TAXATION OF A GMWB VARIABLE ANNUITY IN A STOCHASTIC INTEREST RATE MODEL

BY

ANDREA MOLENT

ABSTRACT

Modeling taxation of Variable Annuities has been frequently neglected, but accounting for it can significantly improve the explanation of the withdrawal dynamics and lead to a better modeling of the financial cost of these insurance products. The importance of including a model for taxation has first been observed by Moenig and Bauer (2016) while considering a Guaranteed Minimum Withdrawal Benefit (GMWB) Variable Annuity. In particular, they consider the simple Black–Scholes dynamics to describe the underlying security. Nevertheless, GMWB are long-term products, and thus accounting for stochastic interest rate has relevant effects on both the financial evaluation and the policyholder behavior, as observed by Goudenège *et al.* (2018). In this paper, we investigate the outcomes of these two elements together on GMWB evaluation. To this aim, we develop a numerical framework which allows one to efficiently compute the fair value of a policy. Numerical results show that accounting for both taxation and stochastic interest rate has a determinant impact on the withdrawal strategy and on the cost of GMWB contracts. In addition, it can explain why these products are so popular with people looking for a protected form of investment for retirement.

KEYWORDS

Variable Annuities, taxation, stochastic interest rate, optimal withdrawal, tree method.

JEL codes: G2, G22, G12, G13, C6.

Astin Bulletin 50(3), 1001-1035. doi:10.1017/asb.2020.29

1. Introduction

Variable Annuities are tax-deferred investment contracts with insurance coverage. The market for such products has been steadily growing in the past years all around the world, and 2019 has set best sales year since 2008 in Unites States. According to the Secure Retirement Institute (2019), the Variable Annuity sales in 2019 amounted to over $100 billions, which represents almost half of the total annuity sales. In this paper, we focus on a particular type of Variable Annuity, called Guaranteed Minimum Withdrawal Benefit (GMWB) which promises to return the entire initial investment by means of cash withdrawals during the policy life, plus a final payment amounting to the remaining account value at the contract maturity. Usually, the policyholder (hereinafter PH) pays the whole premium as a lumpsum, and he is entitled to withdraw at each contract anniversary a variable amount, with a minimum guaranteed. Thanks to the guarantee included in the policy, the PH can withdraw money from his account even if it has run out. Moreover, if the PH death occurs before the contract maturity, then his heirs receive the remaining account value as a lumpsum payout. The premium paid at contract inception determines the risky account, which changes over time according to a financial index (usually a fund) but it is also reduced due to the fees applied by the insurer and by withdrawals made by the PH.

In order to manage GMWB contracts, insurers usually employ hedging techniques which rely on the computation of the fair prices of the policies in a risk neutral probability framework. In addition, the hedging costs are offset by deducting a proportional fee from the risky asset account. Moreover, the mortality risk is hedged by using the law of large numbers (see Bernard and Kwak, 2016 and Lin *et al.*, 2016 for an explanation of move-based and semi-static hedging of Variable Annuities). Price and Greeks calculation usually relies on numerical computations, which are based on a convenient model of the product, of the financial market, and nonetheless of the behavior of the PH. In fact, since the PH can choose (within certain limits established by the contract) the amount to be withdrawn, he can decisively drive the total payoff of the contract. Anyway, ordinary techniques for pricing American and Bermudan options lead to prices which differ significantly from market observations (Moenig and Bauer, 2016). In particular, the typical prices for Variable Annuities in the marketplace are often much less than the no-arbitrage value. A possible simple explanation for this anomaly is that the price gap derives from errors in the internal models of insurance companies or it is caused by (dangerous) short-term marketing decisions, forced by market competition. Anyway, various academics have postulated more fascinating reasons. For example, Piscopo and Haberman (2011) prove that neglecting randomness of mortality rates can lead to mispricing, while Sun *et al.* (2018) attribute mispricing to the lack of a correct model for management fees. Furthermore, Kling *et al.* (2014) find that the price of the guarantee strongly depends on the considered model for the PH's behavior. In this regard, an element that affects the withdrawal strategy of the PH and that supports the theoretical-empirical price gap is the correct modeling of taxation dynamics. In particular, Moenig and Bauer (2016) propose to model taxation imposed to the PH and to consider a subjective valuation of the contract. Specifically, they show that when accounting for taxation, PH withdraws less frequently than without taxes and by employing ordinary pricing techniques, one can obtain prices which are in line with empirical observations. Moreover, Moenig and Zhu (2018) observe that the preferential tax treatment has been one of the key factors that have made Variable Annuities such a popular instrument and thus correctly modeling taxation can improve the explanation of the still unclear mechanisms about these products. We stress out that the investigations in Moenig and Bauer (2016) and Moenig and Zhu (2018) have been performed by assuming the Black–Scholes model for the underlying fund.

Interest rates are another relevant factor in Variable Annuities evaluation. As observed by Goudenège *et al.* (2018), since GMWB contracts have long maturities that could last almost 25 years, the Black–Scholes model seems to be unsuitable for such a long time interval as it assumes constant interest rate and volatility. Several authors have investigated the possibility of evaluating GMWB contracts while considering a stochastic interest rate. For example, Peng *et al.* (2012) develop an analytic approximation of the fair value of the GMWB under the Vasicek stochastic interest rate model. Donnelly *et al.* (2014) consider pricing and Greeks calculation through an Alternating Direction Implicit method in the advanced Heston–Hull–White model. Dai *et al.* (2015) develop a tree-based model to include both stochastic interest rate and mortality in their evaluation framework. Gudkov *et al.* (2019) employ the operator splitting method to price GMWB products under stochastic interest rate, volatility and mortality. Shevchenko and Luo (2017) employ high-order Gauss-Hermite quadrature to evaluate the GMWB contract under the Vasicek interest rate model. Recently, Goudenège *et al.* (2019) exploit a hybrid tree-PDE method together with Machine Learning techniques to efficiently evaluate the GMWB contract in a model that considers both stochastic interest rate and stochastic volatility. More generally, as far as pricing of Variable Annuities in a stochastic interest rate framework is considered, it is worth mentioning the work of Bacinello and Zoccolan (2019) that develops a Monte Carlo flexible approach to study the impact of threshold fee on the optimal surrender strategy about a product including accumulation and death guaranteed benefits under a model which considers stochastic interest rate, volatility and mortality. We also mention Goudenège *et al.* (2016), who employ the hybrid tree-PDE method to evaluate a GLWB contract under stochastic interest rate.

In this paper, we present an investigation about GMWB pricing and PH behavior when both tax treatment and stochastic interest rate are considered. In particular, following Moenig and Bauer (2016) and Moenig and Zhu (2018), we model taxation of GMWB through a constant marginal income tax rate on all policy earnings and a constant marginal tax rate on capital gains from investments outside of the policy. Moreover, we also include a premium-based model for taxation of the insurer, which was neglected in previous researches. Because of taxation, the evaluation of the contract is not straightforward, so we exploit the same subjective risk-neutral valuation methodology employed in Moenig and Bauer (2016). In particular, in this framework, the value of a given post-tax cash flow is the amount necessary to set up a pre-tax portfolio that replicates the considered cash flow. This causes the insurer and the PH to evaluate the policy differently, and we investigate both the perspectives. As far as the stochastic interest rate is concerned, we consider the Hull–White model (Hull and White, 1994), which is often employed by both academics and practitioners for its easiness of calibration and simple probability distribution. This model has already been employed in other research works concerning GMWB Variable Annuities (e.g. Donnelly *et al.*, 2014; Dai *et al.*, 2015; Goudenège *et al.*, 2018, 2019). We stress out that considering both taxation and stochastic interest rate is a challenging task because of the computational effort required to consider many factors together. In particular, evaluating a GMWB policy in the considered model is a four (plus time)-dimensional problem, which means a high computational cost in terms of both computing time and working memory required. Moreover, the evaluation of a policy through the subjective risk-neutral valuation methodology requires the resolution of many fixed point problems, and this increases even more the computational cost. Finally, we assume the PH to employ an optimal withdrawal strategy, which implies the numerical resolution of a dynamic control problem. In order to manage such a computational effort, we use a backward dynamic approach that exploits a tree approach to compute the fair contract price. In particular, we employ a trinomial tree to approximate the stochastic interest rate process through a Markov chain, which represents an efficient numerical solution already used by Goudenège *et al.* (2019). It is worth noting that tree methods have already been used to study the GMWB contract. In this regard, we mention the works of Costabile (2017) and of Costabile *et al.* (2020) that employ a trinomial tree to evaluate a GMWB policy and to investigate the PH decisions while including exogenous factors in the model.

In order to test our approach, we perform some numerical experiments. Specifically, we study how the evaluation of the policy varies according to the insurer and to the PH perspectives and how the withdrawal strategy is modified, by including or not including taxation and by changing the parameters of the interest rate and the fund. Numerical results show many interesting findings. First of all, if taxation is considered, the fair value of the policy for the PH is higher than the fair value for the insurer. This means that the PH attributes a higher price to the policy than the insurer does, so buying and selling the contract can be a good deal for both of them. Secondly, we observe that taxation and interest rate modeling have a significant impact on the withdrawal strategies of the PH.

To the best of our knowledge, this is the first analysis about GMWB pricing and withdrawal strategy which accounts for both taxation and stochastic interest rate. Our research could be useful both for the qualitative observations obtained and for the numerical solutions adopted.

The reminder of the paper is organized as follows. Section 2 introduces the stochastic model for the underlying and the interest rate processes. Section 3 describes the GMWB contract and the taxation model. Section 4 presents pricing assumptions. Section 5 describes the pricing method and the technical measures. Section 6 shows numerical results on various examples. Finally, Section 7 draws the conclusions.

2. The stochastic model

In order to define the notation used throughout the rest of the paper, let us introduce the Black–Scholes Hull–White model. The Hull–White model (Hull and White, 1994) is one of the historically most important interest rate models, which is nowadays often used for option pricing purposes. In particular, the existence of closed formulas for the price of bonds, caplets and swaptions is one of the important advantages of this model. Furthermore, it is capable of generating negative interest rates, actually observed in the markets in recent years. We report the dynamics of the Black–Scholes Hull–White model, which combines the dynamics of the interest rate with the dynamics of the underlying:

$$\begin{cases} dS_t = r_t S_t dt + \sigma S_t dZ_t^S \\ dr_t = k\left(\theta\left(t\right) - r_t\right) dt + \omega dZ_t^r, \end{cases} \tag{2.1}$$

where Z^S and Z^r are Brownian motions with $d\left\langle Z_t^S, Z_t^r\right\rangle = \rho dt$. Moreover σ, k and ω are positive values, and the initial values $S_0 > 0$ and r_0 are given. Furthermore, $\theta\left(t\right)$ is a deterministic function which is completely determined by the market values of the zero-coupon bonds by calibration (see Brigo and Mercurio, 2007) so that the theoretical prices of the zero-coupon bonds match exactly the market prices.

It is well known that the (short) interest rate process r can be written as

$$r_t = Y_t + \beta\left(t\right), \tag{2.2}$$

where Y is a stochastic process whose dynamics is given by

$$dY_t = -kY_t dt + \omega dZ_t^r, \;\; Y_0 = 0, \tag{2.3}$$

and $\beta\left(t\right)$ is a real-valued function given by

$$\beta\left(t\right) = r_0 e^{-kt} + k\int_0^t \theta\left(s\right) e^{-k(t-s)} ds. \tag{2.4}$$

Moreover, $\beta\left(t\right)$ can be estimated directly from market data as

$$\beta\left(t\right) = -\frac{\partial \ln P^M\left(0, t\right)}{\partial t} + \frac{\omega^2}{2k^2}\left(1 - \exp\left(-kt\right)\right)^2, \tag{2.5}$$

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

where $P^M(0, T)$ stands for the market price of the zero-coupon bonds at time 0 for the maturity t.

The Black–Scholes Hull–White model can be described by the following relations:

$$\begin{cases} dS_t = r_t S_t dt + \sigma S_t dZ_t^S & S_0 = \bar{S}_0, \\ dY_t = -kY_t dt + \omega dZ_t^r & Y_0 = 0, \\ r_t = Y_t + \beta(t). \end{cases} \tag{2.6}$$

The *flat curve* case is a particular case for the market price of a zero-coupon bonds: in this specific case, the price at time t of a zero-coupon bond with maturity $\bar{t}$ is given by

$$P^M\left(t, \bar{t}\right) = e^{-r_0\left(\bar{t}-t\right)}, \tag{2.7}$$

and the function β is given by

$$\beta(t) = r_0 + \frac{\omega^2}{2k^2}\left(1 - \exp(-kt)\right)^2. \tag{2.8}$$

We stress out that assuming a flat curve for the price of bonds is not essential for the development of our model, but it simplifies the numerical settings.

3. Modeling the contract

3.1. Modeling taxation

In order to model taxation, we follow the same approach proposed by Moenig and Bauer (2016), which in turn is a simplified version of the model currently in force in the Unites States.

Variable Annuities are tax-deferred products, which means the PH does not pay federal taxes on the income and on the investment gains from the annuity until withdrawals are made. Moreover, taxes are due on future policy gains and not on the invested amount. In this regard, we assume a constant marginal income tax rate τ to be applied on all policy earnings. Moreover, we assume that earnings from the policy are withdrawn before the initial premium, following a last-in first-out approach. On the contrary, capital gains form investments held by the PH outside the policy are taxed annually at a constant marginal tax rate κ. Thus, if the PH sets up a multi-asset portfolio made of bonds and stocks, then κ is the tax rate applied on portfolio gains.

In order to complete tax modeling, we have to consider taxation concerning the insurer, which is usually of two types: premium taxation and net income taxation (see Skipper, 2001). Determining life insurer profit is a challenge because of the difference in timing between premium payments and claim payments, so premium taxes are the most common. Furthermore, as far as

Unites States life insurance system is concerned, the insurance companies can elect to be taxed based on either premiums or net income (see Nissim, 2010). For sake of simplicity, we assume premium-based taxation, that is the insurer pays a certain percentage of the gross premium GP as taxes. So, the tax due by the insurer is thus $\chi \cdot GP$, where χ is the premium tax rate. Such a rate usually varies between 0.5% and 3% (see Moran, 2017). Obviously, the insurer has to recover this tax cost; therefore, we assume that such an amount is applied indirectly to the customer as an entry cost, which reduces the gross premium and determines the net premium P, given by $P = GP \cdot (1 - \chi)$.

3.2. The GMWB contract

We study here a simple version of the GMWB contract which was first investigated by Moenig and Bauer (2016). We consider an x-year-old individual that purchased a GMWB policy with a finite integer maturity T against the payment of a single gross premium GP. Then, entry expenses are deducted from the gross premium, and the net premium P is credited to the policy's account. There are three variables which determine the state of a policy at time t, namely the account value X_t, the benefit base G_t and the tax base H_t whose values at time $t = 0$ are equal to the policy net premium, that is

$$X_0 = G_0 = H_0 = P. \tag{3.1}$$

In particular, the account value X represents the risky account of the policy, which changes as if it were invested in a market fund, aside from being reduced by withdrawals and management costs. The benefit base G represents the guarantee inherent in the policy as it regulates the maximum withdrawal that the PH can make, while the tax base H represents the amount that may still be withdrawn from the policy free of tax.

Let t_i denote the time of the i-th contract anniversary, that is $t_i = i$. The variables G_t and H_t do not change during the time between two consecutive anniversaries, that is for $t \in]t_{i-1}, t_i[$, while X_t varies according to an underlying investment fund changes. This fund is usually chosen by the customer from a list proposed by the insurer. Specifically, let us term S_t the value of the underlying fund, which evolves according to (2.1). Then, for $t \in]t_{i-1}, t_i[$, X_t follows the same dynamics of S_t with the exception that fees are subtracted continuously, that is

$$dX_t = \frac{X_t}{S_t} dS_t - \varphi X_t dt. \tag{3.2}$$

The variable φ in (3.2) is the (constant) fee rate and it controls the fees withdrawn by the account value.

At each anniversary time t_i, the continuation of the policy is determined according to the survival of the PH during the last year of the contract. In order to describe the policy revaluation mechanisms, let us denote with $X_{t_i}^-$ and $X_{t_i}^+$

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

the account values just before and after any cash flow at time t_i (we use the same notation for G_{t_i} and H_{t_i}). If the PH has passed away during the previous year, then his heirs receive the death benefit b_i, which is paid at time t_i and it is given by the residual account value net of taxation, that is

$$b_i = X_{t_i}^- - \tau \left(X_{t_i}^- - H_{t_i}^- \right)_+, \tag{3.3}$$

where τ is the income tax rate and $(x)_+ = \max(x, 0)$. After the payment of the death benefit, the contract ends and it has no residual value. On the contrary, if the PH has not passed away, then he is entitled to withdraw an amount w_i within some limits. According to the contract, the withdrawal amount w_i selected by the PH must satisfy the following relation:

$$0 \leq w_i \leq \max\left\{X_{t_i}^-, \min\left\{g^W, G_{t_i}^-\right\}\right\}, \tag{3.4}$$

where g^W is a positive constant value called the annual guaranteed amount and it is stated in the contract. In particular, if $g^W = {}^P/_T$, then the PH is entitled to withdraw at each contract anniversary exactly an amount equal to g^W throughout the duration of the contract. We stress out that the alive PH is entitled to perform withdrawals at times $t_i = 1, \ldots, T$, for a total of T events, thus a minimum amount is guaranteed from the first anniversary t_1 up to maturity T. After the withdrawal has been performed, the new account value is given by

$$X_{t_i}^+ = \left(X_{t_i}^- - w_i \right)_+, \tag{3.5}$$

while the new benefit base and tax base are given by

$$G_{t_{i+1}}^- = G_{t_i}^+ = \begin{cases} \left(G_{t_i}^- - w_i \right)_+, & \text{if } w_i \leq g^W \\ \left(\min\left\{ G_{t_i}^- - w_i, G_{t_i}^- \cdot \frac{X_{t_i}^+}{X_{t_i}^-} \right\} \right)_+, & \text{if } w_i > g^W \end{cases} \tag{3.6}$$

and

$$H_{t_{i+1}}^- = H_{t_i}^+ = H_{t_i}^- - \left(w_i - \left(X_{t_i}^- - H_{t_i}^- \right)_+ \right)_+ \tag{3.7}$$

respectively.

The PH does not receive the whole amount withdrawn w_i because some fees and tax may be applied. Specifically, the PH receives the withdrawn amount reduced by the fees due to the insurer for withdrawing more than the guaranteed amount g^W and also reduced by a penalty for early withdrawals and by the taxation. Specifically, the net amount he receives is given by

$$w_i - fee_i - pen_i - tax_i \tag{3.8}$$

being fee_i the cost for withdrawing an amount exceeding $\min\left\{g^W, G_{t_i}\right\}$, pen_i an early withdrawal penalty for any withdrawal before the age of 59.5 years and tax_i the income taxes associated with the withdrawal. In particular,

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

$$fee_i = s_i \cdot \left(w_i - \min\left\{g^W, G_{t_i}^-\right\}\right)_+, \tag{3.9}$$

$$pen_i = s^g \cdot (w_t - fee_i) \cdot 1_{\{x+t_i<59.5\}}, \tag{3.10}$$

and

$$tax_i = \tau \cdot \min\left\{w_i - fee_i - pen_i, \left(X_{t_i}^- - H_{t_i}^-\right)_+\right\}. \tag{3.11}$$

The coefficient s_i in (3.9) is a non-negative coefficient called surrender charge, which usually decreases with time and it is zero within the term of the contract. Moreover, s^g in (3.10) is another non-negative coefficient that determines the penalty for an early withdrawal. In particular, since these contracts are usually employed as a supplement to the retirement pension, we assume that when the contract maturity is achieved, the PH must be older than 59.5 years, so penalty is not applied at last withdrawal at time T.

Finally, after the last withdrawal has been made at time T, the alive PH receives the remaining account value net of taxes, that is

$$X_T^+ - \tau\left(X_T^+ - H_T^+\right)_+, \tag{3.12}$$

and the contract ends.

4. Pricing assumptions

In this Section, we present the pricing framework.

First of all, we present the main pricing tool, that is the subjective risk-neutral valuation. In a nutshell, such an approach defines the value of the policy as the amount of money that an agent requires to set up a pre-tax financial portfolio such that, after taxation, it replicates the post-tax policy cash flows.

After introducing the subjective risk-neutral valuation, we focus on the appraisal of the GMWB contract by considering PH's subjective valuation and then we present the same while assuming insurer's subjective valuation. The main differences in the two perspectives are due to taxation and to control on withdrawals. As far as taxation is considered, the PH has to pay taxes on both policy earnings and capital gains outside the policy. On the contrary, the taxation applied to the insurer is much simpler: a percentage of the gross premium. As far as withdrawals are concerned, the PH selects optimal withdrawals in order to maximize the expected value of its assets, net of taxation: if taxation is applied, such a value is not equal to insurer's liability. Thus, the amount withdrawn by the PH is optimal for him, but it could be different from the worst amount computed considering the insurer's point of view, that is the amount that maximizes insurer's liability to the PH. This means that the PH withdraws money trying to maximize his economic return, rather than trying to maximize the outputs of the insurer: since these two strategies do not coincide, the costs for the insurer are lower than the worst withdrawal case.

Finally, we underline that the considered framework captures an interesting feature of insurance products. Taxation makes GMWB policies particularly attractive to customers: although taxes are applied on the earnings of the policy, the tax regime is particularly favorable for this type of product and therefore it is more convenient for the customer buying the policy rather than reproducing it through a replicating portfolio.

In the next Subsections, we show how to compute the initial contract value according to the PH and to the insurer's subjective valuation. We stress out that in both cases we compute the cost of the replicating portfolio under the same risk neutral measure $\mathbb{Q}$ for the Black–Scholes Hull–White model (see Brigo and Mercurio, 2007).

4.1. Subjective risk-neutral valuation

The subjective risk-neutral valuation was introduced in the context of Variable Annuities by Moenig and Bauer (2016) by drawing inspiration from the approach of Sibley (2002) originally employed for pricing insurance products with deterministic cash flows. Such an approach solves the problem of defining a pricing framework when accounting for taxation. In fact, when taxation is applied, the well-known standard risk-neutral valuation is not suitable: as observed by Ross (1987), taxation leads to the loss of uniqueness of prices of contingent claims since the valuation of a specific cash flow depends on the personal endowment and on the tax rates applied to the agent that owns the claim.

While assessing his financial position, each agent (in our case the PH or the insurer) must consider the taxation applied to the various instruments he owns, by distinguishing between the pre-tax and post-tax amounts. Obviously, these two amounts cannot be directly compared with each other since the former ones, unlike the latter ones, still have to discount the taxation before they are actually available for consumption.

The key idea of the subjective risk-neutral valuation is that in a complete pre-tax market, any agent can replicate all post-tax cash flows from the policy by investing a pre-tax amount in a replicating portfolio including securities such as shares of the underlying fund, a bank account for cash money and other interest rate products. It is important to note that all of these financial instruments must discount capital gain taxes. According to the subjective risk-neutral valuation, the amount required to set up such a replicating portfolio defines the contract value. Clearly, such a value depends on the specific taxation applied to the agent and this can be significantly different with respect to the PH and to the insurer: that's why the valuation is termed subjective. Furthermore, taxation of policy cash flows may be different from taxation of the replicating portfolio: a lighter taxation is usually applied to insurance products (such as Variable Annuities policies) and a heavier taxation is imposed to financial products (such as the securities in the replicating portfolio).

Finally, we point out that, when a positive tax rate κ is applied on gains of the replicating portfolio, the subjective risk-neutral value of the policy must be computed by solving a non-linear equation which involves the post-tax value of the policy together with the risk-neutral expected post-tax cash flow. On the contrary, if no taxation is applied to the replicating portfolio, then a linear equation is obtained, which can be solved by computing the expected discounted value under the risk-neutral probability measure as in the usual case without taxation.

4.2. Policyholder's subjective valuation

We focus now on the PH's subjective valuation of the contract, that is we compute the amount of money that a PH needs to set a hypothetical replicating portfolio, which replicates the post-tax policy cash flows. Specifically, let $\mathcal{V}(t, r, X, G, H)$ denote the fair value according to an alive PH of a GMWB contract at time t, being r the interest rate, X the account value, G the guarantee base and H the tax base. Specifically, following the same approach of Moenig and Bauer (2016), $\mathcal{V}$ represents the average option value across the many policies sold to the customers that are still alive at time t.

Finally, in order to compute PH's subjective value of the contract at time $t = 0$, we proceed backward in time, starting from contract's maturity at time T and by taking into account the changes that occur to the policy status parameters.

4.2.1. *Value function at a contract anniversary.*
First of all, let us denote with $\mathcal{V}^+\left(T, r_T, X_T^+, G_T^+, H_T^+\right)$ the policy value at maturity, after the last withdrawal is performed. Such an amount is given by the final payoff, that is

$$\mathcal{V}^+\left(T, r_T, X_T^+, G_T^+, H_T^+\right) = X_T^+ - \tau\left(X_T^+ - H_T^+\right)_+. \tag{4.1}$$

Now, let us focus on the i-th contract anniversary, at time t_i. Since we are assuming that the PH is alive, then he is entitled to perform a withdrawal from his account. Let $\mathcal{V}^-\left(t_i, r_{t_i}, X_{t_i}^-, G_{t_i}^-, H_{t_i}^-\right)$ and $\mathcal{V}^+\left(t_i, r_{t_i}, X_{t_i}^+, G_{t_i}^+, H_{t_i}^+\right)$ represent the values of the policy just before and after the PH has withdraw money, respectively. In particular, $r_{t_i}, X_{t_i}^-, G_{t_i}^-, H_{t_i}^-$ are the state parameters before withdrawing at time t_i, while $r_{t_i}, X_{t_i}^+, G_{t_i}^+, H_{t_i}^+$ are the state parameters after withdrawing at time t_i. Please, observe that there is no need to distinguish between the value of the interest rate before and after the withdrawal because such a value is not modified by the withdrawal, so we simply write r_{t_i} in both cases. We can write the relation between the two policy values in the general form

$$\begin{aligned}\mathcal{V}^-\left(t_i, r_{t_i}, X_{t_i}^-, G_{t_i}^-, H_{t_i}^-\right) = {} & \mathcal{V}^+\left(t_i, r_{t_i}, X_{t_i}^+(w_i), G_{t_i}^+(w_i), H_{t_i}^+(w_i)\right) \\ & + \left(w_i - fee_i(w_i) - pen_i(w_i) - tax_i(w_i)\right),\end{aligned} \tag{4.2}$$

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

where we underline the dependence of many variables on the withdrawal w_i by denoting them as a function of w_i. In particular, Equations (3.5), (3.6), (3.7), (3.9), (3.10) and (3.11) express the dependence of $X_{t_i}^+, G_{t_i}^+, H_{t_i}^+, fee_i, pen_i$ and tax_i on w_i, respectively.

The PH might adopt a static withdrawal strategy, which means he withdraws an amount w_i equal to g^W, regardless of the value taken from the policy state parameters. Such a strategy is easy to be implemented, and it may be interesting in practice in the sense that retirees often desire stable (real) cash flows to fund expenses. However, in the case of an investor interested in maximizing his financial return, a strategy based on constant withdrawals may not be the best one. Moreover, for such investors, a dynamic strategy is not only desirable but also possible, thanks to the indications of the financial advisors which could direct their withdrawal strategies as already pointed out by Kling *et al.* (2014) and by Moenig and Bauer (2016). Thus, since we are interested in investigating optimal withdrawal strategies, we consider a value-maximizing approach. Specifically, we assume that the PH selects the amount w_i in order to maximize the expected value of his assets – contract plus net withdrawal–, that is

$$\begin{aligned} w_i = \operatorname*{argmax}_{w \in [0, W_{\max}]} & \mathcal{V}^+\left(t_i, r_{t_i}, X_{t_i}^+(w), G_{t_i}^+(w), H_{t_i}^+(w)\right) \\ & + (w - fee_i(w) - pen_i(w) - tax_i(w)), \end{aligned} \tag{4.3}$$

where

$$W_{\max} = \max\left\{X_{t_i}^-, \min\left\{g^W, G_{t_i}^-\right\}\right\} \tag{4.4}$$

is the maximum withdrawal allowed by the contract. We observe that, at maturity, the optimization problem (4.3) can be easily solved as the continuation value after the payment is given by the final payoff, which has a closed formulation. Specifically, one can prove that the optimal withdrawal in this particular case is given by

$$w_T = \min\left\{g^W, G_T^-\right\}. \tag{4.5}$$

Moreover, by using Equations (4.1), (4.2) and (4.5), one can obtain the following expression:

$$\begin{aligned} \mathcal{V}^-\left(T, r_T, X_T^-, G_T^-, H_T^-\right) = & \max\left\{X_T^-, w_T\right\} - \tau \min\left\{w_T, \left(X_T^- - H_T^-\right)_+\right\} \\ & - \tau\left(\left(w_T - \left(X_T^- - H_T^-\right)_+\right)_+ + \left(X_T^- - w_T\right)_+ - H_T^-\right)_+. \end{aligned} \tag{4.6}$$

When the optimal withdrawal w_i at time $t_i < T$ is concerned, there is no closed formula as for the last anniversary T, so w_i must be approximated by a numerical procedure.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

Remark. *The model that we have considered here assumes the PH to determine the amount to be withdrawn in order to maximize the expected value of his assets. A more complicated approach would be to specify a utility model to determine the withdrawal strategy. Such an approach has been investigated by Moenig (2012) (in the Black–Scholes model) and it requires the resolution of a life cycle utility optimization problem that incorporates the relevant decision variables. Anyway, as observed by Campbell (2006) and by Moenig and Bauer (2016), including all important elements and risk factors within a life cycle model is a demanding task, so that formulating a model for a real-world GMWB contract may not be possible. Furthermore, as observed by Moenig (2012), it appears that a life cycle model does not bring significant improvements in modeling the behavior of the PH compared to the use of taxation alone and it gives similar results as far as withdrawal strategies and pricing results are concerned.*

4.2.2. *Dynamics of the value function between two anniversaries.*
During the time between two contract anniversaries t_i and t_{i+1}, the variables G and H do not change. Changes of the policy value are solely due to the passage of time and to the changes of the account value X and of the interest rate r. Following Moenig and Bauer (2016), the subjective risk-neutral value at time t_i of $\mathcal{V}^+$ is given via a nonlinear implicit equation:

$$\mathcal{V}^+ = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_{t_i}^{t_{i+1}} r_s ds}\left(q_{x+t_i} b_{i+1} + p_{x+t_i}\mathcal{V}^-\right)\right] + \frac{\kappa}{1-\kappa}\cdot\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_{t_i}^{t_{i+1}} r_s ds}\left(q_{x+t_i} b_{i+1} + p_{x+t_i}\mathcal{V}^- - \mathcal{V}^+\right)_+\right], \tag{4.7}$$

where $\mathcal{V}^+$ stands for $\mathcal{V}^+(t_i, r_{t_i}, X^+_{t_i}, G^+_{t_i}, H^+_{t_i})$ and $\mathcal{V}^-$ stands for $\mathcal{V}^-(t_{i+1}, r_{t_{i+1}}, X^-_{t_{i+1}}, G^-_{t_{i+1}}, H^-_{t_{i+1}})$. Furthermore, b_{i+1} is the death benefit that may be paid at time t_{i+1} in case of death and it is computed according to (3.3). Moreover, q_{x+t_i} is the probability that the alive PH, aged exactly $x+t_i$ at time t_i, will die in 1 year, while $p_{x+t_{i+1}}$ is the probability that he will survive at least one more year. We stress out that the use of death and survival probabilities is possible if a large number of contract holders is assumed: in this case, mortality risk is diversifiable.

4.3. Insurer's subjective valuation

Let $\mathcal{U}(t, r, X, G, H)$ denote the fair value of the GMWB contract but according to insurer's subjective value, that is the amount of money that the insurer needs to set a replicating portfolio. The valuation according to the insurer differs from the valuation according to the PH for some reasons. First of all, the taxation applied to the insurer only concerns the initial gross premium and it is not applied to the replicating portfolio. Secondly, the insurer must shell out an amount gross of taxes, while the PH receives the net amount. Finally, the

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

insurer has no decision-making power and undergoes the PH's choices regarding the amount to be withdrawn. Just as done for the PH's subjective valuation, in order to compute insurer's subjective value at contract inception, we proceed backward in time.

4.3.1. *Value function at a contract anniversary.*
Let $\mathcal{U}^+\left(T, r_T, X_T^+, G_T^+, H_T^+\right)$ be the policy value at maturity according to the insurer, after the last withdrawal is performed. Such an amount is given by the final payoff before tax, that is the residual account value:

$$\mathcal{U}^+\left(T, r_T, X_T^+, G_T^+, H_T^+\right) = X_T^+. \tag{4.8}$$

Moreover, since the optimal withdrawal w_T at time T is given by (4.5), one can prove the following relation:

$$\mathcal{U}^-\left(T, r_T, X_T^-, G_T^-, H_T^-\right) = \max\left\{X_T^-, \min\left\{g^W, G_T^-\right\}\right\}. \tag{4.9}$$

Now, let us focus on the i-th contract anniversary at time t_i. The functions $\mathcal{U}^-\left(t_i, r_{t_i}, X_{t_i}^-, G_{t_i}^-, H_{t_i}^-\right)$ and $\mathcal{U}^+\left(t_i, r_{t_i}, X_{t_i}^+, G_{t_i}^+, H_{t_i}^+\right)$ represent the value of the contract just before and after the PH has withdrawn the amount w_i, which is the solution of problem (4.3). The following relation holds,

$$\begin{aligned}\mathcal{U}^-\left(t_i, r_{t_i}, X_{t_i}^-, G_{t_i}^-, H_{t_i}^-\right) = &\,\mathcal{U}^+\left(t_i, r_{t_i}, X_{t_i}^+(w_i), G_{t_i}^+(w_i), H_{t_i}^+(w_i)\right)\\ &+ \left(w_i - fee_i(w_i) - pen_i(w_i)\right).\end{aligned} \tag{4.10}$$

Equation (4.10) is similar to Equation (4.2) but taxes are not subtracted because the insurer has to pay the amount before taxation.

Remark. *In Subsection 4.2.1, we have assumed the PH to withdraw money in order to maximize his total wealth according to his subjective valuation, that is according to Equation (4.3) which defines the withdrawal strategy. Another interesting withdrawal strategy is the so-called worst-case strategy, which assumes the PH to make the worst withdrawal according to insurer's subjective valuation. The term "worst" is to be understood as "the most expensive", that is, the one that makes the replication of the contract the most money demanding according to the insurer's subjective valuation. Please, observe that the PH has no incentive to act in such a way as to maximize the worst-case hedging cost for the insurance company, nonetheless his withdrawal strategy may be different from the expected one (irrational customer) and therefore could be worse than expected for the insurer. Therefore, although the worst-case strategy should not be realized in practice, it is important to assess whether the insurer has the economic coverage necessary to cope with any possible withdrawal strategy. In order to deal with this state, it is sufficient to replace Equation (4.3) with the following one:*

$$w_i = \underset{w\in[0, W_{\max}]}{argmax}\, \mathcal{U}^+\left(t_i, r_{t_i}, X_{t_i}^+(w), G_{t_i}^+(w), H_{t_i}^+(w)\right) + \left(w - fee_i(w) - pen_i(w)\right). \tag{4.11}$$

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

4.3.2. *Dynamics of the value function between two anniversaries.*
As opposed to the PH, the insurer pays no taxes on the replicating portfolio. The subjective risk-neutral value at time t_i^+ of $\mathcal{U}$ is given by the discounted expected future value of the death benefit plus the value of the policy, that is

$$\mathcal{U}^+ = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_{t_i}^{t_{i+1}} r_s ds}\left(q_{x+t_i} b_{i+1} + p_{x+t_i}\mathcal{U}^-\right)\right], \tag{4.12}$$

where $\mathcal{U}^+$ stands for $\mathcal{U}^+\left(t_i, r_{t_i}, X_{t_i}^+, G_{t_i}^+, H_{t_i}^+\right)$ and $\mathcal{U}^-$ stands for $\mathcal{U}^-(t_{i+1}, r_{t_{i+1}}, X_{t_{i+1}}^-, G_{t_{i+1}}^-, H_{t_{i+1}}^-)$.

5. Pricing method

The fair value of the GMWB contract at time $t = 0$ according to the PH's subjective perspective, denoted by $\mathcal{V}(0, r_0, P, P, P)$, can be computed by moving backward in time. The terminal condition is expressed by (4.1). In order to proceed backward, we have to solve the nonlinear implicit Equation (4.7) in $]t_i, t_{i+1}[$ for $t_i = T-1, \ldots, 0$, and apply relations (4.2) and (4.3) to handle the jumps due to withdrawals at each contract anniversary.

With a similar approach, the initial fair value of the contract according to the insurer's perspective, denoted by $\mathcal{U}(0, r_0, P, P, P)$, can be computed by starting from the terminal condition (4.8), by solving backward Equation (4.12) and by applying relation (4.10). We observe that computing $\mathcal{U}(0, r_0, P, P, P)$ requires the knowledge of the optimal withdrawals, which can be achieved through the parallel computation of $\mathcal{V}(0, r_0, P, P, P)$.

We stress out that the evaluation problems of $\mathcal{V}$ and $\mathcal{U}$ are four-dimensional problems (plus the time variable) and this represents a non-trivial challenge which requires an efficient numerical method to be solved.

5.1. Problem discretisation

The variables that determine the state of the policy at any time are the G, H, X and r. To tackle the problem numerically, we prefer to replace r with Y, since the dynamics of Y is simpler and one can easily compute r from Y through (2.2). We consider a set of discrete values $\mathcal{G}_Y$ for Y, $\mathcal{G}_X$ for X, $\mathcal{G}_G$ for G and $\mathcal{G}_H$ for H, and we define a four-dimensional grid $\mathcal{G} = \mathcal{G}_Y \times \mathcal{G}_X \times \mathcal{G}_G \times \mathcal{G}_H$.

First of all, since the benefit base G and the tax base H are non-negative values that do not exceed P, it is worth exploiting an uniform partition of the interval $[0, P]$ to define $\mathcal{G}_G$ and $\mathcal{G}_H$. In particular, we set

$$\mathcal{G}_G = \left\{g_j = \frac{j}{N_G}P, j = 0, \ldots, N_G\right\} \tag{5.1}$$

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

and

$$\mathcal{G}_H = \left\{ h_j = \frac{j}{N_H} P, j = 0, \ldots, N_H \right\}, \tag{5.2}$$

where N_G and N_H are two positive integers.

As opposed to G and H, the account value X assumes non-negative unbounded values. Anyway, because of withdrawals and fees applied by the insurer, such a value should not grow too much during the life of the policy. In fact, as observed by MacKay *et al.* (2017) and by Bacinello and Zoccolan (2019) in a similar context, when the account value is very high there is a great incentive for the PH to surrender the contract by withdrawing all the money. So, following same principle of the spatial grid employed by Haentjens and In't Hout (2012), we consider $\mathcal{G}_X$ as a non-uniform distribution of points which is more dense where the process X is more likely to be. Specifically, we consider two sets of points: the first set

$$\mathcal{G}_{X_1} = \left\{ x_j^1 = 2.5 \cdot \frac{j}{N_{X_1}} P, j = 0, \ldots, N_{X_1} \right\} \tag{5.3}$$

is made of $N_{X_1} + 1$ uniformly distributed points between 0 and $2.5 \cdot P$ and the second one

$$\mathcal{G}_{X_2} = \left\{ x_j^2 = 2.5 \cdot P \cdot \exp\left((\ln(30) - \ln(2.5)) \frac{j}{N_{X_2}} \right),\ j = 1, \ldots, N_{X_2} \right\} \tag{5.4}$$

are made of N_{X_2} points which are distributed uniformly in log between $2.5 \cdot P$ and $30 \cdot P$. Then, $\mathcal{G}_X = \mathcal{G}_{X_1} \cup \mathcal{G}_{X_2}$ and we term x_j the j-th point of $\mathcal{G}_X$. Moreover, for seek of simplicity, we consider $N_{X_1} = N_{X_2}$ and we term N_X the number of elements of $\mathcal{G}_X$. We stress out that the coefficients 2.5 and 30 are determined empirically in order to give accurate results and their small variations do not produce impacts on the numerical results.

Finally, the construction of the set $\mathcal{G}_Y$ relies on the trinomial tree proposed by Goudenège *et al.* in (2019). Such a tree defines a discrete Markov chain $\bar{Y}^{\Delta t}$ that matched the first two moments of the process Y. We set

$$\mathcal{G}_Y = \left\{ y_j = \frac{3}{2} (j - N_Y) \sigma_Y^{\Delta t}, j = 0, \ldots 2N_Y \right\}, \tag{5.5}$$

where $\sigma_Y^{\Delta t}$ is a positive coefficient that depends on the standard deviation of the process Y and N_Y is a suitable integer value, thus $\mathcal{G}_Y$ is made of $2N_Y + 1$ points uniformly distributed in $\left[-\frac{3}{2} N_Y \sigma_Y^{\Delta t}; \frac{3}{2} N_Y \sigma_Y^{\Delta t}\right]$. Appendix A presents technical details about the process $\bar{Y}^{\Delta t}$, the coefficient $\sigma_Y^{\Delta t}$ and the integer N_Y.

5.2. Backward evaluation of $\mathcal{V}$

Once the grid $\mathcal{G}$ has been build, we can start the computation of the numerical approximation of $\mathcal{V}$ defined on $\mathcal{G}$ at any time t_i. In particular, for every policy anniversary t_i, we compute a function $\bar{\mathcal{V}}_i^+ : \mathcal{G} \to \mathbb{R}$ such that for any point (y, x, g, h) of $\mathcal{G}$, $\bar{\mathcal{V}}_i^+ (y, x, g, h)$ approximates $\mathcal{V}^+ (t_i, y + \beta (t_i), x, g, h)$. Moreover, we also compute a function $\bar{\mathcal{V}}_i^- : \mathcal{G} \to \mathbb{R}$ such that for any point (y, x, g, h) of $\mathcal{G}$, $\bar{\mathcal{V}}_i^- (y, x, g, h)$ approximates $\mathcal{V}^- (t_i, y + \beta (t_i), x, g, h)$. According to (4.6), the terminal condition at each point (y, x, g, h) of $\mathcal{G}$ is given by:

$$\begin{aligned} \bar{\mathcal{V}}_T^- (y, x, g, h) = \max \{x, w_T (g)\} - \tau \min \left\{w_T (g), (x - h)_+\right\} \\ - \tau \left(\left(w_T (g) - (x - h)_+ \right)_+ + (x - w_T (g))_+ - h \right)_+, \end{aligned} \tag{5.6}$$

where $w_T (g) = \min \left\{g^W, g\right\}$.

Suppose now the function $\bar{\mathcal{V}}_{i+1}^-$ to be known on $\mathcal{G}$. Let us fix $(y, x, g, h) \in \mathcal{G}$ and let us focus on the computation of $\bar{\mathcal{V}}_i^+ (y, x, g, h)$ by solving Equation (4.7). Furthermore, according to Equation (4.7), $\bar{\mathcal{V}}_i^+ (y, x, g, h)$ can be interpreted as the solution of a fixed point problem:

$$v = f (v), \tag{5.7}$$

with

$$f (v) = \mathbb{E}^{\mathbb{Q}} \left[e^{- \int_{t_i}^{t_{i+1}} Y_s + \beta(s) ds} \left(F + \frac{\kappa}{1 - \kappa} \cdot (F - v)_+ \right) \Big| Y_{t_i} = y, X_{t_i} = x \right], \tag{5.8}$$

where F stands for

$$F = q_{x+t_i} \left(X_{t_i+1} - \tau \left(X_{t_{i+1}} - h \right)_+ \right) + p_{x+t_{i+1}} \bar{\mathcal{V}}_{i+1}^- \left(Y_{t_{i+1}}, X_{t_{i+1}}, g, h \right). \tag{5.9}$$

Following the same approach employed by Moenig and Bauer (2016) under the Black–Scholes model, one can verify that the solution of Equation (5.7) exists and is unique for $0 \leq \kappa < 1$. In particular, existence follows from $\lim_{v \to +\infty} v - f (v) = +\infty, f (0) > 0$ and from continuity of f by the Intermediate Value Theorem, while uniqueness is guaranteed by the monotony of $v - f (v)$. Moreover, one can prove that, under additional hypotheses easily met, F is a contraction; thus, Equation (5.8) can be faced by fixed point iterations, which can be started by considering $\bar{\mathcal{V}}_{i+1}^- (y, x, g, h)$ as the initial guess. In this regard, we employ a stopping criterion which is a combination of an absolute and a relative tolerance. Specifically, the solver stops when $|v - f (v)| < TOL_v (1 + |v|)$, where TOL_v is a given tolerance. We emphasize that, as an alternative to fixed point iterations, one may employ Newton-type methods to accelerate the convergence to the solution of Equation (5.7) (see e.g. Judd, 1998). Regardless of the chosen fixed point solver, the key point consists in calculating the expected

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

values that appears in (5.8). In order to tackle such a problem, we employ a tree approach. Technical details are explained in Appendix B.

Once the function $\bar{\mathcal{V}}_i^+$ is known, we can compute $\bar{\mathcal{V}}_i^-$ by solving the optimal withdrawal problem related to Equation (4.3). So, let us fix again $(y, x, g, h) \in \mathcal{G}$ and let us focus on solving the following problem:

$$\bar{\mathcal{V}}_i^- (y, x, g, h) = \max_{w \in [0, W_{\max}]} \hat{f}(w) \tag{5.10}$$

with

$$\hat{f}(w) = \bar{\mathcal{V}}_i^+ \left(y, x^+(w), g^+(w), h^+(w)\right) + (w - fee_i(w) - pen_i(w) - tax_i(w)), \tag{5.11}$$

$$W_{\max} = \max\left\{x, \min\left\{g^W, g\right\}\right\}, \tag{5.12}$$

$$x^+(w) = (x - w)_+ \tag{5.13}$$

$$g^+(w) = \begin{cases} (g - w)_+, & \text{if } w \leq g^W \\ \left(\min\left\{g - w, g \cdot \frac{x^+(w)}{x}\right\}\right)_+, & \text{if } w > g^W \end{cases} \tag{5.14}$$

$$h^+(w) = h - \left(w - (x - h)_+\right)_+ \tag{5.15}$$

$$fee_i(w) = s_i \cdot \left(w - \min\left\{g^W, g\right\}\right)_+, \tag{5.16}$$

$$pen_i(w) = s^g \cdot (w - fee_i(w)) \cdot 1_{\{x + t_i < 59.5\}}, \tag{5.17}$$

$$tax_i(w) = \tau \cdot \min\left\{w - fee_i(w) - pen_i(w), (x - h)_+\right\}. \tag{5.18}$$

The resolution of (5.10) is not trivial as the function $\hat{f}$ has no smoothness properties. In particular, $\hat{f}$ has singular points, due to the presence of the positive part function, as well as a discontinuity point, due to the function $g^+(w)$ at $w = g^W$. Therefore, we approach the solution of the maximization problem (5.10) through a very simple approach: we evaluate the target function in a set W of points and record the maximum value achieved on these points. In particular, we consider W as the union of two sets, W_1 and W_2. The first set $W_1 = \{n \cdot \frac{P}{N_W}, n \in \mathbb{N}\} \cap [0, W_{\max}]$ is a set of uniformly distributed values, with N_W a positive integer, while the second set $W_2 = \left\{g^W, g^W + 10^{-6}, W_{\max}\right\} \cap [0, W_{\max}]$ is a set of critical values that might not be included in W_1. In particular, g^W and $g^W + 10^{-6}$ are considered in order to handle the discontinuity at g^W.

Evaluating the function $\hat{f}$ requires the calculation of the function $\bar{\mathcal{V}}_i^+$ at $\left(y, x^+(w), g^+(w), h^+(w)\right)$ for any $w \in W$. These points may not belong to the grid $\mathcal{G}$ as the values $x^+(w), g^+(w), h^+(w)$ may not belong to $\mathcal{G}_X, \mathcal{G}_G$ and $\mathcal{G}_H$,

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

respectively. So, in order to compute $\bar{\mathcal{V}}_i^+\left(y, x^+(w), g^+(w), h^+(w)\right)$, interpolation on $\mathcal{G}$ of $\bar{\mathcal{V}}_i^+$ is required. To this aim, we employ trilinear interpolation (Gomes *et al.*, 2019).

Finally, we observe that if we are interested in computing the policy cost by assuming the worst-case withdrawal strategy, adapting Equation (5.10) to this aim is straightforward.

5.3. Backward evaluation of $\mathcal{U}$

Once an approximation $\bar{\mathcal{V}}$ of $\mathcal{V}$ is available, we can tackle the policy evaluation according to the insurer's perspective, that is computing $\mathcal{U}$. Approximating $\mathcal{U}$ is easier than approximating $\mathcal{V}$ because of two reasons. First of all, the implicit nonlinear Equation (4.7) is replaced by an explicit Equation (4.12). Secondly, the problem of computing the best withdrawal has been already solved while approximating $\mathcal{V}$, so we have just to recover the optimal withdrawals already computed.

Similarly to what we have done for $\mathcal{V}$, we consider a function $\bar{\mathcal{U}}_i^+ : \mathcal{G} \to \mathbb{R}$ such that for any point (y, x, g, h) of $\mathcal{G}$, $\bar{\mathcal{U}}_i^+(y, x, g, h)$ approximates $\mathcal{U}^+(t_i, y+\beta(t_i), x, g, h)$ and a function $\bar{\mathcal{U}}_i^- : \mathcal{G} \to \mathbb{R}$ such that for any point (y, x, g, h) of $\mathcal{G}$, $\bar{\mathcal{U}}_i^-(y, x, g, h)$ approximates $\mathcal{U}^-(t_i, y+\beta(t_i), x, g, h)$. According to (4.9), for any point (y, x, g, h) of $\mathcal{G}$, the terminal condition is given by:

$$\bar{\mathcal{U}}_T^-(y, x, g, h) = \max\left\{x, \min\left\{g^W, g\right\}\right\}. \tag{5.19}$$

Suppose now the function $\bar{\mathcal{U}}_{i+1}^-$ to be known on $\mathcal{G}$. Let us fix $(y, x, g, h) \in \mathcal{G}$ and let us focus on the computation of $\bar{\mathcal{U}}_i^+(y, x, g, h)$ by computing the following expression:

$$\begin{aligned}\bar{\mathcal{U}}_i^+(y, x, g, h) = \mathbb{E}^{\mathbb{Q}}\Big[e^{-\int_{t_i}^{t_{i+1}} Y_s+\beta(s)ds}\left(q_{x+t_i}X_{t_{i+1}} + p_{x+t_{i+1}}\bar{\mathcal{U}}_{i+1}^-\left(Y_{t_{i+1}}, X_{t_{i+1}}, g, h\right)\right)\,\Big|\, Y_{t_i} \\ = y, X_{t_i} = x\Big].\end{aligned} \tag{5.20}$$

We compute such an expression by using the same tree approach employed to compute (5.8). Please observe that in this case, no fix point iterations are required because (5.20) gives $\mathcal{U}_i^+$ through an explicit equation.

Suppose now the function $\bar{\mathcal{U}}_i^+$ to be known on $\mathcal{G}$. Let us fix $(y, x, g, h) \in \mathcal{G}$ and let us focus on the computation of $\bar{\mathcal{U}}_i^-(y, x, g, h)$. Let w_i be the maximum point for the problem (5.10) for the point $(y, x, g, h) \in \mathcal{G}$. The following relation holds:

$$\bar{\mathcal{U}}_i^-(y, x, g, h) = \bar{\mathcal{U}}_i^+\left(y, x^+(w_i), g^+(w_i), h^+(w_i)\right) + (w - fee_i(w_i) - pen_i(w_i)), \tag{5.21}$$

where x^+, g^+, h^+, fee_i and pen_i are defined as in (5.13)–(5.17). Also in this case, interpolation is required and we employ again trilinear interpolation.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

5.4. Sketch of the algorithm

We present the sketch of the algorithm to approximate the initial fair contract values $\mathcal{V}(0, r_0, P, P, P)$ and $\mathcal{U}(0, r_0, P, P, P)$.

1. Set the terminal values $\bar{\mathcal{V}}_T^-(y, x, g, h)$ and $\bar{\mathcal{U}}_T^-(y, x, g, h)$ according to Equations (5.6) and (5.19) for every point (y, x, g, h) in $\mathcal{G}$.
2. For all $i = T - 1, \ldots 1$
   (a) Compute $\bar{\mathcal{V}}_i^+(y, x, g, h)$ and $\bar{\mathcal{U}}_i^+(y, x, g, h)$ by solving Equations (5.8) and in (5.20) for every point (y, x, g, h) in $\mathcal{G}$;
   (b) Compute $\bar{\mathcal{V}}_i^-(y, x, g, h)$ by solving Equation (5.10) for every point (y, x, g, h) in $\mathcal{G}$;
   (c) Compute $\bar{\mathcal{U}}_i^-(y, x, g, h)$ by solving Equation (5.21) for every point (y, x, g, h) in $\mathcal{G}$;
3. Compute $\bar{\mathcal{V}}_0^+(0, P, P, P)$ and $\bar{\mathcal{U}}_0^+(0, P, P, P)$ by solving Equations (5.8) and (5.20).

Values $\bar{\mathcal{V}}_0^+(0, P, P, P)$ and $\bar{\mathcal{U}}_0^+(0, P, P, P)$ approximate $\mathcal{V}(0, r_0, P, P, P)$ and $\mathcal{U}(0, r_0, P, P, P)$, respectively. We point out that the algorithm is fully parallelizable: in fact the computations for every point in $\mathcal{G}$ are independent of each other.

A common practice in the context of Variable Annuities (see, e.g., Forsyth and Vetzal, 2014) consists in computing the fair policy cost φ_{IN}^*, that is the particular value of φ that makes the insurer's initial value of the policy $\mathcal{U}(0, r_0, P, P, P)$ equal to the net premium P. To this aim, the algorithm can be plugged into the Secant method to solve the equation $\mathcal{U}(0, r_0, P, P, P)(\varphi) = P$. The Secant method must be equipped with a suitable pair of initial values and a suitable stopping criterion. In this regard, we suggest to use $\varphi_0 = 0$, and $\varphi_1 = 100$ bps (basis points) as the initial values. The stopping criterion that we consider is as a combination of a function tolerance TOL_f and a step tolerance TOL_φ, which means that the algorithm stops when both the two following conditions are satisfied:

$$|\mathcal{U}(0, r_0, P, P, P)(\varphi_n) - P| < TOL_f \text{ and } |\varphi_n - \varphi_{n-1}| < TOL_\varphi. \tag{5.22}$$

6. Numerical results

In this Section, we report the results of some numerical tests. Tables 1 and 2 report the employed parameters for the GMWB product and for the stochastic model, respectively. We underline that these parameters, with the exception of those for the interest rate process, are the same employed by Moenig and Bauer (2016). Table 3 shows the discretization parameters for the numerical method. In particular, in order to investigate the convergence of the proposed method, we vary the discretization parameters by considering seven different parameter configurations. Specifically, the basic configuration is termed $\mathcal{C}_{base}$.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

TABLE 1

PARAMETER CHOICES FOR THE PH AND CONTRACT SPECIFICATIONS.

Description	Parameter	Value
Age at inception	x	55
Premium	P	100
Years to maturity	T	15
Annual guaranteed amount	g^W	7
Excess withdrawal fee	s_i	8%, 7%, . . . , 1%, 0%, 0%, . . .
Fee rate	φ	to be determined
Income tax rate	τ	0%, or 30%
Capital gain tax rate	κ	0%, or 23%
Early withdrawal penalty	s^g	10%

TABLE 2

PARAMETER CHOICES FOR THE BLACK–SCHOLES HULL–WHITE MODEL.

Description	Parameter	Value
Initial fund value	S_0	100
Fund volatility	σ	0.1, 0.3
Initial interest rate	r_0	0.03, 0.05
Interest rate mean reversion speed	k	1
Interest rate mean	θ_t	*flat*
Interest rate volatility	ω	0.05, 0.1
Correlation	ρ	0.2

TABLE 3

PARAMETER CONFIGURATIONS FOR THE NUMERICAL METHOD.

		Parameter configurations						
Description	Parameter	$\mathcal{C}_{T-}$	$\mathcal{C}_{X-}$	$\mathcal{C}_{GHW-}$	$\mathcal{C}_{base}$	$\mathcal{C}_{T+}$	$\mathcal{C}_{X+}$	$\mathcal{C}_{GHW+}$
Time step per year	N_T	12	25	25	25	50	25	25
Points in $\mathcal{G}_X$	N_X	500	250	500	500	500	1000	500
Points in $\mathcal{G}_G$	N_G	100	100	50	100	100	100	200
Points in $\mathcal{G}_H$	N_H	100	100	50	100	100	100	200
Withdrawal step	N_W	100	100	50	100	100	100	200

Configurations $\mathcal{C}_{T-}$, $\mathcal{C}_{X-}$ and $\mathcal{C}_{GHW-}$ employ a smaller number of discretization points than $\mathcal{C}_{base}$, while $\mathcal{C}_{T+}$, $\mathcal{C}_{X+}$ and $\mathcal{C}_{GHW+}$ employ a larger number of discretization steps. In particular, $\mathcal{C}_{T-}$ and $\mathcal{C}_{T+}$ vary N_T, that is the number of time steps. Configurations $\mathcal{C}_{X-}$ and $\mathcal{C}_{X+}$ vary N_X, that is the number of points that discretize the account value. Finally, configurations $\mathcal{C}_{GHW-}$ and $\mathcal{C}_{GHW+}$ vary the parameters N_G, N_H and N_W together, which rule the discretization of the product bases G_t and H_t, and of the discretization of the withdrawal step respectively.

TABLE 4

FAIR FEE RATE φ^*_{IN} (IN BASIS POINTS) ACCORDING TO THE INSURER'S SUBJECTIVE VALUATION, FOR DIFFERENT VALUES OF r_0, σ AND ω.

		$r_0 = 0.03,\ \sigma = 0.16$		$r_0 = 0.05,\ \sigma = 0.19$	
	Configuration	$\omega = 0.05$	$\omega = 0.1$	$\omega = 0.05$	$\omega = 0.1$
No taxation	$\mathcal{C}_{T-}$	69.40	94.81	41.87	56.96
	$\mathcal{C}_{X-}$	69.93	95.35	42.24	57.33
	$\mathcal{C}_{GHW-}$	66.57	91.19	40.59	55.25
	$\mathcal{C}_{base}$	69.36	94.78	41.90	56.98
	$\mathcal{C}_{T+}$	69.35	94.78	41.91	56.97
	$\mathcal{C}_{X+}$	69.34	94.77	41.89	56.97
	$\mathcal{C}_{GHW+}$	69.37	94.79	41.91	56.99
With taxation	$\mathcal{C}_{T-}$	43.10	60.26	23.83	33.46
	$\mathcal{C}_{X-}$	43.45	60.70	24.16	33.80
	$\mathcal{C}_{GHW-}$	40.78	57.17	22.84	32.12
	$\mathcal{C}_{base}$	43.12	60.29	23.91	33.54
	$\mathcal{C}_{T+}$	43.18	60.31	23.96	33.58
	$\mathcal{C}_{X+}$	43.08	60.26	23.90	33.50
	$\mathcal{C}_{GHW+}$	43.19	60.40	23.99	33.64

Moreover, as far as the solution of Equation (5.7) is concerned, we set $TOL_v = 10^{-6}$, while, with regard to the Secant method, we set $TOL_f = 10^{-2}$ and $TOL_\varphi = 10^{-1}$ bps.

If not explicitly stated otherwise, we assume that the PH adopts the withdrawal strategy that maximizes his total wealth, according to Equation (5.10). Furthermore, in Subsection 6.2, we also investigate the worst-case cost of the hedge, that is the PH determines the withdrawal amount according to Equation (4.11).

Finally, in order to estimate the mortality and survival probabilities q and p, we employ the 2007 Period Life Table for the Social Security Area Population for the USA (Social Security Administration).

6.1. Computing the fair fee rate

We start by computing the fair fee rate φ^*_{IN} according to the insurer's subjective valuation. In particular, we consider some test cases with different values of the initial interest rate r_0, the volatility of the interest rate ω and the volatility of the underlying fund σ. Numerical results are reported in Table 4.

First of all, we observe that the numerical results are very stable. In particular, results about the configurations with an increased number of discretization points (that is $\mathcal{C}_{T+}$, $\mathcal{C}_{X+}$, and $\mathcal{C}_{GHW+}$), always vary less than 0.2 bps against the basic configuration $\mathcal{C}_{base}$.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

By comparing the results with and without taxation, we observe that, in all the considered cases, including taxation decrease the fair fee that reduces the account value, that is the policy cost. The reason for this reduction lies in the withdrawal strategy: if taxation is applied, the optimal withdrawal strategy from PH's perspective changes and it does not overlap anymore with the worst-case strategy according to insurer's perspective, thus the withdrawal strategy adopted by the PH becomes less expensive for the insurer. In particular, as shown in Subsection 6.3 and as already noted by Moenig and Bauer (2016) in the Black–Scholes model, the PH tends in general to defer withdrawals when taxation is applied: withdrawing may be suboptimal for the PH since the withdrawals are subject to income taxes, whereas the sums invested in the policy grow tax-deferred. Moreover, we observe that the higher the interest rate volatility, the greater the policy cost. This is probably due to the fact that, by increasing the volatility of the interest rate, it is easier to observe very low (or negative) interest rates which make replicating the policy very expensive. Thus, both taxation and interest rate modeling have a sensitive impact on policy evaluation.

6.2. Comparing policy initial values

We compute now the PH's initial subjective policy value with φ equal to φ^*_{IN}, that is the break-even fee, as in Table 4: this is the amount of money the agent needs to replicate the policy on its own. To this aim, we employ parameter configuration $\mathcal{C}_{base}$. Numerical results are reported in Table 5. We observe that if no taxation is applied, the subjective valuations of the PH and of the insurer equate the initial premium P and the contract is fair for both the two agents. Instead, when taxation is applied, the contract values according to the subjective valuations of the two agents both increase. In particular, as far as the insurer is concerned, we have considered $\chi = 3\%$ as the premium tax rate (which is a common value, see Moran, 2017). With such a premium tax rate, the gross premium GP that the insurer requires to cover all the costs is 103.09, so that the net premium P is 100 and the contract is fair for the insurer. As far as the PH is concerned, the increase in the contract value is due to the tax regime applied to the policy, which is advantageous compared to the tax regime applied to investment outside the policy. Moreover, according to Table 5, in all the considered cases, the customer will be willing to pay much more than 103.90 to buy the policy: for example, if $r_0 = 0.03$, $\sigma = 0.16$ and $\omega = 0.05$ then the PH's subjective valuation of the policy is 110.15. Therefore, if the insurer sets a sale price between 103.09 and 110.15, then the sale will be advantageous for both the PH and the insurer.

The proposed model allows us to recreate a framework that makes Variable Annuities particularly interesting to customers: although taxes are applied on the earnings of the policy, the tax regime is particularly favorable for this type of products. Therefore, the GMWB policy is attractive for the customer and profitable for the insurer.

TABLE 5

FAIR INITIAL OPTION VALUE ACCORDING TO THE SUBJECTIVE VALUATIONS OF THE INSURER AND THE PH, WITH $\varphi = \varphi^*_{IN}$ AS IN TABLE 4, FOR DIFFERENT VALUES OF r_0, σ AND ω.

		$r_0 = 0.03,\ \sigma = 0.16$		$r_0 = 0.05,\ \sigma = 0.19$	
Agent	Tax	$\omega = 0.05$	$\omega = 0.1$	$\omega = 0.05$	$\omega = 0.1$
Insurer	No	100.00	100.00	100.00	100.00
	With	103.09	103.09	103.09	103.09
PH	No	100.00	100.00	100.00	100.00
	With	110.15	110.70	114.99	115.94

TABLE 6

FAIR INITIAL OPTION VALUE ACCORDING TO THE SUBJECTIVE VALUATIONS OF THE INSURER AND OF THE PH WHILE ASSUMING THE WORST-CASE WITHDRAWAL STRATEGY, WITH $\varphi = \varphi^*_{IN}$ AS IN TABLE 4, FOR DIFFERENT VALUES OF r_0, σ AND ω.

		$r_0 = 0.03,\ \sigma = 0.16$		$r_0 = 0.05,\ \sigma = 0.19$	
Agent	Tax	$\omega = 0.05$	$\omega = 0.1$	$\omega = 0.05$	$\omega = 0.1$
Insurer	No	100.00	100.00	100.00	100.00
	With	104.75	105.16	104.39	104.73
PH	No	100.00	100.00	100.00	100.00
	With	105.62	106.32	107.00	107.69

To complete our investigation about the subjective evaluation of the GMWB policy, we also consider the worst-case of the hedge. In this particular case, we assume that the PH withdraws as to maximize the hedging cost for the insurance company. To this aim, we compute the gross fee (before tax) required to fund policy replication. In order to compare the results with those of Table 5, we consider again $\varphi = \varphi^*_{IN}$ (that is the the value of φ which makes the contract fair according to insurer's subjective valuation, while assuming a PH that maximizes his own wealth). Numerical results in Table 6 show that, in the absence of taxation, the worst strategy for the insurer overlaps with the best strategy for the PH: in fact, since $\varphi = \varphi^*_{IN}$, the contract is worth 100 and it is fair for both the two agents. On the contrary, when the taxation is applied, we can observe that the cost for the insurer increases (this is in fact the maximum possible cost) and the subjective value for the PH decreases (he no longer employs the best strategy). In particular, with reference to the case $r_0 = 0.03$, $\sigma = 0.16$ and $\omega = 0.05$, the cost for the insurer increases to 104.75, while the value for the PH decreases down to 105.62. Again, it is worth noting that, in all the considered cases, the worst-case cost of hedging never exceeds the subjective value of the PH; therefore, even in this case, it is still possible to find an advantageous price for both the insurer and the PH. In particular, if the insurer sets

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

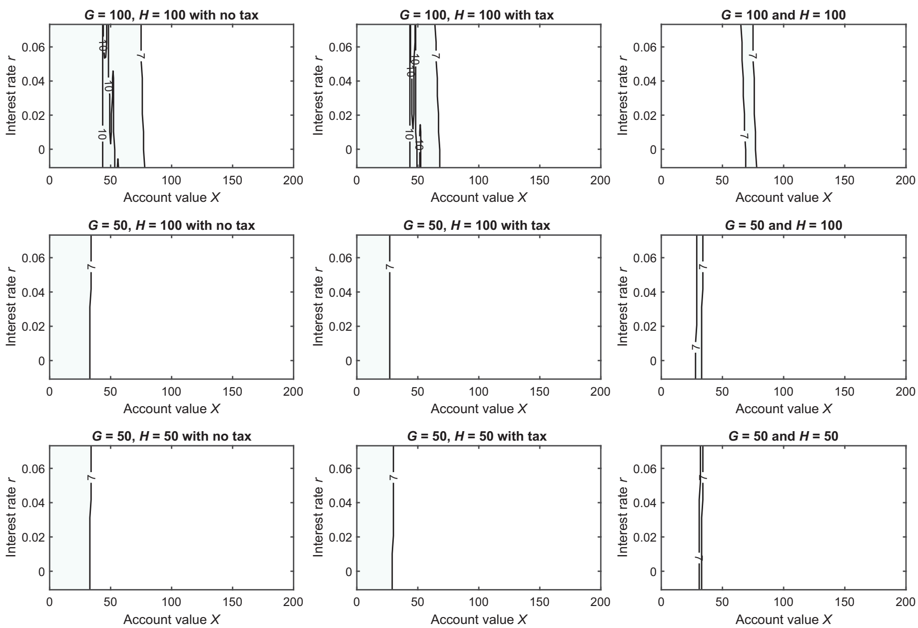

FIGURE 1: In the first two columns, the contour plot with respect to X_4^- (x-axis) and r_4 (y-axis) of the optimal withdrawal w_4 at time $t_4 = 4$ without or applying taxation. In the third column, the difference between the optimal withdrawals without and with taxation reported in the first two columns. Green areas denote positive values.

a sale price between 104.75 and 110.15, he will surely cover the hedging costs regardless of the withdrawal strategy, and the PH will have a profit margin.

6.3. Comparing withdrawal strategies with and without taxation

The last numerical test we propose consists in comparing the optimal withdrawals performed by the PH while considering or not taxation. In particular, we consider the case with $r_0 = 0.03$, $\sigma = 0.16$ and $\omega = 0.05$ (results for other parameters combinations are similar). For this test, we employ parameter configuration $\mathcal{C}_{T+}$, which provides a better discretization of the process Y_t than the others. Moreover, the value of φ is set as the break-even fee with taxes, that is $\varphi^*_{IN} = 43.18$ basis points. Optimal withdrawal amounts w_i at different anniversaries are reported in Figures 1, 2 and 3. The first column represents the optimal amount as a function of $X_{t_i}^-$ and r_{t_i}, by considering different values for G and H and a zero tax rate, while in the second column by considering both a positive income tax rate and a positive capital gain tax rate. The area where the color is darker identifies higher withdrawals. Finally, in the third column, the difference between the optimal amount without taxation and with taxation. Here, green areas indicate that withdrawals without tax are higher, while red areas (not visible) indicate that withdrawals with tax are higher.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

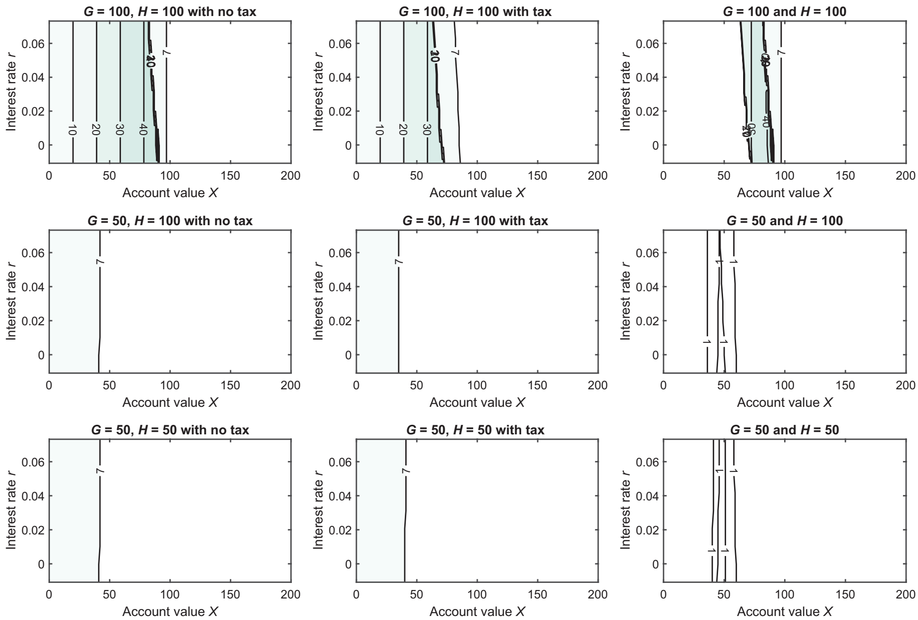

FIGURE 2: Same contour plots as in Figure 1, but considering t_8 in place of t_4.

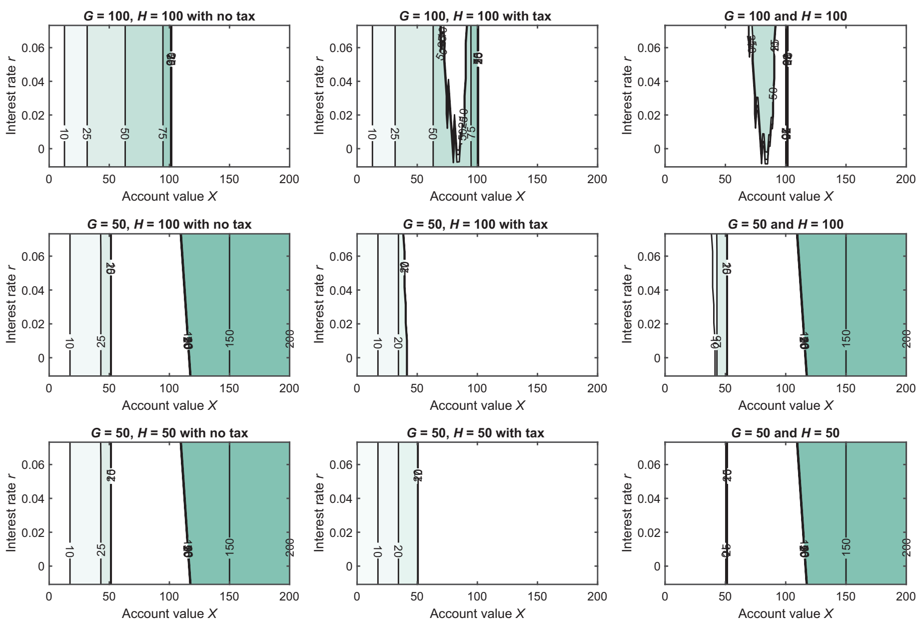

FIGURE 3: Same contour plots as in Figure 1, but considering t_{12} in place of t_4.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

We can observe that the optimal amount depends on all the considered parameters. In particular, the withdrawal strategy may significantly change according to the actual interest rate, as shown in Figure 3 for $G_{t_{12}} = H_{t_{12}} = 100$. As far as the impact of taxation on withdrawal strategy is concerned, we find the same effect observed by Moenig and Bauer (2016): when taxation is applied, the PH withdraws less than when taxation is not considered. In fact, for all the numerical cases considered, the last columns in Figures 1, 2 and 3 show only positive values, which means the amount withdrawn with no taxation is higher. As observed in Moenig and Bauer (2016), in the absence of taxes, as the account value increases more and more, the PH is motivated to withdraw money instead of leaving it in the policy where it is reduced by fees. Conversely, if taxation is applied, withdrawals are taxed as ordinary income and they are subject to capital gain tax if invested in other products. Therefore, it is more convenient for the PH not to withdraw the money, letting it grow within the policy. Such a difference in the withdrawal strategy is particularly clear in Figure 3: if $G = H = 50$, then the PH withdraws large amount of money when X is high if taxation is neglected, whereas no money if taxation is applied.

7. Conclusions

In this paper we have investigate the impact of taxation on a GMWB Variable Annuity when stochastic interest rate is considered. We modeled taxation following the approach of Moenig and Bauer (2016): we have considered a subjective risk-neutral valuation methodology that considers differences in the taxation for both different products and market agents. Moreover, we have modeled stochastic interest rate through the Hull–White model. This analysis combines the effects of taxation and of the variable interest rate which, as already shown separately in other research work, can have a significant impact on the withdrawal choices and thus on the hedging costs. This analysis has been possible thanks to the use of an efficient numerical method based on a tree approach (Goudenège *et al.*, 2019). Numerical results show many interesting facts. First of all, both taxation and interest rate modeling can have a relevant impact on policy evaluation: the break-even fee can change of several basis points when the parameters of these two factors change. Then, applying different taxation to insurer and to PH can lead to different policy evaluations: in particular PH's valuation is higher than insurer's valuation and this makes buying and selling the policy convenient for both the two agents. Moreover, numerical tests show that taxation clearly impacts on withdrawal strategy: it discourages the PH to perform withdrawals. This is useful to match theoretical prices to those actually observed on the real market. In conclusion, the model presented here represents an important extension in the evaluation of GMWB-type policies.

REFERENCES

BACINELLO, A.R. and ZOCCOLAN, I. (2019) Variable annuities with a threshold fee: Valuation, numerical implementation and comparative static analysis. *Decisions in Economics and Finance*, **42**(1), 21–49.

BERNARD, C. and KWAK, M. (2016) Semi-static hedging of variable annuities. *Insurance: Mathematics and Economics*, **67**, 173–186.

BRIGO, D. and MERCURIO, F. (2007) *Interest Rate Models-Theory and Practice: with Smile, Inflation and Credit*. Springer, Berlin, Heidelberg: Springer Science & Business Media.

CAMPBELL, J.Y. (2006) Household finance. *The Journal of Finance*, **61**(4), 1553–1604.

COSTABILE, M. (2017) A lattice-based model to evaluate variable annuities with guaranteed minimum withdrawal benefits under a regime-switching model. *Scandinavian Actuarial Journal*, **2017**(3), 231–244.

COSTABILE, M., MASSABÒ, I. and RUSSO, E. (2020) Evaluating variable annuities with GMWB when exogenous factors influence the policy-holder's withdrawals. *The European Journal of Finance*, **26**(2–3), 238–257.

DAI, T.-S., YANG, S.S. and LIU, L.-C. (2015) Pricing guaranteed minimum/lifetime withdrawal benefits with various provisions under investment, interest rate and mortality risks. *Insurance: Mathematics and Economics*, **64**, 364–379.

DONNELLY, R.F., JAIMUNGAL, S. and RUBISOV, D. (2014) Valuing guaranteed withdrawal benefits with stochastic interest rates and volatility. *Quantitative Finance*, **14**(2), 369–382.

EKVALL, N. (1996) A lattice approach for pricing of multivariate contingent claims. *European Journal of Operational Research*, **91**(2), 214–228.

FORSYTH, P. and VETZAL, K. (2014) An optimal stochastic control framework for determining the cost of hedging of variable annuities. *Journal of Economic Dynamics and Control*, **44**, 29–53.

GOMES, A., VOICULESCU, I., JORGE, J., WYVILL, B. and GALBRAITH, C. (2009) *Implicit Curves and Surfaces: Mathematics, Data Structures and Algorithms*. Springer-Verlag London: Springer Science & Business Media.

GOUDENÈGE, L., MOLENT, A. and ZANETTE, A. (2016) Pricing and hedging GLWB in the Heston and in the Black–Scholes with stochastic interest rate models. *Insurance: Mathematics and Economics*, **70**, 38–57.

GOUDENÈGE, L., MOLENT, A. and ZANETTE, A. (2018) Pricing and hedging GMWB in the Heston and in the Black–Scholes with stochastic interest rate models. *Computational Management Science*, **16**(1).

GOUDENÈGE, L., MOLENT, A. and ZANETTE, A. (2019) Gaussian process regression for pricing variable annuities with stochastic volatility and interest rate. arXiv preprint arXiv:1903.00369.

GUDKOV, N., IGNATIEVA, K. and ZIVEYI, J. (2019) Pricing of guaranteed minimum withdrawal benefits in variable annuities under stochastic volatility, stochastic interest rates and stochastic mortality via the componentwise splitting method. *Quantitative Finance*, **19**(3), 501–518.

HAENTJENS, T. and IN'T HOUT, K.J. (2012) Alternating direction implicit finite difference schemes for the Heston–Hull–White partial differential equation. *The Journal of Computational Finance*, **16**(1), 83.

HULL, J. and WHITE, A. (1994) Numerical procedures for implementing term structure models I: Single-factor models. *Journal of Derivatives*, **2**(1), 7–16.

JUDD, K.L. (1998) *Numerical Methods in Economics*. Cambridge, Massachusetts London, England: MIT Press.

KLING, A., RUEZ, F. and RUSS, J. (2014) The impact of policyholder behavior on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities. *European Actuarial Journal*, **4**(2), 281–314.

LEHMANN, E.L. (2004) *Elements of Large-Sample Theory*. Springer-Verlag New York: Springer Science & Business Media.

LIN, X.S., WU, P. and WANG, X. (2016) Move-based hedging of variable annuities: A semi-analytic approach. *Insurance: Mathematics and Economics*, **71**, 40–49.

MACKAY, A., AUGUSTYNIAK, M., BERNARD, C. and HARDY, M.R. (2017) Risk management of policyholder behavior in equity-linked life insurance. *Journal of Risk and Insurance*, **84**(2), 661–690.

MOENIG, T. (2012) *Optimal policyholder behavior in personal savings products and its impact on valuation*. Ph.D. Thesis, Georgia State University.
MOENIG, T. and BAUER, D. (2016) Revisiting the risk-neutral approach to optimal policyholder behavior: A study of withdrawal guarantees in Variable Annuities. *Review of Finance*, **20**(2), 759–794.
MOENIG, T. and ZHU, N. (2018) Lapse-and-reentry in Variable Annuities. *Journal of Risk and Insurance*, **85**(4), 911–938.
MORAN, S. (2017) Taxation of insurance companies. https://docs.legis.wisconsin.gov/misc/lfb/informational_papers. Wisconsin Legislative Fiscal Bureau.
NELSON, D.B. and RAMASWAMY, K. (1990) Simple binomial processes as diffusion approximations in financial models. *The Review of Financial Studies*, **3**(3), 393–430.
NISSIM, D. (2010) Analysis and valuation of insurance companies. *Center for Excellence in Accounting and Security Analysis*, (2).
PENG, J., LEUNG, K.S. and KWOK, Y.K. (2012) Pricing guaranteed minimum withdrawal benefits under stochastic interest rates. *Quantitative Finance*, **12**(6), 933–941.
PISCOPO, G. and HABERMAN, S. (2011) The valuation of guaranteed lifelong withdrawal benefit options in variable annuity contracts and the impact of mortality risk. *North American Actuarial Journal*, **15**(1), 59–76.
POLLARD, D. (2012) *Convergence of Stochastic Processes*. Springer-Verlag New York: Springer Science & Business Media.
ROSS, S. (1987) Arbitrage and martingales with taxation. *Journal of Political Economy*, **95**(2), 371–393.
SECURE RETIREMENT INSTITUTE (2019) U.S. individual annuity sales survey, third quarter. https://www.limra.com/globalassets/limra/newsroom/fact-tank/sales-data/2019/q4/4q-2019-annuity-sales-estimates-vfinal.pdf. Accessed: 01 Apr 2020.
SHEVCHENKO, P.V. and LUO, X. (2017) Valuation of variable annuities with guaranteed minimum withdrawal benefit under stochastic interest rate. *Insurance: Mathematics and Economics*, **76**, 104–117.
SIBLEY, M. (2002) On the valuation of tax-advantaged retirement accounts. *Financial Services Review*, **11**(3), 233.
SKIPPER JR., H.D. (2001) The taxation of life insurance policies in OECD countries: Implications for tax policy and planning. In *Insurance and Private Pensions Compendium for Emerging Economies*. Paris: OECD.
SOCIAL SECURITY ADMINISTRATION. Actuarial Life Table. https://www.ssa.gov/oact/STATS/table4c6_2007.html.
SUN, J., SHEVCHENKO, P.V. and FUNG, M.C. (2018) The impact of management fees on the pricing of variable annuity guarantees. *Risks*, **6**(3), 103.
VAN DEN BERG, I. (2000) *Principles of Infinitesimal Stochastic and Financial Analysis*. Singapore: World Scientific.

ANDREA MOLENT
Dipartimento di Scienze Economiche e Statistiche
Università degli Studi di Udine
Udine, Italy
E-Mail: andrea.molent@uniud.it

APPENDIX A

A.1. Markov chain to approximate Y

In this Appendix, we explain how to design a discrete time Markov chain that approximates the process Y, based on the trinomial tree introduced in Goudenège *et al.* (2019). First of

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

all, we consider a partition of the time interval [0, T] in $T \cdot N_T$ sub-intervals, that is N_T per year. We define $\Delta t = \frac{1}{N_T}$ as the time increment and we term $\bar{t}_n = n \cdot \Delta t$ the n-th time step for $n = 0, \ldots, T \cdot N_T$. Please observe that policy anniversaries $t_0, \ldots, t_T$ are included in the time steps $\bar{t}_0, \bar{t}_1 \ldots, \bar{t}_{T \cdot N_T}$ and in particular $t_i = \bar{t}_{i \cdot N_T} = i$. We consider the set $\mathcal{Y}$ given by

$$\mathcal{Y} = \left\{ \upsilon_j = \frac{3}{2} j \sigma_Y^{\Delta t},\ j \in \mathbb{Z} \right\}, \tag{A1}$$

where the coefficient $\sigma_Y^{\Delta t}$ is the standard deviation of the random variable $Y_{\bar{t}_{n+1}} - Y_{\bar{t}_n}$ (which is the same for all $\bar{t}_n$ values) and it is given by

$$\sigma_Y^{\Delta t} = \omega \sqrt{\frac{1 - \exp\left(-2k \cdot \Delta t\right)}{2k}}. \tag{A2}$$

We define now a discrete time Markov chain $\bar{Y}^{\Delta t} = \left\{ \bar{Y}_n^{\Delta t}, n = 0, \ldots, T \cdot N_T \right\}$ whose state space is an opportune subset of $\mathcal{Y}$ and that matches the first two moments of the process $Y = \{ Y_t, 0 \leq t \leq T \}$. The process $\bar{Y}^{\Delta t}$ is designed so it weakly converges to the process Y: in particular $\bar{Y}_n^{\Delta t}$ converges to $Y_{\bar{t}_n}$.

The initial value is $\bar{Y}_0^{\Delta t} = \upsilon_0 = 0$, so $\bar{Y}_0^{\Delta t} = Y_0$. Now, let us fix a value $n \in \{0, \ldots, T \cdot N_T - 1\}$ and suppose $\bar{Y}_n^{\Delta t} = \upsilon_m$ for a certain integer $m \in \mathbb{Z}$. Let

$$\mu_Y^{\Delta t}\left(\upsilon_m\right) = \mathbb{E}\left[Y_{\bar{t}_{n+1}} \,\middle|\, Y_{\bar{t}_n} = \upsilon_m \right] = \upsilon_m \cdot \exp\left(-k \cdot \Delta t\right) \tag{A3}$$

be the expected value of the random variable $Y_{\bar{t}_{n+1}} \,|\, Y_{\bar{t}_n} = \upsilon_m$. We term

$$j_A = \text{ceil}\left[\frac{2}{3\sigma_Y^{\Delta t}} \mu_Y^{\Delta t}\left(\upsilon_m\right) \right] \tag{A4}$$

the index of the first element of $\mathcal{Y}$ whose value is bigger than the expected value of the process $Y_{\bar{t}_{n+1}} \,|\, Y_{\bar{t}_n} = \upsilon_m$. Moreover, we also consider these three indices:

$$j_B = j_A - 1,\ j_C = j_A + 1,\ j_D = j_A - 2. \tag{A5}$$

In particular, if we define the variables

$$\Delta^A = \upsilon_{j_A} - \mu_Y^{\Delta t}\left(\upsilon_m\right) \tag{A6}$$

and

$$\Delta^B = \mu_Y^{\Delta t}\left(\upsilon_m\right) - \upsilon_{j_B} \tag{A7}$$

then $0 \leq \Delta^A < \frac{3}{2}\sigma_Y^{\Delta t}$ and $0 < \Delta^B \leq \frac{3}{2}\sigma_Y^{\Delta t}$.

There are two alternatives for the future states of the process $\bar{Y}^{\Delta t}$: it can move from υ_m either to υ_{j_A}, υ_{j_B}, υ_{j_C}, or to υ_{j_A}, υ_{j_B}, υ_{j_D}. Transition probabilities p_A, p_B, p_C, p_D for both of these two alternatives are stated in Table A.1. In particular, it is possible to prove that if $0 \leq \Delta^A \leq \frac{\sqrt{5}}{2}\sigma_Y^{\Delta t}$ then p_A, p_B, $p_C \in [0, 1]$, while if $\frac{3-\sqrt{5}}{2}\sigma_Y^{\Delta t} \leq \Delta^A < \frac{3}{2}\sigma_Y^{\Delta t}$ then p_A, p_B, $p_D \in [0, 1]$. Since $\frac{3-\sqrt{5}}{2}\sigma_Y^{\Delta t} < \frac{\sqrt{5}}{2}\sigma_Y^{\Delta t}$, at least one of the two sets of probabilities is well defined.

TABLE A.1

TRANSITION PROBABILITIES FOR THE PROCESS $\bar{Y}^{\Delta t}$.

	Transition to $\upsilon_{j_A}, \upsilon_{j_B}, \upsilon_{j_C}$	Transition to $\upsilon_{j_A}, \upsilon_{j_B}, \upsilon_{j_D}$
p_A	$\frac{5\left(\sigma_Y^{\Delta t}\right)^2-4\left(\Delta^A\right)^2}{9\left(\sigma_Y^{\Delta t}\right)^2}$	$\frac{2\left(\Delta^B\right)^2+3\left(\sigma_Y^{\Delta t}\right)\left(\Delta^B\right)+2\left(\sigma_Y^{\Delta t}\right)^2}{9\left(\sigma_Y^{\Delta t}\right)^2}$
p_B	$\frac{2\left(\Delta^A\right)^2+3\left(\sigma_Y^{\Delta t}\right)\left(\Delta^A\right)+2\left(\sigma_Y^{\Delta t}\right)^2}{9\left(\sigma_Y^{\Delta t}\right)^2}$	$\frac{5\left(\sigma_Y^{\Delta t}\right)^2-4\left(\Delta^B\right)^2}{9\left(\sigma_Y^{\Delta t}\right)^2}$
p_C	$\frac{2\left(\Delta^A\right)^2-3\left(\sigma_Y^{\Delta t}\right)\left(\Delta^A\right)+2\left(\sigma_Y^{\Delta t}\right)^2}{9\left(\sigma_Y^{\Delta t}\right)^2}$	0
p_D	0	$\frac{2\left(\Delta^B\right)^2-3\left(\sigma_Y^{\Delta t}\right)\left(\Delta^B\right)+2\left(\sigma_Y^{\Delta t}\right)^2}{9\left(\sigma_Y^{\Delta t}\right)^2}$

Moreover, transition probabilities in Table A.1 have been computed in order to match the first two moments of the process Y: this means that the random vectors $\bar{Y}_{n+1}^{\Delta t} - \bar{Y}_n^{\Delta t}$ and $Y_{\bar{t}_{n+1}} - Y_{\bar{t}_n}$, given $\bar{Y}_n^{\Delta t} = Y_{\bar{t}_n} = y_m$, have the same mean and variance.

The choice between the two alternatives – υ_{j_A}, υ_{j_B}, υ_{j_C} or υ_{j_A}, υ_{j_B}, υ_{j_D} – is made in order to reduce the number of points connected with υ_0, which is the starting point. Since Y reverts to 0, it is sufficient, when possible, to choose the set with the points closest to υ_0. Specifically, if $\Delta^A < \frac{3-\sqrt{5}}{2}\sigma_Y^{\Delta t}$, then $\bar{Y}^{\Delta t}$ can only move to υ_{j_A}, υ_{j_B}, υ_{j_C} (in fact at least one among p_A, p_B and p_D is not in [0, 1]). If $\frac{\sqrt{5}}{2}\sigma_Y^{\Delta t} < \Delta^A$, then $\bar{Y}^{\Delta t}$ can only move to υ_{j_A}, υ_{j_B}, υ_{j_D} (in fact at least one among p_A, p_B and p_C is not in [0, 1]). Finally, if $\frac{3-\sqrt{5}}{2}\sigma_Y^{\Delta t} \le \Delta^A \le \frac{\sqrt{5}}{2}\sigma_Y^{\Delta t}$ both choices are admissible: if $\left|\upsilon_{j_C}\right| \le \left|\upsilon_{j_D}\right|$, then $\bar{Y}^{\Delta t}$ moves to $\upsilon_{j_A}, \upsilon_{j_B}, \upsilon_{j_C}$ otherwise to $\upsilon_{j_A}, \upsilon_{j_B}, \upsilon_{j_D}$.

The state space of $\bar{Y}^{\Delta t}$ is the connected component of υ_0, that is the set $\mathcal{Y}_0 \subset \mathcal{Y}$ of points that the process $\bar{Y}^{\Delta t}$ can reach. Taking advantage of the symmetry and mean reversion properties of the process Y, and thus of $\bar{Y}^{\Delta t}$, one can prove that $\mathcal{Y}_0 = \left\{\upsilon_j, j = -N_Y, \ldots, N_Y\right\}$ where N_Y is an integer. Moreover, by exploiting the definition of $\bar{Y}^{\Delta t}$, one can prove that

$$N_Y \le \frac{\left(3-\sqrt{5}\right)e^{k\cdot\Delta t}}{3\left(e^{k\cdot\Delta t}-1\right)} + 1, \tag{A8}$$

thus $N_Y \le \frac{3-\sqrt{5}}{3k} \cdot N_T$ as $N_T \to +\infty$.

Finally, we stress out that $\bar{Y}^{\Delta t}$ matches the first two moments of the process Y, thus weak convergence for $N_T \to +\infty$ is guaranteed and it can be proved as done by Nelson and Ramaswamy (1990).

To conclude, we observe that the set $\mathcal{G}_Y$ defined in (5.5) is equal to $\mathcal{Y}_0$: the only difference concerns the indexing of the elements and in particular $y_j = \upsilon_{j-N_Y}$ for $j \in \{0, \ldots, 2N_Y\}$.

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

APPENDIX B

B.1. Computing expected value (5.8)

B.1.1. *The binomial tree approach.*

In this Appendix, we explain how to efficiently compute the expectation in (5.8). Such a computation can be seen as a particular case of a more general problem: computing

$$E = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_{t_i}^{t_{i+1}} Y_s+\beta(s)ds}\phi\left(Y_{t_{i+1}}, X_{t_{i+1}}, g, h\right)\middle| Y_{t_i}=y, X_{t_i}=x\right], \tag{B1}$$

where $\phi : A \subset \mathbb{R}^4 \to \mathbb{R}$ is a given continuous function. Moreover, t_{i+1} and t_i are two consecutive policy anniversaries times and thus $t_{i+1} - t_i = 1$. Furthermore, $y \in \mathcal{G}_Y$ and $x \in \mathcal{G}_X$.

First of all, let us consider the Gaussian vector $\Lambda = (\Lambda_1, \Lambda_2, \Lambda_3)^\top$ given by

$$\Lambda = \left(Y_{t_{i+1}}, \ln\left(X_{t_{i+1}}\right), \int_{t_i}^{t_{i+1}} Y_s + \beta\left(s\right) ds\right)^\top \left| Y_{t_i} = y, X_{t_i} = x.\right. \tag{B2}$$

It is possible to prove that the mean vector μ of Λ is given by

$$\mu = \left(\mu_1, \mu_2, \mu_3\right)^\top,$$

where

$$\mu_1 = y \cdot e^{-k},\ \mu_2 = \ln\left(x\right) + \mu_3 - \varphi - \frac{1}{2}\sigma^2,\ \mu_3 = y\frac{1-e^{-k}}{k} + \int_{t_i}^{t_{i+1}} \beta\left(s\right) ds. \tag{B3}$$

Moreover, the covariance matrix Π of Λ is given by

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{12} & \Pi_{22} & \Pi_{23} \\ \Pi_{13} & \Pi_{23} & \Pi_{33} \end{pmatrix}, \tag{B4}$$

with

$$\Pi_{11} = \frac{1}{2}\omega^2\frac{(1-e^{-2k})}{k}, \tag{B5}$$

$$\Pi_{33} = \left(\frac{\omega}{k}\right)^2\left(1 + \frac{2e^{-k}}{k} - \frac{e^{-2k}}{2k} - \frac{3}{2k}\right), \tag{B6}$$

$$\Pi_{22} = \Pi_{33} + \sigma^2 + 2\sigma\rho\frac{\omega}{k}\left(1 - \frac{1-e^{-k}}{k}\right), \tag{B7}$$

$$\Pi_{13} = \frac{\omega^2}{2}\left(\frac{1-e^{-k}}{k}\right)^2, \tag{B8}$$

$$\Pi_{12} = \Pi_{13} + \sigma\rho\omega\frac{1-e^{-k}}{k}, \tag{B9}$$

$$\Pi_{23} = \Pi_{33} + \sigma\rho\frac{\omega}{k}\left(1 - \frac{1-e^{-k}}{k}\right). \tag{B10}$$

Let Γ be the lower triangular Cholesky decomposition of Π, and suppose $\tilde{G} = (\tilde{G}_1, \tilde{G}_2, \tilde{G}_3)^\top$ to be a Gaussian standard vector. So, the random vector $\mu + \Gamma\tilde{G}$ has the same law of Λ. Following the same approach of Ekvall (1996) for multidimensional simulation, we can develop a binomial tree method to compute (B1). Such a method exploits three independent binomial approximations of $\tilde{G}_1$, $\tilde{G}_2$ and $\tilde{G}_3$. In particular, we consider the binomial random variable

$$B^N \sim Bi\left(N_T, \frac{1}{2}\right) \tag{B11}$$

and define

$$\hat{G}^{N_T} = \frac{B^{N_T} - \frac{N_T}{2}}{\sqrt{\frac{N_T}{4}}}. \tag{B12}$$

It is well known that $\hat{G}^{N_T}$ converges in distribution to a standard normal distribution so, if $\hat{G}_1^{N_T}, \hat{G}_2^{N_T}, \hat{G}_3^{N_T}$ are i.i.d. random variables that have the same law of $\hat{G}^{N_T}$, then the vector $\hat{\Lambda}^{N_T} = \left(\hat{\Lambda}_1^{N_T}, \hat{\Lambda}_2^{N_T}, \hat{\Lambda}_3^{N_T}\right)^\top$ given by

$$\hat{\Lambda}^{N_T} = \mu + \Gamma\left(\hat{G}_1^{N_T}, \hat{G}_2^{N_T}, \hat{G}_3^{N_t}\right)^\top \tag{B13}$$

converges in distribution to Λ. Let $\left\{\hat{g}^0, \ldots, \hat{g}^{N_T}\right\}$ be the support of $\hat{G}^{N_T}$ and let

$$\hat{p}^m = \mathbb{P}\left[\hat{G}^{N_t} = \hat{g}^l\right] = \binom{N_T}{l}\left(\frac{1}{2}\right)^{N_T} \tag{B14}$$

for $l = 0, \ldots, N_T$ be the associated probabilities. Let ξ_1, ξ_2, ξ_3 be three integers in $\{0, \ldots, N_T\}$ and let $\left(\hat{\lambda}_1^{\xi_1}, \hat{\lambda}_2^{\xi_1,\xi_2}, \hat{\lambda}_3^{\xi_1,\xi_2,\xi_3}\right)^\top$ be the vector defined by

$$\left(\hat{\lambda}_1^{\xi_1}, \hat{\lambda}_2^{\xi_1,\xi_2}, \hat{\lambda}_3^{\xi_1,\xi_2,\xi_3}\right)^\top = \mu + \Gamma\left(\hat{g}^{\xi_1}, \hat{g}^{\xi_2}, \hat{g}^{\xi_3}\right)^\top. \tag{B15}$$

Please observe that, since Γ is lower triangular, $\hat{\lambda}_1^{\xi_1}$ does not depend on $\hat{g}^{\xi_1}$ and $\hat{g}^{\xi_2}$, while $\hat{\lambda}_2^{\xi_1,\xi_2}$ does not depend on $\hat{g}^{\xi_3}$. Moreover

$$\mathbb{P}\left[\hat{\Lambda}_1^N = \hat{\lambda}_1^{\xi_1}, \hat{\Lambda}_2^N = \hat{\lambda}_2^{\xi_1,\xi_2}, \hat{\Lambda}_3^N = \hat{\lambda}_3^{\xi_1,\xi_2,\xi_3}\right] = \hat{p}^{\xi_1} \cdot \hat{p}^{\xi_2} \cdot \hat{p}^{\xi_3}. \tag{B16}$$

In order to approximate (B1), we replace $Y_{t_{i+1}}, \ln\left(X_{t_{i+1}}\right)$ and $\int_{t_i}^{t_{i+1}} Y_s + \beta(s)\,ds$ with $\hat{\Lambda}_1^N$, $\hat{\Lambda}_2^N$ and $\hat{\Lambda}_3^N$, respectively. We obtain

$$\hat{E} = \mathbb{E}\left[e^{-\hat{\Lambda}_3^{N_T}}\phi\left(\hat{\Lambda}_1^{N_T}, \hat{\Lambda}_2^{N_T}, g, h\right)\right] \tag{B17}$$

$$= \sum_{\xi_1=0}^{N_T}\sum_{\xi_2=0}^{N_T}\sum_{\xi_3=0}^{N_T}\hat{p}^{\xi_1}\hat{p}^{\xi_2}\hat{p}^{\xi_3}\exp\left(-\hat{\lambda}_3^{\xi_1,\xi_2,\xi_3}\right)\phi\left(\hat{\lambda}_1^{\xi_1}, \exp\left(\hat{\lambda}_2^{\xi_1,\xi_2}\right), g, h\right) \tag{B18}$$

$$= \sum_{\xi_1=0}^{N_T}\hat{p}^{\xi_1}\sum_{\xi_2=0}^{N_T}\hat{p}^{\xi_2}\phi\left(\hat{\lambda}_1^{\xi_1}, \exp\left(\hat{\lambda}_2^{\xi_1,\xi_2}\right), g, h\right)\sum_{\xi_3=0}^{N_T}\hat{p}^{\xi_3}\exp\left(-\hat{\lambda}_3^{\xi_1,\xi_2,\xi_3}\right) \tag{B19}$$

Such an expression converges to E thanks to the properties of convergence in distribution for expected values (see for example Pollard, 2012). Please observe that, by leaving the random variable $\int_{t_i}^{t_{i+1}} Y_s + \beta(s)\, ds$ as the third component in Λ, the two variables $\hat{\lambda}_1^{\xi_1}$ and $\hat{\lambda}_2^{\xi_1,\xi_2}$ do not depend on ξ_3. So, in order to evaluate (B1), the function ϕ needs to be evaluated only $(N_T)^2$ times in place of $(N_T)^3$. This is a relevant improvement because evaluating the function ϕ many times can be time demanding. Moreover, if the function ϕ is known only on the grid $\mathcal{G}$ – this is what happens for the functions $\bar{\mathcal{V}}^-$ and $\bar{\mathcal{U}}^-$ – then a two-dimensional interpolation is required.

Finally, it is well known that a vector of random independent binomial variables, suitably standardized, converges weakly to a vector of independent standard Gaussian variables (see Lehmann, 2004). This property, together with the assumption of growth at the most exponential of the integrand function with respect to the random variables $\Lambda_1, \Lambda_2, \Lambda_3$, guarantees the convergence of $\hat{E}$ to the expected value E (see Van den Berg, 2000).

B.1.2. *Improving computational efficiency.*

The Markov chain $\bar{Y}^{\Delta t}$ introduced in Appendix A does not only provide a way to define the set $\mathcal{G}_Y$ but it can also be used to improve the evaluation of (B1). Suppose now y in (B1) to be equal to $y_m \in \mathcal{G}_Y$ for a particular integer m. Let $n = N_T \cdot i$ so that $\bar{t}_n = t_i$ and $\bar{t}_{n+N_T} = t_{i+1}$. In order to improve the discretisation of the random variable $Y_{t_{i+1}} \mid Y_{t_i} = y_m$, we replace $\hat{\Lambda}_1^{N_T}$ in (B17) with $\bar{Y}^{\Delta t}_{n+N_T} \mid \bar{Y}^{\Delta t}_n = y_m$. The transition probabilities

$$\bar{p}_{m,l} = P\left(\bar{Y}^{\Delta t}_{n+N_T} = y_l \,\middle|\, \bar{Y}^{\Delta t}_n = y_m\right) \tag{B20}$$

for $m, l \in \{0, \ldots, 2N_Y\}$ can be obtained by computing the N_T-power of transition matrix of $\bar{Y}^{\Delta t}$, whose elements are determined according to Table A.1. Finally, we conclude by observing that the support of the random variable $\bar{Y}^{\Delta t}_{n+N_T} \mid \bar{Y}^{\Delta t}_n = y_m$ is a subset of $\mathcal{G}_Y$ for every value y_m in $\mathcal{G}_Y$. We stress out that the support of the random variable $\hat{\Lambda}_1^{N_T}$ has $N_T + 1$ elements, while the support of $\bar{Y}^{\Delta t}_{n+N_T} \mid \bar{Y}^{\Delta t}_n = y_m$ has at most $2N_Y + 1$ elements. Numerical tests show that $2N_Y + 1$ is usually smaller than $N_T + 1$, so computational efficiency is improved: for example, with respect to our tests in Section 6, we have $N_T + 1 = 51$ and $2N_Y + 1 = 27$.

The Markov chain $\bar{Y}^{\Delta t}$ helps us to discretize the process Y, but in order to compute (B19), we also have to simulate the whole random vector $\hat{\Lambda}^{N_T}$ in (B2). To do so, we have to compute the normal Gaussian increments associated with the transitions of $\bar{Y}^{\Delta t}$. Let us define the discrete random variable $\bar{G}^{N_T}$ as the standard score of $\bar{Y}^{\Delta t}_{n+N_T} \mid \bar{Y}^{\Delta t}_n = y_m$, that is

$$\bar{G}^{N_T} = \frac{\bar{Y}^{\Delta t}_{n+N_T} - \mu_Y^1(y_m)}{\sigma_Y^1}. \tag{B21}$$

https://doi.org/10.1017/asb.2020.29 Published online by Cambridge University Press

Since $\bar{Y}^{\Delta t}_{n+N_T} \,|\, \bar{Y}^{\Delta t}_n = y_m$ matches the first two moment of the random Gaussian variable $Y_{t_{i+1}} \,|\, Y_{t_i} = y_m$, then $\bar{G}^{N_T}$ matches the first two moments of a standard Gaussian variable and so it can be employed in place of $\hat{G}^{N_T}_1$. Moreover, $\mu^1_Y(y_m) = \mu_1$ and $\left(\sigma^1_Y\right)^2 = \Pi_{11}$. Then, we define the vector $\bar{\Lambda}^{N_T} = \left(\bar{\Lambda}^{N_T}_1, \bar{\Lambda}^{N_T}_2, \bar{\Lambda}^{N_T}_3\right)^\top$ given by

$$\bar{\Lambda}^{N_T} = \mu + \Gamma\left(\bar{G}^{N_T}_1, \hat{G}^{N_T}_2, \hat{G}^{N_t}_3\right)^\top, \tag{B22}$$

which converges to Λ and in particular $\bar{\Lambda}^{N_T}_1 = \bar{Y}^{\Delta t}_{n+N_T} \,|\, \bar{Y}^{\Delta t}_n = y_m$. Let ξ_1, ξ_2, ξ_3 be three integers such that ξ_1 is in $\{0, \ldots 2N_Y\}$ and ξ_2, ξ_3 are in $\{0, \ldots, N_T\}$. Let $\left(\bar{\lambda}^{\xi_1}_1, \bar{\lambda}^{\xi_1,\xi_2}_2, \bar{\lambda}^{\xi_1,\xi_2,\xi_3}_3\right)^\top$ be the vector defined by

$$\left(\bar{\lambda}^{\xi_1}_1, \bar{\lambda}^{\xi_1,\xi_2}_2, \bar{\lambda}^{\xi_1,\xi_2,\xi_3}_3\right)^\top = \mu + \Gamma\left(\bar{g}^{\xi_1}, \hat{g}^{\xi_2}, \hat{g}^{\xi_3}\right)^\top, \tag{B23}$$

where $\left\{\bar{g}^0, \ldots \bar{g}^{2N_Y}\right\}$ is the support of $\bar{G}^{N_T}$. Please note that $\bar{\lambda}^{\xi_1}_1$ is equal to y_{ξ_1} which is in $\mathcal{G}_Y$. Thus, we obtain the following approximation of E, based on the Markov chain $\bar{Y}^{\Delta t}$:

$$\bar{E} = \sum_{\xi_1=0}^{2N_Y} \bar{p}_{m,\xi_1} \sum_{\xi_2=0}^{N_T} \hat{p}^{\xi_2} \phi\left(y_{\xi_1}, \exp\left(\bar{\lambda}^{\xi_1,\xi_2}_2\right), g, h\right) \sum_{\xi_3=0}^{N_T} \hat{p}^{\xi_3} \exp\left(-\bar{\lambda}^{\xi_1,\xi_2,\xi_3}_3\right). \tag{B24}$$

Because of the Markov chain, $\bar{Y}^{\Delta t}$ matches the first two moments of the process Y, weak convergence with respect to $\bar{G}^{N_T}_1$ is guaranteed (see Nelson and Ramaswamy, 1990). Moreover, the weak convergence of the whole discretization is guaranteed by weak convergence on each variable $\bar{G}^{N_T}_1, \hat{G}^{N_T}_2, \hat{G}^{N_t}_3$ to standard Gaussian variables and by their independence.

We conclude by observing that Equation (B24) has one important advantage over Equation (B19), that improves computational efficiency when the function ϕ is known only at the points of $\mathcal{G}$. Specifically, the computation of (B19) requires a two-dimensional interpolation to evaluate the function ϕ outside $\mathcal{G}$, while (B24) requires only a one-dimensional interpolation because, as opposed to $\hat{\lambda}^{\xi_1}_1$, y_{ξ_1} is an element of $\mathcal{G}_Y$. To this aim, we employ one-dimensional cubic spline interpolation, which is very fast and accurate.