 60%, consistent with the "rule of thumb" that investors should optimally hold 60% of their assets in stocks and 40% in bonds (cf. Gerber and Shiu (2000)).



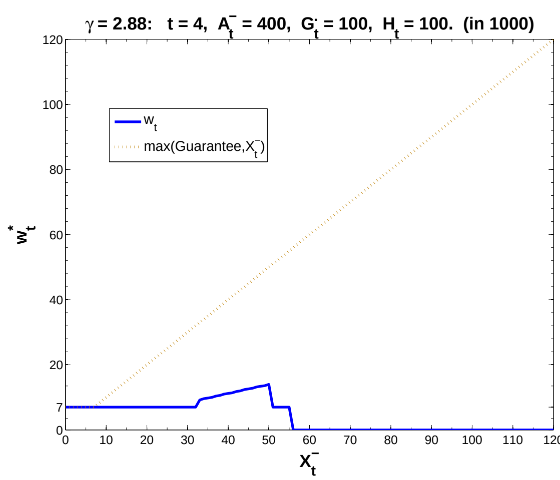
(a) Benchmark Case



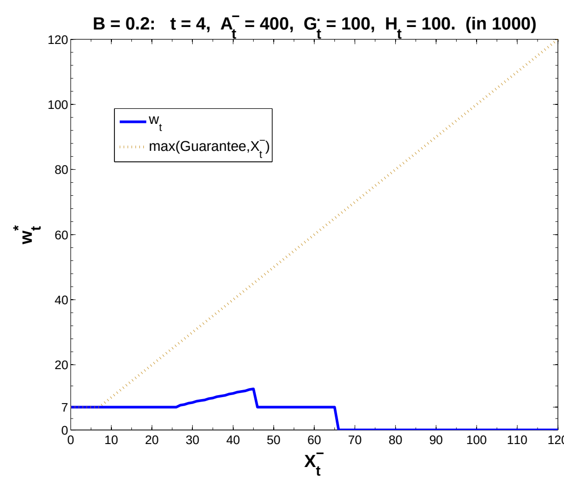
(b) More Wealth, i.e. Larger Outside Account



(c) Larger Income



(d) More Risk Averse



(e) Lower Bequest Motive

Figure 3: Withdrawal behavior at $t = 4$ when key unobservables change.

methodological flaw.

To cope with tax considerations, in this section we develop a general methodology for valuing cash flows when tax rates differ over investment opportunities. Subsequently, in order to vindicate – or rather rectify – the risk-neutral valuation approach to optimal policyholder behavior, we describe how to implement the method in the context of this paper. Finally, we present numerical results and contrast them with those from the life-cycle model.

5.1 Valuation of Cash Flows under Different Taxation Schemes

Arbitrage pricing in the presence of taxation is an intricate issue. For instance, as demonstrated by Ross (1987), no universal pricing measure exists when tax rates vary for different agents. In fact, an agent's valuation of a given cash flow depends on his individual endowment. Our primary idea is that we can nevertheless identify a unique – though individual – valuation methodology if the pre-tax financial market for "ordinary" investments such as stocks and bonds is complete.[17] Then, consistent with standard arbitrage pricing arguments, we define the time-zero value of a post-tax cash flow X as the amount necessary to set up a pre-tax portfolio that – after taxes – replicates X. This valuation rule is individual in the sense that it depends on the investor's current position. For instance, if the investor has additional investments that may offset tax responsibilities for the replicating portfolio, the relative value of the replicating portfolio will increase.

More formally, we consider an individual with endowment A and access to underlying securities such as stocks and bonds subject to capital gains taxation as described in Section 2.4. We assume that the pre-tax market is complete. Hence, there exists a unique equivalent martingale measure, denoted by $\mathbb{Q}$, such that the cost for a replicating portfolio for any pre-tax cash flow is given by its expected discounted value under $\mathbb{Q}$ with respect to the numeraire $(B_t)_{t\geq 0}$ (savings account). Furthermore, for simplicity and without much loss of generality, we assume that all cash flows are realized at the end of years only.

We first focus on a single year $(t, t+1]$ and consider the (post-tax) cash flow $X_{t+1} \equiv X$ at time $t+1$ potentially originating from a separate investment opportunity subject to different tax rules. We are interested in its value at time t. Denote by A_{t+1} the investor's (state-specific, post-tax) endowment at time $t+1$, with known value A_t at time t. Then, if an amount V_t is necessary to replicate the post-tax cash flow $Y \equiv X_{t+1} + A_{t+1}$, then the (marginal) value of X_{t+1} is given by $X_t \equiv V_t - A_t$. Hence, the valuation problem reduces to determining the (pre-tax) replicating portfolio, and thereby V_t.

Define Z as the corresponding *pre-tax* cash flow required to attain Y after tax payments, i.e.

$$Y = Z - \kappa \cdot [Z - V_t]^+ . \tag{15}$$

Inverting the function on the right-hand side, Equation (15) may be restated as

$$Z = Y + \frac{\kappa}{1-\kappa} \cdot [Y - V_t]^+ . \tag{16}$$

[17] Note that even in the absence of taxes, arbitrage pricing theory does not give a unique pricing rule if the market is incomplete.

On the other hand, since V_t is the cost of setting up the pre-tax cash flow Z and the pre-tax market is complete, we have:

$$V_t = \mathbb{E}_t^{\mathbb{Q}}\left[\frac{B_t}{B_{t+1}} \cdot Z\right].$$

Thus, with Equation (16), we obtain

$$V_t = \mathbb{E}_t^{\mathbb{Q}}\left[\frac{B_t}{B_{t+1}} \cdot Y\right] + \frac{\kappa}{1-\kappa} \cdot \mathbb{E}_t^{\mathbb{Q}}\left[\frac{B_t}{B_{t+1}} \cdot (Y - V_t)^+\right], \tag{17}$$

with the unknown V_t, depending on the state of the world at time t. Hence, Equation (17) presents a (non-linear) valuation rule for Y and, thus, X_{t+1}, which gives a unique value V_t as shown by the following result.[18]

Proposition 5.1. *Any time $t+1$ post-tax cash flow X_{t+1} can be valued uniquely by the policyholder at time t, and its time-t value is given by $V_t - A_t$, where V_t is the unique solution to Equation* (17)*.*

To generalize this method for payoffs multiple years ahead, consider again the cash flow $X_{t+1} \equiv X$ at time $t+1$. As we have argued, its time t value is given by $X_t = V_t - A_t$, which again can be interpreted as a post-tax cash flow. Hence, its value at time $t-1$ is given by $X_{t-1} \equiv V_{t-1} - A_{t-1}$, where V_{t-1} is the setup cost for a replicating portfolio for the post-tax cash flow $X_t + A_t = V_t$ as above.[19] Hence, the time $t-2$ value – and similarly the values at times $t-3, t-4, \ldots$, and eventually the time-zero value – can be determined recursively by serially solving the corresponding Equation (17).

This procedure allows us to evaluate every combination of post-tax cash flows – for any given outside investment and consumption strategy – uniquely as the marginal increase required in today's outside portfolio in order to replicate the aggregate cash flow. In particular, it can be applied to analyze the cash flows from the VA contract introduced in the previous sections.

5.2 The Policyholder's Optimization Problem under Risk-Neutral Valuation

Following the previous analysis, we implement a risk-neutral valuation approach for our representative VA policyholder assuming he maximizes the value of all benefits. Akin to the life-cycle model, the problem will be set up recursively, period by period.

However, since the valuation introduced in the previous subsection applies only *locally*, i.e. given investment and consumption decisions, a "proper" risk-neutral valuation methodology formally still requires us to consider the policyholder's entire portfolio. This implies that the value-maximization approach loses one of its key benefits: the ability to focus solely on the cash flows associated with the VA, and to ignore all other factors. In particular, this means that the model essentially possesses roughly the same level of complexity as the utility-based framework and the numerical implementation of this approach will be equally cumbersome.

[18] The proof is provided in Appendix B.2.

[19] Note that for simplicity of exposition we disregard consumption and income that would alter A_t. Generalizations are straight forward.

The complexity greatly reduces when we rely on an exogenous assumption about (outside) investment and allocation decisions. While formally imposing such assumptions appears problematic, it is important to note that their effect in regards to the VA solely comes into play when there is a possibility to offset gains and losses. In all other case, the outside account is immaterial.[20] Hence, supposing that the effects are minor, we consider the simplest possible case when no offsets are possible at all, which can be represented by simply setting the outside account to zero.

Under the same specifications as in Section 2, at each policy anniversary date, the policyholder's decision is then based solely on observing the concurrent state variables X_t^-, $G_t^\cdot$ and H_t, and only entails choosing the withdrawal amount w_t. Again, we implement the model numerically in a Black-Scholes framework using recursive dynamic programming. In doing so, we compute the value of the payoff at maturity, and then recursively proceed similar to Algorithm 3.1. More specifically, for each $t = T-1, \ldots, 1$, the policyholder chooses the withdrawal amount that maximizes the continuation value, i.e.

$$V_t(X_t^-, G_t^\cdot, H_t) = \max_{w_t}(w_t - \text{fee}_I - \text{fee}_G - \text{taxes}) + \tilde{V}_t, \tag{18}$$

where $\tilde{V}_t$ is given implicitly by (cf. Equation (17))

$$e^r \cdot \tilde{V}_t - \mathbb{E}_t^{\mathbb{Q}}[q_{x+t}\, b_{t+1} + p_{x+t}\, V_{t+1}(.)] - \frac{\kappa}{1-\kappa}\mathbb{E}_t^{\mathbb{Q}}\left[(q_{x+t}\, b_{t+1} + p_{x+t}\, V_{t+1}(.) - \tilde{V}_t)^+\right] = 0.$$

Clearly, this optimization problem is subject to a variety of constraints regarding account evolution and updating, which are similar to the implementation of the life-cycle framework. For more details, see Appendix A.3.

5.3 Results II: Withdrawal behavior under the Risk-Neutral Framework

Table 4 shows the resulting values and aggregate withdrawal statistics for the risk-neutral valuation approach (RNV) in comparison to the life-cycle (LC) benchmark model in the case with (Column [1]) and without (Column [2]) taxes. In addition, we present the respective results for contracts that differ from the benchmark case, by assuming there are no fees on excess withdrawals (Column [3]), by reducing the equity exposure within the VA account (Column [4]), and by varying the income tax rate (Column [5]). While we observe considerable differences *between* the latter results and the benchmark case – which are discussed in detail in Section 6 below – we find that *within* each specification the differences between the results from the two approaches are quite small.

Therefore, overall our presumption that the results for the two approaches will be similar proves true. The same conclusion can be drawn from Figure 4, where the optimal withdrawal behavior at times $t = 10$ and $t = 4$ is displayed for both approaches in the case with and without taxes. In particular, we do not have any withdrawals in the presence of taxes if the account value significantly exceeds the tax base (Figures 4(a) and 4(b)). In contrast, in the absence of taxes, it is optimal to withdraw when the guarantee is out-of-the money under both optimality criteria, indicating their alignment (Figures 4(e) and 4(f)).

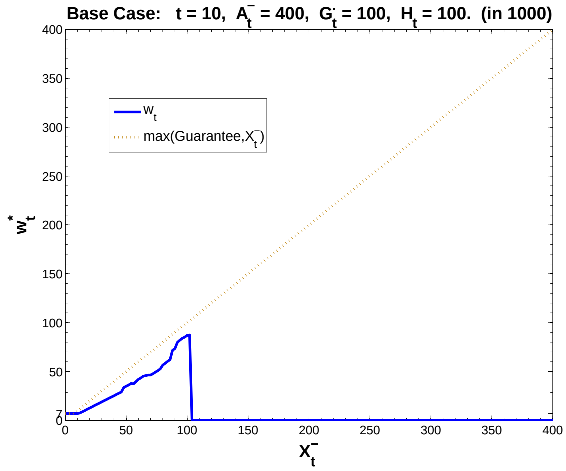
[20] Obviously, this is quite different from the life-cycle model, where the outside account and consumption decisions directly enter the value function. In contrast, here we have only an indirect effect through potential tax offsets.

Sensitivities to Contract Specifications and Tax Rates

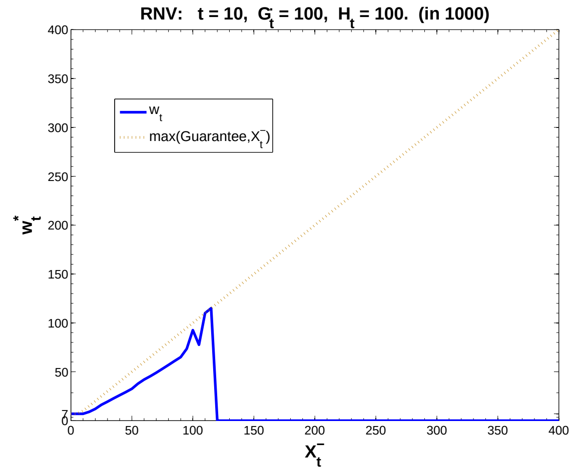
	[1]		[2]		[3]		[4]		[5]		[6]
	BM		BM w/o taxes		$s = 0$		$\nu^X = 90\%$		$\tau = 30\%$		ITM
	LC	RNV	LC	RNV	LC	RNV	LC	RNV	LC	RNV	
$\mathbb{E}^{\mathbb{Q}}$[Fees] [a]	5,971	5,708	3,265	3,299	5,366	5,388	5,987	5,682	5,779	5,495	5,735
$\mathbb{E}^{\mathbb{Q}}$[Excess-Fee]	30	162	18	10	0	0	38	142	96	258	0
$\mathbb{E}^{\mathbb{Q}}$[GMWB]	1,480	2,094	3,068	3,163	1,557	1,974	1,033	1,517	1,591	2,179	1,924
$\mathbb{E}$[agg. w/d] [b]	13,449	19,240	190,340	191,320	22,322	24,020	12,232	18,640	18,380	23,140	13,435
$\mathbb{E}$[excess w/d]	8,555	13,029	135,110	136,140	19,903	18,811	8,369	13,562	13,830	16,498	0
$\mathbb{E}$[w/d , $t \leq 4$]	151	1,094	27,400	27,400	65	1,025	63	1,060	150	1,780	6,685
$\mathbb{E}$[w/d , $5 \leq t \leq 8$]	5,263	7,639	27,110	25,980	18,406	14,600	5,139	6,540	7,993	9,610	3,549
$\mathbb{E}$[w/d , $t \geq 9$]	8,035	10,507	135,830	137,940	3,851	8,395	7,030	11,040	10,237	11,750	3,201
$\mathbb{E}[G_T^{\cdot}]$ [c]	85,961	80,974	6,687	6,794	76,891	76,153	87,421	81,851	81,689	77,503	86,565
$\mathbb{P}(G_T^{\cdot} == 0)$	4.9%	9.3%	84.0%	83.6%	12.0%	13.5%	5.4%	11.4%	10.6%	13.2%	2.5%
$\mathbb{P}(G_T^{\cdot} < P_0)$	23.6%	13.0%	89.5%	88.7%	24.0%	27.4%	20.9%	24.5%	25.4%	29.3%	47.9%
$\mathbb{E}[H_T]$	86,865	81,809	–	–	78,076	76,863	88,122	82,324	82,503	78,116	86,565
V_0 [d]	−3.9397	100,064	–	–	–	–	−3.9458	99,301	–	–	–

Table 4: Withdrawal statistics in the life-cycle model (LC) and under risk-neutral valuation (RNV): for the benchmark case with (Column [1]) and without ([2]) taxation, for different contracts ([3], [4]) and tax rates ([5]). Column [6] displays results when the policyholder withdraws the guaranteed amount if and only if the guarantee is in the money.

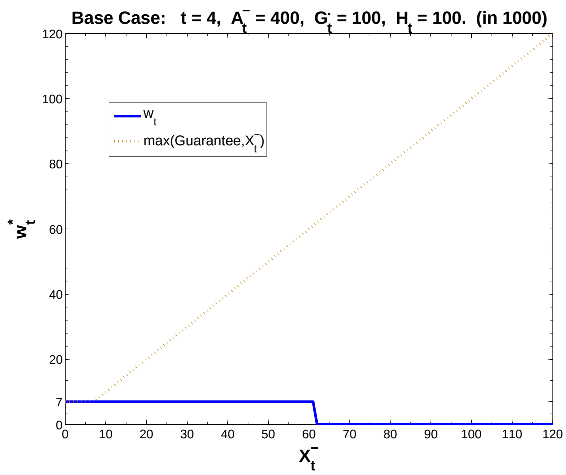
[a], [b], [c] See Table 3.
[d] Time zero expected discounted lifetime-utility / value.



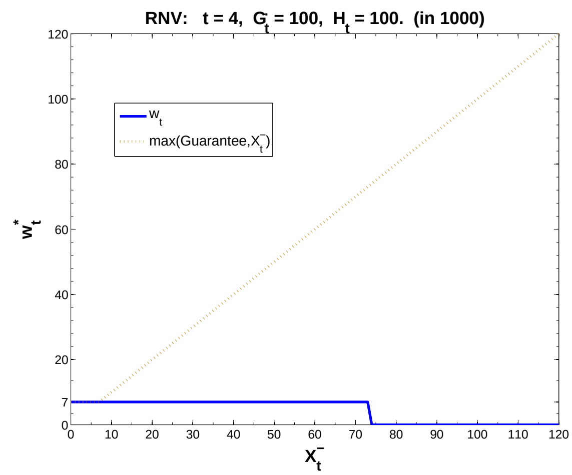
(a) Benchmark Case, $t = 10$



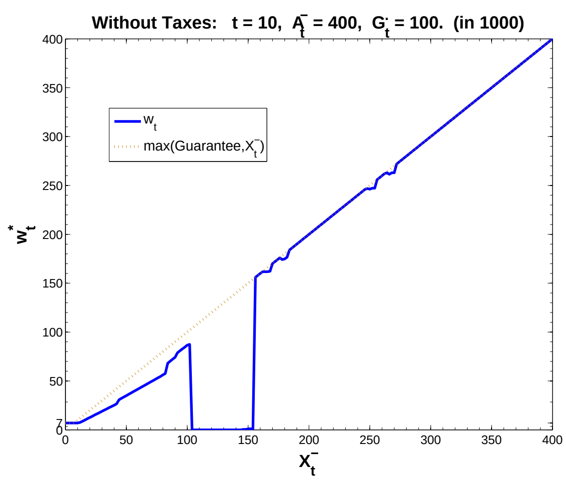
(b) Risk-Neutral Valuation, $t = 10$



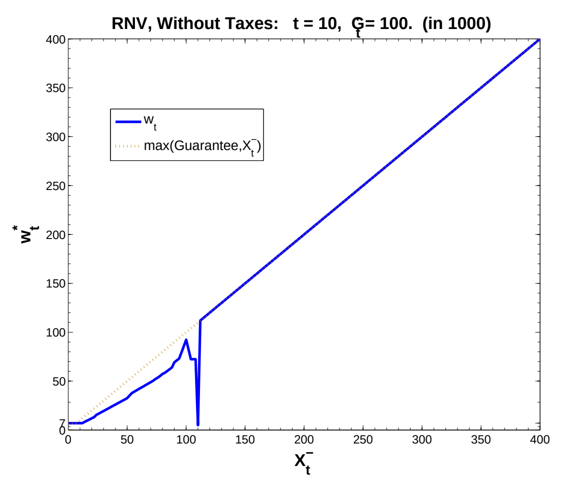
(c) Benchmark Case, $t = 4$



(d) Risk-Neutral Valuation, $t = 4$



(e) Benchmark Case, no tax



(f) Risk-Neutral Valuation, no tax

Figure 4: Withdrawal behavior under risk-neutral valuation.

The rather slight deviations between the two approaches are in line with the – rather slight – sensitivities of the results from the life-cycle model to wealth and preferences, as analyzed in Section 4.2. There we demonstrate that the response in withdrawal behavior to changes in wealth, income, or risk aversion are motivated by the allocation to death and life states. The risk-neutral approach can now be interpreted as the limiting case of a policyholder with infinite wealth or zero risk aversion. For instance, in Section 4.2 we explain that early in the contract ($t = 4$), an increase in wealth yields a wider withdrawal range (Figure 3(b) vs. 3(a)). Accordingly, the risk-neutral approach results in an even wider range (Figure 4(d) vs. Figures 3(b) and 4(c)). A similar relationship holds when varying risk aversion: High risk aversion leads to a smaller withdrawal range than lower risk aversion (Figure 3(d) vs. 3(a)), which in turn leads to a smaller withdrawal range than the risk-neutral approach (Figure 4(d) vs. 4(c)).

These deviations then also lead to a slight increase in the value of the GMWB rider. In particular, the difference in the values of collected fees and benefits associated with the GMWB – which may be interpreted as the value from the insurer's perspective – decreases from $4,521$ to $3,776$. It is important to note, however, that this value in the benchmark case relies on our specific assumptions on preferences, income and wealth — and therefore may be even smaller for alternative choices. Also, it may change in the presence of alternative investment options. For instance, if the policyholder has access to other life-contingent contracts, withdrawals may be even less affected by the policyholder's allocation motive and, thus, his preferences. In the absence of taxes, the difference in values is even less pronounced (215 in the benchmark case and 146 for risk-neutral valuation), and the big gap between the two tax regimes again resonates the deficiencies of risk-neutral valuation approaches in the previous literature on policyholder exercise behavior within VAs.

6 Results III: Implications for Life Insurance Practice

The results from the previous sections have important – and encouraging – implications for life insurers offering GMWBs. On the one hand, this paper provides theoretical insights into optimal withdrawal patterns that appear to be in line with actual observations. On the other hand, we demonstrate that optimal withdrawal behavior can be analyzed based on the – relative to a life cycle model – simple risk-neutral valuation approach.

However, of course questions arise regarding the generality of the detected patterns, for instance with respect to changes in VA design or the underlying tax rates as different policyholders may fall into different tax brackets. Furthermore, eventual quantitative results depend on specifics of considered contracts, yet a practical implementation of the risk-neutral approach to determine optimal policyholder behavior – despite its relative simplicity – may still be too ambitious for most insurers. In fact, current industry practice is to rely on historic exercise probabilities or static exercise rules, although some insurers indicate they started to use simple dynamic assumptions (cf. Society of Actuaries (2009)).

Therefore, in this section we first analyze the robustness of the uncovered exercise patterns with respect to the underlying contract specification and parameters. Subsequently, to appraise the advancement of imposing dynamic exercise rules in the context of our model, we compare our results to the usage of a simple rule based on the "moneyness" of the guarantee.

6.1 Robustness of the Results

Columns [3] and [4] of Table 4 show the effects when modifying the contract's specifications vis-à-vis the benchmark case. As may be anticipated, removing the excess withdrawal fees (Column [3]) induces more frequent withdrawals early in the contract period. This reduces collected fees and, at the same time, increases the "usage" and therefore the value of the guarantee. These findings suggest that the "actual value" of the excess withdrawal fee to the insurer is larger than the rather moderate amount that is directly attributed to excessive withdrawals in the benchmark case (30) due to changes in policyholder behavior.

In practice, insurers sometimes limit equity exposure in the VA in order to contain the risks. Column [4] of Table 4 provides some evidence to that effect: When reducing equity exposure to 90%, the value of the guarantee diminishes, while collected fees remain relatively unaffected. This of course is a direct consequence of a reduced likelihood of adverse scenarios under the more conservative allocation strategy, since withdrawals are optimal only when the guarantee is "in the money". In other words, the downside protection is more valuable for a more risky investment strategy, so that is is also no surprise that the policyholder, ceteris paribus, will prefer a 100% equity exposure inside the VA (under both models), as shown in the bottom line of Table 4.

Nevertheless, in both cases, when modifying the surrender fee structure or when adjusting the equity exposure, the general withdrawal patterns prevail. In particular, withdrawals still are the exception rather than the norm and mostly occur in adverse scenarios for the underlying index. The driver behind these observations of course is the deferred taxation of investments inside the VA.

Results change dramatically in the absence of taxation (cf. Table 4, Column [2]). More precisely, investment in the VA loses its comparative advantage, withdrawals increase substantially, and – as a result – so does the guarantee value. A similar argument applies if the tax advantage becomes less important, e.g. due to an increase in the income tax rate: The VA becomes less attractive, and the policyholder has a greater incentive to withdraw and invest it outside of the VA. Accordingly, Column [5] of Table 4 shows that an increase from 25% to 30% leads to an increase in withdrawals. However, we find that even under a tax rate of 30% on withdrawals, the deferred taxation feature is valuable enough to sustain an absence of withdrawals in most scenarios. Therefore, we again have a positive answer to the question of generality of our results: Even for relatively high tax brackets, the observed patterns prevail.

6.2 Simple Reduced-Form Exercise Strategies

As indicated above, insurers have started to rely on simple dynamic exercise rules in their calculations, where approaches based on the "moneyness" of the guarantee appear particularly popular. More formally, these rules assume the policyholder withdraws the guaranteed annual amount whenever the guarantee is "in the money" (that is: the account value lies below the benefits base), and zero otherwise, regardless of tax and fee considerations:

$$w_t \equiv \min\{g_t^W, G_t^W\} \cdot \mathbb{1}_{[\{X_t^- \leq G_t^W\}]}.$$

The results in the context of our model are provided in Column [6] of Table 4, and we find that overall the withdrawal statistics are very close to the benchmark case. Given the described withdrawal patterns for the benchmark case, the relatively good performance of the "moneyness" assumption should come as no surprise. Still, the suitability of this assumption – particularly in view of the valuation results – bears good news for the insurance industry. In particular, our models endorse the use of the simple "in-the-money" rule.

This result is also consistent with empirical findings from Knoller et al. (2011), who determine in the case of Japanese VA products that "moneyness (...) has the largest explanatory power for the rate at which policyholders surrender their policies". It is worth noting, however, that in our setting the validity of "moneyness" as a proxy for optimal policyholder behavior is mainly a consequence of the similarities between tax and benefits base.

7 Conclusions and Future Research

The present paper concerns the optimal policyholder exercise behavior within embedded options in life insurance contracts. More specifically, our focus is on withdrawal behavior for the holder of a VA contract including a GMWB rider, even though our insights are not limited to this popular but rather specific product.

Our main conclusion is that the key driver for exercise behavior in the pre-retirement period is value maximization. More precisely, we contrast two approaches to optimal policyholder behavior, namely a lifetime utility model that explicitly allows for outside investment and a risk-neutral value-maximization approach. We find that despite their conceptual differences, both approaches yield very similar withdrawal patterns and aggregate withdrawal statistics. In particular, the possibility to defer investment taxation within the VA seems to be the dominating factor, inducing policyholders to only withdraw when the guarantee is in the money, i.e. in adverse market conditions. As a consequence, under our parameter assumptions, a guarantee fee of 50 bps, which is close to levels encountered in practice, more than sufficiently provides for a return-of-investment GMWB.

The latter result is in stark contrast to findings in the actuarial literature based on value-maximizing approaches, which suggest much higher guarantee fees. In fact, it is exactly this disparity that has led researchers to gravitate towards (more complex) life-cycle models, supposing that the disparity has to be attributed to the incompleteness of the individual insurance market. In line with this hypothesis, in our utility-based framework, policyholders respond to changes in parameters by balancing payouts in the case of death and survival – which of course is the source of the incompleteness of the market. However, these effects are far less pronounced relative to the motive to maximize the value of the embedded option. In contrast, we show that the primary reason for the alluded disparity is the negligence of tax effects: A value-maximizing approach that correctly accounts for tax benefits will produce very similar outcomes as the considerably more complex life cycle model. Furthermore, the difference between the two approaches might be even smaller if we additionally included a market for life contingencies such as term life insurance in our lifetime utility framework. Thus, our findings suggest that the supplemental insights provided by a utility-based framework do not justify the additional complexity relative to the risk-neutral approach – if the latter is taken out properly, that is from the policyholder's perspective.

Beyond taxation rules, this distinction of perspective can also be important for the underlying assumptions of the approach. For instance, in the present paper we assume a given set of mortality rates that are identical both from the individual and the company's perspective. In future work, it may be worthwhile to study the effects of an asymmetry in the mortality assumptions on withdrawal behavior. Here, the asymmetry may originate from informational asymmetries between the policyholder and the company during the contract phase due to individual mortality risk, or may be simply caused by the policyholder's poor understanding of his own mortality risk as suggested in the behavioral economics literature (cf. Harrison and Rutström (2006)). The latter may be particularly interesting since, to our knowledge, so far there are no attempts to quantify the financial impact of such "behavioral anomalies" on exercise-dependent retirement savings products.

Another obvious direction of future research is the generalization of our results to a more general model frameworks and a more general set of life insurance products. Specifically, while the life-cycle model considered here is rather simple, we see evidence that modifying the preference assumptions or adding additional risk factors will not affect withdrawal behavior – provided that these risks are separately insurable. However, non-insurable exogenous expenditure or liquidity shocks may well yield a difference in patterns. This is akin to Carpenter (1998), who in the context of employee stock options proposes a risk-neutral valuation model with an exogenous withdrawal state. Similarly, including additional guaranteed benefits and extending the contract period may provide further insights. For instance, it is conceivable that for other guarantees the pooling with a savings product could provide payoff profiles that impair the performance of the risk-neutral approach relative to the utility maximization framework, especially later in life.

Finally, it is important to further elaborate on the practical significance of our results. While we already highlighted that our results endorse simple reduced-form exercise rules based on the "moneyness" of the guarantee, the question arises if there are better performing reduced-form rules – both in the context of our model and in view of empirical exercise patterns.

References

Bauer, Daniel, Alexander Kling and Jochen Russ, 2008. A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities. ASTIN Bulletin 38, 621-651.

Bauer, Daniel, Daniela Bergmann and Rüdiger Kiesel, 2010. On the Risk-Neutral Valuation of Life Insurance Contracts with Numerical Methods in View. ASTIN Bulletin 40, 65-95.

Black, Kenneth, Jr. and Harold D. Skipper Jr., 2000. Life and Health Insurance, 13th Edition. Upper Saddle River, NJ: Prentice Hall.

Boyle, Phelim and Mary Hardy, 2003. Guaranteed Annuity Options. ASTIN Bulletin 33, 125-152.

Carpenter, Jennifer, 1998. The Exercise and Valuation of Executive Stock Options. Journal of Financial Economics 48, 127-158.

Chen, Zhang, and Peter A. Forsyth, 2008. A numerical scheme for the impulse control formulation for pricing variable annuities with a guaranteed minimum withdrawal benefit (GMWB). Numerische Mathematik 109, 535-569.

Chen, Zhang, Ken Vetzal and Peter A. Forsyth, 2008. The Effect of Modelling Parameters on the Value of GMWB Guarantees. Insurance: Mathematics and Economics 43, 165-173.

Dai, Min, Yue Kuen Kwok and Jianping Zong, 2008. Guaranteed Minimum Withdrawal Benefit in Variable Annuities. Mathematical Finance 18, 595-611.

De Giovanni, Domenico, 2010. Lapse Rate Modeling: a Rational Expectations Approach. Scandinavian Actuarial Journal 2010, 56-67.

Epstein, Larry G. and Stanley E. Zin, 1989. Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. Econometrica 57, 937-969.

Gao, Jin and Eric Ulm, 2011. Optimal Allocation and Consumption with Guaranteed Minimum Death Benefits with Labor Income and Term Life Insurance. Working Paper, Lingnan University and Georgia State University.

Gerber, Hans U. and Elias S.W. Shiu, 2000. Investing for Retirement: Optimal Capital Growth and Dynamic Asset Allocation. North American Actuarial Journal 4, 42-62.

Grosen, Anders and Peter Løchte Jørgensen, 2000. Fair Valuation of Life Insurance Liabilities: The Impact of Interest Rate Guarantees, Surrender Options, and Bonus Policies. Insurance: Mathematics and Economics 26, 37-57.

Harrison, Glenn W. and Rutström, E. Elisabet, 2006. Eliciting Subjective Beliefs about Mortality Risk Orderings. Environmental and Resource Economics 33, 325-346.

U.S. Department of the Treasury, Internal Revenue Service, 2003. General Rule for Pensions and Annuities (Publication 939). Available at `http://www.irs.gov/pub/irs-pdf/p939.pdf`.

Kim, Changki, 2005. Modeling Surrender and Lapse Rates with Economic Variables. North American Actuarial Journal 9, 56-70.

Knoller, Christian, Gunther Kraut and Pascal Schoenmaekers, 2011. On the Propensity to Surrender a Variable Annuity Contract. Working Paper, Ludwig Maximilian Universitaet Munich.

Kuo, Weiyu, Chenghsien Tsai and Wei-Kuang Chen, 2003. An Empirical Study on the Lapse Rate: The Cointegration Approach. Journal of Risk and Insurance 70, 489-508.

Merton, Robert C., 1969. Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. Review of Economics and Statistics 51, 247-257.

Milevsky, Moshe A. and Kamphol Panyagometh, 2001. Variable Annuities versus Mutual Funds: A Monte Carlo Analysis of the Options. Financial Services Review 10, 145-161.

Milevsky, Moshe A. and Steven E. Posner, 2001. The Titanic Option: Valuation of the Guaranteed Minimum Death Benefit in Variable Annuities and Mutual Funds. Journal of Risk and Insurance 68, 93-128.

Milevsky, Moshe A. and Thomas S. Salisbury, 2006. Financial Valuation of Guaranteed Minimum Withdrawal Benefits. Insurance: Mathematics and Economics 38, 21-38.

Nishiyama, Shinichi, and Kent Smetters, 2005. Consumption Taxes and Economic Efficiency with Idiosyncratic Wage Shocks. Journal of Political Economy 113, 1088-1115.

Ross, Stephen A., 1987. Arbitrage and Martingales with Taxation. Journal of Political Economy 95, 371-393.

Society of Actuaries, 2009. Policyholder Behavior in the Tail – Variable Annuity Guaranteed Benefits Survey/C3 Phase II, 2009 Results

Stanton, Richard, 1995. Rational Prepayment and the Valuation of Mortgage-Backed Securities. Review of Financial Studies 8, 677-708.

Steinorth, Petra and Olivia S. Mitchell, 2011. Valuing Variable Annuities with Guaranteed Minimum Lifetime Withdrawal Benefits. Working Paper, St. John's University and University of Pennsylvania.

Ulm, Eric, 2006 The Effect of the Real Option to Transfer on the Value of Guaranteed Minimum Death Benefits. Journal of Risk and Insurance 73, 43-69.

Zaglauer, Katharina and Daniel Bauer, 2008. Risk-Neutral Valuation of Participating Life Insurance Contracts in a Stochastic Interest Rate Environment. Insurance: Mathematics and Economics 43, 29-40.

Appendices

A Model Details

A.1 Timeline

Starting at the end of policy year $t \in \{1, \ldots, T-1\}$, just prior to the t-th policy anniversary date, the timeline of events is as follows:

1. The policyholder observes the annual asset returns S_t/S_{t-1}, and thus the level of his current state variables, A_t^- and X_t^-.

2. The policyholder dies with probability q_{x+t-1}.

3. In case of death, he leaves bequest $b_t = A_t^- + \max\{X_t^-\ ,\ G_t^D\}$, and no further actions are taken. If he survives:

4. The policyholder receives income I_t for the new period.

5. Based on this information, he chooses how much to withdraw from the VA account, w_t, consume, C_t, and how to allocate his outside portfolio, ν_t.

6. The policyholder consumes C_t. Account values (A_t^+, X_t^+), the benefit base ($G^{\cdot}_{t+1}$) and the tax base (H_{t+1}) are updated accordingly.

7. Between t and $t+1$, the account values evolve in line with the asset(s).

8. (1.) The policyholder observes S_{t+1}/S_t, etc.

A.2 Bellman Equation in the Black-Scholes Framework

In a Black-Scholes environment, as described in Section 3, optimization problem (6) takes the form

$$V_t^-(y_t) = \max_{C_t,w_t,\nu_t} u_C(C_t) + e^{-\beta} \int_{-\infty}^{\infty} \psi(\gamma) \left[q_{x+t} \cdot u_B(b_{t+1}|S'(\gamma)) + p_{x+t} \cdot V_{t+1}^-(y_{t+1}|S'(\gamma)) \right] d\gamma,$$

where $\psi(\gamma) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\gamma^2}{2})$ is the standard normal probability density function, and $S'(\gamma) = S_t \cdot e^{\sigma\gamma+\mu-\frac{1}{2}\sigma^2}$ is the annual gross return of the risky asset, subject to

$$\begin{aligned}
y_t &= \{A_t^-, X_t^-, G^{\cdot}_t, H_t\}, \\
X_t^+ &= \left(X_t^- - w_t\right)^+, \\
A_t^+ &= A_t^- + I_t + w_t - \text{fee}_I - \text{fee}_G - \text{taxes} - C_t, \\
\text{fee}_I &= s \cdot \max\left\{w_t - \min(g_t^W, G_t^W)\right\}, \\
\text{fee}_G &= s^g \cdot (w_t - \text{fee}_I) \cdot \mathbf{1}_{\{x+t<59.5\}}, \\
\text{taxes} &= \tau \cdot \min\{w_t - \text{fee}_I - \text{fee}_G, (X_t^- - H_t)^+\}, \\
G^{\cdot}_{t+1} &= \begin{cases} (G^{\cdot}_t - w)^+ & : w \le g_t^W \\ \left(\min\left\{G^{\cdot}_t - w \,,\, G^{\cdot}_t \cdot \frac{X_t^+}{X_t^-}\right\} \right)^+ & : w > g_t^W, \end{cases} \\
H_{t+1} &= H_t - \left(w_t - \left(X_t^- - H_t\right)^+\right)^+, \\
A_{t+1}^- &= A_t^+ \cdot \left[\nu_t \cdot e^{\sigma\gamma+\mu-\frac{1}{2}\sigma^2} + (1-\nu_t) \cdot e^r - \kappa \cdot \left(\nu_t \cdot e^{\sigma\gamma+\mu-\frac{1}{2}\sigma^2} + (1-\nu_t) \cdot e^r - 1\right)^+\right], \\
X_{t+1}^- &= X_t^+ \cdot e^{-\phi} \cdot \left[\nu^X \cdot e^{\sigma\gamma+\mu-\frac{1}{2}\sigma^2} + (1-\nu^X) \cdot e^r\right], \\
b_X &= \max\left\{X_{t+1}^-, G_{t+1}^D\right\}, \\
b_{t+1} &= A_{t+1}^- + b_X - \tau \cdot (b_X - H_t, 0), \\
0 \le C_t &\le A_t^- + I_t + w_t - \text{fee}_I - \text{fee}_G - \text{taxes}, \\
0 \le w_t &\le \max\left\{ X_t^- \,,\, \min\left\{ g_t^W \,,\, G_t^W \right\} \right\}, \quad \text{and} \\
0 \le \nu_t &\le 1.
\end{aligned}$$

A.3 Bellman Equation in the Black-Scholes Framework under Risk-Neutral Valuation

In a Black-Scholes environment, the risk-neutral valuation problem takes the form

$$V_t(X_t^-, G_t^\cdot, H_t) = \max_{w_t}(w_t - \text{fee}_I - \text{fee}_G - \text{taxes}) + X_0,$$

where $X_0 = V$ is given implicitly by

$$e^r \cdot V - \mathbb{E}^{\mathbb{Q}}[Y] - \frac{\kappa}{1-\kappa}\mathbb{E}^{\mathbb{Q}}[Y - V)^+] = 0,$$

and subject to

$$\begin{aligned}
Y &= q_{x+t} b_{t+1} + p_{x+t} V_{t+1}(X_{t+1}^-, G_{t+1}^\cdot, H_{t+1}), \\
X_t^+ &= \left(X_t^- - w_t\right)^+, \\
\text{fee}_I &= s \cdot \max\left\{w_t - \min(g_t^W, G_t^W)\right\}, \\
\text{fee}_G &= s^g \cdot (w_t - \text{fee}_I) \cdot \mathbf{1}_{\{x+t<59.5\}}, \\
\text{taxes} &= \tau \cdot \min\{w_t - \text{fee}_I - \text{fee}_G, (X_t^- - H_t)^+\}, \\
G_{t+1}^\cdot &= \begin{cases} (G_t^\cdot - w)^+ & : w \le g_t^W \\ \left(\ \min\left\{G_t^\cdot - w\ ,\ G_t^\cdot \cdot \frac{X_t^+}{X_t^-}\right\}\ \right)^+ & : w > g_t^W, \end{cases} \\
H_{t+1} &= H_t - \left(w_t - \left(X_t^- - H_t\right)^+\right)^+, \\
X_{t+1}^- &= X_t^+ \cdot e^{-\phi} \cdot \left[\nu^X \cdot e^{\sigma\gamma + r - \frac{1}{2}\sigma^2} + (1-\nu^X)\cdot e^r\right], \\
b_X &= \max\left\{X_{t+1}^-, G_{t+1}^D\right\}, \\
b_{t+1} &= b_X - \tau \cdot (b_X - H_t, 0), \quad \text{and} \\
0 \le w_t &\le \max\left\{\ X_t^-\ ,\ \min\left\{\ g_t^W\ ,\ G_t^W\ \right\}\ \right\}.
\end{aligned}$$

where γ follows a standard normal distribution.

B Derivations and Proofs

B.1 Derivation of Approximation (14)

For given M and $u_k, k = 0, \ldots, M$, we can compute $x_k = \lambda(u_k) = \exp(\sigma \cdot u_k + r - \frac{1}{2}\sigma^2)$ and $\psi_k = F(x_k)$, if necessary by interpolating and/or extrapolating $F(.)$ over the state space grid. Note that the domain of gross return variable x by definition is $(0, \infty)$.

Then, for arbitrary $0 < x < \infty$, we can approximate the corresponding function value linearly by

$$\begin{aligned}
F(x) &\approx \sum_{k=0}^{M-1} \left(\psi_k + \tfrac{x - x_k}{x_{k+1} - x_k} \cdot (\psi_{k+1} - \psi_k)\right) \cdot \mathbf{1}_{[x_k, x_{k+1})}(x) \\
&= \sum_{k=0}^{M-1} (a_k + b_k \cdot x) \cdot \mathbf{1}_{[x_k, x_{k+1})}(x),
\end{aligned}$$

where for $k = 0, \ldots, M-1$

$$a_k \equiv \frac{x_{k+1} \cdot \psi_k - x_k \cdot \psi_{k+1}}{x_{k+1} - x_k}, \quad \text{and} \quad b_k \equiv \frac{\psi_{k+1} - \psi_k}{x_{k+1} - x_k}.$$

In addition, we define $a_M = b_M \equiv 0$.

Plugging this into Equation (13), we obtain

$$\begin{aligned} K &= \int_{-\infty}^{\infty} \phi(u) F(\lambda(u))\, du \quad \approx \int_{-\infty}^{\infty} \phi(u) \cdot \sum_{k=0}^{M-1} (a_k + b_k \cdot x) \cdot \mathbf{1}_{[x_k, x_{k+1})}(x)\, du \\ &= \sum_{k=0}^{M-1} \int_{u_k}^{u_{k+1}} \phi(u) \cdot (a_k + b_k \cdot \lambda(u))\, du \quad = \sum_{k=0}^{M-1} a_k \cdot \int_{u_k}^{u_{k+1}} \phi(u)\, du + b_k \cdot \int_{u_k}^{u_{k+1}} \phi(u) \cdot \lambda(u)\, du, \end{aligned}$$

and since

$$\phi(u) \cdot \lambda(u) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}u^2\right) \cdot \exp\left(\sigma u + \mu - \frac{1}{2}\sigma^2\right) = e^{\mu} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u-\sigma)^2\right) = e^{\mu} \phi(u - \sigma),$$

we obtain

$$K \approx \sum_{k=0}^{M-1} a_k \cdot [\Phi(u_{k+1}) - \Phi(u_k)] + \exp(\mu) \cdot b_k \cdot [\Phi(u_{k+1} - \sigma) - \Phi(u_k - \sigma)].$$

Finally, reordering of the summation terms yields Equation (14). □

B.2 Proof of Proposition 5.1

Proof. All that is left to show is the existence and uniqueness of the solution to Equation (17), which can be written as

$$V_t - \mathbb{E}_t^{\mathbb{Q}}\left[\frac{B_t}{B_{t+1}} \cdot Y\right] - \frac{\kappa}{1-\kappa} \int \frac{B_t}{B_{t+1}} \cdot (Y - V_t) \cdot \mathbb{1}_{[Y > V_t]}\, dF^{\mathbb{Q}} = 0,$$

whereby $0 \leq \kappa \leq 1$, and $0 \leq B_t \leq B_{t+1}$. Denote the left hand side by $f(V_t)$. To complete the proof, we only need to demonstrate that the equation $f(V) = 0$ has exactly one solution.

Let us first demonstrate existence: Since $f(.)$ is continuous, $f(-\infty) = -\infty$ and $f(\infty) = \infty$, by the Intermediate Value Theorem, there needs to exist $-\infty < V < \infty$ such that $f(V) = 0$.

To show uniqueness, it suffices to show that $f(.)$ is strictly increasing. For that matter, consider $V^2 > V^1$. Then:

$$\begin{aligned} f(V^2) - f(V^1) &= V^2 - V^1 - \tfrac{\kappa}{1-\kappa}\left[\int \tfrac{B_t}{B_{t+1}}(Y - V^2)\mathbb{1}_{[Y > V^2]} dF^{\mathbb{Q}} - \int \tfrac{B_t}{B_{t+1}}(Y - V^1)\mathbb{1}_{[Y > V^1]} dF^{\mathbb{Q}}\right] \\ &= V^2 - V^1 + \tfrac{\kappa}{1-\kappa}\left[\int \tfrac{B_t}{B_{t+1}} \cdot \left\{(Y - V^1)\mathbb{1}_{[V^1 < Y \leq V^2]} + (V^2 - V^1)\mathbb{1}_{[Y > V^2]}\right\} dF^{\mathbb{Q}}\right] \\ &> 0. \end{aligned}$$

□