

X. o	NORTH AME	RICAN
TAN	ACTUARIAL	JOURNAL
X10		
17		
-Dell		
2		
• /\		
-7-	SOCIETY OF	D Routledge
	Actualles	IS territori

North American Actuarial Journal

ISSN: 1092-0277 (Print) 2325-0453 (Online) Journal homepage: https://www.tandfonline.com/loi/uaaj20

## **Efficient Nested Simulation for Conditional Tail Expectation of Variable Annuities**

Ou Dang, Mingbin Feng & Mary R. Hardy

To cite this article: Ou Dang, Mingbin Feng & Mary R. Hardy (2020) Efficient Nested Simulation for Conditional Tail Expectation of Variable Annuities, North American Actuarial Journal, 24:2, 187-210, DOI: 10.1080/10920277.2019.1636399

To link to this article: https://doi.org/10.1080/10920277.2019.1636399



Published online: 18 Oct 2019.



Submit your article to this journal 🗗

Article views: 122



View related articles



🌔 🛛 View Crossmark data 🗹



Check for updates

# Efficient Nested Simulation for Conditional Tail Expectation of Variable Annuities

Ou Dang, 🝺 Mingbin Feng, 🝺 and Mary R. Hardy 🝺

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Monte Carlo simulations—in particular, nested Monte Carlo simulations—are commonly used in variable annuity (VA) risk modeling. However, the computational burden associated with nested simulations is substantial. We propose an Importance-Allocated Nested Simulation (IANS) method to reduce the computational burden, using a two-stage process. The first stage uses a low-cost analytic proxy to identify the tail scenarios most likely to contribute to the Conditional Tail Expectation risk measure. In the second stage we allocate the entire inner simulation computational budget to the scenarios identified in the first stage. Our numerical experiments show that, in the VA context, IANS can be up to 30 times more efficient than a standard Monte Carlo experiment, measured by relative mean squared errors, when both are given the same computational budget.

#### 1. INTRODUCTION

Variable annuity (VA) guarantees range from the apparently straightforward embedded put options of standard guaranteed minimum maturity benefits (GMMB), through to complex combinations of path dependent, exotic lookback and tandem options, for example, in the guaranteed minimum income benefit (GMIB) studied by Marshall, Hardy, and Saunders (2010). For practitioners who must model the liabilities for valuation and hedging purposes, even fairly simple options are complex in terms of risk measurement and management. One major reason is that the hedge calculations generally use stochastic volatility models because of the long-term nature of the policies and because the guarantee costs are highly dependent on the tail outcomes of the underlying asset distribution. Constant or deterministic volatility models do not provide a good fit to stock prices in the long term and are particularly poor at capturing tail dynamics (Hardy 2001). The introduction of more sophisticated, dynamic models of policyholder behavior also adds to the computational burden associated with VA risk modeling.

In general, this level of complexity and dimensionality can be solved only using stochastic simulation, and it is now standard in the insurance industry to use Monte Carlo simulation to determine hedge portfolios. The computational burden is substantial but not impossible.

The problem comes when the insurer is required to do a higher level analysis of the distribution of future cash flows, including hedge costs, for a portfolio of VA policies. The context of this type of calculation would be in determining regulatory or economic capital. For example, in a Solvency II regulatory environment, insurers are required to project all their future cash flows for current policies, calculate the aggregate discounted net liability cash flows at each year end, for each simulated path of asset and liability experience, and evaluate the 99.5% Value at Risk (VaR) of the change in surplus each year. In North America, companies use a similar approach under their Own Risk Solvency Assessment obligations, but with the Conditional Tail Expectation (CTE)<sup>1</sup> risk measure taking the place of the VaR, to be consistent with the regulatory capital requirements. In Canada, VA reserves are generally set between the 60% and 80% CTE of the liability values from Monte Carlo simulation (Canadian Institute of Actuaries, 2017). The total gross calculated requirement is generally set at the 95% CTE so that the required capital is the difference between the 95%

Address correspondence to Ou Dang, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. E-mail: jessica.dang@uwaterloo.ca

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaaj.

<sup>&</sup>lt;sup>1</sup>The  $\alpha$ -CTE, which was introduced in Wirch and Hardy (1999), is identical to the TailVaR<sub> $\alpha$ </sub>, which is based on (but not identical to) a measure introduced in Artzner et al. (1999). Both terms are used to mean the expected value of a loss, conditional on the loss falling in the upper  $(1 - \alpha)$  part of its distribution.

CTE and the reserve (say, 80% CTE) (Office of the Superintendent of Financial Institutes Canada 2017). In the United States, the stochastic component of the reserve of VAs uses a 70% CTE and the minimum required capital uses a 90% CTE.

Projecting future hedge cash flows for economic or regulatory capital then requires a *nested Monte Carlo simulation* (also known as two-tier or stochastic-on-stochastic simulation). In a nested simulation for VA economic capital calculations, the first level, or *outer simulations*, are the simulated real-world paths for the underlying risk factors; in the case of VA guarantees, this would include the rate of growth of the policyholder's fund values as well as, potentially, policyholder behavior and interest rates. The time step for these projections would be at least as frequent as the expected interval between hedge rebalancing points, and the time horizon would typically be sufficient to run-off the current business. The second level, or *inner simulations* are used to determine the cost of hedging the guarantees at each future time point, based on the simulated risk factors under the outer simulation. Under the ideal conditions of the Black-Scholes-Merton world, VA hedges would be self-financing, with no requirement for additional economic or regulatory capital. In practice there are slippages in hedge portfolios, arising from basis risk (as the real-world stock price movements do not exactly follow the assumed model), from discrete hedge rebalancing intervals, and from the effects of policyholder behavior. At each rebalancing point, the value of the hedging portfolio brought forward from the previous period may be different from the value of the hedging portfolio required for the subsequent period. As a result, at each rebalancing point, the insurer may incur additional costs if the hedge brought forward from the previous period for the next period.

Nested simulations are computationally very burdensome. Consider a single VA contract with 20-year maturity and is dynamically hedged monthly. A Monte Carlo projection, based on a two-level nested simulation with 5,000 outer scenarios and 1,000 single-step inner simulations at each monthly rebalancing point, will require  $20 \times 12 \times 5,000 \times 1,000 = 1.2 \times 10^9$  total simulated asset or liability values; If the inner simulations are single step, and if each simulated value takes 1 µs  $(10^{-6} \text{ seconds})$  to complete, then it would take around 20 minutes to simulate the cash flows for a single policy. A typical block of business would involve potentially tens of thousands of VA contracts. If the inner simulations are stepwise to the end of the 20-year term, the total number of simulated cash flows increases by a factor of around 120. Successful management of VA risks requires the insurer to be able to run the economic capital models quickly and frequently; a model that takes multiple days to complete a single set of projections is not as useful as one that is marginally less accurate but can run in a few hours. It is not surprising that there is considerable industry interest in techniques for reducing the number of simulation points required for VA risk measurement; see, for example, Cathcart and Morrison (2009) and Feng, Cui, and Li (2016), which was commissioned by the Society of Actuaries. In addition, there are other contexts in financial risk management where nested simulations are required or desirable. Risk measurement in derivatives portfolios of banks, using nested simulation, is the topic of Gordy and Juneja (2010), Liu and Staum (2010) and Broadie, Du, and Moallemi (2011). Solvency 2 Solvency Capital Requirement calculations are the topic of Bauer, Reuss, and Singer (2012).

The number of individual simulated asset/liability cash flows required in a nested simulation is the product of (1) the number of contracts in a portfolio, (2) the number of outer paths, (3) the number of inner paths, and (4) the average number of time steps in each inner simulation. Recent research efforts to address the computation challenges in nested simulation focus on reducing one or more of the above four factors. Gan (2013, 2015a,b); Gan and Lin (2015, 2017) proposed using clustering algorithms to select *representative policies*, then use functional approximations to predict the values of other contracts, to reduce the number of model points for a portfolio. In this article we are interested in improving the efficiency of the simulation for each model policy so this work can be combined with the representative policy methods.

The literature on nested simulations suggests two main strands of thought. The first is to use a pilot exercise to generate an empirical *metamodel* to replace the inner simulations in the subsequent full Monte Carlo simulation. The metamodel might be based on interpolation (Hardy 2003) or stochastic kriging (a more sophisticated interpolation method; Liu and Staum 2010), or using least-squares regression (e.g., Broadie, Du, and Moallemi 2015, Cathcart and Morrison 2009), or on a generic partial differential equation approach (Feng 2014). Each of these may be referred to as a *proxy model*, as the pilot exercise is used to develop an empirical proxy model, which is subsequently used to replace the full inner simulation distribution.

The second strand in the nested simulation literature focuses on the allocation of a *computational budget* between the outer and inner simulations. Fixing the computational budget means fixing the number of individual simulated asset or liability values, where each simulated value from an inner path is assumed to have essentially the same cost (in terms of computing time) as each simulated value from an outer path. Gordy and Juneja (2010) demonstrated that outer simulations are more important than inner simulations in an accurate estimation of tail risk measures so that at some point the advantage gained from additional inner simulations is minimal. They then proposed a method of strategic allocation of the budget between inner and outer simulations. They used a uniform approach to inner simulations—that is, all inner simulations use the same number of paths. Subsequently, Broadie et al. (2011) developed a dynamic allocation algorithm that was not uniform, in which more inner simulations were applied to some outer paths than others. *Screening* refers to the more extreme form of allocation, under which no

inner simulations are applied at all to some outer paths, based on the probability that these outer loops would not contribute to the risk measure of interest—typically VaR, CTE, or probability of a shortfall based on some specified threshold. Liu and Staum (2010) used an iterative approach in three stages involving stochastic kriging and screening. Lan, Nelson, and Staum (2010) also use screening, based on a pilot study.

In this work we use ideas from some of the articles cited above. Our problem is more complex than most of the previous studies, including Broadie et al. (2011) and Lan et al. (2010), as these articles assume a single-step outer simulation, whereas we require multiple time-step outer simulations, which means that the nature and importance of the inner simulations can change over time for each outer path. Although proxy methods have many advantages, and have some traction in practice, we are interested in moving beyond the empirical proxy approach. One reason is that the pilot exercise adds to the computational budget, and there are risks in methods that rely on pilot simulations. For example, how often must the pilot be rerun? How do we know if or when the proxy has shifted too far from the underlying model? In addition, in our path-dependent setting, running pilot exercises at all durations will be time-consuming.

Our proposed method, the Importance-Allocated Nested Simulation (IANS) approach, uses two stages for the inner simulations. The first stage uses a low-cost analytic (not empirical) proxy to identify the tail scenarios most likely to contribute to the CTE risk measure. In the second stage we then allocate the entire inner simulation computational budget to the scenarios identified in the first stage. Our application is specifically the estimation of the CTE for cash flows associated with embedded VA options, but the methodology should be applicable to a wider range of problems. In particular, there is a high flexibility in the choice of proxy models, as they are used only to identify tail scenarios; it is not necessary for the proxy to accurately measure the costs arising in these scenarios, as that will be determined using the inner simulation. Different tail risk measures, such as VaR and probability of loss above threshold, can also be accommodated. Our numerical experiments show that, in the VA context, IANS can be up to 30 times more efficient than a standard Monte Carlo experiment, measured by relative mean squared errors (RMSEs), when both are given the same computational budget.

The remainder of this article is organized as follows. Section 2 discusses dynamic hedging for common types of VA riders and describes the process of a standard nested simulation. Section 3 presents our new approach, the IANS method. Section 4 illustrates the performance of the IANS method in numerical experiments. Section 5 concludes the article.

#### 2. MODELING VARIABLE ANNUITY COSTS USING NESTED SIMULATIONS

In this section we introduce our notation and assumptions. We present the common types of variable annuities riders that we use in the rest of the article. We also describe the standard two-level nested simulation, with path-dependent outer simulations, which is the base method used as a benchmark for our new approach. For more information on VA contracts and different types of guarantees, see Hardy (2003).

#### 2.1. Dynamic Hedging for VA via Nested Simulation

In a dynamic hedging program, a hedging portfolio is set up for a block of VA contracts using stocks, bonds, futures, and possibly options. The hedging portfolio is rebalanced periodically, responding to changes in market conditions and in the demographics of the block of contracts. In this article we consider a delta hedge for a single VA contract. For transparency, we ignore management fees and all other charges and expenses. We are concerned only with the costs of delta hedging the embedded option. In our illustrations we also consider only guaranteed maturity benefits, and we ignore mortality. All of these assumptions can easily be relaxed, but because the main contribution to the costs of most VA guarantees is the cost of hedging the maturity guarantee, that is our focus here, and eliminating the other factors helps focus on the primary issue.

We assume that the option matures at *T*; for a GMMB this would be the expiry date of the policy. At any  $t \le T$ , let S(t) be the underlying stock price at time *t* in an outer scenario. We assume that the delta hedge for the embedded option is composed of  $\Delta(t)$  units in the underlying stock and a sum B(t) in a risk-free zero-coupon bond maturing at *T*. The delta hedge portfolio at t - 1 is then

$$H(t-1) = \Delta(t-1)S(t-1) + B(t-1).$$

Given a risk-free force of interest of r per time unit, at the end of the th time period, the value of this hedge has changed to

$$H^{BF}(t) = \Delta(t-1)S(t) + B(t-1)e^{r},$$
(1)

and this is the hedge brought forward at time t (we assume no rebalancing between times t - 1 and t). The cash flow incurred by the insurer, which we call the *hedging error*, is the difference between the cost of the hedge at time t and the value of the hedge brought forward:

O. DANG ET AL.

$$HE(t) = H(t) - H^{BF}(t).$$
(2)

The costs to set up the initial hedging portfolio, the periodic hedging losses because of rebalancing, and the final unwinding of the hedge are recognized as part of the profit and loss of the VA contract. The present value of these cash flows, discounted at the risk-free rate of interest, constitutes the liability of the VA to the insurer; this is the loss random variable to which we apply a suitable risk measure.

For a GMMB, without a ratchet option, the liability can be decomposed as follows. Let F(t) denote the value of the policyholder's funds at *t*. The funds increase in proportion to a stock index with value S(t) at *t* (as we are ignoring fees and expenses), so for convenience we can scale the stock price index and assume that F(0) = S(0). The guaranteed minimum benefit is assumed to be a fixed value, *G*, say. Let  $H(0) = \Delta(0)S(0) + B(0)$  denote the cost of the initial hedge, and let H(T) denote the ultimate guarantee payoff at *T*, that is,  $H(T) = (G - S(T))^+$ .

$$L = H(0) + \sum_{t=1}^{T} e^{-rt} HE(t)$$

$$= H(0) + \sum_{t=1}^{T} e^{-rt} [H(t) - H^{BF}(t)]$$

$$= B(0) + \Delta(0)S(0) + \sum_{t=1}^{T-1} e^{-rt} \Big( B(t) + S(t)\Delta(t) - B(t-1)e^{t} - S(t)\Delta(t-1) \Big) + H(T) - H^{BF}(T)$$

$$= S(0)\Delta(0) + \sum_{t=1}^{T-1} e^{-rt} S(t) \Big( \Delta(t) - \Delta(t-1) \Big) + e^{-rT} \Big( \Big( G - S(T) \Big)^{+} - S(T)\Delta(T-1) \Big).$$
(3)
(3)
(4)

Equation (4) shows that the liability from dynamic hedging can be decomposed into the sum of three components:

- The present value of the payoff at maturity (t = T) minus the stock part of the hedge brought forward.
- The cost of the stock part of the initial hedge portfolio.
- The present value of the changes in the value of the stock holding because of periodic rebalancing.

In the bond holdings of the hedging portfolio, the interest rate at which the bond value accumulates is the same as the rate at which cash flows from bond trades are discounted. Therefore all the interim bond trades can be reduced to one trade, which is to set up a bond holding at the time of valuation so that it can accumulate to the bond holding required at maturity—that is,  $e^{-rT}G I(G > S(T))$ , where I(E) is an indicator random variable for event *E*. In the stock holdings of the hedging portfolio, the liability arises from the initial setup of the stock future holding, as well as the cash flows from every stock future trade. Computationally, Equation (4) is more efficient than (3) because the interim hedging portfolio values H(t) for t = 0, ..., T - 1 are not required.

For guaranteed minimum accumulation benefits (GMAB), there are fixed renewal or ratchet points in the policy term. We consider a VA contract that matures at  $T_2$  and that has a single renewal date at time  $T_1 < T_2$ . Assume an initial guarantee of  $G_1$ , which will be applied to the fund at  $T_1$ , and which is fixed at the inception of the policy. At time  $T_1$ , if the fund is less than the initial guarantee, then the insurer deposits the difference into the policyholder's fund, and the policy continues with the same guarantee applying at the final maturity  $T_2$ . If the fund value is greater than the guarantee at  $T_1$ , then there is no payment, but the guarantee applying at  $T_2$  is increased from  $G_1$  up to the fund value at  $T_1$ . If we assume, as above, that F(0) = S(0), then the cash flows and fund values are as follows:

For 
$$t < T_1$$
:  $F(t) = S(t)$   
For  $t > T_1$ :  $F(t) = S(t) \max\left(1, \frac{G_1}{S(T_1)}\right)$   
Insurer payout at  $T_1$ :  $\left(G_1 - S(T_1)\right)^+$   
Insurer payout at  $T_2$ :  $\left(\max(G_1, S(T_1)) - S(T_2) \max\left(1, \frac{G_1}{S(T_1)}\right)\right)^+$ 

At time  $T_1^+$ , immediately after any payment at the time, the insurer must fund a new hedge portfolio for the option maturing at time  $T_2$ , which we denote  $H(T_1^+)$ . Then the liability can be written as follows, where  $\Delta(T_1)$  is understood to mean the delta of

 $H(T_1^+)$ :

$$L = \Delta(0)S(0) + \sum_{t=1}^{T_2 - 1} e^{-rt} S(t) \left( \Delta(t) - \Delta(t-1) \right) + e^{-rT_1} \left( G_1 - S(T_1) \right)^+ + e^{-rT_2} \left( \left( \max(G_1, S(T_1)) - S(T_2) \max\left(1, \frac{G_1}{S(T_1)}\right) \right)^+ - S(T_2) \Delta(T_2 - 1) \right).$$
(5)

Using Monte Carlo simulation to assess the costs of hedging, we generate, say, J simulated liability values and sort them such that  $L_{(j)}$  is the *j*th smallest value. We then estimate the 100 $\alpha$ %-CTE as

$$CTE_{\alpha} = \frac{1}{J(1-\alpha)} \sum_{j=J\alpha+1}^{J} L_{(j)}.$$
 (6)

A single-tier simulation is required to generate real-world simulated paths for the stock prices, S(t). The second level of simulation may be required to determine the  $\Delta(t)$  values. In very simple models, this may not be necessary, but it will usually be required when we use a stochastic volatility model or when we incorporate dynamic lapses or other complications.

Algorithm 1 outlines the steps of a standard two-level nested simulation for estimating  $CTE_{\alpha}$  of losses for a Delta-hedged VA contract, using Equation (3) for a generic process, allowing for path dependent guarantees and for fund values that differ from the underlying stock prices—for example, because of fees or because of renewals or other policyholder options.

Algorithm 1: Two-level nested simulation of losses for a Delta-hedged VA contract with a single payout date T.

input: r: Per-period risk-free interest rate.  $T_k$ : Potential renewal and maturity dates of the VA contract.  $T = \max(T_k)$ . G(0): Guarantee value at the time of valuation. F(0): Subaccount value at the time of valuation. S(0): Underlying stock price at the time of valuation.  $H(0), \Delta(0), B(0)$ : Hedging portfolio value, delta, and bond value.  $\alpha$ : Confidence level (e.g.,  $\alpha = 95\%$ ) at which the CTE of losses is required. N, J: Number of replications in inner and outer simulations, respectively. **output**:  $CTE_{\alpha}$  for the losses of Delta-hedging the VA contract of interest. 1 for j = 1, ..., J do 2 Initialization for outer simulations: 3 Set  $S_i(0) \leftarrow S(0), F_i(0) \leftarrow F(0), G_i(0) \leftarrow G(0)$ . 4 for t = 1, ..., T do 5 Simulate  $S_i(t)$  using the real-world asset model. 6 Calculate subaccount value  $F_i(t)$  and guarantee values  $G_i(t)$ . 7 Calculate guarantee payouts at renewal and maturity dates, e.g.,  $(G_i(T) - F_i(T))^+$ . 8 for i = 1, ..., N do 9 Initialization for inner simulations: 10 Set  $S_{ij}(t) \leftarrow S_j(t), F_{ij}(t) \leftarrow F_j(t), G_{ij}(t) \leftarrow G_j(t)$ . Perform the *i*th inner simulation under a risk-neutral measure. 11 1. Simulate  $S_{ii}(T)$  and  $F_{ii}(T)$  given  $S_{ii}(t)$ . 2. Determine the guarantee payoff at T given  $F_{ii}(T)$ , that is  $(G_{ii}(T) - F_{ii}(T))^+$ . 3. The *i*th inner simulated value of the hedging portfolio at time *t* is  $H_{ij}(t) = e^{-r(T-t)}(G_{ii}(T) - F_{ii}(T))^+$ . 4. The *i*th inner simulated delta of the hedge portfolio is  $\Delta_{ii}(t)$ , which measures the sensitivity of  $H_{ii}(t)$ with respect to  $S_{ii}(t)$ . 12 end

```
191
```

#### 13 Estimate the hedging portfolio value, the delta, and the hedge bond value at time t.

$$H_j(t) = \frac{1}{N} \sum_{i=1}^N H_{ij}(t), \quad \Delta_j(t) = \frac{1}{N} \sum_{i=1}^N \Delta_{ij}(t), \quad B_j(t) = H_j(t) - \Delta_j(t) S_j(t).$$

Calculate  $H_{j(t)}^{BF} = B_j(t)e^r + \Delta_j(t-1)S_j(t)$ , the brought-forward portfolio value. Calculate  $HE_j(t) = H_{j(t)}^{BF} - H_j(t)$ , the hedging loss at time *t*. 15

16

17 end

Calculate the loss random variable for the *j*th outer path,  $L_i$ , following Equations (3), (4), or (5) as appropriate. 18

19 end

20 Sort the simulated liabilities, such that  $L_{(j)}$  is the *j*th smallest value and estimate the CTE using Equation (6).

Algorithm 1 can easily be extended to hedging strategies that depend on other sensitivities (e.g., Gamma, Rho, Theta). In these cases, the inner simulation model would be extended to estimate the relevant Greeks, resulting in hedging portfolios that may consist of additional assets such as options, forwards, and VIX. See L'Ecuyer (1990), Glasserman (2013), and Fu (2016) for more information on estimating greeks using a Monte Carlo simulation. We use the Infinite Perturbation Analysis (Broadie and Glasserman, 1996; Glasserman, 2013) method for sensitivity estimation in our numerical studies.

The different purposes of the outer and inner simulations result in different stochastic asset models being applied. A realworld model is used in outer simulations (line 5) to examine losses associated with the VA contract under realistic scenarios. A risk-neutral model is used in inner simulations (line 11) only for evaluating the hedge costs for each time step for each scenario.

The evolution of subaccount and guarantee values in Line 6 and the inner simulation model in line 11 of Algorithm 1 can be adapted to a range of VA guarantees and assumptions. In some cases, the hedge portfolio at each time point can be determined analytically, as we demonstrate in the following section.

#### 2.2. Analytic Hedge Calculations Using Black-Scholes

In the case where the risk-neutral measure is assumed to be Geometric Brownian Motion, and where the guarantee is a GMMB with fixed guarantee, or GMAB with fixed initial guarantee, then the hedge portfolio can be determined analytically, without requiring the inner simulation step.

Consider first the GMMB, with a fixed guarantee G. We ignore mortality, fees, and expenses and assume for convenience that F(0) = S(0). The maturity payoff is a simple European put option, so the hedge at t under the jth outer simulation,  $H_i(t)$ , can be determined from the Black-Scholes formula for a put option, where r is the risk-free rate of interest continuously compounded, per time unit, and  $\sigma$  is the volatility of the risk-neutral GBM, expressed per time unit:

$$H_j(t) = G(t)e^{-r(T-t)}\Phi(-d_2) - S_j(t)\Phi(-d_1), \quad \Delta_j(t) = -\Phi(-d_1),$$
(7)

where  $\Phi(x)$  is the cumulative function of the standard Normal random variable and

$$d_1(t,T) = \frac{\ln\left(\frac{S_j(t)}{G(t)}\right) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \quad d_2(t,T) = d_1(t,T) - \sigma\sqrt{T - t}.$$
(8)

The GMAB payoff(s) have the same structure as a tandem put option in finance. We present formulas for the hedge for a European tandem option, with F(0) = S(0), with a single renewal point  $T_1$  and final maturity  $T_2$ . The initial guarantee is  $G_1$ ; the second guarantee is  $G_2 = \max(G_1, S(T_1))$ . For details of the derivation, see Hardy (2003, Appendix B).

For  $t < T_1$ , we have  $F_j(t) = S_j(t)$ , and the value and the Delta of the corresponding European tandem option, given  $S_j(t)$ are given by the following equations. First, we define  $p^*$  to denote the value at  $T_1$  of an at-the-money option, with unit strike and stock price, expiring at  $T_2$ . That is,

$$p^* = e^{-r(T_2 - T_1)} \Phi(-d_2^*) - \Phi(-d_1^*), \tag{9}$$

where

$$d_1^* = \frac{(r + \sigma^2/2)(T_2 - T_1)}{\sigma\sqrt{T_2 - T_1}} \qquad d_2^* = d_1^* - \sigma\sqrt{T_2 - T_1}$$
(10)

This is a useful function because at time  $T_1$ , after any payment or adjustment of guarantee made at that time, the remaining guarantee is an at-the-money option, with term  $T_2 - T_1$  years, with value at  $T_1$  of

$$H_j(T_1^+) = \max(S(T_1), G_1)p^*.$$

Note that  $p^*$  is a constant with respect to the underlying stock price process. For the full GMAB guarantee hedge at  $t < T_1$  we have

$$H_{j}(t) = \left(G_{1}(t)e^{-r(T_{1}-t)}\Phi(-d_{2}(t,T_{1})) - S_{j}(t)[\Phi(-d_{1}(t,T_{1})) - 1]\right) \ (1+p^{*}) - S_{j}(t), \tag{11}$$

$$\Delta_j(t) = -\Big(\Phi(-d_1(t,T_1)) - 1\Big)(1+p^*) - 1, \tag{12}$$

$$= p^* \Phi(d_1(t, T_1)) - \Phi(-d_1(t, T_1)), \tag{13}$$

where  $d_1(t, T_1)$  and  $d_2(t, T_1)$  are as defined in Equation (8).

If  $t > T_1$  then the hedge portfolio is identical to the GMMB case, with guarantee  $G_2 = \max(G_1, S(T_1))$  and with maturity  $T_2$ . The fund value at  $t > T_1$  is  $F_j(t) = S_j(t)(\max(1, \frac{G_1}{S_j(T_1)}))$ , and this replaces  $S_j(t)$  in Equations (7) and (8).

In practice, the analytic expressions from the Black-Scholes model may not be sufficiently accurate for tail risk measures of the hedge costs. Introducing a stochastic volatility model for the hedge costs can make analytic evaluation unwieldy or impossible, and when dynamic lapse assumptions are incorporated, the only feasible approach is the Monte Carlo simulation. Nonetheless the analytic Black-Scholes hedge costs are expected to be correlated with the true values, so we will use the analytic expressions as our first stage analysis to screen out the outer scenarios that are very unlikely to contribute to the CTE and run the inner simulation part of the nested simulation algorithm only for those scenarios deemed sufficiently important after the first screening. The two-stage process is described more fully in the following section.

#### 3. IANS METHOD

In this section, we present our IANS method for estimating the  $CTE_{\alpha}$  of a VA GMMB or GMAB, using a nested simulation with screening approach. We provide an outline in Section 3.1 and supply the details in subsequent sections.

### 3.1. Outline of the IANS Method

The IANS method replaces the inner simulation steps in Algorithm 1 with a two-stage process. As before, T denotes the final expiration date of the guarantee.

Algorithm 2: Importance-Allocated Nested Simulation of losses for a Delta-hedged VA contract with a single payout date T.

input: - Underlying real-world and risk-neutral asset models with parameters.

- VA contract, term *T*, and fully specified dynamic hedging program.
- The risk measure and level, e.g.,  $CTE_{\alpha}$ .

**output**:  $CTE_{\alpha}$  for the losses of Delta-hedging the VA contract of interest.

**Initialization:** Simulate J outer scenarios, each is a T-period simulated stock price sample path under the real-world measure.

#### Stage I: Identification of proxy tail scenarios

- (I.1) Select a proxy *financial derivative* and associated *asset model* that provide tractable, analytic hedge costs and for which the payoff is expected to be well correlated to the VA guarantee costs. See Section 3.2.
- (I.2) Calibrate the proxy asset model to the underlying risk-neutral asset model in inner-level simulations. See Section 3.3.
- (I.3) Implement Algorithm 1 but with the analytic hedge calculations for the proxy derivative and asset model replacing the inner simulation step.

193

#### O. DANG ET AL.

(I.4) Identify  $(1 - \xi)J$  proxy tail scenarios with the largest simulated loss in step (I.3) for some  $\xi \in [0, \alpha]$ . See Section 3.4. Stage II: Nested simulation with concentrated computation budget

- (II.1) Allocate remaining computational budget to the  $(1 \xi)J$  proxy tail scenarios.
- (II.2) Implement the inner simulation step of Algorithm 1 with the original risk-neutral asset model and VA payoff, but only for the  $(1 \xi)J$  outer scenarios identified in step (I.4).
- (II.3) Identify the  $(1 \alpha)J$  largest liability values based on the inner simulations.
- (II.4) Compute  $CTE_{\alpha}$  in Equation (6) as the output.

The user must specify some parameters and experiment design choices that govern the behavior of the IANS method. In many actuarial and financial applications the inner-level simulation dominates the computational cost because of the need for repetition of the full set at each time point. Appropriate selection of the proxy asset model and financial derivative in step (I.1) greatly improves the simulation efficiency for the IANS method compared with a full nested simulation. Hence, in practice, the most important choices for this algorithm are the proxy asset model and financial derivative used in stage I. Both the proxy asset model and proxy financial derivative could be different from the underlying asset model and the VA contract of interest, which will be referred to as the true models hereinafter for convenience. Although using proxy model to replace the inner-level simulation is an inspiration to our work, the IANS method employs the proxies in a distinctive way. In particular, IANS does not directly use the estimated losses from the proxies to estimate CTE, but instead uses them for identifying the outer scenarios that will be relevant to the tail risk measure. These are referred to as the proxy tail scenarios. Guided by such identification, the full inner simulation paths are then run for the proxy tail scenarios. Ideally, the selection of proxies should result in fast computation of the proxy tail scenarios and should be sufficiently correlated with the true costs, at least in terms of ranking of the costs associated with the outer scenarios, so that the proxy tail scenarios accurately capture the true tail scenarios that would have been identified with a full inner simulation process. The proxy calculations take negligible computation effort; if the proxy tail scenarios contain the true tail scenarios, then the IANS method considers all the relevant outer scenarios without compromising the granularity in the inner simulations.

#### 3.2. Selection of Proxies

Unlike a standard proxy approach, the proxy tail scenarios do not need to accurately assess the liability values for those scenarios; we use the proxy step to ascertain a ranking of the liabilities by outer scenarios. This means that the IANS method is expected to perform well as long as the *rankings of losses* between the proxies and original models are highly correlated, even if the *losses* themselves are not.

In most VA portfolios, the key benefits contributing to the risk are the living benefits. GMMB and GMAB are among the simpler forms of living benefits, and we consider more complex living benefits such as GMIB and GLWB in future work. For GMMB and GMAB, the put option and tandem put option derivatives identified in Section 2.2 are obvious proxy derivatives, as the option payoffs are identical to the guarantee payoffs, if we ignore complications of policyholder behavior. Using a Black-Scholes (risk-neutral Geometric Brownian Motion [GBM]) model as the proxy asset model allows us to use the analytic option formulas from Section 2.2 in proxy calculations at negligible computational cost. To ensure that the proxy liability values are as highly correlated as possible with the true values under the inner simulation asset model, we dynamically calibrate the Black-Scholes volatility in the proxy model to the conditional expected volatility of the real-world model, based on the scenario path up to the valuation. We explain this further in the next section.

#### 3.3. Calibration of Proxy Asset Model

GBM is inconsistent with important features of observed stock returns, including extreme left-tail events, volatility clustering, and association of high volatility and low returns. Therefore, it is not a good model to use (in real-world or equivalent risk-neutral form) in any practical long-term application in which tail risk is the main consideration. To capture the correlations and fatter tails of real-world stock price growth, more advanced asset models such as regime-switching lognormal (RSLN) (Hardy 2001) and generalized autoregressive conditional heteroskedasticity (GARCH) models (Bollerslev 1986; Duan 1995) are often used. We consider both of these asset models as the true asset models in our numerical studies, both for the outer real-world scenarios and in an equivalent risk-neutral form, for the inner scenarios. See the Appendix for details.

For the proxy model volatility at time t, say, we set the GBM volatility equal to the expected volatility based on the full inner simulation model, which may be dependent on the outer simulation path up to time t for stochastic volatility models. In particular,

Notation	Value
Т	240 months
$T_1$	End of the 120th month
F(0)	\$1,000
$G_1$	100% of <i>F</i> (0)
	Notation T $T_1$ F(0) $G_1$

 TABLE 1

 Parameters for Variable Annuity Contracts

*Note:* GMAB = guaranteed minimum accumulation benefits.

- For VA contracts that are not renewable, or where the time of valuation *t* is beyond the last renewal, we set the Black-Scholes volatility between *t* and maturity *T* to the expected volatility of the true asset model in the same period, conditioning on the state variables at time *t*.
- For VA contracts with a valuation date prior to the renewal date  $t < T_1$ , we calibrate two volatilities: (1) the expected volatility of the true asset model between *t*, and *T*<sub>1</sub>, conditioning on the state variables at time *t*, and (2) the expected volatility of the true asset model between *T*<sub>1</sub> and *T*<sub>2</sub>, conditioning on the state variables at time *t*. These two volatilities can be different and are used to calculate *d*'s in Equation (8) and *d*\*s in Equation (10).

Detailed descriptions and the corresponding volatility calibrations of an RSLN model with two regimes and a GARCH(1,1) model are given in the Appendix.

#### 3.4. Safety Margin for Tail Scenario Identifications

The proxies selected in step (I.2) cannot perfectly capture the complexities of the original asset model and VA contract of interest, resulting in potential misclassification of tail scenarios. Therefore we select a proxy confidence level  $\xi$  in step (I.4) with some safety margin, so that  $\alpha - \xi \ge 0$ . This means that the proxy tail scenarios are the  $(1 - \xi)J$  outer scenarios with the largest simulated liabilities based on the proxy calculations. We use these to identify the largest  $(1 - \alpha)J$  simulated liabilities based on the inner simulations, assuming that, with high confidence, the  $(1 - \alpha)J$  true tail scenarios are a subset of the  $(1 - \xi)J$  proxy tail scenarios.

This proxy confidence level  $\xi$  is an experiment design parameter in IANS. If  $\xi$  is very small, the likelihood of capturing the true tail scenarios is high, but at the cost of running the inner simulations on a large number of outer scenarios. With a fixed budget for the inner simulations, this will generate higher mean square errors in the liability values and CTE estimates. On the other hand, if  $\xi$  is close to  $\alpha$ , the inner simulation budget is focused on fewer scenarios, so those that are included will have more accurate liability valuations, but some tail scenarios will be wrongly omitted because the proxy liability ranking is not comonotonic with the true liability ranking. Hence there is a trade-off between a high likelihood of including the true tail scenarios ( $\xi \rightarrow 0$ ) and high concentration of simulation budget in stage II ( $\xi \rightarrow \alpha$ ). In this work an arbitrary safety margin of 5% is included in the numerical examples, that is,  $\xi = \alpha - 5\%$ . In future work we are considering a more structured approach to the safety margin. Optimization of this trade-off is a subject of future study.

#### 4. NUMERICAL EXPERIMENTS

To illustrate the performance of the IANS method, we use it to estimate CTEs at different confidence levels for GMMB and GMAB liabilities, using different true asset models, and under different lapse assumptions. A few simplifying assumptions are made, consistently with the development of the previous sections, specifically

- No transaction costs are in the hedging program.
- The initial premium is invested in a stock index, with no transfers between funds.
- There are no subsequent premiums.
- We ignore mortality and other decrements unless otherwise stated.
- No management or guarantee rider fees are deducted from the fund.
- The risk is delta hedged at monthly intervals.

Monthly Rate	Real-World	Risk-Neutral
	0.000	0.000
Risk-free rate: r	0.002	0.002
Mean - Regime 1 ( $\rho = 1$ ): $\mu_1$	0.0085	0.0013875
Mean - Regime 2 ( $\rho = 2$ ): $\mu_2$	-0.0200	-0.0012000
Std Dev Regime 1: $\sigma_1$	0.035	0.035
Std Dev Regime 2: $\sigma_2$	0.080	0.080
Transition probability from Regime 1: $p_{12}$	0.04	0.04
Transition probability from Regime 2: $p_{21}$	0.20	0.20

 TABLE 2

 Parameters for the Regime-Switching Model Used in Section 4

Under these assumptions, the liabilities of the VA contracts consist only of the liability from the hedging program, that is, the initial cost of the hedging portfolio and the present value of periodic hedging errors. In practice, fee income, expenses, commissions, decrements, and costs because of basis risk (difference between real-world and hedging models) are likely to make up a proportion of the liability. Nonetheless, the IANS method still offers useful insights because the liability from hedging is the only quantity we need from the inner simulation part of a nested simulation.

We consider two risk measures,  $CTE_{80\%}$  and  $CTE_{95\%}$ , as these are commonly used in valuation and economic capital setting in Canada, consistent with regulatory standards. As discussed above, we consider a GMMB and GMAB in the numerical experiments, using the parameters specified in Table 1. Two popular stochastic asset price models, RSLN with two regimes and GARCH(1,1), are considered in our experiments. The model parameters are provided in Table 2 and Table 3.

The financial market is incomplete in the regime-switching model, thus its risk-neutral measure is not unique (Hardy 2001). Given the real-world measure in the regime-switching model, we employ the risk-neutral model studied in Bollen (1998) and Hardy (2001), whose mean conditional log return is  $r - \sigma_i^2/2$  for i = 1, 2. All other parameters are the same in the real-world and risk-neutral models.

The parameters for the real-world GARCH(1,1) model used in our numerical studies are summarized in Table 2. The value of  $\alpha_0$  is chosen such that long-run average volatility in this GARCH(1,1) models is equal to the long-run average volatility in the RSLN model with parameters in Table 2. The risk-neutral version of the model is attained by Esscher transform (Ng and Li 2013).

We assess the IANS method by assuming a fixed computational budget for simulation, and we compare the accuracy of the resulting CTE estimates with estimators produced with the same computational budget for simulation, using the standard nested Monte Carlo (SMC) simulation in Algorithm 1. Section 4.2 presents numerical experiments under static lapse, and experiments for dynamic lapse, are presented in Section 4.3.

#### 4.1. Benchmarking Large-Scale Nested Simulations

To assess the relative mean squared errors (RMSEs) of different estimators, we first conduct a large-scale nested simulation, with 10,000 inner-level simulations and 10,000 outer-level simulations, to obtain accurate estimates for the CTEs of interest. We say that this large-scale nested simulation takes a computational budget of  $10,000 \times 10,000 \times (1 + 12 \times 20) \times (12 \times 20) \div 2 = 2.892 \times 10^{12}$ . Hereinafter these estimates are referred to as the *true means* for different CTE estimators. To illustrate the first stage of the IANS method, we replace the inner simulations with closed-form formulas based on the put option (GMMB) and tandem put option (GMAB) proxy derivatives, with the Black-Scholes asset model and examine how many true tail scenarios are correctly identified by the proxies.

Figure 1 depicts the comparisons between the losses that are simulated by the true nested simulation and those by the IANS method's proxy simulation. We can see graphically that the values of the simulated losses produced by these two methods are highly correlated. This indicates that stage I in the IANS methods is able to correctly identify most true tail scenarios without any inner simulation. Table 4 summarizes this observation quantitatively. We see that the closed-form proxy calculation along with the safety margin in stage I of the IANS method identifies the true tail scenarios in the nested simulation very accurately, for different CTE levels, different true asset models, and different VA types. Such robust and accurate identification of tail scenarios leads to the high performance of the IANS method, as showcased in subsequent experiments.

Monthly Rate	Real-World	Risk-Neutral
r	0.002	0.002
μ	0.00375	n/a
$\alpha_0$	0.0002094225	0.0002094225
$\alpha_1$	0.1	0.1
β	0.8	0.8
$\sigma(0)$	0.0457627	n/a
$\epsilon(0)$	0	n/a

 TABLE 3

 Parameters GARCH(1,1) Model Used in Section 4

*Note:* GARCH = generalized autoregressive conditional heteroskedasticity.



Misclas	sification
<ul> <li>Worst 5%: by Sim, not by Proxy</li> <li>Worst 5%: by Proxy, not by Sim</li> </ul>	<ul><li>Worst 20%: by Sim, not by Proxy</li><li>Worst 20%: by Proxy, not by Sim</li></ul>

FIGURE 1. Simulated Losses in 10,000 Outer Scenarios. *Note:* The x and y coordinates of each point in the figures represent the loss in a scenario, simulated by the IANS proxy simulation and by the true nested simulation, respectively. GMMB = guaranteed minimum maturity benefits; GMAB = guaranteed minimum accumulation benefits.

TABLE 4 Tail Scenario Identification by the Proxy Simulation (with Safety Margin) in Stage I of the IANS Method Under Static Lapse Assumption

CTE Level	#No. Tail Scen.	#No. (%) of Correctly Identified Tail Scen.				
		RSLN		GARCH(1,1)		
		GMMB	GMAB	GMMB	GMAB	
80%	2,000	2,000 (100.00%)	1,998 (99,90%)	1,992 (99,60%)	1,985 (99.25%)	
95%	500	500 (100.00%)	500 (100.00%)	500 (100.00%)	500 (100.00%)	

*Note:* IANS = Importance-Allocated Nested Simulation; Scen. = scenario; RSLN = regime switching lognormal; GARCH = generalized autoregressive conditional heteroskedasticity; CTE = Conditional Tail Expectation; GMMB = guaranteed minimum maturity benefits; GMAB = guaranteed minimum accumulation benefits.

Simulations in Numerical Experiments					
(a) J	(b) N	$(c) \\ (1-\xi)J$	Nested Simulation Computational Budget		
5,000	200		$(a) \times (b) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$		
1,000	1,000	_	$(a) \times (b) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$		
200	5,000		$(a) \times (b) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$		
5,000	800	1,250	$(b) \times (c) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$		
5,000	2,000	500	$(b) \times (c) \times \frac{(1+12 \times 20) \times (12 \times 20)}{2} = 2.892 \times 10^{10}$		
	( <i>a</i> ) <i>J</i> 5,000 1,000 200 5,000 5,000	(a)         (b)           J         N           5,000         200           1,000         1,000           200         5,000           5,000         800           5,000         2,000	(a)(b)(c)JN $(1-\xi)J$ 5,0002001,0001,0002005,0005,0008001,2505,0002,000500		

TABLE 5 Simulations in Numerical Experiments

Note. SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation.

#### 4.2. Dynamic Hedging Under Static Lapse

To demonstrate the efficiency of the IANS method, we compare it to three standard nested simulation experiments that use the same computational budget but with different allocation between inner- and outer-level simulations, as shown in Table 5. The fixed simulation budget in all cases is 1% of that in the benchmark simulation in Section 4.1. By design, the *SMC-5,000-200* experiment has a large number of outer-level projections and the *SMC-200-5,000* experiment has a larger number of inner projections. The *SMC-1,000-1,000* experiment is designed with a more balanced number of inner-and outer-level projections. For the IANS estimators, we set J = 5, 000. Column (c) in Table 5 indicates the number of inner projections required using IANS with a margin of  $\alpha - \xi = 5\%$ . Each of the experiment designs is repeated independently 100 times to produce 100 estimates of CTEs at both 80% and 95% levels, for both the GMMB and GMAB contracts.

Figures 2 and 3 depict the CTE estimates in different experiment designs for the GMMB where the true asset model is RSLN. The solid red line in each graph indicates the true value estimated by the large-scale simulation discussed in Section 4.1. Comparing Figure 2a with Figures 2b and 2c, and Figure 3a with Figures 3b and 3c, we see that a sufficient number of outer-level simulation reduces the variance, whereas inner-level simulation appears to reduce the bias in estimating tail risk measures. These results are consistent with, for example, Broadie et al. (2011) and Gordy and Juneja (2010). A sufficient number of outer-level simulation reduces the variation in extreme losses simulated from one experiment to another, which reduces the variance. On the other hand, a sufficient number of inner-level simulation ensures that a more consistent distribution of number of contracts renews or matures in-the-money, which reduces the bias. Figures 2 and 3 show that the IANS method achieves both low bias and low variance compared with the three standard nested simulation experiments. Figures 4 and 5 illustrate the results from the same experiments for the GMAB contract under the GARCH(1,1) asset model.

Table 6 summarizes the RMSEs for different experiment designs. Each RMSE is calculated as  $\frac{1}{100}\sum_{i=1}^{100} \frac{(\hat{\mu}_i^{est} - \mu)^2}{\mu}$ , where  $\hat{\mu}_i^{est}$  is the estimated CTE in the *i*th independent repeated experiment and  $\mu$  is the corresponding true CTE value estimated by the large-scale nested simulation discussed in Section 4.1. The RMSEs are then decomposed into relative bias in Table 7 and rela-

tive variance in Table 8 for different experiment designs. Each relative bias is calculated as  $\frac{\left(\frac{1}{100}\sum_{i=1}^{100}\hat{\mu}_{i}^{est}-\mu\right)}{\mu}$ , whereas each relative bias is calculated as  $\frac{\left(\frac{1}{100}\sum_{i=1}^{100}\hat{\mu}_{i}^{est}-\mu\right)}{\mu}$ , whereas each relative bias is calculated as  $\frac{\left(\frac{1}{100}\sum_{i=1}^{100}\hat{\mu}_{i}^{est}-\mu\right)}{\mu}$ , whereas each relative bias is calculated as  $\frac{\left(\frac{1}{100}\sum_{i=1}^{100}\hat{\mu}_{i}^{est}-\mu\right)}{\mu}$ , whereas each relative bias is calculated as  $\frac{\left(\frac{1}{100}\sum_{i=1}^{100}\hat{\mu}_{i}^{est}-\mu\right)}{\mu}$ .

tive variance is calculated as  $\frac{1}{100} \sum_{i=1}^{100} \frac{(\hat{\mu}_i^{est} - \frac{1}{100} \sum_{i=1}^{100} \hat{\mu}_i^{est})^2}{\mu}$ .

Table 6 demonstrates that, for these examples, the IANS method achieves smaller RMSEs compared with straightforward nested simulation, for the same simulation budget. For both 80% and 95% confidence levels, we see from Table 6 that the SMC-5,000-200 experiments have significantly smaller RMSEs than those of the other corresponding SMC experiments. Table 8 further shows that the smaller RMSEs in the SMC-5,000-200 experiments are mostly attributed to the smaller relative variance. This indicates the importance of sufficient outer-level simulation relative to inner-level simulations; this observation is consistent with other studies in nested simulations (Gordy and Juneja 2010; Broadie et al., 2011). The RMSEs indicate that the mean squared error in the IANS method experiments are within 0%–5% of the true CTE values, whereas the mean squared error in a few SMC experiments are much higher relative to the true CTE value. Compared to the SMC-1,000-1,000 and SMC-200-5,000 experiments, we observe in Table 6 and Table 8 that the reduction in RMSEs in the IANS experiments are mostly due to the reduction in relative variance. In fact, the level of reduction in this case is similar for both 80% and 95% confidence level experiments because the reduction in relative variance is driven by the increase in the number of outer scenarios



(c) SMC-200-5,000, CTE 80%

(d) IANS, CTE 80%

FIGURE 2. Estimated  $CTE_{80\%}$  of Simulated GMMB Losses under the Regime-Switching Model in 100 Independent Repeated Experiments. *Note:* The solid red line in each graph indicates the true value estimated by the large-scale simulation discussed in Section 4.1. SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation; GMMB = guaranteed minimum maturity benefits.

considered in the IANS experiments compared to the SMC-1,000-1,000 and SMC-200-5,000 experiments, which is the same at 80% and 95% confidence levels. In contrast, compared to the SMC-5,000-200 experiments, we observe in Table 6 and Table 7 that the reduction in RMSEs in the IANS experiments are mostly due to the reduction in relative bias. The IANS experiments result in smaller bias because of more inner simulations. Given the fixed computation budget, and the same number of outer-loop simulations in the IANS experiments at 80% and 95% confidence levels, more inner-loop simulations are used in the 95% confidence level experiments than in the 80% confidence level experiments. However, the reduction in relative bias and RMSEs in the IANS experiments. This is because the sensitivity of bias in the CTE estimates to the number of inner simulations varies by CTE levels and by contract types. This is also evident when comparing the relative bias among SMC experiments with different inner-loop simulations. The results for GMAB contracts in Table 6 are commensurate



FIGURE 3. Estimated  $CTE_{95\%}$  of Simulated GMMB Losses under the Regime-Switching Model in 100 Independent Repeated Experiments. *Note:* The solid red line in each graph indicates the true value estimated by the large-scale simulation discussed in Section 4.1. SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation; GMMB = guaranteed minimum maturity benefits.

with those of GMMB contracts. The GMAB contract is more complicated than GMMB by design, so the simulated losses are more volatile, resulting in higher RMSEs in general.

#### 4.3 Dynamic Hedging of VAs under Dynamic Lapse

To further demonstrate the robustness of the IANS method, we also conducted large-scale nested simulations, similar to those in Section 4.1, for both GMMB and GMAB contracts under dynamic lapse by policyholders in the liability projection. We also simulate such contracts under both the regime-switching and the GARCH models. The goal of these experiments is to test the effectiveness of the proxies in identifying the tail scenarios in more realistic settings. The IANS method is shown to be highly robust, which suggests that its practical value is significant and promising.



(c) SMC-200-5,000, CTE 80%

(d) IANS, CTE 80%

FIGURE 4. Estimated  $CTE_{80\%}$  of Simulated GMAB Losses under GARCH(1,1) Model in 100 Independent Repeated Experiments. *Note:* The solid red line in each graph indicates the true value estimated by the large-scale simulation discussed in Section 4.1. SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation; GMAB = guaranteed minimum accumulation benefits; GARCH = generalized autoregressive conditional heteroskedasticity.

The dynamic lapse behavior by policyholders are modeled as follows.

- The fund value F and guarantee value G are reduced proportionally by lapse.
- $q_{x+t}$ , the monthly lapse rate as of time t is

$$q_{x+t} = \min\left(1, \max\left(0.5, 1 - 1.25 \times \left(\frac{G(t)}{F(t)} - 1.1\right)\right)\right) \times q_{x+t}^{base}$$
(14)



(c) SMC-200-5,000, CTE 95%

(d) IANS, CTE 95%

FIGURE 5. Estimated  $CTE_{95\%}$  of Simulated GMAB Losses under GARCH(1,1) Model in 100 Independent Repeated Experiments. *Note.* The solid red line in each graph indicates the true value estimated by the large-scale simulation discussed in Section 4.1. SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation; GMAB = guaranteed minimum accumulation benefits; GARCH = generalized autoregressive conditional heteroskedasticity.

where

$$q_{x+t}^{base} = \begin{cases} 0.00417 & \text{if } t < 84, \\ 0.00833 & \text{if } t \ge 84. \end{cases}$$
(15)

This dynamic lapse multiplier applied to the base lapse rate is taken from the National Association of Insurance Commissioners' (NAIC) Valuation Manual 21. Dynamic lapse multiplier of this form is commonly used in practice to model simpler VA contracts such as GMMB and GMAB. Whether this dynamic lapse assumption is the most suitable for modeling GMMB and GMAB

#### NESTED SIMULATION FOR CTE OF VARIABLE ANNUITIES

Experiment Design	RSLN		GARCH(1,1)	
	GMMB	GMAB	GMMB	GMAB
SMC-5,000-200, CTE 80%	2.56%	6.57%	7.61%	13.17%
SMC-1,000-1,000, CTE 80%	1.96%	5.51%	1.57%	4.30%
SMC-200-5,000, CTE 80%	10.79%	21.95%	10.18%	19.29%
IANS, CTE 80%	0.44%	0.94%	0.41%	1.07%
SMC-5,000-200, CTE 95%	8.32%	15.44%	10.65%	26.33%
SMC-1,000-1,000, CTE 95%	5.38%	25.93%	5.78%	16.58%
SMC-200-5,000, CTE 95%	25.50%	101.76%	42.69%	99.67%
IANS, CTE 95%	1.28%	5.34%	1.32%	4.40%

 TABLE 6

 Relative Mean Square Errors for Different Experiment Designs Under Static Lapse

*Note:* RSLN = regime switching lognormal; GARCH = generalized autoregressive conditional heteroskedasticity; GMMB = guaranteed minimum maturity benefits; GMAB = guaranteed minimum accumulation benefits; SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation.

 TABLE 7

 Relative Bias for Different Experiment Designs Under Static Lapse

	RSLN		GARCH(1,1)	
Experiment Design	GMMB	GMAB	GMMB	GMAB
SMC-5,000-200, CTE 80%	1.42%	1.46%	2.69%	2.24%
SMC-1,000-1,000, CTE 80%	0.09%	-0.07%	0.32%	0.38%
SMC-200-5,000, CTE 80%	-0.05%	-0.22%	-0.39%	0.34%
IANS, CTE 80%	0.01%	0.04%	0.39%	-0.05%
SMC-5,000-200, CTE 95%	2.34%	1.89%	2.82%	2.70%
SMC-1,000-1,000, CTE 95%	0.51%	-0.47%	0.25%	0.06%
SMC-200-5,000, CTE 95%	-0.24%	-1.18%	-1.06%	-0.17%
IANS, CTE 95%	0.17%	-0.41%	-0.21%	0.20%

*Note:* RSLN = regime switching lognormal; GARCH = generalized autoregressive conditional heteroskedasticity; GMMB = guaranteed minimum maturity benefits; GMAB = guaranteed minimum accumulation benefits; SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation.

TABLE 8	
Relative Variance for Different Experiment Designs Under Static Lap	pse

Experiment Design	RSLN		GARCH(1,1)	
	GMMB	GMAB	GMMB	GMAB
SMC-5,000-200, CTE 80%	0.34%	1.00%	0.25%	0.97%
SMC-1,000-1,000, CTE 80%	1.95%	5.50%	1.47%	3.96%
SMC-200-5,000, CTE 80%	10.79%	21.83%	10.02%	19.01%
IANS, CTE 80%	0.44%	0.94%	0.26%	1.06%
SMC-5,000-200, CTE 95%	1.20%	4.32%	1.07%	5.26%
SMC-1,000-1,000, CTE 95%	5.05%	25.26%	5.70%	16.57%
SMC-200-5,000, CTE 95%	25.43%	97.46%	41.34%	99.58%
IANS, CTE 95%	1.24%	4.81%	1.26%	4.28%

*Note:* RSLN = regime switching lognormal; GARCH = generalized autoregressive conditional heteroskedasticity; GMMB = guaranteed minimum maturity benefits; GMAB = guaranteed minimum accumulation benefits; SMC = standard nested Monte Carlo; CTE = Conditional Tail Expectation; IANS = Importance-Allocated Nested Simulation.



FIGURE 6. Simulated Losses in 10,000 Outer Scenarios. *Note:* The x and y coordinates of each point in the figures represent the loss in a scenario, simulated by the Importance-Allocated Nested Simulation proxy simulation and by the true nested simulation, respectively. r.v. = random variable; GMMB = guaranteed minimum maturity benefits; RSLN = regime switching lognormal; GMAB = guaranteed minimum accumulation benefits; GARCH = generalized autoregressive conditional heteroskedasticity.

contracts is outside the scope of this article. Our focus is to demonstrate the effectiveness of the IANS method based on a model that is similar to industry practice.

For demonstration, we use the same proxy calculations as those in Section 4.2 despite the additional complexity of dynamic lapse. The proxy calculation uses base lapse assumptions. We observe that large periodic hedging gains and losses are more likely to result in extreme values of the loss random variable L. In addition, many large periodic hedging gains or losses occur in periods when the contracts are close to at-the-money and are close to renewal and maturity. Under such circumstances, the fund values are more likely to deplete at the base lapse rate.

CTE Level		#No. (%) of Correctly Identified Tail Scen.				
		RSLN		GARCH(1,1)		
	#No. Tail Scen.	GMMB	GMAB	GMMB	GMAB	
80%	2,000	1,989 (99.45%)	1,987 (99.35%)	1,957 (97.85%)	1,919 (95.95%)	
95%	500	500 (100.00%)	492 (98.40%)	499 (99.80%)	478 (95.6%)	

TABLE 9
Tail Scenario Identification by the Proxy Simulation (with Safety Margin) in
Stage I of the IANS Method Under Dynamic Lapse Assumption

*Note:* Scen. = scenario; RSLN = regime switching lognormal; GARCH = generalized autoregressive conditional heteroskedasticity; CTE = Conditional Tail Expectation; GMMB = guaranteed minimum maturity benefits; GMAB = guaranteed minimum accumulation benefits.

Figure 6 depicts the comparisons between the losses that are simulated by the true nested simulation and those by the IANS method's proxy simulation, for GMMB and GMAB contracts under dynamic lapse assumptions.

Similar to the results of the benchmark runs with no lapses, for both GMMB and GMAB contracts and under both asset models, the  $(1 - \alpha)J$  tail scenarios from nested simulations overlap almost entirely with the  $(1 - \xi)J$  proxy tail scenarios. This is also illustrated numerically in Table 9. Such overlapping suggest that the IANS method remains effective in this realistic setting using only simple proxy calculations. Another intriguing observation in Figure 6 is that, after modeling dynamic lapse, the simulated losses by the nested simulation and those by the proxy simulation can be significantly different in value. Nonetheless, the rankings of these simulated losses remain similar, so the proxy model can still effectively identify the true tail scenarios.

#### 5. CONCLUSION

In this article, we illustrate a simulation procedure for estimating the CTE of liabilities of a VA dynamic hedging strategy. The IANS method we propose takes advantage of the special structure of the CTE by first identifying a small set of potential tail scenarios from the first tier of simulation based on a proxy for liabilities calculated from a closed-form solution. We then focus the simulation budget on only those scenarios. We conduct extensive numerical experiments on GMMB and GMAB contracts. The numerical results show significant improvement in efficiency using the IANS method compared to a standard nested simulation.

The IANS method also inspires efficient experiment designs in other financial and actuarial applications where the CTE is estimated by Monte Carlo simulation. For future work, we consider a more rigorous and systematic approach in selecting  $\xi$ , the threshold for tail scenarios to be considered for nested simulations. We also explore solutions to improve the efficiency of nested simulation of Guaranteed Minimum Income Benefits and Guaranteed Minimum Withdrawal Benefits products.

#### ACKNOWLEDGMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada, funding reference number 203271 (Hardy) and 03755 (Feng). This work was also supported by the Society of Actuaries through a Center of Actuarial Excellence Research Grant and through the Hickman Scholarship held by Ou Dang. Ou Dang also receives support from the Ontario Graduate Scholarship.

#### ORCID

Ou Dang (b) http://orcid.org/0000-0002-9284-6550 Mingbin Feng (b) http://orcid.org/0000-0002-9748-6435 Mary R. Hardy (b) http://orcid.org/0000-0002-3199-4728

#### REFERENCES

- Artzner, P., F. Delbaen, J. M. Eber, and D. Heath. 1999. Coherent measures of risk. Mathematical Finance, 9(3): 203-28.
- Bauer, D., A. Reuss, and D. Singer. 2012. On the calculation of the solvency capital requirement based on nested simulations. ASTIN Bulletin: The Journal of the IAA 42(2): 453–99.
- Bollen, N. P. 1998. Valuing options in regime-switching models. The Journal of Derivatives 6(1): 38-49.
- Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31(3): 307-27.
- Broadie, M., Y. Du, and C. C. Moallemi. 2011. Efficient risk estimation via nested sequential simulation. Management Science 57(6): 1172-94.
- Broadie, M., Y. Du, and C. C. Moallemi. 2015. Risk estimation via regression. Operations Research 63(5): 1077-97.
- Broadie, M., and P. Glasserman. 1996. Estimating security price derivatives using simulation. Management Science 42(2): 269-85.
- Cathcart, M., and S. Morrison. 2009. Variable annuity economic capital: The least-squares Monte Carlo approach. Life & Pensions, pp. 36-40.
- Canadian Institute of Actuaries. 2017. Standards of Practice, Part 2000— Insurance. Technical report, Canadian Institute of Actuaries. Available at http:// www.cia-ica.ca/docs/default-source/standards/si020317e.pdf?sfvrsn=2.
- Duan, J. C. 1995. The GARCH option pricing model. Mathematical Finance 5(1): 13-32.
- Feng, R 2014. A comparative study of risk measures for guaranteed minimum maturity benefits by a PDE method. *North American Actuarial Journal* 18(4): 445–61.
- Feng, R., Z. Cui, and P. Li. 2016. Nested stochastic modeling for insurance companies. Technical report, Society of Actuaries. Available at https://www.soa. org/research/2015-nested-stochastic-modeling.pdf.
- Fu, M. C. ed. 2016. Handbook of simulation optimization. New York: Springer.
- Gan, G. 2013. Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics* 53(3): 795–801. doi:10.1016/j.insmatheco.2013.09.021.
- Gan, G. 2015a. Application of metamodeling to the valuation of large variable annuity portfolios. In *Proceedings of the 2015 Winter Simulation Conference*, 1103–14. doi:10.1109/WSC.2015.7408237.
- Gan, G. 2015b. A multi-asset Monte Carlo simulation model for the valuation of variable annuities. In *Proceedings of the 2015 Winter Simulation Conference*, 3162–63. doi:10.1109/WSC.2015.7408450.
- Gan, G., and X. S. Lin. 2015. Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics* 62: 138–50.
- Gan, G., and X. S. Lin. 2017. Efficient greek calculation of variable annuity portfolios for dynamic hedging: A two-level metamodeling approach. North American Actuarial Journal 21(2): 161–77.
- Glasserman, P. 2013. Monte Carlo methods in financial engineering, Vol. 53. New York: Springer.
- Gordy, M. B., and S. Juneja. 2010. Nested simulation in portfolio risk measurement. Management Science 56(10): 1833-48.
- Hardy, M. R. 2001. A regime-switching model of long-term stock returns. North American Actuarial Journal 5(2): 41-53.
- Hardy, M. R. 2003. Investment guarantees: Modeling and risk management for equity-linked life insurance, Vol. 215. New York: Wiley & Sons.
- Lan, H., B. L. Nelson, and J. Staum. 2010. A confidence interval procedure for expected shortfall risk measurement via two-level simulation. *Operations Research* 58(5): 1481–90.
- L'Ecuyer, P. 1990. A unified view of the IPA, SF, and LR gradient estimation techniques. Management Science 36(11): 1364-83.
- Liu M., and J. Staum. 2010. Stochastic kriging for efficient nested simulation of expected shortfall. The Journal of Risk 12(3): 3.
- Marshall C., M. R. Hardy, and D. Saunders. 2010. Valuation of a guaranteed minimum income benefit. North American Actuarial Journal 14(1): 38-58.
- Ng A. C. Y., and J. S. H. Li. 2013. Pricing and hedging variable annuity guarantees with multiasset stochastic investment models. *North American Actuarial Journal* 17(1): 41–62.
- Office of the Superintendent of Financial Institutes Canada. 2017. Life Insurance Capital Adequacy Test. Technical report, Office of the Superintendent of Financial Institutes Canada. Available at http://www.osfi-bsif.gc.ca/Eng/Docs/LICAT18.pdf.
- Wirch J. L., and M. R. Hardy. 1999. A synthesis of risk measures for capital adequacy. Insurance: Mathematics and Economics 25(3): 337-47.

Discussions on this article can be submitted until January 1, 2021. The authors reserve the right to reply to any discussion. Please see the Instructions for Authors found online at http://www.tandfonline.com/uaaj for submission instructions.

#### APPENDIX. ADVANCED ASSET MODELS AND VOLATILITY CALIBRATION

To implement the IANS method, we are interested in calibrating the volatility of the proxy Black-Scholes model to that in the true asset model. In this appendix we consider two advanced stochastic asset models: the two-regime RSLN model and GARCH(1,1). Besides brief summaries of the two asset models, we show calculations of  $\tilde{\sigma}(t_1, t_2|t_0)$ , the average volatility in the period  $[t_1, t_2)$  conditioning on the current state at time  $t_0$ , where  $t_0 \le t_1 < t_2$ .

#### A.1. RSLN Model

In general, regime-switching models assume that a discrete process switches between regimes randomly. It is common to assume that switching process is Markovian (Hardy 2003), that is, conditioning on the current regime, the probability of changing regime is independent of the history of the switching process.

Each regime in a regime-switching model is characterized by a different parameter set or even a different model. For simplicity we consider the two-regime RSLN, which "provides a very good fit to the stock index data relevant to equity-linked insurance" (Hardy 2003, 31).

Let  $\rho(t) = 1, 2$ , denote the regime for period *t*, that is, [t, t+1), and let S(t) be the underlying stock price at time *t*. Let  $Y(t) = \ln (S(t+1)/S(t))$  be the log-return process, then  $Y(t)|\rho(t) \sim N(\mu_{\rho(t)}, \sigma_{\rho(t)}^2)$  where  $\mu_h, \sigma_h^2$  are the mean and variance parameter of the *h*th regime. The transition probabilities

$$p_{ij} = \Pr[\rho(t+1) = j | \rho(t) = i], \quad i, j = 1, 2$$

denote the probabilities of switching regimes, which is assumed to take place at the end of each period. Besides the riskfree rate *r*, there are six parameters in an RSLN model with two regimes, which are assumed to be accurately estimated in our setting. The parameters for the two-regime RSLN model used in our numerical studies are summarized in Table 2.

The financial market is incomplete in the regime-switching model, thus its risk neutral measure is not unique (Hardy 2001). Given the real-world measure in the regime-switching model, we employ the risk-neutral model studied in Bollen (1998) and Hardy (2001), whose mean conditional log return is  $r - \sigma_i^2/2$  for i = 1, 2. Let  $Q(t_1, t_2)$  be the number of sojourns in regime 1 in  $[t_1, t_2)$ , its conditional expected value and variance given the regime at time  $t_1$ , that is,  $E[Q(t_1, t_2)|\rho(t_1)]$  and  $Var[Q(t_1, t_2)|\rho(t_1)]$ , can be calculated via backward recursion (Hardy 2003, Chapter 6.1). These are useful quantities for volatility calibrations in the IANS method.

**Proposition A.1.** Consider a two-regime RSLN model. For any  $t_0 \le t_1 < t_2$ , denote the average volatility in the period  $[t_1, t_2)$  conditioning on the current state at time  $t_0$  by  $\tilde{\sigma}(t_1, t_2|t_0) = \sqrt{\tilde{\sigma}^2(t_1, t_2|t_0)}$ .

If  $t_0 = t_1$ , then

$$\tilde{\sigma}^{2}(t_{1}, t_{2}|t_{1}) = \frac{1}{t_{2} - t_{1}} \times ((t_{2} - t_{1})\sigma_{2}^{2} + (\sigma_{1}^{2} - \sigma_{2}^{2})E[Q(t_{1}, t_{2})|\rho(t_{1})] + \frac{1}{4}(\sigma_{1}^{2} - \sigma_{2}^{2})^{2}Var[Q(t_{1}, t_{2})|\rho(t_{1})]).$$

If  $t_0 < t_1$ , then

$$\tilde{\sigma}^{2}(t_{1}, t_{2}|t_{0}) = \frac{1}{t_{2} - t_{1}} \times ((t_{2} - t_{1})\sigma_{2}^{2} + (\sigma_{1}^{2} - \sigma_{2}^{2})E[E[Q(t_{1}, t_{2})|\rho(t_{1})]|\rho(t_{0})] \\ + \frac{1}{4}(\sigma_{1}^{2} - \sigma_{2}^{2})^{2}(E[Var[Q(t_{1}, t_{2})|\rho(t_{1})]|\rho(t_{0})] + Var[E[Q(t_{1}, t_{2})|\rho(t_{1})]|\rho(t_{0})])).$$

where

$$\begin{split} E[E[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)] &= \sum_{h=1}^{2} E[Q(t_1, t_2)|\rho(t_1) = h] \times \Pr[\rho(t_1) = h|\rho(t_0)], \\ E[Var[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)] &= \sum_{h=1}^{2} Var[Q(t_1, t_2)|\rho(t_1) = h] \times \Pr[\rho(t_1) = h|\rho(t_0)], \text{ and} \\ Var[E[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)] &= \sum_{h=1}^{2} E\Big[Q(t_1, t_2)^2|\rho(t_1) = h\Big] \times \Pr[\rho(t_1) = h|\rho(t_0)] \\ &\quad - E[E[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)]^2 \end{split}$$

*Proof.* If  $t_0 < t_1$ , then

$$\tilde{\sigma}^2(t_1, t_2|t_0) = \frac{1}{t_2 - t_1} Var\left[\sum_{t=t_1}^{t_2 - 1} Y(t)|\rho(t_0)\right]$$

where

$$Var\left[\sum_{t=t_1}^{t_2-1} Y(t)|\rho(t_0)\right] = E\left[Var\left[\sum_{t=t_1}^{t_2-1} Y(t)|Q(t_1,t_2)\right]|\rho(t_0)\right] + Var\left[E\left[\sum_{t=t_1}^{t_2-1} Y(t)|Q(t_1,t_2)\right]|\rho(t_0)\right]$$

In a regime-switching model, we have

$$\begin{split} E\left[\sum_{i=t_1}^{t_2-1} Y(t) | \mathcal{Q}(t_1, t_2)\right] &= \left(r - \frac{\sigma_1^2}{2}\right) \mathcal{Q}(t_1, t_2) + \left(r - \frac{\sigma_2^2}{2}\right) [(t_2 - t_1) - \mathcal{Q}(t_1, t_2)] \\ &= \frac{1}{2} \left(\sigma_2^2 - \sigma_1^2\right) \mathcal{Q}(t_1, t_2) + \left(r - \frac{\sigma_2^2}{2}\right) (t_2 - t_1) \\ Var\left[\sum_{t=t_1}^{t_2-1} Y(t) | \mathcal{Q}(t_1, t_2)\right] &= \sigma_1^2 \mathcal{Q}(t_1, t_2) + \sigma_2^2 [(t_2 - t_1) - \mathcal{Q}(t_1, t_2)] \\ &= \left(\sigma_1^2 - \sigma_2^2\right) \mathcal{Q}(t_1, t_2) + \sigma_2^2 (t_2 - t_1) \\ E\left[Var\left[\sum_{t=t_1}^{t_2-1} Y(t) | \mathcal{Q}(t_1, t_2)\right] | \rho(t_0)\right] &= E\left[(\sigma_1^2 - \sigma_2^2) \mathcal{Q}(t_1, t_2) + \sigma_2^2 (t_2 - t_1) | \rho(t_0)\right] \\ &= (t_2 - t_1)\sigma_2^2 + (\sigma_1^2 - \sigma_2^2) E[\mathcal{Q}(t_1, t_2) | \rho(t_0)] \\ &= (t_2 - t_1)\sigma_2^2 + (\sigma_1^2 - \sigma_2^2) E[\mathcal{Q}(t_1, t_2) | \rho(t_0)] \\ &= (t_2 - t_1)\sigma_2^2 + (\sigma_1^2 - \sigma_2^2) E[\mathcal{Q}(t_1, t_2) | \rho(t_0)] \\ Var\left[E\left[\sum_{t=t_1}^{t_2-1} Y(t) | \mathcal{Q}(t_1, t_2)\right] | \rho(t_0)\right] &= Var\left[\frac{1}{2} (\sigma_2^2 - \sigma_1^2) \mathcal{Q}(t_1, t_2) + \left(r - \frac{\sigma_2^2}{2}\right) (t_2 - t_1) | \rho(t_0)\right] \\ &= \frac{1}{4} (\sigma_1^2 - \sigma_2^2)^2 (E[Var[\mathcal{Q}(t_1, t_2) | \rho(t_0)] \\ &= \frac{1}{4} (\sigma_1^2 - \sigma_2^2)^2 (E[Var[\mathcal{Q}(t_1, t_2) | \rho(t_0)] | \rho(t_0)] \end{split}$$

If 
$$t_0 = t_1$$
,

$$\begin{split} E[E[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)] &= E[Q(t_1, t_2)|\rho(t_1)]\\ E[Var[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)] &= E[Var[Q(t_1, t_2)|\rho(t_1)]]\\ Var[E[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)] &= Var[E[Q(t_1, t_2)|\rho(t_1)]] \end{split}$$

+  $Var[E[Q(t_1, t_2)|\rho(t_1)]|\rho(t_0)]))$ 

This completes the proof.

#### A.2. GARCH(1,1) Model

The GARCH model was first developed by Bollerslev (1986) and is one of the most popular asset models because of its flexibility and good fit for many econometric applications. For simplicity we consider GARCH(1,1) in our numerical study to model the monthly log return of the stock price and its variance:

$$Y(t) = \mu + \sigma(t)\varepsilon(t), \quad \varepsilon(t) \text{ iid } \sim N(0,1),$$
  

$$\sigma^{2}(t) = \alpha_{0} + \alpha_{1}\sigma^{2}(t-1)\varepsilon^{2}(t-1) + \beta\sigma^{2}(t-1).$$

The current state in the above GARCH(1,1) model at time *t* includes Y(t) and  $\sigma(t)$ . In addition, the log return of the stock price and variance under the risk-neutral measure are given by Ng and Li (2013):

$$\begin{split} \tilde{Y}(t) &= r - \frac{1}{2}\sigma^2(t) + \sigma(t)\tilde{\varepsilon}(t), \quad \tilde{\varepsilon}(t) \text{ iid } \sim N(0,1), \\ \sigma^2(t) &= \alpha_0 + \alpha_1\sigma^2(t-1)\tilde{\varepsilon}^2(t-1) + \beta\sigma^2(t-1). \end{split}$$

The parameters for the GARCH(1,1) model used in our numerical studies are summarized in Table 3. The value of  $\alpha_0$  is chosen such that long-run average volatility in this GARCH(1,1) models equals to the long-run average volatility in the RSLN model with parameters in Table 2.

To calibrate the volatility in the Black-Scholes model to the conditional expected volatility in the GARCH(1,1) model, one needs the average volatility in the period  $[t_1, t_2)$  conditioning on the current state at time  $t_0$ , where  $t_0 \le t_1 < t_2$ , which is summarized below.

**Proposition A.2.** Consider a GARCH(1,1) model. For any  $t_0 \le t_1 < t_2$ , denote the average volatility in the period  $[t_1, t_2)$  conditioning on the current state at time  $t_0$  by  $\tilde{\sigma}(t_1, t_2|t_0) = \sqrt{\tilde{\sigma}^2(t_1, t_2|t_0)}$ .

If 
$$t_0 = t_1$$
,

$$\tilde{\sigma}^{2}(t_{1},t_{2}|t_{1}) = \frac{1}{t_{2}-t_{1}} \times \left(\sigma^{2}(t_{1}) + \frac{\alpha_{0}(t_{2}-t_{1}-1)}{1-\alpha_{1}-\beta} + \left(\sigma^{2}(t_{1}+1) - \frac{\alpha_{0}}{1-\alpha_{1}-\beta}\right) \frac{1-(\alpha_{1}+\beta)^{t_{2}-t_{1}-1}}{1-\alpha_{1}-\beta}\right)$$

*If*  $t_0 < t_1$ ,

$$\tilde{\sigma}^{2}(t_{1},t_{2}|t_{0}) = \frac{1}{t_{2}-t_{1}} \times \left(\frac{\alpha_{0}(t_{2}-t_{1})}{1-\alpha_{1}-\beta} + (\alpha_{1}+\beta)^{t_{1}-t_{0}-1} \left(\sigma^{2}(t_{0}+1) - \frac{\alpha_{0}}{1-\alpha_{1}-\beta}\right) \frac{1-(\alpha_{1}+\beta)^{t_{2}-t_{1}}}{1-\alpha_{1}-\beta}\right)$$

*Proof.* Let  $\mathcal{F}_t$  denote the filtration at time *t*. More specifically,  $\mathcal{F}_t$  represents information at time *t*, such as  $\sigma(t)$  and  $\varepsilon(t)$ . We define

$$ilde{\sigma}^2(t_1, t_2|t_0) = rac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2 - 1} E \Big[ \sigma^2(t) |\mathcal{F}_{t_0} \Big]$$

For  $t \ge t_1$ ,

$$\begin{split} E[\sigma^{2}(t)|\mathcal{F}_{t_{0}}] &= E[\alpha_{0} + \alpha_{1}\sigma^{2}(t-1)\tilde{\epsilon}^{2}(t-1) + \beta\sigma^{2}(t-1)|\mathcal{F}_{t_{0}}] \\ &= \alpha_{0} + E[(\alpha_{1}\tilde{\epsilon}^{2}(t-1) + \beta)\sigma^{2}(t-1)|\mathcal{F}_{t_{0}}] \\ &= \alpha_{0} + E[(\alpha_{1}\tilde{\epsilon}^{2}(t-1) + \beta)|\mathcal{F}_{t_{0}}] \times E[\sigma^{2}(t-1)|\mathcal{F}_{t_{0}}] \\ &= \alpha_{0} + (\alpha_{1} + \beta)E[\sigma^{2}(t-1)|\mathcal{F}_{t_{0}}] \\ &\vdots \\ &= \alpha_{0}\sum_{i=0}^{t-t_{0}-2}(\alpha_{1} + \beta)^{i} + (\alpha_{1} + \beta)^{t-t_{0}-1}E[\sigma^{2}(t_{0} + 1)|\mathcal{F}_{t_{0}}] \\ &= \frac{\alpha_{0}(1 - (\alpha_{1} + \beta)^{t-t_{0}-1})}{1 - \alpha_{1} - \beta} + (\alpha_{1} + \beta)^{t-t_{0}-1}\sigma^{2}(t_{0} + 1) \end{split}$$

If  $t_0 = t_1$ ,  $\sum_{t=t_1}^{t_2-1} E[\sigma^2(t)|\mathcal{F}_{t_1}] = \sigma^2(t_1) + \sum_{t=t_1+1}^{t_2-1} \left( \frac{\alpha_0(1-(\alpha_1+\beta)^{t-t_1-1})}{1-\alpha_1-\beta} + (\alpha_1+\beta)^{t-t_1-1}\sigma^2(t_1+1) \right)$   $= \sigma^2(t_1) + \frac{\alpha_0(t_2-t_1-1)}{1-\alpha_1-\beta} + \left( \sigma^2(t_1+1) - \frac{\alpha_0}{1-\alpha_1-\beta} \right) \frac{1-(\alpha_1+\beta)^{t_2-t_1-1}}{1-\alpha_1-\beta}$ 

If  $t_0 < t_1$ ,

$$\begin{split} \sum_{t=t_1}^{t_2-1} & E\left[\sigma^2(t)|\mathcal{F}_{t_0}\right] &= \sum_{t=t_1}^{t_2-1} \left(\frac{\alpha_0(1-(\alpha_1+\beta)^{t-t_0-1})}{1-\alpha_1-\beta} + (\alpha_1+\beta)^{t-t_0-1}\sigma^2(t_0+1)\right) \\ &= \frac{\alpha_0(t_2-t_1)}{1-\alpha_1-\beta} + (\alpha_1+\beta)^{t_1-t_0-1} \left(\sigma^2(t_0+1) - \frac{\alpha_0}{1-\alpha_1-\beta}\right) \frac{1-(\alpha_1+\beta)^{t_2-t_1}}{1-\alpha_1-\beta} \end{split}$$

This completes the proof.