Efficient Nested Simulation of Tail Risk Measures for Variable Annuities

by

Ou Dang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Runhuan Feng Professor, Dept. of Mathematics University of Illinois at Urbana-Champaign
Supervisor(s):	Mingbin Feng Assistant Professor, Dept. of Statistics and Actuarial Science University of Waterloo
	Mary R. Hardy Professor, Dept. of Statistics and Actuarial Science University of Waterloo
Internal Member:	Paul Marriott Professor, Dept. of Statistics and Actuarial Science University of Waterloo
	Ruodu Wang Associate Professor, Dept. of Statistics and Actuarial Science University of Waterloo
Internal-External Member:	Justin Wan Professor, David R. Cheriton School of Computer Science University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Estimating tail risk measures for portfolios of complex Variable Annuities (VA) is an important enterprise risk management task which usually requires nested simulation. In the nested simulation, the outer simulation stage involves projecting scenarios of key risk factors under the real world measure, while the inner stage is used to value payoffs under guarantees of varying complexity, under a risk neutral measure.

In this thesis we propose and analyze three different two-stage simulation procedures that improve the computation efficiency of nested simulation. All three proposals allocate the inner simulations to the specific outer scenarios that are most likely to generate larger losses. These scenarios are identified using a proxy evaluation in the Stage 1 simulation. The proxy evaluation is used only to rank the outer scenarios, not to estimate the tail risk measure directly. The proxy evaluation can be based on a closed-form calculation which works very efficiently for simpler contracts, or a pilot nested simulation using likelihood ratio estimators which accommodates very complex path-dependent contracts. Then in the Stage 2 simulation we allocate the remaining computational budget to the scenarios identified in Stage 1. Our numerical experiments show that, in the VA context, our proposals are significantly more efficient than a standard Monte Carlo experiment, measured by relative mean squared errors (RMSE), when both are given the same computational budget.

Acknowledgements

I would like to express my greatest appreciation to my thesis supervisors, Professor Ben Feng and Professor Mary Hardy, who patiently guided me through the long journey toward this degree, and made it an enjoyable and rewarding experience for me. I thank them for their expert advice, their care for my career and well-being, and for always being available to help me.

I would also like to thank all the examining committee members, Professor Runhuan Feng, Paul Marriott, Justin Wan and Ruodu Wang, for their valuable time and their insightful critiques and suggestions. They made this thesis a better work, and pointed to some interesting directions for future research.

I am grateful for the encouragement from Professor Chris Groendyke at Robert Morris University and my now supervisor Professor Mary Hardy at the start of this endeavour. Thanks to their support for my "crazy" idea of pursuing an academic career ten years after graduating from university, I have had a fun and fruitful experience at graduate school.

Many thanks to the faculty and staff in the Department of Statistics and Actuarial Science, for all their help and advice during the past six years.

I also want to thank my friend Lin and Lulu for helping me get through the most difficult period of this Ph.D.

I am indebted to my parents for all their love and care, and for their unreserved support for my education. I am also thankful for the love and support from my children, Jasper and Stanley, particularly during "mommy's important meetings" in this past year of pandemic.

I dedicate this thesis to my husband, Toby, who provided tremendous amount of emotional support, and took on many family responsibilities throughout my Ph.D. journey. Without him, this thesis would not have existed.

Table of Contents

List of Tables

List of Figures

Abbreviations

CTE Conditional Tail Expectation

DIANS Dynamic Importance-Allocated Nested Simulation

GBM Geometric Brownian Motion

GLWB Guaranteed Lifetime Withdrawal Benefit

GMAB Guaranteed Minimum Accumulation Benefit

GMDB Guaranteed Minimum Death Benefit

GMIB Guaranteed Minimum Income Benefit

GMMB Guaranteed Minimum Maturity Benefit

GMWB Guaranteed Minimum Withdrawal Benefit

IANS Importance-Allocated Nested Simulation

IPA Infinite Perturbation Analysis

LSMC Least-squares Monte Carlo

MLR Mixture Likelihood Ratio

ORSA Own Risk Solvency Assessment

P&L Profit-and-Loss

RMSE Relative Mean Squared Errors

RSLN Regime-switching Log-normal

SMC Standard Monte Carlo

VA Variable Annuity

VaR Value at Risk

List of Symbols

B_t Amount of risk free zero coupon bond maturing at time T that is held in hedging portolio at time t

$C_{\mathbf{LR}}$ Number of calculations required to calculate one likelihood ratio adjusted inner sample

$C_{\mathbf{inner}}$ Number of calculations required at each time step in one inner simulation sample

D_t Present value at time 0 of \$1 payable at time t, discounted at risk-free rate

F_t Sub-account value at time t

G_t Guarantee value at time t

HE_t Hedging error incurred at time t

H_t^{bf} Value at time t of the hedging portolio brought forward from time $t-1$

H_t Value of the hedging portolio at time t

I_t Guaranteed minimum withdrawal benefit amount at time t

L Loss random variable of a VA contract

M Number of outer scenarios in a nested simulation

N Number of inner simulations in a nested simulation

R_t Log-return of stock price between time t and $t+1$

S_t Price of the underlying stock index at time t

T^M Time to maturity of the VA contract

$\mathcal{T}_m$ The subset containing the scenarios associated with the m largest losses

T Time to expiration of the VA contract, $T = \min(T^M, \tau)$

V_t The unhedged liability of a VA contract as of time t

X_t Outer scenario state variables at time t

$X_t^{(i)}$ Sample value of X_t in the ith outer scenario

$\boldsymbol{X}^{(i)}$ The ith outer scenario

$\boldsymbol{Y}_t$ Entire path of inner simulation state variables, conditional on X_t

$Y_{t,t'}$ Inner simulation state variable at time t', conditional on X_t

$Y_{t,t'}^{(i,j)}$ Sample value at time t' on the jth inner simulation sample path, conditional on $X_t^{(i)}$

$\boldsymbol{Y}_t^{(i,j)}$ The j-th inner sample path, conditional on $X_t^{(i)}$

$Y_{t,t'}^{(i)}$ Random variable $Y_{t,t'}$, conditional on $X_t^{(i)}$

$\boldsymbol{Y}_t^{(i)}$ Entire paths of random vector $Y_{t,t'}^{(i)}$, $t' = t, \ldots, T$

$\Delta_t^{(i)}$ Number of units of the underlying stock held in hedging portfolio at time t in ith outer scenario

Δ_t Number of units of the underlying stock held in hedging portfolio at time t

Γ Total number of inner simulation replications in a simulation procedure

$g_t^{(i)}(y_t)$ Probability density function of y_t, conditional on $X_t^{(i)}$

η^g Gross rate of management fee deduction

η^n Net rate of management fee income

γ Guaranteed minimum withdrawal rate every time period in a GMWB contract

$\mu_t(X_t)$ Pricing or valuation functional evaluated at time t given X_t

τ Time to death of the policyholder

Chapter 1

Introduction

1.1 Variable Annuities

Variable Annuities (VA) are long-term insurance contracts that are widely used for wealth accumulation and for providing retirement income. Annual sales of VAs in the U.S. in 2019 were $102 billion, and assets under management of inforce VA contracts at the end of 2019 totaled around $2.2 trillion (Insurance Information Institute, 2020).

A VA contract works in many ways like a mutual fund investment with an embedded option. The premium of the contract, which, typically, is a lump sum at the start of the contract, is deposited into a mutual-fund like vehicle called the sub-account. The sub-account is invested in financial assets such as bonds and stocks. The insurer periodically deducts fees from the sub-account, and in exchange offers death and/or living benefits to the policyholder. The VA benefits are linked to the value of underlying investment in the sub-account, subject to a guaranteed minimum. The guaranteed minimum benefits protect policyholders from downside market risk of their investment.

1.1.1 Types of VA guarantees

The major types of benefits provided by Variable Annuities in the market today include:

Guaranteed Minimum Maturity Benefit (GMMB) In the simplest GMMB, the contract pays a maturity benefit equal to the greater of the sub-account value and a fixed guarantee value. In many cases, the guarantee value is set as a percentage, say 75% or 100%, of the initial premium.

Guaranteed Minimum Death Benefit (GMDB) A GMDB contract is similar to a GMMB contract with the only difference being the timing of payoff. The payoff of a GMDB contract occurs at the time of policyholder's death rather than at maturity of the contract. GMDBs are sometimes offered as a standalone VA contract, but more often, they are offered together with other guaranteed benefits in a VA contract.

Guaranteed Minimum Accumulation Benefit (GMAB) In a GMAB contract, renewals are scheduled during the term of the policy. At each renewal, if the sub-account value exceeds the guarantee value, then the guarantee value is reset to the level of fund value. On the other hand, if the guarantee value exceeds the fund value at renewal, then the insurer will pay to increase the fund value to that of the guarantee value. The contract also has a maturity benefit equal to the greater of the sub-account value and the guarantee value. Compared to a GMMB contract, the GMAB contract gives the policyholder an opportunity to lock-in investment gain, and requires the insurer to make up for some investment loss during the course of the contract. As such, the guaranteed benefit in GMAB is richer than the guaranteed benefit in GMMB.

Guaranteed Minimum Withdrawal Benefit (GMWB) A GMWB contract guarantees the minimum amount of periodic withdrawal the policyholder can take from the sub-account until maturity. The insurer guarantees to pay the minimum withdrawal benefit each year even if the sub-account value reduces to zero. The minimum withdrawal benefit is typically a fixed percentage of the guarantee value. The guarantee value will decrease accordingly if the withdrawal exceeds the guaranteed minimum. The GMWB is typically offered with a flexible or fixed accumulation period. During the accumulation period, no withdrawals are made. A GMDB benefit is usually offered during the accumulation period. Additional features offered with the GMWB include roll-up, ratchet, and reset (Geneva Association, 2013). With roll-up, the guarantee value grows at a pre-determined rate. Typically, the roll-up feature is only offered during the accumulation period. With ratchet, the guarantee value is periodically set to the sub-account value if the sub-account value becomes higher similarly to the GMAB. With reset, the guarantee value can be reset to the higher of the current and the original sub-account value. The guaranteed minimum periodic withdrawal feature makes the GMWB contract a useful tool for providing retirement income.

Guaranteed Lifetime Withdrawal Benefit (GLWB) A GLWB contract is similar to a GMWB contract, with the only difference being the length of the payoff period. The payoff of a GLWB contract ends at the time of policyholder's death rather than at maturity of the contract. Compared to the GMWB contract, the GLWB also protects the policyholder from longevity risk.

Guaranteed Minimum Income Benefit (GMIB) A GMIB contract is similar to a GLWB contract. Instead of providing a guaranteed minimum withdrawal benefit for life, GMIB provides a guaranteed income benefit for life through annuitization. After annuitization, the contract becomes a fixed payout annuity. The policyholder cannot withdraw any funds exceeding the annuity benefit, nor can they access any remaining sub-account value at the time of death. Therefore, compared to a GLWB contract, the GMIB contract restricts liquidity significantly from policyholder's perspective, but offers higher income on average.

According to Geneva Association (2013), by the third quarter of 2012, among all VA sales in the U.S., GMWBs and GLWBs accounted for 41%, GMIBs accounted for 16%, GMABs accounted for 3% while non-living benefit accounted for 35%.

1.1.2 Pricing and valuation of Variable Annuities

From the insurers' perspective, the various guaranteed benefits listed above can be viewed as financial options embedded in these VA contracts. Early work on pricing and valuation of VA includes Boyle and Schwartz (1977); Boyle and Hardy (1997, 2003). Boyle and Schwartz (1977) pioneer the theoretical framework for pricing death and maturity benefit guarantees in equity-linked life insurance contracts as financial options. They also present how a hedging strategy can be constructed by insurers to minimize investment risk exposure in these equity-linked contracts. Boyle and Hardy (1997) present two approaches to reserve for maturity benefit guarantees in VA (also known as Segregated Funds in Canada): a stochastic simulation approach and an option pricing approach. Under the stochastic simulation approach, the distribution of portfolio values and the value of guaranteed maturity benefit is projected using Monte Carlo simulation. Then the reserve is set at an level such that with high probability, e.g. 99%, the reserve will be sufficient to meet the obligation of maturity benefit guarantee. Under the option pricing approach, the maturity benefit guarantee is viewed as a put option, and the underlying investment risk can be hedged accordingly. The reserve is then set as the sum of the price of put option plus the transaction cost associated with hedging. In Boyle and Hardy (2003), the authors present a pricing model for Guaranteed Annuity Options, which is similar to a GMIB VA contract. The pricing model incorporates interest rate risk, equity risk, and mortality risk of the guaranteed benefits.

Bauer *et al.* (2008) propose a universal pricing framework for various guaranteed benefits, including GMDB, GMAB, GMWB and GMIB. The proposed pricing framework can accommodate modeling of stochastic policyholder behavior. Marshall *et al.* (2010) focus on

the pricing of GMIBs. They extend the work of Bauer *et al.* (2008) by including stochastic interest rate model, reflecting the practical fee structure, and separately quantifying the fair price of different product features such as roll-up or ratchet. They conclude that roll-up is an expensive feature while ratchet is not. They also show interest rate and lapse rate has significant impact on the fair price of the GMIB. Dai *et al.* (2008) develop a stochastic control model for pricing GMWBs and consider policyholders' optimal withdrawal strategy to maximize the expected discounted cashflow from the contract. They suggest that the flexibility for policyholders to dynamically withdraw in a GMWB contract is an expensive option offered by the insurer. They also discussed several optimal withdrawal strategies derived from the model. Piscopo and Haberman (2011) introduce a valuation model for GLWBs which incorporates a stochastic mortality model. They show that the fair value of the GLWB is sensitive to mortality risk as well as model and parameter selection.

1.1.3 Risk factors of Variable Annuities

As alluded to in the works listed above, the various guaranteed minimum benefits pose significant risks to insurance companies issuing VAs. The most prominent risk factors include equity market performance, interest rate, policyholder behavior, and longevity experience. Fung *et al.* (2014) generalizes the valuation framework for GMWB presented in Kolkiewicz and Liu (2012) to the GLWB under static withdrawal assumption. They integrate a stochastic mortality model in the valuation framework to reflect longevity risk. The sensitivity of the fair price, risk measure and profit-and-loss (P&L) of the GLWB contract are examined. The authors find that the fair market price of the GLWB increases significantly with increasing mortality volatility, longevity risk, and equity exposure. The fair market price is negatively correlated to interest rates and is highly sensitive in a low interest rate environment. They also show that tail losses of GLWBs increase when longevity risk is present. In addition, the loss distribution is sensitive to parameter risk and model risk.

Geneva Association (2013) also recognizes the key risks associated with VAs as longevity risk, equity risk, interest rate risk, persistency risk, and benefit utilization risk. Drexler *et al.* (2017); Chahboun and Hoover (2019) echo these findings, and identify hedge risk as another key risk of VA. We will discuss this next. Drexler *et al.* (2017) also point out low interest rates exacerbate the risks associated with VAs.

The risk associated with policyholder behavior refers to the deviation of assumed and actual policyholder behavior, such as lapse/surrender of the contract, the amount of withdrawal benefit made (Steinorth and Mitchell, 2015; Moenig and Bauer, 2016), the length of

the accumulation period (Huang *et al.*, 2014), etc. A unique feature of VAs in terms of policyholder behavior, is policyholders' propensity to lapse the contract when the guarantee value is substantially below the sub-account value (i.e. when the contract is out-of-the-money), in order to obtain a new policy with a higher guarantee value. Conversely, when the guarantee value is above the sub-account value (i.e. when the contract is in-the-money), policyholders tend to be more reluctant to lapse the contract. The impact of this feature of VAs is that the policyholder's behavior is dependent on financial market performance. This is a well studied topic in VA research. See, for example, Bernard *et al.* (2014b); Knoller *et al.* (2016); Moenig and Zhu (2018); Bernard *et al.* (2014a).

1.1.4 Dynamic hedging of Variable Annuities

Given the significant market risk exposure of VA, insurance companies very often use dynamic hedging to mitigate their risks. Dynamic hedging uses a portfolio of assets whose value moves in tandem to the direction of the value of liability associated with the guaranteed minimum benefits. As a result, the gain or loss from the hedging portfolio is expected to offset, at least to some extent, the liability under the guaranteed minimum benefit. The hedging portfolio is rebalanced frequently due to changes in the liability and hedging portfolio component values. The composition of hedge portfolio is determined based on the "Greeks" of the liability. Typical Greeks used in a dynamic hedging program of VAs include:

Delta Sensitivity of the liability associated with guaranteed minimum benefit to changes in the price of underlying asset, e.g. equity investment in the sub-account.

Gamma Sensitivity of the Delta to changes in the price of underlying asset.

Rho Sensitivity of the liability associated with guaranteed minimum benefit to changes in interest rate.

Vega Sensitivity of the liability associated with guaranteed minimum benefit to changes in the market implied volatility of underlying asset.

1.1.5 Hedging effectiveness of Variable Annuities

Under the ideal conditions of the Black-Scholes-Merton world, the dynamic hedging strategy would be self financing, with no requirement for additional economic or regulatory

capital. In practice, there will always be gaps between the evolution of the hedge portfolio and the hedge position required for the liability, which makes slippages in hedge portfolios inevitable. At each rebalancing point of the hedging portfolio, the value of the hedging portfolio brought forward from the previous period may be different from value of the hedging portfolio required for the subsequent period. Hence, 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. Hardy (2003) discusses in detail dynamic hedging of Variable Annuities.

The study done by Sun *et al.* (2016) found that "hedging programs are 92% effective in reducing P&L volatility and 96% effective in offsetting losses from market movements". Causes for slippage in hedge portfolio effectiveness include basis risk, discrete hedge rebalancing intervals, and the effects of policyholder behavior.

Basis risk refers to the risk that the price movements of underlying investment in the sub-account of VA contracts do not exactly follow the price movement of the hedging instrument. Trottier *et al.* (2018) and Ankirchner *et al.* (2014) examine basis risk and liquidity risk in VA hedging. Liquidity risk refers to the risk that insurers need to borrow to finance the set up of VA hedging portfolio, given the premium structure of VAs (periodic rather than upfront fee deduction). They propose a numerical framework for quantifying hedging errors in presence of basis and liquidity risk. They show that, based on their model, even with a high correlation of 0.99 between the risk source and the hedging instrument, 14% of the volatility in price movement still cannot be hedged.

In Kling *et al.* (2011), the authors study various features of GLWB benefits and impact of different hedging strategies on an insurer's P&L. They find that different product features have vastly different impact on the composition of hedging portfolio and the insurers risk exposure after hedging. They also show that the insurer's risk exposure could vary significantly in presence of model risk and stochastic volatility.

Kling *et al.* (2014) extend the work of Kling *et al.* (2011) and incorporate policyholder behavior in their valuation framework for GLWB. The policyholder behavior they consider is the policyholder's decision to either withdraw the guaranteed minimum withdrawal amount, or to withdraw the entire sub-account value, i.e. a full surrender of the VA contract. They use this model to study the fair price of different product features in presence of policyholder behavior. They also study the impact of policyholder behavior on hedging efficiency, insurer's expected profit and risk exposure. They find that deviation between assumed and actual policyholder behavior has a material impact on insurers' expected profit and risk exposure after hedging, particularly for product without a ratchet feature. Augustyniak and Boudreault (2015) use a GMMB example to study the impact

on hedging error under a few scenarios including (1) not reflecting policyholder behavior in hedging; (2) reflecting the policyholder behavior perfectly in hedging; (3) reflecting the policyholder behavior in hedging with some imperfection. They found significant impact on hedging error, especially in the tail risk measure when policyholder behavior is not reflected in hedging. Reflecting the policyholder behavior in hedging, even with some imperfection, provides sizable relief to hedging error.

1.2 Risk Measures

Quantifying tail risk is an important task in risk management of insurance and other financial industries. Two commonly used tail risk measures are Value-at-Risk (VaR) and Conditional Tail Expectation (CTE[1]).

For a random loss L, the α-VaR, where $0 \leq \alpha \leq 1$, is defined as

$$\text{VaR}_\alpha = \inf \{Q : Pr[L \leq Q] \geq \alpha\}$$

Let Q_v denote the v-VaR of L, then the CTE_α is defined as

$$\text{CTE}_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 Q_v dv$$

In the above definitions, and for the rest of the thesis, we assume that losses lie in the right tail of the distribution whereas gains lie in the left tail of the distribution.

Tail risk measures such as VaR and CTE are widely used for setting regulatory and economic capital. Regulatory capital is the amount of capital each institution must hold as prescribed by regulator, whereas economic capital is the amount of capital that the institution judges is necessary to cover its risk (Tiesset and Troussard, 2005).

In the insurance industry Solvency II EIOPA (2014) is the regulatory capital requirement applicable to insurers in the European Union. Solvency II requires evaluating the 99.5% Value at Risk (VaR) of the change in surplus each year. In Canada, the regulatory capital requirement for Variable Annuities, as prescribed in OSFI (2017), is the difference

[1]The CTE_α, which was introduced in Wirch and Hardy (1999), is identical to the TailVaR_α or Expected $\text{Shortfall}_{1-\alpha}$, 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.

between the Total Gross Calculated Requirement, which is set at the 95% CTE, and the reserve, which is set between the 60% and 80% CTEs (CIA, 2017). In the US, the stochastic component of the reserve of VAs uses a 70% CTE (NAIC, 2020) and the Total Asset Requirement uses a 98% CTE (NAIC, 2016).

In the banking industry, under Basel III (Basel Committee on Banking Supervision, 2019), the regulatory capital requirement for market risk is the 97.5% Expected Shortfall over a 10-day horizon under the internal models approach, replacing the 99% VaR required under Basel II.

1.3 Nested Simulation

Estimating the tail risk measures of VA liabilities, including the hedging error, is of prime interest to risk management and regulatory capital assessment. In most cases the evaluation of these risk measures is computationally burdensome, requiring nested, path-dependent Monte Carlo simulation (also known as two-tier or stochastic-on-stochastic simulation). Therefore, developing more efficient and accurate methods for the valuation and risk management of embedded options is a topic of considerable interest to insurers , and has applications more broadly in financial risk management. In Feng (2016), a Society of Actuaries' survey participated by 18 insurance companies with over $100 billion in assets combined, none of the participants adopted nested stochastic modeling at the time the survey was conducted. The majority of the participants identified run time and difficulty of modeling as the primary reasons for not doing so, but most expressed a desire to implement it in the future.

The nested Monte Carlo simulation process for assessing the tail risk measure of a dynamically hedged VA requires two levels of simulation.

- The outer level simulation projects the underlying risk factors to quantify the loss distribution under the real world measure. In many finance applications, the projection involves only a single step, but in our context the outer level projections are multi-period, e.g. 20 or 40 years, with a time step based on the assumed hedge rebalancing frequency, e.g. monthly. The outer level simulated paths are known as the *scenarios*, and may include simulated asset returns, guarantee value, interest rates, policyholder behaviour, and longevity experience.

- The inner level simulations are used to estimate the fair market value of the guaranteed liability, i.e. the embedded options of VA contracts, at each future date,

conditional on the outer level scenarios up to each valuation date. In other words, the inner level simulation target is an expectation, conditional on the outer scenario state variables. As this is a valuation step, a risk neutral pricing measure is used.

In Figure 1.1 we illustrate the nested simulation process both for the single period and multi-period case. The entire figure represents the multi-period case considered in this thesis while the portion circled in the green represents the single period case considered in most other research into this subject, including Gordy and Juneja (2010); Lan *et al.* (2010); Broadie *et al.* (2011, 2015); Bauer *et al.* (2012).

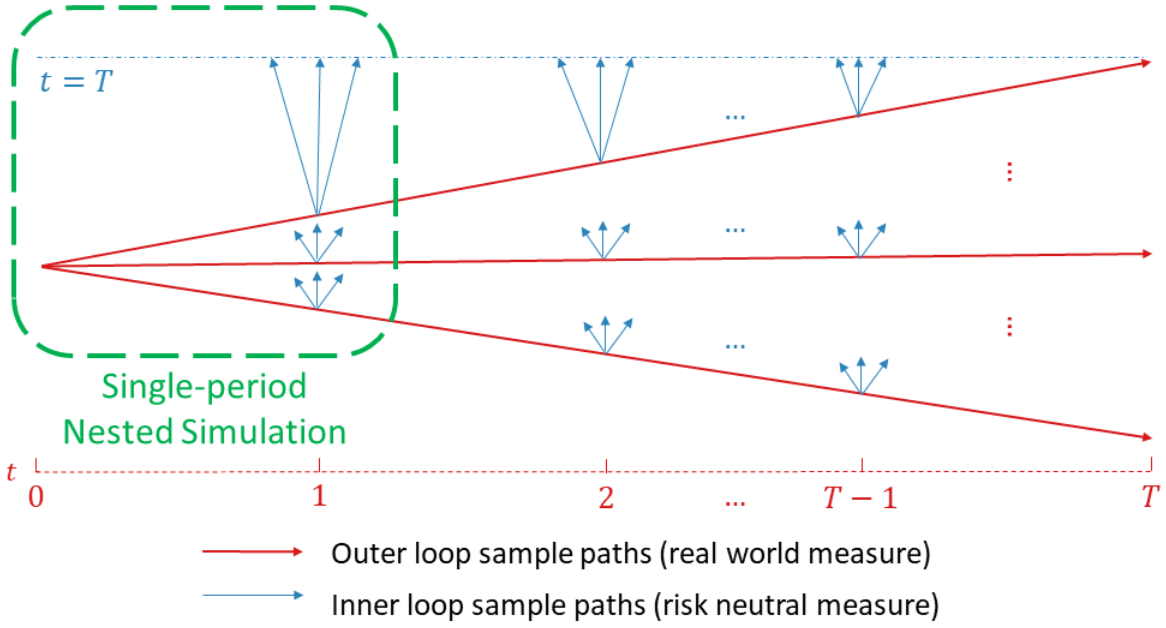


Figure 1.1: Nested simulation structure

One major reason that nested simulations are required is that the hedge calculations of VA generally use stochastic volatility models, because of the long term nature of the contracts, 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 only be solved using

stochastic simulation, and it is now standard practice in the insurance industry to use Monte Carlo simulation to model Variable Annuities.

For insurers, large scale nested simulations for assessing VA losses will take considerable run time. A large number of outer level simulations are needed in order to estimate the extreme tails needed for VaR or CTE calculations, and a large number of inner simulations are needed at each time step, because the embedded options are often far out-of-the-money. Furthermore, the calculations have to be repeated for each VA contract, or cluster of contracts, in force. Consider a single VA contract with 20-year maturity, which is dynamically hedged with rebalancing at monthly intervals. 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 each simulated value takes 1 μs (10^{-6} seconds) to complete, then it would take around 20 minutes to simulate the cash flows for a single policy. 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. Given the computational burden, many insurers are very interested in variance reduction techniques for nested simulation models that can achieve accurate results within a limited computational budget (Feng, 2016). Feng (2018) offers a detailed description of nested simulation of VA and some existing techniques to address the computational challenge of nested simulation. We will detail a few of these techniques in Section 1.4.

Similar nested simulation challenges arise in banking, where exotic options and intractable pricing measures make the assessment of the VaR or CTE measures too complex for analytic calculation. Holton (2003) points out that large scale nested simulation is too time consuming for practical, everyday risk analysis, where risk exposures may be required with only a few minutes notice. Compromises, such as limiting the choice of pricing models, or prematurely terminating simulations, are usually insufficiently accurate for tail risks, and can produce unacceptably biased estimators.

1.4 Literature Review

Research aimed at improving the computational efficiency of nested simulations mainly addresses the issue from two different angles: (1) by considering nested simulation of each model contract, or (2) by considering large portfolios of contracts. Proposals in the first category typically involve a modified simulation procedure than a standard nested simulation. Proposals in the second category typically retain the standard nested simulation

structure but try to reduce the number of contracts in the portfolio to be processed through nested simulation. Our proposed work fits in the first category.

The literature on nested simulations of model contracts focuses either on an efficient allocation of computational budget between outer and inner simulations (e.g. Gordy and Juneja (2010)), or on methods to improve the efficiency of inner simulations. In Gordy and Juneja (2010), the authors demonstrate how a fixed computation budget can be allocated between inner and outer simulations to minimize the mean squared error of various tail risk measures. They derive convergence analysis results for the mean squared errors of the tail risk measures. In addition, they illustrate how the jackknife method can be applied to reduce bias in the estimated tail risk measure. Their methods are useful in nested simulations where the number of outer scenarios are not pre-determined and can vary from one experiment to another.

On the other hand, among the work that aims at improving the efficiency of inner simulations, two different approaches have been proposed. The first uses proxy models to replace the inner simulation step, and the second uses a dynamic, non-uniform allocation of the inner simulation budget. Some work use a combination of the two approaches.

Proxy models are tractable analytic functions that replace the inner simulation stage of a nested simulation. Proxies may be empirical – i.e. intrinsic to the simulation process, or may be extrinsic. Extrinsic proxy functions are selected to be close to the inner simulation values, and therefore require detailed information about the payoffs that are evaluated in the inner simulation step. See Aggarwal *et al.* (2016) for other examples.

Empirical proxies are constructed using an initial pilot simulation to develop factors or functionals that can subsequently be used in place of the inner simulation. Hardy and Wirch (2004) use a factor approach. They propose a linear interpolation for calculating risk measures over multiple periods in the time horizon for profit testing purposes. A grid of risk measures factors are pre-calculated using standard nested simulation based on different term-to-maturity and moneyness ratio. Then the risk measure at future times are projected as linear interpolation of the pre-calculated risk measures. This approach is useful in simpler contracts where the relationship between risk measures and moneyness ratio is close to being linear.

Another prominent empirical proxy method discussed in research is the least squares Monte Carlo (LSMC) method (following Longstaff and Schwartz (2001)), for example in Cathcart and Morrison (2009); Bauer and Ha (2015); Krah *et al.* (2018); Broadie *et al.* (2015). The work in Cathcart and Morrison (2009) is motivated by the nested simulation required for risk based capital requirement of VAs such as Solvency II. They demonstrate how the LSMC method can be used to replace inner simulations. They show that with

simple input variables, such as the sub-account value, interest rate, and guarantee value, the LSMC method can estimate liabilities accurately for complex VA contracts such as GMWBs. Bauer and Ha (2015) apply the LSMC method in a nested simulation of GMIBs. They study the convergence of the LSMC algorithm, particularly for estimating VaR. They show that the left singular functions of the valuation operator present the optimal basis functions in the LSMC method. The authors point out that the choice of basis function could significantly affect the accuracy of results, particularly when the complexity and dimensionality of the problem increase. Complexity and high dimensionality are indeed features of the multi-period nested simulation that we consider in this thesis. Thus we are exploring other options for improving the efficiency of multi-period nested simulation in this thesis. Krah *et al.* (2018) consider the LSMC method in a nested simulation for quantifying the Statutory Capital Requirement under Solvency II for life insurance contracts. They present an adaptive algorithm to build up basis functions. They also discuss processes for validating the proxy built with the basis functions. Broadie *et al.* (2015) apply the LSMC method in a nested simulation to quantify risk measures in a portfolio of financial assets whose prices are driven by multiple risk factors such as stock and commodity prices, interest rates, currency exchange rates, etc. They analyze the convergence of MSE of the risk measure estimator based on LSMC method, and show that the performance of the estimator does not depend on the dimension of the problem *if* basis functions are well-chosen. They also suggest that expert knowledge about the problem is helpful in choosing basis functions. In addition, the authors present a weighted LSMC method for tail risk measure estimation that can further reduce bias. In the weighted method, more weights are given to scenario in the tail region of the loss distribution in the least squares regression. This is along a similar line of thought as the methods we are proposing in this thesis. More specifically, in implementing the weighted method, the authors suggest a two-stage process: The first stage carries out an unweighted regression whose output is used to determine the weights assigned to each outer scenario in the weighted regression in the second stage.

Other forms of intrinsic proxy include Gaussian process regression (Risk and Ludkovski, 2018), analytical solutions for the inner simulation output (Feng and Jing, 2017), and replacing the inner simulation by solving PDEs (Li and Feng, 2021). The focus of Feng and Jing (2017) is not nested simulation but the authors derive analytical solution for the risk-neutral liability and delta of a plain-vanilla GLWB contract in a geometric Brownian motion asset model. The analytical solution uses a technique of fitting exponential sums to a mortality density function. The analytical solution presented can be used as an intrinsic proxy to replace inner simulations of a nested simulation. Li and Feng (2021) provide an overview of multi-period nested simulation for regulatory purposes in insurance applications, and propose a PDE method to replace inner simulation of nested simula-

tions for complex GMWB contracts in a realistic modeling setting. The study finds the PDE methods superior in computational efficiency when compared with standard Monte Carlo method, least-squares Monte Carlo, or preprocessed inner simulations. Our proposals in this thesis for improving the efficiency of nested simulations remain in the traditional nested simulation framework, for its adaptability and widespread utilization in the insurance industry. Nevertheless, the PDE method proposed by Li and Feng (2021) can be incorporated into the proxy method that we propose.

In general, empirical proxies requiring trial inner simulations do not transfer well to path-dependent VA contracts, because the VA options are typically far out-of-the-money. As a result, a small number of trial inner simulations will not give an adequate assessment of the hedging losses in the general case, although in some specific cases importance sampling within the inner simulations could mitigate this problem.

The second approach to improve the efficiency of inner simulations, namely methods utilising a dynamics allocation of the computational budget of a nested simulation, have been developed by Lan *et al.* (2010), Liu and Staum (2010), Bauer *et al.* (2012),Broadie *et al.* (2011) and Risk and Ludkovski (2018). Assume that we have a fixed number of inner simulations that will be distributed across the outer scenarios. This is the inner simulation budget. The uniform nested simulation method allocates the inner simulation budget equally across all the scenarios. We discuss this process in more detail in Chapter 2. Dynamic allocation involves non-uniform allocation of the inner simulation budget. Lan *et al.* (2010) suggest a two stage process: In the first stage, results from a small number of initial inner simulations, uniformly allocated, are used to signal which scenarios were likely to have the most impact on the risk measure. The remainder of the inner simulation budget is then allocated only to these scenarios in the second stage simulation. After that the simulation output from the second stage is used to construct a confidence interval for expected shortfall. This is also similar to our proposal. We will discuss in Section 6.2 the applicability of their method in the multi-period nested simulation problem we study. Liu and Staum (2010) propose a three-stage process based on stochastic kriging: In the first stage, a stochastic kriging metamodel is built with a pilot inner simulation, on a set of design points (outer scenarios). In the second stage, an additional set of design points is identified, based on the metamodel from the first stage. In the third stage, the remaining computation budget is spent on inner simulation for all the design points identified in the first two stages. The resulting metalmodel is then used to estimate the losses of each prediction point (outer scenario), and subsequently to estimate the tail risk measure. This method is useful when there is flexibility in the choice of outer scenarios in the experiment design. This may not be the case in the context of regulatory reporting for which multi-period nested simulation is often required. Risk and Ludkovski (2018) develop

a k-round sequential algorithm that uses a Gaussian Process emulator as empirical proxy in each round and adaptively allocates inner simulation budget. The allocation of simulation budget is determined based on results from the previous round of simulation. Methods using Gaussian Process regression or stochastic kriging can also be useful in multi-period nested simulation. However, we need to first study how to choose appropriate design points for stochastic kriging at each time step in a multi-period nested simulation. We do not address this issue in the thesis but it will be a topic of future research. Broadie *et al.* (2011) also use a smaller number of initial trial simulations, but their method then proceeds sequentially, determining which individual scenario should be allocated the next simulation from the remaining inner simulation budget. Bauer *et al.* (2012) adapt and extend the idea of Gordy and Juneja (2010) and Lan *et al.* (2010) to nested simulation for quantifying the Solvency II Statutory Capital Requirement for equity-linked life insurance contracts.

It is worth noting that in all these papers, except Li and Feng (2021), the problem involves a single-step outer scenario, compared with the multi-period problem that we are considering. Typically, an insurer might project risk factors many years ahead, in weekly or monthly time steps, and new inner level simulations (which may be single or multi-period, depending on path dependency of the embedded options) are required at each time step of each scenario, as illustrated in Figure 1.1. The main challenge in applying the methods developed for single-period nested simulation to multi-period problem is the increase in dimension of the problem being considered, due to the number of time steps at which the inner simulation needs to be repeated. The multi-period problem has significantly higher variance arising from multiple iterations of inner simulations. Therefore in the pilot simulation, in order to utilize methods proposed in the literature listed above, a significant amount of computation would be required to produce any meaningful signal as to which scenarios belong to the tail. This would dampen the computational savings from these methods. Feng *et al.* (2016) summarize a few existing methods discussed above, including optimal budget allocation, sequential allocation of inner loops, linear interpolation of preprocessed inner loops, least-squares Monte Carlo and using PDE methods to replace inner loop simulation. They demonstrate how they can be applied in multi-period nested simulation for life insurance contracts and evaluate their performances. Their main findings are: (1) In modeling simple contracts with small computation budget, all these methods are more accurate than a standard Monte Carlo nested simulation, although the improvement is less prominent when a large computation budget is allowed. (2) In modeling complex contracts, analytical and numerical PDE methods have the best performance in terms of efficiency and accuracy. Linear interpolation of preprocessed inner loops and least squares Monte Carlo work well in low dimension problems but suffer from the curse

of dimensionality.

Among the work discussed above, Bauer *et al.* (2012); Cathcart and Morrison (2009); Bauer and Ha (2015); Krah *et al.* (2018); Feng *et al.* (2016); Feng and Jing (2017); Li and Feng (2021) consider nested simulation in the life insurance context while others consider nested simulation for portfolio risk management in a finance context.

All the papers discussed above consider nested simulation of a single model contract. As discussed, the other research angle on VA nested simulation is simulation of a large portfolio of insurance policies. Gan (2013, 2015); Gan and Lin (2015, 2017); Gan and Valdez (2019) propose various metamodeling approaches to select representative policies and use functional approximations to predict the values of the entire portfolio, to reduce the number of model points for the portfolio. Lin and Yang (2020b,a) also consider nested simulation of a large portfolio, with Lin and Yang (2020b) in the single-period and Lin and Yang (2020a) in the multi-period nested simulation setting. They propose using a cube sampling algorithm to select representative policies and using clustering to select representative outer scenarios. Functional approximations are then used to predict liability of the portfolio. In their work, the vast majority (approximately 98%) of computational savings arise from policy reduction in the simulation model.

In our work, 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.

Chapter 2

Standard Nested Simulation for Tail Risk Measures of Variable Annuity

In this chapter, we present a few common types of Variable Annuity guarantees, and discuss how a VA contract with a dynamic hedging program is modeled via a nested simulation. We also introduce the notation and assumptions used throughout this thesis. We limit discussion to research pertaining to the topic of this thesis. For a comprehensive review of different types of guarantees offered in VA contracts, hedging of the embedded guarantees, and modeling of the contracts, we refer readers to Hardy (2003).

2.1 Variable Annuity Payouts

Consider a generic VA contract whose time to maturity is $T^M \geq 0$ periods, e.g. T^M months. Suppose the policyholder of the VA contract dies at time τ. Let $T = \min(T^M, \tau)$, the earlier of the contract maturity and death of the policyholder, be the time the contract expires. Note that at any given time t, the time to death τ is a random variable whereas the time to maturity T^M is a constant.

Let F_t and G_t be the sub-account value and the guarantee value, respectively, at time $t = 0, 1, 2, \ldots, T-1, T$.

Let η^g be the gross rate at which management fee is deducted from the fund value each period. We assume the amount of management fee collected as income by the insurer is different from the gross amount deducted from the fund, with the difference offsetting the

regular investment expenses incurred. Let $\eta^n < \eta^g$ be the net rate at which management fee income is received by the insurer each period.

Let D_t be the present value at time 0 of \$1 payable at time t, discounted at the risk-free rate.

We will describe the payouts of the major types of VA benefits described in Section 1.1 more formally.

Guaranteed Minimum Maturity Benefit (GMMB) In the simplest form of the guarantee, the GMMB provides a fixed guarantee value, G, say, at the contract expiration. Typically $G = F_0$, the initial premium of the contract. The insurer's payout on the GMMB contract is $(G - F_T)^+$ at time T (the expiry date), and insurer's liability, net of fee income, at time t, $t = 1, 2, \ldots, T$, is

$$V_t = \frac{D_T}{D_t}(G - F_T)^+ - \sum_{s=t+1}^{T} \frac{D_s}{D_t} F_s \eta^n$$

This is similar to the payoff of a put option on the underlying investment.

Guaranteed Minimum Death Benefit (GMDB) The insurer's payout on the GMDB contract is $(G - F_\tau)^+$ at time τ and insurer's liability net of fee income at time t is

$$V_t = \frac{D_\tau}{D_t}(G - F_\tau)^+ - \sum_{s=t+1}^{\tau} \frac{D_s}{D_t} F_s \eta^n, \qquad t < \tau \leq T^M$$

Guaranteed Minimum Accumulation Benefit (GMAB) Suppose the GMAB contract has one renewal at time $T_1 < T$, then

$$\begin{aligned}
&\text{For } t \leq T_1: && G_t = G_0 \\
&\text{For } t > T_1: && G_t = \max(F_{T_1}, G_0) \\
&\text{Insurer payout at } T_1: && (G_0 - F_{T_1})^+ \\
&\text{Insurer payout at } T: && (\max(F_{T_1}, G_0) - F_T)^+
\end{aligned}$$

Thus, the insurer's liability of the GMAB contract at time t is

$$V_t = \frac{D_{T_1}}{D_t}(G_0 - F_{T_1})^+ + \frac{D_T}{D_t}(\max(F_{T_1}, G_0) - F_T)^+ - \sum_{s=t+1}^{T} \frac{D_s}{D_t} F_s \eta^n$$

Guaranteed Minimum Withdrawal Benefit (GMWB) Suppose the guaranteed minimum withdrawal rate is γ per period. Let $I_t = \gamma G_t$ denote the amount of guaranteed minimum withdrawal benefit payable at time t. The insurer's payout on the GMWB contract consists of all the guaranteed minimum withdrawal benefit paid after the sub-account value depletes, i.e. $\sum_{s=1}^{T}(I_s - F_s)^+$. Then the present value at $t < T$ of the GMWB liability is V_t where

$$V_t = \sum_{s=t+1}^{T} \frac{D_s}{D_t}(I_s - F_s)^+ - \sum_{s=t+1}^{T} \frac{D_s}{D_t} F_s \eta^n \mathbb{1}\{F_s > 0\} \tag{2.1}$$

Here we assume that the deduction of management fee coincides with the payment of withdrawal benefit. $\mathbb{1}\{\cdot\}$ is an indicator function.

Guaranteed Lifetime Withdrawal Benefit (GLWB) Suppose the policyholder of the GLWB contract dies at time τ, then the insurer's payout of the GLWB contract is $\sum_{s=1}^{\tau}(I_s - F_s)^+$, and insurer's liability net of fee of the GLWB contract V_t at time t is

$$V_t = \sum_{s=t+1}^{\tau} \frac{D_s}{D_t}(I_s - F_s)^+ - \sum_{s=t+1}^{\tau} \frac{D_s}{D_t} F_s \eta^n \mathbb{1}\{F_s > 0\}, \qquad t < \tau \leq T^M \tag{2.2}$$

In this thesis, we study the nested simulation of GMMB, GMAB, and GMWB.

2.2 Dynamic Hedging for Variable Annuities

As discussed in Section 1.1, insurers commonly use dynamic hedging to mitigate the market risk associated with the embedded options in VA contracts. In a dynamic hedging program, a hedging portfolio is set up for a block of VA contracts using futures and other derivatives (Geneva Association, 2013). The hedging portfolio is rebalanced periodically, responding to changes in market conditions and in the demographics of the block of contracts. More specifically, we consider a delta hedge for a single VA contract in this thesis. We choose to focus on delta hedge in this thesis since the most common hedge strategy in VA involves hedging delta and rho (Sun *et al.*, 2016). The proposals in this thesis can be adapted to model a delta/rho hedging strategy.

In this section we present the mechanism of dynamic hedging for a generic VA liability. We assume the time to expiration of the VA is T months. We also assume that the VA sub-account invests in a stock index, and that the hedging portfolio is rebalanced every month, coinciding with the deduction of management fee and the payment of benefits when

applicable. At any $t \leq T$, let S_t be the underlying stock price at time t. Then in absence of any withdrawal, the sub-account value F_t evolves as

$$F_t = F_{t-1} \times \frac{S_t}{S_{t-1}} \times (1 - \eta^g).$$

The evolution of sub-account value involving withdrawals is discussed in Section 5.4.

We assume that the delta hedge for the VA embedded option is composed of Δ_t units in the underlying stock, and a sum B_t in a risk free zero coupon bond maturing at T. The value of the delta hedge portfolio at $t-1$ is then

$$H_{t-1} = \Delta_{t-1} S_{t-1} + B_{t-1}$$

At time t, the value of this hedge has changed to

$$H_t^{bf} = \Delta_{t-1} S_t + B_{t-1} \frac{D_{t-1}}{D_t}$$

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.

$$HE_t = H_t - H_t^{bf}. \tag{2.3}$$

The costs to set up the initial hedge portfolio, the periodic hedging gains and losses due to rebalancing at each date, $t = 1, 2, \ldots, T$, the final unwinding of the hedge, the payment of guaranteed benefit, and the management fee income, 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 insurer's overall (gain)/loss from the VA; this is the loss random variable to which we apply a suitable risk measure.

More concretely, at time 0, the insurer's overall (gain)/loss L of the VA contract and

its dynamic hedging program is:

$$L = H_0 + \sum_{t=1}^{T-1} HE_t \cdot D_t - H_T^{bf} D_T + V_0 \tag{2.4}$$

$$= H_0 + \sum_{t=1}^{T-1} D_t \left(H_t - H_t^{bf} \right) - H_T^{bf} D_T + V_0$$

$$= B_0 + S_0\Delta_0 + \sum_{t=1}^{T-1} D_t \left(B_t + S_t\Delta_t - B_{t-1}\frac{D_{t-1}}{D_t} - S_t\Delta_{t-1} \right) - D_T S_T \Delta_{T-1} + V_0$$

$$= \sum_{t=0}^{T-1} \Delta_t (D_t S_t - D_{t+1} S_{t+1}) + V_0 \tag{2.5}$$

Note, as indicated in the previous section, V_0 represents the discounted payoff of the VA contract as of time 0, net of fee income, in the absence of hedging. The first three terms $H_0 + \sum_{t=1}^{T-1} HE_t \cdot D_t - H_T^{bf} D_T$ in Equation (2.4) represent the (gain)/loss from the hedging program: H_0 is the initial cost of setting up the hedge, $\sum_{t=1}^{T-1} HE_t \cdot D_t$ represents the (gain)/loss from all the rebalancing trades in the dynamic hedging program, while $H_T^{bf} D_T$ represents the proceeds from unwinding the hedging program.

Equation (2.5) shows that in the bond holdings of the hedging portfolio, all the interim bond trades cancel out, because the interest rate at which the bond value accumulates is the same as the rate at which profit and loss from bond trades are discounted. In the stock holdings of the hedging portfolio, the (gain)/loss arises from the initial set up of the stock future holding, as well as the profit and loss from each stock future trade. Computationally, Equation (2.5) is more efficient than (2.4) because the interim hedging portfolio values H_t for $t = 0, \ldots, T-1$ are not required.

In Equation (2.5), Δ_t is determined using risk neutral pricing under probability measure $\mathbb{Q}$, while the tail risk measure of the loss random variable L is determined under the real world probability measure $\mathbb{P}$. If we are interested in estimating the tail risk measure of L using Monte Carlo simulation, then we need either an analytical solution for Δ_t, or a nested simulation where the outer scenarios are generated under real world measures to estimate the loss distribution, and the inner sample paths are generated under a risk neutral measure to estimate Δ_t.

2.3 Multi-period uniform nested simulation for tail risk measures

In this section, we describe a standard multi-period nested simulation for the tail risk measures that we consider in this thesis. We refer to this kind of nested simulation as a uniform nested simulation because it allocates the same number of inner simulations at each time step to each outer scenario.

In a nested simulation, we define the following notation:

- $X_t \in \mathbb{R}^d$, for $d \geq 1, t = 1, 2, \ldots, T$ is a random vector, representing the outer scenario state variable at time t in the nested simulation.

  In the VA example we discussed in Section 2.2, each X_t is a vector consisting of the stock index price, and possibly the sub-account and guarantee values under the $\mathbb{P}$ measure at time t.

- $Y_{t,t'}$ for $t = 1, 2, \ldots, T$ and $t' = t, \ldots, T$, is the inner simulation state variable at time t', originating from X_t.

  In the VA example,$Y_{t,t'}$ is a vector consists of the stock index price, and possibly the sub-account and guarantee values under the $\mathbb{Q}$ measure at time t', given X_t. It is not an essential detail that X and Y are evaluated under different probability measures, but it is a common feature of losses relating to dynamic hedging programs.

- $\boldsymbol{Y}_t = (Y_{t,t}, \ldots, Y_{t,T})$ is the entire path of inner simulation state variables, given X_t.

- $\mu_t(X_t)$ is a pricing or valuation functional evaluated at time t given X_t.

  In the VA example, $\mu_t(X_t) = \Delta_t$, the delta of the contract at time t, which is an expectation of a function of the inner simulation state variables $\boldsymbol{Y}_t$, conditional on X_t. From here on, we specify that $\mu_t(X_t) = \Delta_t = \mathbb{E}\left[f(\boldsymbol{Y}_t|X_t)\right]$. We use the Infinite Perturbation Analysis (IPA) method for sensitivity estimation Broadie and Glasserman (1996); Glasserman (2013).

- L is the loss random variable. In particular, we assume L is a function of $\mu_t(X_t)$, for $t = 0, 1, \ldots, T$. That is,

$$L = h\left(\mu_0(X_0), \ldots, \mu_T(X_T)\right)$$

  where h is a function that maps $\mu_0(X_0), \ldots, \mu_T(X_T)$ to L.

In the VA example, L is the overall loss of the VA contract and its dynamic hedging program, as defined in Equation (2.5). More specifically, L is a function of the stock index prices and the delta's at each future hedge rebalancing date, up to the expiry of the contract.

We assume X_t and $Y_{t,t'}$ both follow discrete time Markov processes. The discrete time assumption is fitting in our simulation context. In addition, most models used in practice for evaluating insurance portfolios are Markov or can be treated as Markov by expanding the state variable space.

Our objective is to estimate the appropriate tail risk measure of L. Here $\mu_t(X_t)$ can be evaluated analytically where feasible, or using Monte Carlo simulation where analytical solutions are not available. If we use Monte Carlo simulation for this step, then we require inner level simulations to estimate $\mu_t(X_t)$, $t = 0, 1, \ldots, T$.

Assuming no analytical solution for $\mu_t(X_t)$ exists, we use a multi-period nested simulation with a horizon of T periods and time step of 1 period to estimate the tail risk measure of L. In the context of a dynamically hedged VA contract, we assume the time steps coincide with the re-balancing frequency of the hedging portfolio and with the payment frequency of guaranteed benefits.

We define some more notations for the nested simulation:

- M is the number of outer scenarios sample path in the nested simulation.
- N is the number of inner simulations invoked at each time step along each of the M outer scenarios.
- $X_t^{(i)}$, for $t = 1, 2, \ldots, T$, $i = 1, \ldots, M$, is the sample value of X_t in the ith outer scenario.
- $\boldsymbol{X}^{(i)} = (X_0^{(i)}, X_1^{(i)}, X_2^{(i)}, \ldots, X_T^{(i)})$, $i = 1, \ldots, M$ is the ith outer scenario.
- $Y_{t,t'}^{(i)}$, for $t' = t, \ldots, T$, is random variable $Y_{t,t'}$, conditional on $X_t^{(i)}$.
- $\boldsymbol{Y}_t^{(i)} = \left(Y_{t,t}^{(i)}, Y_{t,t+1}^{(i)}, \ldots, Y_{t,T}^{(i)}\right)$ is the entire path of random vector $Y_{t,t'}^{(i)}$, $t' = t, \ldots, T$.
- $Y_{t,t'}^{(i,j)}$, $t = 1, 2, \ldots, T$, $t' = t, \ldots, T$, $i = 1, \ldots, M$, $j = 1, \ldots, N$, is the sample value at time t' on the jth inner simulation sample path, conditional on $X_t^{(i)}$.

- $\boldsymbol{Y}_t^{(i,j)} = \left(Y_{t,t}^{(i,j)}, \ldots, Y_{t,T}^{(i,j)}\right)$, $t = 1, 2, \ldots, T$, $i = 1, \ldots, M$, $j = 1, \ldots, N$, is the entire j-th inner sample path, conditional on $X_t^{(i)}$.
- $g_t^{(i)}(\boldsymbol{Y}_t^{(i,j)})$ is the probability density function of $\boldsymbol{Y}_t^{(i,j)}$, conditional on $X_t^{(i)}$, that is, $g_t^{(i)}(\boldsymbol{y}) = g(\boldsymbol{y}|X_t^{(i)})$.

  In a Monte Carlo simulation, $\boldsymbol{Y}_t^{(i,j)} \overset{i.i.d.}{\sim} g_t^{(i)}(\boldsymbol{y})$, for all $j = 1, \ldots, N$.

Then in a standard Monte Carlo simulation with N inner simulations at time t in outer scenario i, $\mu_t\left(X_t^{(i)}\right)$ can be estimated as $\widehat{\mu}_t\left(X_t^{(i)}\right) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} f\left(\boldsymbol{Y}_t^{(i,j)}\right)$.

Suppose the tail risk measure of interest is CTE_α. Given M equally likely scenarios from a Monte Carlo simulation, denote the (unknown) true loss in each given scenario by $L_i = h\left(\mu_0(X_0^{(i)}), \ldots, \mu_0(X_T^{(i)})\right)$, $i = 1, \ldots, M$, and let $L_{(1)} \leq L_{(2)} \leq \cdots \leq L_{(M)}$ be the corresponding ordered losses. For a prescribed confidence level α (assume αM is an integer), the CTE_α is given by

$$CTE_\alpha = \frac{1}{(1-\alpha)M}\sum_{i=\alpha M+1}^{M} L_{(i)} = \frac{1}{(1-\alpha)M}\sum_{i\in\mathcal{T}_{(1-\alpha)M}} L_i, \tag{2.6}$$

where $\mathcal{T}_{(1-\alpha)M} = \{\boldsymbol{X}^{(i)} : L_i > L_{(\alpha M)}\}$ is the set of the $(1-\alpha)M$ *true tail scenarios* that are included in the calculation of CTE_α in (2.6).

Using the estimated inner simulation outputs $\mu_t(X_t^{(i)})$ for $i = 1, \ldots, M$ and $t = 0, \ldots, T$, the loss random variables L_i, $i = 1, \ldots, M$ can be estimated by

$$\widehat{L}_i^{\text{NS}} = h\left(\mu_0^{\text{NS}}(X_0^{(i)}), \mu_1^{\text{NS}}(X_1^{(i)}), \ldots, \mu_T^{\text{NS}}(X_T^{(i)})\right), i = 1 \ldots, M.$$

Denote the corresponding ordered losses by $\widehat{L}_{(1)^{\text{NS}}}^{\text{NS}} \leq \widehat{L}_{(2)^{\text{NS}}}^{\text{NS}} \leq \cdots \leq \widehat{L}_{(M)^{\text{NS}}}^{\text{NS}}$, then the CTE_α can be estimated by

$$\widehat{CTE}_\alpha^{\text{NS}} = \frac{1}{(1-\alpha)M}\sum_{i=\alpha M+1}^{M} \widehat{L}_{(i)^{\text{NS}}}^{\text{NS}} = \frac{1}{(1-\alpha)M}\sum_{i\in\widehat{\mathcal{T}}_{(1-\alpha)M}^{\text{NS}}} \widehat{L}_i^{\text{NS}}, \tag{2.7}$$

where $\widehat{\mathcal{T}}_{(1-\alpha)M}^{\text{NS}} = \{\boldsymbol{X}^{(i)} : \widehat{L}_i^{\text{NS}} > \widehat{L}_{(\alpha M)^{\text{NS}}}^{\text{NS}}\}$ is the set of the $(1-\alpha)M$ *nested simulation tail scenarios* that are included in the calculation of $\widehat{CTE}_\alpha^{\text{NS}}$ in (2.7)

Note that both (2.7) and (2.6) are based on the same given set of M outer scenarios, so their difference is only due to the inner simulation noise. The inner simulation noise affects the accuracy of (2.7) in two ways:

1. Classification of tail scenarios. Due to the inner simulation noise, the estimated losses $\widehat{L}_i^{\text{NS}}$ may differ from true losses L_i. As a result, these losses may have different rankings so the nested simulation tail scenarios may be different from the true tail scenarios. The consequence is two-fold: some true tail scenarios are missed from the estimated CTE and some true non-tail scenarios are included.

2. Estimation of tail losses. If the tail scenarios could somehow be identified, we see that it is important to accurately estimate the losses for the tail scenarios. Non-tail scenarios' losses are irrelevant for estimating the CTE; this is also true for other tail risk measures.

The simulation procedures we propose in this thesis are specifically designed to address both the above observations. We will discuss them in detail in later chapters.

We refer to the multi-period uniform nested simulation that we described in this section as standard Monte Carlo (SMC) nested simulation. Algorithm 1 outlines the steps of a SMC nested simulation for estimating the CTE_α of losses for a Delta-hedged VA contract.

For example, in a GMMB contract, $\boldsymbol{X}^{(i)}$ is the sub-account value in the i-th outer scenario while $\boldsymbol{Y}_t^{(i,j)}$ is the sub-account value in the j-th inner simulation sample path. $\boldsymbol{Y}_t^{(i,j)}$ originates from $X_t^{(i)}$, the sub-account value at time t in outer scenario i. Using the IPA method, the j-th inner sample derivative of VA payoff at time t in outer scenario i is

$$f(\boldsymbol{Y}_t^{(i,j)}) = -\frac{D_T}{D_t} \cdot \frac{Y_{t,T}^{(i,j)}}{S_t^{(i)}} \cdot \mathbb{1}\{G > Y_{t,T}^{(i,j)}\} - \sum_{s=t+1}^{T} \frac{D_s}{D_t} \frac{Y_{t,s}^{(i,j)}}{S_t^{(i)}} \eta^n.$$

Algorithm 1 can easily 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.* (2015) for more information on estimating greeks using Monte Carlo simulation.

The evolution in Algorithm 1 of state variables such as F_t and G_t in Line 2, and the inner simulation model in Line 3, 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 Section 2.4 below.

Algorithm 1: Standard multi-period nested simulation for estimating CTE_α of losses in a dynamically hedged VA contract.

Input: α : CTE confidence level
M : numbers of outer scenarios.
N : numbers of inner simulations.
X_0 : initial state variables of the VA.
T : Number of time periods in the VA contract term.

for $i = 1, \ldots, M$ **do**
  Set $X_0^{(i)} = X_0$ then simulate under the real-world measure the i-th outer scenario $\boldsymbol{X}^{(i)} = (X_0^{(i)}, X_1^{(i)}, \ldots, X_T^{(i)})$.
  for $t = 0, 1, \ldots, T$ **do**
    for $j = 1, \ldots, N$ **do**
      Set $Y_{t,t}^{(i,j)} = X_t^{(i)}$, then simulate under the risk-neutral measure the j-th inner sample path of state variables: $\boldsymbol{Y}_t^{(i,j)} = \left(Y_{t,t}^{(i,j)}, Y_{t,t+1}^{(i,j)}, \ldots, Y_{t,T}^{(i,j)}\right)$.
      Calculate the j-th inner sample derivative of VA payoff $f\left(\boldsymbol{Y}_t^{(i,j)}\right)$.
    end
    Estimate the Greeks of VA at t in outer scenario i using the IPA method as $\widehat{\mu}\left(X_t^{(i)}\right) = \frac{1}{N}\sum_{j=1}^{N} f\left(\boldsymbol{Y}_t^{(i,j)}\right)$.
  end
  Estimate the VA loss in outer scenario i as $\widehat{L}_i = h\left(\widehat{\mu}_0\left(X_0^{(i)}\right), \ldots, \widehat{\mu}_T\left(X_T^{(i)}\right)\right)$.
end
Sort the M simulated losses in ascending order to give $\widehat{L}_{(1)} \leq \widehat{L}_{(2)} \leq \cdots \leq \widehat{L}_{(M)}$.
Estimate the CTE_α of the loss by $\widehat{CTE}_\alpha = \dfrac{1}{(1-\alpha)M}\displaystyle\sum_{i=\alpha M+1}^{M} \widehat{L}_{(i)}$.

We refer to the total number of inner simulation replications in a simulation procedure as the simulation budget and denote it by Γ. Let N_{SMC} denote the number of inner simulations required at each time step in each outer scenario, and let C_{inner} denote the number of calculations required at each time step in a inner simulation for estimating the conditional expectation $\mu_t(X_t)$. Then to estimate the CTE_α, using the standard Monte Carlo nested simulation in Algorithm 1, the simulation budget required is:

$$\Gamma_{\text{SMC}} = \frac{M \times N_{\text{SMC}} \times T \times (T+1)}{2} \times C_{\text{inner}} \tag{2.8}$$

Here the factor $\frac{T \times (T+1)}{2}$ represents the sum of number of times the inner simulation needs to be repeated, which is the sum of the sequence $T, T-1, \ldots, 1$.

As discussed in Section 1.3, the simulation budget Γ required in a typical VA valuation using the standard nested simulation procedure is prohibitively large.

2.4 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 ith outer simulation, $H_t^{(i)}$, 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:

$$H_t^{(i)} = Ge^{-r(T-t)}\Phi(-d_2(t,T)) - S_t^{(i)}\Phi(-d_1(t,T)), \qquad \Delta_t^{(i)} = -\Phi(-d_1(t,T)) \tag{2.9}$$

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

$$d_1(t,T) = \frac{\ln\left(\frac{S_t^{(i)}}{G}\right) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \qquad d_2(t,T) = d_1(t,T) - \sigma\sqrt{T-t}. \tag{2.10}$$

The GMAB payoff(s) have the same structure as a tandem put option in finance. We derive formulas for the hedge of a European tandem option, with $F_0 = S_0$, with a single renewal point T_1 and final maturity T. The initial guarantee is G_0; the second guarantee is $\max(G_0, S_{T_1})$. We consider two different cases: hedge at $t < T_1$ and hedge at $t \geq T_1$.

I. $H_t^{(i)}$ when $t < T_1$

When $t < T_1$, we have $F_t^{(i)} = S_t^{(i)}$. Then the payout at time T_1 is $\left(G_0 - S_{T_1}^{(i)}\right)^+$. The sub-account value immediately after the payout at time T_1 is $F_{T_1^+}^{(i)} = \max\left(G_0, S_{T_1}^{(i)}\right)$,

and the guarantee value after T_1 is also $\max\left(G_0, S_{T_1}^{(i)}\right)$. Then the payout at time T is

$$\begin{aligned}\left(\max\left(G_0, S_{T_1}^{(i)}\right) - F_T^{(i)}\right)^+ &= \left(\max\left(G_0, S_{T_1}^{(i)}\right) - \max\left(G_0, S_{T_1}^{(i)}\right) \times \frac{S_T^{(i)}}{S_{T_1}^{(i)}}\right)^+ \\ &= \max\left(G_0, S_{T_1}^{(i)}\right) \times \left(1 - \frac{S_T^{(i)}}{S_{T_1}^{(i)}}\right)^+\end{aligned}$$

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. That is

$$p^* = e^{-r(T-T_1)}\,\Phi(-d_2^*) - \Phi(-d_1^*) \tag{2.11}$$

$$\text{where} \quad d_1^* = \frac{(r+\sigma^2/2)(T-T_1)}{\sigma\sqrt{T-T_1}} \qquad d_2^* = d_1^* - \sigma\sqrt{T-T_1} \tag{2.12}$$

Note that p^* is a constant with respect to the underlying stock price process. This is a useful function in our derivation below.

Hence, the full GMAB guarantee hedge at $t < T_1$ is the expectation of the sum of discounted payout at time T_1 and T, under risk neutral probability measure.

$$\begin{aligned}H_t^{(i)} =& \mathbb{E}_t\left[e^{-r(T_1-t)}\left(G_0 - S_{T_1}^{(i)}\right)^+ + e^{-r(T-t)}\max\left(G_0, S_{T_1}^{(i)}\right)\left(1 - \frac{S_T^{(i)}}{S_{T_1}^{(i)}}\right)^+\right] \qquad (2.13)\\ =& \mathbb{E}_t\left[e^{-r(T_1-t)}\left(G_0 - S_{T_1}^{(i)}\right)^+\right] + \mathbb{E}_t\left[e^{-r(T_1-t)}\left(\left(G_0 - S_{T_1}^{(i)}\right)^+ + S_{T_1}^{(i)}\right)p^*\right]\\ =& \mathbb{E}_t\left[e^{-r(T_1-t)}\left(G_0 - S_{T_1}^{(i)}\right)^+\right](1+p^*) + \mathbb{E}_t\left[e^{-r(T_1-t)}S_{T_1}^{(i)}\right]p^*\\ =& \left(G_0 e^{-r(T_1-t)}\Phi\left(-d_2(t,T_1)\right) - S_t^{(i)}\Phi\left(-d_1(t,T_1)\right) + S_t^{(i)}\right)(1+p^*) - S_t^{(i)}\end{aligned}$$

Hence, the Delta of the GMAB guarantee is

$$\Delta_t^{(i)} = -\left(\Phi(-d_1(t,T_1)) - 1\right)(1+p^*) - 1 = p^*\Phi(d_1(t,T_1)) - \Phi(-d_1(t,T_1)) \tag{2.14}$$

where $d_1(t,T_1)$ and $d_2(t,T_1)$ are as defined in Equations (2.10).

II. $H_t^{(i)}$ when $t \geq T_1$

When $t \geq T_1$, the hedge portfolio is identical to the GMMB case, with guarantee $G = \max(G_0, S_{T_1})$ and with maturity T. The fund value at $t \geq T_1$ is $F_t^{(i)} = S_t^{(i)} \left(\max \left(1, \frac{G_0}{S_{T_1}^{(i)}} \right) \right)$, and this replaces $S_t^{(i)}$ in Equations (2.9) and (2.10).

This derivation can also be found in (Hardy, 2003, Appendix B).

We can see from Equations (2.10) that delta of the GMAB contract could be positive when $t < T_1$ and the contract is out of the money. However, immediately after the same contract renews at time T_1, the contract becomes a GMMB contract, which has has a negative delta. This reveals significant gamma risk embedded in the GMAB contract, particularly around the time of renewal.

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 Monte Carlo simulation. But the analytic Black-Scholes hedge costs are expected to be correlated with the true values, so we could 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 chapter.

Chapter 3

Importance-Allocated Nested Simulation

3.1 Introduction

In this chapter, we will present the Importance-Allocated Nested Simulation (IANS) procedure. The IANS procedure is a two-stage process which exploits the fact that the CTE calculation only uses the largest $(1-\alpha)M$ simulated loss values, so the inner simulation budget can be concentrated on the scenarios which are most likely to generate the largest losses. In the IANS procedure, the Stage 1 simulation uses an analytical proxy evaluation to identify the scenarios most likely to contribute to the CTE risk measure, while Stage 2 involves inner simulation only for the scenarios identified in the first stage.

For many VA guarantees, especially those with a simpler structure, we can identify an extrinsic proxy which is likely to be adequate for ranking scenarios. The proxy model values are not used directly in the estimation of the risk measure, they are only used to determine the allocation of the inner simulation budget. Hence, the proxy model does not have to provide an accurate valuation of the underlying losses; it only has to provide an accurate ranking of the values of the underlying losses. If the proxy is perfect at ranking scenarios, because the proxy valuation and the full inner simulation valuation are comonotonic, then we can strategically allocate the inner simulation budget to exactly the scenarios that generate the losses required for the CTE calculation. If, as is likely, the proxy is not perfect, then we use it to identify a larger set of scenarios that will, with high probability, contain the true tail scenarios. In this case, there is some wastage of inner simulation budget. If the proxy is very poor, then the scenario set may have to be very

large to ensure that the tail scenarios are included, and there may be minimal or zero efficiency gains compared to the full uniform nested simulation method.

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 only used 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.

Our numerical experiments show that, in the VA context, IANS can be up to 30 times more accurate than a standard Monte Carlo experiment, measured by relative mean squared errors (RMSE), when both are assigned the same computational budget.

3.2 Importance-Allocated Nested Simulation (IANS) Method

In this section, we present the Importance-Allocated Nested Simulation (IANS) procedure for estimating the CTE_α of a VA GMMB or GMAB, using a nested simulation with screening approach. We provide an outline in Section 3.2.1 and supply the details in subsequent sections.

3.2.1 Outline of the IANS procedure

Algorithm 2: Importance-Allocated Nested Simulation of losses for a Delta-hedged VA contract.

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 from Delta-hedging the VA contract of interest.

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

Algorithm 2: Importance-Allocated Nested Simulation of losses for a Delta-hedged VA contract *(continued)*.

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. See Section 3.2.2.

(I.2) Calibrate the proxy asset model to the underlying risk-neutral asset model in inner-level simulations. See Section 3.2.3.

(I.3) Implement Algorithm 1 but with the analytic hedge calculations for the proxy derivative and asset model replacing the inner simulation step. Calculate loss, L_i^P, given scenario $\boldsymbol{X}^{(i)}$, $i = 1, 2, \ldots, M$, based on the proxy model.

(I.4) Identify a set of m proxy tail scenarios, $\mathcal{T}_m^P$, where

$$\boldsymbol{X}^{(i)} \in \mathcal{T}_m^P \Leftrightarrow L_i^P > L_{(M-m)}^P$$

for some $m \geq (1-\alpha)M$, where $L_{(j)}^P$ denotes the jth ranked value of L_i^P. See Section 3.2.6.

Stage II: Nested simulation with concentrated computation budget

(II.1) Allocate the remaining computational budget to the m proxy tail scenarios in $\mathcal{T}_m^P$.

(II.2) Implement the inner simulation step of Algorithm 1, with the original risk neutral asset model and VA payoff, but only for the m scenarios identified in Step (I.4). Denote the simulated loss in this step as ${}^m\widehat{L}_i$, for each $\boldsymbol{X}^{(i)} \in \mathcal{T}_m^P$.

(II.3) Identify the $(1-\alpha)M$ largest values of ${}^m\widehat{L}_i$.

(II.4) Compute the $\widehat{CTE}_\alpha$ as the output using

$$\widehat{CTE}_\alpha = \frac{1}{(1-\alpha)M} \sum_{j=m-(1-\alpha)M+1}^{m} {}^m\widehat{L}_{(j)} \tag{3.1}$$

3.2.2 Selection of proxies

Unlike the proxy approaches in the existing literature, such as least-squares Monte Carlo, preprocessed inner simulations, analytical or PDE solutions, the proxy tail scenarios 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 procedure 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.

To identify a suitable proxy model, we first consider why the inner simulation step is needed. Typically, the complexity in the valuation, leading to the need for the guarantee cost to be determined using simulation rather than analytically, comes from some combination of the following issues.

(1) An intractable risk-neutral measure; this is a common problem, as the contracts are very long term, and models used often involve stochastic volatility, for which analytic valuation formulae may not be available.

(2) Dynamic lapse assumptions; insurers typically assume that lapses are (somewhat) dependent on the moneyness of the guarantee. A popular version of the dynamic lapse model, from NAIC (2020), is described in Section 3.3.3. Incorporating dynamic lapses creates a path dependent option valuation that is not analytically tractable.

(3) The option payoff is too complex for analytic valuation.

The proxy model should be a tractable model that is close enough to the more complex model to give an approximate ranking of the scenarios. We might construct the proxy by using a tractable risk-neutral measure in place of the stochastic volatility model, to cover point (1) above; we might use a simplified lapse rate assumption, to deal with point (2) above, and we might replace a complex payoff with a simpler one that captures most of the costs to cover point (3) above. We reiterate that the proxy does not have to give an accurate estimate of the option costs based on the more complex assumptions; it is sufficient that the scenarios generating the highest losses under the proxy model overlap significantly with the scenarios generating the highest losses under the full inner simulation approach. If a suitable extrinsic proxy is not available, it may be replaced with an intrinsic proxy based, for example, on pilot simulations. We consider this issue in more detail in Chapter 5.

In most VA portfolios, the key benefits contributing to the risk are the living benefits. The GMMB and GMAB are among the simpler forms of living benefits, and we will

consider these here. More complex living benefits, in particular those with annuity-type benefits, require more complicated proxy models, such as the semi-analytical solutions for risk-neutral pricing of GMWB and GLWB in Feng and Volkmer (2016) and Feng and Jing (2017), and maybe less suited to this approach. We consider an extension for these cases in Chapter 5, in the context of a GMWB. For the GMMB and GMAB, the put option and tandem put option identified in Section 2.4 are obvious proxy derivatives, as the option payoffs are identical to the guarantee payoffs, if we ignore complications of policyholder behaviour.

3.2.3 Calibration of proxy asset model

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 2.4 in proxy calculations at negligible computational cost. GBM is inconsistent with important features of observed stock returns, including extreme left-tail (extreme loss) 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 where 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 in our numerical studies, both for the outer, real world scenarios, and also, in an equivalent risk neutral form, for the inner scenarios.

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. Detailed descriptions of a RSLN model with two regimes and a GARCH(1,1) model are given next in Section 3.2.4. The corresponding volatility calibrations of these two models are given in Section 3.2.5.

3.2.4 Advanced asset models

In this section we describe the two advanced stochastic asset models we consider in this chapter: two-regime switching lognormal model and GARCH(1,1) model. Let $R_t = \ln(S_{t+1}/S_t)$ be the log-return process of stock price. The stochastic asset models describe the process of R_t.

Regime Switching Lognormal (RSLN) Model

The first stochastic asset process we consider is the regime-switching lognormal (RSLN) model with two regimes. Each regime in the regime-switching model is characterized by a lognormal model with a different set of parameters. The regime-switching model was introduced to the actuarial literature by Hardy (2001).

Definition 3.2.1. *Let $\rho_t = 1, 2$, be the regime in the period $[t, t+1)$.*

In a RSLN model with two regimes, the log return process is defined as

$$R_t|\rho_t \sim N\left(\mu_{\rho_t}, \sigma^2_{\rho_t}\right)$$

where μ_h, σ^2_h are the mean and variance parameter of the hth regime.

We assume the switching between two regimes takes place at the end of each period, following a Markov process. The transition probabilities of regime switching is

$$p_{ij} = Pr\left[\rho_{t+1} = j|\rho_t = i\right], \qquad i, j = 1, 2.$$

The two-regime RSLN model "*provides a very good fit to the stock index data relevant to equity-linked insurance*" (Hardy, 2003).

There are six parameters in an RSLN model with two regimes under real world measure $\mathbb{P}$. The six parameters used in our numerical studies are summarized in Table 3.2.

The financial market is incomplete in the regime-switching model, thus its risk neutral measure $\mathbb{Q}$ 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). The change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is achieved by changing the log return mean parameter from μ_i under the $\mathbb{P}$ measure to $r - \sigma_i^2/2$ for $i = 1, 2$.

GARCH(1,1) Model

The generalized autoregressive conditional heteroskedasticity (GARCH) model was first developed by Bollerslev (1986), and is still one of the most popular asset models due to its flexibility and good fit for many econometric applications. In this chapter we consider a GARCH(1,1) model of the monthly log return of the stock price. The GARCH(1,1) model is a class of GARCH model frequently considered in the econometrics and option pricing literature (Augustyniak *et al.*, 2017).

Definition 3.2.2. *In a GARCH(1,1) model, the log return process is defined as*

$$\begin{aligned} R_t &= \mu + \sigma_t \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} N(0,1), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

Given the log return process under the $\mathbb{P}$ measure in Definition 3.2.2, the log return of the stock price and variance under the risk-neutral measure $\mathbb{Q}$ are given by (Duan, 1995):

$$\begin{aligned} \widetilde{R}_t &= r - \frac{1}{2}\sigma_t^2 + \sigma_t \widetilde{\varepsilon}_t, \qquad \widetilde{\varepsilon}_t \overset{i.i.d.}{\sim} N(0,1), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 \widetilde{\varepsilon}_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned} \tag{3.2}$$

The parameters for the GARCH(1,1) model used in our numerical studies are summarized in Table 3.3.

3.2.5 Proxy model volatility calibration

To implement the IANS procedure, we are interested in calibrating the volatility of the proxy Black-Scholes model to that in the true asset model. We show the calculation of the average volatility in the period $[t_1, t_2)$ conditioning on the current state at time t_0, where $t_0 \leq t_1 < t_2$. This is the Black-Scholes implied volatility used in the proxy model.

For GMMBs and GMABs with no further renewal date, we set the Black-Scholes volatility between t and maturity T to the average volatility of the true asset model in the same period, conditioning on the state variables at time t.

For GMABs whose valuation date is prior to the renewal date $t < T_1$, we calibrate two volatilities: (1) the average volatility of the true asset model between t and T_1, conditioning on the state variables at time t and (2) the average volatility of the true asset model between T_1 and T, conditioning on the state variables at time t. These two volatilities can be different and are used to calculate d's in Equation (2.10) and d^*'s in Equation (2.12).

Regime Switching Lognormal (RSLN) Model

Let $Q(t_1, t_2)$ be the number of sojourns in regime 1 in $[t_1, t_2)$. The probability function of $Q(t_1, t_2)$ given the regime at time t_1, can be calculated via backward recursion (Hardy, 2001). These are useful quantities for volatility calibrations in the IANS method.

To calibrate the implied volatility in the Black-Scholes proxy model to the RSLN model, we consider the *average volatility* in the period $[t_1, t_2)$ conditioning on the information at time t_0, where $t_0 \leq t_1 < t_2$.

Let $\mathcal{F}_t$ denote the information at time t.

Definition 3.2.3. *For any time t_1 and t_2 such that $t_0 \leq t_1 < t_2$, we define $R(t_1, t_2)$ to be the log-return process between t_1 and t_2 so that $R(t_1, t_2) = \ln(S_{t_2}/S_{t_1})$. We also define*

$$\widetilde{\sigma}(t_1, t_2|t_0) = \sqrt{\frac{1}{t_2 - t_1}\mathbb{V}[R(t_1, t_2)|\mathcal{F}_{t_0}]}$$

*to be the **average volatility** of $R(t_1, t_2)$ conditioning on the information at time t_0.*

Proposition 3.2.4. *In two-regime risk-neutral regime-switching lognormal model, whose mean conditional log return is $r - \sigma_i^2/2$ for $i = 1, 2$*

$$\begin{aligned}\widetilde{\sigma}^2(t_1, t_2|t_0) =& \sigma_2^2 + \frac{(\sigma_1^2 - \sigma_2^2)}{t_2 - t_1}\Big(\mathbb{E}\big[Q(t_1, t_2)|\rho_{t_0}\big] \\ &+ \frac{1}{4}(\sigma_1^2 - \sigma_2^2)\big(\mathbb{E}\big[\mathbb{V}\left[Q(t_1, t_2)|\rho_{t_1}\right]|\rho_{t_0}\big] + \mathbb{V}\big[\mathbb{E}[Q(t_1, t_2)|\rho_{t_1}]|\rho_{t_0}\big]\big)\Big),\end{aligned}$$

where

$$\mathbb{E}\big[Q(t_1, t_2)\big|\rho_{t_0}\big] = \mathbb{E}\big[\mathbb{E}[Q(t_1, t_2)|\rho_{t_1}]\big|\rho_{t_0}\big] = \sum_{i=1}^{2}\mathbb{E}[Q(t_1, t_2)|\rho_{t_1} = i] \times Pr[\rho_{t_1} = i|\rho_{t_0}],$$

$$\mathbb{E}\big[\mathbb{V}\left[Q(t_1, t_2)|\rho_{t_1}\right]\big|\rho_{t_0}\big] = \sum_{i=1}^{2}\mathbb{V}\left[Q(t_1, t_2)|\rho_{t_1} = i\right] \times Pr[\rho_{t_1} = i|\rho_{t_0}], \text{ and}$$

$$\mathbb{V}\big[\mathbb{E}[Q(t_1, t_2)|\rho_{t_1}]\big|\rho_{t_0}\big] = \sum_{i=1}^{2}\mathbb{E}[Q(t_1, t_2)^2|\rho_{t_1} = i] \times Pr[\rho_{t_1} = i|\rho_{t_0}] - \mathbb{E}\big[Q(t_1, t_2)\big|\rho_{t_0}\big]^2$$

Proof. Let $n = t_2 - t_1$.

In a two-regime regime-switching lognormal model, where μ_i and σ_i^2 are the mean and variance parameter of the ith regime, $i = 1, 2$, by definition

$$R(t_1, t_2)|Q(t_1, t_2) \sim N\left(Q(t_1, t_2)\mu_1 + (n - Q(t_1, t_2))\,\mu_2, Q(t_1, t_2)\sigma_1^2 + (n - Q(t_1, t_2))\,\sigma_2^2\right)$$

Hence

$$\begin{aligned}
&\mathbb{V}[R(t_1,t_2)]\\
=&\mathbb{E}\left[\mathbb{V}[R(t_1,t_2)|Q(t_1,t_2)]\right]+\mathbb{V}\left[\mathbb{E}[R(t_1,t_2)|Q(t_1,t_2)]\right]\\
=&\mathbb{E}\left[Q(t_1,t_2)\sigma_1^2+(n-Q(t_1,t_2))\,\sigma_2^2\right]+\mathbb{V}\left[Q(t_1,t_2)\mu_1+(n-Q(t_1,t_2))\,\mu_2\right]\\
=&n\sigma_2^2+(\sigma_1^2-\sigma_2^2)\mathbb{E}\left[Q(t_1,t_2)\right]+(\mu_1-\mu_2)^2\mathbb{V}\left[Q(t_1,t_2)\right]\\
=&n\sigma_2^2+(\sigma_1^2-\sigma_2^2)\mathbb{E}\left[Q(t_1,t_2)\right]+(\mu_1-\mu_2)^2\mathbb{V}\left[Q(t_1,t_2)\right]\\
=&n\sigma_2^2+(\sigma_1^2-\sigma_2^2)\mathbb{E}\left[Q(t_1,t_2)\right]+\frac{1}{4}(\sigma_1^2-\sigma_2^2)^2\mathbb{V}\left[Q(t_1,t_2)\right]\qquad\text{since } \mu_i=r-\sigma_i^2/2
\end{aligned}$$

Now if we consider $\mathbb{V}[R(t_1,t_2)|\rho_{t_0}]$, we have

$$\begin{aligned}
&\mathbb{V}[R(t_1,t_2)|\rho_{t_0}]\\
=&n\sigma_2^2+(\sigma_1^2-\sigma_2^2)\mathbb{E}\left[Q(t_1,t_2)|\rho_{t_0}\right]+\frac{1}{4}(\sigma_1^2-\sigma_2^2)^2\mathbb{V}\left[Q(t_1,t_2)\big|\rho_{t_0}\right]\\
=&(t_2-t_1)\sigma_2^2+(\sigma_1^2-\sigma_2^2)\mathbb{E}\left[\mathbb{E}[Q(t_1,t_2)|\rho_{t_1}]\big|\rho_{t_0}\right]\\
&+\frac{1}{4}(\sigma_1^2-\sigma_2^2)^2\left(\mathbb{V}\left[\mathbb{E}[Q(t_1,t_2)|\rho_{t_1}]\big|\rho_{t_0}\right]+\mathbb{E}\left[\mathbb{V}[Q(t_1,t_2)|\rho_{t_1}]\big|\rho_{t_0}\right]\right)
\end{aligned}$$

Thus

$$\begin{aligned}
&\widetilde{\sigma}^2(t_1,t_2|t_0)=\frac{1}{t_2-t_1}\mathbb{V}[R(t_1,t_2)|\rho_{t_0}]\\
=&\sigma_2^2+\frac{(\sigma_1^2-\sigma_2^2)}{t_2-t_1}\Big(\mathbb{E}\big[Q(t_1,t_2)|\rho_{t_0}\big]\\
&+\frac{1}{4}(\sigma_1^2-\sigma_2^2)\big(\mathbb{E}\big[\mathbb{V}\left[Q(t_1,t_2)|\rho_{t_1}\right]|\rho_{t_0}\big]+\mathbb{V}\big[\mathbb{E}[Q(t_1,t_2)|\rho_{t_1}]|\rho_{t_0}\big]\big)\Big)
\end{aligned}$$

This completes the proof.

□

GARCH(1,1) Model

To calibrate the implied volatility in the Black-Scholes proxy model to the GARCH(1,1) model, we consider the *average expected volatility* in the period $[t_1,t_2)$ conditioning on the information at time t_0, where $t_0\le t_1<t_2$.

Definition 3.2.5. *In the GARCH(1,1) model under risk neutral measure in Equation (3.2), for any time t_1 and t_2 such that $t_0 \leq t_1 < t_2$, we define*

$$\bar{\sigma}(t_1, t_2|t_0) = \sqrt{\frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \mathbb{E}\left[\mathbb{V}[\widetilde{R}_t|\mathcal{F}_{t-1}]\middle|\mathcal{F}_{t_0}\right]}$$

to be the ***average expected volatility*** *of $R(t_1, t_2)$ conditioning on the information at time t_0.*

Note that the average expected volatility $\bar{\sigma}(t_1, t_2|t_0)$ for calibrating the proxy implied volatility to the GARCH(1,1) model is defined differently from the average volatility $\widetilde{\sigma}(t_1, t_2|t_0)$ for calibrating the proxy implied volatility to the RSLN model. We choose a different definition for different asset models so that the proxy implied volatility can be calculated with little computation and is sufficiently accurate to identify tail scenarios in the IANS methods. Note that in the GARCH(1,1) model under the risk neutral measure defined in Equation (3.2),

$$\bar{\sigma}(t_1, t_2|t_0) = \sqrt{\frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \mathbb{E}\left[\mathbb{V}[\widetilde{R}_t|\mathcal{F}_{t-1}]\middle|\mathcal{F}_{t_0}\right]} < \sqrt{\frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \mathbb{V}\left[\widetilde{R}_t\middle|\mathcal{F}_{t_0}\right]}$$

As a result, the Black-Scholes implied volatility in the proxy model is slightly understated. However, as we will show in numerical examples, the proxy implied volatility still performs well in the IANS method.

Proposition 3.2.6. *In a GARCH(1,1) model of the monthly log return of the stock price:*

$$\begin{aligned} \widetilde{R}_t &= r - \frac{1}{2}\sigma_t^2 + \sigma_t \widetilde{\varepsilon}_t, \qquad \widetilde{\varepsilon}_t \overset{i.i.d.}{\sim} N(0,1), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 \widetilde{\varepsilon}_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

For any time t_1 and t_2 such that $t_0 \leq t_1 < t_2$,

$$\bar{\sigma}^2(t_1, t_2|t_0) = \frac{\alpha_0}{1 - \alpha_1 - \beta} + \frac{(\alpha_1 + \beta)^{t_1 - t_0 - 1}}{t_2 - t_1} \left(\sigma_{t_0+1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta}\right) \frac{1 - (\alpha_1 + \beta)^{t_2 - t_1}}{1 - \alpha_1 - \beta}$$

Proof. For $t \geq t_1$,

$$
\begin{aligned}
\mathbb{E}\left[\sigma_t^2|\mathcal{F}_{t_0}\right] =& \mathbb{E}\left[\alpha_0 + \alpha_1\sigma_{t-1}^2\widetilde{\varepsilon}_{t-1}^2 + \beta\sigma_{t-1}^2|\mathcal{F}_{t_0}\right] \\
=& \alpha_0 + \mathbb{E}\left[(\alpha_1\widetilde{\varepsilon}_{t-1}^2 + \beta)\sigma_{t-1}^2|\mathcal{F}_{t_0}\right] \\
=& \alpha_0 + \mathbb{E}\left[(\alpha_1\widetilde{\varepsilon}_{t-1}^2 + \beta)|\mathcal{F}_{t_0}\right] \times \mathbb{E}\left[\sigma_{t-1}^2|\mathcal{F}_{t_0}\right] \\
=& \alpha_0 + (\alpha_1 + \beta)\mathbb{E}\left[\sigma_{t-1}^2|\mathcal{F}_{t_0}\right] \qquad \text{since } \mathbb{E}[\widetilde{\varepsilon}_t^2] = 1 \\
& \vdots \\
=& \alpha_0 \sum_{i=0}^{t-t_0-2} (\alpha_1 + \beta)^i + (\alpha_1 + \beta)^{t-t_0-1}\mathbb{E}\left[\sigma_{t_0+1}^2|\mathcal{F}_{t_0}\right] \\
=& \frac{\alpha_0(1 - (\alpha_1 + \beta)^{t-t_0-1})}{1 - \alpha_1 - \beta} + (\alpha_1 + \beta)^{t-t_0-1}\sigma_{t_0+1}^2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\bar{\sigma}^2(t_1, t_2|t_0) =& \frac{1}{t_2 - t_1}\sum_{t=t_1}^{t_2}\mathbb{E}\left[\mathbb{V}[\widetilde{R}_t|\mathcal{F}_{t-1}]\big|\mathcal{F}_{t_0}\right] = \frac{1}{t_2 - t_1}\sum_{t=t_1}^{t_2-1}\mathbb{E}[\sigma_t^2|\mathcal{F}_{t_0}] \\
=& \frac{1}{t_2 - t_1}\sum_{t=t_1}^{t_2-1}\left(\frac{\alpha_0(1 - (\alpha_1 + \beta)^{t-t_0-1})}{1 - \alpha_1 - \beta} + (\alpha_1 + \beta)^{t-t_0-1}\sigma_{t_0+1}^2\right) \\
=& \frac{\alpha_0}{1 - \alpha_1 - \beta} + \frac{(\alpha_1 + \beta)^{t_1-t_0-1}}{t_2 - t_1}\left(\sigma_{t_0+1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta}\right)\frac{1 - (\alpha_1 + \beta)^{t_2-t_1}}{1 - \alpha_1 - \beta}
\end{aligned}
$$

□

3.2.6 Safety margin for tail scenario identifications

The proxies selected in Step (I.2) of Algorithm 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 the proxy tail scenario set $\mathcal{T}_m^P$ in Step (I.4) with some safety margin, that is, $m \geq (1 - \alpha)M$, to ensure the outer scenarios with the largest $(1 - \alpha)M$ losses based on inner simulations are included in $\mathcal{T}_m^P$. So, the proxy tail scenarios are the $m > (1 - \alpha)M$ outer scenarios with the largest simulated losses based on the proxy calculations; we use these to identify the largest $(1 - \alpha)M$ simulated losses based on the inner simulations, assuming that, with high confidence, the $(1 - \alpha)M$ true tail scenarios are a subset of the m proxy tail scenarios.

The size of the proxy tail scenario set m is an experiment design parameter in IANS. If m is very large, the likelihood of capturing the true tail scenarios is high, but at the cost of spreading the inner simulations over a larger number of 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 m is close to $(1-\alpha)M$,then the inner simulation budget is focused on fewer scenarios, so those that are included will have more accurate loss valuations, but some tail scenarios will be wrongly omitted, because the proxy loss ranking is not comonotonic with the true loss ranking. Hence there is a tradeoff between a high likelihood of including the true tail scenarios (when $m \to M$) and high concentration of simulation budget in Stage 2 (when $m \to (1-\alpha)M$). In this work an arbitrary selection of $m = (1-\alpha+5\%)M$ for the proxy tail scenario set is included in the numerical examples. In Chapter 4 we consider a more structured approach to the selection of m.

3.3 Numerical Experiments

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

- No transaction costs 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.

Under these assumptions, the loss random variable of the VA contracts consist only of the gain and loss from the hedging program, i.e. the initial cost of the hedging portfolio and the present value of periodic hedging errors, plus any payout under the guaranteed benefit. In practice, fee income, expenses, commissions, decrements, and costs due to basis risk are likely to make up a proportion of the loss. Nonetheless, the IANS procedure still offers useful insights because the loss from hedging is the only quantity for which we need the inner simulation part of a nested simulation.

We consider two risk measures, $\text{CTE}_{80\%}$ and $\text{CTE}_{95\%}$, as these are commonly used in valuation and economic capital setting in Canada, consistent with regulatory standards. See OSFI (2017) and CIA (2017). As discussed above, we consider a GMMB and GMAB in the numerical experiments, using the parameters specified in Table 3.1.

Description	Notation	Value
Maturity of Contract and Projection Period	T	240 months
Time of Renewal (for GMAB only)	T_1	End of the 120th month
Initial Fund Value	F_0	1000
Initial Level of Guarantee	G_0	100% of F_0

Table 3.1: Parameters for VA Contracts

The RSLN and GARCH(1,1) model parameters under the $\mathbb{P}$ measure are provided in Table 3.2 and Table 3.3. We derive the parameters under the $\mathbb{Q}$ measure accordingly based on the change of measure described in Section 3.2.4. Parameters in the RSLN model were taken from those estimated based on TSE 300 data in Hardy (2001), with some modification to the log mean return slightly more distinctive between the two regimes. The value of α_0 is chosen such that long-run average volatility in this GARCH(1,1) models equals the long-run average volatility in the RSLN model, with the parameters in Table 3.2.

(Monthly rate)	Real World
Risk-free Rate: r	0.002
Mean - Regime 1 ($\rho = 1$): μ_1	0.0085
Mean - Regime 2 ($\rho = 2$): μ_2	-0.0200
Standard Deviation - Regime 1: σ_1	0.035
Standard Deviation - Regime 2: σ_2	0.080
Transition Probability - from Regime 1: p_{12}	0.04
Transition Probability - from Regime 2: p_{21}	0.20

Table 3.2: Parameters in the regime-switching lognormal model in numerical examples in Section 3.3 under the $\mathbb{P}$ measure

The use of a stochastic interest rate model can also play a critical role in the valuation of VAs (Peng *et al.*, 2012; Shevchenko and Luo, 2017; Augustyniak and Boudreault, 2017;

(Monthly rate)	Real World
r	0.002
μ	0.00375
α_0	0.0002094225
α_1	0.1
β	0.8
σ_0	0.0457627
ε_0	0

Table 3.3: Parameters in the GARCH(1,1) model in numerical examples in Section 3.3 under the $\mathbb{P}$ measure

Molent, 2020). We use a constant risk-free rate in this work. In future work, we will extend our proposal to incorporate stochastic interest rate models.

We assess the IANS procedure by assuming a fixed computational budget for simulation, and compare the accuracy of the resulting CTE estimates with estimators produced with the same computational budget, using the standard nested Monte Carlo (SMC) simulation in Algorithm 1. Section 3.3.2 presents numerical experiments under static lapse, and experiments for dynamic lapse are presented in Section 3.3.3.

3.3.1 Benchmarking Large Scale Nested Simulations

In order to assess the relative mean squared errors (RMSE) 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 (for convenience) as the *true* value of the CTEs; and the scenarios generating the largest $(1-\alpha)M$ losses are referred to as the *true* tail scenarios.

To illustrate the first stage of the IANS procedure, we replace the inner simulations with closed-form formulas based on the put option (GMMB) and tandem put option (GMAB) proxy derivatives, with the GBM asset model and examine how many true tail scenarios are correctly identified by the proxies.

In Figure 3.1 we compare the losses that are simulated by the full nested simulation and those generated by the proxy simulation. We can see that the values of the simulated

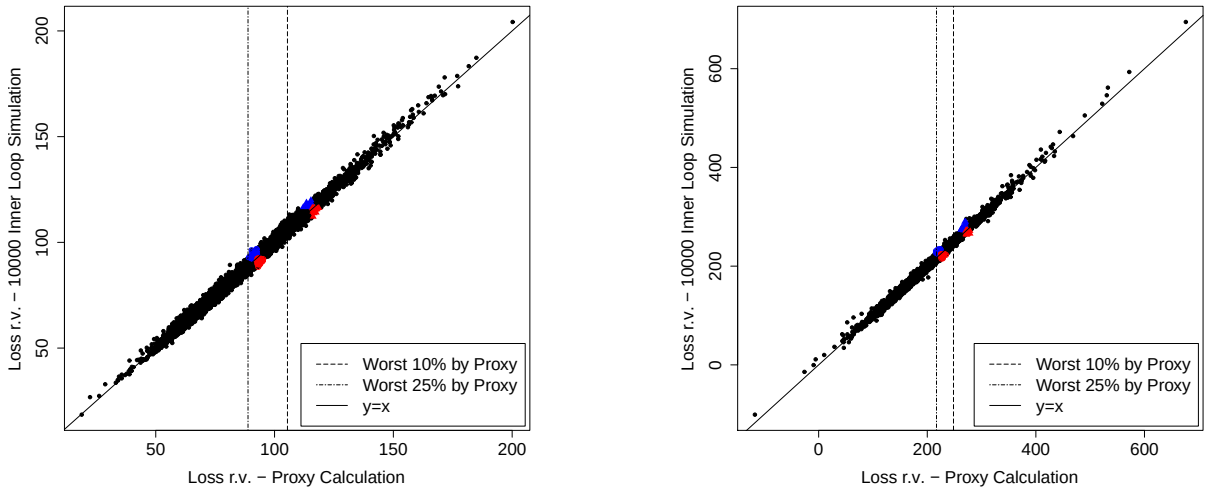
losses produced by these two methods are highly correlated. This indicates that Stage 1 of the IANS procedure is able to correctly identify most true tail scenarios without any inner simulation. Table 3.4 summarizes this observation quantitatively. We see that the closed-form proxy calculation, along with the safety margin in Stage 1 of the IANS procedure, identifies the true tail scenarios in the nested simulation very accurately, for different CTE levels, different asset models, and different VA types. This robust and accurate identification of tail scenarios leads to the high performance of the IANS procedure, as showcased in subsequent experiments. Note though the accurate identification of tail scenarios does come at the cost of the additional scenarios included in the safety margin. For example, for $\text{CTE}_{95\%}$ estimation, 10% or 1,000 outer scenarios are included in the safety margin. Table 3.4 shows that all 500 true tail scenarios are included in the safety margin, which suggests that 500 non-tail scenarios are included in the Stage 2 simulation. Similarly, for $\text{CTE}_{80\%}$ estimation, 25% or 2,500 outer scenarios are included in the safety margin. Data in Table 3.4 then imply between 500 to 515 non-tail scenarios are included in the Stage 2 simulation.

		#(%) of Correctly Identified Tail Scen.			
		RSLN		GARCH(1,1)	
CTE Level	# Tail Scen.	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%)

Table 3.4: Tail scenario identification by the proxy simulation (with safety margin) in Stage 1 of the IANS procedure under static lapse assumption.Correctly identified tail scenarios are the true tail scenarios included in the safety margin in Stage 1 of the IANS procedure.

3.3.2 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 3.5. The computation effort required in the proxy calculation is negligible. Thus it is omitted in the calculation of computational budget. The fixed simulation budget in all cases is 1% of that in the bench-



(a) Simulated losses for GMMB.

(b) Simulated losses for GMAB.

Figure 3.1: Simulated losses in 10,000 outer scenarios under the RSLN model. 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.

mark simulation in Section 3.3.1. By design, the *SMC-5,000-200* experiment has a larger 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 $M = 5,000$. The number of inner projections required using IANS is set to $m = (1 - \alpha + 5\%)M$. 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.

Figure 3.2 and 3.3 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 using the large scale simulation discussed in Section 3.3.1. Comparing Figure 3.2a with Figure 3.2b and 3.2c, and Figure 3.3a with Figure 3.3b and 3.3c, we see

Experiment	(a) M	(b) N	(c) m	Nested Simulation Computational Budget Γ
SMC-5,000-200, CTE α	5,000	200	n/a	$(a) \times (b) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$
SMC-1000-1000, CTE α	1,000	1,000	n/a	$(a) \times (b) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$
SMC-200-5,000 CTE α	200	5,000	n/a	$(a) \times (b) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$
IANS, CTE $\alpha = 80\%$	5,000	800	1,250	$(b) \times (c) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$
IANS, CTE $\alpha = 95\%$	5,000	2,000	500	$(b) \times (c) \times \frac{(1+12\times20)\times(12\times20)}{2} = 2.892 \times 10^{10}$

Table 3.5: Number of simulations in each numerical experiment shown in Section 3.3.2

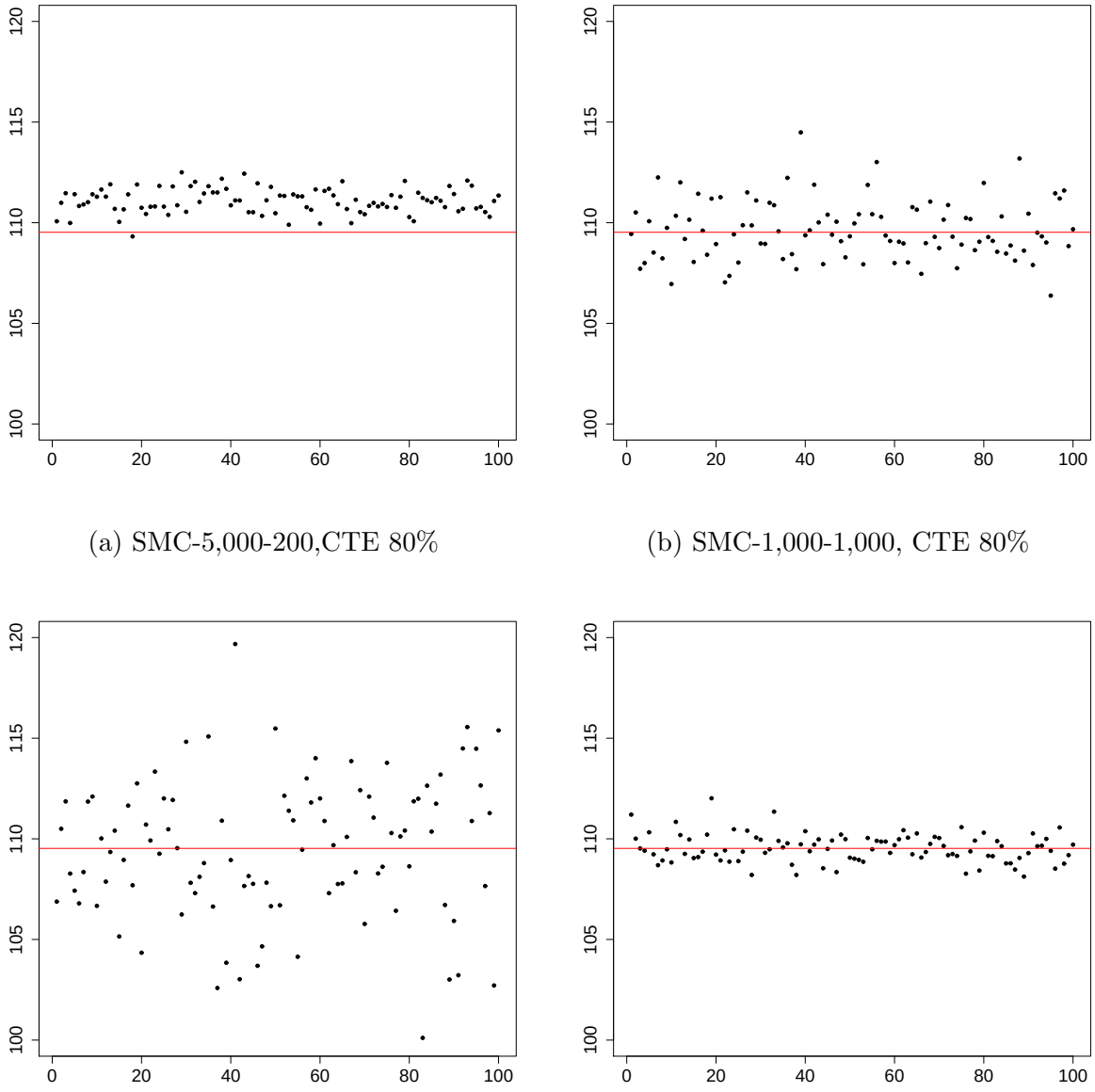
that using a larger number of outer-level simulations reduces the variance, while using a larger number of inner-level simulations appears to reduce the bias. These results are consistent with, for example, Broadie *et al.* (2011), Gordy and Juneja (2010). A larger number of outer-level simulations reduces variation in extreme losses simulated from one experiment to another, which reduces the variance. On the other hand, a larger number of inner-level simulations ensures a more consistent distribution of number of contracts renews or matures in-the-money, which reduces the bias in the hedging error. Figures 3.2 and 3.3 show that the IANS method achieves lower bias than the SMC-5,000-200 experiment, and lower variance than the SMC-1,000-1,000 and SMC-200-5,000 experiment. Figure 3.4 and 3.5 illustrate the results from the same experiments for the GMAB contract under the GARCH(1,1) asset model, where similar patterns are found.

Table 3.6 summarizes the RMSE for different experiment designs. Each RMSE is calculated as

$$\text{RMSE} = \frac{1}{n}\sum_{i=1}^{n}\frac{(\widehat{\mu}_i^{est} - \mu)^2}{\mu} \tag{3.3}$$

where $n = 100$, $\widehat{\mu}_i^{est}$ is the estimated CTE in the ith independent repeated experiment and μ is the corresponding CTE value estimated by large scale nested simulation discussed in Section 3.3.1. The RMSEs are then decomposed into relative bias in Table 3.7 and relative variance in Table 3.8, for the different experiment designs. The relative bias is calculated as $\frac{(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}_i^{est} - \mu)}{\mu}$, and the relative variance is calculated as $\frac{1}{n}\sum_{i=1}^{n}\frac{(\widehat{\mu}_i^{est} - \frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}_i^{est})^2}{\mu}$.

Table 3.6 demonstrates that, for these examples, the IANS procedure achieves smaller RMSEs compared with straightforward nested simulation, given the same simulation budget. For both 80% and 95% confidence levels, we see from Table 3.6 that the SMC-5,000-



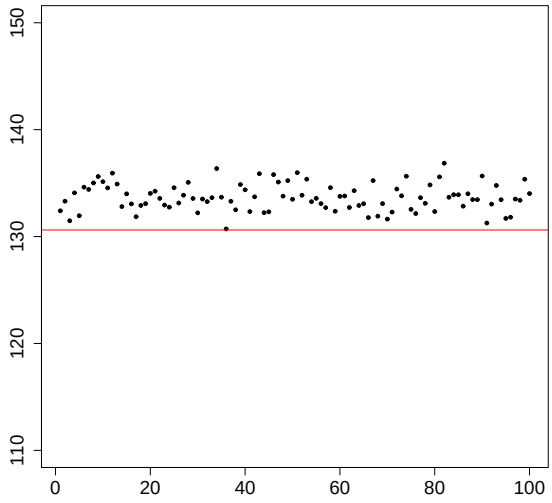
(a) SMC-5,000-200,CTE 80%

(b) SMC-1,000-1,000, CTE 80%

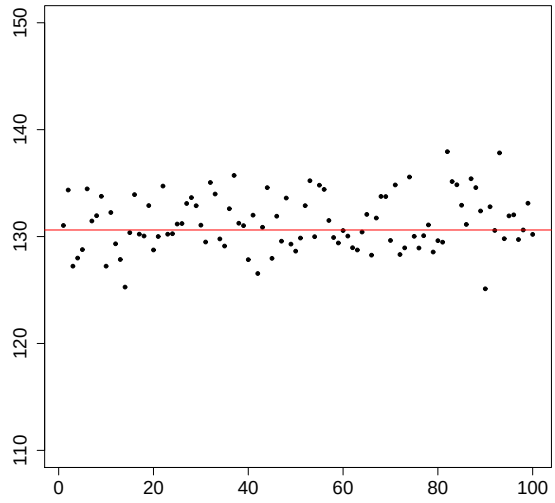
(c) SMC-200-5,000, CTE 80%

(d) IANS, CTE 80%

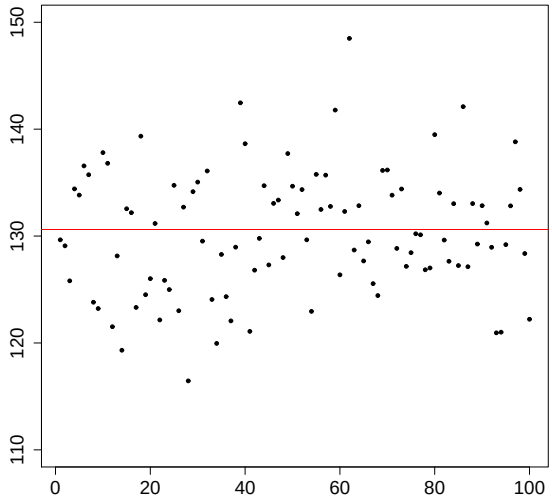
Figure 3.2: Estimated $\text{CTE}_{80\%}$ of simulated GMMB losses under Regime-Switching Model in 100 independent repeated experiments. The solid red line in each graph indicates the true value estimated by the large scale simulation discussed in Section 3.3.1.



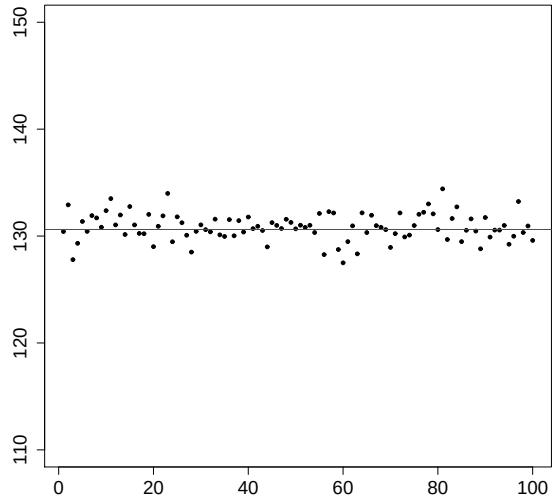
(a) SMC-5,000-200, CTE 95%



(b) SMC-1,000-1,000, CTE 95%

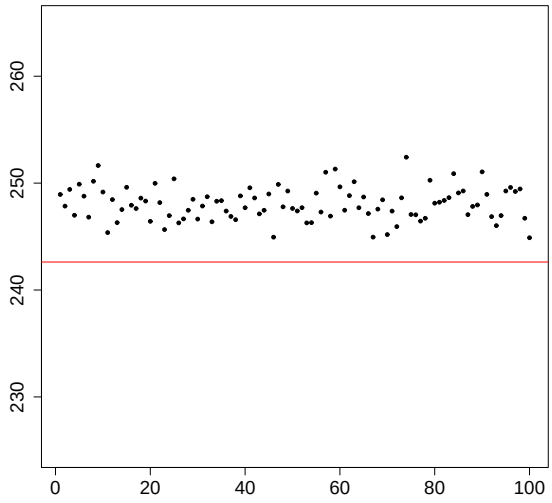


(c) SMC-200-5,000, CTE 95%

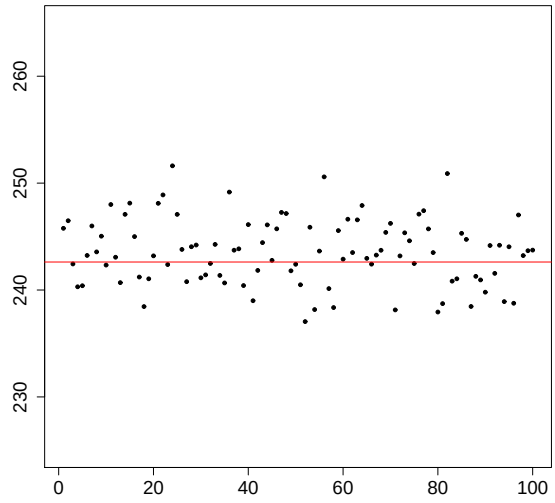


(d) IANS, CTE 95%

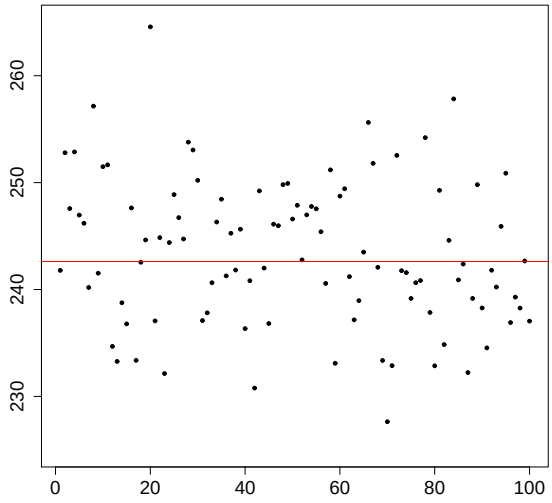
Figure 3.3: Estimated $\text{CTE}_{95\%}$ of simulated GMMB losses under Regime-Switching Model in 100 independent repeated experiments. The solid red line in each graph indicates the true value estimated by the large scale simulation discussed in Section 3.3.1.



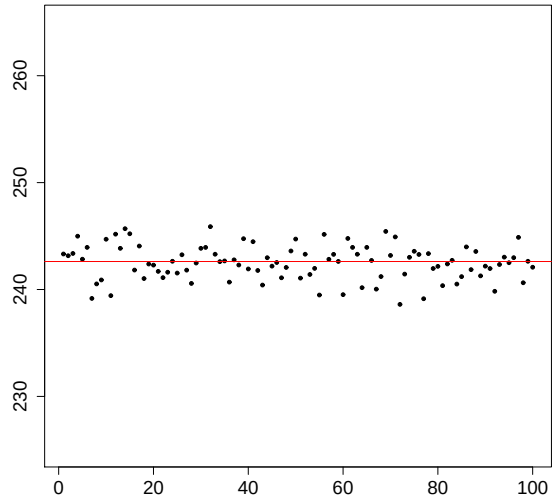
(a) SMC-5,000-200, CTE 80%



(b) SMC-1,000-1,000, CTE 80%

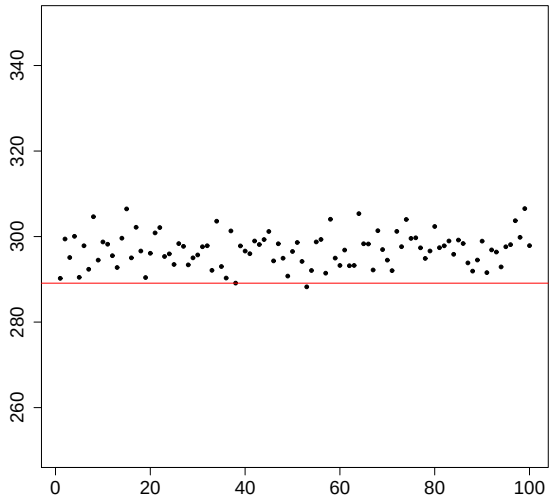


(c) SMC-200-5,000, CTE 80%

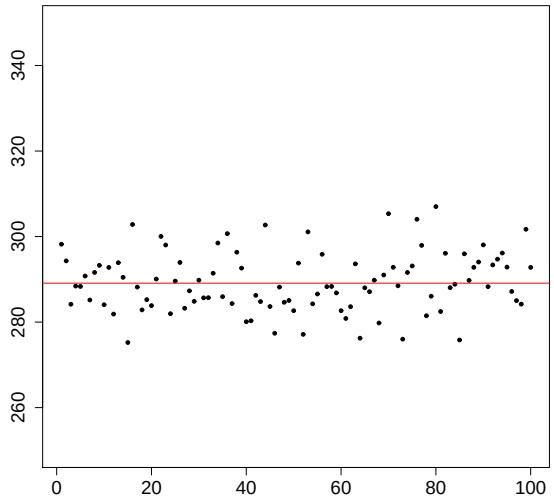


(d) IANS, CTE 80%

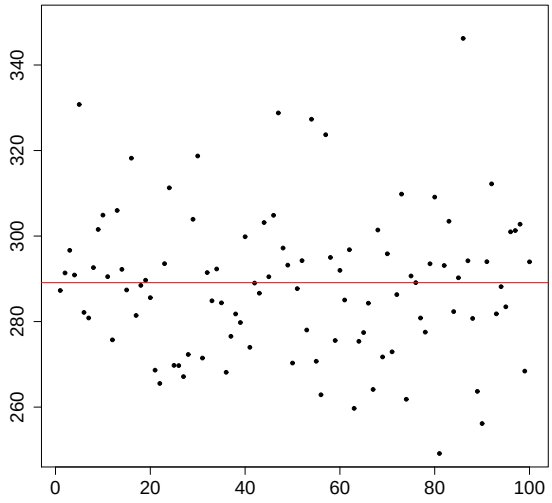
Figure 3.4: Estimated $\text{CTE}_{80\%}$ of simulated GMAB losses under GARCH(1,1) Model in 100 independent repeated experiments. The solid red line in each graph indicates the true value estimated by the large scale simulation discussed in Section 3.3.1.



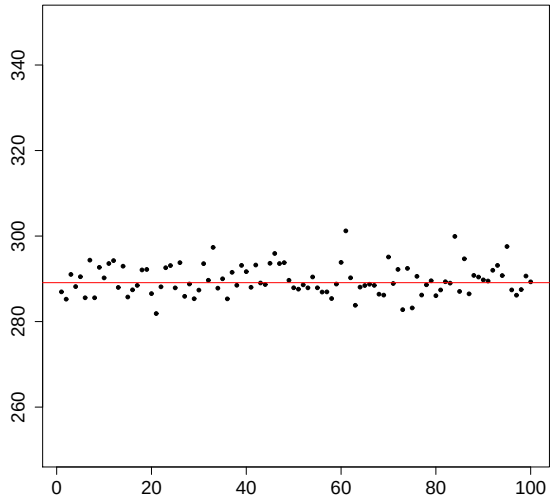
(a) SMC-5,000-200, CTE 95%



(b) SMC-1,000-1,000, CTE 95%



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



(d) IANS, CTE 95%

Figure 3.5: Estimated $\text{CTE}_{95\%}$ of simulated GMAB losses under GARCH(1,1) Model in 100 independent repeated experiments. The solid red line in each graph indicates the true value estimated by the large scale simulation discussed in Section 3.3.1.

	RSLN		GARCH(1,1)	
Experiment Design	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 3.6: Relative mean square errors (RMSEs) in the CTEs at $\alpha = 80\%$ and 95%, for different experiment designs assuming static lapses.

	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%

Table 3.7: Relative bias in the CTEs at $\alpha = 80\%$ and 95%, for different experiment designs assuming static lapses.

200 experiments have significantly smaller RMSEs than those of the other corresponding SMC experiments. Table 3.8 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 the outer-level simulations relative to the inner-level simulations; this observation is consistent with other studies in nested simulations (Broadie *et al.*, 2011; Gordy and Juneja, 2010).

	RSLN		GARCH(1,1)	
Experiment Design	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%

Table 3.8: Relative variance in the CTEs at $\alpha = 80\%$ and 95%, for different experiment designs assuming static lapses.

The RMSEs indicate that the mean squared errors in the IANS experiments are within 0-5% of the true CTE values, whereas the mean squared errors in a few SMC experiments are much higher relative to the true values. Compared to the SMC-1,000-1,000 and SMC-200-5,000 experiments, we observe in Table 3.6 and Table 3.8 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 considered in the IANS experiments, compared with 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 3.6 and Table 3.7 that the reduction in RMSEs in the IANS experiments are mostly due to the reduction in relative bias, because they use more inner simulations.

Among the IANS experiments, the 95% confidence level experiments use more inner simulations than the 80% confidence level experiments. However, the magnitude of improvement in relative bias and RMSEs in the 95% confidence level IANS experiment is smaller than those in the 80% confidence level experiments. This is due to the fact that the sensitivity of bias 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 simulations.

The results for GMAB contracts in Table 3.6 are consistent with those of GMMB contracts. The GMAB contract is more complicated than the GMMB, so the simulated

losses are more volatile, resulting in higher RMSEs in general.

3.3.3 Dynamic Hedging of VAs under Dynamic Lapse

To further demonstrate the robustness of the IANS procedure, we conducted large scale nested simulations, similar to those in Section 3.3.1, for both GMMB and GMAB contracts using a dynamic lapse model. We simulate the cashflows under both the regime-switching and the GARCH models. The goal of these experiments is to test the effectiveness of the proxy model in identifying the tail scenarios in more realistic settings.

The dynamic lapse model is as follows.

- The fund value F and guarantee value G are reduced proportionally by lapse.
- ${}_{\frac{1}{12}}q^{l}_{x+\frac{t}{12}}$, the monthly lapse rate from time t to $t+1$ is:

$$ {}_{\frac{1}{12}}q^{l}_{x+\frac{t}{12}} = \min\left(1, \max\left(0.5, 1 - 1.25 \times \left(\frac{G_t}{F_t} - 1.1\right)\right)\right) \times {}_{\frac{1}{12}}q^{base}_{x+\frac{t}{12}} \tag{3.4} $$

where

$$ {}_{\frac{1}{12}}q^{base}_{x+\frac{t}{12}} = \begin{cases} 0.00417 & \text{if } t < 84, \\ 0.00833 & \text{if } t \geq 84. \end{cases} \tag{3.5} $$

This dynamic lapse multiplier applied to the base lapse rate is taken from the NAIC's Valuation Manual 21 (NAIC, 2020). A dynamic lapse multiplier of this form is commonly used in practice to model simpler VA contracts such as GMMBs and GMABs. Whether or not this dynamic lapse assumption is the most suitable for modeling GMMB and GMAB contract is outside the scope of this thesis; our focus is to demonstrate the effectiveness of the IANS procedure based on a model that is similar to industry practice, and more complex than the plain vanilla cases above.

We use the same proxy calculations as those in Section 3.3.2 despite the additional complexity of dynamic lapse; the proxy calculation uses a static lapse assumption equal to the base lapse rate. The reason why a static lapse rate works well in the proxy calculation is that the tail scenarios we focus on are scenarios with large hedging losses. Furthermore, many large hedging losses occur during periods with large stock price movement, particularly when the contracts are close to at-the-money and are close to renewal and maturity. According to the dynamic lapse model, under these circumstances, the fund values are more likely to deplete at the base lapse rate.

In Figure 3.6, we compare the losses simulated by the full nested simulation and those from the IANS procedure's proxy simulation, for both GMMB and GMAB contracts, under the dynamic lapse model.

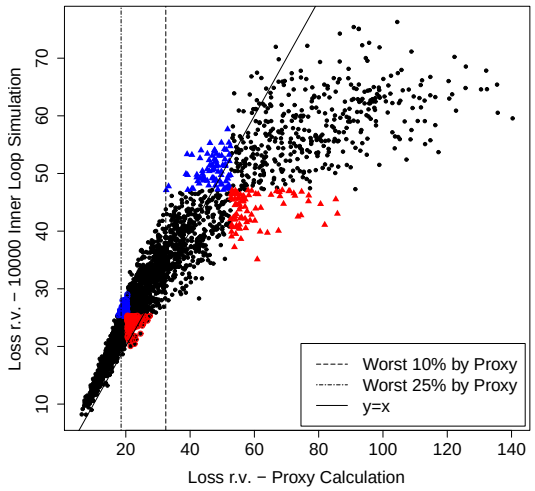
Similar to the results of the benchmark runs with no lapses, for both GMMB and GMAB contracts, and under both asset models, the $(1-\alpha)M$ tail scenarios from the nested simulations overlap almost entirely with the m proxy tail scenarios. This is also illustrated numerically in Table 3.9. Such overlapping suggest that the IANS method remains effective in this more complex setting, using the simple proxy model. Another intriguing observation in Figure 3.6 is that the simulated losses under the nested simulation and under the proxy calculation can be significantly different. Nonetheless, the rankings of the simulated losses remains similar, so the proxy model can still effectively identify the true tail scenarios. Comparison of different experiment designs under the same computational budget with the dynamic lapse assumption, similar to those illustrated in Section 3.3.2, are shown in Section 4.3 using the DIANS method in Chapter 4.

		#(%) of Correctly Identified Tail Scen.			
		RSLN		GARCH(1,1)	
CTE Level	# 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%)

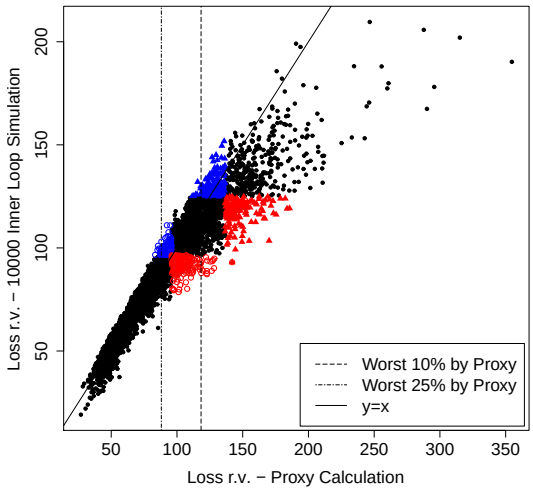
Table 3.9: Tail scenario identification by the proxy simulation (with safety margin) in Stage I of the IANS procedure under dynamic lapse assumption. Correctly identified tail scenarios are the true tail scenarios included in the safety margin in Stage 1 of the IANS procedure.

3.4 Conclusion

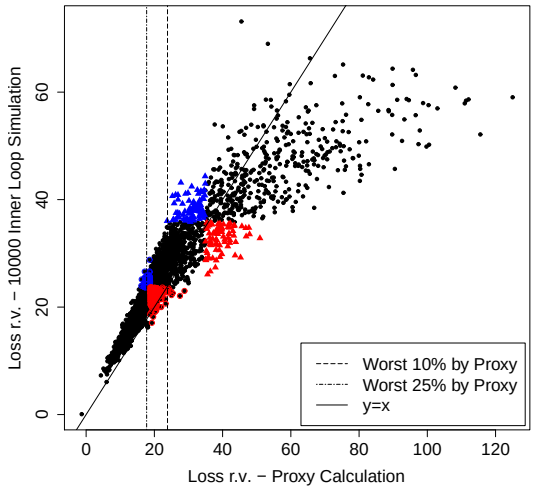
In this chapter, we illustrated a simulation procedure for estimating the CTE of loss in a VA dynamic hedging strategy. The Importance-Allocated Nested Simulation procedure takes advantage of the special structure of the CTE by first identifying a small set of potential tail scenarios, based on a proxy model of losses, calculated from a closed-form



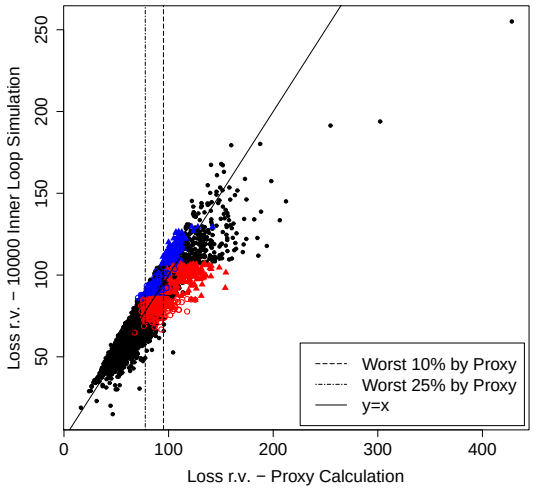
(a) GMMB RSLN



(b) GMAB RSLN



(c) GMMB GARCH(1,1)



(d) GMAB GARCH(1,1)

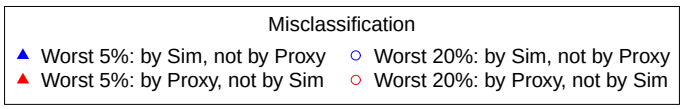


Figure 3.6: Simulated losses in 10,000 outer scenarios with dynamic lapses. The x and y coordinates of each point in the figures represent the loss in a scenario, simulated by the IANS proxy simulation and by the true nested simulation, respectively.

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 procedure compared to a standard nested simulation.

One area for improvement in the IANS method is the choice for the size of the proxy tail scenario set. In this chapter, an arbitrary selection of m is used. In the next chapter, we will consider a more rigorous and systematic approach in selecting m, the size of the proxy tail scenario set to be considered for nested simulations.

Chapter 4

Dynamic Importance-Allocated Nested Simulation

4.1 Introduction

In this chapter we introduce a dynamic importance allocated nested simulation (DIANS) methodology, which is an extension of the IANS method. In Chapter 3, a fixed margin of $5\%M$ was used to determine the number of proxy scenarios to include in the proxy tail scenario set. In this chapter, we remove that arbitrary margin, and replace it with a dynamic algorithm, based on the relationship between the proxy model and inner simulation rankings for a trial subset of scenarios; if the relationship is not sufficiently close, the subset of scenarios assigned to the tail set is increased iteratively. The closeness of rankings is measured using the empirical copula to generate moments of concomitant of order statistics (David, 1973) for the proxy model in relation to the inner simulation model. The dynamic methodology not only reduces the need for subjective input, it also provides a measure for assessing the performance of the extrinsic proxy. If the iterative process indicates a need to incorporate a very large number of scenarios into the tail set, that signals that the proxy is not performing adequately, reducing the reliance on backtesting.

4.2 Dynamic Importance Allocated Nested Simulation (DIANS)

The DIANS procedure is similar to the IANS procedure except for the determination of m, the size of the proxy tail scenario set. In the IANS procedure, the proxy is successful if the ranking of scenarios based on the proxy model corresponds closely with the true ranking of the scenarios, which we estimate using simulation. A natural way to explore how close the rankings are is through the empirical copulas of the bivariate random variables comprised of the proxy loss and the simulated loss.

Consider the bivariate random variable (L_i^P, L_i), representing the proxy loss and the random loss generated by the ith scenario, where $\mathbf{X}^{(i)} \in \mathcal{T}_m^P$. Let (U_i^P, U_i) represent the uniform random variables generated by applying the marginal distribution functions to L_i^P and L_i, that is

$$(U_i^P, U_i) = \big(F_{L^P}(L_i^P), (F_L(L_i)\big)$$

We assume, for convenience, that the losses are continuous; it is straightforward to adapt the method for mixed distributions. If we take a random sample of m pairs of (U_i^P, U_i), and order them by the U_i^P values, from smallest to largest, we will have the ordered sample

$$(U_{(1)}^P, U_{[1]}), (U_{(2)}^P, U_{[2]}), \ldots, (U_{(m)}^P, U_{[m]})$$

where $U_{(j)}^P$ is the j-th smallest (or j-th order statistic) of the U_i^P values, and $U_{[j]}$ is known as the concomitant of the jth order statistic (David, 1973).

Let $R_{j:m}$ denote the rank of the value of U_i corresponding to the jth smallest value in a sample of m values of U_i^P (or equivalently, the rank of the value of L_i corresponding to the jth smallest value of L_i^P). Then, for example, the rank of $U_{[1]}$ amongst the U_i sample is denoted $R_{1:m}$, and, in general,

$$\big(U_{(j)}^P, U_{[j]}\big) = \big(U_{(j)}^P, U_{(R_{j:m})}\big)$$

If U^P and U are comonotonic, then $R_{j:m} = j$, and we have a perfect proxy in terms of ranking of losses. If not, then we can use results from David *et al.* (1977) and O'Connell (1974) to derive formulae for moments of $R_{j:m}$, in terms of the copula function and the copula density function of U^P and U, denoted $C(u^P, u)$ and $c(u^P, u)$ respectively. We also need the density function of the rth order statistic among an i.i.d. sample of m Uniform(0,1) random variables, which is

$$f_{j:m}(u) = \frac{m!}{(j-1)!(m-j)!} u^{j-1}(1-u)^{m-j}.$$

Proposition 4.2.1. *Let U, V denote $U(0,1)$ random variables, with joint distribution function $C(U,V)$ and joint density function $c(u,v)$. Let $R_{j:m}$ denote the rank of the concomitant of the jth order statistic of U, from a random sample of size m, and let $f_{j:m}$ denote the density function defined above.*

$$E[R_{j:m}] = 1 + m\left\{\int_0^1\left[\int_0^1 C(u,v)c(u,v)dv\right] f_{j-1:m-1}(u)du + \int_0^1\left[\int_0^1 \big(v - C(u,v)\big)c(u,v)dv\right] f_{j:m-1}(u)du\right\}$$

$$E[R_{j:m}^2] = 3E[R_{j:m}] - 2 + m(m-1)\left\{\int_0^1\left[\int_0^1 \big(C(u,v)\big)^2 c(u,v)dv\right] f_{j-2:m-2}(u)du + \int_0^1\left[\int_0^1 \big(v - C(u,v)\big)^2 c(u,v)dv\right] f_{j:m-2}(u)du + 2\int_0^1\left[\int_0^1 C(u,v)\big(v - C(u,v)\big)c(u,v)dv\right] f_{j-1:m-2}(u)du\right\}$$

The proof of Proposition 4.2.1 is included in Section 4.2.1.

We use this proposition to quantify how well the proxy works at ranking the losses. First, we set an initial value for the number of proxy tail scenarios, denoted m_0. We allocate a portion of the inner simulation budget to run, say, N_0 inner simulations for each scenario in $\mathcal{T}_{m_0}^P$. We then assess the closeness of the ranking of proxy losses and simulated losses for scenarios in $\mathcal{T}_{m_0}^P$, using the moments of rank of the concomitant for a specific order statistic. If the test (described below) is not satisfied, then we increase the number of scenarios in the set and apply N_0 inner simulations to the newly added proxy tail scenarios. If the proxy is working, our process will cease after a few iterations, and the remaining inner simulation budget is applied to the final proxy tail scenario set $\mathcal{T}_{\widetilde{m}}^P$. If the iterations continue, creating tail scenario sets that are larger than a prescribed maximum, this will signal that the proxy is inadequate.

The test for increasing the sample size, or not, at each iteration proceeds as follows. Assume that there are m_k scenarios in the proxy tail scenario set on the kth iteration. We

have preliminary inner simulation results for each scenario, from which we can construct the empirical copula and copula density functions.

From these we can calculate the mean and standard deviation of the rank of a simulated loss concomitant to any given ranked proxy loss. We choose to consider the $d = m_0 - (1-\alpha)M$th ranked proxy loss (this value stays the same through the iterations of m_k). For the initial iteration of the process, the number of scenarios in the tail scenario set is $m_0 = (1-\alpha)M + d$, so d represents the additional scenarios included beyond the minimum of $(1-\alpha)M$ in the initial iteration.

We use the mean and standard deviation of the rank of concomitant of the dth ranked proxy loss to calculate a one-sided upper 95% bound for the concomitant rank,

$$b_k = E[R_{d:m_k}] + 1.645\sqrt{V[R_{d:m_k}]}. \tag{4.1}$$

If this upper bound is greater than $m_k - (1-\alpha)M$, then we increase the sample size, to

$$m_{k+1} = (1-\alpha)M + b_k,$$

and repeat the test.

Note that Equation (4.1) effectively assumes that $R_{d:m_k}$ has a normal distribution. Even though this is a loose assumption, we will demonstrate in Section 4.3 that the upper bound in Equation (4.1) is very effective and reasonably precise in capturing true tail scenarios. An alternative to making the normality assumption is to use the Cantelli's inequality to build a wider upper bound for $R_{d:m_k}$. More specifically, based on Cantelli's inequality, we have

$$Pr\left[R_{d:m_k} \leq E[R_{d:m_k}] + \sqrt{\frac{0.95}{1-0.95} \times V[R_{d:m_k}]}\right] \geq 0.95 \tag{4.2}$$

Hence we can use an alternative one-sided upper 95% bound for the concomitant rank

$$b_k^{\text{Cantelli}} = E[R_{d:m_k}] + \sqrt{19 \times V[R_{d:m_k}]}. \tag{4.3}$$

With the wider upper bound in Equation (4.3), the DIANS procedure will include many more proxy tail scenarios in nested simulation. Under a fixed computation budget, this would dilute the simulation budget for each proxy tail scenarios and result in less accurate estimate of CTEs. We will demonstrate this in numerical examples in Section 4.3.

Algorithm 3 describes the Dynamic Importance-Allocated Nested Simulation procedure.

Algorithm 3: Dynamic proxy tail scenario selection procedure.

Input: $\boldsymbol{X}^{(i)}, i = 1, \ldots, M$: all outer scenarios.
$m_0 \geq (1-\alpha)M$: initial number of scenarios in proxy tail scenario set.
$d = m_0 - (1-\alpha)M$: initial threshold scenario ranking for CTE calculation.
$m^{\max}$: max. number of proxy tail scenarios for an acceptable proxy model.
Γ : overall simulation budget.
N_0 : number of inner simulations in pilot runs.
$h^P(\boldsymbol{X})$: function that returns the proxy loss of an outer scenario $\boldsymbol{X}$.

Calculate the proxy losses $L_i^P = h^P(\boldsymbol{X}^{(i)})$ for all scenarios $\boldsymbol{X}^{(i)}$, $i = 1, \ldots, M$.
Identify the proxy tail scenario set $\mathcal{T}_{m_0}^P$ corresponding to the m_0 largest proxy losses.
Set $m_{-1} = 0$, $\mathcal{T}_{m_{-1}}^P = \emptyset$, and $k \leftarrow 0$.
while $m_k > m_{k-1}$ **do**
    for $\boldsymbol{X}^{(i)} \in \mathcal{T}_{m_k}^P \setminus \mathcal{T}_{m_{k-1}}^P$ **do**
        Invoke the multi-period inner simulation procedure for scenario $\boldsymbol{X}^{(i)}$, using N_0 independent replications at every time step of size h.
        Pair the simulated loss $\widehat{L}_i$ with the proxy loss L_i^P for $(L_i^P, \widehat{L}_i)$.
    end
    Convert the pairs $(L_i^P, \widehat{L}_i)$ to (U_i^P, U_i) by applying the marginal distribution function of L_i^P and $\widehat{L}_i$ respectively, for all $\boldsymbol{X}^{(i)} \in \mathcal{T}_{m_k}^P$; sort the pairs in ascending order of U_i^P's to get $(U_{(j)}^P, U_{[j]})$, $j = 1, \ldots, m_k$.
    Calibrate the empirical copula $C(u^P, u)$ and copula density function $c(u^P, u)$ using (U_i^P, U_i), for all $\boldsymbol{X}^{(i)} \in \mathcal{T}_{m_k}^P$.
    Calculate an approximate upper 95% bound for $R_{d:m_k}$ as in Equation (4.1), i.e., $b_k = E[R_{d:m_k}] + 1.645\sqrt{V[R_{d:m_k}]}$.
    Calculate the required number of proxy tail scenarios $m_{k+1} = (1-\alpha)M + b_k$
    if $m_{k+1} \geq m^{\max}$ **then**
        Stop. Proxy model is inadequate.
    end
    $k \leftarrow k + 1$
end
Return proxy tail scenario set $\mathcal{T}_{\widetilde{m}}^P = \mathcal{T}_{m_k}^P$, simulated losses $\widehat{L}_i$ for all $\boldsymbol{X}^{(i)} \in \mathcal{T}_{\widetilde{m}}^P$, and remaining simulation budget $\Gamma' = \Gamma - \widetilde{m} \times N_0 \times \frac{T/h \times (T/h+1)}{2}$.
for $\boldsymbol{X}^{(i)} \in \mathcal{T}_{\widetilde{m}}^P$ **do**
    Invoke the multi-period inner simulation procedure for scenario $\boldsymbol{X}^{(i)}$, using $N' = \lfloor \frac{2\Gamma'}{T/h \times (T/h+1) \times \widetilde{m}} \rfloor$ inner simulations.
    Store simulated loss $\widehat{L}_i'$ scenario $\boldsymbol{X}^{(i)}$.
    Update $\widehat{L}_i := \frac{N_0}{N_0+N'}\widehat{L}_i + \frac{N'}{N'+N_0}\widehat{L}_i'$
end
Sort the $\widetilde{m}$ simulated losses in ascending order to give $\widehat{L}_{(1)} \leq \widehat{L}_{(2)} \leq \cdots \leq \widehat{L}_{(\widetilde{m})}$.
Estimate the CTE_α of the loss by $\widehat{CTE}_\alpha = \frac{1}{(1-\alpha)M} \sum_{i=\widetilde{m}-(1-\alpha)M+1}^{\widetilde{m}} \widehat{L}_{(i)}$.

The choice of $d = m_0 - (1-\alpha)M$ in the test for adequacy of $\mathcal{T}^P_{m_k}$ is a convenient heuristic. Intuitively, (assuming positive correlation between proxy and true losses), we might prefer to use the lowest available order statistic (the minimum for which the mean and variance of the rank of the concomitant can be calculated is $d = 3$), but this will be more unstable, due to the higher uncertainty at the boundary of the empirical copula. In the numerical experiments illustrated in Section 4.3, we chose $d = 150$, which is $0.03M$, as the initial and minimum margin.

The initial number of inner simulations, N_0, is a design variable that needs to be chosen carefully. If N_0 is too low, then the simulated losses will be subject to greater sampling variability, leading to greater variability in $R_{j:m}$. This will tend to generate a higher number of tail scenarios (m_k), which is wasteful of the inner simulation budget. On the other hand, if N_0 is too high, then we may run out of computation budget. One approach is to set N_0 to be the minimum number of inner simulations required for an adequate assessment of the losses, which will depend on the nature and moneyness of the embedded option.

Note from Line 14 of the algorithm that it may stop without calculating the CTE, if the number of proxy tail scenarios selected exceeds $m^{\max}$. A valuable feature of the algorithm is that it signals when the proxy and the simulated losses have diverged such that the number of proxy tail scenarios required to capture the true tail scenarios would be large, meaning that the inner simulation budget would be spread too thinly for accurate evaluation of the tail risk measure. An example is shown in the following section. Setting $m^{\max} = M$ removes the stopping point, so that, if necessary, the algorithm continues until all M scenarios are included in the inner simulation set. In this case, the DIANS procedure becomes a standard nested simulation.

Another useful feature of the algorithm is each round of iteration provides the sample size for the next round. Even though statistics such as Spearman's rho or Kendall's tau also give an indication of the level of rank dependency, they provide no guidance on the sample size increment, nor do they offer any objective criteria for when the iteration could stop.

The algorithm specifically targets the CTE_α estimate, but it can be easily adapted to other tail risk measures such as the VaR_α. The only change required in the algorithm is to replace the CTE estimator in Line 25 by a VaR estimator such as the one proposed by Hyndman and Fan (1996),

$$\widehat{VaR}_\alpha = (1-\gamma)\widehat{L}_{(\widetilde{m}-(M-g))} + \gamma\widehat{L}_{(\widetilde{m}-(M-g)+1)} \tag{4.4}$$

where $g = \lfloor (M+\frac{1}{3})\alpha + \frac{1}{3} \rfloor$ and $\gamma = (M+\frac{1}{3})\alpha + \frac{1}{3} - g$. See Kim and Hardy (2007) or Risk and Ludkovski (2018) for a fuller account of bias reduction in VaR estimation.

4.2.1 Proof of Proposition 4.2.1

For a general bivariate distribution of (U, V), from David *et al.* (1977) we have

$$E[R_{j:m}] = 1 + m\left(\int_{\infty}^{-\infty} \left[\int_{\infty}^{-\infty} \theta_1 f(v|u)dv \right] f_{U_{j-1:m-1}}(u)du + \int_{\infty}^{-\infty} \left[\int_{\infty}^{-\infty} \theta_3 f(v|u)dv \right] f_{U_{j:m-1}}(u)du \right) \tag{4.5}$$

where

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

If (U, V) has a bivariate uniform distribution, we have

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

where

$C(u, v)$ is the copula function of $U = u$ and $V = v$.

$c(u, v)$ is the density function of $C(U, V)$.

$f_{U_{j:m}}(u)$ represents the density function of the jth order statistic among m U's:

$$f_{U_{j:m}}(u) = \frac{m!}{(j-1)!(m-j)!} u^{j-1}(1-u)^{m-j}.$$

Therefore, in this case Equation (4.5) is equivalent to

$$\begin{aligned} &E[R_{j:m}] \\ &= 1 + m\left(\int_0^1 \left[\int_0^1 C(u,v)c(u,v)dv \right] f_{U_{j-1:m-1}}(u)du + \int_0^1 \left[\int_0^1 \big(v - C(u,v)\big)c(u,v)dv \right] f_{U_{j:m-1}}(u)du \right) \end{aligned} \tag{4.6}$$

To derive the second moment of $R_{j:m}$, we first derive the second moment of $R_{j:m}$ for a general bivariate pair, (X, Y), i.e. not specifically bivariate uniform distributed. We use the same factorial moment method as in O'Connell (1974).

First we have,

$$\begin{aligned} E[R_{j:m}^2] =& \sum_{s=1}^{m} s^2 P[R_{j:m} = s] \\ =& \sum_{s=0}^{m-1} (s+1)^2 P[R_{j:m} = s+1] \\ =& \sum_{s=0}^{m-1} P[R_{j:m} = s+1] + 2\sum_{s=0}^{m-1} sP[R_{j:m} = s+1] + \sum_{s=0}^{m-1} s^2 P[R_{j:m} = s+1] \\ =& 1 + 2(E[R_{j:m}] - 1) + \sum_{s=0}^{m-1} s^2 P[R_{j:m} = s+1] \end{aligned} \tag{4.7}$$

More specifically,

$$\sum_{s=0}^{m-1} s^2 P[R_{j:m} = s+1] = \sum_{s=0}^{m-1} s^2 m \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sum_{k=0}^{t} C_k \theta_1^k \theta_2^{j-1-k} \theta_3^{s-k} \theta_4^{m-j-s+k} f(x,y)dxdy \tag{4.8}$$

where $C_k = \binom{m-1}{j-1}\binom{j-1}{k}\binom{m-j}{s-k}$

Let $i = s - k$, then

$$\begin{aligned} & \sum_{s=0}^{m-1} s^2 P[R_{j:m} = s+1] \\ =& \sum_{s=0}^{m-1} s^2 m \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sum_{k=0}^{t} C_k \theta_1^k \theta_2^{j-1-k} \theta_3^{s-k} \theta_4^{m-j-s+k} f(x,y)dxdy \\ =& n\binom{m-1}{j-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sum_{k=0}^{j-1}\sum_{i=0}^{m-j} (k+i)^2 \binom{j-1}{k}\binom{m-j}{i} \theta_1^k \theta_2^{j-1-k} \theta_3^{i} \theta_4^{m-j-i} f(x,y)dxdy \\ =& m\binom{m-1}{j-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sum_{k=0}^{j-1}\sum_{i=0}^{m-j} (k^2+i^2+2ki) \binom{j-1}{k}\binom{m-j}{i} \theta_1^k \theta_2^{j-1-k} \theta_3^{i} \theta_4^{m-j-i} f(x,y)dxdy \end{aligned} \tag{4.9}$$

Furthermore,

$$m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}\sum_{i=0}^{m-j}k^2\binom{j-1}{k}\binom{m-j}{i}\theta_1^k\theta_2^{j-1-k}\theta_3^i\theta_4^{m-j-i}f(x,y)dxdy \qquad (4.10)$$

$$\begin{aligned}
= \quad & m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}\sum_{i=0}^{m-j}k(k-1)\binom{j-1}{k}\binom{m-j}{i}\theta_1^k\theta_2^{j-1-k}\theta_3^i\theta_4^{m-j-i}f(x,y)dxd \\
& + m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}\sum_{i=0}^{m-j}k\binom{j-1}{k}\binom{m-j}{i}\theta_1^k\theta_2^{j-1-k}\theta_3^i\theta_4^{m-j-i}f(x,y)dxdy \\
= \quad & m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}k(k-1)\binom{j-1}{k}\theta_1^k\theta_2^{j-1-k}\sum_{i=0}^{m-j}\binom{m-j}{i}\theta_3^i\theta_4^{m-j-i}f(x,y)dxdy \\
& + m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}k\binom{j-1}{k}\theta_1^k\theta_2^{j-1-k}\sum_{i=0}^{m-j}\binom{m-j}{i}\theta_3^i\theta_4^{m-j-i}f(x,y)dxdy \\
= \quad & m\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\binom{m-1}{j-1}(j-1)(j-2)\theta_1^2[F_X(x)]^{j-3}[1-F_X(x)]^{m-j}f(x,y)dxdy \\
& + m\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\binom{m-1}{j-1}(j-1)\theta_1[F_X(x)]^{j-2}[1-F_X(x)]^{m-j}f(x,y)dxdy \\
= \quad & m(m-1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1^2\frac{f(x,y)}{f_X(x)}f_{j-2:m-2}(x)dxdy + m\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1\frac{f(x,y)}{f_X(x)}f_{j-1:m-1}(x)dxdy \\
= \quad & m(m-1)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1^2f(y|x)dy\right]f_{j-2:m-2}(x)dx + m\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1f(y|x)dy\right]dy]f_{j-1:m-1}(x)dx
\end{aligned}$$

Similarly,

$$m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}\sum_{i=0}^{m-j} i^2\binom{j-1}{k}\binom{m-j}{i}\theta_1^k\theta_2^{j-1-k}\theta_3^i\theta_4^{m-j-i}f(x,y)dxdy \tag{4.11}$$

$$= m(m-1)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_3^2 f(y|x)dy\right]f_{j:m-2}(x)dx + m\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_3 f(y|x)dy\right]dy]f_{j:m-1}(x)dx$$

And

$$m\binom{m-1}{j-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{k=0}^{j-1}\sum_{i=0}^{m-j} ki\binom{j-1}{k}\binom{m-j}{i}\theta_1^k\theta_2^{j-1-k}\theta_3^i\theta_4^{m-j-i}f(x,y)dxdy \tag{4.12}$$

$$= m(m-1)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1\theta_3 f(y|x)dy\right]f_{j-1:m-2}(x)dx$$

Substitute (4.10), (4.11) and (4.12) back in (4.9), we have

$$\sum_{s=0}^{m-1} s^2 P[R_{j:m}=s+1] \tag{4.13}$$

$$= m(m-1)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1^2 f(y|x)dy\right]f_{j-2:m-2}(x)dx + m\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1 f(y|x)dy\right]dy]f_{j-1:m-1}(x)dx$$
$$+ m(m-1)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_3^2 f(y|x)dy\right]f_{j:m-2}(x)dx + m\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_3 f(y|x)dy\right]dy]f_{j:m-1}(x)dx$$
$$+ 2m(m-1)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1\theta_3 f(y|x)dy\right]f_{j-1:m-2}(x)dx$$

$$= m(m-1)\Big(\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1^2 f(y|x)dy\right]f_{j-2:m-2}(x)dx + \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_3^2 f(y|x)dy\right]f_{j:m-2}(x)dx$$
$$+ 2\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\theta_1\theta_3 f(y|x)dy\right]f_{j-1:m-2}(x)dx\Big) + E[R_{r:m}] - 1$$

Substitute (4.13) back in (4.7), we have

$$E[R_{j:m}^2] = 3E[R_{j:m}] - 2 + m(m-1) \times \Big(\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1^2 f(y|x) dy \right] f_{j-2:m-2}(x) dx \tag{4.14}$$
$$+ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_3^2 f(y|x) dy \right] f_{j:m-2}(x) dx + 2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \theta_1 \theta_3 f(y|x) dy \right] f_{j-1:m-2}(x) dx \Big)$$

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

$$E[R_{j:m}^2] = 3E[R_{j:m}] - 2 + m(m-1) \times \Big(\int_0^1 \left[\int_0^1 \big(C(u,v)\big)^2 c(u,v) dv \right] f_{U_{j-2:m-2}}(u) du$$
$$+ \int_0^1 \left[\int_0^1 \big(v - C(u,v)\big)^2 c(u,v) dv \right] f_{U_{j:m-2}}(u) du$$
$$+ 2 \int_0^1 \left[\int_0^1 C(u,v)\big(v - C(u,v)\big) c(u,v) dv \right] f_{U_{j-1:m-2}}(u) du \Big)$$

4.3 Numerical Experiments

4.3.1 Setup and assumptions

We illustrate the DIANS procedure by applying it to estimate the $\text{CTE}_{95\%}$ of the hedging costs for a GMMB and GMAB contract under a Markov regime-switching lognormal asset model, with a dynamic lapse assumption.

Both the GMMB and GMAB contracts are 240-month, single premium policies. The premium is 1,000. The risk is managed using delta hedging, rebalanced at monthly intervals.

In the GMMB example, the contract has a guaranteed return of premium at maturity. A gross management fee of 1.75% per annum is deducted monthly from the sub-account value, of which 0.30% per annum is returned as net fee income for the insurer. The

remaining fee of 1.45% per annum is assumed to pay for expenses that are not modeled explicitly.

In the GMAB example, the contract has a renewal in 120-months' time, and matures in 240 months. A gross management fee of 2.00% per annum is deducted monthly from the sub-account value, of which 0.60% per annum is counted as net fee income for the insurer. The remaining fee of 1.40% per annum is assumed to pay for expenses that are not modeled explicitly.

To simplify the presentation, we assume that there are no transactions costs and we ignore mortality.

Returns on the sub-account funds, under the real world measure, are modelled as a 2-regime lognormal process with monthly time steps. The model parameters are given in Table 3.2 in Section 3.3. The dynamic lapse behaviour of policyholders is modelled using the NAIC formula described in Section 3.3.3.

The proxy liabilities are calculated using the Black-Scholes put option formula, so the proxy model assumes geometric Brownian motion for the stock return process with volatility recalibrated at each time t, depending on $X_t^{(i)}$. Under the proxy model, lapse rates are assumed to be constant, equal to the base rates of the dynamic lapse rate model.

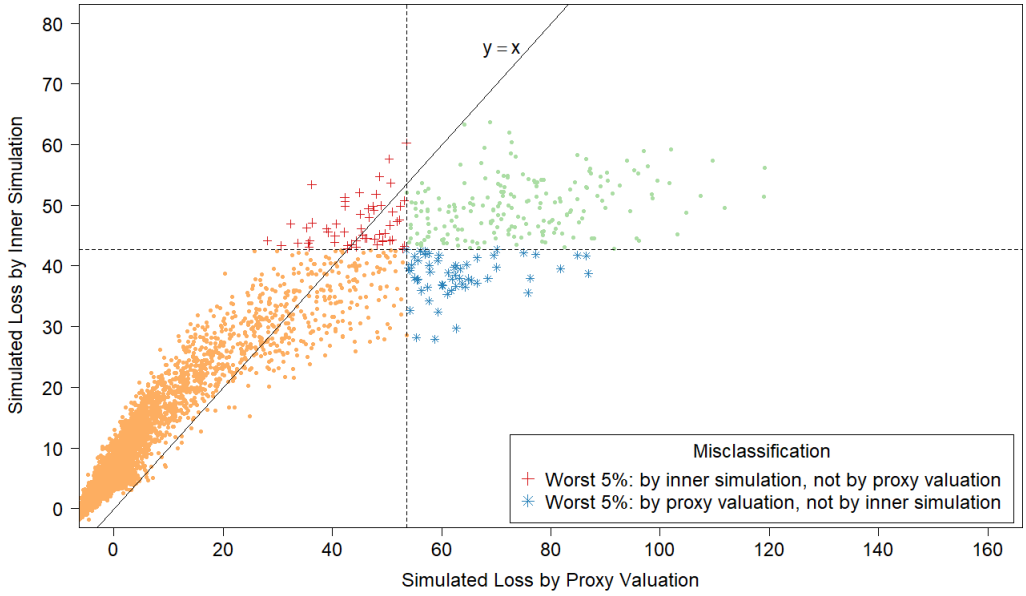
4.3.2 Proxy losses versus true losses

We conduct a large scale, full uniform nested simulation as a benchmark, against which we will compare the results of the DIANS method.

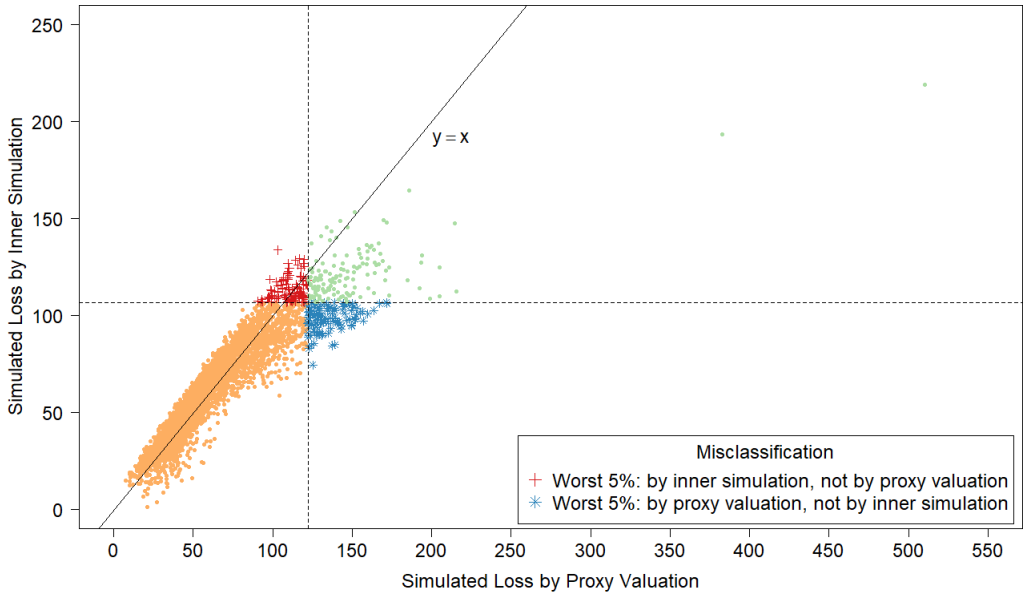
We use 5,000 outer scenarios, with 10,000 inner simulations at each time step of each scenario. We assume (after some testing) that this is sufficient to give a very accurate evaluation of the loss for each scenario $\boldsymbol{X}^{(i)}$, so for convenience, we will designate these the "true" losses associated with each $\boldsymbol{X}^{(i)}$, denoted L_i.

We also apply the proxy model to each of the 5,000 scenarios, generating proxy losses, L_i^P. In Figure 4.1 we show the proxy losses (x-axis) plotted against the true losses (y-axis).

We assume that we are interested in the 95% CTE, which involves the largest 250 losses from the 5,000 scenarios. The "+"'s in Figure 4.1 represent $\mathcal{T}_{250} \setminus \mathcal{T}_{250}^P$, that is, the losses that are ranked in the top 5% of the L_i, but are not in the top 5% of the proxy loss estimates. The "$*$"'s represent $\mathcal{T}_{250}^P \setminus \mathcal{T}_{250}$. Losses that lie on the right of the vertical line correspond to the worst 5% proxy losses, while losses that lie above the horizontal line correspond to the worst 5% true losses.



(a) GMMB



(b) GMAB

Figure 4.1: Simulated losses in 5,000 outer scenarios, by proxy valuation (x axis) and by inner simulation (y axis). Region above the horizontal line indicates the worst 5% loss by inner simulation. Region to the right of the vertical line indicates the worst 5% loss by proxy valuations.

In Figure 4.2 we show the P-P plots of the simulated losses presented in Figure 4.1. It suggests that the ranking of losses from inner simulation are reasonably well correlated to the ranking of losses from the proxy model.

We note from Figure 4.1 and Figure 4.2 that although the proxy model produces similar ranking of losses to the 'true' losses, the actual values of the losses produced by the proxy model are very different to the accurate inner simulation model losses; the points in Figure 4.1 deviate significantly from the $y = x$ line in each plot. This means that we cannot simply use the proxy estimates of loss in the risk measure – we must proceed to the inner simulation step of the algorithm.

In addition, Figure 4.3 illustrates the final empirical copula used in applying the DIANS procedure to the 5,000 scenarios presented in Figure 4.1. The empirical copula suggests the proxy and inner simulation losses within the proxy tail scenarios set $\mathcal{T}^P_{\widetilde{m}}$ are also fairly well correlated, with the correlation in the empirical copula of the GMMB being stronger than that of GMAB. This has an impact on the number of proxy tail scenarios identified by the DIANS algorithm for the two different types of contracts, as we will see in Section 4.3.4.

4.3.3 Identifying $\mathcal{T}^P_{\widetilde{m}}$

We explore the variables m^* and $\widetilde{m}$, where $\mathcal{T}^P_{m^*}$ is the smallest set of proxy tail scenarios containing $\mathcal{T}_{250}$, and $\widetilde{m}$ is the number of proxy tail scenarios identified by the DIANS algorithm. To do this, we run 20 repetitions of the DIANS algorithm, each using the same set of $M = 5,000$ scenarios as used in the full uniform nested simulation, and each with the following input parameters:

$$\Gamma = 5,000 \times 200 = 10^6,\ N_0 = 1000,\ m_0 = 400,\ m^{\max} = 5,000,\ d = 150$$

Note that we have set $m^{\max} = M$, which means that we have allowed the algorithm to continue to find an unconstrained value of $\widetilde{m}$. In practice, $m^{\max}$ is a design variable that the user of the DIANS procedure would choose, based on the minimum acceptable number of inner simulations.

From the full scale uniform nested simulation, we know that the minimum number of proxy tail scenarios required to capture all the true tail scenarios, $m^* = \min\{m : \mathcal{T}_{250} \subset \mathcal{T}^P_m\}$, is $m^* = 557$ in the GMMB example and $m^* = 787$ in the GMAB example.

From the DIANS algorithm, for each repetition we record $\widetilde{m}$, which is the final number of scenarios in the proxy tail scenario set.

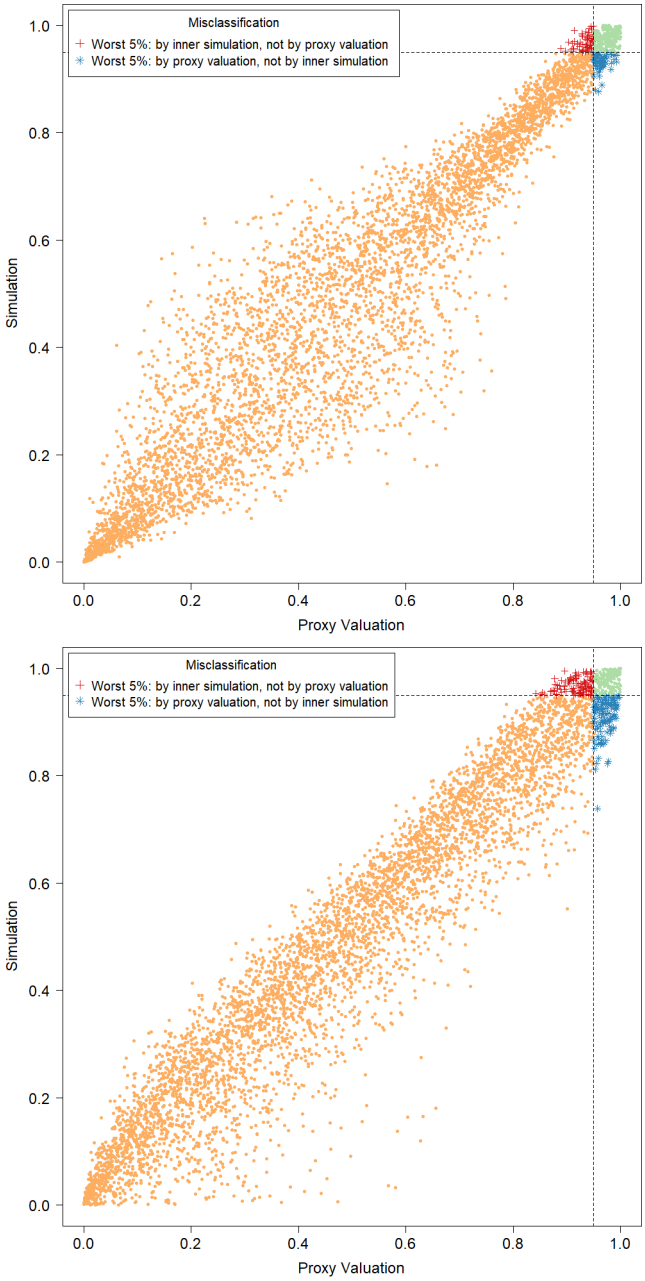
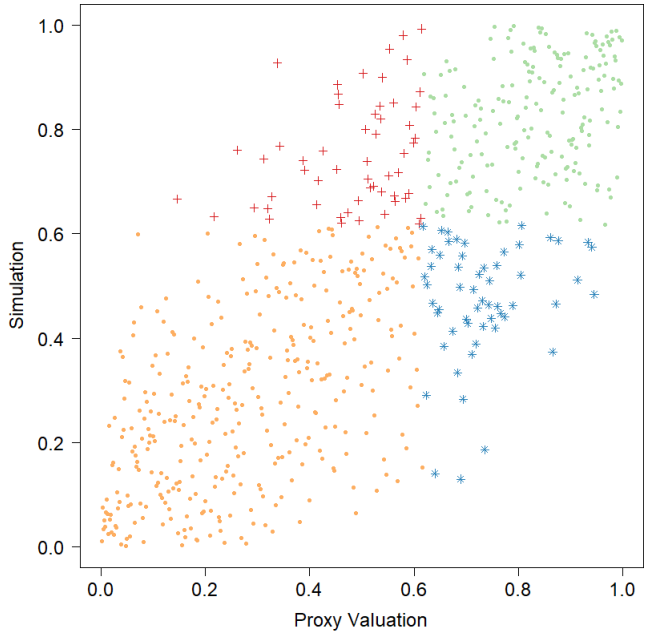
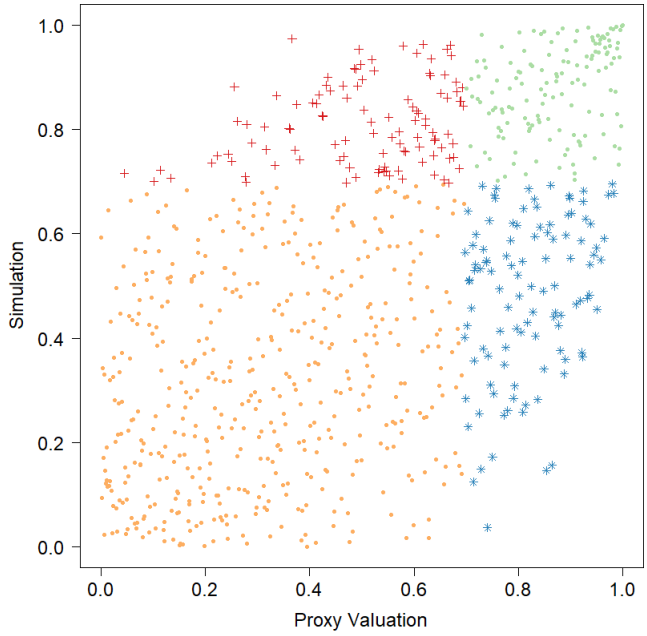


Figure 4.2: P-P plots of the simulated loss cumulative distribution functions in 5,000 outer scenarios, by proxy valuation (x axis) and by inner simulation (y axis); GMMB (top figure) and GMAB (bottom figure). The vertical and horizontal line represent the respective 95% quantile on the x and y axis.



(a) GMMB



(b) GMAB

Figure 4.3: Empirical copula of simulated losses within the proxy tail scenarios set $\mathcal{T}_{\tilde{m}}^{P}$. The same legends as in Figure 4.1 are used.

The results are illustrated in Figure 4.4. Each column represents a separate repetition of the DIANS valuation. In each column, the triangle represents $\widetilde{m}$, which is the number of scenarios included in the tail scenario set using DIANS. The dots (which are the same in each column) represent the quantities $M - R_{j:M}$, for $M = 5000$, and $j = 4751, 4752, \ldots, 5000$. Here, $R_{j,M}$ is the concomitant rank of the jth true tail losses, so $M - R_{j:M}$ indicates the number of proxy tail scenarios required to capture the top $M - j$ true tail scenarios. The maximum value of $M - R_{j:M}$ (i.e. the top dot in each column) is m^*, which is the number of proxy tail scenarios required to capture the scenarios generating the top 5% of true losses.

We see from the figure that $\widetilde{m}$ remains relatively stable across the experiments. We also see that for both the GMMB and the GMAB, in each of the 20 experiments the threshold generated by the DIANS algorithm (the triangle) lies above the maximum required to capture all the true tail scenarios (represented by the uppermost dot), meaning that the algorithm generated a proxy tail set that included all the true tail scenarios. On the one hand, this is encouraging - the algorithm, here, does a good job of capturing all the true tail scenarios. On the other hand, in the GMMB case, the number of proxy tail scenarios selected is significantly greater than the number required to capture the true tail scenarios, signalling that we might be wasting computational effort. There is a trade-off here, between ensuring that the true tail scenarios are all captured, and ensuring that the inner simulation budget is sufficiently concentrated to give reliable results. The balance between these competing objectives can be adjusted by increasing or decreasing the confidence level used for the bound in Algorithm 3.

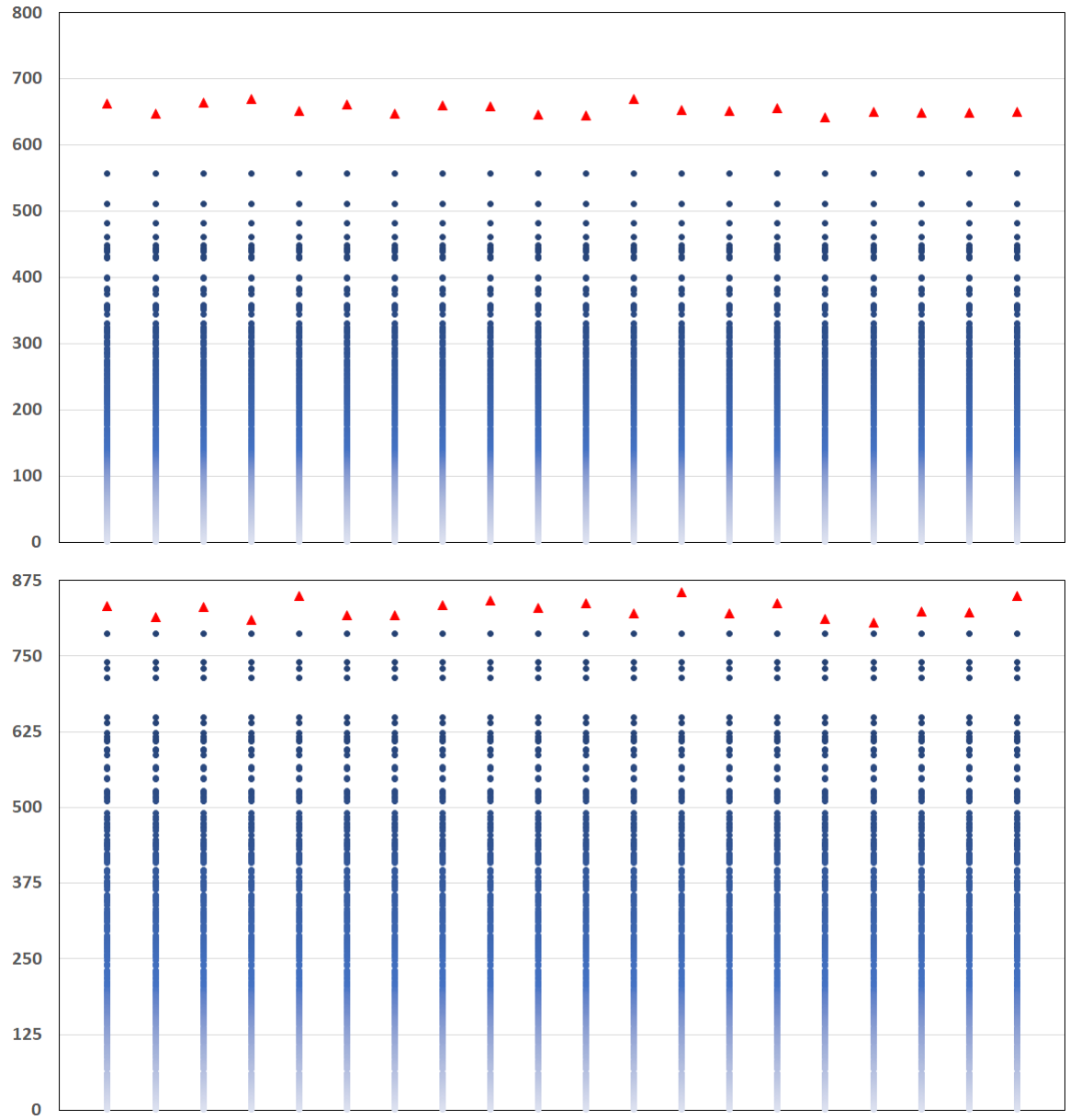


Figure 4.4: Actual *inverse* rank of concomitant of true tail losses and $\widetilde{m}$ (threshold generated by DIANS), for 20 repeated experiments described in Section 4.3.3; GMMB (top), GMAB (bottom).

4.3.4 CTE Estimation

In this section we compare $\text{CTE}_{95\%}$ estimates of the losses for both the GMMB and the GMAB described in Section 4.3.1. We fix the inner simulation budget, and apply the following estimation methods:

(a) DIANS, as described in Algorithm 3, parameters as in the previous section.
(b) DIANS, as in (a) but using the upper bound based on Cantelli's inequality in Equation (4.3).
(c) Fixed (non-dynamic) importance allocation nested simulation (See Chapter 3) with
  (c1) $m = 0.15 \times M = 750$.
  (c2) $m = 0.10 \times M = 500$.
  (c3) $m = 0.05 \times M = 250$.
(d) Standard nested simulation with equal number of inner simulation.

Each experiment is repeated 100 times. The outer scenarios used for each repetition of each experiment are the same, so the differences between the results arise solely from tail scenario selection, and sampling variability, at the inner simulation stage. The scenarios are also the same as those used for the large scale nested simulations illustrated in Figure 4.1, which were used to calculate the accurate CTE estimate used as the basis for the RMSE values below.

The inner simulation budget is fixed at $\Gamma = 10^6$. Using importance allocated nested simulation, with a fixed number of scenarios, m say, in the proxy tail scenario set, means that each scenario in the tail scenario set would be allocated $10^6/m$ inner simulations, at each time step of the scenario. Rather than a prescribed fixed number of proxy tail scenarios, a main improvement in DIANS is to search for the proxy tail scenarios dynamically based on concomitants. Computationally, we observe in our numerical experiments that this dynamic search algorithm takes negligible runtime (roughly the simulation runtime for 2 outer scenarios in the uniform inner simulation experiment with 200 inner simulations). So running the different experiments with the same simulation budget is a fair comparison.

The results for the GMMB experiments are summarized in Table 4.1, and in the box-and-whisker plot in Figure 4.5. The results for the GMAB experiments are summarized in Table 4.2, and are illustrated in the box-and-whisker plot in Figure 4.6.

Experiment	m	N	RMSE	Average % of True Tail Scenarios Scenarios Used in CTE Estimation
(a) Dynamic IANS, $m_0 = 400$	≈ 654	$\approx 1,528$	0.0072% (0.0010%)	96% (0.1%)
(b) DIANS (Cantelli), $m_0 = 400$	≈ 948	$\approx 1,055$	0.0130% (0.0019%)	95% (0.1%)
(c1) Fixed IANS	750	1,333	0.0088% (0.0013%)	96% (0.1%)
(c2) Fixed IANS	500	2,000	0.0069% (0.0011%)	96% (0.1%)
(c3) Fixed IANS	250	4,000	7.8604% (0.0274%)	78% (0.0%)
(d) Uniform inner simulation	5,000	200	0.2234% (0.0212%)	90% (0.1%)

Table 4.1: Results from 100 repetitions of fixed and dynamic IANS process, and standard nested simulation, GMMB example, with standard errors. All values are based on a single outer scenario set.

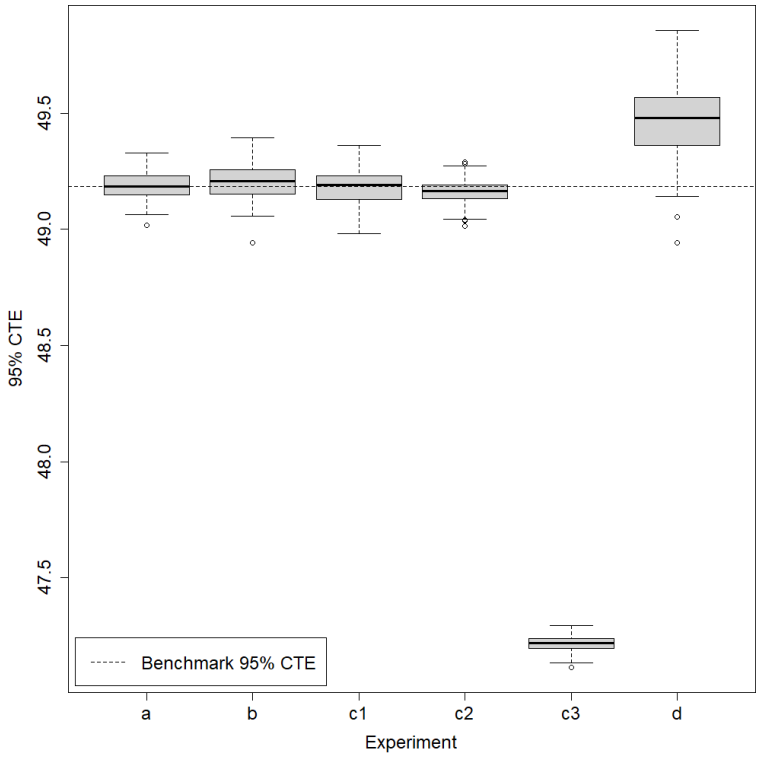


Figure 4.5: Box-and-whisker plot of results from 100 repetitions of fixed and dynamic IANS process, and standard nested simulation, GMMB example.

In the final column of Table 4.1 we show the percentage of true tail scenarios used in the $\text{CTE}_{95\%}$ estimation of the GMMB experiments. In experiments (a), (b) and (c1), although $\mathcal{T}_{\widetilde{m}}$ captured all the true tail scenarios, the ranking of the tail scenarios in each case is not identical to the benchmark run, due to inner simulation noise. As a result, only $95-96\%$ of the true tail scenarios were used in the actual CTE calculation. We note from Figure 4.1 that, close to the threshold of the top 5% of true losses, the values of the losses immediately above the threshold are very close to the losses immediately below the threshold, so a small amount of replacement, in this example, makes little difference to the CTE estimation.

In this example, the RMSE results suggest that the dynamic IANS procedure achieves significantly higher accuracy than IANS with fixed $m = 250$. The DIANS results are similar compared with IANS, with fixed $m = 750$ or $m = 500$, and significantly outperforms the uniform inner simulation method.

The RMSEs in methods (a), (b), (c1) and (c2) are similar because the minimum number of proxy tail scenarios required to capture the full set of true tail scenarios is $m^* = 557$ (for this set of $\boldsymbol{X}$). Thus, any importance allocation method with $m > 557$ would capture all the true tail scenarios, and that includes the DIANS case (m exceeded 557 in each of the 100 repetitions) and the fixed IANS case with $m = 750$. For the fixed IANS case with $m = 500 < 557$, some true tail scenarios are omitted from the inner simulation stage; from the top plot in Figure 4.4, by looking at the number of dots lying above the $y = 500$ line, we see that using 500 proxy tail scenarios will miss just two true tail scenarios. Even though the true tail scenarios are all, or almost all captured in experiments (a), (b), (c1) and (c2), the losses for the tail scenarios are estimated with different numbers of inner simulations in each of the four experiments. The difference in RMSE between experiments (a), (b) and (c1) is driven entirely by the difference in the number of inner simulations deployed to each scenario in $\mathcal{T}_m$; all methods capture all the true tail scenarios, but the DIANS method in experiment (a) does so with less redundancy, and therefore more accuracy in the loss estimation. In contrast, the DIANS method based on the Cantelli inequality in experiment (b) uses more conservative set of proxy tail scenarios, and results in less accurate loss estimation. This is illustrated in Figure 4.5, which shows that both experiments appear to generate unbiased estimators, but the variance of (c1) is a little greater than the variance of (a). Experiment (c2) misses two true tail scenarios, but achieves more accurate results for those that it does capture. Because it misses some true tail scenarios, the CTE estimate is biased low (as we can see in Figure 4.5). However, in this case, the bias is compensated by the low variance in the RMSE calculation.

In contrast, the RMSE under experiment (c3) is close to 1,000 times that of the DIANS result. Experiment (c3) uses a fixed m of only 250, allowing no cushion for losses that

are in the top 5% under the accurate calculation, but are below the top 5% by the proxy calculation. The evaluation of loss for each scenario in the proxy tail set will be more accurate, using 4,000 inner simulations, but many true tail scenarios are missed in this experiment. The missed tail scenarios are replaced with others that are lesser ranked, based on the initial simulation values, so the CTE estimate is, again, biased low – much more significantly than in (c2). Note that, comparing the result from method (c3) with the uniform nested simulation result, in method (d), we see that if the importance allocation method misses too many true tail scenarios the result is actually worse than using a uniform allocation of inner simulation under the same budget. This underscores the usefulness of using the dynamic IANS procedure to ensure sufficient tail scenario coverage, rather than a fixed IANS method. In practice, we do not know the value of m^*; the advantage of the dynamic IANS procedure is that we eliminate the subjectivity involved in selecting a fixed m.

Note that the positive bias indicated in the uniform nested simulation approach (method (d)), results from evaluating discrete hedging errors for out-of-the-money options with a small number of simulations (Boyle and Emanuel, 1980).

Experiment	m	N	RMSE	Average % of True Tail Scenarios Scenarios Used in CTE Estimation
(a) Dynamic IANS, $m_0 = 400$	≈ 826	$\approx 1,210$	0.0515% (0.0052%)	92% (0.1%)
(b) DIANS (Cantelli), $m_0 = 400$	1,000	1,000	0.0668% (0.0074%)	91% (0.1%)
(c1) Fixed IANS	750	1,333	0.0329% (0.0042%)	92% (0.1%)
(c2) Fixed IANS	500	2,000	0.1978% (0.0121%)	89% (0.1%)
(c3) Fixed IANS	250	4,000	38.1046% (0.1007%)	57% (0.0%)
(d) Uniform inner simulation	5,000	200	1.7094% (0.0849%)	82% (0.2%)

Table 4.2: Results from 100 repetitions of fixed and dynamic IANS process, and standard nested simulation, GMAB example, Standard errors in brackets.

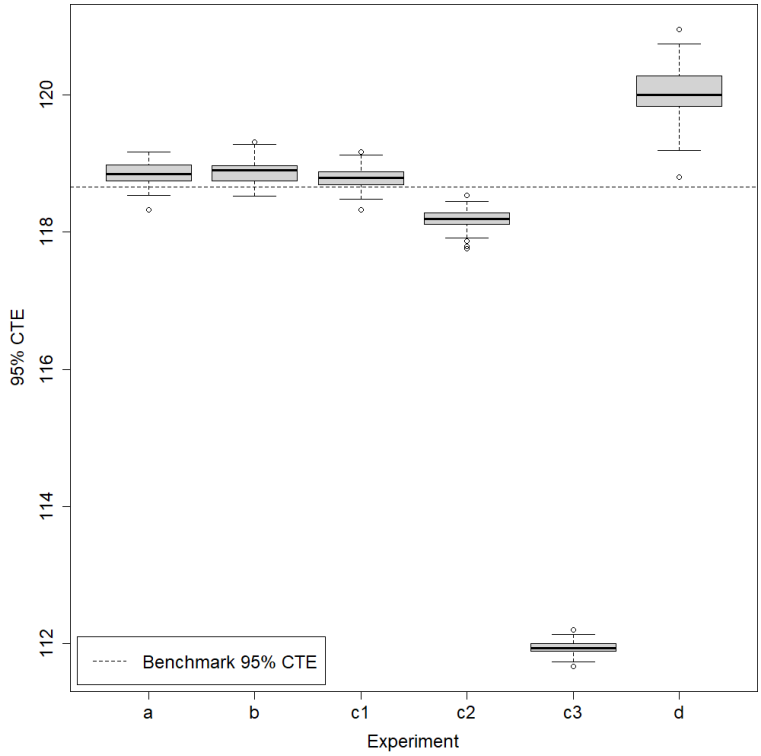


Figure 4.6: Box-and-whisker plot of results from 100 repetitions of fixed and dynamic IANS process, and standard nested simulation, GMAB example.

In the GMAB case, the greater volatility in estimated losses for each scenario means that a larger number of additional proxy tail scenarios are required to capture all 250 true tail scenarios. GMABs involve very significant gamma risk (Hardy, 2003), particularly at the renewal dates, when the delta of the option can decline sharply from positive to negative. There is, therefore, significantly more hedging error from a delta hedge of a GMAB than of the GMMB. For this set of scenarios, the number of proxy tail scenarios required to capture all 250 true tail scenarios is $m^* = 787$. The DIANS method in (a) captures all these scenarios, with an average m of 826, but with only (on average) 1210 inner simulations allocated to the valuation for each step in each scenario; this number is relatively small, leading to a small positive bias in the estimation. The DIANS method based on Cantelli's inequality for GMAB effectively becomes a fixed IANS experiment with $m = 1,000$ because in all 100 repetitions, the search procedure stops at $m = m^{\max} = 1,000$ because of the more conservative bound for rank of concomitant. The fixed IANS method with $m = 750$ captures all but one of the true tail scenarios, and has a slightly higher inner

simulation budget than the DIANS method, with the result that the RMSE is slightly better than the DIANS method. Experiments (c2) and (c3) both have values of m that are too small, missing a significant number of true tail scenarios, creating an estimate that is biased low, with much larger RMSEs. The result for the uniform nested simulation method is similar to the GMMB case.

In both GMMB and GMAB examples, we see that the DIANS method using Cantelli inequality is less efficient than the DIANS method proposed in Algorithm 3. Under the DIANS method using Cantelli inequality, the proxy tail scenario set identified is overly conservative, which results in wastage of computation budget on too many non-tail scenarios.

4.3.5 VaR Estimation

In this section, we repeat the DIANS (experiment (a)) and standard nested simulation (experiment (d)) in Section 4.3.4, but for a $\text{VaR}_{99\%}$ estimation. The purpose of these experiments is the demonstrate the gain in computation efficiency of applying the DIANS procedure in a quantile risk measure estimation further into the tail region of the loss distribution.

We apply the same experiment setting, and used the same sets of outer scenarios for GMMB and GMAB, respectively, as in Section 4.3.4. The DIANS experiment was conducted using Algorithm 3 with $\alpha = 99\%$ and the VaR estimator in Equation (4.4).

The parameters and results of the experiments are summarized in Table 4.3. The results are also illustrated in the box-and-whisker plots in Figure 4.7.

Experiment	m	N	RMSE
GMMB			
Dynamic IANS, $m_0 = 200$	≈ 324	$\approx 3,083$	0.0861% (0.0117%)
Uniform inner simulation	5,000	200	0.4316% (0.0549%)
GMAB			
Dynamic IANS, $m_0 = 200$	≈ 368	$\approx 2,773$	0.4390% (0.0507%)
Uniform inner simulation	5,000	200	2.8224% (0.2387%)

Table 4.3: 99% VaR results from 100 repetitions of dynamic IANS process and standard nested simulation. Standard error of the results indicated in bracket. All values are based on a single outer scenario set.

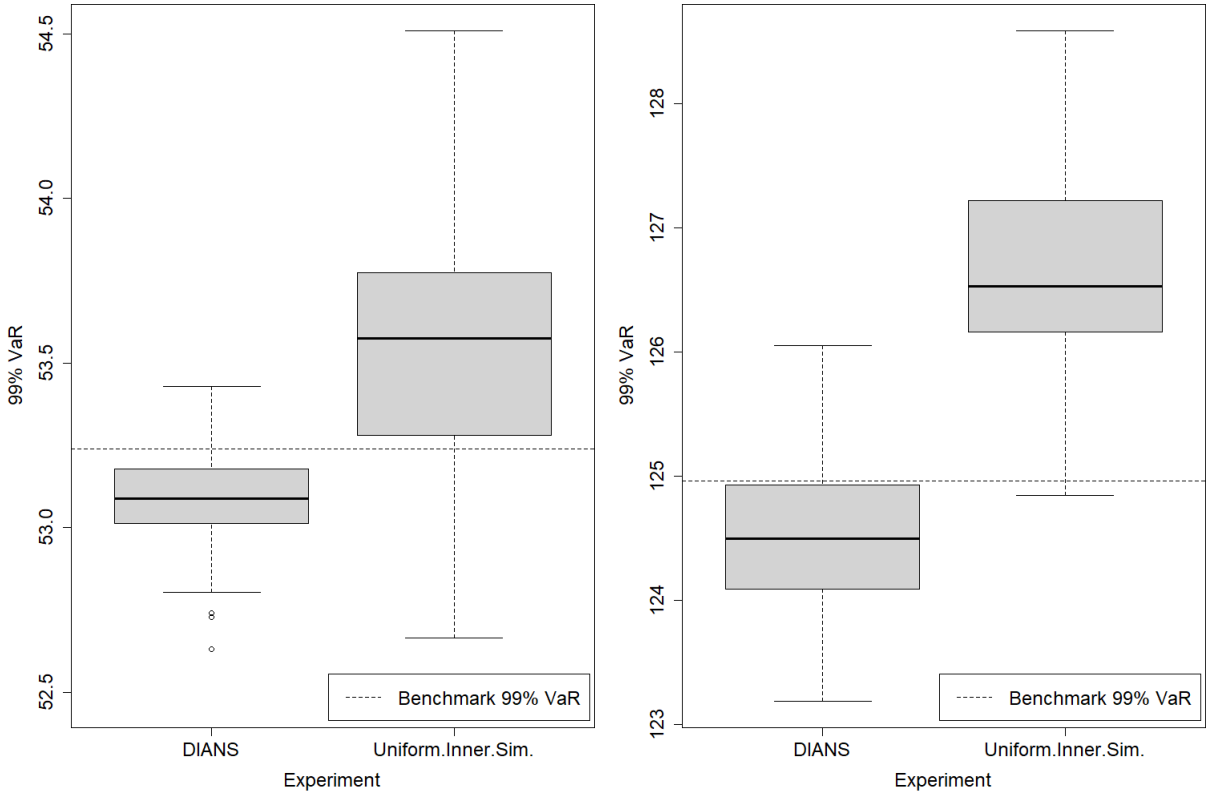


Figure 4.7: Box-and-whisker plot of 99% VaR results from 100 repetitions of dynamic IANS process, and standard nested simulation, GMMB (left) and GMAB (right) example.

Given the same computation budget, the DIANS procedure achieves an RMSE 5-6 times smaller than using the standard nested simulation method. The improvement is less significant than that observed in the CTE estimation. There are two main reasons for this. The first is that in nested simulations, as the number of inner simulation increases, the bias of the estimated CTE reduces faster than bias of the estimated VaR. See the convergence results derived in Gordy and Juneja (2010). The second reason is that there is more variance in the VaR estimate than the CTE estimate. CTE is estimated by taking the average of all relevant tail scenario loss, which smooths out some variations in losses of individual tail scenarios. In contrast, VaR is only a weighted average of losses in two individual tail scenarios.

4.3.6 Identifying a bad proxy

With all methods based on proxy modelling, there is a risk that, over time, the relationship between the true loss and the proxy loss can deteriorate; the simplifications used in the proxy model may drift too far from the real world experience, or the proxy model parameters may need to be updated. An advantage of the DIANS approach is that the suitability of the proxy model can be assessed directly using the DIANS output, without the need for additional backtesting.

To illustrate, we repeat the GMMB experiment from Section 4.3.3, but with very different parameters for the fund returns, specified in Table 4.4. These parameters generate prolonged periods of very poor mean returns and very high volatility.

(Monthly rate)	Real World
Risk-free Rate: r	0.002
Mean - Regime 1 ($\rho = 1$): μ_1	0.0085
Mean - Regime 2 ($\rho = 2$): μ_2	**-0.0500**
Standard Deviation - Regime 1: σ_1	0.035
Standard Deviation - Regime 2: σ_2	**0.200**
Transition Probability - from Regime 1: p_{12}	**0.10**
Transition Probability - from Regime 2: p_{21}	0.20

Table 4.4: Parameters in Shocked Regime-Switching Model

The results are shown in Figure 4.8. As in Figure 4.4, each column represents a repetition of the DIANS procedure under shocked parameters. The highest dot marks m^*, the minimum number of proxy scenarios required to capture the true 5% tail scenarios, and the triangle represents the cut-off identified by the DIANS algorithm, but without the constraint in line 14 of Algorithm 3, that $m_k \leq m^{\max}$. In fact, in all cases, the algorithm would be stopped as m_k exceeds the maximum of 1,000 scenarios. A small amount of investigation in this case indicates that the constant lapse assumption used in the proxy is not sufficiently accurate when the fund returns are consistently poor for long periods, as is the case under the new parameters. A small change to the lapse assumption restores the proxy model as an adequate signal for the tail scenarios.

Checking how close $\widetilde{m}$ is to $m^{\max}$ is only one way that the DIANS procedure signals an inadequate proxy. Other indicators include the following.

- The Spearman rank correlation can be calculated for the $\widetilde{m}$ proxy losses and simulated losses in the proxy tail scenario set. A strong proxy will have a rank correlation,

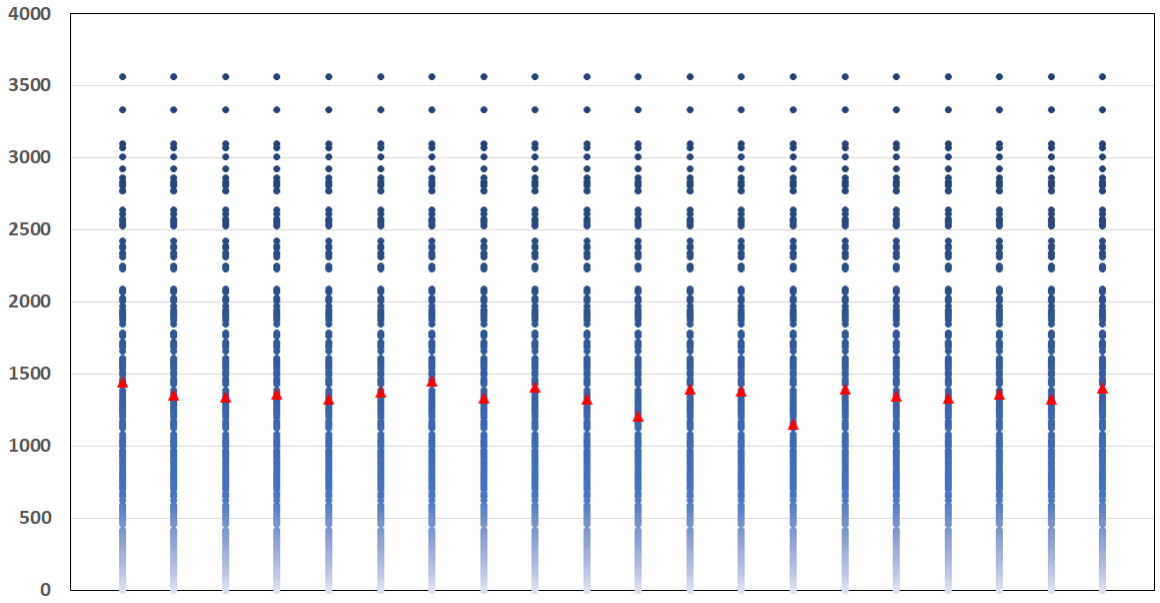


Figure 4.8: Actual *inverse* rank of concomitant of true tail scenarios and and $\widetilde{m}$ (threshold generated by DIANS), in 20 repeated experiments in Section 4.3.6 (sensitivity test). GMMB example.

greater than, say, 0.75. An adequate proxy will have a rank correlation of at least around 0.6. Lower correlations indicate that the proxy needs to be updated.

- The plots of $(U_j^P, \widehat{U}_j)$ (which are $p-p$ plots for the proxy and simulated losses), generated by successive iterations of m_k, can give a visual signal of the proxy model adequacy. If the proxy is working well, then the $p-p$ plots will show strong clustering around the $y = x$ line through successive iterations. If there are a significant number of outliers, that could indicate that the proxy is systematically missing some of the true tail scenarios.

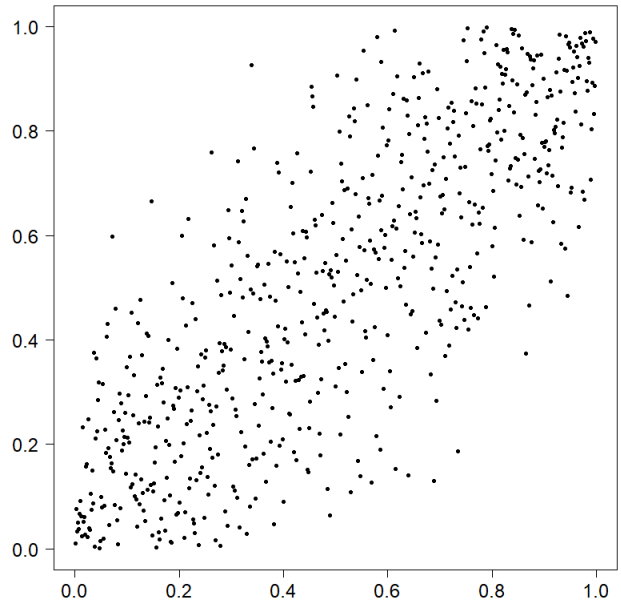
In Figure 4.9 we show examples of copulas generated by successful and unsuccessful proxies. In each case, $M = 5,000$ and $m_0 = 400$. Figure 4.9a shows the same copula in the GMMB experiment as in Figure 4.3a. As discussed in Section 4.3.2, the proxy is a good indicator of the ranking of the losses. The Spearman correlation coefficient was $\rho^s = 0.6$ on the first iteration of the algorithm, and $\rho^s = 0.8$ on the third and final iteration, and $\widetilde{m}$ ended at 652 scenarios.

Figure 4.9b shows the copula from the GMAB experiment. As mentioned earlier, this

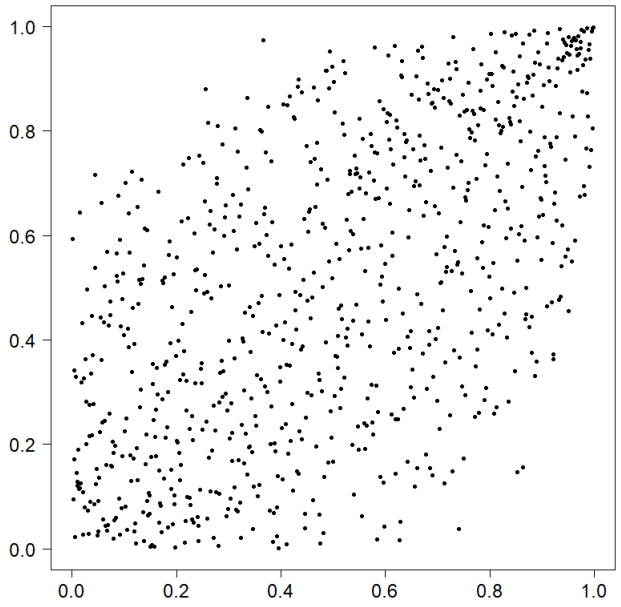
is a less accurate proxy than the GMMB case. The Spearman correlation coefficient was $\rho^s = 0.4$ on the first iteration of the algorithm, and $\rho^s = 0.6$ on the third and final iteration, and $\widetilde{m}$ ended at 823 scenarios.

In Figure 4.9c, the P-P plot indicates that the proxy is not a good indicator of the ranking of the simulated losses. The Spearman correlation coefficient in this example was $\rho^s = 0.2$ on the first iteration of the algorithm, and $\rho^s = 0.3$ on the fourth and final iteration. The final $\widetilde{m}$ was around 1,300 scenarios.

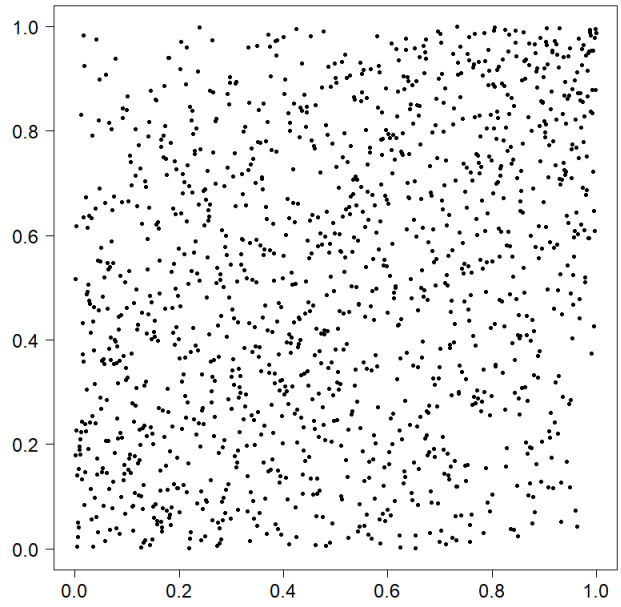
In Figure 4.9d, the copula indicates that the proxy is capturing some of the tail scenarios, but is also misclassifying some. This can happen, for example, for more complex payouts with two triggers, and where the proxy only captures one trigger. The Spearman correlation coefficient in this case was $\rho^s = 0.55$ on the first iteration, and $\rho^s = 0.56$ on the seventh and final iteration. The final number of scenarios in $\mathcal{T}^P_{\widetilde{m}}$ was just under 600, which would not indicate that the proxy was inadequate. The only signals here of an inadequate proxy are the Spearman's rho and the P-P plot.



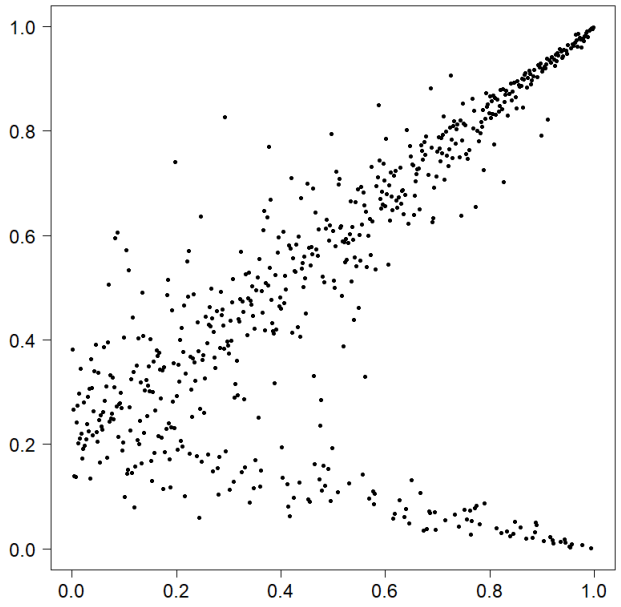
(a) GMMB benchmark example



(b) GMAB benchmark example



(c) Unsuccessful proxy 1



(d) Unsuccessful proxy 2

Figure 4.9: Examples of empirical copulas for Proxy Tail Scenario Sets

4.4 Conclusion

The DIANS method provides a powerful tool for nested simulation in insurance liability measurement, and has potential for application to a wider range of problems, including, for example, assessing semi-static hedging strategies, estimating multi-period risk measures (Hardy and Wirch (2004), Devolder and Lebègue (2017)), and calculating Solvency II regulatory capital requirements (Bauer *et al.*, 2012). It is particularly useful for path dependent problems, where the non-uniform allocation approaches of Gordy and Juneja (2010) and Broadie *et al.* (2011) are not directly applicable. Compared with full proxy model approaches, the DIANS offers more accurate calculation, and also signals an inadequate or ineffective proxy. The extra calculations involved in the process of finding the appropriate size for the proxy tail scenario set is minor, compared with the computational cost of additional simulations.

The identification and implementation of a suitable proxy model is a large part of this methodology. For most VA guarantees with lump sum benefits (that is, excluding GMWBs and GMIBs) the Black-Scholes option pricing framework provides the obvious resource. Complex VA guarantees can be mapped to formulas or numerical approximations developed for exotic options; for example, the GMMB with resets is very similar to a 'put-on-the-max', or high water mark option, for which a valuation formula was provided in Goldman *et al.* (1979).

Where no tractable Black-Scholes valuation approach is available, the proxy model is most likely to be generated intrinsically. This may involve a regular calibration exercise to construct empirical valuation functions based on key scenario variables. The PDE valuation method of Feng (2014) could be used to construct a proxy model, which can then be combined with targeted inner simulations. In the next chapter, the proxy model is replaced with a pilot simulation using common set of inner simulations, with the inner simulation probabilities adjusted for each scenario using a likelihood ratio approach.

Chapter 5

Nested Simulation using Likelihood Ratio Methods

5.1 Introduction

In this chapter, we use the likelihood ratio method to improve the efficiency of nested simulation. Inspired by the green simulation (Feng and Staum, 2017) design paradigm, which reuses simulation outputs in repeated experiments, we reuse the inner sample paths in nested simulation procedures across scenarios. To do this, we employ the so-called mixture likelihood ratio (MLR) estimator that pools inner replications across all outer scenarios to improve the accuracy for each scenario's inner simulation output. The MLR estimator is applicable to both single and multi period nested simulations.

As in the IANS and DIANS method, we propose a two-stage nested simulation procedure for tail risk estimation. The first stage uses a fraction of the simulation budget to identify some tail scenarios. Unlike the IANS and DIANS method, which uses an extrinsic proxy model to identify tail scenarios, here we use an intrinsic proxy based on results from the first stage pilot simulation. In the second stage, we concentrate the remaining simulation budget on the tail scenarios identified in the first stage. We do not use the standard Monte Carlo method in the second stage, as in the case of the IANS and DIANS method. Instead, we apply the MLR estimator to simulation output in the second stage.

Recall from Section 1.1.1, the GMWB contract offers annuity-type of guaranteed benefits where the benefit amount could potentially increase over time, depending on the stock price path. In this case, the GMWBs are path-dependent contracts for which the

modeling is materially different from the GMMBs and GMABs we study in the previous chapters. The IANS and DIANS method we present in previous chapters do not work well on path-dependent annuity-type contracts like the GMWBs, because there is no clear effective proxy model for these contracts. In this chapter, we first adapt the MLR estimators to accommodate path-dependency in complex VA contract such as the Guaranteed Minimum Withdrawal Benefit (GMWB). Then we apply the adapted MLR estimator in the two-stage simulation procedure for tail risk estimation.

Our proposed procedure only requires the asset model to be a Markov process. It can be applied in other path-dependent contract such as the GLWBs, GMIBs, and financial derivatives with path-dependent payoffs. Numerical studies show that the proposed procedure can produce much more accurate estimates of tail risk measures than a standard nested simulation, given the same computation budget.

5.2 Nested Simulation for Tail Risk Measures Using the Likelihood Ratio Method

5.2.1 Nested simulations using likelihood ratios

The *green simulation* design paradigm proposed by Feng and Staum (2017) reuses simulation outputs in temporally repeated experiments so simulation outputs in past simulation experiments can be used to improve the efficiency in future experiments. Here we describe how green simulation estimators, particularly the likelihood ratio based estimators, can be applied in the context of nested simulations.

We presented details of the standard Monte Carlo nested simulation in Section 2.3 and discussed the computation burden of the procedure. To improve the efficiency of the standard Monte Carlo nested simulation, we propose applying likelihood ratio estimators to reuse inner simulation outputs in all scenarios at each time $t = 0, \ldots, T-1$.

Recall the following definitions and notations from Section 2.3 for describing nested simulations.

- M is the number of outer scenarios sample paths in the nested simulation.
- N is the number of inner simulation invoked at each time step along each of the M outer scenarios.

- $X_t^{(i)}$, for $t = 1, 2, \ldots, T$, $i = 1, \ldots, M$, is the sample value of X_t in the ith outer scenario.
- $\boldsymbol{X}^{(i)} = (X_0^{(i)}, X_1^{(i)}, X_2^{(i)}, \ldots, X_T^{(i)})$, $i = 1, \ldots, M$ is the ith outer scenario.
- $Y_{t,t'}^{(i)}$, for $t' = t, \ldots, T$, is random variable $Y_{t,t'}$, conditional on $X_t^{(i)}$.
- $\boldsymbol{Y}_t^{(i)} = \left(Y_{t,t}^{(i)}, Y_{t,t+1}^{(i)}, \ldots, Y_{t,T}^{(i)}\right)$ is the entire paths of random vector $Y_{t,t'}^{(i)}$, $t' = t, \ldots, T$.
- $Y_{t,t'}^{(i,j)}$, $t = 1, 2, \ldots, T$, $t' = t, \ldots, T$, $i = 1, \ldots, M$, $j = 1, \ldots, N$, is the sample value at time t' on the jth inner simulation sample path, conditional on $X_t^{(i)}$.
- $\boldsymbol{Y}_t^{(i,j)} = \left(Y_{t,t}^{(i,j)}, \ldots, Y_{t,T}^{(i,j)}\right)$, $t = 1, 2, \ldots, T$, $i = 1, \ldots, M$, $j = 1, \ldots, N$, is the entire j-th inner sample path, conditional on $X_t^{(i)}$.
- $g_t^{(i)}(\boldsymbol{Y}_t^{(i,j)})$ is the probability density function of $\boldsymbol{Y}_t^{(i,j)}$, conditional on $X_t^{(i)}$), that is, $g_t^{(i)}(\boldsymbol{y}) = g(\boldsymbol{y}|X_t^{(i)})$. In a Monte Carlo simulation, $\boldsymbol{Y}_t^{(i,j)} \overset{i.i.d.}{\sim} g_t^{(i)}(\boldsymbol{y})$, for all $j = 1, \ldots, N$.
- $\mu_t(X_t^{(i)}) = \mathbb{E}\left[f(\boldsymbol{Y}_t^{(i)}|X_t^{(i)})\right]$ is the pricing or valuation functional evaluated via inner simulation at time t given $X_t^{(i)}$. In the VA hedging loss context, $\mu_t(X_t^{(i)}) = \Delta_t^{(i)}$, the delta of the hedging portfolio at time t in outer scenario i.

Assumption 5.2.1 makes sure that the likelihood ratio estimators in this thesis are well-defined.

Assumption 5.2.1. *The conditional probability density functions, $g_t^{(i)}(\boldsymbol{y})$, are well-defined and can be calculated for all $t = 0, 1, \ldots, T-1$ and all $i = 1, \ldots, M$. Moreover, for each $t = 0, 1, \ldots, T-1$, the supports of $g_t^{(i)}(\boldsymbol{y})$ are identical for all $i = 1, \ldots, M$.*

Given Assumptions 5.2.1, we have the following identity

$$\begin{aligned}
\mu_t(X_t^{(i)}) &= \mathbb{E}\left[f(\boldsymbol{Y}_t|X_t^{(i)})\right] = \mathbb{E}\left[f(\boldsymbol{Y}_t^{(i)})\right] \\
&= \int f(\boldsymbol{y}) g_t^{(i)}(\boldsymbol{y}) d\boldsymbol{y} = \int f(\boldsymbol{y}) \frac{g_t^{(i)}(\boldsymbol{y})}{g_t^{(k)}(\boldsymbol{y})} g_t^{(k)}(\boldsymbol{y}) d\boldsymbol{y} \\
&= \mathbb{E}\left[f(\boldsymbol{Y}_t) \frac{g_t^{(i)}(\boldsymbol{Y}_t)}{g_t^{(k)}(\boldsymbol{Y}_t)} \middle| X_t^{(k)}\right], \qquad (5.1)
\end{aligned}$$

Equation (5.1) leads to the likelihood ratio (LR) estimator of $\mu_t(X_t^{(i)})$, using scenario $X_t^{(k)}$ inner simulations:

$$\widehat{\mu}_{t,k}^{\mathrm{LR}}(X_t^{(i)}) = \frac{1}{N}\sum_{j=1}^{N} f\left(\boldsymbol{Y}_t^{(k,j)}\right)\frac{g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}{g_t^{(k)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}, \quad \boldsymbol{Y}_t^{(k,j)} \overset{i.i.d.}{\sim} g_t^{(k)}(\boldsymbol{y}), \forall j = 1,\ldots,N. \tag{5.2}$$

We refer to scenario $X_t^{(k)}$ in this case as the *sampling scenario* and scenario $X_t^{(i)}$ as the *target scenario*.

The intuition behind the likelihood ratio estimators is appropriately reweighting the inner simulation outputs in scenario $X_t^{(k)}$ (scenario k), i.e., $f\left(\boldsymbol{Y}_t^{(k,j)}\right)$, $j = 1,\ldots,N$, to estimate scenario i's expected inner simulation output $\mu_t(X_t^{(i)}) = \mathbb{E}\left[f(\boldsymbol{Y}_t|X_t^{(i)})\right]$.

We see from Equation (5.2) that the likelihood ratio method is mathematically identical to the importance sampling variance reduction method, but they differ in means and goals. The main difference in means is the choice of the sampling distribution. Importance sampling seeks the best sampling distribution to reduce the estimation variance. In contrast, the likelihood ratio method has no control over the sampling distribution but reweight previously simulated outputs via likelihood ratios to form unbiased estimators for a different target distribution. In the context of nested simulation, our goal is to reuse inner simulation outputs across different outer scenarios to improve the estimation accuracy for each scenario, whereas in importance sampling, the goal is to reduce estimation variance.

Figure 5.1 illustrates the reuse of simulation outputs. The solid lines represent simulated sample paths. Scenario $X_t^{(i)}$ represents the target scenario whereas scenario $X_t^{(k)}$ and $X_t^{(M)}$ represent two different sampling scenarios. The dashed lines represent the inner sample outputs being reused from the sampling scenarios to the target scenario. As shown in Figure 5.1, reusing simulation outputs in different sampling scenarios effectively increases the number of inner sample paths applied to the target scenario.

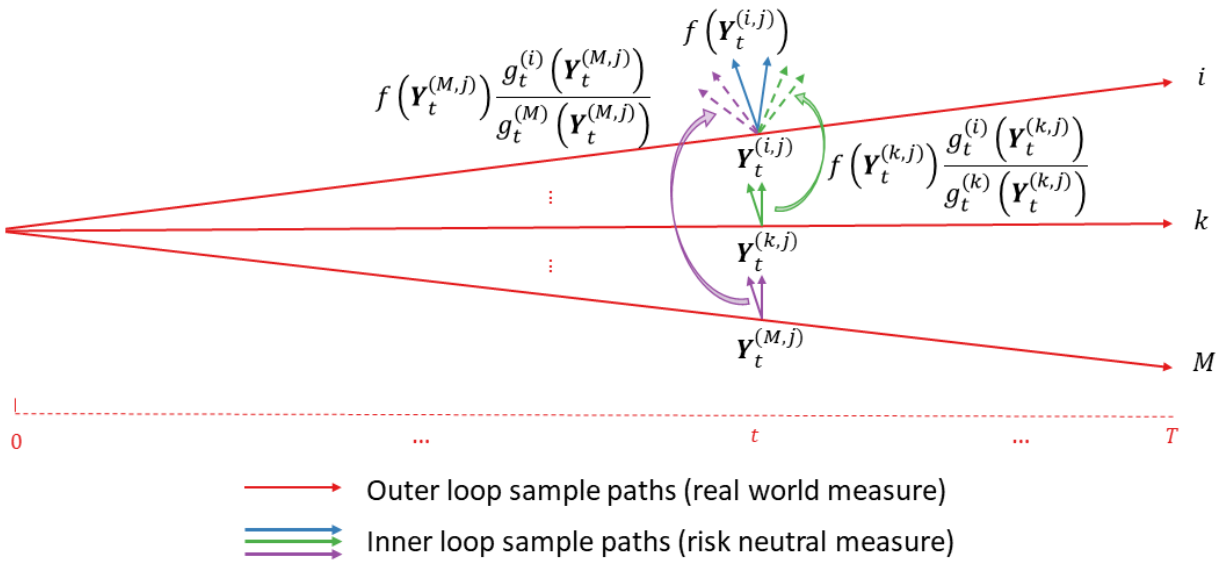


Figure 5.1: Likelihood Ratio Estimator

Computationally, reusing the output $f\left(\boldsymbol{Y}_t^{(k,j)}\right)$ for different target scenarios requires the likelihood ratio $\dfrac{g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}{g_t^{(k)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}$. If the calculation of the likelihood ratio $\dfrac{g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}{g_t^{(k)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}$ is faster than simulating a new inner path $\boldsymbol{Y}_t^{(i)}$ and calculating its output $f\left(\boldsymbol{Y}_t^{(i)}\right)$, then reusing the simulation output via the likelihood ratio estimator is more efficient than increasing the number of inner simulations for each scenario.

The likelihood ratio $\dfrac{g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}{g_t^{(k)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}$ requires calculating the joint pdf for the *whole inner sample path*, $\boldsymbol{Y}_t^{(k,j)}$, conditioning on the *outer scenario* $X_t^{(i)}$ or $X_t^{(k)}$. But this calculation simplifies to $\dfrac{g_t^{(i)}\left(Y_{t,t+1}^{(k,j)}\right)}{g_t^{(k)}\left(Y_{t,t+1}^{(k,j)}\right)}$ when the state process is Markov. This simplification further reduces the computations needed in our reusing of simulation outputs.

Proposition 5.2.2. *If Assumption 5.2.1 holds, and the concatenated stochastic process* $(X_t, Y_{t,t+1}, \ldots, Y_{t,T})$ *is Markov, then the likelihood ratio estimator in* (5.2) *can be written*

as

$$\widehat{\mu}_{t,k}^{LR}(X_t^{(i)}) = \frac{1}{N}\sum_{j=1}^{N} f\left(\boldsymbol{Y}_t^{(k,j)}\right)\frac{g_t^{(i)}\left(Y_{t,t+1}^{(k,j)}\right)}{g_t^{(k)}\left(Y_{t,t+1}^{(k,j)}\right)}, \quad \boldsymbol{Y}_t^{(k,j)} \overset{i.i.d.}{\sim} g_t^{(k)}(\boldsymbol{y}), \forall j = 1,\ldots,N. \tag{5.3}$$

Proof. By the Markov property of $(X_t, Y_{t,t+1}, \ldots, Y_{t,T})$, the likelihood ratio in (5.3) can be simplified as

$$\begin{aligned}\frac{g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}{g_t^{(k)}\left(\boldsymbol{Y}_t^{(k,j)}\right)} &= \frac{g\left(Y_{t,t+1}^{(k,j)},\ldots,Y_{t,T}^{(k,j)}\middle|X_t^{(i)}\right)}{g\left(Y_{t,t+1}^{(k,j)},\ldots,Y_{t,T}^{(k,j)}\middle|X_t^{(k)}\right)} = \frac{g\left(Y_{t,t+2}^{(k,j)},\ldots,Y_{t,T}^{(k,j)}\middle|Y_{t,t+1}^{(k,j)}\right)\cdot g\left(Y_{t,t+1}^{(k,j)}\middle|X_t^{(i)}\right)}{g\left(Y_{t,t+2}^{(k,j)},\ldots,Y_{t,T}^{(k,j)}\middle|Y_{t,t+1}^{(k,j)}\right)\cdot g\left(Y_{t,t+1}^{(k,j)}\middle|X_t^{(k)}\right)} \\ &= \frac{g_t^{(i)}\left(Y_{t,t+1}^{(k,j)}\right)}{g_t^{(k)}\left(Y_{t,t+1}^{(k,j)}\right)},\end{aligned}$$

as desired. □

Proposition 5.2.2 shows that, even though the entire inner simulation path $\boldsymbol{Y}_t^{(k,j)}$ is simulated and fed through the inner simulation model $f(\boldsymbol{Y}_t)$, we only need to calculate the likelihood ratio for the one-step transition densities from time t to $t+1$. This means that the likelihood ratio calculation can be very efficient even for embedded options with complex path-dependent payoff structures.

We propose to reuse the simulation outputs from *every* sampling scenario $k = 1,\ldots,M$, for *each* target scenario $i = 1,\ldots,M$. That is, each scenario serves as a sampling scenario for all other scenarios and every scenario reuses samples from all other scenarios. This way, every simulation output is calculated only once but is reused M times, once for each target scenario.

One way to reuse the outputs from all scenarios is to equally weight all these likelihood ratio estimators, i.e., $\widehat{\mu}_t^{\mathrm{ILR}}(X_t^{(i)}) = \frac{1}{M}\sum_{k=1}^{M}\widehat{\mu}_{t,k}^{\mathrm{LR}}(X_t^{(i)})$. This is the *individual likelihood ratio (ILR)* estimator in Feng and Staum (2017). The main drawback of using the ILR estimator is its potentially large or even infinite variance when one or more sampling scenarios are significantly different from the target scenario (Feng and Staum, 2017). We will illustrate an example where the ILR esimator results in large estimation variance in Section 5.2.2.

5.2.2 Nested simulations using mixture likelihood ratio

A remedy proposed by Feng and Staum (2017) is the *mixture likelihood ratio (MLR)* estimator, which is given by

$$\widehat{\mu}_t^{\mathrm{MLR}}(X_t^{(i)}) = \frac{1}{MN}\sum_{k=1}^{M}\sum_{j=1}^{N} f\left(\boldsymbol{Y}_t^{(k,j)}\right)\frac{g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right)}{\bar{g}_M\left(\boldsymbol{Y}_t^{(k,j)}\right)}, \tag{5.4}$$

where the *mixture density function* for the inner paths is defined as

$$\bar{g}_M\left(\boldsymbol{Y}_t^{(k,j)}\right) = \frac{1}{M}\sum_{i=1}^{M} g_t^{(i)}\left(\boldsymbol{Y}_t^{(k,j)}\right).$$

In this chapter, we assume the outer scenarios $\boldsymbol{X}^{(i)}$, $i = 1, \ldots, M$ are given, so for notational convenience they are omitted from the mixture density function.

Proposition 5.2.3. *If Assumption 5.2.1 holds, and the concatenated stochastic process $(X_t, Y_{t,t+1}, \ldots, Y_{t,T})$ is Markov, then the likelihood ratio estimator in (5.4) can be written as*

$$\widehat{\mu}_t^{MLR}(X_t^{(i)}) = \frac{1}{MN}\sum_{k=1}^{M}\sum_{j=1}^{N} f\left(\boldsymbol{Y}_t^{(k,j)}\right)\frac{g_t^{(i)}\left(Y_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(Y_{t,t+1}^{(k,j)}\right)},$$

where $\bar{g}_M(y) = \frac{1}{M}\sum_{i=1}^{M} g_t^{(i)}(y)$, *and* $\boldsymbol{Y}_t^{(k,j)} \overset{i.i.d.}{\sim} g_t^{(k)}(\boldsymbol{y})$ *for all* $j = 1, \ldots, N$, *for every* $k = 1, \ldots, M$.

Proof. The proof is almost identical to the proof for Proposition 5.2.2, by applying the Markov property to all conditional densities (including those in the mixture density) in the likelihood ratio calculation. □

The intuition behind the MLR estimator (5.4) is similar to that for the likelihood ratio estimator (5.2), where we reweight the simulation outputs by the appropriate likelihood ratios, but they differ in the interpretations of the sampling distributions for inner sample paths. The likelihood ratio estimator (5.2), and the ILR estimator, view the inner sample paths $\left\{\boldsymbol{Y}_t^{(k,j)}, j = 1, \ldots, N\right\}$, for $k = 1, 2, \ldots, M$ separately, as M different samples, with conditional distributions $g_t^{(k)}(\boldsymbol{y})$ for sampling scenario k. The MLR estimator, in contrast,

views all the inner sample paths $\left\{\boldsymbol{Y}_t^{(k,j)}, k = 1, \ldots, M,\ j = 1, \ldots, N\right\}$, collectively, as one (stratified) sample from the mixture distribution $\bar{g}_M(\boldsymbol{y})$.

The MLR produces more stable estimates than the ILR because the mixture likelihood ratio is always bounded when reusing samples from finite number of sampling distributions while the individual likelihood ratio can be unbounded. Consider a GMMB. Under the Black-Scholes model with $r = 2\%$ and $\sigma = 30\%$, the inner simulation (risk-neutral) conditional distribution $g(S_{t,t+1}|S_t)$ is a lognormal distribution with mean $\ln(S_t) + r - \frac{\sigma^2}{2}$ and standard deviation σ. Suppose we have two outer scenarios at time $t = 1$, $S_1^{(1)} = 651$ and $S_1^{(2)} = 1525$, and an inner simulation sample path $\boldsymbol{S}_1^{(2,1)} = (S_1^{(2)}, S_{1,2}^{(2,1)}) = (1525, 592)$ which originates from scenario $k = 2$ at time $t = 1$. Suppose we reuse the inner sample path $\boldsymbol{S}_1^{(2,1)}$ to estimate delta for target scenario $i = 1$ at time 1, that is, $\Delta_1^{(1)}$. In the sampling scenario $k = 2$, transitioning from $S_1^{(2)} = 1525$ to $S_{1,2}^{(2,1)} = 592$ is unlikely but plausible in an inner simulation model. In this case, the individual likelihood ratio (ILR) is

$$\frac{g(S_{1,2}^{(2,1)} = 592|S_1^{(1)} = 651)}{g(S_{1,2}^{(2,1)} = 592|S_1^{(2)} = 1525)} = \frac{1.2941}{0.0119} = 109.$$

These extremely high likelihood ratios occur when the two outer scenarios are significantly different ($S_1^{(1)} = 651$ and $S_1^{(2)} = 1525$ in this example), and they may distort the ILR estimate. We can avoid this problem by using the MLR estimator; the mixture likelihood ratio (MLR) for reusing the above sample path is

$$\frac{g(S_{1,2}^{(2,1)} = 592|S_1^{(1)} = 651)}{\frac{1}{2}\left[g(S_{1,2}^{(2,1)} = 592|S_1^{(1)} = 651) + g(S_{1,2}^{(2,1)} = 592|S_1^{(2)} = 1525)\right]} = 1.98,$$

which is a much more reasonable value. The MLR estimator has been shown to have superior accuracy, compared to the ILR estimator. See Hesterberg (1988); Veach and Guibas (1995), and Feng and Staum (2017), for examples.

We use the MLR estimator to improve simulation efficiency in our two-stage procedure, which is described in Section 5.3.

Applying the MLR estimator in a nested simulation involves a small scale, standard nested simulation at first, to generate the inner simulation sample paths, and then applying the MLR estimator to estimate each $\mu_t(X_t^{(i)})$. Suppose the number of calculations required to calculate one likelihood ratio adjusted inner sample is C_{LR}. Then the total simulation

budget required for a nested simulation using the MLR estimator is

$$\Gamma_{\text{MLR}} = \underbrace{\frac{M \times N_{\text{MLR}} \times T \times (T+1)}{2} \times C_{\text{inner}}}_{\text{sample paths generation}} + \underbrace{M^2 \times N_{\text{MLR}} \times T \times C_{\text{LR}}}_{\text{likelihood calculation}} \tag{5.5}$$

Here M is typically large compared to T. Therefore, the likelihood calculation portion dominates the overall computation cost of nested simulation using the MLR estimator. In fact, Equation (5.5) suggests that $\Gamma_{\text{MLR}} = \mathcal{O}(M^2)$, which means that the computation cost of nested simulations using the MLR estimator can be significant if M is large. This is a disadvantage in using the MLR estimator, and provides motivation for the two-stage nested simulation design, which is discussed in detail in Section 5.3.

Compared to the computation cost for a standard nested simulation, the nested simulation using the MLR estimator can achieve significant savings in computation when T is relatively large and N_{MLR} is relatively small. More specifically, assuming $\frac{C_{\text{LR}}}{C_{\text{inner}}} = \mathcal{O}(1)$ and $N_{\text{MLR}} \ll N_{\text{SMC}}$, then

$$\frac{\Gamma_{\text{MLR}}}{\Gamma_{\text{SMC}}} \propto \frac{M \times N_{\text{MLR}}}{N_{\text{SMC}} \times T} \tag{5.6}$$

The impact on the computational efficiency through using the MLR estimator, compared to a standard nested simulation, is illustrated in the numerical experiments in Section 5.5.

5.3 Two-stage Nested Simulation for Tail Risk Measures Using Mixture Likelihood Ratio Estimator

The motivation behind the two-stage simulation design is as follows. First of all, similar to the rationale behind the IANS and DIANS procedure, we want to focus the computation budget on tail scenarios that are relevant to tail risk measure estimation. Secondly, since the computation required in the MLR estimation grows quadratically with the number of outer scenarios being considered, it is more efficient to consider all M outer scenarios in the first stage to get a reasonable estimate of the *ranking* of each scenario's loss, for which fewer inner simulations are required, and then to consider only the outer scenarios that are highly likely to be included in the CTE estimation in the second stage, when we estimate the value of losses in these highly likely tail scenarios more accurately.

We assume that the set of outer scenarios $\boldsymbol{X}^{(i)}$, $i = 1, \ldots, M$, and the simulation budget Γ, are given.

The procedure can be broken down into the following steps.

1. **Classification of tail scenarios**. In this stage, we use a fraction of the simulation budget to identify a set of tail scenarios that is highly likely to include the true tail scenarios set $\mathcal{T}_{(1-\alpha)M}$. The number of inner simulations N_1 is a design parameter selected by the user.

   (1.A) For each outer scenario $\boldsymbol{X}^{(i)}$, $i = 1, \ldots, M$, simulate N_1 inner sample paths at each time $t = 0, 1, \ldots, T-1$.

   (1.B) Use the MLR estimator $\widehat{\mu}_t^{\text{MLR}}(X_t^{(i)})$ to estimate $\mu_t(X_t^{(i)})$, for all $t = 0, 1, \ldots, T-1$. Note that the inner sample size for each scenario is N_1 and the MLR estimator reuses inner sample paths in all M outer scenarios.

   (1.C) Use the estimates $\widehat{\mu}_t^{\text{MLR}}(X_t^{(i)})$, $t = 0, 1, \ldots, T-1$ in Step (1.B) to estimate the loss L_i in scenario $\boldsymbol{X}^{(i)}$. In the VA dynamic hedging context, $\widehat{\mu}_t^{\text{MLR}}(X_t^{(i)}) = \widehat{\Delta}_t^{(i)\text{MLR}}$.

$$\widehat{L}_i^{\text{MLR}} = \sum_{t=0}^{T-1} \widehat{\Delta}_t^{(i)\text{MLR}} \left(D_t S_t^{(i)} - D_{t+1} S_{t+1}^{(i)} \right) + V_0^{(i)}, \ i = 1, \ldots, M.$$

   (1.D) Sort the estimated losses $\widehat{L}_{(1)^{\text{HL}}}^{\text{MLR}} \leq \widehat{L}_{(2)^{\text{HL}}}^{\text{MLR}} \leq \cdots \leq \widehat{L}_{(M)^{\text{HL}}}^{\text{MLR}}$ then identify the set of *highly likely (HL) tail scenarios* $\widehat{\mathcal{T}}_{M^*}^{\text{HL}} = \{\boldsymbol{X}^{(i)} : \widehat{L}_i^{\text{MLR}} > \widehat{L}_{(M-M^*)^{\text{HL}}}^{\text{MLR}}\}$, where the number tail size $M^* \geq (1-\alpha)M$ is a design parameter selected by the user.

2. **Estimation of tail losses**. In this stage, we concentrate the remaining simulation budget to the highly likely tail scenarios to improve the estimation of these tail losses. We solve for N_{MLR} using Equation (5.5), given the remaining computation budget and with $M = M^*$. Then let $N_2 = N_{\text{MLR}}$ be the (rounded) average inner sample paths for the highly likely tail scenarios.

   (2.A) Similar to Step (1.A), but only run inner simulations for the highly likely tail scenarios, i.e., $\boldsymbol{X}^{(i)}$ for $i \in \widehat{\mathcal{T}}_{M^*}^{\text{HL}}$.

   (2.B) Similar to Step (1.B), but use the MLR estimator $\widehat{\mu}_t^{\text{MLR}}(X_t^{(i)})$ to estimate $\mu_t(X_t^{(i)})$ for $\boldsymbol{X}^{(i)} \in \widehat{\mathcal{T}}_{M^*}^{\text{HL}}$; also, only reuse the inner sample paths in these scenarios. The inner sample size for each scenario in $\widehat{\mathcal{T}}_{M^*}^{\text{HL}}$ is $N_1 + N_2$, as we also reuse the inner sample paths simulated in Stage 1.

   (2.C) Similar to Step (1.C), but estimate the losses $\widehat{L}_i^{(\text{MLR2})}$ for all $\boldsymbol{X}^{(i)} \in \widehat{\mathcal{T}}_{M^*}^{\text{HL}}$ using $\mu_t(X_t^{(i)})$ estimated in Step (2.B).

(2.D) Sort the estimated losses $\widehat{L}^{\text{MLR2}}_{(1)^{\text{MLR}}} \le \widehat{L}^{\text{MLR2}}_{(2)^{\text{MLR}}} \le \cdots \le \widehat{L}^{\text{MLR2}}_{(M^*)^{\text{MLR}}}$ for the highly likely tail scenarios. Then estimate the α-CTE by

$$\widehat{CTE}^{\text{MLR}}_{\alpha} = \frac{1}{(1-\alpha)M} \sum_{i=\alpha M+1}^{M} \widehat{L}^{\text{MLR2}}_{(i)^{\text{MLR}}} = \frac{1}{(1-\alpha)M} \sum_{i \in \widehat{\mathcal{T}}^{\text{MLR}}_{(1-\alpha)M}} \widehat{L}^{\text{MLR2}}_{i}, \tag{5.7}$$

where the set of *MLR tail scenarios* is

$$\widehat{\mathcal{T}}^{\text{MLR}}_{(1-\alpha)M} = \{\boldsymbol{X}^{(i)} \in \widehat{\mathcal{T}}^{\text{HL}}_{M^*} : \widehat{L}^{\text{MLR2}}_{i} > \widehat{L}^{\text{MLR2}}_{(M^*-(1-\alpha)M)^{\text{MLR}}}\}.$$

In other words, $\widehat{\mathcal{T}}^{\text{MLR}}_{(1-\alpha)M}$ are the $(1-\alpha)M$ scenarios, among the M^* highly likely tail scenarios, with the largest estimated losses.

We see that Steps (2.A)–(2.C) are similar to Steps (1.A)–(1.C), but the simulation and estimation focus only on the highly likely tail scenarios. Also, the CTE_α estimator (5.7) averages the largest $(1-\alpha)M$ losses among those in the highly likely tail scenarios.

We find that a judicious choice of the design parameters N_1 and M^* can be valuable. Here we provide some guidelines based on our experience:

- The number of inner simulations per scenario in Stage 1, N_1, should be very small, for example, 1 or 2. In Stage 1, the goal is to identify the highly likely tail scenarios. So we can afford coarse estimates for the *values* of the scenario losses, as long as their *relative rankings* are similar to the rankings of the true losses. The MLR estimator reuses the inner sample paths from all M scenarios, thus every $\mu_t^{\text{MLR}}(X_t)$ is estimated using MN_1 inner sample paths; so a small N_1 suffices for our purpose in Stage 1. Moreover, the smaller N_1 is, the larger the remaining simulation budget is for more accurate estimation in Stage 2.

- The number of tail scenarios, M^*, can be the desired number of true tail scenarios, i.e., $(1-\alpha)M$, plus a small safety margin, e.g., 5% of the outer scenarios. Due to simulation noise, the rankings of the MLR estimates of losses may be different from the rankings of the true losses. So, we should select $M^* \ge (1-\alpha)M$ to, with high confidence, include the $(1-\alpha)M$ true tail scenarios in $\widehat{\mathcal{T}}^{\text{HL}}_{M*}$. However, if M^* is too large, the remaining simulation budget will be less concentrated in Stage 2. The appropriate safety margin varies for different applications. In our experiments, we test a few safety margins and find that $M^* = (1-\alpha)M + 5\%M$ to $M^* = (1-\alpha)M + 10\%M$ strikes a good balance for including the true tail scenarios in Stage 1 and concentrating the simulation budget in Stage 2.

We choose to use an arbitrary selection of safety margin in the two-stage nested simulation using MLR estimator in this chapter. This is the drawback of the IANS method that we addressed in the DIANS method. However, the DIANS method is not helpful in this case because by applying the MLR estimator in each iteration of the search for proxy tail scenarios in the DIANS method, we need to re-calculate the likelihood ratio in every sampling scenario for each target scenario in the proxy tail scenario set, even if only a handful of scenarios are added to the proxy tail scenario set in the iteration. The computation required to re-calculate the likelihood ratio can be significant. Thus, we choose not to apply the DIANS method in the two-stage nested simulation using MLR estimator.

We see that both design parameters N_1 and M^* affects the classification of tail scenarios and estimation of tail losses. The optimal selection of these parameters will be considered in future studies.

5.4 Nested Simulation for Tail Risk Measures of Guaranteed Minimum Withdrawal Benefit (GMWB)

In this section we apply the two-stage nested simulation procedure using MLR estimators to estimate the tail risk of the hedging loss for Guaranteed Minimum Withdrawal Benefit (GMWB). We see that, even though the financial modeling of the embedded option (Section 5.4.1) is much more complicated than the GMMBs or GMABs, little adjustment (Section 5.4.2) is needed to apply our procedure.

5.4.1 Financial Modeling and Dynamic Hedging for GMWB

In Section 2.1, we briefly discussed the mechanism of various VA guarantees including the GMWB. In this section, we will describe in more detail the VA GMWB contract with ratchet feature that we focus on in this chapter. We will then illustrate how the liability of this contract, including the gain and loss from a dynamic delta hedging portfolio, is modeled.

Consider a single premium GMWB contract with a term to maturity of $T = 240$ months. We assume the sub-account value is denoted by F_t at time t, and the sub-account invests a stock index with price S_t at time t. Let η^g be the gross rate at which management fee is deducted from the sub-account each month. Let $\eta^n < \eta^g$ be the net rate at which

management fee income is received by the insurer each month. For simplicity, we assume that the withdrawal benefit starts immediately after the contract commences. Let G_t be the guarantee value at time t. The contract offers a guaranteed monthly withdrawal benefit in the amount of $I_t = \gamma G_t$ for some fixed percentage γ.

At the inception of the contract, i.e., $t = 0$, we assume that the whole premium is invested in the stock index and the guarantee base is set to the sub-account value, so mathematically,

$$F_0 = S_0 = G_0.$$

Also, set $I_0 = 0$ as the first withdrawal starts at time 1. For clarity, we use F_t and F_{t^+} to denote the sub-account value just before and just after the withdrawal at time t.

At each time $t = 1, \ldots, T$, the follow events take place in order:

1. The (remaining) sub-account value changes according to the growth of the underlying stock and the management fee is deducted, that is,

$$F_t = (F_{t-1} - I_{t-1})^+ \cdot \frac{S_t}{S_{t-1}}(1 - \eta^g) = (F_{t-1^+})^+ \cdot \frac{S_t}{S_{t-1}}(1 - \eta^g), \tag{5.8}$$

   Thus the insurer's income at time t is $F_t \eta^n$.

2. The guarantee value ratchets up if the sub-account value exceeds the previous guarantee value, i.e.,

$$G_t = \max\{G_{t-1}, F_t\}. \tag{5.9}$$

3. The withdrawal is made and is deducted from the sub-account value, i.e.,

$$F_{t^+} = F_t - I_t, \text{ where } I_t = \gamma G_t. \tag{5.10}$$

We see from the above modeling that the status of a GMWB at any time t can be summarized by the triplet $X_t = (S_t, F_t, G_t)$, so we treat this triplet as the state variable for modeling. The evolution of the triplet is driven by the stochasticity of the stock price process S_t. This means the likelihood ratio calculations are minimally impacted by the more complicated payoff structure.

For clarity of exposition, we do not consider mortality, lapse, or excess withdrawal. We assume all these factors are deterministic given the stock price sample paths; this is an assumption consistent with current industry modeling practice. Our procedure easily applies to simulation models that incorporate these factors, by adding them to the state

variable. Interested readers are encouraged to refer to, e.g. Fung *et al.* (2014); Piscopo and Haberman (2011); Bauer *et al.* (2017); Bernard *et al.* (2014b); Moenig and Zhu (2018); Knoller *et al.* (2016) for the impacts of these factors to the risk profile of VAs.

The insurer's liability in a GMWB contract is the present value of all the withdrawal benefits paid after the depletion of the sub-account F_t, offset by the present value of all the net management fees $F_t\eta^n$ collected as income. See Section 2.1. Mathematically, the PV of the unhedged liability of the GMWB contract at time t is

$$V_t = \sum_{s=t+1}^{T} e^{-r(s-t)} \left[(I_s - F_s)^+ - F_s\eta^n \right].$$

Given the evolution of $X_t = (S_t, F_t, G_t)$, shown in Equation (5.8)-(5.10), we see that the GMWB liability is path-dependent. That is, the value depends on the entire path $\{X_t, X_{t+1}, \ldots, X_T\}$.

Consider a dynamic delta hedging program for this GMWB contract. The insurer constructs and maintains (by periodic rebalancing) a hedge portfolio at times $t = 0, \ldots, T-1$ to offset the delta of the GMWB's future liabilities. Similar to delta hedging the GMMB and GMAB examples in previous chapters, the time t hedge portfolio consists of Δ_t units of underlying stock S_t, where Δ_t is the sensitivity of the GMWB's future liability beyond time t with respect to the time t stock price S_t.

Following Cathcart *et al.* (2015), given an outer scenario at time t, $X_t = (S_t, F_t, G_t)$, the pathwise estimator of Δ_t of the GMWB liability is estimated from a single inner simulation path as

$$f\left(\boldsymbol{S}_t, \boldsymbol{F}_t, \boldsymbol{G}_t\right) = \sum_{s=t+1}^{T} e^{-r(s-t)} \left[\mathbb{1}\left\{I_{t,s} > F_{t,s}\right\} \cdot \left(\frac{dI_{t,s}}{dS_t} - \frac{dF_{t,s}}{dS_t} \right) - \frac{dF_{t,s}}{dS_t}\eta^n \right] \tag{5.11}$$

where $\boldsymbol{S}_t = \{S_{t,t+1}, \ldots, S_{t,T}\}$, $\boldsymbol{F}_t = \{F_{t,t+1}, \ldots, F_{t,T}\}$ and $\boldsymbol{G}_t = \{G_{t,t+1}, \ldots, G_{t,T}\}$ are the inner simulation sample paths for stock price, sub-account value, and guarantee value, respectively. Using Equations (5.8)–(5.10), at each time step $s = t+1, \ldots, T$ (in the inner sample paths), the sensitivities $\frac{dF_s}{dS_t}$, $\frac{dG_s}{dS_t}$ and $\frac{dI_s}{dS_t}$ are calculated recursively as

$$\frac{dF_{t,s}}{dS_t} = \mathbb{1}\{I_{t,s-1} < F_{t,s-1}\} \left(\frac{dF_{t,s-1}}{dS_t} - \frac{dI_{s-1}}{dS_t} \right) \cdot \frac{S_{t,s}}{S_{t,s-1}} \cdot (1-\eta^g), \tag{5.12}$$

$$\frac{dG_{t,s}}{dS_t} = \mathbb{1}\{G_{t,s-1} < F_s\} \frac{dF_{t,s}}{dS_t} + \mathbb{1}\{G_{t,s-1} \geq F_{t,s}\} \frac{dG_{t,s-1}}{dS_t}, \tag{5.13}$$

$$\frac{dI_{t,s}}{dS_t} = \gamma \frac{dG_{t,s}}{dS_t}. \tag{5.14}$$

The initial condition of these recursions are determined as follows.

- Each inner simulation sample path at time t is initialized by setting
$$(S_{t,t}, F_{t,t}, G_{t,t}) = (S_t, F_t, G_t).$$
- Before any withdrawal at time t, F_t is proportional to S_t. Hence $\dfrac{dF_{t,t}}{dS_t} = \dfrac{dF_t}{dS_t} = \dfrac{F_t}{S_t}$.
- Before any withdrawal at time t, G_t and I_t are constant. Hence $\dfrac{dG_{t,t}}{dS_t} = \dfrac{dG_t}{dS_t} = 0$ and $\dfrac{dI_{t,t}}{dS_t} = \gamma \dfrac{dG_{t,t}}{dS_t} = 0$.

To summarize, the initial conditions of the above recursion are
$$\frac{dF_{t,t}}{dS_t} = \frac{F_t}{S_t}, \ \frac{dG_{t,t}}{dS_t} = 0, \text{ and } \frac{dI_{t,t}}{dS_t} = 0.$$
Interestingly, we observe that calculations are unnecessary when $F_t \leq \gamma G_t = I_t$. When $F_t \leq I_t$, the sub-account is depleted at or before time t, so the GMWB's future liabilities beyond time t no longer depend on the stock; and thus no hedge is needed. In such cases, we can skip the recursive calculations and simply set $\Delta_t = 0$ without any simulation (or set all the corresponding inner simulation outputs $f(\boldsymbol{S}_t, \boldsymbol{F}_t, \boldsymbol{G}_t) = 0$ for reusing purpose). In summary, for each outer scenario $\left(S_t^{(i)}, F_t^{(i)}, G_t^{(i)}\right)$, the time t pathwise delta estimator for GMWB in a standard nested simulation is given by
$$\widehat{\Delta}_t^{(\mathrm{i})\mathrm{NS}} = \begin{cases} 0, & \text{if } F_t^{(i)} \leq \gamma G_t^{(i)}, \\ \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} f\left(\boldsymbol{S}_t^{(i,j)}, \boldsymbol{F}_t^{(i,j)}, \boldsymbol{G}_t^{(i,j)}\right), & \text{if } F_t^{(i)} > \gamma G_t^{(i)}, \end{cases} \tag{5.15}$$
where $\boldsymbol{S}_t^{(i,j)} \overset{i.i.d.}{\sim} g_t^{(i)}(\boldsymbol{s}_t)$ all all $j = 1, \ldots, N$.

Using the this delta estimator for GMWB, a standard multi-period nested simulation procedure for the tail risk estimation for the hedging loss for this GMWB contract is analogous to that for the GMMBs and GMABs, but with the extended state variable and the new liability structure. Specifically, for each of the given scenarios $\boldsymbol{X}^{(i)}$, $i = 1, \ldots, M$, the GMWB's hedging loss is
$$\widehat{L}_i^{\mathrm{NS}} = \sum_{t=0}^{T-1} \widehat{\Delta}_t^{(\mathrm{i})\mathrm{NS}} [e^{-rt} S_t^{(i)} - e^{-r(t+1)} S_{t+1}^{(i)}] + \sum_{t=1}^{T} e^{-rt} \left[(I_t^{(i)} - F_t^{(i)})^+ - F_t^{(i)} \eta^n \right]. \tag{5.16}$$

These estimated losses are sorted and the CTE_α is estimated by averaging the largest $(1-\alpha)M$ estimated loses.

5.4.2 Adapting the Likelihood Ratio Method to the GMWB Loss

Due to some special features of the GMWB's liability, a few subtle yet important adaptations to the likelihood ratio method are needed before it can be applied to estimating the tail risk of the hedging loss.

Recall from Figure 5.1 that, when reusing inner simulation outputs from sampling scenarios k to a target scenario i, we effectively *concatenate* the inner sample paths from time t to T in scenarios k, $\boldsymbol{Y}_t^{(k,j)}$, to the outer scenario i, $X_t^{(i)}$. In the modeling of GMWB (Section 5.4.1) we established the state variable $X_t = (F_t, G_t, S_t)$. Therefore, concatenating the jth inner sample path from sampling scenario k to target scenario i we get

$$\underbrace{\left(S_t^{(i)}, F_t^{(i)}, G_t^{(i)}\right)}_{X_t^{(i)}}, \underbrace{\left(S_{t,t+1}^{(k,j)}, F_{t,t+1}^{(k,j)}, G_{t,t+1}^{(k,j)}\right), \ldots, \left(S_{t,T}^{(k,j)}, F_{t,T}^{(k,j)}, G_{t,T}^{(k,j)}\right)}_{\boldsymbol{Y}_t^{(k,j)}}.$$

The problem is that some concatenated paths are invalid, as they are impossible in the context of the GMWB. According to Equation (5.9), the guarantee value G_t is the running maximum of the sub-account value F_t. A concatenated sample path can violate this condition in two ways:

(I) $G_t^{(i)} > G_{t,t+1}^{(k,j)}$: Since G_t is the running maximum of all sub-account values, it should never decrease, therefore any concatenated sample path with $G_t^{(i)} > G_{t,t+1}^{(k,j)}$ is invalid.

(II) $G_t^{(i)} < G_{t,t+1}^{(k,j)}$ but $G_{t,t+1}^{(k,j)} \neq F_{t,t+1}^{(k,j)}$: Since in the concatenated sample path, $G_{t,t+1}^{(k,j)} = \max\left(G_t^{(i)}, F_{t,t+1}^{(k,j)}\right)$, $G_t^{(i)} < G_{t,t+1}^{(k,j)}$ requires that $G_{t,t+1}^{(k,j)} = F_{t,t+1}^{(k,j)}$.

We do not want to remove all the invalid concatenated paths, as this wastes simulation outputs. Instead, when reusing inner sample paths from sampling scenario k to target scenario i, we adjust the inner sample paths as

$$\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}, \widetilde{\boldsymbol{F}}_t^{(k,j)}, \widetilde{\boldsymbol{G}}_t^{(k,j)}\right) = \begin{pmatrix} S_t^{(i)}, & \boldsymbol{S}_{t+1}^{(k,j)} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(i)}}{S_t^{(k)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+}, \\ F_t^{(i)}, & \boldsymbol{F}_{t+1}^{(k,j)} \cdot \frac{G_t^{(i)}}{G_t^{(k)}}, \\ G_t^{(i)}, & \boldsymbol{G}_{t+1}^{(k,j)} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \end{pmatrix} \tag{5.17}$$

With this adjustment, in the concatenated sample path, $\widetilde{G}_{t,t+1}^{(k,j)} = G_{t,t+1}^{(k,j)} \times \dfrac{G_t^{(i)}}{G_t^{(k)}} \geq G_t^{(i)}$ always holds because $G_{t,t+1}^{(k,j)} \geq G_t^{(k)}$ by definition of G_t. We re-examine the two cases of violation in the adjusted concatenated sample paths.

(I) $G_t^{(i)} > \widetilde{G}_{t,t+1}^{(k,j)}$: This case no longer occurs in the adjusted concatenated sample paths.

(II) $G_t^{(i)} < \widetilde{G}_{t,t+1}^{(k,j)}$ but $\widetilde{G}_{t,t+1}^{(k,j)} \neq \widetilde{F}_{t,t+1}^{(k,j)}$: In this case, even if $\widetilde{G}_{t,t+1}^{(k,j)} \neq \widetilde{F}_{t,t+1}^{(k,j)}$, $\widetilde{G}_{t,t+1}^{(k,j)} \geq G_t^{(i)}$ still holds so the contradiction no longer exists.

The adjustment made to the stock price path $\boldsymbol{S}_{t+1}^{(k,j)}$ is to align with the adjustments to $\boldsymbol{F}_{t+1}^{(k,j)}$ and $\boldsymbol{G}_{t+1}^{(k,j)}$ described above. More specifically, by Equation (5.8) and 5.17, we have

$$\widetilde{F}_{t,t+1}^{(k,j)} = (F_t^{(i)} - I_t^{(i)})^+ \cdot \frac{\widetilde{S}_{t,t+1}^{(k,j)}}{S_t^{(i)}}(1-\eta^g) = F_{t,t+1}^{(k,j)} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} = (F_t^{(k)} - I_t^{(k)})^+ \cdot \frac{S_{t,t+1}^{(k,j)}}{S_t^{(k)}}(1-\eta^g) \cdot \frac{G_t^{(i)}}{G_t^{(k)}}.$$

Then we can deduct that

$$\widetilde{S}_{t,t+1}^{(k,j)} = S_{t,t+1}^{(k,j)} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(i)}}{S_t^{(k)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+} \tag{5.18}$$

For subsequent stock price $\widetilde{S}_{t,s}^{(k,j)}$ where $s \geq t+2$, we assume the log-return of $\widetilde{S}_{t,s}^{(k,j)}$ is identical to that of $S_{t,s}^{(k,j)}$.

We need to consider the effects of this adjustment on the likelihood ratio calculation and on the estimated deltas. As alluded to in Section 5.4.1, the stochasticity of GMWB state variable is driven by the underlying stock S_t, so the likelihood ratio calculation is based on the conditional densities of inner sample paths $\boldsymbol{S}_t$, given the outer scenarios. Moreover, both the conditional density function and the mixture density function remain unchanged under the adjustment. The arguments for the likelihood calculations change from the unadjusted inner path $\boldsymbol{S}_t^{(k,j)}$ to the adjusted inner path $\widetilde{\boldsymbol{S}}_t^{(k,j)}$. Mathematically, the adjusted likelihood ratio is then given by $\dfrac{g_t^{(i)}\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}$. For the delta calculations, note that the proposed adjustment implies $\dfrac{d\widetilde{F}_{t,s}^{(k,j)}}{d\widetilde{F}_{t,t+1}^{(k,j)}} = \dfrac{dF_{t,s}^{(k,j)}}{dF_{t,t+1}^{(k,j)}}$ and $\dfrac{d\widetilde{F}_{t,s}^{(k,j)}}{d\widetilde{G}_{t,s}^{(k,j)}} = \dfrac{dF_{t,s}^{(k,j)}}{dG_{t,s}^{(k,j)}}$ for all

$j = 1, \ldots, N$ and all $s = t+1, \ldots, T$. Therefore, for all $j = 1, \ldots, N$ and all $s = t+1, \ldots, T$, using the chain rule, we have

$$\begin{aligned}\frac{d\widetilde{F}_{t,s}^{(k,j)}}{dS_t^{(i)}} &= \frac{d\widetilde{F}_{t,s}^{(k,j)}}{d\widetilde{F}_{t,t+1}^{(k,j)}} \cdot \frac{d\widetilde{F}_{t,t+1}^{(k,j)}}{dS_t^{(i)}} = \frac{d\widetilde{F}_{t,s}^{(k,j)}}{d\widetilde{F}_{t,t+1}^{(k,j)}} \cdot (\frac{dF_t^{(i)}}{dS_t^{(i)}} - \frac{dI_t^{(i)}}{dS_t^{(i)}})^+ \cdot \frac{\widetilde{S}_{t,t+1}^{(k,j)}}{S_t^{(i)}}(1-\eta^g) \\ &= \frac{d\widetilde{F}_{t,s}^{(k,j)}}{d\widetilde{F}_{t,t+1}^{(k,j)}} \cdot (\frac{F_t^{(i)}}{S_t^{(i)}} - 0)^+ \cdot \frac{\widetilde{F}_{t,t+1}^{(k,j)}}{(F_t^{(i)} - I_t^{(i)})^+} = \frac{d\widetilde{F}_{t,s}^{(k,j)}}{d\widetilde{F}_{t,t+1}^{(k,j)}} \cdot \frac{F_t^{(i)}}{S_t^{(i)}} \cdot \frac{F_{t,t+1}^{(k,j)}}{(F_t^{(i)} - I_t^{(i)})^+} \cdot \frac{G_t^{(i)}}{G_t^{(k)}},\end{aligned}$$

and $\dfrac{dF_{t,s}^{(k,j)}}{dS_t^{(k)}} = \dfrac{dF_{t,s}^{(k,j)}}{dF_{t,t+1}^{(k,j)}} \cdot \dfrac{dF_{t,t+1}^{(k,j)}}{dS_t^{(k)}} = \dfrac{dF_{t,s}^{(k,j)}}{dF_{t,t+1}^{(k,j)}} \cdot \dfrac{F_t^{(k)}}{S_t^{(k)}} \cdot \dfrac{F_{t,t+1}^{(k,j)}}{(F_t^{(k)} - I_t^{(k)})^+}$. Thus,

$$\frac{d\widetilde{F}_{t,s}^{(k,j)}}{dS_t^{(i)}} = \frac{dF_{t,s}^{(k,j)}}{dS_t^{(k)}} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+}$$

Using the chain rule, we also have, for all $j = 1, \ldots, N$ and all $s = t+1, \ldots, T$, $\dfrac{d\widetilde{G}_{t,s}^{(k,j)}}{dS_t^{(i)}} = \dfrac{d\widetilde{G}_{t,s}^{(k,j)}}{d\widetilde{F}_{t,s}^{(k,j)}} \cdot \dfrac{d\widetilde{F}_{t,s}^{(k,j)}}{dS_t^{(i)}}$ and $\dfrac{dG_{t,s}^{(k,j)}}{dS_t^{(k)}} = \dfrac{dG_{t,s}^{(k,j)}}{dF_{t,s}^{(k,j)}} \cdot \dfrac{dF_{t,s}^{(k,j)}}{dS_t^{(k)}}$. Therefore,

$$\frac{d\widetilde{G}_{t,s}^{(k,j)}}{dS_t^{(i)}} = \frac{dG_{t,s}^{(k,j)}}{dS_t^{(k)}} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+},$$

and

$$\begin{aligned}\frac{d\widetilde{I}_{t,s}^{(k,j)}}{dS_t^{(i)}} &= \gamma \frac{d\widetilde{G}_{t,s}^{(k,j)}}{dS_t^{(i)}} = \gamma \frac{dG_{t,s}^{(k,j)}}{dS_t^{(k)}} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+} \\ &= \frac{dI_{t,s}^{(k,j)}}{dS_t^{(k)}} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+},\end{aligned}$$

Consequently, the adjusted simulation output for scenario k and target scenario i is

$$\begin{aligned}&f\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}, \widetilde{\boldsymbol{F}}_t^{(k,j)}, \widetilde{\boldsymbol{G}}_t^{(k,j)}\right) \\ =&f\left(\boldsymbol{S}_t^{(k,j)}, \boldsymbol{F}_t^{(k,j)}, \boldsymbol{G}_t^{(k,j)}\right) \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+}, \ \forall j = 1, \ldots, N,\end{aligned} \tag{5.19}$$

or equivalently in the context of modeling a delta-hedged VA contract,

$$\widetilde{\Delta}_t^{(k,j)} = \widehat{\Delta}_t^{(k,j)} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+}, \ \forall j = 1, \ldots, N.$$

Based on the above discussions, the *adjusted* MLR estimator for the time t delta in scenario i is given by

$$\begin{aligned}
\widehat{\Delta}_t^{(i)\text{MLR}} &= \frac{1}{MN} \sum_{k=1}^{M} \sum_{j=1}^{N} \widetilde{\Delta}_t^{(k,j)} \frac{g_t^{(i)}\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)} \\
&= \frac{1}{MN} \sum_{k=1}^{M} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+} \sum_{j=1}^{N} \widehat{\Delta}_t^{(k,j)} \frac{g_t^{(i)}\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}.
\end{aligned} \tag{5.20}$$

Computationally, we only need to compute the *unadjusted* outputs $\widehat{\Delta}_t^{(k,j)}$ once, then reuse them M times in different target scenarios; the adjustment $\cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+}$ adds minimal computations. Lastly, the adjustment does not affect the Markov property of the state process, so the likelihood ratio calculation can be further simplified for Markov asset models; see Proposition 5.4.1.

Proposition 5.4.1. *If Assumption 5.2.1 holds, and the concatenated stochastic process $(X_t, Y_{t,t+1}, \ldots, Y_{t,T})$ is Markov, then the likelihood ratio estimator in (5.20) can be written as*

$$\begin{aligned}
\widehat{\Delta}_t^{(i)MLR} &= \frac{1}{MN} \sum_{k=1}^{M} \sum_{j=1}^{N} \widetilde{\Delta}_t^{(k,j)} \frac{g_t^{(i)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)} \\
&= \frac{1}{MN} \sum_{k=1}^{M} \cdot \frac{F_t^{(i)}}{F_t^{(k)}} \cdot \frac{G_t^{(i)}}{G_t^{(k)}} \cdot \frac{S_t^{(k)}}{S_t^{(i)}} \cdot \frac{(F_t^{(k)} - I_t^{(k)})^+}{(F_t^{(i)} - I_t^{(i)})^+} \sum_{j=1}^{N} \widehat{\Delta}_t^{(k,j)} \frac{g_t^{(i)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)},
\end{aligned} \tag{5.21}$$

where $\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right) = \frac{1}{M} \sum_{i=1}^{M} g_t^{(i)}(\widetilde{S}_{t,t+1}^{(k,j)})$, *and* $\widetilde{S}_{t,t+1}^{(k,j)} \overset{i.i.d.}{\sim} g_t^{(k)}(s)$, $\widetilde{S}_{t,t+1}^{(k,j)} = S_{t,t+1}^{(k,j)} \cdot \frac{G_t^{(i)}}{G_t^{(k)}}$ *for all* $j = 1, \ldots, N$, *for every* $k = 1, \ldots, M$.

Proof. The proof is identical to the proof for Proposition 5.2.3. □

In Appendix A, we use a simplified and hypothetical GMWB example to demonstrate step-by-step the calculations required in Stage 1 inner simulation of the Section 5.3 algorithm using the MLR estimator in Equation (5.21), as well as all the necessary adjustments outlined in Section 5.4.2.

In Section 5.5, we conduct numerical studies to examine the performance of this two-stage procedure, with adjustments, for the tail risk evaluation of a GMWB.

5.5 Numerical Experiments

We use a standard, multi-period nested simulation and our two-stage procedure to estimate the $\text{CTE}_{95\%}$ of the hedging losses for a delta-hedged GMWB contract. We set up the experiments so that these two procedures use similar runtimes to compare the accuracy of the results.

The GMWB allows the policyholder to make monthly withdrawals of $\gamma = 0.375\%$ of the guarantee value G_t, which is reset (ratcheted) monthly to the higher of the previous month's guarantee value and the current month's sub-account value prior to the withdrawal. The initial sub-account value is F_0 =1,000, which is also the initial guarantee value G_0 =1,000. The GMWB contract has a 20-year maturity and is hedged monthly, so $T = 240$. A management fee of $\eta^g = 0.2\%$ is deducted monthly from the sub-account value and half of it is treated as income for the contract guarantees, i.e., $\eta^n = 0.1\%$.

Similar to the numerical experiments in previous chapters, we assume that the underlying stock follows a regime-switching lognormal (RSLN) asset model with two regimes and parameters as specified in Table 3.2 in Section 3.3.

We first simulated $M = 10,000$ scenarios $\boldsymbol{S}^{(i)}$, $i = 1, \ldots, M$ for the underlying stock, and use these scenarios to compute the scenario paths for the sub-account $\boldsymbol{F}^{(i)}$ and the guarantee value $\boldsymbol{G}^{(i)}$. All experiments in this section will use these scenarios.

As the hedging loss for the GMWB contract under the RSLN model cannot be calculated analytically, we first run a large scale standard nested simulation with $N = 10,000$ inner simulations in each scenario to obtain accurate estimates for the hedging loss. These accurate estimates are used as benchmarks to assess the accuracies for different simulation procedures. We are interested in estimating the $\text{CTE}_{95\%}$ of these benchmark losses, which is the average of the largest 500 benchmark losses corresponding to the "true tail scenarios".

We first provide a holistic view of the performance of the two-stage procedure. Consider a two-stage procedure with $N_1 = 2$ inner sample paths per scenario in Stage 1, and $M^* = 1,500$ highly likely tail scenarios (that is, $10\%M$ safety margin). Under this design, in Stage 1, although only $N_1 = 2$ inner sample paths are generated in each of the $M = 10,000$ outer scenarios, all $N_1 \cdot M = 20,000$ inner sample paths are reused, through the likelihood ratio method, in estimating every outer scenario's loss. Similarly, in Stage 2, $(N_1 + N_2) \cdot M^* = 1,500(N_2 + N_1)$ inner sample paths are reused for each outer scenario.

A common diagnostic for the likelihood ratio estimator is the so-called *effective sample size*, which is the number of i.i.d. samples that would have achieved the same level of accuracy, measured by variance or mean squared error, as the likelihood ratio estimator of interest. Interested readers can refer to Owen (2013), Liu (1996) and Kong (1992).

In Kong (1992) and Liu (1996), under some technical assumptions, it can be shown that the effective sample size, $\widetilde{N}$, of a likelihood ratio estimator can be estimated by

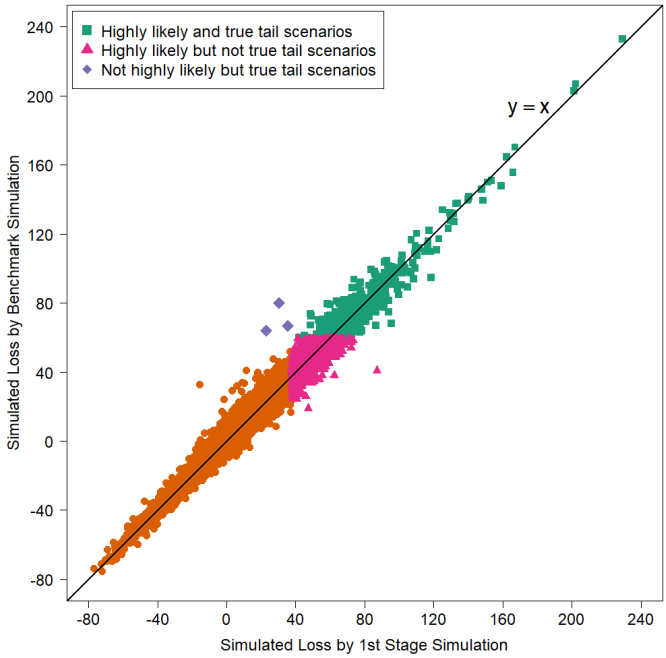
$$\widetilde{N}_t^{(i)} = \frac{\left(\sum_{k=1}^{M}\sum_{j=1}^{N} w_t^{(k,j)}\right)^2}{\sum_{k=1}^{M}\sum_{j=1}^{N}\left(w_t^{(k,j)}\right)^2}$$

where $w_t^{(k,j)} = \dfrac{g_t^{(i)}\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{\boldsymbol{S}}_t^{(k,j)}\right)}$ is the likelihood ratio in the MLR estimator context. We will use this to estimate the effective sample size of the MLR estimator in our numerical studies.

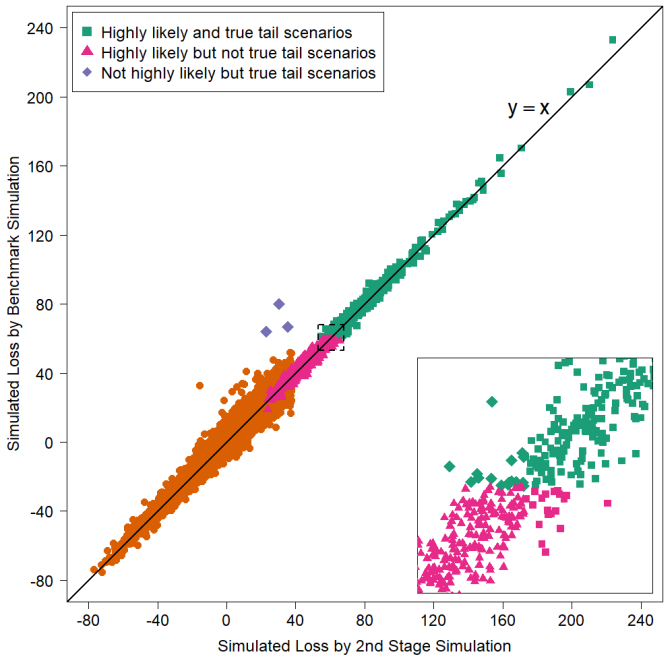
Figure 5.2a depicts the estimated losses in Stage 1 versus the benchmark losses. We see that, with only 2 inner simulations per scenario, the MLR estimates of the losses are close to the benchmark losses, as the points in Figure 5.2a are near the 45-degree line. Moreover, 497 of the 500 true tail scenarios are included in the $M^* = 1,500$ highly likely tail scenarios.

In Stage 2, each of the 1,500 highly likely tail scenarios are assigned an additional $N_2 = 80$ inner simulations. Figure 5.2b shows the estimated losses in Stage 2 versus the benchmark losses. We see that, with the concentrated simulation budget, the Stage 2 simulation significantly improves the accuracy of the loss estimates in the highly likely tail scenarios, which will in turn improve the accuracy of the CTE estimate.

The subfigure in Figure 5.2b zooms into the border between tail and non-tail scenarios: the scenarios that are included in the CTE estimate in our two-stage procedure are labeled by squares. In this illustration, the CTE estimate included 481 of the 500 true tail scenarios.



(a) Simulated losses by Stage 1 simulation (x axis) and by the benchmark simulation (y axis).



(b) Simulated losses by Stage 2 simulation (x axis) and by the benchmark simulation (y axis).

Figure 5.2: Illustration of the proposed two-stage simulation procedure.

Among the 19 tail scenarios that were not included in the CTE estimate, 3 were missed in Stage 1, and 16 were missed in Stage 2.

Next we examine the performance of the two-stage procedure in more details. We repeat the following four experiments 100 times:

Experiment (a): The proposed two-stage procedure with

(a1) $N_1 = 2$, $M^* = 1,500$, and $N_2 = 80$ (the same configuration as in Figures 5.2a and 5.2b),

(a2) $N_1 = 2$, $M^* = 1,000$, and $N_2 = 170$, and

(a3) $N_1 = 2$, $M^* = 500$, and $N_2 = 620$.

Experiment (b): Standard multi-period nested simulation with $N = 350$ inner simulations for each outer scenario.

We use the same set of $M = 10,000$ outer scenarios for each repetition of each method listed above. They are also the same set of outer scenarios in the benchmark experiment shown in Figure 5.2a-5.2b. Thus, the repetitions are solely capturing the sampling variability at the inner simulation stage. In addition, all experiments using two-stage nested simulation with MLR estimators share the same 100 independent repetitions of Stage 1 output.

We selected the above configurations so that the average runtime for all the experiments are similar. Within each experiment design, we illustrate in Table 5.1 the proportion of total computation budget spent on the sample path generation steps, Step (1.A) and (2.A), and the likelihood calculation steps, Step (1.B) and (2.B) of the algorithm in Section 5.3, respectively.

Experiment Design	Stage 1		Stage 2	
	Sample Paths Generation	Likelihood Calculation	Sample Paths Generation	Likelihood Calculation
(a) Two-stage using MLR:				
a1. $N_1 = 2$, $M^* = 1,500$, $N_2 = 80$	1%	51%	4%	45%
a2. $N_1 = 2$, $M^* = 1,000$, $N_2 = 170$	1%	51%	5%	44%
a3. $N_1 = 2$, $M^* = 500$, $N_2 = 620$	1%	51%	9%	39%
(b) Standard nested sim. with $N = 350$	100%	n/a	n/a	n/a

Table 5.1: Proportion of total computation budget deployed in each step of the simulation procedure.

We can see from Table 5.1 that in experiment (a1), (a2) and (a3), Stage 1 uses about 52% of total computation budget while the remaining 48% are spent in Stage 2. In both stages, the majority of the computation budget is spent on likelihood calculation. The amount of computation budget spent in Stage 1 is the same for all experiments using MLR estimators because they use the same stage 1 output. In Stage 2, as the number of highly likely tail scenarios M^* decreases, a higher proportion of computation budget is allocated to the sample paths generation step in this stage because the amount of computation required in likelihood calculation decreases with M^*.

Each of the 100 repetitions of the above experiments produces a $\text{CTE}_{95\%}$ estimate. By comparing these estimates with the benchmark $\text{CTE}_{95\%}$ estimate we can estimate the RMSE for each of the four experiments. The relative biases and relative variances for the four experiments can also be estimated. Table 5.2 summarizes these performance measures. The same results are also presented in Figure 5.3.

We make the following observations from Table 5.2.

Experiment (b), standard nested simulation, has high relative bias and high relative variance. In our experiment, this is caused by insufficient number of inner simulations, which is much smaller than the number of outer simulations. The standard Monte Carlo experiment has a significant positive bias because of insufficient number of inner simulations used, which is consistent with findings in Gordy and Juneja (2010). The intuition is that the small number of inner simulations causes significant noise in the number of out-of-the-money inner simulation paths, and subsequently in the individual loss estimation from each outer scenario. Because the CTE estimation picks up the highest ranked $(1 - \alpha)M$

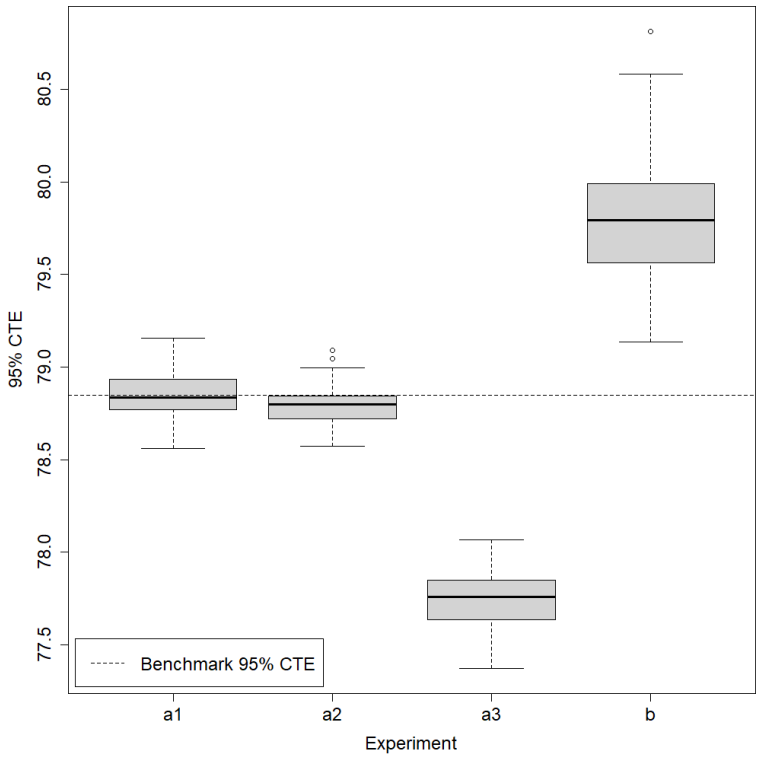


Figure 5.3: Box-and-whisker plot of results from 100 repetitions of two-stage nested simulation using MLR estimators, and standard nested simulation, GMWB example.

Experiment Design	RMSE	Relative Bias	Relative Variance
(a) Two-stage procedures			
a1. $N_1 = 2$, $M^* = 1,500$, $N_2 = 80$	0.0168%	-0.002%	0.017%
a2. $N_1 = 2$, $M^* = 1,000$, $N_2 = 170$	0.0171%	-0.069%	0.013%
a3. $N_1 = 2$, $M^* = 500$, $N_2 = 620$	1.5826%	-1.405%	0.026%
(b) Standard nested sim. with $N = 350$	1.2578%	1.201%	0.121%

Table 5.2: Accuracy measures from 100 repetitions of different configurations of the two-stage procedure and the standard nested simulation

individual losses, the individual losses end up in the CTE estimation tend to be the ones that are higher than its true loss value due to inner simulation noise. Therefore, the CTE estimation is inherently biased high when there are insufficient number of inner simulations.

The higher variance in the standard nested simulation experiments is also due to insufficient number of inner simulations. Gordy and Juneja (2010) found that the variance in nested simulation is primarily caused by an insufficient number of outer scenarios in independent repetitions with newly generated outer scenarios in each repetition. However, in our experiment, we limit ourselves to the same set of 10,000 outer scenarios in all experiments. In our SMC experiment, the inner simulation noise causes sampling variance in the loss estimation of each scenario.

In Experiment (a1) and (a2), even though the number of inner simulations in the highly likely tail scenarios ($N_1 + N_2 = 82$ and 172, respectively) are smaller than that in the standard nested simulation ($N = 350$), the $\text{CTE}_{95\%}$ estimate in experiment (a1) and (a2) are more accurate than those in experiment (b), because the estimated effective sample size of inner simulations in (a1) and (a2) is bigger than $N = 350$. In Figure 5.4, we illustrate the estimated effective sample size over time in one of the repetitions under different experiment designs.

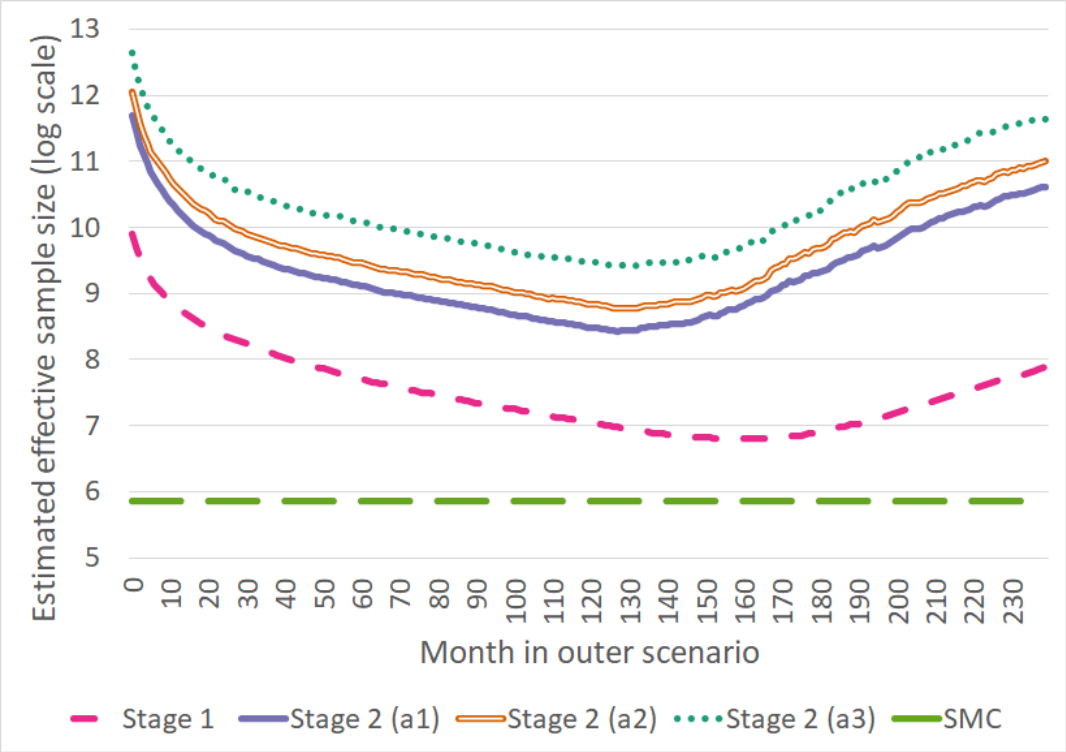


Figure 5.4: Estimated effective sample size (in log scale) in one repetition of Stage 1 MLR, Stage 2 MLR, and SMC experiment.

We can see from Figure 5.4 that in the Stage 1 experiment, which is shared among all three MLR experiment designs, the estimated effective sample size starts at $N_1 \cdot M = 20,000$ ($\ln(20,000) = 9.90$) and decreases as time progresses along the outer scenarios. The state variables in outer scenarios cause the effective sample size to decrease over time. Towards the end of the simulation, the effective sample size increases because in many scenario, the sub-account of the GMWB has depleted so the state variables start to converge at this point. The smallest estimated effective sample size in Stage 1 is slightly below 1,000, still higher than the $N = 350$ inner simulations used in the standard Monte Carlo simulation. In the Stage 2 experiment, the estimated effective sample size follows similar trend as in Stage 1 in all three experiment designs. The lowest estimated effective sample size is close to 5,000 in experiment (a1) and (a2), and slightly above 12,000 in experiment (a3).

With a reasonable safety margin, the two-stage procedure produces accurate $\text{CTE}_{95\%}$ estimates. The number of highly likely tail scenarios in Experiments (a1) and (a2) are $M^* = 1,500$ and $M^* = 1,000$, which represent $5\%M$ and $10\%M$ safety margins, respectively.

Both of these two experiments produces $\text{CTE}_{95\%}$ estimates that are orders of magnitudes more accurate than that in the standard nested simulation (74 and 94 times smaller MSEs, respectively). The RMSEs, relative biases, and relative variances for Experiments (a1) and (a2) are all significantly smaller than those for the standard nested simulation. But if no safety margin is given, as in Experiment (a3), the two-stage procedure could miss many true tail scenario in Stage 1, thus poorly estimate tail risk measures.

The negative bias in the two-stage procedure is caused by the misclassification of tail scenario. The highly likely tail scenarios in Stage 1 may exclude some true tail scenarios (with large losses). Some non-tail scenarios' small losses will be accurately estimated in Stage 2 and be included in the $\text{CTE}_{95\%}$ estimate. This observation reiterates the importance of setting reasonable safety margins.

To further examine the misclassification of tail scenarios in the two-stage procedure, Table 5.3 summarizes the number of the true tail scenarios included in Stage 2 simulation and in CTE estimation, in each of the 100 repeated experiments. In the standard nested simulation, the tail scenario set in the CTE estimate includes merely 434.4 true tail scenarios on average. When no safety margin is added, as in Experiment (a3), none of the 100 repeated experiments captures all 500 true tail scenarios in the highly likely tail scenario set $\widehat{\mathcal{T}}^{HL}_{M^*}$, which on average includes 438.6 of the 500 true tail scenarios. While this is number higher (thus better) than the standard nested simulation, there is significant room for improvement. Increasing the safety margin to $5\%M$, as in Experiment (a2), drastically reduces the misclassification of tail scenarios. On average, 498.0 (99.6%) true tail scenarios are included in the highly likely tail scenario set $\widehat{\mathcal{T}}^{HL}_{M^*}$. Moreover, in 17 of 100 repeated experiments, the highly likely tail scenario set $\widehat{\mathcal{T}}^{HL}_{M^*}$ includes all true tail scenarios. Experiment (a2) includes 478.9 true tail scenarios in the CTE estimate on average, which is much higher than that in the standard nested simulation. Lastly, when the number of highly likely tail scenarios is further increased to $M^* = 1,500$, as in Experiment (a1), all tail scenarios are included in $\widehat{\mathcal{T}}^{HL}_{M^*}$ in 66 of the 100 repeated experiments. As we see from Table 5.2, the slightly different misclassification between Experiments (a1) and (a2) has little effect on their RMSEs.

5.6 Conclusion

In this chapter, we present a two-stage nested simulation procedure for estimating the tail risks of dynamically hedged complex VA contracts. The mixture likelihood ratio estimator is used in both stages to reuse simulation outputs and to improve the estimation accuracy.

Experiment Design	# true tail scen. in $\widehat{\mathcal{T}}^{HL}_{M^*}$	# of repetitions $\widehat{\mathcal{T}}^{HL}_{M^*}$ includes all true tail scenarios	# true tail scen. in $\widehat{\mathcal{T}}^{MLR}_{(1-\alpha)M}$
(a) Two-stage procedures			
a1. $N_1 = 2$, $M^* = 1,500$, $N_2 = 80$	499.6	66/100	476.9
a2. $N_1 = 2$, $M^* = 1,000$, $N_2 = 170$	498.0	17/100	478.9
a3. $N_1 = 2$, $M^* = 500$, $N_2 = 620$	438.6	0/100	438.6
(b) Standard nested sim. with $N = 350$	n/a	n/a	434.4

Table 5.3: True tail scenarios captured in 100 repetitions of two-stage nested simulation using MLR estimator of different configurations of the two-stage procedure and the standard nested simulation. Standard error indicated in brackets.

It eliminates the need for extrinsic proxy models used in the IANS and DIANS method. In an in-depth numerical study, we apply the proposed two-stage procedure on the Guaranteed Minimum Withdrawal Benefit (GMWB). With a similar runtime, our procedure produces CTE estimates that are significantly more accurate than that of the standard nested simulation.

Chapter 6

Future Work

In this chapter, we discuss a few methods that could potentially further improve the methods of the previous chapters. We first introduce a simulation procedure that combines the IANS procedure with the MLR estimator. Then we discuss a screening procedure that is similar to Stage 1 of our proposed nested simulation procedures. After that, we illustrate how the jackknife method proposed in the literature can be applied to reduce bias and further improve the accuracy of tail risk measure estimation.

6.1 IANS Procedure with MLR Estimator

As we presented in Algorithm 2, Stage 1 of the IANS procedure uses a closed-form proxy to identify a set of proxy tail scenarios, while Stage 2 carries out standard nested simulation on these proxy tail scenarios. In fact, when we have a reliable extrinsic proxy, we can replace the Stage 2 simulation in the IANS procedure with nested simulation using the MLR estimator on the proxy tail scenarios to achieve even greater computation efficiency. In other words, we can replace Line (II.1) to (II.4) Algorithm 2 with Step (2.A) to (2.D) of the two-stage nested simulation using MLR estimators.

We have conducted a numerical experiment using this IANS procedure with MLR estimator. We use the same GMMB with dynamic lapse example as in Section 4.3. Similar to all experiments in Section 4.3, the experiment uses the same $M = 5,000$ outer scenarios as in the benchmark experiment in Section 4.3.2. The proxy tail scenario set $\mathcal{T}_m^P$ has $m = 750$, that is $m = 15\%M$. In the Stage 2 simulation, $N_2 = 90$ inner simulations were conducted for each scenario in $\mathcal{T}_{750}^P$. This experiment uses about 50% of the computation

budget as the experiments listed in Table 4.1. We repeat this experiment 100 times using the IANS procedure with MLR estimator, same as the experiments listed in Table 4.1.

Table 6.1 summarizes results from the 100 repetitions. Compared to the results in

Experiment	m	N_2	RMSE
Fixed IANS with MLR estimator	750	90	0.0140% (0.0022%)
(a) Dynamic IANS, $m_0 = 400$	≈ 654	$\approx 1,528$	0.0072% (0.0010%)
(c1) Fixed IANS	750	1,333	0.0088% (0.0013%)
(c2) Fixed IANS	500	2,000	0.0069% (0.0011%)
(c3) Fixed IANS	250	4,000	7.8604% (0.0274%)

Table 6.1: Results from 100 repetitions of fixed IANS procedure with MLR estimator, fixed and dynamic IANS process, GMMB example, Standard error of the results indicated in bracket. All values are based on a single outer scenario set, $\boldsymbol{X}$. Results from the fixed and dynamic IANS process are identical to those in Table 4.1.

all experiments listed in Table 4.1, the IANS procedure with MLR estimator achieves similar level of accuracy as the experiments with the most accurate CTE estimate in Table 4.1 (experiment (a), (c1) and (c2)), but uses only 50% of the computation budget used by experiments in Table 4.1. This improvement demonstrates the additional gain in computation efficiency by incorporating MLR estimator in the IANS procedure.

Compared to the fixed IANS procedure, we can choose a more relaxed proxy tail scenario set size m in IANS procedure using MLR estimator because the computation effort spent on the extra proxy tail scenarios is not entirely "wasted" when the inner simulation output is estimated by an MLR estimator. Inner simulation output from these extra proxy tail scenarios is also included in the MLR estimator.

Note that even though as discussed in Chapter 5, the DIANS procedure is more robust than the IANS procedure in the sense that it gives statistical assurance that sufficient number of true tail scenarios have been captured, it is inefficient to incorporate MLR estimator in the DIANS procedure because of the iterative additive nature of the DIANS process.

6.2 Screening Procedure for Likely Tail Scenarios

Lan *et al.* (2010) introduced a two-stage nested simulation design similar to our proposal,

but in a single-period nested simulation setting. In this setting, the loss is defined as an expectation, e.g. the expected value of a portfolio based on risk neutral valuation. In their design, the Stage 1 simulation uses results from a small pilot standard nested simulation (as few as 30 inner simulations based on their suggestion) to select the scenarios that are highly likely to be the true tail scenarios. In Stage 2 more standard nested simulation is carried out only to the highly likely tail scenarios to build a confidence interval for the CTE estimate.

In theory, their screening procedure for highly likely tail scenarios could replace the Stage 1 simulation of our proposed procedure. More precisely in the context of our multi-period nested simulation problem, the loss random variable L is a linear function of various conditional expectations $\mu_t(X_t) = \mathbb{E}[f(\boldsymbol{Y}|X_t)]$. In Equation (2.5), given a outer scenario, L is a linear function of Δ_t's, which are conditional expectations given the outer scenario state variables. As such, in cases where all the Δ_t's are estimated with the same number of inner simulations, we can express the loss random variable L as an expectation of a linear function of $f(\boldsymbol{Y}|X_t)$. That is, based on Equation (2.5)

$$L = \sum_{t=0}^{T-1} \Delta_t (D_t S_t - D_{t+1} S_{t+1}) + V_0 = \sum_{t=0}^{T-1} \mathbb{E}[f(\boldsymbol{Y}|X_t)](D_t S_t - D_{t+1} S_{t+1}) + V_0 \tag{6.1}$$

$$= \mathbb{E}\left[\sum_{t=0}^{T-1} f(\boldsymbol{Y}|X_t)(D_t S_t - D_{t+1} S_{t+1})\right] + V_0 \tag{6.2}$$

According to Lan *et al.* (2010), since L is an expectation, we can apply a pair-wise student-t test to each $\widehat{L}_i, i = 1, \ldots, M$ from Stage 1 inner simulation samples, regardless of the nested simulation procedure used. The student-t test will then indicate, at a given confidence level, which outer scenario has a sample mean smaller than $(1-\alpha)M$ or fewer outer scenarios, which means that the scenario is in $\mathcal{T}_{(1-\alpha)M}$ at the specified confidence level.

However, in practice, when we applied the screening procedure in a multi-period setting, it ends up screening out only a small portion of the outer scenarios, given a small computation budget allocated to the screening procedure. For the GMWB contract, we apply the Lan *et al.* (2010) screening procedure to Stage 1 simulation output from the MLR procedure described in Section 5.5. The resulting highly likely tail scenario set has a size of $53\%M$, which does not allow for sufficient Stage 2 inner simulations under a fixed computation budget.

Similarly, when we apply the screening procedure to the GMMB with dynamic lapses from Section 4.3, with 30 inner simulation in the pilot simulation, the screening procedure did not screen out any scenarios. In other words, the suggested highly likely tail scenario set has a size of M. When we apply 900 inner simulations in the pilot simulation, the screening procedure gives a set of 467 scenarios out of 5,000 scenarios at 99% confidence level. This set includes all true tail scenarios. but this pilot simulation alone uses 4.5 times as much computation budget as the experiments in Section 4.3.4.

Moreover, the pairwise student-t test requires estimating the standard errors for $\frac{M(M-1)}{2}$ pairs of scenarios. For $M = 10,000$ in our numerical study, the screening procedure itself takes longer to run than our entire two-stage procedure.

It is obvious based on these experiments that the screening procedure proposed in Lan *et al.* (2010) is not very useful in the multi-period nested simulation setting that we are considering. As we consider dynamic hedging for complex variable annuities, i.e., multi-period nested simulation with path-dependent embedded option, the estimated losses in Stage 1 have large sample variances. As noted in Lan *et al.* (2010), large sample variances make it more difficult to screen out scenarios using their pairwise student-t test.

We also note that not all losses in nested simulation can be expressed as an expectation of linear function. For example, in the NAIC Valuation Manual 21, the loss is defined as the greatest present value of the projected accumulated deficiencies over the projection horizon (NAIC, 2020). In this case, the screening procedure cannot be applied in the multi-period nested simulation case because the overall loss L cannot be defined as an expectation.

6.3 Jackknife Method

In Gordy and Juneja (2010), the authors show how the jackknife method can be applied in a single-period nested simulation setting, to reduce bias in the tail risk measure estimation. Again, in this setting, the loss is defined as an expectation.

According to Gordy and Juneja (2010), the jackknife method is applied in a nested simulation of tail risk measure as follows. Given our nested simulation procedure with N inner simulations, we divide the N inner simulation sample output into Z partition, and we have N/Z inner samples in each partition (assuming N/Z is an integer). Let CTE^{-z} denote the CTE estimate by omitting the inner samples in the z-th partition, where $z = 1, \dots Z$, that is, by pretending the N/Z inner simulation samples in the z-th partition do not exist. Let CTE^{N} denote the CTE estimate based on all N inner simulation sample output. Then

the bias-corrected jackknife CTE estimate is

$$\mathrm{CTE}_{\mathrm{JK}} = \mathrm{CTE}^{N} + \frac{Z-1}{Z}\sum_{z=1}^{Z}(\mathrm{CTE}^{N} - \mathrm{CTE}^{-z}) \tag{6.3}$$

As discussed in Section 6.2, the loss L that we consider in this thesis can be defined as an expectation of linear function, so the jackknife estimator can be applied at virtually no additional computation cost in standard nested simulations, and all of our proposed procedures.

We have conducted some numerical experiments using the jackknife method on GMWB contract described in Section 5.5. We conducted 20 repetitions for each of the following experiment:

(a) Two-stage nested simulation using MLR estimator, with $M^* = 1,500$, $N_1 = 2$, $N_2 = 10$

(b) Standard nested simulation with $N = 200$ inner simulations for all outer scenarios

These two sets of experiments have similar run time. We do not apply the jackknife method directly to the simulation output from numerical experiments in Section 5.5 because the bias in the MLR experiments are either negligible or are caused by the missing true tail scenarios, which cannot be remedied by the jackknife method.

The RMSE, relative bias and relative variance of $\mathrm{CTE}_{95\%}$ from the two sets of experiments are summarized in Table 6.2.

Experiment Design	RMSE		Bias	Variance
Without jackknife				
(a) Two-stage MLR	0.4850%	(0.0842%)	0.701%	0.098%
(b) SMC	4.3849%	(0.3994%)	2.309%	0.180%
With jackknife				
(a) Two-stage MLR	0.1279%	(0.0417%)	-0.191%	0.099%
(b) SMC	0.2653%	(0.0885%)	0.058%	0.263%

Table 6.2: Results from 20 repetitions of two-stage nested simulation using MLR estimator, and standard nested simulation, GMWB example, jackknife method applied in CTE estimate. Standard error of RMSE are indicated in bracket. All values are based on the same outer scenario set, $\boldsymbol{X}$.

We can see from Table 6.2 that the jackknife method reduces bias, and consequently, the RMSE of the CTE estimation in both standard nested simulation and nested simulation using MLR. However, the reduction in bias and RMSE is more prominent in standard nested simulation than in the MLR case. Designing a more effective bias reduction method for the MLR procedure, as well as the impact of applying the jackknife estimator on RMSE based on different configurations of the MLR and standard nested simulation procedures will be a topic for future study.

6.4 Other Future Work

Future research topics that continue this work can be roughly split into two categories:

1. Further refinements to the methods developed in this thesis.
2. Adapting or developing methodology for a wider range of financial risk management applications.

Further refinements to our proposals in this thesis include the following.

- We have worked under the premise of a fixed set of outer scenarios in this thesis. This is consistent with commonly used practice in insurance companies. However, it is worthwhile to consider the impact of outer-level simulation design on computation efficiency in a nested simulation.
- We will consider convergence analysis in each stage of our proposed procedure. This will help us identify the optimal design variables, such us the number of outer and inner simulations, and the size of proxy tail scenario set.
- We will consider additional variance and bias reduction techniques that complement our procedures, to further improve accuracy.

To adapt the procedures for a wider range of financial applications, we will focus on extending the MLR method to other simulation-based calculations in financial reporting and risk management practice. To facilitate decision-making and fulfill regulatory requirement, many practice areas in insurance companies and other financial institutions now rely on real-time or near real-time calculation of various financial and risk metrics. Due to the sophisticated nature of today's financial contracts, many of these calculations are based

on simulation models. The likelihood ratio method can be extended to pre-process the simulation and calculate the financial and risk metrics based on real-time market variables whenever results are required. This will help significantly reduce the time required to complete simulation-based calculations. In addition to the likelihood ratio method, machine learning techniques such as stochastic kriging and neural networks may be explored to recycle simulation output for fast computation.

In addition, we will consider other types of nested simulation. As we described in this thesis, in most of the existing literature on nested simulation, the first-level simulation estimates the tail risk measure of a loss distribution or probability of large loss while the second-level simulation estimates the mean of a distribution. However, in other nested simulation applications, for example, under the Principle-Based Reserving regime in the U.S., some pricing calculations require the first-level simulation to estimate the mean of a distribution and the second-level simulation to estimate the tail of a distribution, while other applications require both levels of simulation to estimate the tail of a distribution (iterated risk measures). Under these circumstances, existing proposals in the literature that use various regression-based proxy methods to replace the second-level simulation are likely to fail, because they are less effective for estimating the tail of a distribution rather than the mean. In these cases, methods to accurately identify tail scenarios will be more general and useful. For example, machine learning methods can be used for choosing tail scenarios versus non-tail scenarios. Such methods will then be useful not only in a nested simulation setting, but also in applications where only one layer of Monte Carlo simulation is used to determine tail risk measures.

Chapter 7

Conclusion

In both insurance and finance, risk management of more complex products, with longer horizons, and with more complex economic capital requirements is creating computational complexity that challenge even the most sophisticated computing environments. The development of efficient, accurate and implementable computational tools is important and timely.

In this thesis, we have presented two variations of two-stage nested simulation procedures for estimating tail risk measures of Variable Annuities: an Importance Allocated Nested Simulation (IANS) procedures in Chapter 3 and 4, and the mixture likelihood ratio estimator in Chapter 5. Both procedures take advantage of the special structure of the CTE by first identifying a small set of potential tail scenarios from the Stage 1 simulation, based on a proxy for loss. The proxy is either evaluated from a closed-form calculation, or based on a pilot nested simulation using MLR estimator. We then focus the simulation budget on only those scenarios in the Stage 2 simulation.

In cases where a closed-form proxy can be easily derived, as we've shown for the GMMB and GMAB VA contracts, the IANS procedure is significantly more efficient than standard nested simulation. Furthermore, we have developed the Dynamic Importance Allocated Nested Simulation procedure to dynamically determine the number of proxy tail scenarios to be included in Stage 2 simulation. The DIANS method is a systematic approach that helps us find all true tail scenarios with high confidence.

In cases where a closed-form proxy cannot be easily derived, for example in the case of a GMWB contract with ratchet feature, we adapt the two-stage procedure to use only intrinsic information, with the MLR estimator. Nested simulation with MLR is not limited to any contract type, as long as the asset and/or any other stochastic state variables follow

a Markov process. We first use output from a rough MLR nested simulation with very few inner simulations as a proxy for identifying tail scenarios. Then in Stage 2 simulation, we apply the MLR nested simulation again but with more inner simulations to achieve a more accurate tail risk measure estimate. The two-stage nested simulation using MLR estimator is also much more efficient than the standard nested simulation procedure.

These new proposals are developed for multi-period nested simulations used in insurance applications, particularly for estimating tail risk measures of VAs. They reduce the computational burden of multi-period nested simulations that many existing techniques cannot address or at least not as efficiently because most of the existing techniques are designed for single-period nested simulation.

References

Aggarwal A, Beck MB, Cann M, Ford T, Georgescu D, Morjaria N, Smith A, Taylor Y, Tsanakas A, Witts L, *et al.* (2016). "Model risk-daring to open up the black box." *British Actuarial Journal*, **21**(2), 229–296.

Ankirchner S, Schneider JC, Schweizer N (2014). "Cross-hedging minimum return guarantees: Basis and liquidity risks." *Journal of Economic Dynamics and Control*, **41**, 93–109.

Artzner P, Delbaen F, Eber JM, Heath D (1999). "Coherent measures of risk." *Mathematical Finance*, **9**(3), 203–228.

Augustyniak M, Boudreault M (2015). "On the importance of hedging dynamic lapses in variable annuities." *Risks & Rewards, Society of Actuaries*, **66**, 12–16.

Augustyniak M, Boudreault M (2017). "Mitigating interest rate risk in variable annuities: An analysis of hedging effectiveness under model risk." *North American Actuarial Journal*, **21**(4), 502–525.

Augustyniak M, Godin F, Simard C (2017). "Assessing the effectiveness of local and global quadratic hedging under GARCH models." *Quantitative Finance*, **17**(9), 1305–1318.

Basel Committee on Banking Supervision (2019). "Minimum capital requirements for market risk." *Technical report*, Basel Committee on Banking Supervision. URL https://www.bis.org/bcbs/publ/d457.pdf.

Bauer D, Gao J, Moenig T, Ulm ER, Zhu N (2017). "Policyholder exercise behavior in life insurance: the state of affairs." *North American Actuarial Journal*, **21**(4), 485–501.

Bauer D, Ha H (2015). "A least-squares Monte Carlo approach to the calculation of capital requirements." In *World Risk and Insurance Economics Congress, Munich, Germany,*

August, pp. 2–6. URL https://danielbaueracademic.files.wordpress.com/2018/02/habauer_lsm.pdf.

Bauer D, Kling A, Russ J (2008). "A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities." *ASTIN Bulletin: The Journal of the IAA*, **38**(2), 621–651.

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.

Bernard C, Hardy M, MacKay A (2014a). "State-dependent fees for variable annuity guarantees." *ASTIN Bulletin: The Journal of the IAA*, **44**(3), 559–585.

Bernard C, MacKay A, Muehlbeyer M (2014b). "Optimal surrender policy for variable annuity guarantees." *Insurance: Mathematics and Economics*, **55**, 116–128.

Bollen NP (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), 307–327.

Boyle P, Hardy M (2003). "Guaranteed annuity options." *ASTIN Bulletin: The Journal of the IAA*, **33**(2), 125–152.

Boyle PP, Emanuel D (1980). "Discretely adjusted option hedges." *Journal of Financial Economics*, **8**(3), 259–282.

Boyle PP, Hardy MR (1997). "Reserving for maturity guarantees: Two approaches." *Insurance: Mathematics and Economics*, **21**(2), 113–127.

Boyle PP, Schwartz ES (1977). "Equilibrium prices of guarantees under equity-linked contracts." *Journal of Risk and Insurance*, pp. 639–660.

Broadie M, Du Y, Moallemi CC (2011). "Efficient risk estimation via nested sequential simulation." *Management Science*, **57**(6), 1172–1194.

Broadie M, Du Y, Moallemi CC (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.

Cathcart MJ, Lok HY, McNeil AJ, Morrison S (2015). "Calculating variable annuity liability greeks using Monte Carlo simulation." *ASTIN Bulletin: The Journal of the IAA*, **45**(2), 239–266.

Chahboun I, Hoover N (2019). "Variable Annuities: Underlying Risks and Sensitivities." *FRB Boston Risk and Policy Analysis Unit Paper No. RPA*, pp. 19–1.

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.

David H (1973). "Concomitants of order statistics." *Bulletin of the International Statistical Institute*, **45**(1), 295–300.

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.

Devolder P, Lebègue A (2017). "Iterated VaR or CTE measures: A false good idea?" *Scandinavian Actuarial Journal*, **2017**(4), 287–318.

Drexler A, Plestis T, Rosen RJ, *et al.* (2017). "How much risk do variable annuity guarantees pose to life insurers." *Chicago Fed Letter*, **384**.

Duan JC (1995). "The GARCH option pricing model." *Mathematical Finance*, **5**(1), 13–32.

EIOPA (2014). "The underlying assumptions in the standard formula for the Solvency Capital Requirement calculation." *Technical report*, European Insurance and Occupational Pensions Authority. URL https://eiopa.europa.eu/Publications/Standards/EIOPA-14-322_Underlying_Assumptions.pdf.

Feng M, Staum J (2017). "Green simulation: Reusing the output of repeated experiments." *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **27**(4), 1–28.

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 (2016). "2015 Society of Actuaries Survey on Nested Stochastic Modeling." *Technical report*, Society of Actuaries. URL https://www.soa.org/globalassets/assets/files/static-pages/research/2015-nested-stochastic-modeling.pdf.

Feng R (2018). *An introduction to computational risk management of equity-linked insurance*. CRC Press.

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.

Feng R, Jing X (2017). "Analytical valuation and hedging of variable annuity guaranteed lifetime withdrawal benefits." *Insurance: Mathematics and Economics*, **72**, 36–48.

Feng R, Volkmer HW (2016). "An identity of hitting times and its application to the valuation of guaranteed minimum withdrawal benefit." *Mathematics and Financial Economics*, **10**(2), 127–149.

Fu MC, *et al.* (2015). *Handbook of simulation optimization*, volume 216. Springer.

Fung MC, Ignatieva K, Sherris M (2014). "Systematic mortality risk: An analysis of guaranteed lifetime withdrawal benefits in variable annuities." *Insurance: Mathematics and Economics*, **58**, 103–115.

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 (2015). "Application of Metamodeling to the Valuation of Large Variable Annuity Portfolios." In *Proceedings of the 2015 Winter Simulation Conference*, pp. 1103–1114. doi:10.1109/WSC.2015.7408237.

Gan G, Lin XS (2015). "Valuation of large variable annuity portfolios under nested simulation: A functional data approach." *Insurance: Mathematics and Economics*, **62**, 138–150.

Gan G, Lin XS (2017). "Efficient greek calculation of variable annuity portfolios for dynamic hedging: A two-level metamodeling approach." *North American Actuarial Journal*, **21**(2), 161–177.

Gan G, Valdez EA (2019). *Metamodeling for Variable Annuities*. CRC Press.

Geneva Association (2013). "Variable Annuities-an analysis of financial stability." *The Geneva Association, Geneva*.

Glasserman P (2013). *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media.

Goldman MB, Sosin HB, Gatto MA (1979). "Path dependent options: Buy at the low, sell at the high." *The Journal of Finance*, **34**(5), 1111–1127.

Gordy MB, Juneja S (2010). "Nested simulation in portfolio risk measurement." *Management Science*, **56**(10), 1833–1848.

Hardy MR (2001). "A regime-switching model of long-term stock returns." *North American Actuarial Journal*, **5**(2), 41–53.

Hardy MR (2003). *Investment guarantees: modeling and risk management for equity-linked life insurance*, volume 215. John Wiley & Sons.

Hardy MR, Wirch JL (2004). "The iterated CTE: a dynamic risk measure." *North American Actuarial Journal*, **8**(4), 62–75.

Hesterberg TC (1988). *Advances in importance sampling*. Ph.D. thesis, Stanford University.

Holton G (2003). *Value-at-risk: Theory and Practice*. Academic Press advanced finance series. Academic Press. ISBN 9780123540102.

Huang H, Milevsky MA, Salisbury TS (2014). "Optimal initiation of a GLWB in a variable annuity: No Arbitrage approach." *Insurance: Mathematics and Economics*, **56**, 102–111.

Hyndman RJ, Fan Y (1996). "Sample quantiles in statistical packages." *The American Statistician*, **50**(4), 361–365.

Insurance Information Institute (2020). "Facts + Statistics: Annuities." *Technical report*.

Kim JHT, Hardy MR (2007). "Quantifying and correcting the bias in estimated risk measures." *ASTIN Bulletin: The Journal of the IAA*, **37**(2), 365–386.

Kling A, Ruez F, Ruß J (2011). "The impact of stochastic volatility on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities." *ASTIN Bulletin: The Journal of the IAA*, **41**(2), 511–545.

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

Knoller C, Kraut G, Schoenmaekers P (2016). "On the propensity to surrender a variable annuity contract: An empirical analysis of dynamic policyholder behavior." *Journal of Risk and Insurance*, **83**(4), 979–1006.

Kolkiewicz A, Liu Y (2012). "Semi-static hedging for GMWB in variable annuities." *North American Actuarial Journal*, **16**(1), 112–140.

Kong A (1992). "A Note on Importance Sampling using Standardized Weights." *Technical report*, Department of Statistics, University of Chicago.

Krah AS, Nikolić Z, Korn R (2018). "A least-squares Monte Carlo framework in proxy modeling of life insurance companies." *Risks*, **6**(2), 62.

Lan H, Nelson BL, 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.

Li P, Feng R (2021). "Nested Monte Carlo simulation in financial reporting: a review and a new hybrid approach." *Scandinavian Actuarial Journal*, pp. 1–35.

Lin XS, Yang S (2020a). "Efficient dynamic hedging for larget variable annuity portfolio with multiple underlying assets." *ASTIN Bulletin: The Journal of the IAA*, **50**, 913–957.

Lin XS, Yang S (2020b). "Fast and efficient nested simulation for large variable annuity portfolios: A surrogate modeling approach." *Insurance: Mathematics and Economics*, **91**, 85–103.

Liu JS (1996). "Metropolized independent sampling with comparisons to rejection sampling and importance sampling." *Statistics and computing*, **6**(2), 113–119.

Liu M, Staum J (2010). "Stochastic kriging for efficient nested simulation of expected shortfall." *Journal of Risk*, **12**(3), 3.

Longstaff FA, Schwartz ES (2001). "Valuing American options by simulation: a simple least-squares approach." *The Review of Financial Studies*, **14**(1), 113–147.

Marshall C, Hardy M, Saunders D (2010). "Valuation of a guaranteed minimum income benefit." *North American Actuarial Journal*, **14**(1), 38–58.

Moenig T, Bauer D (2016). "Revisiting the risk-neutral approach to optimal policyholder behavior: A study of withdrawal guarantees in variable annuities." *Review of Finance*, **20**(2), 759–794.

Moenig T, Zhu N (2018). "Lapse-and-Reentry in Variable Annuities." *Journal of Risk and Insurance*, **85**(4), 911–938.

Molent A (2020). "Taxation of a GMWB Variable Annuity in a Stochastic Interest Rate Model." *ASTIN Bulletin: The Journal of the IAA*, **50**(3), 1001–1035.

NAIC (2016). "Actuarial Guideline XLIII." *Technical report*, National Association of Insurance Commissioners. URL https://www.naic.org/documents/cmte_e_va_issues_wg_related_redlined_ag43_160926.pdf.

NAIC (2020). "NAIC Valuation Manual." *Technical report*, National Association of Insurance Commissioners. URL https://content.naic.org/sites/default/files/pbr_data_valuation_manual_current_edition.pdf.

O'Connell MJ (1974). *Theory and applications of concomitants of order statistics*. Digital Repository, Iowa State University. URL http://lib.dr.iastate.edu/.

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.

Owen AB (2013). *Monte Carlo theory, methods and examples*.

Peng J, Leung KS, Kwok YK (2012). "Pricing guaranteed minimum withdrawal benefits under stochastic interest rates." *Quantitative Finance*, **12**(6), 933–941.

Piscopo G, Haberman S (2011). "The valuation of guaranteed lifelong withdrawal benefit options in variable annuity contracts and the impact of mortality risk." *North American Actuarial Journal*, **15**(1), 59–76.

Risk J, Ludkovski M (2018). "Sequential design and spatial modeling for portfolio tail risk measurement." *SIAM Journal on Financial Mathematics*, **9**(4), 1137–1174.

Shevchenko PV, Luo X (2017). "Valuation of variable annuities with guaranteed minimum withdrawal benefit under stochastic interest rate." *Insurance: Mathematics and Economics*, **76**, 104–117.

Steinorth P, Mitchell OS (2015). "Valuing variable annuities with guaranteed minimum lifetime withdrawal benefits." *Insurance: Mathematics and Economics*, **64**, 246–258.

Sun P, Kelkar R, Dai J, Huang V (2016). "How effective is variable annuity guarantee hedging." *Technical report*, Milliman Research Report. URL https://au.milliman.com/en/insight/how-effective-is-variable-annuity-guarantee-hedging.

Tiesset M, Troussard P (2005). "Regulatory capital and economic capital." *Banque de France Financial Stability Review*, **7**, 59–74.

Trottier DA, Godin F, Hamel E (2018). "Local hedging of variable annuities in the presence of basis risk." *ASTIN Bulletin: The Journal of the IAA*, **48**(2), 611–646.

Veach E, Guibas LJ (1995). "Optimally combining sampling techniques for Monte Carlo rendering." In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pp. 419–428. ACM.

Wirch JL, Hardy MR (1999). "A synthesis of risk measures for capital adequacy." *Insurance: mathematics and Economics*, **25**(3), 337–347.

Appendix A

Calculation Illustration: GMWB with Dynamic Hedging Using Various Nested Simulation Procedure

In this GMWB example, we make the following assumptions:

- $S_0 = F_0 = G_0 = 1,000$
- $T = 3$ years
- $I_t = 30\%G_t$ for $t = 1, 2, 3$
- The guarantee value G_t ratchets up annually such that $G_t = \max(G_{t-1}, F_t)$ for $t = 1, 2, 3$
- No management fee deduction or other expenses, i.e. $\eta^g = \eta^n = 0$.
- No decrements due to lapse or mortality
- Delta hedging program is rebalanced every year

Other than the above assumptions, we assume the GMWB contract is as defined in Section 5.4.

We use a standard nested simulation with $M = 2$ outer scenarios and $N = 3$ inner simulations to illustrate the calculation of estimated loss $\widehat{L}$ of the delta-hedged GMWB

contract. We assume a risk free rate of $r = 2\%$, and a Black-Scholes model with volatility $\sigma = 30\%$ for the inner simulation stock price. Hence, the inner simulation (risk-neutral) conditional distribution $g(S_{t,t+1}|S_t)$ is a lognormal distribution with mean $\ln(S_t)+r-\frac{\sigma^2}{2} = \ln S_t - 0.025$ and standard deviation $\sigma = 30\%$.

Again, given the highly simplified assumptions and very rough simulation model, the following calculation illustration should only be used to help understand the mechanism of various likelihood ratio estimators presented in this section. The calculation results should not be interpreted as characteristics of a GMWB contract or of dynamic hedging program.

Table A.1 shows the simulated outer scenarios $\boldsymbol{X}^{(i)} = \left(S^{(i)}, F^{(i)}, G^{(i)}\right)$. The stochasticity of the outer scenarios comes from the stock price S_t. Other components can be calculated deterministically based on our model and the stock price. See Equation (5.8)-(5.10).

For example,

$$
\begin{aligned}
F_2^{(1)} &= \left(F_1^{(1)} - I_1^{(1)}\right)^+ \times \frac{S_2^{(1)}}{S_1^{(1)}} = (512 - 300)^+ \times \frac{786}{512} = 325 \\
G_2^{(1)} &= \max\left(G_1^{(1)}, F_2^{(1)}\right) = \max(1,000, 325) = 1,000 \\
I_2^{(1)} &= \gamma G_2^{(1)} = 30\% \times 1000 = 300
\end{aligned}
$$

Table A.2 shows the inner simulation sample paths: $S_{t,t'}^{(i,j)}$, $F_{t,t'}^{(i,j)}$, $G_{t,t'}^{(i,j)}$, and $I_{t,t'}^{(i,j)}$. The inner simulation stock price sample path $S_{t,t'}^{(i,j)}$ is simulated according to Step (1.A) in the Section 5.3 algorithm. Each inner simulation starts at $t' = t$, so $S_{t,t'}^{(i,j)} = S_t^{(i)}$. The sub-account value $F_{t,t'}^{(i,j)}$, guarantee value $G_{t,t'}^{(i,j)}$, and withdrawal benefit amount $I_{t,t'}^{(i,j)}$, in the inner simulation sample paths are calculated using Equation (5.8)-(5.10). For example,

$$
\begin{aligned}
F_{1,2}^{(2,3)} &= \left(F_{1,1}^{(2,3)} - I_{1,1}^{(2,3)}\right)^+ \times \frac{S_{1,2}^{(2,3)}}{S_{1,1}^{(2,3)}} = (1,024 - 307)^+ \times \frac{1,945}{1,024} = 1,362 \\
G_{1,2}^{(2,3)} &= \max\left(F_{1,2}^{(2,3)}, G_{1,1}^{(2,3)}\right) = \max(1,362, 1,024) = 1,362 \\
I_{1,2}^{(2,3)} &= \gamma G_{1,2}^{(2,3)} = 30\% \times 1,362 = 408
\end{aligned}
$$

Table A.3 show the value of partial derivatives calculated based on inner simulation

$S_t^{(i)}$	$t=0$	$t=1$	$t=2$	$t=3$
$i=1$	1,000	512	786	928
$i=2$	1,000	1,024	1,237	1,176

(a) Stock price $S_t^{(i)}$

$F_t^{(i)}$	$t=0$	$t=1$	$t=2$	$t=3$
$i=1$	1,000	512	325	30
$i=2$	1,000	1,024	866	531

(b) Fund value $F_t^{(i)}$

$G_t^{(i)}$	$t=0$	$t=1$	$t=2$	$t=3$
$i=1$	1,000	1,000	1,000	1,000
$i=2$	1,000	1,024	1,024	1,024

(c) Guarantee value $G_t^{(i)}$

$I_t^{(i)}$	$t=0$	$t=1$	$t=2$	$t=3$
$i=1$	0	300	300	300
$i=2$	0	307	307	307

(d) Withdrawal benefit $I_t^{(i)}$

Table A.1: Outer Scenario

sample path values using Equation (5.12)-(5.14). For example,

$$\frac{dF_{1,2}^{(2,3)}}{dS_1^{(2)}} = \mathbb{1}_{\left\{I_{1,1}^{(2,3)} < F_{1,1}^{(2,3)}\right\}} \left(\frac{dF_{1,1}^{(2,3)}}{dS_1^{(2)}} - \frac{dI_{1,1}^{(2,3)}}{dS_1^{(2)}}\right) \times \frac{S_{1,2}^{(2,3)}}{S_{1,1}^{(2,3)}} = \mathbb{1}_{\{307<1024\}} (1-0) \times \frac{1,945}{1,024} = 1.899$$

$$\begin{aligned}\frac{dG_{1,2}^{(2,3)}}{dS_1^{(2)}} &= \mathbb{1}_{\left\{G_{1,1}^{(2,3)} < F_{1,2}^{(2,3)}\right\}} \frac{dF_{1,2}^{(2,3)}}{dS_1^{(2)}} + \mathbb{1}_{\left\{G_{1,1}^{(2,3)} \geq F_{1,2}^{(2,3)}\right\}} \frac{dG_{1,1}^{(2,3)}}{dS_1^{(2)}} \\ &= \mathbb{1}_{\{1,024<1,362\}} \times 1.899 + \mathbb{1}_{\{1,024\geq 1,362\}} \times 0 = 1.899\end{aligned}$$

$$\frac{dI_{1,2}^{(2,3)}}{dS_1^{(2)}} = \gamma \frac{dG_{1,2}^{(2,3)}}{dS_1^{(2)}} = 30\% \times 1.899 = 0.570$$

These partial derivatives are then used to calculate the estimated delta's.

Table A.4 shows the unadjusted sample delta $\widehat{\Delta}_t^{(k,j)}$, as well as delta $\widehat{\Delta}_t^{(i)\text{SMC}}$ and loss $\widehat{L}_2^{\text{SMC}}$ estimated using standard Monte Carlo simulation. They are calculated using Equation (5.11), (5.15) and (5.16), respectively. For example,

$$\begin{aligned}
\widehat{\Delta}_1^{(2,3)} &= \sum_{s=2}^{3} e^{-2\%(s-1)} \times \mathbb{1}_{\left\{I_{1,s}^{(2,3)} > F_{1,s}^{(2,3)}\right\}} \left(\frac{dI_{1,s}^{(2,3)}}{dS_1^{(2)}} - \frac{dF_{1,s}^{(2,3)}}{dS_1^{(2)}} \right) \\
&= \left(e^{-2\%\times(2-1)} \times \mathbb{1}_{\{408>1{,}362\}} \left(0.570 - 1.899\right) + e^{-2\%\times(3-1)} \times \mathbb{1}_{\{408>1{,}018\}} \left(0.570 - 1.420\right) \right) = 0.000 \\
\widehat{\Delta}_1^{(2)\text{SMC}} &= \frac{1}{N} \sum_{j=1}^{N} \widehat{\Delta}_1^{(2,j)} = \frac{1}{3} \times (-0.713 - 0 - 0) = -0.238 \\
\widehat{L}_2^{\text{SMC}} &= \sum_{t=0}^{T-1} \widehat{\Delta}_t^{(2)\text{SMC}} \left(e^{-rt} S_t^{(2)} - e^{-r(t+1)} S_{t+1}^{(2)} \right) + \sum_{t=1}^{T} e^{-rt} \left(I_t^{(2)} - F_t^{(2)} \right)^+ \\
&= 0.000 \times (1{,}000 - e^{-2\%} \times 1{,}024) - 0.238 \times (e^{-2\%} \times 1{,}024 - e^{-2\times 2\%} \times 1{,}237) \\
&\quad + 0.000 \times (e^{-2\times 2\%} \times 1{,}237 - e^{-3\times 2\%} \times 1{,}176) + e^{-2\%}(307 - 1{,}024)^+ \\
&\quad + e^{-2\times 2\%}(307 - 1{,}237)^+ + e^{-3\times 2\%}(307 - 1{,}176)^+ = 44
\end{aligned}$$

Next we demonstrate how the MLR estimator is applied in this example.

If we are interested in estimating, say $\widetilde{\Delta}_1^{(2)}$, using the MLR estimator, we will reweight each inner simulation sample paths at $t = 1$ for target outer scenario $i = 2$. As an example, the inner sample paths for target scenario $i = 2$ at time $t = 1$ without any adjustment are shown in Table A.5. It is obvious the unadjusted inner sample paths $\left(S_{1,t'}^{(1,j)}, F_{1,t'}^{(1,j)}, G_{1,t'}^{(1,j)}\right)$ in Table A.5 are inconsistent with the dynamic of sub-account value and guarantee value described in Equation (5.8)-(5.10). For example, the guarantee value decreases from $t' = 1$ to $t' = 2$. Hence we need to apply the adjustments discussed earlier in Section 5.4.2 to these inner sample paths before we can use them in the MLR estimator.

Table A.6 shows the adjusted stock price $\widetilde{S}_{t,t+1}^{(k,j)}$, the adjusted sub-account value $\widetilde{F}_{t,t+1}^{(k,j)}$, the adjusted guarantee base $\widetilde{G}_{t,t+1}^{(k,j)}$ and adjusted sample delta $\widetilde{\Delta}_t^{(k,j)}$ used in the MLR estimator. The adjusted inner sample paths $\left(\widetilde{S}_{t,t+1}^{(k,j)}, \widetilde{F}_{t,t+1}^{(k,j)}, \widetilde{G}_{t,t+1}^{(k,j)}\right)$ are calculated using Equation (5.17). The adjusted sample deltas are calculated using Equation (5.19). For

example,

$$\widetilde{S}_{1,2}^{(1,2)} = S_{1,2}^{(1,2)} \times \frac{G_1^{(2)}}{G_1^{(1)}} = 1,152 \times \frac{1,024}{1,000} = 1,180$$

$$\widetilde{F}_{1,2}^{(1,2)} = F_{1,2}^{(1,2)} \times \frac{G_1^{(2)}}{G_1^{(1)}} = 477 \times \frac{1,024}{1,000} = 488$$

$$\widetilde{G}_{1,2}^{(1,2)} = G_{1,2}^{(1,2)} \times \frac{G_1^{(2)}}{G_1^{(1)}} = 1,000 \times \frac{1,024}{1,000} = 1,024$$

$$\widetilde{\Delta}_1^{(1,2)} = \widehat{\Delta}_1^{(1,2)} \times \frac{G_1^{(2)}}{G_1^{(1)}} \times \frac{F_1^{(2)}}{F_1^{(1)}} \times \frac{S_1^{(1)}}{S_1^{(2)}} = -2.533 \times \frac{1,024}{1,000} \times \frac{1,024}{512} \times \frac{512}{1,024} = -2.594$$

From Table A.6, we can verify that these adjusted inner sample paths are consistent with the dynamic of sub-account value and guarantee value in Equation (5.8)-(5.10). In addition, we can also verify that the return on stock price remain the same as the unadjusted inner sample paths after time $t+1$ for inner sample paths starting at time t. In other words, the Markov property of the inner sample path is preserved.

The probability density function $g_t^{(i)}(\widetilde{S}_{t,t+1}^{(k,j)})$ in Table A.7 are calculated based on the the Black-Scholes inner simulation asset model, i.e. the lognormal model with a standard deviation of $\sigma = 30\%$ and mean of $\ln(S_t^{(i)}) + r - \frac{\sigma^2}{2} = \ln(S_t^{(i)}) - 0.025$. Subsequently, the probability density function $\bar{g}_M(\widetilde{S}_{t,t+1}^{(k,j)})$ is calculated based on Proposition 5.2.3. For example, for target scenario $i = 2$,

$$\begin{aligned} g_1^{(2)}\left(\widetilde{S}_{1,2}^{(1,2)}\right) &= \phi\left(\ln\left(\widetilde{S}_{1,2}^{(1,2)}\right); \ln\left(S_1^{(2)}\right) - 0.025, 30\%\right) \\ &= \phi\left(\ln(1,180); \ln(1,024) - 0.025, 30\%\right) = 1.1400 \end{aligned}$$

where $\phi(x; \mu, \sigma)$ is the probability density function of a normal distribution with mean of μ and standard deviation of σ. And

$$g_m\left(\widetilde{S}_{1,2}^{(1,2)}\right) = \frac{1}{2}\left(g_1^{(1)}\left(\widetilde{S}_{1,2}^{(1,2)}\right) + g_1^{(2)}\left(\widetilde{S}_{1,2}^{(1,2)}\right)\right) = \frac{1}{2}\left(0.0274 + 1.1400\right) = 0.5837$$

Table A.8 shows the MLR adjusted sample delta $\widetilde{\Delta}_t^{(k,j)} \frac{g_t^{(1)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}$, delta estimated using MLR method $\widehat{\Delta}_t^{(i)\text{MLR}}$, and the loss estimated using MLR method $\widehat{L}_i^{\text{MLR}}$. These values are

calculated based on Proposition 5.4.1. For example, for target scenario $i = 2$,

$$\widetilde{\Delta}_1^{(1,3)} \frac{g_1^{(2)}\left(\widetilde{S}_{1,2}^{(1,3)}\right)}{\bar{g}_M\left(\widetilde{S}_{1,2}^{(1,3)}\right)} = -1.720 \times \frac{1.2649}{0.7387} = -2.9448$$

$$\begin{aligned}
\widehat{\Delta}_1^{(2)\mathrm{MLR}} =& \frac{1}{MN}\sum_{k=1}^{M}\sum_{j=1}^{N}\widetilde{\Delta}_t^{(k,j)}\frac{g_t^{(2)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)} \\
=& \frac{1}{2\times 3}(-2.9577 - 5.0665 - 2.9448 - 0.3204 + 0.0000 + 0.0000) = -1.882
\end{aligned}$$

$$\begin{aligned}
\widehat{L}_2^{\mathrm{MLR}} =& \sum_{t=0}^{T-1}\widehat{\Delta}_t^{(2)\mathrm{MLR}}\left(e^{-rt}S_t^{(2)} - e^{-r(t+1)}S_{t+1}^{(2)}\right) + \sum_{t=1}^{T}e^{-rt}\left(I_t^{(2)} - F_t^{(2)}\right)^+ \\
=& 0.000 \times (1,000 - e^{-2\%} \times 1,024) - 1.882 \times (e^{-2\%} \times 1,024 - e^{-2\times 2\%} \times 1,237) \\
& - 0.226 \times (e^{-2\times 2\%} \times 1,237 - e^{-3\times 2\%} \times 1,176) + e^{-2\%}(307 - 1,024)^+ \\
& + e^{-2\times 2\%}(307 - 1,237)^+ + e^{-3\times 2\%}(307 - 1,176)^+ = 329
\end{aligned}$$

	$t=0$			$t=1$			$t=2$		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\boldsymbol{S}^{(i,j)}_{t,t'}$, $i=1$									
$t'=0$	1,000	1,000	1,000						
$t'=1$	993	985	1,040	512	512	512			
$t'=2$	1,019	1,132	1,054	924	1,152	887	786	786	786
$t'=3$	978	1,312	1,273	865	1,350	895	651	800	919
$i=2$									
$t'=0$	1,000	1,000	1,000						
$t'=1$	969	1,218	1,150	1,024	1,024	1,024			
$t'=2$	973	1,337	1,421	605	1,845	1,945	1,237	1,237	1,237
$t'=3$	837	1,258	1,494	760	1,950	2,077	1,036	1,324	1,171
$\boldsymbol{F}^{(i,j)}_{t,t'}$, $i=1$									
$t'=0$	1,000	1,000	1,000						
$t'=1$	993	985	1,040	512	512	512			
$t'=2$	711	787	738	383	477	367	325	325	325
$t'=3$	395	565	514	77	207	68	21	26	30
$i=2$									
$t'=0$	1,000	1,000	1,000						
$t'=1$	969	1,218	1,150	1,024	1,024	1,024			
$t'=2$	672	936	995	424	1,292	1,362	866	866	866
$t'=3$	320	537	683	146	956	1,018	468	598	529
$\boldsymbol{G}^{(i,j)}_{t,t'}$, $i=1$									
$t'=0$	1,000	1,000	1,000						
$t'=1$	1,000	1,000	1,040	1,000	1,000	1,000			
$t'=2$	1,000	1,000	1,040	1,000	1,000	1,000	1,000	1,000	1,000
$t'=3$	1,000	1,000	1,040	1,000	1,000	1,000	1,000	1,000	1,000
$i=2$									
$t'=0$	1,000	1,000	1,000						
$t'=1$	1,000	1,218	1,150	1,024	1,024	1,024			
$t'=2$	1,000	1,218	1,150	1,024	1,292	1,362	1,024	1,024	1,024
$t'=3$	1,000	1,218	1,150	1,024	1,292	1,362	1,024	1,024	1,024
$\boldsymbol{I}^{(i,j)}_{t,t'}$, $i=1$									
$t'=0$	0	0	0						
$t'=1$	300	300	312	300	300	300			
$t'=2$	300	300	312	300	300	300	300	300	300
$t'=3$	300	300	312	300	300	300	300	300	300
$i=2$									
$t'=0$	0	0	0	138					
$t'=1$	300	365	345	307	307	307			
$t'=2$	300	365	345	307	387	408	307	307	307
$t'=3$	300	365	345	307	387	408	307	307	307

Table A.2: Inner simulation sample paths

		$t=0$			$t=1$			$t=2$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\frac{\boldsymbol{dF}_{\boldsymbol{t,t'}}^{\boldsymbol{(i,j)}}}{\boldsymbol{dS}_{\boldsymbol{t}}^{\boldsymbol{(i)}}}$, $i=1$										
	$t'=0$	1.000	1.000	1.000						
	$t'=1$	0.993	0.985	1.040	1.000	1.000	1.000			
	$t'=2$	1.019	1.132	0.738	1.805	2.250	1.732	0.414	0.414	0.414
	$t'=3$	0.978	1.312	0.514	1.689	2.637	1.748	0.343	0.421	0.484
	$i=2$									
	$t'=0$	1.000	1.000	1.000						
	$t'=1$	0.969	1.218	1.150	1.000	1.000	1.000			
	$t'=2$	0.973	0.936	0.995	0.591	1.802	1.899	0.700	0.700	0.700
	$t'=3$	0.837	0.537	0.683	0.742	1.333	1.420	0.586	0.749	0.663
$\frac{\boldsymbol{dG}_{\boldsymbol{t,t'}}^{\boldsymbol{(i,j)}}}{\boldsymbol{dS}_{\boldsymbol{t}}^{\boldsymbol{(i)}}}$, $i=1$										
	$t'=0$	0.000	0.000	0.000						
	$t'=1$	0.000	0.000	1.040	0.000	0.000	0.000			
	$t'=2$	0.000	0.000	1.040	0.000	0.000	0.000	0.000	0.000	0.000
	$t'=3$	0.000	0.000	1.040	0.000	0.000	0.000	0.000	0.000	0.000
	$i=2$									
	$t'=0$	0.000	0.000	0.000						
	$t'=1$	0.000	1.218	1.150	0.000	0.000	0.000			
	$t'=2$	0.000	1.218	1.150	0.000	1.802	1.899	0.000	0.000	0.000
	$t'=3$	0.000	1.218	1.150	0.000	1.802	1.899	0.000	0.000	0.000
$\frac{\boldsymbol{dI}_{\boldsymbol{t,t'}}^{\boldsymbol{(i,j)}}}{\boldsymbol{dS}_{\boldsymbol{t}}^{\boldsymbol{(i)}}}$, $i=1$										
	$t'=0$	0.000	0.000	0.000						
	$t'=1$	0.000	0.000	0.312	0.000	0.000	0.000			
	$t'=2$	0.000	0.000	0.312	0.000	0.000	0.000	0.000	0.000	0.000
	$t'=3$	0.000	0.000	0.312	0.000	0.000	0.000	0.000	0.000	0.000
	$i=2$									
	$t'=0$	0.000	0.000	0.000						
	$t'=1$	0.000	0.365	0.345	0.000	0.000	0.000			
	$t'=2$	0.000	0.365	0.345	0.000	0.541	0.570	0.000	0.000	0.000
	$t'=3$	0.000	0.365	0.345	0.000	0.541	0.570	0.000	0.000	0.000

Table A.3: Partial derivatives in inner simulation sample paths

	$t=0$			$t=1$			$t=2$		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\boldsymbol{i=1}$									
$\widehat{\Delta}_t^{(1,j)}$	0.000	0.000	0.000	-1.623	-2.533	-1.680	-0.336	-0.413	-0.475
	$\widehat{\Delta}_0^{(1)\mathrm{SMC}} = 0.000$			$\widehat{\Delta}_1^{(1)\mathrm{SMC}} = -1.945$			$\widehat{\Delta}_2^{(1)\mathrm{SMC}} = -0.408$		
	$\widehat{L}_1^{\mathrm{SMC}} = 795$								
$\boldsymbol{i=2}$									
$\widehat{\Delta}_t^{(2,j)}$	0.000	0.000	0.000	-0.713	0.000	0.000	0.000	0.000	0.000
	$\widehat{\Delta}_0^{(2)\mathrm{SMC}} = 0.000$			$\widehat{\Delta}_1^{(2)\mathrm{SMC}} = -0.238$			$\widehat{\Delta}_2^{(2)\mathrm{SMC}} = 0.000$		
	$\widehat{L}_2^{\mathrm{SMC}} = 44$								

Table A.4: Standard nested simulation output: $\widehat{\Delta}_t^{(i,j)}$, $\widehat{\Delta}_t^{(i)\mathrm{SMC}}$ and $\widehat{L}_i^{\mathrm{SMC}}$

	$t=1$		
	$j=1$	$j=2$	$j=3$
$k=1$			
$t'=1$	(1,024, 1,024, 1,024)	(1,024, 1,024, 1,024)	(1,024, 1,024, 1,024)
$t'=2$	(924, 383, 1,000)	(1,152, 477, 1,000)	(887, 367, 1,000)
$t'=3$	(865, 77, 1,000)	(1,350, 207, 1,000)	(895, 68, 1,000)
$k=2$			
$t'=1$	(1,024, 1,024, 1,024)	(1,024, 1,024, 1,024)	(1,024, 1,024, 1,024)
$t'=2$	(605, 424, 1,024)	(1,845, 1,292, 1,292)	(1,945, 1,362, 1,362)
$t'=3$	(760, 146, 1,024)	(1,950, 956, 1,292)	(2,077, 1,018, 1,362)

Table A.5: Unadjusted inner sample paths $\left(S_{1,t'}^{(k,j)}, F_{1,t'}^{(k,j)}, G_{1,t'}^{(k,j)}\right)$ for target scenario $i=2$

	$t=0$			$t=1$			$t=2$		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\widetilde{\boldsymbol{S}}^{(\boldsymbol{k},\boldsymbol{j})}_{\boldsymbol{t},\boldsymbol{t+1}}$									
$i=1,k=1$	993	985	1,040	924	1,152	887	651	800	919
$i=1,k=2$	969	1,218	1,150	591	1,802	1,899	1,012	1,293	1,144
$i=2,k=1$	993	985	1,040	946	1,180	908	667	819	941
$i=2,k=2$	969	1,218	1,150	605	1,845	1,945	1,036	1,324	1,171
$\widetilde{\boldsymbol{F}}^{(\boldsymbol{k},\boldsymbol{j})}_{\boldsymbol{t},\boldsymbol{t+1}}$									
$i=1,k=1$	993	985	1,040	383	477	367	21	26	30
$i=1,k=2$	969	1,218	1,150	414	1,261	1,330	457	584	516
$i=2,k=1$	993	985	1,040	392	488	376	22	27	30
$i=2,k=2$	969	1,218	1,150	424	1,292	1,362	468	598	529
$\widetilde{\boldsymbol{G}}^{(\boldsymbol{k},\boldsymbol{j})}_{\boldsymbol{t},\boldsymbol{t+1}}$									
$i=1,k=1$	1,000	1,000	1,040	1,000	1,000	1,000	1,000	1,000	1,000
$i=1,k=2$	1,000	1,218	1,150	1,000	1,261	1,330	1,000	1,000	1,000
$i=2,k=1$	1,000	1,000	1,040	1,024	1,024	1,024	1,024	1,024	1,024
$i=2,k=2$	1,000	1,218	1,150	1,024	1,292	1,362	1,024	1,024	1,024
$\widetilde{\boldsymbol{\Delta}}^{(\boldsymbol{k},\boldsymbol{j})}_{\boldsymbol{t}}$									
$i=1,k=1$	0.000	0.000	0.000	-1.623	-2.533	-1.680	-0.336	-0.413	-0.475
$i=1,k=2$	0.000	0.000	0.000	-0.696	0.000	0.000	0.000	0.000	0.000
$i=2,k=1$	0.000	0.000	0.000	-1.662	-2.594	-1.720	-0.582	-0.715	-0.821
$i=2,k=2$	0.000	0.000	0.000	-0.713	0.000	0.000	0.000	0.000	0.000

Table A.6: Mixture Likelihood Ratio methods: adjusted inner sample paths $\left(\widetilde{S}^{(k,j)}_{t,t+1}, \widetilde{F}^{(k,j)}_{t,t+1}, \widetilde{G}^{(k,j)}_{t,t+1}\right)$ and adjusted sample delta $\widetilde{\Delta}^{(k,j)}_{t}$

	$t=0$			$t=1$			$t=2$		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\boldsymbol{g_t^{(i)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}$									
$i=1$									
$k=1$	1.3274	1.3291	1.2997	0.1622	0.0274	0.2125	1.1464	1.3164	1.1078
$k=2$	1.3295	1.0108	1.1436	1.1364	0.0001	0.0001	0.8671	0.2914	0.5468
$i=2$									
$k=1$	1.3274	1.3291	1.2997	1.3084	1.1400	1.2649	0.1882	0.5784	0.9438
$k=2$	1.3295	1.0108	1.1436	0.3293	0.1640	0.1127	1.1690	1.2675	1.3232
$\boldsymbol{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}$									
$k=1$	1.3274	1.3291	1.2997	0.7353	0.5837	0.7387	0.6673	0.9474	1.0258
$k=2$	1.3295	1.0108	1.1436	0.7329	0.0821	0.0564	1.0180	0.7794	0.9350

Table A.7: Mixture Likelihood Ratio method: probability density function $g_t^{(i)}\left(\widetilde{S}_{t,t}^{(k,j)}\right)$ and $\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)$

	$t=0$			$t=1$			$t=2$		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\boldsymbol{i=1}$									
$\widetilde{\Delta}_t^{(k,j)}\frac{g_t^{(1)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}$									
$k=1$	0.0000	0.0000	0.0000	-0.3581	-0.1189	-0.4832	-0.5775	-0.5740	-0.5125
$k=2$	0.0000	0.0000	0.0000	-1.0798	0.0000	0.0000	0.0000	0.0000	0.0000
	$\widehat{\Delta}_0^{(1)\mathrm{MLR}}=0.000$			$\widehat{\Delta}_1^{(1)\mathrm{MLR}}=-0.340$			$\widehat{\Delta}_2^{(1)\mathrm{MLR}}=-0.277$		
	$\widehat{L}_1^{\mathrm{MLR}}=373$								
$\boldsymbol{i=2}$									
$\widetilde{\Delta}_t^{(k,j)}\frac{g_t^{(2)}\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}{\bar{g}_M\left(\widetilde{S}_{t,t+1}^{(k,j)}\right)}$									
$k=1$	0.0000	0.0000	0.0000	-2.9577	-5.0665	-2.9448	-0.1642	-0.4366	-0.7558
$k=2$	0.0000	0.0000	0.0000	-0.3204	0.0000	0.0000	0.0000	0.0000	0.0000
	$\widehat{\Delta}_0^{(2)\mathrm{MLR}}=0.000$			$\widehat{\Delta}_1^{(2)\mathrm{MLR}}=-1.882$			$\widehat{\Delta}_2^{(2)\mathrm{MLR}}=-0.226$		
	$\widehat{L}_2^{\mathrm{MLR}}=329$								

Table A.8: MLR nested simulation output