

Using Concomitant and Nested Simulation for Tail Risk Measure Estimation

Jessica Ou Dang ^{*†}M. Ben Feng ^{*}

Abstract

Tail risk measures is of critical importance for enterprise risk management, especially for managing large portfolios of complex financial instruments. The computational burdens required by such simulation can be substantial or even unbearable, depending on the complexity of the underlying economic model and the risk management objective. This paper proposes, analyzes, and tests an efficient nested simulation procedure for estimating tail risk measures. The procedure first uses proxy models and their concomitants to quickly and accurately identify tail scenarios where the given computational budget is concentrated. We demonstrate the proposed procedure in estimating tail risk measures of variable annuities. Our results show that, given a fixed computational budget, the proposed procedure can be an order of magnitude more accurate than a standard nested simulation procedure.

Key Words: Nested Simulation, Conditional Tail Expectation, Variable Annuities, Concomitant, Tail Value at Risk

1 Introduction

Variable annuity (VA) are insurance contracts that are widely used to provide retirement income. Its annual sales in the U.S. market in recent years is around \$100 billion¹. In addition to death and survival benefits to the policyholder, a VA contract also offers benefits that are linked to the performance of certain financial assets, e.g. stocks and bonds. Such equity-linked benefits are offered in the form of riders, which provides different types guarantees to protect the policyholder from downside market risks. From the insurer's perspective, these guarantees can be viewed as embedded options on the underlying financial assets offered to the policyholder. The complexities of the embedded options varies: from European put options of standard Guaranteed Minimum Maturity Benefits (GMMB) to complex combinations of path dependent, exotic look-back and tandem options of the Guaranteed Minimum Income Benefit (GMIB) studied by Marshall et al. (2010). Similar to the risk management of financial options, in practice insurers commonly use a dynamic hedging strategy to mitigate the financial risks of the guarantees in a VA contract. For insurers who must model the liabilities for valuation and hedging purposes, depending on the complexity of the economic model and the options under consideration, risk measurement and management could be difficult tasks. For example, due to the long term nature of insurance policies such as VAs, simple economic models such as log-normal returns with constant volatility are often insufficient for risk management purpose, which often concerns the tail risk of the loss distribution. In addition, as alluded above, some embedded options are complicated in nature. Lastly, as a unique feature to VAs, insurers need to consider the policyholder behaviors, e.g. full/partial withdrawal (lapse) of the contract, under different economic scenarios. Due to these complexities, Monte Carlo simulation is often the only viable way for measuring and managing the risks of a VA contract. Indeed, it is now standard in the insurance industry to use Monte Carlo simulation to determine hedge portfolios. The computational burden is substantial and can sometimes become the bottleneck of making high-quality risk management decisions. The goal of this research is to provide an efficient and reliable nested simulation framework that can quickly and accurately estimate the tail risk measures of a dynamic hedging program of VA.

^{*}Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1

[†]Corresponding author. E-mail: jessica.dang@uwaterloo.ca

¹Source: LIMRA Data Bank

Tail risk measures are important enterprise risk management for insurance companies. Such importance is evidenced by their ubiquitous presence in regulations around the globe: As a main regulatory directive in the European Union, Solvency II requires insurers to project all their future cashflows 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, insurers use a similar approach under their Own Risk Solvency Assessment (ORSA) obligations, but with the Conditional Tail Expectation (CTE) 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 (CIA 2017). The total gross calculated requirement (TGCR) is generally set at the 95% CTE, so that the required capital is the difference between the 95% CTE and the reserve (OSFI 2017). In the US, 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 behaviour 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. In some simple economic models, 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 behaviour. 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 last period is insufficient to fund the hedge required for the next time 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 $20 \times 12 \times 5000 \times 1000 = 1.2 \times 10^9$ total simulated asset or liability values. If the inner simulations were single-step, and if each simulated value takes $1 \mu s$ (10^{-6} 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. 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 & Morrison (2009) and Feng et al. (2016) which were 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 using nested simulation is the topic of Gordy & Juneja (2010), Liu & Staum (2010), and Broadie et al. (2011). Solvency II SCR (Solvency Capital Requirement) calculation is the topic of Bauer et al. (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 by reducing one or more of the above four factors. Gan (2013, 2015a,b), Gan & Lin (2015, 2017) propose 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 paper 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)), stochastic kriging (a more sophisticated interpolation method, see Liu & Staum (2010) for example), least-squares regression, (see Broadie et al. (2015) and Cathcart & Morrison (2009), for example), or a generic partial differential equation approach (Feng 2014). Each of these may be referred to as a *proxy model*, as a pilot simulation 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 efficiently allocate a given computational budget between the outer and inner simulations. Gordy & Juneja (2010) demonstrate that outer simulations are more important than inner simulations in accurate estimation of tail risk measures, so that at some point the advantage gained from additional inner simulations is minimal. They then propose a method of strategic allocation of the budget between inner and outer simulations. They use 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 which was not uniform, where more inner simulations were applied to some outer paths than others. *Screening* refers to the more extreme form of allocation, under which some 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 & Staum (2010) uses an iterative approach involving stochastic kriging and screening, in three stages. Lan et al. (2010) also use screening, based on a pilot study.

In Dang et al. (2018) we proposed an efficient nested simulation procedure, called the *Importance-Allocated Nested Simulation* (IANS) method. The IANS method uses two stages for the inner simulations. The first stage uses a low-cost proxy model to identify the *potential tail scenarios* that are most likely to contribute to the CTE risk measure. In the second stage the entire computational budget is concentrated on the tail scenarios identified in the first stage. Some important questions remains unresolved in Dang et al. (2018) and only ad-hoc steps were taken in its numerical demonstrations. More specifically, in the first stage, how many potential tail scenarios should be identified and why?

In this paper we derive the expected value and variance of rank of concomitant in bivariate order statistics in the context of nested simulation. Based on these results, we then design an algorithm to search for the appropriate number of potential tail scenarios to be identified in the first stage of the IANS method. The proposed algorithm allows us to verify if the potential tail scenarios identified in the first stage of the IANS method sufficiently cover the tail of the true distribution of inner simulation loss. More broadly, the method to construct confidence interval of rank of the inner simulation loss may also be applied in other proxy approaches for nested simulation. Our numerical experiments show that, the proposed algorithm generates a cut-off for potential tail scenarios that capture the true tail scenarios in terms of inner simulation 94% of the time and achieves comparable level of efficiency improvement as that shown in the numerical experiments in Dang et al. (2018).

The remainder of this article is organized as follows: Section 2 provides an overview of Variable Annuity, its dynamic hedging practice and the process of a standard nested simulation. Section 3 recaps the Important-Allocated Nested Simulation method presented in Dang et al. (2018). Section 4 derives the first two moments of the rank of concomitant and presents the algorithm to find the optimal cut-off for potential tail scenarios. Section 5 illustrates the performance of the proposed method in numerical experiments. Section 6 concludes.

2 Modelling 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 paper. We also describe the standard two level nested simulation, with path dependent outer simulations, which is the base method used as a benchmark for the IANS approach.

2.1 Overview of Variable Annuities

A variable annuity (VA) is a deferred annuity that allows the policyholder to invest contributions into mutual funds or sub-accounts, subject to an explicit or implicit guarantee fee in addition to the regular mutual fund management fee. The underlying investment in the mutual funds include equities and fixed income assets. Lapse is allowed and the fund value less any applicable surrender charges are paid upon lapse. As insurance contracts, unlike mutual fund investments, a VA contract provides downside protections from the fluctuation of the financial markets in the form of minimum guarantees.

Consider a generic VA contract whose *time to maturity* is $T \geq 0$. Let $F(t)$ and $G(t)$ be the sub-account fund value and the guarantee value, respectively, at time $t = 0, \dots, T$. The major types of benefits provided by Variable Annuities in the market today include:

Guaranteed Minimum Maturity Benefit (GMMB) Provided survival of the policyholder, GMMB contract pays $\max(F(T), G(T))$ to the policyholder at maturity T , of which the insurer is liable for paying $\max(G(T) - F(T), 0)$. In other words, from the insurer's perspective, issuing a GMMB contract is equivalent to holding a short position on a put option of the underlying investment. This is one of the simplest VA contracts. For ease of exposure we consider GMMB in our numerical experiments.

Guaranteed Minimum Death Benefit (GMDB) A GMDB contract pays $\max(F(t), G(t))$ upon the policyholder's death at time t . $G(t)$ is typically 75% of 100% of the original premium, if we ignore any previous partial surrender or subsequent premium paid. Under simplifying assumptions (e.g., no lapse), GMDB's payoff coincides with that of the GMMB.

Guaranteed Minimum Accumulation Benefit (GMAB) A GMAB contract guarantees a minimum fund value at both renewal, say at time $R < T$, and maturity of the contract. Let R^- and R^+ be the instants before and after the contract renewal, respectively. Upon renewal at R , a new guarantee value $G(R^+)$ may be set according to the following rules and will be effective from R^+ to T :

- If $F(R) \leq G(R)$, the new guarantee value $G(R^+) = G(R)$. In this case the insurer pays the deficit $G(R) - F(R)$ into the sub-account to meet the new guarantee value.
- If $F(R) > G(R)$, then $G(R^+) = F(R)$. In this case the insurer does not need to pay additional cash flows into the sub-account.

In essence, the new guarantee value is $G(R^+) = \max(F(R), G(R))$. At the maturity T the insurer pays $\max(F(T), G(T))$. In practice, minimum term applies (typically 10 years) on renewal and there may be a limit to the number of renewals allowed.

Guaranteed Minimum Income Benefit (GMIB) A GMIB contract guarantees the minimum annual income rate at which the policyholder can convert the fund value to an annuity benefit. The contract typically pays $\max(F(t), G(t))$ upon the policyholder's death at time t or $\max(F(T), G(T))$ at maturity T (Dai et al. (2008)). This is one of the complicated types of VA contracts.

For more information on VA contracts and different types of guarantees, readers are encouraged to refer to Hardy (2003).

2.2 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. For instance, for a GMMB contract whose underlying asset is a stock, the contract is equivalent to a short position on the put option of the stock from the insurer's perspective. As a result, the insurer will short the appropriate amount of future contracts of this stock in the hedging portfolio. The hedging portfolio is rebalanced periodically, responding to changes in market conditions and in the demographics of the block of contracts.

In this paper 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 only consider 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 guaranteed minimum maturity benefit (GMMB) this would be the expiry date of the policy. At any $t \leq T$, 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 t th time period, the value of this hedge has changed to

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

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

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

The costs to set up the initial hedging portfolio, the periodic hedging losses due to rebalancing, and the final unwinding of the hedge are recognized as part of the profit and loss (P&L) 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 straightforward GMMB 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 $F(0) = S(0)$. For simplicity, the guaranteed minimum benefit is assumed to be a fixed value, G . 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 $H(T) = (G - S(T))^+$. The discounted hedging loss random variable is given by

$$L = H(0) + \sum_{t=1}^T e^{-rt} HE(t) \quad (3)$$

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

Please refer to Dang et al. (2018) for derivation and interpretation of (4).

Using Monte Carlo simulation to assess the costs of hedging, we generate J simulated loss values L_1, \dots, L_J . The order statistics of the simulated losses are denoted by $L_{(1)}, \dots, L_{(J)}$, where $L_{(j)}$ is the j th smallest value. An estimator of the $100\alpha\%$ -CTE is given by

$$CTE_\alpha = \frac{1}{J(1-\alpha)} \sum_{j=N\alpha+1}^J L_{(j)}. \quad (5)$$

Algorithm 1 outlines the steps of a standard two-level nested simulation for estimating CTE_α of losses for a Delta-hedged VA contract. A generic stock price model is considered and path-dependent guarantees are allowed. While omitted for clarity, practical considerations such as fees, renewals, and policyholder behaviors can be incorporated into Algorithm 1 easily.

The different purposes of the outer- and inner-simulations result in different stochastic asset models being applied. For the purpose of estimating CTE of L , a real-world model is used in outer simulations of stock price paths (Line 5) to examine losses associated with the VA contract under realistic scenarios. Meanwhile, a risk-neutral model is used in inner simulations (Line 11) for evaluating the $\Delta(t)$ values and subsequently the hedge costs for each time step for each scenario. In simple models, inner simulations may not be necessary and the hedge cost at each time point can be determined analytically, as we demonstrate in the following section. However, when we use a stochastic volatility model, or when we incorporate dynamic lapses or other complications, inner simulations will usually be required.

The evolution of sub-account 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. Algorithm 1 can also be extended to hedging strategies that depend on other sensitivities, e.g. Gamma, Rho, Theta, etc. 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, VIX and others. See L'Ecuyer (1990), Glasserman (2013), Fu et al. (2016) for more information on estimating greeks using Monte Carlo simulation. We use the Infinite Perturbation Analysis (IPA) (Broadie & Glasserman 1996, Glasserman 2013) method for sensitivity estimation in our numerical studies.

2.3 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, then the hedge portfolio can be determined analytically, without requiring the inner simulation step.

Consider 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 j th outer simulation, $H_j(t)$, 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 σ is the volatility of the risk neutral GBM, expressed per time unit:

$$H_j(t) = G(t)e^{-r(T-t)}\Phi(-d_2) - S_j(t)\Phi(-d_1), \quad \Delta_j(t) = -\Phi(-d_1) \quad (6)$$

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

$$d_1(t, T) = \frac{\ln\left(\frac{S_j(t)}{G(t)}\right) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \quad d_2(t, T) = d_1(t, T) - \sigma\sqrt{T-t}. \quad (7)$$

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

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_α .

output: CTE_α 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* which provide tractable, analytic hedge costs, and for which the payoff which is expected to be well-correlated to the VA guarantee costs.
- (I.2) Calibrate the proxy asset model to the underlying risk-neutral asset model in inner-level simulations.
- (I.3) Implement Algorithm 1 but with the analytic hedge calculations for the proxy derivative and asset model replacing the inner simulation step.
- (I.4) Identify $(1 - \xi)J$ proxy tail scenarios with the largest simulated loss in Step (I.3) for some $\xi \in [0, \alpha]$.

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_α in Equation (5) as the output.
-

feasible approach is Monte Carlo simulation. But 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 Importance-Allocated Nested Simulation (IANS) Method

In this section, we present an outline of *Importance-Allocated Nested Simulation (IANS)* method for estimating the CTE_α of a VA GMMB, using a nested simulation with screening approach. More detailed explanation of this approach can be found in Dang et al. (2018).

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. The user must specify some parameters and experiment design choices that govern the behavior of the IANS method.

Unlike a standard proxy approach, the proxy tail scenarios in the IANS method do not need to accurately assess the liability values for those scenarios – what we use the proxy step for is 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.

For GMMB, the put option identified in Section 2.3 is an obvious proxy derivative, as the option payoffs are identical to the guarantee payoffs, if we ignore complications of policyholder behaviour. 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.

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 ξ in Step (I.4) with some safety margin, so that $\alpha - \xi \geq 0$. This means that the proxy tail scenarios are the $(1 - \xi)J$ outer scenarios with the largest simulated loss based on the proxy calculations. We use these to identify the largest $(1 - \alpha)J$ simulated loss 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 ξ is an experiment design parameter in IANS. If ξ 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 loss values and CTE estimates. On the other hand, if ξ is close to α , the inner simulation budget is focused on fewer scenarios, so those 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 Dang et al. (2018), the work was based on an arbitrary safety margin of 5%. Here in this paper, we propose a structured approach to determine the optimal proxy confidence level ξ in Section 4.

4 Concomitant of Order Statistics in Application of IANS Method

A nested simulation using the IANS method generates a sample with $(1 - \xi)J$ pairs of bivariate output, corresponding to the $(1 - \xi)J$ proxy tail scenarios. Each pair of output consists of hedging loss estimated by proxy and by inner simulation, denoted by (\widehat{L}_k, L_k) , respectively, for $k = 1, \dots, (1 - \xi)J$. (\widehat{L}_k, L_k) are referred to as proxy loss and inner simulation loss hereinafter. Given this sample, we try to determine if the $(1 - \alpha)J$ true tail scenarios are a subset of the $(1 - \xi)J$ proxy tail scenarios. For this purpose, we only need to focus on the quantile of proxy loss and inner simulation loss, rather than the actual value of such losses, which simplifies the problem to one with a bivariate uniform distribution. Because the order statistics of the proxy loss among all scenarios are known, we can study the quantile of inner simulation loss as a concomitant of the quantile of proxy loss. We first derive the first and second moment of the rank of concomitant in a bivariate uniform distribution. We then illustrate how they can be applied in the selection of the optimal proxy confidence level ξ in the IANS method.

4.1 Rank of Concomitant of Order Statistics

Let $(U_i, V_i), i = 1, \dots, n$ be n independent pairs of r.v. with a common bivariate distribution. We denote the r^{th} order statistic among n U_i 's as $U_{r:n}$. Then the V-variate associated with $U_{r:n}$ is called the concomitant of the r^{th} order statistic (David et al. 1977), and is denoted as $V_{[r:n]}$. The rank of $V_{[r:n]}$ among all n V_i 's are denoted as $R_{r:n}$. In other words, $R_{r:n} = s$ implies $V_{[r:n]} = V_{s:n}$. In Figure 1, we use a sample with 5 pairs of (U_i, V_i) 's as an example to illustrate the value of $U_{r:n}$, $V_{[r:n]}$, and $R_{r:n}$ in this case.

David et al. (1977) derived a general expression for the expected value of $R_{r:n}$, $E[R_{r:n}]$, for any bivariate distribution of (U, V) . Since the bivariate random variables we focus on are quantiles of the proxy loss and inner simulation loss, (U, V) has a bivariate uniform distribution in this case. This allows us to use the copula function $C(U, V)$ to denote the distribution function of (U, V) . We also denote the density function of the copula as $c(U, V)$.

Using the results in David et al. (1977), we derive Proposition 4.1

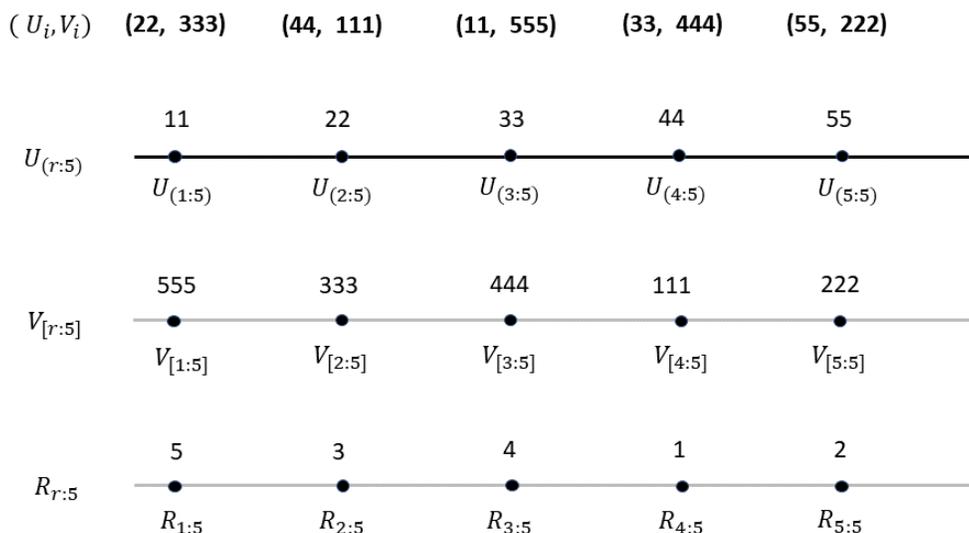


Figure 1: Example: Rank of Concomitant of Order Statistics

Proposition 4.1. Suppose (U, V) has bivariate uniform distribution. In a sample with n pairs of (U, V) 's, the expected value of the rank of concomitant of U 's r^{th} order statistic is

$$E[R_{r:n}] = 1 + n \left(\int_0^1 \left[\int_0^1 C(u, v) c(u, v) dv \right] f_{U_{r-1:n-1}}(u) du + \int_0^1 \left[\int_0^1 (v - C(u, v)) c(u, v) dv \right] f_{U_{r:n-1}}(u) du \right) \tag{8}$$

where $f_{U_{r:n}}(u)$ is the density function of $U_{r:n}$. More specifically, $f_{U_{r:n}}(u) = \frac{n!}{(r-1)!(n-r)!} u^{r-1} (1-u)^{n-r}$.

Furthermore, in O'Connell (1974), the author derived an expression for $E[R_{r:n}^2]$, where U and V are linearly correlated. Following the same methodology as in O'Connell (1974), we derive Proposition 4.2

Proposition 4.2. Suppose (U, V) has bivariate uniform distribution. In a sample with n pairs of (U, V) 's, the second moment of the rank of concomitant of U 's r^{th} order statistic is

$$E[R_{r:n}^2] = 3E[R_{r:n}] - 2 + n(n-1) \times \left(\int_0^1 \left[\int_0^1 (C(u, v))^2 c(u, v) dv \right] f_{U_{r-2:n-2}}(u) du + \int_0^1 \left[\int_0^1 (v - C(u, v))^2 c(u, v) dv \right] f_{U_{r:n-2}}(u) du + 2 \int_0^1 \left[\int_0^1 C(u, v)(v - C(u, v)) c(u, v) dv \right] f_{U_{r-1:n-2}}(u) du \right) \tag{9}$$

Given Proposition 4.1 and 4.2, the variance of $R_{r:n}$ can be easily derived. The proof of Proposition 4.1 and 4.2 are shown in the Appendix.

4.2 Application in Importance-Allocated Nested Simulation

As discussed in Section 3, the optimal proxy confidence level ξ in the IANS method is the highest ξ such that the $(1 - \alpha)J$ true tail scenarios are a subset of the $(1 - \xi)J$ proxy tail scenarios. This is equivalent to the highest ξ such that, among all J outer loop scenarios, the lowest rank of all concomitants (i.e. the proxy loss) of the largest $(1 - \alpha)J$ inner simulation losses is greater than ξJ . Since we could not find this optimal ξ without a full nested simulation of all J outer loop scenarios, we attempt to answer a similar question instead. That is, what is the highest value of ξ such that the rank of the concomitant (i.e. the inner simulation loss) of the $(\xi J + 1)$ th proxy loss is higher than αJ ? We propose Algorithm 3 to find this answer.

Algorithm 3 is an iterative process to search for the highest ξ in the IANS method so that within the bivariate sample with $(1 - \xi)J$ pairs of proxy and inner simulation losses, the upper bound of the confidence interval for the rank of concomitant of one of the best proxy losses is less than $(\alpha - \xi)J$. In other words, we are looking for the highest ξ such that we could expect with high confidence at least $(1 - \alpha)J$ inner simulation losses to be larger than the inner simulation losses in the bivariate sample.

In Line 8 of this algorithm, we consider the upper bound of the confidence interval for $R_{(\alpha - \xi_0)J:(1 - \xi)J}$, i.e. the rank of concomitant of the $(\alpha - \xi_0)J$ th proxy loss in the sample. Ideally, the most efficient way to find the optimal ξ would be by considering $R_{3:(1 - \xi)J}$, the rank of concomitant of *the best* proxy loss in the sample whose expected value and variance could be derived. However, we choose to consider the confidence interval for $R_{(\alpha - \xi_0)J:(1 - \xi)J}$ instead, because in the sample of $(1 - \xi)J$ pairs of proxy and inner simulation loss that we consider, the joint distribution of the quantile of the losses at the low end of the distribution is distorted, and is not indicative of the joint distribution of the same quantities among the entire J scenarios. This is observable in the output from a numerical experiment shown in Figure 2. Given we are ultimately concerned about the rank of concomitant among the entire J scenarios for the purpose of finding the optimal ξ , we use a more prudent although less efficient approach of examining $R_{(\alpha - \xi_0)J:(1 - \xi)J}$ instead. In Section 5, we show the difference in success rate of finding the optimal ξ by considering the confidence interval for $R_{(\alpha - \xi_0)J:(1 - \xi)J}$ versus $R_{3:(1 - \xi)J}$ in a numerical experiment.

Algorithm 3 provides an objective way to identify the optimal proxy confidence level ξ in the IANS method. It can be used to replace Stage II in Algorithm 2 as an improved IANS method. It is based on empirical simulation data so it can accommodate the unique feature and structure of different simulation models. Nevertheless, one drawback of this algorithm is that, due to its iterative nature, it is difficult to implement it under a fixed computation budget without any further strategy to allocate uneven number of inner loop simulations across outer loop scenarios. This will be considered in our future work.

5 Numerical Experiments

In Dang et al. (2018), we used extensive numerical experiments to demonstrate the performance of the IANS method. In this paper, however, we limit ourselves to the experiment of GMMB under a regime-switching asset model with dynamic lapse assumption for two reasons. First, this test case is more complex and presents less correlation between the proxy loss and inner simulation loss than other models considered in Dang et al. (2018), so it is more difficult to find the optimal proxy confidence level ξ , which makes it a good test case. Secondly, because Algorithm 3 uses empirical copula to construct the confidence interval for $R_{(\alpha - \xi_0)J:(1 - \xi)J}$, the algorithm should be effective regardless of the underlying asset and liability model used in simulation.

We start with full nested simulations to estimate the CTE 95 hedging loss of the GMMB contract. We use the parameters specified in Table 1.

A few simplifying assumptions are made, consistently with the development of the previous sections; specifically

- No transaction costs in the hedging program.

Algorithm 3: Proxy Confidence Level Identification in Importance-Allocated Nested Simulation.

input : $\xi = \xi_0$: Initial proxy confidence level for inner simulation in the IANS method
 α : Confidence level (e.g., $\alpha = 95\%$) at which CTE of losses is required.
 J : Number of replications in inner and outer simulations.
 $\omega = (\alpha - \xi_0)J$: Initial upper bound of the confidence interval for $R_{(\alpha-\xi_0)J:(1-\xi)J}$.
 $\widehat{L}_i, i = 1, \dots, J$: Proxy loss for all J outer loop scenarios
 Nested Simulation model for the VA contract of interest

output: CTE_α for the losses of Delta-hedging the VA contract of interest.

- 1 **while** $\omega \geq (\alpha - \xi)J$ **do**
- 2 $\xi \leftarrow 1 - \lceil \omega + (1 - \alpha)J \rceil \div J$
- 3 Identify a set \mathcal{K} of $(1 - \xi)J$ proxy tail scenarios with the largest proxy loss \widehat{L}_i from input
- 4 Implement the inner simulation step of Algorithm 1 for loss from hedging L with the original risk neutral asset model and VA payoff, but only for the set of outer scenarios \mathcal{K} identified in Step 3
- 5 Collect the $(1 - \xi)J$ pairs of bivariate output (\widehat{L}_k, L_k) where $k \in \mathcal{K}$ and convert each output to its marginal quantiles within the $(1 - \xi)J$ points. The converted sample is denoted as (U_k, V_k) , for $k = 1, \dots, (1 - \xi)J$
- 6 Estimate an empirical copula function of (U, V) based on (U_k, V_k) 's from Step 5, where

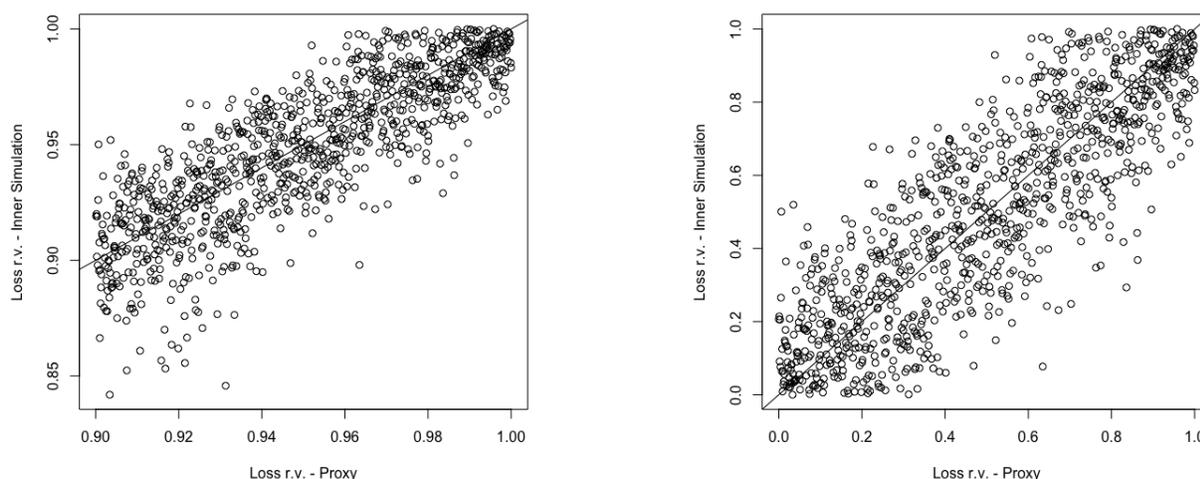
$$C(u, v) \approx \frac{1}{(1 - \xi)J} \sum_{k=1}^{(1-\xi)J} \mathbf{1}_{\{U \leq u, V \leq v\}}$$

(Hofert et al. 2017)
- 7 Estimate an empirical density function of the copula of (U, V) based on (U_k, V_k) 's from Step 5, where $b = \frac{1}{\sqrt{(1-\xi)J}}$ and

$$c(u, v) \approx \frac{1}{4b^2} \left(C(\min(u + b, 1), \min(v + b, 1)) - C(\min(u + b, 1), \max(v - b, 0)) \right. \\ \left. - C(\max(u - b, 0), \min(v + b, 1)) + C(\max(u - b, 0), \max(v - b, 0)) \right)$$
- 8 Calculate $E[R_{(\alpha-\xi_0)J:(1-\xi)J}]$ and $\text{Var}[R_{(\alpha-\xi_0)J:(1-\xi)J}]$ by substituting the empirical Copula and density function of (U, V) from Step 6 and 7 into Equation (8) and (9)
- 9 Construct a 95% one-sided confidence interval for $R_{(\alpha-\xi_0)J:(1-\xi)J}$ as $\left(-\infty, E[R_{(\alpha-\xi_0)J:(1-\xi)J}] + 1.645 \times \sqrt{\text{Var}[R_{(\alpha-\xi_0)J:(1-\xi)J}]} \right)$, and update $\omega = E[R_{(\alpha-\xi_0)J:(1-\xi)J}] + 1.645 \times \sqrt{\text{Var}[R_{(\alpha-\xi_0)J:(1-\xi)J}]}$
- 10 **end**
- 11 Identify the $(1 - \alpha)J$ largest liability values based on the inner simulations from Step 4.
- 12 Compute CTE_α in Equation (5) as the output.

Table 1: Parameters for VA Contracts

Description	Notation	Value
Maturity of Contract and Projection Period	T	240 months
Initial Fund Value	$F(0)$	\$1000
Initial Level of Guarantee	G_1	100% of $F(0)$



(a) QQ Plot of Tail Scenarios as Quantile of the Entire J Scenarios

(b) QQ Plot of Tail Scenarios as Quantile of the Tail Scenarios

Figure 2: Quantile of proxy and simulated losses in 1,000 tail scenarios. GMMB Contract with dynamic lapse under a regime-switching model of 10,000 outer scenarios. See Section 5 for more detail.

- 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.

Under these assumptions, the liabilities of the VA contracts consist only of the liability from the hedging program, i.e. the initial cost of the hedging portfolio and the present value of periodic hedging errors. In practice, fee income, expenses, commissions, decrements, and costs due to basis risk (difference between real world and hedging models) are likely to make up a proportion of the liability.

The model parameters of the regime-switching model are provided 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), 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 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:

$$q_{x+t} = \min \left(1, \max \left(0.5, 1 - 1.25 \times \left(\frac{G(t)}{F(t)} - 1.1 \right) \right) \right) \times q_{x+t}^{base} \quad (10)$$

where

$$q_{x+t}^{base} = \begin{cases} 0.00417 & \text{if } t < 84, \\ 0.00833 & \text{if } t \geq 84. \end{cases} \quad (11)$$

Table 2: Parameters for the real world Regime-Switching Model used in Section 5.

(Monthly rate)	Real World	Risk Neutral
Risk-free Rate: r	0.002	0.002
Mean - Regime 1 ($\rho = 1$): μ_1	0.0085	0.0013875
Mean - Regime 2 ($\rho = 2$): μ_2	-0.0200	-0.0012000
Standard Deviation - Regime 1: σ_1	0.035	0.035
Standard Deviation - Regime 2: σ_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 3: Model in the Full Nested Simulations in Section 5.2

Description	Value
Contract Type	GMMB
Asset Model	Regime-Switching
Lapse Modeling	Dynamic Lapse
Number of Outer Loop Simulation	$J = 5,000$
Number of Inner Loop Simulation	$N = 200$

5.1 IANS Method with $\xi = 10\%$

Figure 3 recaps the comparisons shown in Dang et al. (2018) between the losses that are simulated by true nested simulation and those by the IANS method's proxy simulation, for a GMMB contract under dynamic lapse assumptions.

The results suggest that the $(1 - \alpha)J$ tail scenarios from nested simulations overlap almost entirely with the $(1 - \xi)J$ proxy tail scenarios. Such overlapping suggest that the IANS method is effective in this realistic setting using only simple proxy calculations. Another important observation in Figure 3 is that the simulated losses by the nested simulation and those by the proxy simulation can be significantly different in values. Nonetheless, the rankings of these simulated losses remains similar so the proxy model can still effectively identify the true tail scenarios.

5.2 Finding the Optimal ξ Using Rank of Concomitant

To demonstrate the effectiveness of Algorithm 3, we conduct 100 experiments of CTE 95 estimate based on full nested simulation of the model described in Table 3. We also apply proxy evaluation in all 100 experiments and collect the actual proxy confidence level ξ in each experiment of full nested simulation. We then apply Algorithm 3 to each of the 100 experiments with an initial $\xi_0 = 0.92$ and find the optimal ξ suggested by the confidence interval of $R_{(\alpha - \xi_0)J : (1 - \xi)J}$, where $\alpha = 0.95$ and $J = 5,000$.

As illustrated in the Dang et al. (2018), the CTE 95 estimate in this experiment with only 5,000 outer loop scenarios and 200 inner loop simulations is typically biased. However, our focus in these experiments is not the accuracy of CTE 95 estimate. Instead, we are interested in how well Algorithm 3 can produce ξ 's that capture all the true tail scenarios based on a full nested simulation.

Out of the 100 experiments with $\xi_0 = 92\%$, Algorithm 3 produces proxy confidence level ξ 's that capture all the true tail scenarios in 94 of them. Since Algorithm 3 is based on a 95% confidence interval, we conclude a success rate of 94% is satisfactory. This is compared to a success rate of only 63% in capturing all the true tail scenarios if ξ 's are *fixed* at 90%. These results demonstrates the benefit of using Algorithm 3.

Even in the few experiments where the implied proxy confidence level ξ missed some true tail scenarios, the impact on the accuracy of the CTE estimate is less than 0.03% of the CTE estimate based on the full nested simulation. This is due to the fact for outer loop scenarios that are closer to the center of the distribution, the absolute values of losses are closer to each other despite their difference in rank. This can be observed in Figure 3.

The true proxy confidence level and the implied proxy confidence level based on Algorithm 3 in each of the 100 experiments are shown in Figure 4. The proxy confidence level is expressed in terms of rank of proxy loss among all J outer scenarios. An implied proxy confidence level lower than the true proxy confidence level indicates that the implied proxy confidence level captures all the true tail scenarios in the entire distribution in the experiment.

Figure 4 also shows that the implied proxy confidence level remains relatively stable regardless of the actual proxy confidence level. The reason behind is that the first two moments of $R_{(\alpha-\xi_0)J:(1-\xi)J}$ are affected by the rank of all $(1-\xi)J$ pairs of losses in each sample, which then affect the implied proxy confidence level. In contrast, the actual proxy confidence level can easily be skewed by the rank of concomitant in a single scenario because it depends on the *lowest* rank of proxy losses that correspond to the true tail scenarios. In addition, the underlying asset and liability model are the same across all 100 experiments. Therefore even though the actual proxy confidence level varies from one experiment to the next, the implied proxy confidence level remains relatively stable. Indeed, in cases where the actual proxy confidence level is high, the lower implied proxy confidence level means a “waste” in computation effort. Nonetheless, we argue that the “waste” in computation effort is again a trade-off for a more robust and accurate CTE estimate.

We conducted another set of experiment by applying Algorithm 3 to the same 100 experiments, but with an initial $\xi_0 = 0.9494$. This time none of the 100 experiments produces a proxy confidence level ξ that captures all the true tail scenarios, which demonstrates the necessity to start the experiment with a lower initial ξ_0 such as 92%.

6 Conclusion

In this article, we illustrate an improved Importance-Allocated Nested Simulation procedure in the application of estimating the CTE of liabilities in a VA dynamic hedging strategy. The algorithm for finding the optimal proxy confidence level is proposed based on theories from order statistics. It is a structured approach that is not restricted to the underlying VA contract or models. The numerical results show significant improvement in the accuracy of identifying the true tails scenarios in a nested simulation than the previous proposal of a fixed arbitrary ξ .

For future work, we will expand the improved IANS method to more sophisticated VA contract such as GMIB and GMWB, as well as nested simulation of other financial instruments.

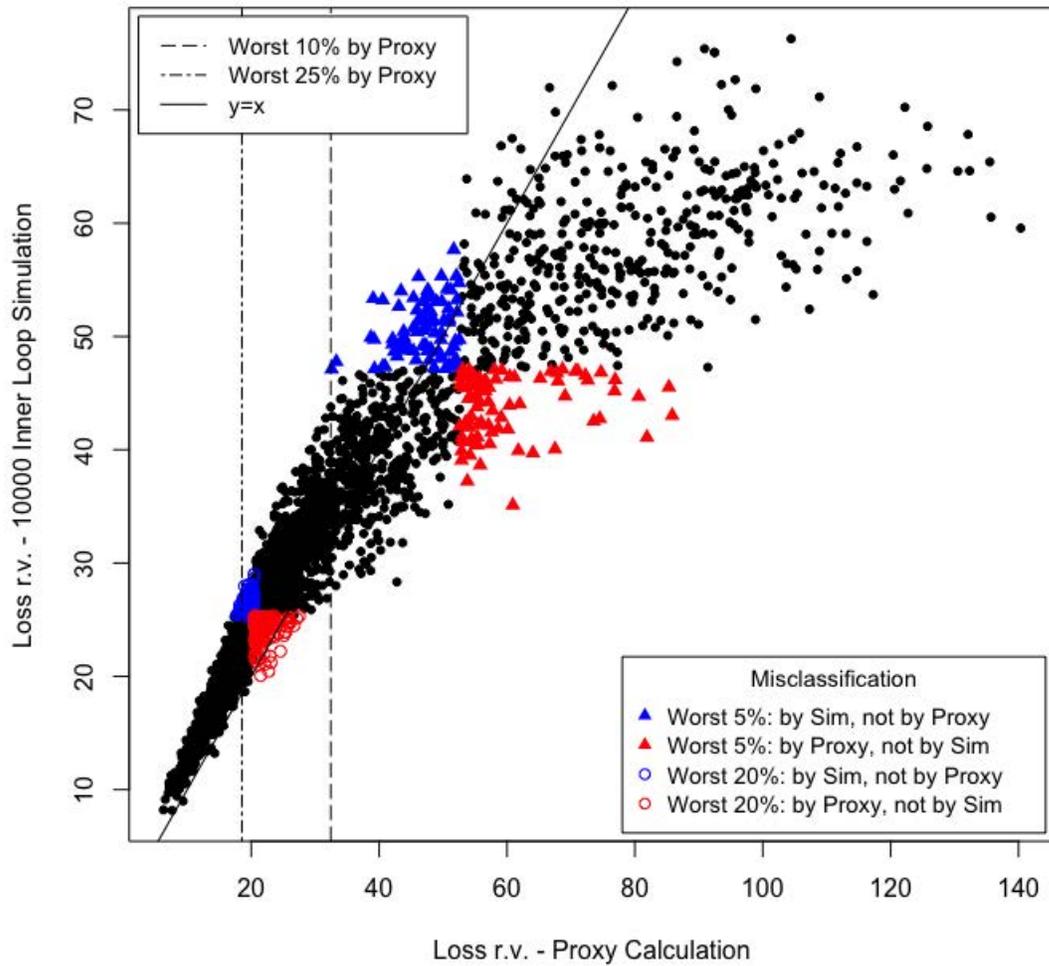


Figure 3: Simulated losses in 10,000 outer scenarios. The x and y coordinates of each point in the figures represent the loss in a scenario, simulated by the IANS proxy simulation and by the true nested simulation, respectively.

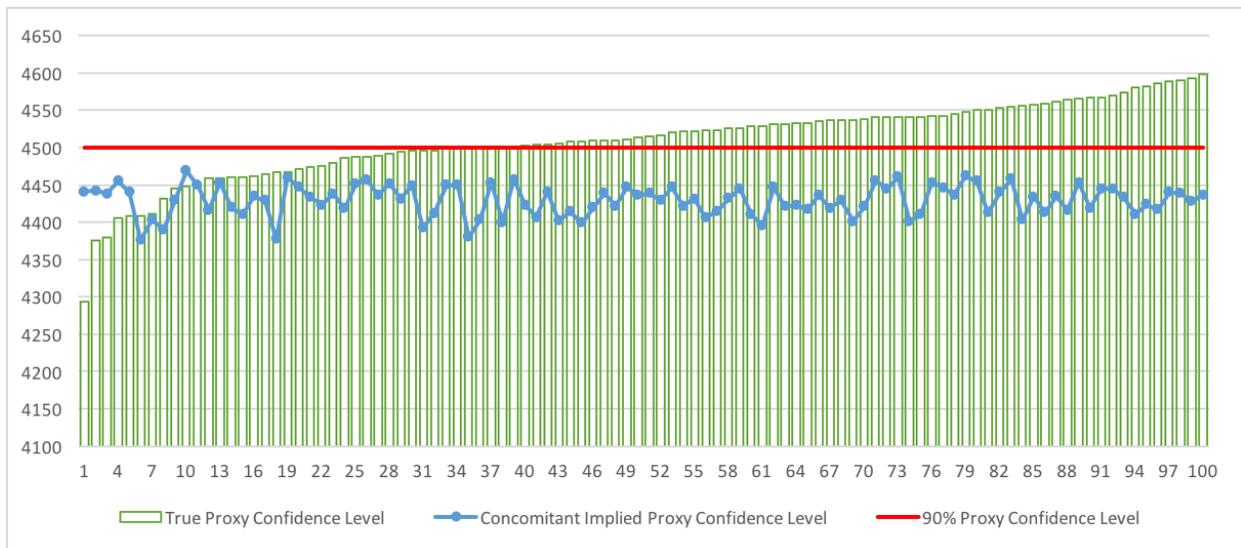


Figure 4: True and implied proxy confidence level in 100 repeated experiments. The output is sorted based on the level of the true proxy confidence level.

References

- Bauer, D., Reuss, A. & Singer, D. (2012), 'On the calculation of the solvency capital requirement based on nested simulations', *ASTIN Bulletin: The Journal of the IAA* **42**(2), 453–499.
- Bollen, N. P. (1998), 'Valuing options in regime-switching models', *The Journal of Derivatives* **6**(1), 38–49.
- Broadie, M., Du, Y. & Moallemi, C. C. (2011), 'Efficient risk estimation via nested sequential simulation', *Management Science* **57**(6), 1172–1194.
- Broadie, M., Du, Y. & Moallemi, C. C. (2015), 'Risk estimation via regression', *Operations Research* **63**(5), 1077–1097.
- Broadie, M. & Glasserman, P. (1996), 'Estimating security price derivatives using simulation', *Management science* **42**(2), 269–285.
- Cathcart, M. & Morrison, S. (2009), 'Variable annuity economic capital: the least-squares monte carlo approach', *Life & Pensions* pp. 36–40.
- CIA (2017), Standards of practice, part 2000 - insurance, Technical report, Canadian Institute of Actuaries.
URL: <http://www.cia-ica.ca/docs/default-source/standards/si020317e.pdf?sfvrsn=2>
- Dai, M., Kuen Kwok, Y. & Zong, J. (2008), 'Guaranteed minimum withdrawal benefit in variable annuities', *Mathematical Finance* **18**(4), 595–611.
- Dang, O., Feng, B. & Hardy, M. R. (2018), Efficient nested simulation for conditional tail expectation of variable annuities.
URL: <https://ssrn.com/abstract=3045430>
- David, H., O'Connell, M. & Yang, S. (1977), 'Distribution and expected value of the rank of a concomitant of an order statistic', *The Annals of Statistics* pp. 216–223.
- Feng, R. (2014), 'A comparative study of risk measures for guaranteed minimum maturity benefits by a pde method', *North American Actuarial Journal* **18**(4), 445–461.
- Feng, R., Cui, Z. & Li, P. (2016), Nested stochastic modeling for insurance companies, Technical report, Society of Actuaries.
URL: <https://www.soa.org/research/2015-nested-stochastic-modeling.pdf>
- Fu, M. C. et al. (2016), *Handbook of simulation optimization*, Springer.
- Gan, G. (2013), 'Application of data clustering and machine learning in variable annuity valuation', *Insurance: Mathematics and Economics* **53**(3), 795–801.
- Gan, G. (2015a), Application of metamodeling to the valuation of large variable annuity portfolios, in 'Proceedings of the 2015 Winter Simulation Conference', pp. 1103–1114.
- Gan, G. (2015b), A multi-asset monte carlo simulation model for the valuation of variable annuities, in 'Proceedings of the 2015 Winter Simulation Conference', pp. 3162–3163.
- Gan, G. & Lin, X. S. (2015), 'Valuation of large variable annuity portfolios under nested simulation: A functional data approach', *Insurance: Mathematics and Economics* **62**, 138–150.

- Gan, G. & Lin, X. S. (2017), 'Efficient greek calculation of variable annuity portfolios for dynamic hedging: A two-level metamodeling approach', *North American Actuarial Journal* **21**(2), 161–177.
- Glasserman, P. (2013), *Monte Carlo methods in financial engineering*, Vol. 53, Springer Science & Business Media.
- Gordy, M. B. & Juneja, S. (2010), 'Nested simulation in portfolio risk measurement', *Management Science* **56**(10), 1833–1848.
- Hardy, M. (2003), *Investment guarantees: modeling and risk management for equity-linked life insurance*, Vol. 215, John Wiley & Sons.
- Hardy, M. R. (2001), 'A regime-switching model of long-term stock returns', *North American Actuarial Journal* **5**(2), 41–53.
- Hofert, M., Kojadinovic, I., Maechler, M. & Yan, J. (2017), *copula: Multivariate Dependence with Copulas*. R package version 0.999-18.
URL: <https://CRAN.R-project.org/package=copula>
- Lan, H., Nelson, B. L. & Staum, J. (2010), 'A confidence interval procedure for expected shortfall risk measurement via two-level simulation', *Operations Research* **58**(5), 1481–1490.
- L'Ecuyer, P. (1990), 'A unified view of the ipa, sf, and lr gradient estimation techniques', *Management Science* **36**(11), 1364–1383.
- Liu, M. & Staum, J. (2010), 'Stochastic kriging for efficient nested simulation of expected shortfall', *The Journal of Risk* **12**(3), 3.
- Marshall, C., Hardy, M. R. & Saunders, D. (2010), 'Valuation of a guaranteed minimum income benefit', *North American Actuarial Journal* **14**(1), 38–58.
- O'Connell, M. J. (1974), 'Theory and applications of concomitants of order statistics'.
- OSFI (2017), Life insurance capital adequacy test, Technical report, Office of the Superintendent of Financial Institutes Canada.
URL: <http://www.osfi-bsif.gc.ca/Eng/Docs/LICAT18.pdf>

A Proof of Proposition 4.1 and 4.2

For a general bivariate distribution of (U, V) ,

$$E[R_{r:n}] = 1 + n \left(\int_{-\infty}^{-\infty} \left[\int_{-\infty}^{-\infty} \theta_1 f(v|u) dv \right] f_{U_{r-1:n-1}}(u) du + \int_{-\infty}^{-\infty} \left[\int_{-\infty}^{-\infty} \theta_3 f(v|u) dv \right] f_{U_{r:n-1}}(u) du \right) \tag{12}$$

(David et al. 1977)

where

- $\theta_1 = P[U < u, V < v]$
- $\theta_2 = P[U < u, V > v]$
- $\theta_3 = P[U > u, V < v]$
- $\theta_4 = P[U > u, V > v]$

In our problem where (U, V) follow a bivariate uniform distribution, we have

- $f(v|u) = \frac{f(u,v)}{f_U(u)} = f(u, v) = c(u, v)$
- $\theta_1 = C(u, v)$
- $\theta_2 = u - C(u, v)$
- $\theta_3 = v - C(u, v)$
- $\theta_4 = 1 - u - v + C(u, v)$

To recap the notations we use:

- $C(u, v)$ is the copula function of $U = u$ and $V = v$.
- $c(u, v)$ is the density function of $C(U, V)$.
- $f_{U_{r:n}}(u)$ represents the density function of the r th order statistics among n U 's. More specifically, $f_{U_{r:n}}(u) = \frac{n!}{(r-1)!(n-r)!} u^{r-1} (1-u)^{n-r}$.

Therefore, in our problem Equation (12) is equivalent to

$$E[R_{r:n}] = 1 + n \left(\int_0^1 \left[\int_0^1 C(u, v) c(u, v) dv \right] f_{U_{r-1:n-1}}(u) du + \int_0^1 \left[\int_0^1 (v - C(u, v)) c(u, v) dv \right] f_{U_{r:n-1}}(u) du \right) \tag{13}$$

To derive the second moment of $R_{r:n}$, we first derive the second moment of $R_{r:n}$ for a general case of bivariate distribution of (X, Y) , i.e. not specific to uniform distribution. We use the same factorial moment method as in O'Connell (1974).

First we have,

$$\begin{aligned}
 E[R_{r:n}^2] &= \sum_{s=1}^n s^2 P[R_{r:n} = s] \\
 &= \sum_{s=0}^{n-1} (s+1)^2 P[R_{r:n} = s+1] \\
 &= \sum_{s=0}^{n-1} P[R_{r:n} = s+1] + 2 \sum_{s=0}^{n-1} s P[R_{r:n} = s+1] + \sum_{s=0}^{n-1} s^2 P[R_{r:n} = s+1] \\
 &= 1 + 2(E[R_{r:n}] - 1) + \sum_{s=0}^{n-1} s^2 P[R_{r:n} = s+1]
 \end{aligned}
 \tag{14}$$

More specifically,

$$\sum_{s=0}^{n-1} s^2 P[R_{r:n} = s+1] = \sum_{s=0}^{n-1} s^2 n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^t C_k \theta_1^k \theta_2^{r-1-k} \theta_3^{s-k} \theta_4^{n-r-s+k} f(x, y) dx dy \tag{15}$$

where $C_k = \binom{n-1}{r-1} \binom{r-1}{k} \binom{n-r}{s-k}$.

Let $j = s - k$, then

$$\begin{aligned}
 &\sum_{s=0}^{n-1} s^2 P[R_{r:n} = s+1] \\
 = &\sum_{s=0}^{n-1} s^2 n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^t C_k \theta_1^k \theta_2^{r-1-k} \theta_3^{s-k} \theta_4^{n-r-s+k} f(x, y) dx dy \\
 = &n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} (k+j)^2 \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy \\
 = &n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} (k^2 + j^2 + 2kj) \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy
 \end{aligned}
 \tag{16}$$

Furthermore,

$$\begin{aligned}
 & n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} k^2 \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy & (17) \\
 = & n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} k(k-1) \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy \\
 & + n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} k \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy \\
 = & n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} k(k-1) \binom{r-1}{k} \theta_1^k \theta_2^{r-1-k} \sum_{j=0}^{n-r} \binom{n-r}{j} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy \\
 & + n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} k \binom{r-1}{k} \theta_1^k \theta_2^{r-1-k} \sum_{j=0}^{n-r} \binom{n-r}{j} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy \\
 = & n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \binom{n-1}{r-1} (r-1)(r-2) \theta_1^2 [F_X(x)]^{r-3} [1-F_X(x)]^{n-r} f(x, y) dx dy \\
 & + n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \binom{n-1}{r-1} (r-1) \theta_1 [F_X(x)]^{r-2} [1-F_X(x)]^{n-r} f(x, y) dx dy \\
 = & n(n-1) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta_1^2 \frac{f(x, y)}{f_X(x)} f_{r-2:n-2}(x) dx dy + n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta_1 \frac{f(x, y)}{f_X(x)} f_{r-1:n-1}(x) dx dy \\
 = & n(n-1) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1^2 f(y|x) dy \right] f_{r-2:n-2}(x) dx + n \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 f(y|x) dy \right] f_{r-1:n-1}(x) dx
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 & n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} j^2 \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy & (18) \\
 = & n(n-1) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3^2 f(y|x) dy \right] f_{r:n-2}(x) dx + n \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3 f(y|x) dy \right] f_{r:n-1}(x) dx
 \end{aligned}$$

And

$$\begin{aligned}
 & n \binom{n-1}{r-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{k=0}^{r-1} \sum_{j=0}^{n-r} kj \binom{r-1}{k} \binom{n-r}{j} \theta_1^k \theta_2^{r-1-k} \theta_3^j \theta_4^{n-r-j} f(x, y) dx dy & (19) \\
 = & n(n-1) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 \theta_3 f(y|x) dy \right] f_{r-1:n-2}(x) dx
 \end{aligned}$$

Substitute (17), (18) and (19) back in (16), we have

$$\begin{aligned}
 & \sum_{s=0}^{n-1} s^2 P[R_{r:n} = s + 1] \tag{20} \\
 = & n(n-1) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1^2 f(y|x) dy \right] f_{r-2:n-2}(x) dx + n \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 f(y|x) dy \right] dy f_{r-1:n-1}(x) dx \\
 & + n(n-1) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3^2 f(y|x) dy \right] f_{r:n-2}(x) dx + n \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3 f(y|x) dy \right] dy f_{r:n-1}(x) dx \\
 & + 2n(n-1) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 \theta_3 f(y|x) dy \right] f_{r-1:n-2}(x) dx \\
 = & n(n-1) \left(\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1^2 f(y|x) dy \right] f_{r-2:n-2}(x) dx + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3^2 f(y|x) dy \right] f_{r:n-2}(x) dx \right. \\
 & \left. + 2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 \theta_3 f(y|x) dy \right] f_{r-1:n-2}(x) dx \right) + E[R_{r:n}] - 1
 \end{aligned}$$

Substitute (20) back in (14), we have

$$\begin{aligned}
 E[R_{r:n}^2] = & 3E[R_{r:n}] - 2 + n(n-1) \times \left(\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1^2 f(y|x) dy \right] f_{r-2:n-2}(x) dx \right. \tag{21} \\
 & \left. + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3^2 f(y|x) dy \right] f_{r:n-2}(x) dx + 2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 \theta_3 f(y|x) dy \right] f_{r-1:n-2}(x) dx \right)
 \end{aligned}$$

In the case of bivariate uniform distribution of (U, V) , we have

$$\begin{aligned}
 E[R_{r:n}^2] = & 3E[R_{r:n}] - 2 + n(n-1) \times \left(\int_0^1 \left[\int_0^1 (C(u, v))^2 c(u, v) dv \right] f_{U_{r-2:n-2}}(u) du \right. \\
 & + \int_0^1 \left[\int_0^1 (v - C(u, v))^2 c(u, v) dv \right] f_{U_{r:n-2}}(u) du \\
 & \left. + 2 \int_0^1 \left[\int_0^1 C(u, v)(v - C(u, v)) c(u, v) dv \right] f_{U_{r-1:n-2}}(u) du \right)
 \end{aligned}$$