A Unified Pricing of Variable Annuity Guarantees under the Optimal Stochastic Control Framework

Pavel V. Shevchenko * and Xiaolin Luo

The Commonwealth Scientific and Industrial Research Organisation, PO BOX 52, North Ryde, NSW 1670, Australia; xiaolin.luo@csiro.au
* Correspondence: pavel.shevchenko@csiro.au; Tel.: +61-2-9325-3218

Abstract: In this paper, we review pricing of the variable annuity living and death guarantees offered to retail investors in many countries. Investors purchase these products to take advantage of market growth and protect savings. We present pricing of these products via an optimal stochastic control framework and review the existing numerical methods. We also discuss pricing under the complete/incomplete financial market models, stochastic mortality and optimal/sub-optimal policyholder behavior, and in the presence of taxes. For numerical valuation of these contracts in the case of simple risky asset process, we develop a direct integration method based on the Gauss-Hermite quadratures with a one-dimensional cubic spline for calculation of the expected contract value, and a bi-cubic spline interpolation for applying the jump conditions across the contract cashflow event times. This method is easier to implement and faster when compared to the partial differential equation methods if the transition density (or its moments) of the risky asset underlying the contract is known in closed form between the event times. We present accurate numerical results for pricing of a Guaranteed Minimum Accumulation Benefit (GMAB) guarantee available on the market that can serve as a numerical benchmark for practitioners and researchers developing pricing of variable annuity guarantees to assess the accuracy of their numerical implementation.

Keywords: variable annuity; guaranteed living and death benefits; guaranteed minimum accumulation benefit; optimal stochastic control; direct integration method

1. Introduction

Many wealth management and insurance companies worldwide are offering investment products known as *variable annuities* (VA) with some guarantees of living and death benefits to assist investors with managing their pre-retirement and post-retirement plans. These products take advantage of market growth while providing a protection of the savings against the market downturns. Insurers started to offer these products from the 1990s in the United States. Later, these products became popular in Europe, UK and Japan and more recently in Australia. The VA contract cashflows received by the policyholder are linked to the investment portfolio choice and performance (e.g., the choice of mutual fund and its strategy) while traditional annuities provide a pre-defined income stream in exchange for the lump sum payment. According to LIMRA (Life Insurance and Market Research Association) reports, the VA market is huge: VA sales in United States were $158 billion in 2011, $147 billion in 2012 and $145 billion in 2013.

The types of VA guarantees (referred in the literature as *VA riders*) offered for investment portfolios are classified as *guaranteed minimum withdrawal benefit* (GMWB), *guaranteed minimum accumulation benefit* (GMAB), *guaranteed minimum income benefit* (GMIB) and *guaranteed minimum death benefit* (GMDB). These guarantees, generically denoted as GMxB, provide different types of protection against the market downturns and policyholder death. GMWB allows the withdrawing of funds from the VA account up

to some pre-defined limit regardless of the investment performance during the contract; GMAB and GMIB both provide a guaranteed investment account balance at the contract maturity that can be taken as a lump sum or standard annuity respectively. *Guaranteed lifelong withdrawal benefit* (GLWB), a specific type of GMWB, allows withdrawing funds at the contractual rate as long as the policyholder is alive. GMDB provides a specified payment if the policyholder dies. Precise specifications of the products within each type can vary across companies and some products may include combinations of these guarantees.

A good overview of the VA products and the development of their market can be found in [1–3]. There have been a number of papers in the academic literature considering pricing of these products. Most of these are focused on pricing VA riders under the pre-determined (*static*) policyholder behaviour in withdrawal and surrender. Some studies include pricing under the active (*dynamic*) strategy when the policyholder '*optimally*' decides the amount of withdrawal at each withdrawal date depending on the information available at that date. Standard Monte Carlo (MC) method can easily be used to estimate price in the case of pre-defined withdrawal strategy but handling the dynamic strategy requires a backward in time solution that can be done only via the partial differential equation (PDE), direct integration or regression type MC methods.

In brief, pricing under the static and dynamic withdrawal strategies via the PDE based methods has been developed in [4–6]. The authors of [1] develop a unified approach with numerical estimation via MC and direct integration methods. The direct integration method was developed further in [7,8] using the Gauss-Hermite quadrature and cubic interpolations. The authors of [9] consider many VA riders under the stochastic interest rate and stochastic volatility if the policyholder withdraws at the pre-defined contractual rate or completely surrenders the contract. Their pricing is accomplished either by the ordinary MC or Least-Squares MC to account for the optimal surrender. Often, pricing of the VA riders is considered under the assumption of a geometric Brownian motion for the risky asset underlying the contract, though a few papers looked at extensions such as stochastic interest rate and/or stochastic volatility, see, e.g., [9–12].

The authors of [13] prove the existence of an optimal *bang-bang* control for GLWB contract when thecontract holder can maximize contract writer's losses by only ever performing non-withdrawal, withdrawal at the contract rate or full surrender. However, they also demonstrate that the related GMWB contract does not satisfy the bang-bang principle other than in certain degenerate cases. The authors of [11] developed a regression-based MC method for pricing GLWB under the bang-bang strategy in the case of stochastic volatility. GMWB pricing under the bang-bang strategy was studied in [14]. The difficulty with applying the well known Least-Squares MC introduced in [15] for pricing VA riders under the optimal strategy is due to the fact that the paths of the underlying VA wealth account are affected by the withdrawals. In principle, one can apply the *control randomization* methods extending Least-Squares MC to handle optimal stochastic control problems with controlled Markov processes recently developed in [16], but the accuracy and robustness of this method for pricing of the VA riders have not been studied yet.

One common observation in the above mentioned literature is that pricing under the optimal strategy often leads to prices significantly higher than observed on the market. These studies rely on the option pricing risk-neutral methodology in quantitative finance to find a *fair fee*. Here, the fundamental idea is to find the cost of a dynamic self-financing replicating portfolio which is designed to provide an amount at least equal to the payoff of the contract. The cost of establishing this hedging strategy is the no-arbitrage price of the contract. This is under the assumption that the contract holder adopts an optimal strategy (exercise strategy maximising the monetary value of the contract). If the purchaser follows any other exercise strategy, the contract writer will generate a guaranteed profit if continuous hedging is performed. Of course the strategy optimal in this sense is not related to the policyholder circumstances. In pricing VA with guarantees, it is reasonable to consider alternative assumptions regarding the investor's withdrawal strategy. This is because an investor may follow what appears to be a *sub-optimal* strategy that does not maximise the monetary value of the option. This could be

due to reasons such as liquidity needs, tax and other personal circumstances. Moreover, the mortality risk is diversified by the contract issuer through selling many contracts while the policyholder cannot do it. Also, there might be no liquid secondary market for VAs on which the policy could be sold (or repurchased) at its fair value. The policyholder may act optimally with respect to his preferences and circumstances but it may be different from the optimal strategy that maximises the monetary value of the contract. In this case we calculate a fair fee to be deducted in order to finance a dynamic replicating portfolio for the guarantees (options) embedded in the contract under the assumption of a particular exercise strategy. The replicating portfolio will provide sufficient funds to meet any future payouts that arise from writing the contract.

However, the fair fee obtained under the assumption that investors behave optimally to maximise the value of the guarantee does offer an important benchmark because it is a worst case scenario for the contract writer. Also, as noted in [17], secondary markets for the equity linked insurance products (where the policyholder can sell their contracts) are growing. Thus, third parties can potentially generate a guaranteed profit through hedging strategies from the financial products such as VA riders which are not priced under the assumption of the optimal withdrawal strategy. The authors of [18] mention several companies recently suffering large losses related to increased surrender rates, indicating that either the charged fees were not sufficiently large or that the hedging program did not perform as expected.

One way to analyze the withdrawal behavior of a VA holder and evaluate the need of these products is to solve the life-cycle utility model accounting for consumption, housing, bequest and other real life circumstances. Developing a full life-cycle model with all preferences and required parameters is challenging but there are already several contributions reporting some interesting findings in this direction: [19–22]. This topic will not be considered in this paper. It is also important to note a recent paper by [23] considering the pricing under the optimal strategy in the presence of taxes via *subjective risk-neutral* valuation methodology. They demonstrated that including taxes significantly affects the value of the VA withdrawal guarantees producing results in line with empirical market prices. This approach will be discussed in Section 5.

Often it is assumed that insurance company can diversify the mortality risk through selling many contracts and the financial risk can be completely hedged using the underlying financial asset. Then the fair price of the product is calculated as the expected present value of the product benefits with respect to the risk-neutral process for the financial risk and the real process for the mortality. In practice, one has to account for a systemic mortality risk that cannot be diversified (often modelled through stochasticity of the death probabilities). If the assets for perfect hedging of the mortality risk would exist then one could calculate the unique fair price as expectation with respect to a risk-neutral process for mortality. However, the current status of the insurance market does not allow investment into such assets. There will be a residual risk for the contract issuer not only due to systemic mortality risk but also due to the finite size of the portfolio and imperfections of the hedging with respect to the financial risk. In this case, one can calculate extra loading to the fair price so that the issuer residual risk (quantified as, e.g., Value-at-Risk or expected shortfall at some confidence level) is no greater than zero. This can be accomplished via simulations of the hedging strategy based on the fair price valuation and quantification of the residual hedging errors. The extra loading to cover the residual risk depends on the risk management strategy for the product and was considered in [24,25]. This topic will be briefly discussed in Section 3.

In this paper we review pricing of living and death benefit guarantees offered with VAs, and present an optimal stochastic control framework for pricing these contracts. The main ideas have been developed and appeared in some forms in a number of other papers. However, we believe that our presentation is easier to understand and implement. We also present a direct integration method based on computing of the expected contract values in a backward time-stepping through the high order Gauss-Hermite integration quadrature applied on the cubic spline interpolation. This method can be applied when transition density of the underlying asset between the contract cashflow event dates

or its moments are known in closed form. We have used this for pricing specific financial derivatives and some simple versions of the VA guarantees in [7,26]. Here, we adapt and extend the method to handle pricing VA riders in general. As a numerical example, we calculate accurate prices of GMAB with possible annual *ratchets* (reset of the guaranteed capital to the investment portfolio value if the latter is larger on the contract anniversary dates) and allowing optimal withdrawals. The contract that we consider is very similar in specifications to the real product marketed in Australia, see for example [27,28]. Numerical difficulties encountered in pricing this VA rider are common across other VA guarantees and at the same time comprehensive numerical pricing results for this product are not available in the literature. These results (reported for a range of parameters) can serve as a numerical benchmark for practitioners and researchers developing pricing of the VA riders to assess the accuracy of their numerical implementations.

In the next section, a general specification of VA riders is given. In Section 3 we discuss different stochastic models used for pricing these products including complete/incomplete financial market models and stochastic mortality. Section 4 provides precise specification for some popular VA riders. In Section 5 we present calculation of the fair price and fair fee as a solution of an optimal stochastic control problem and discuss pricing under the optimal/sub-optimal policyholder behaviour and in the presence of taxes. Section 6 reviews the numerical methods and algorithms for pricing VA riders. In Section 7 we present numerical results for the fair fees of GMAB rider. Concluding remarks are given in Section 8.

2. VA Rider Contract Specification

Consider a VA contract with some guarantees for living and death benefits purchased by an *x*-year old individual at time $t_0 = 0$ with the up-front premium invested in a risky asset (e.g., a mutual fund), denoted as $S(t)$ at time $t \geq 0$. The VA rider specification includes dates when events such as withdrawal, ratchet (*step-up*), bonus (*roll-up*), death benefit payment, etc. may occur. Precise definitions of these events depend on the contract and corresponding examples will be provided in Section 4. The contract holder is allowed to take withdrawals and we assume that these can only take place on the set of the ordered *event times* $\mathcal{T} = \{t_1, \ldots, t_N\}$, where $T = t_N$ is the contract maturity (in the case of lifelong guarantees T corresponds to the maximum age beyond which survival is deemed impossible). Denote the withdrawal amount at time $t_n \in \mathcal{T}$ as γ_n. Also, the set of policy anniversaries when the ratchet event may occur is denoted as $\mathcal{T}_r$ and is assumed to be a subset of $\mathcal{T}$. For simplicity of notation we assume that all other events may only occur on the withdrawal dates. The value of the contract payments at time t_n is determined by three state variables: wealth account $W(t_n)$, guarantee account $A(t_n)$ and life status indicator I_n.

- *Wealth account $W(t)$* is value of the investment account which is linked to the risky asset $S(t)$. Initially, $W(0)$ is set equal to the upfront premium. For a given process of risky asset $S(t)$, $t \geq 0$, the value of the wealth account $W(t)$ evolves as

$$
\begin{aligned}
W(t_n^-) &= \frac{W(t_{n-1}^+)}{S(t_{n-1})} S(t_n) e^{-\alpha dt_n}, \quad n = 1, 2, \ldots, N \\
W(t_n^+) &:= h_n^W \left(W(t_n^-), A(t_n^-), \gamma_n \right) = \max(W(t_n^-) - \gamma_n, 0)
\end{aligned}
\tag{1}
$$

where $dt_n = t_n - t_{n-1}$ and α is the annual fee continuously charged by the contract issuer for the provided guarantee. Here, we denote the values of the wealth account just before and just after t_n as $W(t_n^-)$ and $W(t_n^+)$ respectively, and similarly for other variables. The function $h_n^W(\cdot)$ is introduced to have a generic notation covering cases of discretely charged fees discussed later in this section.

- *Guarantee account $A(t)$* is the value used to calculate benefits provided by the VA rider, also referred in the literature as *benefit base*. It is not changing between the event times but can be stochastic via

stochasticity in $W(t)$ at the event times depending on the contract features. Initially, $A(0)$ is set equal to the upfront premium. At the event time, the benefit base may change as

$$A(t_n^+) := h_n^A\left(W(t_n^-), A(t_n^-), \gamma_n\right) \tag{2}$$

where the function $h_n^A(\cdot)$ is determined by the contract specifications and may also depend on the fee, penalty and other parameters. Specific cases will be provided in Section 4. For example, if only a *ratchet* is possible at $t_n \in \mathcal{T}_r$ and no other contract events, then

$$A(t_n^+) = \max\left(A(t_n^-), W(t_n^-) \times 1_{t_n \in \mathcal{T}_r}\right)$$

Also, if there is only withdrawal at time t_n, then $A(t_n^+) = \max\left(A(t_n^-) - \gamma_n, 0\right)$. In practice, several events such as withdrawal, ratchet, bonus, etc. may occur at the same time, and the contract specification determines the order of these events.

- Life status indicator I_n is a discrete random variable with the states in the set $\mathcal{G} = \{1, 0, -1\}$ corresponding to the policyholder is being alive at t_n, died during $(t_{n-1}, t_n]$, or died before or at t_{n-1} respectively. This variable is needed if there are death benefit features in the contract. It can be extended to include extra states corresponding to the policyholder partner if there is a spouse continuation feature in the contract. This case will not be considered explicitly but is easy to add. One could write the contract payoff function using the random variable τ representing lifetime of the policyholder, but instead we introduce I_n to have a payoff written in the standard form for the optimal stochastic control framework.

An extra state variable is also required to track a tax free base to account for taxes; this will be considered in Section 5.4. In principle, different guarantees included in VA may have different benefit base state variables. For notational simplicity and also from a practical perspective, we assume that all guarantees in VA are linked to the same benefit base account.

The contract product specification also determines:

- the contractual (guaranteed) withdrawal amount G_n for the period $(t_{n-1}, t_n]$ that may depend on the benefit base $A(t_n^-)$ and/or $W(t_n^-)$;
- possible range for withdrawal $\gamma_n \in \mathcal{A}_n\left(W(t_n^-), A(t_n^-)\right)$;
- the payout $P_T(W, A)$ at the contract maturity if policyholder is alive at $t = T$;
- the payout $D_n(W, A)$ to the beneficiary at t_n in the case of the policyholder death during $(t_{n-1}, t_n]$, $n = 1, \ldots, N$;
- the cashflow received by the policyholder $\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n)$ at the event times t_n, $n = 1, \ldots, N-1$, that might be different from γ_n due to penalties.

The specification details typically vary across different companies and are difficult to extract from the very long product specification documents. Moreover, results for specific GMxB riders presented in academic literature often refer to different specifications.

Denote the state variable vector at time t_n before the withdrawal as $X_n = (W(t_n^-), A(t_n^-), I_n)$ and $X = (X_1, \ldots, X_N)$. Then, given the withdrawal strategy $\gamma = (\gamma_1, \ldots, \gamma_{N-1})$, the present value of the overall payoff of the VA contract with a guarantee can be written as

$$H_0(X, \gamma) = B_{0,N} H_N(X_N) + \sum_{n=1}^{N-1} B_{0,n} f_n(X_n, \gamma_n) \tag{3}$$

Here,

$$H_N(X_N) = P_T\left(W(T^-), A(T^-)\right) \times 1_{I_N=1} + D_N\left(W(T^-), A(T^-)\right) \times 1_{I_N=0} \tag{4}$$

is the cashflow at the contract maturity, and

$$f_n(X_n, \gamma_n) = \widetilde{f}_n\left(W(t_n^-), A(t_n^-), \gamma_n\right) \times 1_{I_n=1} + D_n\left(W(t_n^-), A(t_n^-)\right) \times 1_{I_n=0} \tag{5}$$

is the cashflow at time t_n. Also, $B_{i,j}$ is the discounting factor from t_j to t_i

$$B_{i,j} = \exp\left(-\int_{t_i}^{t_j} r(t)dt\right), \; t_j > t_i \tag{6}$$

where $r(t)$ is the risk-free interest rate (possibly stochastic).

In practice, the guarantee fee is charged discretely and proportional to the wealth account that requires some changes to the wealth process (1). Denoting the discretely charged fee with the annual basis as $\widetilde{\alpha}$, the wealth process becomes

$$\begin{aligned}
W(t_n^-) &= \frac{W(t_{n-1}^+)}{S(t_{n-1})}S(t_n), \; n = 1, 2, \ldots, N \\
W(t_n^+) &:= h_n^W\left(W(t_n^-), A(t_n^-), \gamma_n\right) = \max\left(W(t_n^-)(1 - \widetilde{\alpha}dt_n) - \gamma_n, 0\right)
\end{aligned} \tag{7}$$

Typically, the difference between continuously and discretely charged fees is not material as observed in our numerical results given in Section 7.

Another popular fee structure corresponds to fees charged as a proportion of the benefit base, so that

$$\begin{aligned}
W(t_n^-) &= \frac{W(t_{n-1}^+)}{S(t_{n-1})}S(t_n), \; n = 1, 2, \ldots, N \\
W(t_n^+) &:= h_n^W\left(W(t_n^-), A(t_n^-), \gamma_n\right) = \max\left(W(t_n^-) - A(t_n^-)\widetilde{\alpha}dt_n - \gamma_n, 0\right)
\end{aligned} \tag{8}$$

Here, it is assumed that discrete fees are deducted before the withdrawal but it can be vice versa depending on the contract specifications.

Once the above conditions, i.e., functions $h_n^W(\cdot)$, $h_n^A(\cdot)$, $P_T(\cdot)$, $P_D(\cdot)$, $\widetilde{f}_n(\gamma_n)$ and admissible range for withdrawal $\mathcal{A}_n$ are specified by the contract design, and a specific stochastic evolution of the financial risky asset $S(t)$ and death indicator I_n is assumed, then pricing of the contract can be accomplished by numerical methods. In particular, if the withdrawals are optimal then pricing can be accomplished by the PDE, direct integration or regression based MC methods. If the withdrawals are pre-defined, then the standard MC along with PDE and direct integration methods can be used. The use of a particular numerical technique is determined by the complexities of the underlying stochastic model and contract details.

3. Stochastic Model

Commonly in the literature, stochastic models for the financial risky asset $S(t)$ underlying the VA rider assume that there is no arbitrage in the financial market which means that there is a risk-neutral measure $\mathbb{Q}$ under which payment streams can be valued as expected discounted values. Moreover, this means that the cost of a portfolio replicating the contract is given by its expected discounted value under $\mathbb{Q}$. Hence, the fair price of the contract can be expressed as an expectation of the contract discounted cashflows with respect to $\mathbb{Q}$. Some models (such as Black-Scholes model) considered in the literature are *complete* which means that the risk-neutral measure $\mathbb{Q}$ is unique. Other models (such as extending Black-Scholes model to stochastic volatility and/or stochastic interest rate) are *incomplete* implying that $\mathbb{Q}$ is not unique. It is also assumed that the financial market has a risk-free asset that accumulates continuously at the risk-free interest rate. These are typical assumptions in the academic research literature on pricing financial derivatives, for a good textbook in this area we refer the reader to, e.g., [29].

A benchmark model commonly considered in the literature on pricing VA riders is the well-known Black-Scholes dynamics for the reference portfolio of assets $S(t)$ that under the risk-neutral measure $\mathbb{Q}$ is known to be

$$dS(t) = r(t)S(t)dt + \sigma(t)S(t)d\mathcal{B}(t) \tag{9}$$

Here, $\mathcal{B}(t)$ is the standard Wiener process, $r(t)$ is deterministic time dependent risk-free interest rate and $\sigma(t)$ is deterministic time dependent volatility. Under this model the financial market is complete. The solution for this process is just

$$S(t_n) = S(t_{n-1})e^{(r_n - \frac{1}{2}\sigma_n^2)dt_n + \sigma_n\sqrt{dt_n}z_n}$$

where $z_1, \ldots, z_N$ are independent and identically distributed standard Normal random variables, $r_n = \int_{t_{n-1}}^{t_n} r(u)du/dt_n$ and $\sigma_n^2 = \int_{t_{n-1}}^{t_n} \sigma^2(u)du/dt_n$.

Many models have been developed in the literature of pricing financial derivatives extending the Black-Scholes dynamics to the stochastic interest rate and/or stochastic volatility. These are incomplete models where the risk-neutral measure $\mathbb{Q}$ is not unique. It is impossible to review all these models and we refer the reader to the standard textbooks in this area such as [30] or [31]. Some of these models have been applied for pricing VA riders that can be represented as the following risk-neutral diffusion process

$$\begin{aligned}
dr(t) &= \xi_r(\zeta_r - r(t))dt + \sigma_r(r(t))^{a_1}d\mathcal{B}^r(t) \\
dv_t &= \xi_v v_t^{a_2}(\zeta_v - v_t)dt + \sigma_v v_t^{a_3}d\mathcal{B}^v(t), \quad v_t = \sigma^2(t) \\
&\text{with} \quad a_1 = \{0, 1/2\}, \quad a_2 = \{0, 1\}, \quad a_3 = \{1/2, 1, 3/2\}
\end{aligned} \tag{10}$$

where $(\xi_r, \zeta_r, \sigma_r, \xi_v, \zeta_v, \sigma_v)$ are the model parameters, and $\mathcal{B}^r(t)$ and $\mathcal{B}^v(t)$ are the Wiener processes possibly correlated with $\mathcal{B}(t)$. In particular, [9] studied the case of stochastic interest rate and stochastic volatility corresponding to $a_1 = 1/2$, $a_2 = 0$, $a_3 = 1/2$. In [11], the authors considered the case of deterministic interest rate and stochastic volatility with $a_2 = \{0, 1\}$ and $a_3 = \{1/2, 1, 3/2\}$. In [12], the authors developed solution for deterministic volatility and stochastic interest rate with $a_1 = 0$. An alternative way to modelling stochasticity in the interest rate and volatility via diffusion processes is to consider the Markov regime switching models. Under this approach, the interest rate and volatility are assumed to have the finite number of possible values and their evolution in time is driven by the finite state Markov chain variable representing possible regimes of the economy. In the context of pricing VA riders under the optimal withdrawal strategy, this approach was used in, e.g., [10].

The mortality risk is assumed to be independent from the financial risk. A standard way to model mortality is to introduce the *force of mortality* $\lambda(t)$ for a person aged x at time $t = 0$ which may be stochastic or deterministic, such that the person lifetime τ satisfies

$$\Pr[\tau > t' | \tau > t; \lambda(y), t \leq y \leq t'] = e^{-\int_t^{t'} \lambda(y)dy}$$

For calculation of the VA rider payoff we need evolution of a life status indicator variable I_n with the states in the set $\mathcal{G} = \{1, 0, -1\}$ corresponding to the policyholder is being alive at t_n, died during $(t_{n-1}, t_n]$, or died before or at t_{n-1} correspondingly. Then the conditional transition probabilities $q(g_{n-1}, g_n) := \Pr[I_n = g_n | I_{n-1} = g_{n-1}; \lambda(y), t_{n-1} \leq y \leq t_n]$ with $g_{n-1}, g_n \in \mathcal{G}$ can be written as

$$\begin{aligned}
q_n &:= q(1, 0) = 1 - e^{-\int_{t_{n-1}}^{t_n} \lambda(y)dy}, \quad q(1, 1) = 1 - q_n \\
q(1, -1) &= q(0, 1) = q(0, 0) = q(-1, 1) = q(-1, 0) = 0, \quad q(0, -1) = q(-1, -1) = 1
\end{aligned} \tag{11}$$

If $\lambda(t)$ is modelled as stochastic, then the transition probabilities q_n become stochastic due to stochasticity of $\lambda(t)$ and a random variable I_n becomes a *doubly stochastic* random variable. It is common to assume that $\lambda(t)$ is a Markov process. Then the survival function (probability of surviving from time t to t') given the knowledge of the mortality force at time t is

$$p(t, t', \lambda(t)) := \mathrm{E}\left[e^{-\int_t^{t'} \lambda(y)dy} \middle| \lambda(t)\right]$$

There are many mortality models developed in the actuarial literature starting from the Gompertz deterministic model $\lambda(x) = ae^x/b$ introduced in 1825, for a textbook treatment of this topic see, e.g., [32]. If $\lambda(t)$ is deterministic, the mortality risk can be fully diversified via selling the contract to many policyholders, i.e., the contract price averaged over L independent policyholders of the same age converges to the expected value of the contract as L increases. The standard way to estimate deterministic death probabilities q_n is to use the official Life Tables that provide annual death probabilities for each age and gender in a given country; probabilities for time periods within a year are typically estimated assuming constant force of mortality within a year.

Stochastic modelling of $\lambda(t)$ is a more recent development starting from the Lee-Carter model introduced in [33]. Many stochastic mortality models extending the Lee-Carter model have been developed to capture the period, cohort and stochastic volatility features, see, e.g., [34–36]. The stochasticity of $\lambda(t)$ will introduce some dependence between the lifetimes of the policyholders leading to the systemic (undiversified) mortality risk. For pricing VA riders, the force of mortality is often model by a real diffusion process

$$d\lambda(t) = \mu_\lambda(\lambda(t), t)dt + \sigma_\lambda(\lambda(t), t)d\mathcal{B}_\lambda^*(t) \tag{12}$$

where $\mathcal{B}_\lambda^*(t)$ is the Wiener process and some specific forms for the drift $\mu_\lambda(\cdot)$ and volatility $\sigma_\lambda(\cdot)$ are considered in, for example, [9,24]. If we need to model a portfolio of the contracts, then the dependence between the policyholder deaths can be introduced via models for the number of deaths in a given population driven by some common factors as in the Lee-Carter model and its extensions reviewed in, e.g., [34], or under the CreditRisk+ framework for mortality developed recently in [37,38].

Note that the current status of the insurance market does not allow to invest into the assets to hedge the systemic mortality introducing incompleteness in the model. In this case, pricing can be accomplished by either the financial mathematics no-arbitrage approach or the actuarial principle. Under the financial mathematics no-arbitrage approach, ([29] (proposition 15.3)), we find the risk-neutral mortality process corresponding to the real process (12) to be

$$d\lambda(t) = \left\{ \mu_\lambda(\lambda(t), t) - \zeta(\lambda(t), t))\sigma_\lambda(\lambda(t), t) \right\} dt + \sigma_\lambda(\lambda(t), t)d\mathcal{B}^\lambda(t) \tag{13}$$

where $\zeta(\lambda(t), t))$ is the risk premium and $\mathcal{B}^\lambda(t)$ is the Wiener process under the risk-neutral measure. Then the fair price is calculated as an expectation with respect to the risk-neutral measure both for the financial asset and mortality processes. Of course, the risk premium is unknown and should be found using extra condition such as the requirement for the risk of hedging error loss (measured, e.g., as Value-at-Risk or expected shortfall at some confidence level) to be no greater than zero. This is the approach taken in, e.g., [24]. Under the actuarial principle, the fair price is calculated as the expectation under the risk-neutral process for financial asset and under the *real process* for mortality. As a result, there will be a residual risk that the issuer cannot control and the fair price can be adjusted by adding extra loading to the fair fee so that the risk of hedging error loss is no greater than zero, similarly to the estimation of the risk premium under the financial mathematics no-arbitrage approach. This procedure was considered in, e.g., [25].

Even if the force of mortality is deterministic, there will be a residual risk due to the finite size of the portfolio. Moreover, it is important to note that even if there is no mortality risk, in practice, there will be a residual risk due to discrete hedging and other incompleteness of financial markets. To handle this, again, one should quantify the fair price hedging error distribution and adjust the price extra loading under the actuarial principle or adjust risk premium under the no-arbitrage financial mathematics approach so that the risk of hedging error loss will not exceed the required level. These adjustments depend on the risk management strategy for the product.

Remark 1. For simplicity, we do not consider management fees α_m charged by a mutual fund for managing the investment portfolio. If management fees α_m is given exogenously, then it will have an

impact on the fair fee α that should by charged by the VA guarantee issuer. This can be accomplished as described in, e.g., [10] and can be easily incorporated in the framework outlined in our paper. Obviously, α will be larger for given $\alpha_m > 0$ comparing to the case $\alpha_m = 0$. The management fees reduce the performance of the investment account thus increasing the value of the guarantee as reported in, e.g., [39] for GMWB or [10] for GLWB. They commented that insurers wishing to provide the cheapest guarantee could provide the guarantee on the corresponding inexpensive exchange traded index fund rather than on a managed mutual fund account with extra fees.

4. VA Riders

There are many different specifications for the GMWB, GLWB, GMAB, GMIB and GMDB riders in the industry and academic literature. In this section we provide a mathematical formulation for some standard VA rider setups. We assume that the guarantee fee α is charged continuously. If the fee is charged discretely (and before withdrawal and other contract events), then one should make the following adjustment to the formulas in this section:

$$W(t_n^-) \to W(t_n^-)(1 - \widetilde{\alpha} dt_n)$$

if the fee is proportional to the wealth account, or

$$W(t_n^-) \to \max(W(t_n^-) - A(t_n^-)\widetilde{\alpha} dt_n, 0)$$

if the fee is proportional to the benefit base.

4.1. GMWB

A VA contract with GMWB promises to return at least the entire initial investment through cash withdrawals during the policy life plus the remaining account balance at maturity, regardless of the portfolio performance. Often in the academic literature, the studied GMWB type has a very simple structure, where the penalty is applied to the cashflow paid to the contract holder, while the benefit base is reduced by the full withdrawal amount. Specifically,

$$A(t_n^+) := h_n^A(W(t_n^-), A(t_n^-), \gamma_n) = A(t_n^-) - \gamma_n \tag{14}$$

with $\gamma_n \in \mathcal{A}_n, \mathcal{A}_n = [0, A(t_n^-)]$; and the cashflow paid to the contract holder is

$$\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) = \begin{cases} \gamma_n, & \text{if } 0 \le \gamma_n \le G_n \\ G_n + (1-\beta)(\gamma_n - G_n), & \text{if } \gamma_n > G_n \end{cases} \tag{15}$$

where $\beta \in [0,1]$ is the penalty parameter for the excess withdrawal. The contractual amount is defined as $G_n = W(0)(t_n - t_{n-1})/T$ and the maturity condition is

$$P_T(W(t_N^-), A(t_N^-)) = \max(W(t_N^-), \widetilde{f}_n(A(t_N^-)))$$

Note that the above specification does not allow any early surrender which can be included via extending the withdrawal space $\mathcal{A}_n$. Also, there is no death benefit; it is assumed that the beneficiary will maintain the contract in the case of a policyholder death. This contract has only basic features facilitating comparison of results from different academic studies, such as [4,5,7,14].

Specifications common in the industry include cases where the contractual amount G_n is specified to be different from $G_n = W(0)(t_n - t_{n-1})/T$ and a penalty is applied to both the withdrawn

amount and the benefit base. For example, specifications used in [23] to compare with the industry products include:

$$
\begin{aligned}
\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) &= \gamma_n - \delta_{excess} - \delta_{penalty} \\
\delta_{excess} &= \beta_n^{\mathrm{e}} \max\left(\gamma_n - \min(A(t_n^-), G_n), 0\right) \\
\delta_{penalty} &= \beta_n^{\mathrm{g}}(\gamma_n - \delta_{excess}) \times 1_{x+t_n < 59.5}
\end{aligned}
\tag{16}
$$

where x is the age of the policyholder in years at $t_0 = 0$, β_n^{e} and β_n^{g} are excess withdrawal and early withdrawal penalty parameters that can change with time, and

$$
\gamma_n \in \mathcal{A}_n, \ \mathcal{A}_n = \left[0, \max\left(W(t_n^-), \min(A(t_n^-), G_n)\right)\right]
$$

The authors of [23] also considered several specifications for the benefit base jump conditions.

- Specification 1:

$$
A(t_n^+) = \begin{cases}
\max(A(t_n^-) - \gamma_n, 0), & \text{if } \gamma_n \leq G_n \\
\max\left(\min\left(A(t_n^-) - \gamma_n, A(t_n^-)\frac{W(t_n^+)}{W(t_n^-)}\right), 0\right), & \text{if } \gamma_n > G_n
\end{cases}
\tag{17}
$$

- Specification 2:

$$
A(t_n^+) = \begin{cases}
\max(A(t_n^-) - \gamma_n, 0), & \text{if } \gamma_n \leq G_n \\
\max\left(\min\left(A(t_n^-) - \gamma_n, W(t_n^+)\right), 0\right), & \text{if } \gamma_n > G_n
\end{cases}
\tag{18}
$$

- Specification 3:

$$
A(t_n^+) = \begin{cases}
\max(A(t_n^-) - \gamma_n, 0), & \text{if } \gamma_n \leq G_n \\
\max\left(A(t_n^-) - G_n, 0\right)\frac{W(t_n^+)}{\max(W(t_n^-) - G_n, 0)}, & \text{if } \gamma_n > G_n
\end{cases}
\tag{19}
$$

In addition, a ratchet (reset of the benefit base to the wealth account if the latter is higher) can apply at the contract anniversary dates. If it occurs before the withdrawal, then in the above formulas one should make the following adjustment

$$
A(t_n^-) \to \max(A(t_n^-), W(t_n^-)), \text{ if } t_n \in \mathcal{T}_r
$$

If the reset is taking place after the withdrawal, then one should have

$$
A(t_n^+) \to \max(A(t_n^+), W(t_n^+)), \text{ if } t_n \in \mathcal{T}_r
$$

4.2. GLWB

A GLWB rider is similar to GMWB but provides a guaranteed withdrawal for life; upon the policyholder death the remaining wealth account value is paid to the beneficiary. The contractual withdrawal amount G_n is typically based on a fixed proportion g of the benefit base $A(t)$, i.e., $G_n = g \times A(t_n^-)(t_n - t_{n-1})$. The benefit base can increase via ratchet (*step-up*) or bonus (*roll-up*) features. Bonus feature provides an increase of the benefit base if no withdrawal is made on a withdrawal date. Complete surrender refers to the withdrawal of the whole policy account. The withdrawal can exceed the contractual amount and in this case the net amount received by the policyholder is subject to a penalty. Under the typical specification considered, e.g., in [11], the cashflow received by the policyholder is

$$
\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) = \begin{cases}
\gamma_n, & \text{if } 0 \leq \gamma_n \leq G_n \\
G_n + (1 - \beta)(\gamma_n - G_n), & \text{if } \gamma_n > G_n
\end{cases}
\tag{20}
$$

$$
\gamma_n \in \mathcal{A}_n, \quad \mathcal{A}_n = [0, \max(W(t_n^-), G_n)]
$$

where β is the penalty parameter for the excess withdrawal. The benefit base jump condition, including ratchets and bonus features, is given by

$$
\begin{aligned}
A(t_n^+) \quad = \quad & \max\left(A(t_n^-)(1+b_n), W(t_n^-)1_{t_n\in\mathcal{T}_r}\right) \times 1_{\gamma_n=0} \\
& + \max\left(A(t_n^-), \max(W(t_n^-)-\gamma_n,0) \times 1_{t_n\in\mathcal{T}_r}\right) \times 1_{0<\gamma_n\leq G_n} \\
& + \max\left(A(t_n^-)\frac{W(t_n^-)-\gamma_n}{W(t_n^-)-G_n}, (W(t_n^-)-\gamma_n)\times 1_{t_n\in\mathcal{T}_r}\right) \times 1_{G_n<\gamma_n\leq W(t_n^-)} \quad\quad (21)
\end{aligned}
$$

where b_n is the bonus rate parameter that may change in time. Finally, if the policyholder dies during $(t_{n-1}, t_n]$, the beneficiary receives a death benefit payment $D_n(W(t_n^-), A(t_n^-)) = W(t_n^-)$ and t_N corresponds to the maximum age beyond which survival is deemed impossible.

4.3. GMAB

A GMAB rider provides a certainty of capital till some maturity (e.g., 10 or 20 years) and the potential for a capital growth. Typical GMAB products sold on the market do not impose penalty on the policyholder withdrawal amount but can penalise the benefit base (protected capital balance) under some conditions. It is also common to have a ratchet feature, where the protected capital balance increases to the wealth account if the latter is higher on an anniversary date. The withdrawals from the account are allowed subject to a penalty. For example, specifications of the product marketed by [27,28] in Australia are very close to the following formulation:

$$
\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) = \gamma_n \quad\quad (22)
$$

$$
A(t_n^+) := h_n^A(W(t_n^-), A(t_n^-), \gamma_n) = \begin{cases} \max\left(A(t_n^-), W(t_n^-)\right) - C_n(\gamma_n), & \text{if } t_n\in\mathcal{T}_r \\ \max\left(A(t_n^-) - C_n(\gamma_n), 0\right), & \text{otherwise} \end{cases} \quad\quad (23)
$$

where $C_n(\gamma_n)$ is a penalty function that can be larger than γ_n as defined below, and $\gamma_n \in \mathcal{A}_n = [0, W(t_n^-)]$.

The product is offered for the *super* and *pension* account types. The *super* account is designed for an investor being in an accumulation phase, while the *pension* account is for a retired investor in an annuitization phase. The difference between the accounts in terms of technical details is only in the penalty applied to the protected capital after withdrawals; the super account discourages withdrawals more than the pension account. In both cases the penalty is in the form of a reduction of the protected capital larger than the withdrawn amount. The penalty only applies if the wealth account balance is below the protected capital amount. A super account penalizes any amount of withdrawals, while the pension account only penalizes excessive withdrawals.

Specifically, for a *super account*, the function $C_n(\gamma_n)$ is given by

$$
C_n(\gamma_n) = \begin{cases} \gamma_n, & \text{if } W(t_n^-) \geq A(t_n^-) \\ A(t_n^-)\gamma_n/W(t_n^-), & \text{if } W(t_n^-) < A(t_n^-) \end{cases} \quad\quad (24)
$$

and for a *pension account*, the penalty is

$$
C_n(\gamma_n) = \begin{cases} \gamma_n, & \text{if } W(t_n^-) \geq A(t_n^-) \ \text{ or } \ \gamma_n \leq G_n \\ A(t_n^-)\gamma_n/W(t_n^-), & \text{if } W(t_n^-) < A(t_n^-) \ \text{ and } \ \gamma_n > G_n \end{cases} \quad\quad (25)
$$

That is, the penalty for the pension account applies only if the wealth account balance is below the protected capital amount *and* the withdrawal is above a pre-determined amount G_n.

Finally, the terminal condition is given by

$$
P_T(W(t_N^-), A(t_N^-)) = \max(W(t_N^-), A(t_N^-))
$$

A total withdrawal of the wealth account balance effectively terminates the contract, as the penalty mechanism ensures the protected capital is always exhausted to zero by a complete withdrawal.

4.4. GMIB

At maturity, the holder of a GMIB rider can select to take a lump sum of the wealth account $W(T)$ or annuitise this amount at annuitization rate $\ddot{a}_T$ current at the contract maturity or annuitize the benefit base $A(T^-)$ at the pre-specified annuitization rate $\ddot{a}_g$. Annuitization rate is defined as the price of an annuity paying one dollar each year. If the account value is below the benefit base, then the customer cannot take $A(T^-)$ as a lump sum but only as an annuity at the pre-specified rate. Thus, the payoff of VA with GMIB at time T is

$$P_T(W(t_N^-), A(t_N^-)) = \max\left(W(t_N^-), A(t_N^-)\frac{\ddot{a}_T}{\ddot{a}_g} \right)$$

The benefit base may include roll-ups and ratchets similar to as described for the GLWB case. Again, this rider can be offered jointly with other riders. For example, it can be part of GMWB or GMAB contract maturity conditions. For discussion and pricing of GMIB in academic literature, see [1,40].

4.5. GMDB

A GMDB rider provides a death benefit if the policy holder death occurs before or at the contract maturity. Assuming that if the policyholder dies during $(t_{n-1}, t_n]$, then the beneficiary will be paid an amount $D_n(\cdot)$ at t_n, some of the common death benefit types are:

$$D_n(W(t_n^-), A(t_n^-)) = \begin{cases} \max(A(t_n^-), W(t_n^-)), & \text{death benefit type 0} \\ W(0), & \text{death benefit type 1} \\ \max(W(0), W(t_n^-)), & \text{death benefit type 2} \\ W(t_n^-), & \text{death benefit type 3} \end{cases} \quad (26)$$

Some providers adjust the initial premium $W(0)$ for inflation in the death benefit. For some policies, the death benefit type may change at some age, e.g., death benefit type 1 or type 2 may change to type 0, effectively making the death benefit expiring at some age (e.g., at the age of 75 years). The death benefit can be provided on top of some other guarantees and the contract may provide a *spousal continuation option* that allows a surviving spouse to continue the contract. The contract may have accumulation phase where the death benefit may increase, and continuation phase where the death benefit remains constant. Pricing GMDB rider has been considered in, e.g., [8,41,42].

5. Fair Pricing

Given the withdrawal strategy $\gamma = (\gamma_1, \ldots, \gamma_{N-1})$, the present value of the contract cahsflows is given by (3) which is a function of the state vector $X_n = (W(t_n^-), A(t_n^-), I_n)$, $n = 0, \ldots, N$. Let $Q_t(W, A)$ be the price of the VA contract with a guarantee at time t, when $W(t) = W$, $A(t) = A$ and the policyholder is alive. For simplicity of notation, if the policyholder is alive, we drop mortality state variable $I_n = 1$ in the function arguments. Then the contract fair price under the given withdrawal strategy γ can be calculated as

$$Q_0(W(0), A(0)) = \mathrm{E}_{t_0}^{\mathbb{Q},\mathbb{I}}[H_0(\boldsymbol{X}, \boldsymbol{\gamma})] \quad (27)$$

Here, $\mathrm{E}_t^{\mathbb{Q},\mathbb{I}}[\cdot]$ denotes an expectation with respect to the state vector $\boldsymbol{X}$, conditional on information available at time t, i.e., with respect to the financial risky asset process under the risk-neutral probability measure $\mathbb{Q}$ and with respect to the mortality process under the probability measure $\mathbb{I}$. The latter can be

either real or risk-neutral probability measure depending on the pricing approach in the incomplete market as discussed in Section 3.

The fair fee value of α to be charged for the VA guarantee corresponds to $Q_0(W(0), A(0)) = W(0)$. That is, once a pricing of $Q_0(W(0), A(0))$ for a given α is developed, then a numerical root search algorithm is required to find the fair fee.

Remark 2. If volatility, interest rate or force of mortality are stochastic then these variables should be added to the state vector X for valuation of the fair price expectation (27).

5.1. Pricing as Stochastic Control Problem

The withdrawal strategy γ can depend on the time and state variables and is assumed to be given when the price of the contract is calculated in (27). The withdrawal strategies are classified as *static*, *optimal*, and *sub-optimal*.

- **Static strategy.** Under this strategy, the policyholder decisions are deterministically determined at the beginning of the contract and do not depend on the evolution of the wealth and benefit base accounts. For example, policyholder withdraws at the contractual rate only.
- **Optimal strategy.** Under the optimal withdrawal strategy, the decision on the withdrawal amount γ_n depends on the information available at time t_n, i.e., depends on the state variable X_n. The optimal strategy is calculated as

$$\gamma^*(X) = \underset{\gamma \in \mathcal{A}}{\operatorname{argsup}} E_{t_0}^{\mathbb{Q}, \mathbb{I}} [H_0(X, \gamma)] \tag{28}$$

where the supremum is taken over all admissible strategies γ. Any other strategy $\gamma(X)$ different from $\gamma^*(X)$ is called *sub-optimal* and leads to a smaller price.

Given that the state variable $X = (X_1, \ldots, X_N)$ is a Markov process and the contract payoff is represented by the general formula (3), calculation of the contract value (27) under the optimal withdrawal strategy (28) is a standard optimal stochastic control problem for a *controlled Markov process*. Note that, the control variable γ_n affects the transition law of the underlying wealth $W(t)$ process from t_n^- to t_{n+1}^- and thus the process is controlled. For a good textbook treatment of stochastic control problems in finance, see [43]. This type of problems can be solved recursively to find the contract value $Q_{t_n}(x)$ at t_n when $X_n = x$ for $n = N-1, \ldots, 0$ via the backward induction *Bellman equation*

$$Q_{t_n}(X_n) = \sup_{\gamma_n \in \mathcal{A}_n} \left(f_n(X_n, \gamma_n) + E_{t_n}^{\mathbb{Q}, \mathbb{I}, \gamma_n} \left[B_{n,n+1} Q_{t_{n+1}}(X_{n+1}) \Big| X_n \right] \right) \tag{29}$$

starting from the final condition $Q_T(x) = H_N(x)$. The upperscript γ_n in the expectation operator is used to indicate that the transition probability to reach state $X_{n+1} = x'$ at time t_{n+1} if the withdrawal (*action*) γ_n is applied in the state x at time t_n depends on γ_n. Obviously, the above backward induction can also be used to calculate the fair contract price in the case of a static strategy γ; in this case the space of admissible strategies $\mathcal{A}_n$ contains only one pre-defined value and $\sup(\cdot)$ becomes redundant.

For clarity, denote $Q_{t_n^-}(\cdot)$ and $Q_{t_n^+}(\cdot)$ the contract values just before and just after the event time t_n respectively. Then, in the case of deterministic force of mortality, after calculating the expectation with respect to the mortality state variable I_{n+1} in (29), the required backward recursion can be rewritten explicitly as

$$\begin{aligned}
Q_{t_n^+}(W, A) &= (1 - q_{n+1}) E_{t_n^+}^{\mathbb{Q}} \left[B_{n,n+1} Q_{t_{n+1}^-} \left(W(t_{n+1}^-), A \right) \Big| W, A \right] \\
&\quad + q_{n+1} E_{t_n^+}^{\mathbb{Q}} \left[B_{n,n+1} D_{n+1} \left(W(t_{n+1}^-), A \right) \Big| W, A \right]
\end{aligned} \tag{30}$$

with the jump condition

$$Q_{t_n^-}(W, A) = \max_{\gamma_n \in \mathcal{A}_n} \left(\widetilde{f}_n(W, A, \gamma_n) + Q_{t_n^+} \left(h_n^W(W, A, \gamma_n), h_n^A(W, A, \gamma_n) \right) \right) \tag{31}$$

This recursion is solved for $n = N - 1, N - 2, \ldots, 0$, starting from the maturity condition $Q_{t_N^-}(W, A) = P_T(W, A)$.

In the case of a stochastic force of mortality $\lambda(t)$, when $q_n = 1 - \exp(-\int_{t_{n-1}}^{t_n} \lambda(y) dy)$, see (11), the contract value $Q_{t_n}(\cdot)$ and the optimal value of γ_n are not only functions of $W(t_n^-)$ and $A(t_n^-)$ but also functions of a stochastic force of mortality $\lambda(t_n)$. In this case, one has to add expectation with respect to $\lambda(t)$ process conditional on $\lambda(t_n)$ to the right side of Equation (30). Similarly, if the interest rate and/or volatility are modelled as stochastic.

Also, we would like to note that if the risk-free interest rate $r(t)$ is stochastic, then it can be convenient to use change of numéraire technique. In particular, changing the numéraire from the money market account $M(t) = \exp(\int_0^t r(q) dq)$ to the bond price $\mathcal{P}(t_n, t_{n+1})$ at time t_n with maturity t_{n+1} will simplify calculations in (30) using

$$\mathrm{E}_{t_n^+}^{\mathbb{Q}} \left[e^{-\int_{t_n}^{t_{n+1}} r(u) du} Q_{t_{n+1}^-}(W(t_{n+1}^-), A) \,\middle|\, \cdot \right] = \mathcal{P}(t_n, t_{n+1}) \mathrm{E}_{t_n^+}^{\widetilde{\mathbb{Q}}} \left[Q_{t_{n+1}^-}(W(t_{n+1}^-), A) \,\middle|\, \cdot \right]$$

Here, $\widetilde{\mathbb{Q}}$ is a new probability measure obtained from $\mathbb{Q}$ using the Radon-Nikodym derivative $d\widetilde{\mathbb{Q}}/d\mathbb{Q} = (M(t_n)/M(t)) \times (\mathcal{P}(t, t_{n+1})/\mathcal{P}(t_n, t_{n+1}))$, $t \in [t_n, t_{n+1}]$, see, e.g., [12].

5.2. Alternative Solution

Consider the case of deterministic force of mortality $\lambda(t)$ used to calculate death probabilities in (11). Given that the mortality and financial asset processes are independent, and the withdrawal decision does not affect the mortality process, one can calculate the expected value of the payoff (3) with respect to the mortality process, $\widetilde{H}_0(W, A) = \mathrm{E}_{t_0}^{\mathbb{I}}[H_0(X, \gamma)]$, and then calculate the price under the optimal strategy as $\sup_\gamma \mathrm{E}_{t_0}^{\mathbb{Q}}[\widetilde{H}_0(W, A)]$ or under the given strategy as $\mathrm{E}_{t_0}^{\mathbb{Q}}[\widetilde{H}_0(W, A)]$. It is easy to find that

$$\begin{aligned}
\widetilde{H}_0(W, A) &= B_{0,N} \left(p_N P_T \left(W(T^-), A(T^-) \right) + q_N p_{N-1} D_N \left(W(T^-), A(T^-) \right) \right) \\
&\quad + \sum_{n=1}^{N-1} B_{0,n} \left(p_n \widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) + p_{n-1} q_n D_n \left(W(t_n^-), A(t_n^-) \right) \right)
\end{aligned} \tag{32}$$

where $p_n = \Pr[\tau > t_n | \tau > t_0]$ and $q_n p_{n-1} = \Pr[t_{n-1} < \tau \leq t_n | \tau > t_0]$ for a random death time τ, i.e., $p_n = p_{n-1} \times (1 - q_n)$. Note that, previously we defined $q_n = \Pr[t_{n-1} < \tau \leq t_n | \tau > t_{n-1}]$.

The payoff (32) has the same general form as the payoff (3). Thus, the optimal stochastic control problem $\Psi_{t_0}(W(0), A(0)) = \sup_\gamma \mathrm{E}_{t_0}^{\mathbb{Q}}[\widetilde{H}_0(W, A)]$ can be solved using the Bellman Equation (29) leading to the following explicit recursion

$$\begin{aligned}
\Psi_{t_n^+}(W, A) &= \mathrm{E}_{t_n^+}^{\mathbb{Q}} \left[B_{n,n+1} \Psi_{t_{n+1}^-}(W(t_{n+1}^-), A(t_{n+1}^-)) \,|\, W, A \right] \tag{33} \\
\Psi_{t_n^-}(W, A) &= \max_{\gamma_n \in \mathcal{A}_n} \left(p_n \widetilde{f}_n(W, A, \gamma_n) + p_{n-1} q_n D_n(W, A) \right. \\
&\qquad\qquad \left. + \Psi_{t_n^+} \left(h_n^W(W, A, \gamma_n), h_n^A(W, A, \gamma_n) \right) \right) \tag{34}
\end{aligned}$$

for $n = N - 1, N - 2, \ldots, 0$, starting from $\Psi_{t_N^-}(W, A) = p_N P_T(W, A) + p_{N-1} q_N D_N(W, A)$.

It is easy to verify that this recursion leads to the same solution $\Psi_{t_0}(W, A) = Q_{t_0}(W, A)$ and the same optimal strategy for γ as obtained from the recursion (30) and (31), noting that $\Psi_{t_n^-}(W, A) = p_n Q_{t_n^-}(W, A) + p_{n-1} q_n D_n(W, A)$. The result is somewhat obvious because

$$\sup_\gamma \mathrm{E}_{t_0}^{\mathbb{Q}, \mathbb{I}} [H_0(\boldsymbol{X}, \boldsymbol{\gamma})] = \sup_\gamma \mathrm{E}_{t_0}^{\mathbb{Q}} \left[\mathrm{E}_{t_0}^{\mathbb{I}} [H_0(\boldsymbol{X}, \boldsymbol{\gamma})] \right] \tag{35}$$

Note that, $\sup_\gamma \mathrm{E}_{t_0}^{\mathbb{Q}, \mathbb{I}} [H_0(\boldsymbol{X}, \boldsymbol{\gamma})] \neq \mathrm{E}_{t_0}^{\mathbb{I}} \left[\sup_\gamma \mathrm{E}_{t_0}^{\mathbb{Q}} [H_0(\boldsymbol{X}, \boldsymbol{\gamma})] \right]$. That is, one cannot find the price under the optimal strategy conditional on the death time and then average over random death times, that would lead to the result larger than $Q_{t_0}(W, A)$, see [8].

Note that if the force of mortality $\lambda(t)$ is stochastic then equality (35) is not valid, the optimal withdrawal value of γ_n becomes a function of $\lambda(t_n)$ and the formulas in this section cannot be used to calculate the fair price.

5.3. Remarks on Withdrawal Strategy

The guarantee fare fee based on the optimal policyholder withdrawal is the worst case scenario for the issuer, i.e., if the guarantee is hedged then this fee will *ensure* no losses for the issuer (in other words full protection against the policyholder strategy and market uncertainty). Of course this is under the given assumptions about stochastic model for the underlying risky asset, and in the case of deterministic force of mortality assuming that the mortality risk is fully diversified. If the issuer hedges continuously but investors deviate from the optimal strategy, then the issuer will receive a guaranteed profit.

Any strategy different from the optimal is sup-optimal and will lead to smaller fair fees. Of course the strategy optimal in this sense is not related to the policyholder circumstances. The policyholder may act optimally with respect to his preferences and circumstances but it may be different from the optimal strategy calculated in (31). On the other hand, as noted in [17], secondary markets for equity linked insurance products (where the policyholder can sell their contracts) are growing. Thus, financial third parties can potentially generate guaranteed profit through hedging strategies from financial products such as VA riders which are not priced according to the worst case assumption of the optimal withdrawal strategy. Thus the development of secondary markets for VA riders would lead to an increase in the fees charged by the issuing companies. [18] undertakes an empirical study of policyholders behavior in Japanese VA market and they show that the moneyness of the guarantee has the largest explanatory power for the surrender rates.

One way to introduce a reasonable sub-optimal withdrawal model is to assume that the policyholder follows a default strategy withdrawing a contractual amount G_n at each event time t_n unless the extra value from undertaking an optimal withdrawal is greater than $\theta \times G_n$, $\theta \geq 0$. Setting $\theta = 0$ corresponds to the optimal strategy, while $\theta \gg 1$ leads to the strategy of withdrawals at the contract rate. This is the approach considered, e.g., in [10,39]. More complicated approach would specify a life-cycle utility model to determine the strategy optimal for the policyholder with respect to his circumstances and preferences, this is the approach studied in [19–22]. In any case, once the strategy is specified (estimated empirically or by another model), one can use Equation (31) to calculate the fair price and fair fee with the admissible strategy space $\mathcal{A}_n$ restricted to the specified strategy.

5.4. Tax Consideration

Withdrawals from the VA type contracts may attract country and individual specific government taxes. In [23], the authors demonstrated that including taxes significantly affects the value of VA withdrawal guarantees. They developed a *subjective* risk-neutral valuation methodology and produced results in line with empirical market prices. Following closely to [23], we introduce an extra state variable $R(t)$ to present the *tax base* which is the amount that that may still be drawn tax-free, and assume

that all event times $t_n \in \mathcal{T}$ are the policy anniversary dates. The initial premium is assumed to be post-tax and taxes are applied to future investment gains (not the initial investment).

Denote a marginal income tax rate as $\widetilde{\kappa}$ and a marginal capital gain tax from investment outside of VA contract as κ. It is assumed that earnings from VA are treated as ordinary income and withdrawals are taxed on a last-in first-out basis. Thus if the wealth account $W(t_n^-)$ exceeds the tax base $R(t_n^-)$, any withdrawal up to $W(t_n^-) - R(t_n^-)$ will be taxed at the rate $\widetilde{\kappa}$ and will not affect the tax base; larger withdraws will not be subject to tax but will reduce the tax base. Specifically, the tax base will change at the withdrawal time t_n as

$$R(t_n^+) = R(t_n^-) - \max\left(\gamma_t - \max(W(t_n^-) - R(t_n^-), 0), 0\right)$$

The cashflow received by the policyholder will be reduced by taxes,

$$tax = \widetilde{\kappa}\min\left(\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n), \max(W(t_n^-) - R(t_n^-), 0)\right)$$

i.e., one has to make the following change in the contract specifications listed in Section 4,

$$\widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) \to \widetilde{f}_n(W(t_n^-), A(t_n^-), \gamma_n) - tax$$

Using arguments for replicating pre-tax cashlows at t_n with post-tax cashflows at t_{n+1}, it was shown in [23] that $Q_{t_n^+}(W, A, R)$ should be found not as the direct expectation (30) but should be found as the solution of the following nonlinear equation

$$\begin{aligned}
Q_{t_n^+}(W, A, R) &= \mathrm{E}_{t_n^+}^{\mathbb{Q}}\left[V(t_{n+1}^-)|W, A, R\right] \\
&+ \frac{\kappa}{1-\kappa}\mathrm{E}_{t_n^+}^{\mathbb{Q}}\left[\max\left[V(t_{n+1}^-) - B_{n,n+1}Q_{t_n^+}(W, A, R), 0\right]\,\Big|\,W, A, R\right]
\end{aligned} \tag{36}$$

where

$$\begin{aligned}
V(t_{n+1}^-) &= (1 - q_{n+1})B_{n,n+1}Q_{t_{n+1}^-}\left(W(t_{n+1}^-), A(t_{n+1}^-), R(t_{n+1}^-)\right) \\
&+ q_{n+1}B_{n,n+1}D_{n+1}\left(W(t_{n+1}^-), A(t_{n+1}^-), R(t_{n+1}^-)\right)
\end{aligned} \tag{37}$$

This is referred to as *subjective* valuation from the policyholder perspective and depends on the investor current position (including possible offset tax responsibilities) and tax rates. Numerical examples in [23] show that the VA guarantee prices accounting for taxes in the above way are lower than ignoring the taxes (not surprisingly, because it is just a sub-optimal strategy), making the prices overall more aligned with those observed in the market.

6. Numerical Valuation of VA Riders

In the case of realistic VA riders with discrete events such as ratchets and optimal withdrawals, there are no closed form solutions and the fair price has to be calculated numerically, even in the case of simple geometric Brownian motion process for the risky asset. In general, one can use the PDE, direct integration or regression type MC methods, where the backward recursion (30) and (31) is solved numerically. Of course, if the withdrawal strategy is known, then one can always use a standard MC to simulate all state variables forward in time till the contract maturity or the policyholder death and average the payoff discounted cashflows over many independent realizations. This standard procedure is well known and no further discussion is needed.

In this section, we give a brief review of different numerical methods that can be used for valuation of VA riders. Then, we provide detailed description of the direct integration method that can be very efficient and simple to implement, when the transition density of the underlying asset or it's moments

between the event times are known in closed form. Finally, in Section 6.5 we present calculation of the hedging parameters (the so-called *Greeks*).

6.1. Numerical Algorithms

Simulation based Least-Squares MC method introduced in [15] is designed for uncontrolled Markov process problems and can be used to account for the contract early surrender, as, e.g., in [9]. However, it cannot be used to estimate the full optimal withdrawal strategy. This is because dynamic withdrawals affect the paths of the underlying wealth account and one cannot carry out a forward simulation step required for the subsequent regression in the backward induction. However, it should be possible to apply the control randomization methods extending Least-Squares MC to handle the optimal stochastic control problems with controlled Markov processes, as was recently developed in Kharroubi et al. [16]. The idea is to first simulate the control (withdrawals) and the state variables forward in time, where the control is simulated independently from other variables. Then, use a regression on the simulated state variables and control to estimate the expectation in (30) and find the optimal withdrawal using (31). However, the accuracy and robustness of this method for pricing withdrawal benefit type products have not been studied yet. As usual, it is expected that the choice of the basis functions for the required regression step will have significant impact on the performance. We also note that in some simple cases of the withdrawal strategy admissible space such as bang-bang (no withdrawal, withdrawal at the contractual rate, or full surrender), it is possible to develop other modifications of Least-Squares MC such as in [11] for pricing of the GLWB rider.

The expectation in (30) can also be calculated using PDE or direct integration methods. In both cases, the modeller discretizes the space of the state variables and then calculates the contract value for each grid point. The PDE for calculation of expected value (30) under the assumed risk-neutral process for the risky asset $S(t)$ is easily derived using Feynman-Kac theorem; for a good textbook treatment of this topic, see, e.g., [29]. However, the obtained PDE can be difficult or even not practical to solve in the high-dimensional case. In the case of the geometric Brownian motion process for the risky asset (9), the governing PDE in the period between the event times is the one-dimensional Black-Scholes PDE, with the jump conditions (31) at each event time to link the prices at the adjacent periods. Since the benefit base state variable $A(t)$ remains unchanged within the interval (t_{i-1}, t_i), $i = 1, 2, \ldots, N$, the contract value $Q_t(W, A)$ satisfies the following PDE with no explicit dependence on A,

$$\frac{\partial Q}{\partial t} + \frac{\sigma^2}{2} W^2 \frac{\partial^2 Q}{\partial W^2} + (r - \alpha) W \frac{\partial Q}{\partial W} - rQ = 0 \tag{38}$$

This PDE can be solved numerically using, e.g., Crank-Nicholson finite difference scheme for each A backward in time with the jump condition (31) applied at the contract event times. This has been done, e.g., in [4,5] for pricing GMWB with discrete optimal withdrawals. Of course, if the volatility or/and interest rate are modelled as diffusion processes, then extra dimensions will add to the PDE making it more difficult to solve. [10] used the PDE approach to calculate VA rider prices in the case of stochastic regime-switching volatility and interest rate. In this case the required PDE to be solved is still one-dimensional that allows to develop a very efficient numerical valuation.

Under the direct integration approach, the expected value (30) is calculated as an integral approximated by summation over the space grid points, see, e.g., [1]. More efficient quadrature methods (requiring less points to approximate the integral) exist. In particular, in the case of a geometric Brownian motion process for the risky asset, it is very efficient to use the Gauss-Hermite quadrature as developed in [26] and applied for GMWB pricing in [7]. Section 6.3 provides detailed description of the method for pricing VA riders in general. This method can be applied when the transition density of the underlying asset between the event times *or it's moments* are known in closed form. It is relatively easy to implement and computationally faster than PDE method because the latter requires many time steps between the event times. In [12], this method was also used to calculate GMWB in the case of stochastic interest rate under the Vasicek model.

In both PDE and direct integration approaches, one needs some interpolation scheme to implement the jump condition (31), because state variables located at the grid points of the discretized space do not appear on the grid points after the jump event. This will be discussed in detail in Section 6.4. Of course, if the underlying stochastic process is more complicated than geometric Brownian motion (9) and does not allow efficient calculation of the transition density or its moments, one can always resort to PDE method.

In our numerical examples of GMAB pricing in Section 7, we adapt a direct intergation method based on the Gauss-Hermite integration quadrature applied on a cubic spline interpolation, *hereafter referred to as GHQC*. For testing purposes, we also implemented Crank-Nicholson finite difference (FD) scheme solving corresponding PDE (38) with the jump condition (31).

6.2. Overall Algorithm Description

Both PDE and direct integration numerical schemes start from a final condition for the contract value at $t = T^-$. Then, a backward time stepping using (30) or solving corresponding PDE gives solution for the contract value at $t = t_{N-1}^+$. Application of the jump condition (31) to the solution at $t = t_{N-1}^+$ gives the solution at $t = t_{N-1}^-$ from which further backward in time recursion gives solution at t_0. For simplicity assume that there are only $W(t)$ and $A(t)$ state variables. The numerical algorithm then takes the following key steps.

Algorithm 1 Direct Integration or partial differential equation (PDE) method

- Step 1. Generate an auxiliary finite grid $0 = A_1 < A_2 < \cdots < A_J$ to track the benefit base balance A.
- Step 2. Discretize wealth account balance W space as $W_0 < W_1 < \cdots < W_M$ to generate the grid for computing the expectation (30).
- Step 3. At $t = t_N$, apply the final condition at each node point (W_m, A_j), $j = 1, 2, \ldots, J$, $m = 1, 2, \ldots, M$ to get $Q_{t_N^-}(W, A)$.
- Step 4. Evaluate expectation (30) for each A_j, $j = 0, \ldots, J$, to obtain $Q_{t_{N-1}^+}(W, A)$ either using direct integration or solving PDE. In the case of direct integration method, this involves one-dimensional interpolation in W space to find values of $Q_{t_N^-}(W, A)$ at the guadrature points different from the grid points.
- Step 5. Apply the jump condition (31) to obtain $Q_{t_{N-1}^-}(W, A)$ for all possible values of γ_{N-1} and find γ_{N-1} that maximizes $Q_{t_{N-1}^-}(W, A)$. In general, this involves a two-dimensional interpolation in (W, A) space.
- Step 6. Repeat Steps 4 and 5 for $t = t_{N-2}, t_{N-3}, \ldots, t_1$.
- Step 7. Evaluate integration (30) for the backward time step from t_1 to t_0 to obtain solution $Q_0(W, A)$ at $W = W(0)$ and $A = A(0)$, or may be at several points if these are needed for calculation of some hedging sensitivities such as Delta and Gamma discussed in Section 6.5.

In our implementation of the direct integration method based on the Gauss-Hermite quadrature for numerical examples in Section 7, we use a one-dimensional cubic spline interpolation required to handle integration in Step 4 and a bi-cubic spline interpolation to handle jump condition in Step 5.

If the model has other stochastic state variables (similar to W) changing stochastically between the contract event times, such as stochastic volatility and/or stochastic interest rate, then grids for these extra dimensions should be generated and the required integration or PDE to evaluate (30) will have extra dimensions. Also, extra auxiliary state variables (similar to A) unchanged between the contract event times, such as tax base and/or extra benefit base, will require extra dimensions in the grid and interpolation for the jump condition at the event times.

We have to consider the possibility of $W(t)$ goes to zero due to withdrawal and market movement, thus one has to use the lower bound $W_0 = 0$. The upper bound W_M should be set sufficiently far from the initial wealth at time zero $W(0)$. A good choice of such a boundary could be based on the

high quantiles of distribution of $S(T)$. For example, in the case of the geometric Brownian motion process (1), one can set conservatively

$$W_M = W(0)e^{|\text{mean}(\ln(S(T)/S(0)))|+5\times\text{stdev}(\ln(S(T)/S(0)))}$$

Often, it is more efficient to use an equally spaced grid in $\ln W$ space. In this case, W_0 cannot be set to zero and instead should be set to a very small value (e.g., $W_0 = 10^{-10}$). Also, for some VA riders, using equally spaced grid in $\ln A$ space is also more efficient.

6.3. Direct Integration Method

To compute $Q_0(W(0), A(0))$, we have to evaluate the expectations in the recursion (30). Assuming the conditional probability density of $W(t_n^-)$ given $W(t_{n-1}^+)$ is known in closed form $\widetilde{p}_n(w|W(t_{n-1}^+))$, the required expectation (30) can be calculated as

$$Q_{t_{n-1}^+}\left(W(t_{n-1}^+), A\right) = \int_0^{+\infty} \widetilde{p}_n(w|W(t_{n-1}^+))\widetilde{Q}_{t_n^-}(w, A)dw \tag{39}$$

where

$$\widetilde{Q}_{t_n^-}(w, A) = B_{n-1,n} \times \left((1 - q_n)Q_{t_n^-}(w, A) + q_n D_n(w, A)\right)$$

The above integral can be estimated using various numerical integration (*quadrature*) methods. Note that, one can always find $W(t_n^-)$ as a transformation of the standard normal random variable Z as

$$W(t_n^-) = \psi(Z) := F_n^{-1}(\Phi(Z))$$

where $\Phi(\cdot)$ is the standard normal distribution, and $F_n(\cdot)$ and $F_n^{-1}(\cdot)$ are the distribution and its inverse of $W(t_n^-)$. Then, the integral (39) can be rewritten as

$$Q_{t_{n-1}^+}\left(W(t_{n-1}^+), A\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} \widetilde{Q}_{t_n^-}(\psi_n(z), A)dz \tag{40}$$

This type of integrand is well suited for the Gauss-Hermite quadrature that for an arbitrary function $f(x)$ gives the following approximation

$$\int_{-\infty}^{+\infty} e^{-x^2}f(x)dx \approx \sum_{i=1}^{q} \lambda_i^{(q)} f(\xi_i^{(q)}) \tag{41}$$

Here, q is the order of the Hermite polynomial, $\xi_i^{(q)}, i = 1, 2, \ldots, q$ are the roots of the Hermite polynomial $H_q(x)$, and the associated weights $\lambda_i^{(q)}$ are given by

$$\lambda_i^{(q)} = \frac{2^{q-1}q!\sqrt{\pi}}{q^2[H_{q-1}(\xi_i^{(q)})]^2}$$

This approximate integration works very well if function $f(x)$ is without singularities and it calculates the integral exactly if $f(x)$ is represented by a polynomial of degree $2q - 1$ or less.

Note that $\widetilde{Q}_t(w, \cdot)$ is known only at the grid points $W_m, m = 0, 1, \ldots, M$ and interpolation is required to estimate $\widetilde{Q}_t(w, \cdot)$ at the quadrature points. From our experience with pricing different VA guarantees, we recommend the use of the natural cubic spline interpolation which is smooth in the first derivative and continuous in the second derivative; and the second derivative is assumed zero for the extrapolation region above the upper bound.

Of course it can be difficult to find the distribution $F_n(\cdot)$ and its inverse $F_n^{-1}(\cdot)$ in general. In the case of geometric Brownian motion process (1), the transition density $\widetilde{p}_n(\cdot|\cdot)$ is just a lognormal density and

$$W(t_n^-) = \psi_n(Z) := W(t_{n-1}^+) \exp\left((r_n - \alpha - \frac{1}{2}\sigma_n^2)dt_n + \sigma_n\sqrt{dt_n}Z \right)$$

Then, a straightforward application of the Gauss-Hermite quadrature for the evaluation of integral (39) gives

$$Q_{t_{n-1}^+}\left(W(t_{n-1}^+), A\right) \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{q} \lambda_i^{(q)} \widetilde{Q}_{t_n^-}\left(\psi_n(\sqrt{2}\xi_i^{(q)}), A\right) \tag{42}$$

that should be calculated for each grid point $W(t_{n-1}^+) = W_m, m = 0, 1, \ldots, M$. Often, a small number of quadrature points is required to achieve a very good accuracy; in our numerical examples in the next section we use $q = 9$ but very good results are also obtained with $q = 5$.

If the transition density function from $W(t_{n-1}^+)$ to $W(t_n^-)$ is not known in closed form but one can find its moments, then the integration can also be done with similar efficiency and accuracy by the method of matching moments as described in [7,26]. The method also works very well in the two-dimensional case, see, e.g., [12] where it was applied for GMWB pricing in the case of stochastic interest rate.

6.4. Jump Condition Application

Either in PDE or direct integration method, one has to apply the jump condition (31) at the event times to obtain $Q_{t_n^-}(W, A)$. For the optimal strategy, we chose a value of withdrawal $\gamma_n \in \mathcal{A}_n$ maximizing the value $Q_{t_n^-}(W, A)$.

To apply the jump conditions, for each A_j, we associate a continuous solution using (42) and interpolation. In general, as can be seen from (31), the jump condition makes it impractical, if not impossible, to ensure the values of W and A after the jump to always fall on a grid point. Thus a two-dimensional interpolation is required. In this work we adopted the bi-cubic spline interpolation for accuracy and efficiency. Figure 1 illustrates the application of jump conditions.

It is natural to form a uniform grid in A so that optimal withdrawal strategies can be tested on a constant increment $\delta A = A_{j+1} - A_j$, as has been done successfully in [7] for pricing of a basic GMWB specified by (14) and (15). However, extensive numerical tests show that if a uniform grid in A is used for pricing GMAB with ratchets and optimal withdrawals (our numerical example in Section 7), then neither linear interpolation nor cubic interpolation in A can achieve an efficient convergence in pricing results. A very fine mesh has to be used before we see a stable solution, which can take up to several hours to obtain a fair fee, in sharp contrast to the basic GMWB where less than one minute computer time is required. On the other hand, if we make the grid in A uniform in $\ln A$ and use a linear or cubic interpolation based on variable Y, then we obtain a very good convergence on a moderately fine grid and the CPU time for a fair fee is about 30 min (a few minutes for a fair price). The CPU used for all the calculations in this study is Intel(R) Core(TM) i5-2400 @3.1GHz.

As we have already mentioned, a two-dimensional interpolation has to be used for applying the jump condition. We suggest to use either a bi-linear interpolation or a bi-cubic spline interpolation, e.g., see (see Section 3.6 in [44]), in both cases applied on the log-transformed state variables $X = \ln W$ and $Y = \ln A$. For numerical examples in this paper, we have adapted the more accurate bi-cubic spline interpolation for all the numerical results.
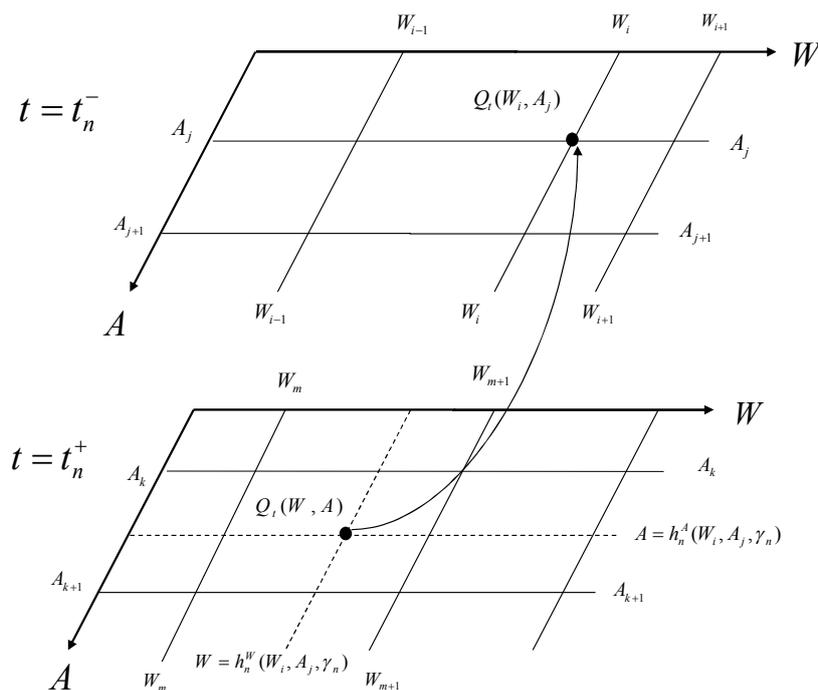
Figure 1. Illustration of the application of jump condition. The value $Q_t(W_i, A_j)$ at $t = t_n^-$ and at the node point (W_i, A_j) equals to $Q_t(W, A)$ at $t = t_n^+$ with $W = h_n^W(W_i, A_j, \gamma_n)$ and $A = h_n^A(W_i, A_j, \gamma_n)$. The point (W, A) is located inside the grid bounded by (W_m, W_{m+1}) and (A_k, A_{k+1}).

For uniform grids, the bi-cubic spline is about five times as expensive in terms of computing time as the one-dimensional cubic spline. Suppose the jump condition requires the value $Q(W, A)$ at the point (W, A) located inside a grid: $W_i \leq W \leq W_{i+1}$ and $A_j \leq A \leq A_{j+1}$. Equivalently, the point $X = \ln W$ and $Y = \ln A$ is inside the grid: $\ln W_i = x_i \leq X \leq x_{i+1} = \ln W_{i+1}$ and $\ln A_j = y_j \leq Y \leq y_{j+1} = \ln A_{j+1}$. Because the grid is uniform in both X and Y variables, the second derivatives $\partial^2 Q / \partial X^2$ and $\partial^2 Q / \partial Y^2$ can be accurately approximated by the three-point central difference, and consequently the one-dimensional cubic spline on a uniform grid involves only four neighboring grid points for any single interpolation. For the bi-cubic spline, we can first obtain $Q(\cdot, \cdot)$ at four points $Q(W, A_{j-1})$, $Q(W, A_j)$, $Q(W, A_{j+1})$, $Q(W, A_{j+2})$ by applying the one-dimensional cubic spline on the dimension $X = \log W$ for each point and then we can use these four values to obtain $Q(W, A)$ through a one-dimensional cubic spline in $Y = \log A$. Thus five one-dimensional cubic spline interpolations are required for a single bi-cubic spline interpolation, which involves sixteen grid points neighboring (W, A) point.

6.5. Calculating Greeks for Hedging

Calculation of the contract price in (30) under the risk-neutral probability measure $\mathbb{Q}$ means that one can find a portfolio replicating the VA guarantee (assuming mortality risk is fully diversified), i.e., perform hedging eliminating the financial risk. Finding correct hedging depends on the underlying stochastic model for the risky asset. The basic hedging is the so-called *delta hedging* eliminating randomness due to stochasticity in the underlying risky asset $S(t)$. Here, we use $S(t)$ as a tradable asset to hedge the exposure of the guarantee to the wealth account $W(t)$. One can construct a portfolio consisting of the money market account and $\Delta_S(t)$ units of $S(t)$, so that $\Delta_S(t)S(t) = \Delta_W(t)W(t)$, where $\Delta_W(t)$ is the number of units of the wealth account referred as *Delta*. Denote the value of the VA guarantee as $U_t(W, A)$ which is just a difference between the contract value with the guarantee $Q_t(W, A)$ and the value of the wealth account W, i.e.,

$$Q_t(W, A) = U_t(W, A) + W \tag{43}$$

Then, under the delta hedging strategy, one has to select

$$\Delta_W(t) = \frac{\partial U_t(W, A)}{\partial W} \Leftrightarrow \Delta_S(t) = \frac{\partial U_t(W, A)}{\partial S} \frac{W}{S}$$

for time t between the contract event times. Of course if there are extra stochasticities in the model such as stochastic interest rate and/or stochastic volatility, delta hedging will not eliminate risk completely and hedging with extra assets will be required which is model specific. See, e.g., [10], for constructing hedging in the case of regime switching stochastic volatility and interest rate. A popular active hedging strategy in the case of extra stochastic factors is the minimum variance hedging strategy, where $\Delta_W(t)$ is selected to minimize the variance of portfolio's instantaneous changes, e.g., applied in [11] for hedging GLWM in the case of stochastic volatility model.

Practitioners also calculate other sensitivities (partial derivatives) of the contract with respect to the interest rate and volatility (referred to as *Rho* and *Vega*) and even second partial derivatives such as *Gamma* $= \partial^2 U_t(W, A)/\partial W^2$ to improve hedging strategies. Here, we refer the reader to the standard textbooks in the area of pricing financial derivatives such as [30] or [31].

Numerical estimation of the contract sensitivities (referred to as *Greeks*) is more difficult than estimation of the contract price. A general standard approach to calculate the Greeks is to perturb the relevant parameter and re-calculate the price. Then one can use a two-point central difference to estimate the first derivatives and a three-point central difference for the second derivatives. In general, the finite difference PDE (or direct integration) methods generally produce superior accuracy in calculating Greeks when compared to the Monte Carlo method (at least for low dimensions when the finite difference method is practical or direct integration is possible). For *Delta* and *Gamma*, the finite difference method (or direct integration method) yields second order accurate values without re-calculating price using prices already calculated at the uniform grid points.

More accurate calculation of the main Greeks, *Delta* and *Gamma*, can be achieved using the so-called *likelihood method* as follows. The contract price at t_0 is calculated in the last time step (t_0, t_1) in backward induction as an integral (39). Differentiating (39) with respect to $W(0) = w_0$, *Delta* can be found as

$$
\begin{aligned}
\frac{\partial Q_{t_0^+}(w_0, A)}{\partial w_0} &= \int \widetilde{Q}_{t_1^-}(w, A) \frac{\partial \widetilde{p}_1(w|w_0)}{\partial w_0} dw \\
&= \int \widetilde{Q}_{t_1^-}(s, A) \frac{\partial \ln \widetilde{p}_1(w|w_0)}{\partial s_0} \widetilde{p}_1(w|w_0) dw \\
&= \mathrm{E}_{t_0^+}^{\mathbb{Q}} \left[\widetilde{Q}_{t_1^-}(W, A) \frac{\partial \ln \widetilde{p}_1(W|w_0)}{\partial w_0} \middle| W(0) = w_0, A \right]
\end{aligned}
\tag{44}
$$

Thus it can be calculated using the same direct integration method as used for $Q_{t_0^+}(w_0, A)$ with the factor $\partial \ln \widetilde{p}_1(W|w_0)/\partial w_0$ added to the integrand. Similarly, the required derivative to calculate *Gamma* can be found as

$$
\begin{aligned}
\frac{\partial^2 Q_{t_0^+}(w_0, A)}{\partial w_0^2} &= \int \widetilde{Q}_{t_1^-}(w, A) \frac{\partial}{\partial w_0} \left[\frac{\partial \ln \widetilde{p}_1(w|w_0)}{\partial w_0} \widetilde{p}_1(w|w_0) \right] dw \\
&= \int \widetilde{Q}_{t_1^-}(w, A) \left[\left(\frac{\partial \ln \widetilde{p}_1(w|w_0)}{\partial w_0} \right)^2 + \frac{\partial^2 \ln \widetilde{p}_1(w|w_0)}{\partial w_0^2} \right] \widetilde{p}_1(w|w_0) dw \\
&= \mathrm{E}_{t_0^+}^{\mathbb{Q}} \left[\widetilde{Q}_{t_1^-}(W, A) \left[\left(\frac{\partial \ln \widetilde{p}_1(w|w_0)}{\partial w_0} \right)^2 + \frac{\partial^2 \ln \widetilde{p}_1(w|w_0)}{\partial w_0^2} \right] \middle| W(0) = w_0, A \right]
\end{aligned}
\tag{45}
$$

Note, the above integrations for Delta and Gamma are only required for the (t_0, t_1) time step and for a single grid point $W(0) = w_0$. Here, t_0 should be understood as the current contract valuation time rather than time when the contract was initiated. Equivalently, for the PDE approach using finite

difference method, one can sometimes derive the corresponding PDEs for the Greeks and solve these PDEs for the last time step, see, e.g., [45]. Similarly for Monte Carlo method, simulations used to calculate the price can be used to calculate *Delta* and *Gamma* weighted with extra factors under the expectations in (44) and (45).

It is illustrative to show how to derive a hedging portfolio under the basic Black-Scholes model. Here, we assume that the underlying risky asset $S(t)$ follows

$$dS(t) = \mu(t)S(t) + \sigma(t)S(t)d\mathcal{B}^*(t) \tag{46}$$

where $\mathcal{B}^*(t)$ is the standard Brownian motion under the physical (*real*) probability measure, and $\mu(t)$ is the real drift. Then the wealth account evolution is

$$dW(t) = (\mu(t) - \alpha)W(t) + \sigma(t)W(t)d\mathcal{B}^*(t)$$

Here, we assume a continuously charged fee proportional to the wealth account but it is not difficult to deal with the case of discretely charged fees.

To hedge, the guarantee writer takes a long position in Δ_S units of $S(t)$, i.e., forms a portfolio

$$\Pi_t = -U_t(W, A) + \Delta_S \times S$$

By Ito's lemma, the changes of portfolio within (t_{n-1}, t_n), $n = 1, \ldots, N$ are

$$d\Pi_t = -\left(\frac{\partial U_t}{\partial t} + \frac{\partial U_t}{\partial W}dW + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 U_t}{\partial W^2}\right)dt - \Delta_S dS + \alpha W dt \tag{47}$$

where the last term $\alpha W dt$ is the fee amount collected over dt. Setting

$$\Delta_S = \frac{\partial U_t(W, A)}{\partial W}\frac{W}{S} = \left(\frac{\partial Q_t(W, A)}{\partial W} - 1\right)\frac{W}{S} \tag{48}$$

eliminates all the random terms in (47), making the portfolio locally riskless. This means that the portfolio earns at the risk-free interest rate $r(t)$, i.e., $d\Pi_t = r\Pi_t dt$, leading to the PDE

$$\frac{\partial U_t}{\partial t} + (r - \alpha)W\frac{\partial U_t}{\partial W} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 U_t}{\partial W^2} - rU_t - \alpha W = 0 \tag{49}$$

Substituting $U_t(W, A) = Q_t(W, A) - W$ in the above gives the PDE for $Q_t(W, A)$, the total value for the contract with the guarantee, i.e., the same as (38). Recalling Feynman-Kac theorem, it is easy to see that the stochastic process for W corresponding to this PDE is the risk-neutral process (9).

7. Numerical Example: GMAB Pricing

Numerical solutions for pricing VA riders involve many complicated numerical procedures and features. These are more involved when compared to pricing of most exotic derivatives in financial markets. It is important that these solutions are properly tested and validated. As a numerical example for illustration, using direct integration method (GHQC), we calculate accurate prices of GMAB with possible annual *ratchets*, allowing optimal withdrawals and no death benefit as specified in Section 4.3. With these features, the GMAB rider is very close to the real product marketed in Australia by, e.g., [27,28]. We assume geometric Brownian motion model for the risky asset (9). When applicable, we compare results with the MC and finite difference PDE methods. Numerical difficulties encountered in pricing this GMAB rider are common across other VA guarantees. Also, comprehensive numerical pricing results for this particular product are not available in the literature. These validated results (reported for a range of parameters) can serve as a numerical benchmark for practitioners and researchers developing numerical pricing of the VA with guarantees. We consider four GMAB types:

1. GMAB with the annual ratchets but no withdrawals. In this case, a standard MC method can be used to compare with GHQC results—this is a good validation of the implemented numerical functions related to the *ratchet* feature, in addition to validating the numerical integration based on the Gauss-Hermite quadrature.
2. GMAB with the annual ratchets and regular withdrawals of a fixed percentage of the wealth account. In this case a standard MC method can also be used to compare with GHQC results—this is a good validation of implemented numerical functions related to jump conditions due to both *ratchet* and *withdrawal* features. In addition, in order to test the numerical functions related to the application of penalties, we assume a pension account where the static withdrawal rate is set above the penalty threshold.
3. GMAB with the optimal quarterly withdrawals and the annual ratchets for a *super account*, where the penalty (24) is applied on any withdrawal γ_n when $W(t_n^-) < A(t_n^-)$.
4. GMAB with the optimal quarterly withdrawals and the annual ratchets for a *pension account*, where the penalty (25) is applied if the withdrawal γ_n is above the penalty threshold G_n and if $W(t_n^-) < A(t_n^-)$.

As a comparison, results from our PDE finite difference implementation will also be shown for Case 4, the most complicated example among the four listed above. In addition, we will also calculate results for Case 4 in the case of the guarantee fee charged discretely (quarterly). All reported GHQC results are based on $q = 9$ quadrature points. We did not observe any material difference in results if q is increased further. Results based on $q = 5$ are also very accurate.

7.1. GMAB with Ratchet and No Withdrawal

Consider a GMAB rider with the annual ratchet and no withdrawals. In this case a standard MC method can be used to compare with GHQC results which is a good validation of implemented numerical functions related to the *ratchet* feature. Table 1 compares GHQC and MC results for the fare fee α of GMAB with the annual ratchet for the interest rate r ranging from 1% to 7% and the volatility $\sigma = 10\%$ and 20%. The maximum relative difference between the two methods is 0.76% at the interest rate $r = 5\%$ and the volatility $\sigma = 10\%$. The maximum absolute difference between the two methods is one basis point at the lowest interest rate $r = 1\%$ where the fee is the highest. On average, the relative difference is 0.52% and the absolute difference is 0.5 basis point, which is 5 cents per year on a one thousand dollar account. The GHQC results are obtained with the mesh size $M = 400$ and $J = 200$, and on the average it takes 22 s per price (calculation of the fare fee requires iterations over several prices). The MC results are obtained using 20 million simulations and it takes about 62 s per price.

Table 1. Fair fee α in bp (1 bp = 0.01%) as a function of the interest rate r for the Guaranteed Minimum Accumulation Benefit (GMAB) rider with the annual ratchet and no withdrawal. The contract maturity is $T = 10$ years. $\widetilde{\delta}$ is the relative difference between Monte Carlo (MC) and GHQC method results.

Interest Rate, r	$\sigma = 10\%$			$\sigma = 20\%$		
	GHQC, bp	MC, bp	$\widetilde{\delta}$	GHQC, bp	MC, bp	$\widetilde{\delta}$
1%	337.2	338.2	0.30%	998.7	999.8	0.11%
2%	186.0	186.8	0.43%	637.1	637.7	0.09%
3%	116.8	117.3	0.43%	458.0	458.5	0.11%
4%	77.94	78.31	0.47%	346.9	347.5	0.17%
5%	53.91	54.32	0.76%	271.1	271.6	0.18%
6%	38.54	38.77	0.60%	216.3	216.7	0.18%
7%	28.11	28.30	0.68%	175.1	175.3	0.34%

The agreement between the two methods at $\sigma = 20\%$ is also very good. In absolute terms, the maximum difference between the two methods is 1.1 basis point at the lowest interest rate $r = 1\%$. In relative terms, the maximum difference between the two methods is 0.34% at the highest interest

rate $r = 7\%$. On average, the relative difference is 0.17% which is significantly smaller than the corresponding case at $\sigma = 10\%$.

7.2. GMAB with Ratchet and Static Withdrawal

Consider a GMAB rider with the annual ratchet and a regular quarterly withdrawals of a fixed percentage of the wealth account. In this case a standard MC method can also be used to compare with the GHQC results which is a good validation of implemented numerical functions related to jump conditions due to both *ratchet* and *withdrawal* features. Here, we consider a pension type account with the penalty given by (25). In this case, regular withdrawals at a pre-determined percentage level are allowed. In order to test the numerical functions related to the application of penalty, we also consider the static withdrawal above a pre-determined threshold level that will attract a penalty. We set the withdrawal threshold at 15% of the wealth account per annum, and the withdrawal frequency is quarterly, i.e., the quarterly withdrawal threshold is $G_n = 3.75\%$ of the wealth account.

In the first test we allow a regular quarterly withdrawal of 3.75% of the wealth account balance, i.e., $\gamma_n = G_n$ and there is no penalty on the withdrawals. Table 2 compares the GHQC and MC results for the fare fee α. In relative terms, the maximum difference between the two methods is 0.08% at interest rate $r = 5\%$. On average, the relative difference is 0.06% and the absolute difference is 0.3 basis point. The GHQC results are obtained with the mesh size $M = 400$ and $J = 400$, and the MC results are obtained with 20 million simulations per price.

Table 2. Fair fee α in bp as a function of interest rate r for the GMAB with the annual ratchet and static quarterly withdrawal of 3.75% and 4% of the wealth account (15% and 16% annually respectively). The penalty threshold (pension type account) is set at 15% annually. The contract maturity is $T = 10$ years and volatility is $\sigma = 20\%$. $\widetilde{\delta}$ is the relative difference between Monte Carlo (MC) and GHQC results.

Interest Rate, r	15% Annual Withdrawal			16% Annual Withdrawal		
	GHQC, bp	MC, bp	$\widetilde{\delta}$	GHQC, bp	MC, bp	$\widetilde{\delta}$
1%	1084	1085	0.09%	185.3	185.3	<0.01%
2%	669.1	669.5	0.06%	152.9	152.9	<0.01%
3%	464.1	464.4	0.06%	126.6	126.6	<0.01%
4%	339.0	339.2	0.06%	105.1	105.1	<0.01%
5%	255.0	255.2	0.08%	87.54	87.51	0.03%
6%	195.7	195.7	<0.01%	73.21	73.14	0.1%
7%	152.1	152.2	0.07%	61.40	61.36	0.07%

Comparing with Table 1, the fair fee for the static withdrawal is about 8% higher than the corresponding no-withdrawal case at the lowest interest rate $r = 1\%$, but it is about 13% lower than the corresponding no-withdrawal case at $r = 7\%$. We have also tested static 2.5% quarterly withdrawals and obtained the same pattern: at the lower interest rate the fair fee of static withdrawal (which is also below the penalty threshold) is higher than the corresponding no-withdrawal case, and at the higher interest it is the opposite. These differences in the fair fees at relatively low and high interest rates can be broadly interpreted as follows. At the lower interest rates, where the expected capital growth is relatively slow, it is better to perform a regular withdrawal at or below the penalty threshold and take the protected capital at the maturity. However, at the higher interest rates, where the expected capital growth rate is also high, it is beneficial not to carry out a regular withdrawal and keep the capital to grow.

The above test also demonstrates that the MC and GHQC methods agree very well for pricing GMAB with a static withdrawal not exceeding the penalty threshold. This confirms the accuracy and efficiency of our numerical implementation of the jump condition using a bi-cubic interpolation in GHQC method.

In the second test of static withdrawal, we allow a regular quarterly withdrawal of 4% (16% per annum), i.e., the annual withdrawal rate is slightly higher than the penalty threshold of 15% per annum and there is a penalty applied for each withdrawal. The GHQC and MC results for this test are also presented in Table 2. In absolute terms, the maximum difference between the two methods is only 0.07 basis point at the interest rate $r = 6\%$, which is less than 1 cent per year for a one thousand dollar account. In relative terms, the maximum difference between the two methods is 0.1% at the interest rate $r = 6\%$. On average the relative difference is only 0.03%, which shows the two methods also agree very well in the case of excessive static withdrawals, where the penalty is applied for each withdrawal. This is a very convincing validation that our GHQC implementation of all numerical functions associated with the jump conditions, including the bi-cubic spline interpolation, is correct and accurate. The above tests are very close to validation of the entire algorithm in the case of optimal withdrawals. This is because in pricing the optimal withdrawal case, exactly the same integration and interpolation functions are used, and the only extra step required is to find the withdrawal rate maximizing the price. Nevertheless, for optimal strategy cases we will carry out some further validations by comparing results between GHQC and finite difference PDE methods.

Comparing Table 2 with Table 1 for the no-withdrawal case, the fair fee for the static withdrawal in excess of the penalty threshold is dramatically reduced: it is reduced by about 80% from the corresponding no-withdrawal case at $r = 1\%$, and it is reduced by about 65% at $r = 7\%$. Thus, a regular withdrawal above the penalty threshold is a very bad strategy regardless the interest rate level. In this instance, the penalty takes away the protected capital on a regular basis. Note that, the penalty is applied in terms of the whole withdrawal amount, not just on the exceeded part, see penalty function (25). Thus a slight excess over the penalty threshold can cause a large change in the price or in the fair fee, as observed in the second test.

The above two tests show that a regular static withdrawal is only slightly beneficial at very low interest rate and only when the withdrawal rate does not exceed the penalty threshold. In the next section, it will be demonstrated by numerical results that an optimal withdrawal is always beneficial regardless the interest rate level and penalties.

7.3. Optimal Withdrawal—Super Account

Consider a GMAB for a super account with the annual ratchet and assume that the policyholder can exercise an optimal withdrawal strategy quarterly. For a super account, any withdrawal will penalize the protected capital amount (benefit base) if the wealth account is below the benefit base according to (23) and (24).

Table 3 shows the fair fee for a super account as a function of the interest rate at $\sigma = 10\%$ and $\sigma = 20\%$. The columns under $\widetilde{\varepsilon}$ show an extra percentage value in the fee due to optimal withdrawal when compared to the static case of no withdrawal in Table 1. The results show, the extra fee is only about one or two percent for most cases, except at the low interest rate and high volatility. This extra fee due to optimal withdrawal is insignificant for the super account in most cases, mainly due to the heavy penalties applied. As will be shown in the next section, if the penalty is less severe as in the case of pension account, the extra fee becomes much more significant. If the penalty is completely removed, then numerical experiments show that the extra fee will be several times, e.g., 300%, larger than the static case, demonstrating the full value of the optimal strategy.

Table 3. Fair fee α in bp (1 bp = 0.01%) as a function of interest rate r for the GMAB on a super account when withdrawals are optimal. The contract maturity is $T = 10$ years. $\widetilde{\varepsilon}$ is the percentage difference between the fair fee in the case optimal withdrawal and the fair fee in the static case (no withdrawal) from Table 1.

Interest Rate, r	$\sigma = 10\%$		$\sigma = 20\%$	
	Fee (b.p)	$\widetilde{\varepsilon}$	Fee (b.p)	$\widetilde{\varepsilon}$
1%	370.7	10%	1235	23.7%
2%	191.2	2.8%	700.1	9.89%
3%	118.1	1.2%	478.8	4.54%
4%	78.52	0.9%	355.5	2.48%
5%	54.47	1.1%	275.2	1.51%
6%	39.00	1.48%	218.8	1.16%
7%	28.38	1.36%	176.9	1.03%

7.4. Optimal Withdrawal—Pension Account

Consider a GMAB for a pension account with the annual ratchet and assume that the policyholder can exercise an optimal withdrawal strategy quarterly. For a pension account, any withdrawal above a pre-defined withdrawal level G_n will penalize the protected capital amount $A(t)$ if the wealth account $W(t)$ is below $A(t)$ according to (23) and (25). Here, we set the annual withdrawal limit at 15% of the wealth account, i.e., $G_n = 0.25 \times 15\% = 3.75\%$ is quarterly withdrawal threshold.

Table 4 shows the fair fee of GMAB for a pension account as a function of the interest rate r when $\sigma = 10\%$ and 20%. The columns under $\widetilde{\varepsilon}$ show an extra percentage value in the fee due to optimal withdrawal when compared to the static case of no withdrawal in Table 1. The results show, the extra fee ranges from about 9% at the highest interest rate $r = 7\%$ to about 39% at the lowest interest rate $r = 1\%$. This extra fee is much more significant than in the case of super account, see Table 3, apparently due to reduced penalties. At $\sigma = 20\%$, the extra fee ranges from about 9% at the highest interest rate $r = 7\%$ to about 46% at the lowest interest rate $r = 1\%$. This extra fee is higher than in the case of lower volatility $\sigma = 10\%$, in both percentage and absolute terms.

Also, in Table 4, the numbers in the parentheses next to the continuous fair fee values α are the GHQC results for the discretely charged fair fee $\alpha_d = -\log(1 - \widetilde{\alpha}dt_n)/dt_n$, where at the end of each quarter t_n, the policyholder wealth account is charged a fee proportional to the account value $\widetilde{\alpha}dt_nW(t_n^-)$, see the wealth process (7). Results show only little difference between the continuous fee α and the discrete fee α_d. On average, the relative difference is 0.15%. Of course, at a higher frequency of charging fee (e.g., a monthly fee), the difference will be even less.

The last column in Table 4 shows the continuous fee calculated using our PDE finite difference (FD) method. The agreement between the GHQC (quadrature method) and FD method is very good. The average relative difference in the fair fees between the two methods is 0.20%. Figure 2 show the curves of the fair fee for a pension account as a function of interest rate r using results from Table 4, in comparison with the static case (no withdrawal) from Table 1.

For some comparison, the market fees offered by [28] are 1.75% for a 10 year capital protection of a "balanced portfolio" and 0.95% for a "conservative growth portfolio"; and fees offered by [27] are 1.3% for a 10 year capital protection of a "balanced strategy" portfolio and 0.95% for a "moderately defensive strategy" portfolio. Though the values of volatility are not known for these market portfolios, it seems that market prices are significantly lower than the *fair fee*, which is also observed in the literature before; e.g., see [1,6,39].

For results in Table 4, the average CPU time per price is 108 s for GHQC and 138 s for FD. Since all the cases have the same maturity and withdrawal frequency, the variation of CPU time among the cases under the same method is negligible. Note the speed advantage of GHQC over FD, although still significant, is not as great as in the case of the basic GMWB reported in [7], where GHQC is several times faster than FD. This is because a much higher proportion of CPU time is spent on the

two-dimensional cubic spline interpolation for applying the jump condition in the case of GMAB, and GHQC and FD use identical interpolation functions.

Table 4. Fair fee α in bp (1 bp = 0.01%) as a function of interest rate r for the GMAB for a pension account when withdrawals are optimal. The contract maturity is $T = 10$ years and quarterly withdrawal limit is $G_n = 0.25 \times 15\%$. $\widetilde{\varepsilon}$ is the percentage difference between the fair fee in the case optimal withdrawal and the fair fee in the static case (no withdrawal) from Table 1. The numbers in the parentheses next to the continuous fair fee are the results for the quarterly charged fees.

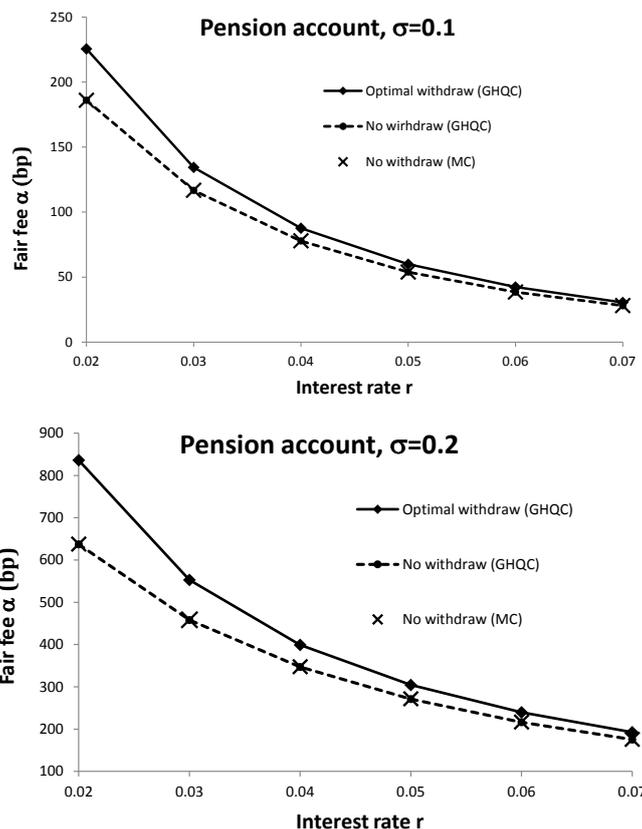| Interest Rate, r	$\sigma = 10\%$		$\sigma = 20\%$		
	Fee (b.p)	$\widetilde{\varepsilon}$	Fee (b.p)	$\widetilde{\varepsilon}$	Fee-FD (b.p)
1%	472.6	39%	1474 (1479)	46%	1466
2%	227.7	21%	836.1 (836.3)	30%	833.7
3%	135.4	15%	552.8 (553.6)	20%	551.7
4%	88.15	13%	399.1 (399.7)	14%	398.6
5%	60.24	11%	304.3 (304.7)	12%	304.0
6%	42.58	10%	239.6 (239.9)	10%	239.4
7%	30.63	9%	192.5 (192.8)	9%	192.4

Figure 2. Fair fee α in bp as a function of the interest rate r at $\sigma = 10\%$ and $\sigma = 20\%$ for the GMAB on a pension account when withdrawals are optimal, in comparison with the static case where no withdrawal is allowed. Values are taken from Tables 1 and 4, respectively.

8. Conclusions

In this paper we have reviewed the pricing VA riders and presented a unified pricing approach via an optimal stochastic control framework. We discussed different models and numerical procedures applicable in general to most of the VA riders with various contractual specifications. To price

these VA riders under the geometric Brownian motion model for the risky asset, often assumed in practice, we have extended and generalized the direct integration method based on the Gauss-Hermite quadrature, introduced earlier in [7] for some specific and simpler product specifications.

As an example, we presented a numerical valuation of capital protection guarantees (GMAB riders), with specifications matching closely the real market products offered in Australia by, e.g., [27,28]. Numerical valuation of this guarantee involves all the main numerical difficulties encountered in pricing other VA riders, such as ratchets and optimal withdrawals. Numerical results have been validated by MC and finite difference PDE methods and can serve as a numerical benchmark for practitioners and researchers developing numerical pricing of VA riders to assess the accuracy of numerical implementations. As expected, we observed that the extra fee that has to be charged to counter the optimal policyholder behavior is most significant at lower interest rates and higher volatility levels, and it is very sensitive to the penalty withdrawal threshold.

As we have already discussed in Section 5.3, the fee based on the optimal policyholder withdrawal is the worst case scenario for the issuer. That is, if the guarantee is hedged perfectly to eliminate the financial risk and mortality risk is fully diversified then this fee will *ensure* no losses for the issuer (in other words a full protection against the policyholder strategy and the market uncertainty). If the issuer hedges continuously but investors deviate from the optimal strategy, then the issuer will receive a guaranteed profit. Any strategy different from the optimal is sup-optimal and will lead to smaller fair fees. Of course the strategy optimal in this sense is not related to the policyholder circumstances. The policyholder may act optimally with respect to his preferences and circumstances which may be different from the optimal strategy maximizing losses for the policy issuer. Life-cycle modelling can be undertaken to analyze and estimate sub-optimality of policyholder behaviors. However, development of the secondary markets for insurance products may expose the policy issuers to some significant risk if a fee for the guarantees is not charged to cover the worst case scenario.

It is important to note that the guarantee could be written on more than one asset (e.g., on several mutual funds). In this case it is still common for practitioners to use a single-asset proxi model to calculate the price and hedging parameters. Obviously such approach has significant drawbacks (e.g., the sum of geometric Brownian motions is not a geometric Brownian motion). The PDE and direct integration methods are not practical in high-dimensions and thus one has to rely on the MC methods to treat multi-asset case accurately. In the case of static withdrawals, it is not difficult to consider a full multi-asset model and calculate the price using a standard MC as in [46]. However, in the case of optimal withdrawal strategies, numerical valuation in the multi-asset case will require the development of a regression type MC solving the backward recursion for processes affected by the withdrawals. One could apply *control randomization* methods extending the Least-Squares MC developed in [16], but the accuracy and robustness of this method for pricing VA riders have not been studied yet.

The specification details of VA riders typically vary across different companies and are difficult to extract and compare from the very long product specification documents. Moreover, results for specific GMxB riders presented in the academic literature often refer to different specifications. As a result, cross-validation and benchmark research studies are rare. Given that numerical solutions used for pricing of VA riders are complex, it is important that these solutions are properly tested and validated. Moreover, new products are appearing in the VA market regularly with increasing complexity that raises an important question, as discussed in [47] and mentioned in [23], whether new complex products are designed to suite the policyholder needs better or introduced for the purpose of *obfuscation*.

Author Contributions: Both authors contributed equally to this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bauer, D.; Kling, A.; Russ, J. A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bull.* **2008**, *38*, 621–651.
2. Kalberer, T.; Ravindran, K. *Variable Annuities: A Global Perspective*; Risk Books: London: UK, 2009.
3. Ledlie, M.; Corry, D.; Finkelstein, G.; Ritchie, A.; Su, K.; Wilson, D. Variable annuities. *Br. Actuar. J.* **2008**, *14*, 327–389.
4. Chen, Z.; Forsyth, P. A numerical scheme for the impulse control formulation for pricing variable annuities with a guaranteed minimum withdrawal benefit (GMWB). *Numer. Math.* **2008**, *109*, 535–569.
5. Dai, M.; Kwok, K.Y.; Zong, J. Guaranteed minimum withdrawal benefit in variable annuities. *Math. Financ.* **2008**, *18*, 595–611.
6. Milevsky, M.A.; Salisbury, T.S. Financial valuation of guaranteed minimum withdrawal benefits. *Insur. Math. Econ.* **2006**, *38*, 21–38.
7. Luo, X.; Shevchenko, P.V. Fast numerical method for pricing of variable annuities with guaranteed minimum withdrawal benefit under optimal withdrawal strategy. *Int. J. Financ. Eng.* **2015**, *2*, doi:10.1142/S2424786315500243.
8. Luo, X.; Shevchenko, P.V. Valuation of variable annuities with guaranteed minimum withdrawal and death benefits via stochastic control optimization. *Insur. Math. Econ.* **2015**, *62*, 5–15.
9. Bacinello, A.; Millossovich, P.; Olivieri, A.; Pitacco, E. Variable annuities: A unifying valuation approach. *Insur. Math. Econ.* **2011**, *49*, 285–297.
10. Forsyth, P.; Vetzal, K. An optimal stochastic control framework for determining the cost of hedging of variable annuities. *J. Econ. Dyn. Control* **2014**, *44*, 29–53.
11. Huang, Y.; Kwok, Y.K. Regression-based Monte Carlo methods for stochastic control models: Variable annuities with lifelong guarantees. *Quant. Financ.* **2015**, doi:10.1080/14697688.2015.1088962.
12. Luo, X.; Shevchenko, P.V. Valuation of variable annuities with guaranteed minimum withdrawal benefit under stochastic interest rate. arXiv:1602.03238, 2016.
13. Azimzadeh, Y.; Forsyth, P.A. The existence of optimal bang-bang controls for GMxB contracts. *SIAM J. Financ. Math.* **2015**, *6*, 117–139.
14. Luo, X.; Shevchenko, P.V. Variable annuity with GMWB: Surrender or not, that is the question. In Proceedings of the 21st International Congress on Modelling and Simulation MODSIM2015, Gold Coast, Australia, 29 November–4 December 2015; Weber, T., McPhee, M.J., Anderssen, R.S., Eds.; Modelling and Simulation Society of Australia and New Zealand: Perth, Australia, 2015; pp. 959–965; ISBN: 978-0-9872143-5-5.
15. Longstaff, F.; Schwartz, E. Valuing American options by simulation: A simple lieast-squares approach. *Rev. Financ. Stud.* **2001**, *14*, 113–147.
16. Kharroubi, I.; Langrené, N.; Pham, H. A numerical algorithm for fully nonlinear HJB equations: An approach by control randomization. *Monte Carlo Methods Appl.* **2014**, *20* 145–165.
17. Hilpert, C.; Li, J.; Szimayer, A. The effect of secondary markets on equity-linked life insurance with surrender guarantees. *J. Risk Insur.* **2014**, *81*, 943–968.
18. Knoller, C.; Kraut, G.; Schoenmaekers, P. On the propensity to surrender a variable annuity contract: An empirical analysis of dynamic policyholder behavior. *J. Risk Insur.* **2015**, doi:10.1111/jori.12076.
19. Gao, J.; Ulm, E.R. Optimal consumption and allocation in variable annuities with guaranteed minimum death benefits. *Insur. Math. Econ.* **2012**, *51*, 586–598.
20. Horneff, V.; Maurer, R.; Mitchell, O.S.; Rogalla, R. Optimal life cycle portfolio choice with variable annuities offering liquidity and investment downside protection. *Insur. Math. Econ.* **2015**, *63*, 91–107.
21. Moenig, T. Optimal Policyholder Behavior in Personal Savings Products and Its Impact on Valuation. Ph.D. Thesis, Risk Management and Insurance Dissertations, Georgia State University, Atlanta, GA, USA, 2012.
22. Steinorth, P.; Mitchell, O.S. Valuing variable annuities with guaranteed minimum lifetime withdrawal benefits. *Insur. Math. Econ.* **2015**, *64*, 246–258.
23. Moenig, T.; Bauer, D. Revisiting the risk-neutral approach to optimal policyholder behavior: A study of withdrawal guarantees in variable annuities. *Rev. Financ.* **2015**, doi:10.1093/rof/rfv018.

24. Fung, M.; Ignatieva, K.; Sherris, M. Systemic mortality risk: An analysis of guaranteed lifetime withdrawal benefits in variable annuities. *Insur. Math. Econ.* **2014**, *58*, 103–115.

25. Gaillardetz, P.; Lakhmiri, J. A new premium principle for equity indexed annuities. *J. Risk Insur.* **2011**, *78*, 245–265.

26. Luo, X.; Shevchenko, P.V. Fast and simple method for pricing exotic options using Gauss-Hermite quadrature on a cubic spline interpolation. *J. Financ. Eng.* **2014**, *1*, doi:10.1142/S2345768614500330.

27. AMP. North Super and Pension Guarantees: Product Disclosure Statement—Part B; AMP; Version 1 July 2014. Available online: http://www.amp.com.au (accessed on 25 March 2015).

28. MLC. MLC MasterKey Super & Pension Fundamentals. MLC. Version 1 July 2014. Available online: http://www.mlc.com.au (accessed on 13 November 2014).

29. Björk, T. *Arbitrage Theory in Continuous Time*; Oxford University Press: Oxford, UK, 2004.

30. Wilmott, P. *Paul Wilmott on Quantitative Finance*; John Wiley & Sons: Chichester, UK, 2000.

31. Hull, J.C. *Options, Futures, and Other Derivatives*; Pearson Education India: Delhi, India, 2006.

32. Pitacco, E.; Denuit, M.; Haberman, S.; Olivieri, A. *Modelling Longevity Dynamics for Pensions and Annuity Business*; Oxford University Press: Oxford, UK, 2009.

33. Lee, R.D.; Carter, L.R. Modeling and forecasting U.S. mortality. *J. Am. Stat. Assoc.* **1992**, *87*, 659–671.

34. Cairns, A.; Blake, D.; Dowd, K.; Coughlan, G.; Epstein, D.; Ong, A.; Balevich, I. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *N. Am. Actuar. J.* **2009**, *13*, 1–35.

35. Fung, M.C.; Peters, G.; Shevchenko, P.V. A unified Approach to Mortality Modelling Using State-Space Framework: Characterization, Identification, Estimation and Forecasting; Preprint. Available online: http://ssrn.com/abstract=2786559 (accessed on 31 May 2016).

36. Renshaw, A.; Haberman, S. A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insur. Math. Econ.* **2006**, *38*, 556–570.

37. Hirz, J.; Schmock, U.; Shevchenko, P.V. Modelling Annuity Portfolios and Longevity Risk with Extended CreditRisk Plus; Preprint. Available online: http://ssrn.com/abstract=2606203 (accessed on 19 January 2016).

38. Shevchenko, P.V.; Hirz, J.; Schmock, U. Forecasting leading death causes in Australia using extended CreditRisk+. In Proceedings of the 21st International Congress on Modelling and Simulation MODSIM2015, Gold Coast, Australia, 29 November–4 December 2015; Weber, T., McPhee, M.J., Anderssen, R.S., Eds.; Modelling and Simulation Society of Australia and New Zealand: Perth, Australia, 2015; pp. 966–972. ISBN: 978-0-9872143-5-5.

39. Chen, Z.; Vetzal, K.R.; Forsyth, P.A. The effect of modelling parameters on the value of GMWB guarantees. *Insur. Math. Econ.* **2008**, *43*, 165–173.

40. Marshall, C.; Hardy, M.; Saunders, D. Valuation of a guaranteed minimum income benefit. *N. Am. Actuar. J.* **2010**, *14*, 38–58.

41. Bélanger, A.; Forsyth, P.; Labahn, G. Valuing the guaranteed minimum death benefit clause with partial withdrawals. *Appl. Math. Financ.* **2009**, *16*, 451–496.

42. Milevsky, M.A.; Posner, S.E. The titanic option: Valuation of the guaranteed minimum death benefit in variable annuities and mutual funds. *J. Risk Insur.* **2001**, *68*, 93–128.

43. Bäuerle, N.; Rieder, U. *Markov Decision Processes with Applications to Finance*; Springer: Berlin, Germany, 2011.

44. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes in C*; Cambridge University Press: Cambridge, UK, 1992.

45. Tavella, D.; Randall, C. *Pricing Financial Instruments—The Finite Difference Method*; John Wiley & Sons: New York, NY, USA, 2000.

46. Ng, A.C.-Y.; Li, J.S.-H. Pricing and hedging variable annuity guarantees with multiasset stochastic investment models. *N. Am. Actuar. J.* **2013**, *17*, 41–62.

47. Carlin, B.I.; Manso, G. Obfuscation, learning, and the evolution of investor sophistication. *Rev. Financ. Stud.* **2011**, *24*, 754–785.