Research Paper Series

Multilevel Monte Carlo for Solvency II SCR: Efficient Nested Simulation

Gianmarco Mori
Francesco Bertonati
Daniel Bruttomesso
Amina Cavazzana

May 2026

iason
ESSENTIAL SERVICES FOR
FINANCIAL INSTITUTIONS

iason

Research Paper Series

Year 2026 - Issue Number 86

Last published issues are available online:
http://www.iasonltd.com/research

Front Cover: **Umberto Boccioni**, *La Strada entra Nella Casa*, 1911.

Executive Summary

The paper proposes the use of Multilevel Monte Carlo as an efficiency-enhancing method for the calculation of the Solvency Capital Requirement within the framework of internal insurance models. The computation of the Solvency Capital Requirement relies on nested Monte Carlo simulations, which become computationally prohibitive for complex portfolios. Commonly adopted proxy methods reduce execution times but introduce model risk and require a complex governance framework. The Multilevel Monte Carlo approach acts as an accelerator that does not rely on proxy functions. Instead, it employs a hierarchy of simulation levels, ranging from coarse to increasingly refined representations, in order to reduce variance and computational cost while preserving the nested simulation structure. The method is risk preserving because it avoids structural bias, ensures near optimal convergence rates, and facilitates regulatory validation through its statistical transparency. The paper also introduces an adaptive version of the Multilevel Monte Carlo method. This extension concentrates computational effort on the most critical scenarios in the tail of the loss distribution, thereby further improving efficiency in the calculation of the capital requirement.

About the Authors

Gianmarco Mori:
Expert
He holds a Bachelor in Banking and Finance and a Master degree in Statistics and Actuarial science. He obtained a specialization in Quantitative Finance at Politecnico di Milano and he had worked in a consulting firm for two year focused in Credit and Liquidity Risk Management. After these years, he moved to banking sector in Risk Management area and currently he has been working in iason as financial engineer for some of the major italian players.

Francesco Bertonati:
Financial Engineer
Graduated in Quantitative Finance and Insurance and holding a Master's degree in Statistics and Applied Mathematics, he worked as a Financial Engineer with the Risk Management team of one of the leading Italian insurance companies, contributing to the development of the internal model. His work focused in particular on modeling dependencies between different risk factors, preparing the related technical documentation, and assessing underwriting risks in the Life insurance sector. He is currently collaborating with the Financial Risk team of one of the major Italian banks, where he focuses on risk pricing analysis and the development of models for financial products.

Daniel Bruttomesso:
Financial Engineer
He holds a degree in Mathematical Engineering. He collaborated as a Financial Engineer with the Risk Management team of one of the leading Italian insurance companies, contributing to the development of the internal model. His work focused in particular on modeling dependencies between risk factors, preparing the related technical documentation, and assessing underwriting risks related to Life insurance. He is currently collaborating with the Financial Risk team of one of the major Italian banks, focusing on risk pricing analysis and development for financial products.

Amina Cavazzana:
Credit Risk Quant
She holds a degree in Statistics for Economics and Business and is currently working as a Credit Risk Analyst, collaborating with the Risk Management team of one of the leading Italian insurance companies, supporting the development and implementation of the internal model. Her main activities include counterparty risk modeling, analysis of dependencies between risk factors, and the preparation of the official model documentation. She is also engaged in managing Underwriting/Life risks.

Table of Content

This paper addresses the problem of calculating the Solvency Capital Requirement (SCR) under Solvency II, which requires insurance companies to hold sufficient capital to withstand extreme losses. The two-layer simulation framework underlying the capital requirement calculation consists of "real-world" simulations of risk factors and, for each of them, corresponding "risk-neutral" valuations of the portfolio. Nested simulation structures of this type are not unique to solvency applications and arise in several other areas of quantitative finance, particularly in the computation of counterparty credit risk (CCR) exposure metrics such as Expected Positive Exposure (EPE), as well as in the valuation and sensitivity analysis of XVA adjustments. While this structure ensures theoretical consistency, it entails a substantial computational burden. The paper discusses the widespread industry adoption of the Least Squares Monte Carlo (LSMC) approach, which relies on proxy functions. However, such approximations may introduce model risk, particularly in the tails of the distribution that are crucial for SCR estimation, and require specific calibration procedures and underlying structural assumptions. The core contribution of the paper is the application of the Multilevel Monte Carlo (MLMC) method, which builds a hierarchy of accuracy levels by combining many low-cost, coarse simulations with a few high-cost, precise ones, and proposing an adaptive version that concentrates computational effort on scenarios close to the critical loss threshold. The methodology is subsequently applied to a index-linked life insurance product with a minimum capital guarantee, combining financial and biometric risks within a nonlinear payoff structure. The paper concludes that MLMC preserves the transparency and accuracy of the standard Nested Monte Carlo approach while significantly reducing computational time, thereby avoiding the need for complex proxy functions. Finally, it points to the potential extension toward Multi-index Monte Carlo (MIMC) to simultaneously address multiple sources of numerical error.

1. SCR Computation under Solvency II

Solvency II introduced a risk-based capital regime for insurers, defining the Solvency Capital Requirement (SCR) as the capital needed to withstand a 1-in-200 year adverse event [6]. The SCR corresponds to the 99.5% one-year Value-at-Risk (VaR) of the loss in Basic Own Funds (BOF). In regulatory terms, the SCR represents the capital needed to ensure that an insurer can absorb extreme but plausible losses over a one-year horizon. Insurers may compute the SCR either via the prescribed Standard Formula or through a company-specific Internal Model, subject to supervisory approval. Within an approved Internal Model, this requirement is computed by simulating the full distribution of one-year changes in BOF and extracting its extreme quantile under market-consistent valuation principles.

Operationally, this entails a two-layer structure. First, real-world ($\mathbb{P}$) simulations generate one-year scenarios for all material risk drivers. Second, for each scenario, the balance sheet must be revalued at the one-year horizon. While assets are often directly observable or analytically tractable, liabilities, especially those embedding financial guarantees, profit-sharing mechanisms, policyholder behaviour and management actions, rarely admit closed-form expressions. Their value at the one-year horizon is typically expressed as a conditional expectation under the risk-neutral ($\mathbb{Q}$) measure. This naturally leads to a Nested Monte Carlo (NMC) architecture, with an outer $\mathbb{P}$-simulation and an inner $\mathbb{Q}$-valuation performed for each real-world scenario.

The resulting computational burden scales proportionally to $N_{\text{out}} \times N_{\text{in}}$. Because the SCR targets a deep tail quantile, both layers must be sufficiently large to control statistical error and avoid distortions in the ordering of extreme scenarios. For realistic life portfolios with

path-dependent features, brute-force nested simulation rapidly becomes impractical in production environments subject to tight reporting deadlines.

To address this issue, the industry has widely adopted proxy techniques, most notably Least Squares Monte Carlo (LSMC) following Longstaff and Schwartz [12], as well as replicating portfolio approaches. These methods replace repeated inner valuations with an approximating function calibrated on a limited training set of fully nested runs. While computationally efficient, proxy approaches introduce structural approximation risk, particularly in the tail region driving the SCR. As documented in the nested simulation literature [10, 3], tail estimation is highly sensitive to model misspecification and extrapolation. Consequently, the computational problem is partly transformed into a governance challenge, requiring extensive validation, stability analysis and documentation of approximation quality.

This paper proposes an alternative that preserves the fully simulation-based structure while materially reducing computational cost: Multilevel Monte Carlo (MLMC), pioneered by Mike Giles [7] and extended to nested risk estimation by Giles and Haji-Ali [8]. Rather than replacing the inner valuation with a parametric surrogate, MLMC applies a hierarchy of simulation accuracies to the $\mathbb{Q}$-valuation layer. Using a telescopic decomposition, most samples are allocated to inexpensive coarse approximations, while only a limited number are used to correct bias at finer levels. Under suitable conditions, this approach achieves near-optimal Monte Carlo complexity for nested problems, including tail probabilities and quantiles.

In contrast to proxy methods, MLMC does not rely on an assumed functional form for the liability value and remains structurally close to full nested simulation. This feature is particularly attractive in an Internal Model context, where transparency, statistical error control and methodological consistency with regulatory expectations are essential. Practitioner evidence in solvency applications indicates that MLMC can deliver substantial runtime reductions while maintaining valuation fidelity [2].

2. SCR and the Nested Monte Carlo Challenge

2.1 The Structure of SCR in an Internal Model

Let $t = 0$ denote the valuation date and $t = 1$ the one-year horizon. Define Basic Own Funds:

$$BOF_t := A_t - L_t, \tag{1}$$

where A_t is the market-consistent value of assets and L_t the market-consistent value of liabilities at time t. Define the one-year loss random variable (sign convention consistent with capital requirement):

$$\Delta BOF := BOF_0 - BOF_1. \tag{2}$$

The Internal Model SCR is:

$$SCR := VaR_{0.995}(\Delta BOF). \tag{3}$$

The definition in (3) embeds several layers of modelling complexity. First, BOF_1 is itself a random variable driven by the joint evolution of all material risk factors over the one-year horizon. These typically include market variables (interest rates, equity indices, credit spreads, implied volatilities), biometric factors (mortality, longevity, lapse), and potentially management actions and dynamic policyholder behaviour. The SCR therefore depends on the full joint distribution of these drivers under the real-world measure $\mathbb{P}$, including their nonlinear interactions. A standard simulation architecture is:

- **Outer simulation ($\mathbb{P}$).** Generate N_{out} independent scenarios of the one-year risk drivers, producing time-one states $X_1^{(i)}$. The state vector X_1 typically contains the entire term structure of interest rates, equity levels, spread curves, volatility surfaces, and relevant insurance risk factors. The outer simulation must be sufficiently rich to capture tail co-movements, as the 99.5% quantile is sensitive to joint stress configurations rather than marginal extremes alone.

- **Inner valuation ($\mathbb{Q}$).** For each $X_1^{(i)}$, compute $L_1^{(i)}$ as a market-consistent value of future cashflows conditional on the time-1 filtration:

$$L_1^{(i)} = \mathbb{E}^{\mathbb{Q}}\left[\sum_{k:t_k>1} D(1,t_k)\,\mathrm{CF}_{t_k} \,\middle|\, X_1^{(i)}\right],$$

  where $D(1,t_k)$ is the stochastic discount factor from 1 to t_k under $\mathbb{Q}$, and CF_{t_k} are liability cashflows.
  The liability value at time 1 can be expressed as a conditional expectation under the risk-neutral measure. In general, no closed-form expression is available when cashflows depend pathwise on future economic scenarios. As a consequence, the conditional expectation must be approximated numerically, most commonly by Monte Carlo simulation under $\mathbb{Q}$. This creates a nested structure: for each outer scenario $X_1^{(i)}$, an inner simulation with N_{in} paths is required to estimate $L_1^{(i)}$.
  It is worth noting that assets may also require revaluation at $t = 1$, especially for path-dependent instruments. However, in many internal model implementations, the dominant computational burden arises from the liability side, where long-dated optionalities and management rules induce high-dimensional path dependence.

- **Quantile estimation.** Compute $\Delta BOF^{(i)}$ via (2) for all outer scenarios, then estimate $VaR_{0.995}$ from the empirical distribution. Since the target is an extreme percentile, the accuracy of the estimator depends critically on the number of outer scenarios and on the noise introduced by the inner estimations of $L_1^{(i)}$. Inner simulation error propagates nonlinearly to the quantile estimator and may induce bias if not properly controlled.

2.2 Why Nested Monte Carlo Arises

Liability valuation at time 1 is a conditional expectation under $\mathbb{Q}$. Given the state $X_1^{(i)}$, the liability value is a functional of the entire future path of risk factors beyond the one-year horizon. Whenever cashflows depend nonlinearly on future asset returns, interest rate paths, policyholder behaviour, or management rules, no tractable closed-form expression is available for this conditional expectation. As a consequence, numerical integration is required, and in practice this is almost always performed via Monte Carlo simulation under the risk-neutral measure. For complex liabilities, one therefore resorts to an inner Monte Carlo estimator (see e.g. Glasserman [9] for a comprehensive treatment of Monte Carlo methods in financial engineering):

$$\widehat{L}_1^{(i)} := \frac{1}{N_{\text{in}}} \sum_{k=1}^{N_{\text{in}}} Y^{(i,k)}, \tag{4}$$

where $Y^{(i,k)}$ is the discounted present value of future cashflows along inner path N_{in}, simulated under $\mathbb{Q}$ from the state $X_1^{(i)}$. The estimator in (4) is unbiased for the conditional

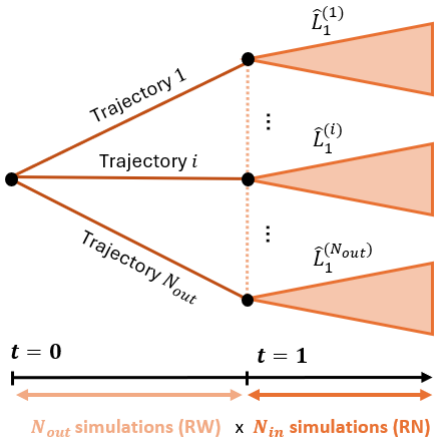

FIGURE 1: Nested Monte Carlo method.

expectation, with variance decreasing at rate $1/N_{\text{in}}$. However, achieving a sufficiently small conditional standard error may require a large N_{in}, especially for long-dated or highly path-dependent contracts.

Since this inner valuation must be repeated for each outer scenario $i = 1, \ldots, N_{\text{out}}$, the resulting structure is inherently nested (Figure 1). The overall computational cost scales as:

$$\text{Cost}_{\text{NMC}} \propto N_{\text{out}} \times N_{\text{in}},$$

up to model-dependent constants capturing projection engine complexity (e.g. number of policy model points, granularity of asset models, time discretization steps). In realistic life portfolios, both N_{out} and N_{in} must be large to control sampling error at the portfolio level, leading to substantial runtime.

The statistical challenge is further compounded by the quantile objective. The SCR targets the 99.5% tail of the loss distribution, and this tail is sensitive not only to outer sampling variability but also to noise in $\widehat{L}_1^{(i)}$. Inner simulation error may induce mis-ordering of scenarios near the tail, thereby distorting the empirical quantile and potentially introducing bias. Importantly, reducing the variance of the inner estimator uniformly across all scenarios is not necessarily optimal, since only a small fraction of outer paths ultimately determine the tail estimate.

The literature on nested risk estimation shows that naive nested Monte Carlo estimators can exhibit slow convergence for risk measures such as VaR, especially when inner noise is not properly controlled relative to the outer sample size. In particular, inefficient allocation between N_{out} and N_{in} can lead to suboptimal complexity and persistent bias effects [10, 3]. These findings motivate more refined allocation strategies and advanced variance-reduction techniques tailored to the nested structure of the SCR problem.

2.3 Least Squares Monte Carlo (LSMC) as an Industrial Compromise

LSMC originates from the Longstaff–Schwartz methodology for American options [12]. In solvency applications, the key idea is to avoid performing a full inner Monte Carlo valuation for every outer scenario. Instead, one constructs a proxy function that approximates the time-1 liability value as a function of the risk factors:

$$L_1 \approx \phi(X_1; \theta),$$

where X_1 is a vector of risk factors at $t = 1$, ϕ is a parametric or semi-parametric function (e.g. polynomial basis expansions, splines, sparse regressions, or neural networks), and θ is calibrated on a training set of scenarios that are fully valued via nested simulation.
The economic rationale is straightforward: while the brute-force nested Monte Carlo approach requires solving a costly conditional valuation problem for each outer scenario, many of these valuations share similar structural dependence on the underlying state variables. If the mapping $X_1 \mapsto L_1$ is sufficiently smooth, it can be learned from a relatively small but carefully designed training set, and then evaluated cheaply across a much larger outer sample. In this way, the expensive inner simulation is performed only during calibration, rather than during every production run of the SCR.
Speed and scalability are the main benefits. Once trained, the proxy function ϕ can be evaluated almost instantaneously across a large set of outer scenarios, reducing computational cost from $\mathcal{O}(N_{\text{out}} \times N_{\text{in}})$ to approximately $\mathcal{O}(N_{\text{train}} \times N_{\text{in}} + N_{\text{out}})$, where N_{train} is the size of the calibration set. This enables frequent recalculations, sensitivity analyses, stress testing, and broad what-if coverage within practical time windows. In large life portfolios, this reduction in runtime is often the primary motivation for adopting proxy techniques.
The efficiency gain comes at the cost of introducing approximation error. Proxy methods concentrate model risk in regions where training density is low and they face extrapolation risk when outer scenarios fall outside the domain well represented by the calibration set. Their performance depends strongly on feature selection, basis design, and regularization choices, which introduces an additional layer of model risk and requires a robust governance framework, including careful calibration design, monitoring of parameter stability, periodic re-training, out-of-sample validation, and explicit assessment of tail accuracy.
A key practical challenge concerns the specification of the proxy function itself. Although convergence of LSMC regressions holds for fairly general families of basis functions, numerical performance can vary substantially depending on the chosen representation, the number and structure of regressors, and the dimensionality of the state space [11]. In solvency applications, where the goal is often to approximate the full distribution rather than only conditional expectations, the choice of basis becomes even more delicate and may require significant tuning. Furthermore, regression-based proxy approaches are exposed to the classical curse of dimensionality: as the dimension of the risk-factor vector increases, the number of basis terms required to maintain accuracy typically grows rapidly, leading to higher computational cost, instability, and potential overfitting issues [1]. This limitation is particularly relevant in ALM and capital models, where the state space can be high-dimensional and non-Markovian, making low-dimensional proxy specifications unable to fully capture the relevant dynamics and limiting scalability with respect to dimensionality.
Conceptually, the core trade-off is that simulation error (which is statistical and can be quantified via confidence intervals) is replaced by approximation error (which is structural and harder to bound rigorously). In particular, the impact of proxy misspecification on extreme quantiles is difficult to characterize with transparent probabilistic guarantees, making validation in a regulatory context both technically and operationally demanding.
In practice, beyond the basic LSMC methodology, corrective and stabilization approaches, as well as advanced techniques for calibration, automatic selection of proxy structure, and computational optimization (see Moody's White Paper, [5]), can be introduced to mitigate these issues; however, the inherent approximation errors and limitations in extreme scenarios remain, requiring ongoing monitoring and robust model governance.

3. Multilevel Monte Carlo

3.1 From Nested Simulation to Multilevel Methods

As explained in Section *The Structure of SCR in an Internal Model*, under an Internal Model, the Solvency Capital Requirement is defined as:

$$SCR = VaR_{0.995}(\Delta BOF),$$

$$\Delta BOF = BOF_0 - BOF_1.$$

Since the VaR is a quantile, it is convenient to work with the associated exceedance probability. For any threshold $x \in \mathbb{R}$ define:

$$I(x) := \mathbb{P}(\Delta BOF > x) = \mathbb{E}^{\mathbb{P}}\left[\mathbf{1}_{\{\Delta BOF > x\}}\right].$$

Then SCR can be characterised as the smallest x such that $I(x) \leq 0.005$.
The nested structure arises because evaluating ΔBOF at time $t = 1$ requires a market-consistent liability value, which is a conditional risk-neutral expectation. Writing:

$$L_1 = \mathbb{E}^{\mathbb{Q}}[Y \mid \mathcal{F}_1],$$

with Y the discounted future cash flows under $\mathbb{Q}$, each outer scenario determines a value of L_1 only through an inner Monte Carlo estimation of this conditional expectation. Hence, estimating $I(x)$ leads to a nested Monte Carlo (NMC) procedure with computational cost proportional to $J \times K$, where J is the number of outer scenarios and K the number of inner simulations per scenario.
For tail quantities driven by discontinuous indicators, crude nested Monte Carlo exhibits an unfavourable accuracy–cost trade-off; in particular, achieving an RMSE of order ε typically entails a complexity of order $\mathcal{O}(\varepsilon^{-3})$, which rapidly becomes computationally prohibitive in realistic insurance applications (see e.g. [10, 3]). This computational barrier motivates replacing a single uniformly high-accuracy nested estimator with a multilevel construction. The MLMC framework has recently been advocated for SCR computation in insurance contexts [2], where it is shown to deliver substantial runtime reductions while preserving a fully simulation-based methodology.

3.2 General Principle of Multilevel Monte Carlo

MLMC replaces a single approximation with a hierarchy of increasingly accurate (and increasingly expensive) approximations, combined through a telescoping sum. In our setting, the quantity of interest is the indicator in (**??**), which we write as:

$$P(x) := \mathbf{1}_{\{\Delta BOF > x\}}.$$

The dependence on the inner valuation is implicit through $L_1 = \mathbb{E}^{\mathbb{Q}}[Y \mid \mathcal{F}_1]$.
Let $(P_\ell)_{\ell=0,\dots,L}$ be a sequence of approximations of P with increasing accuracy, indexed by a simulation level ℓ. MLMC relies on the the telescopic decomposition:

$$\mathbb{E}[P_L(x)] = \mathbb{E}[P_0(x)] + \sum_{\ell=1}^{L} \mathbb{E}\left[P_\ell(x) - P_{\ell-1}(x)\right].$$

Each term is estimated by Monte Carlo with sample size J_ℓ:

$$\widehat{I}_{\mathrm{MLMC}}(x) = \frac{1}{J_0}\sum_{j=1}^{J_0} P_0^{(j)}(x) + \sum_{\ell=1}^{L}\frac{1}{J_\ell}\sum_{j=1}^{J_\ell}\Big(P_\ell^{(j)}(x) - P_{\ell-1}^{(j)}(x)\Big).$$

The essential mechanism is the following:

- Low levels (ℓ small) are computationally cheap but biased. They are simulated many times to reduce variance.
- High levels (ℓ large) are accurate but expensive. They are simulated only a few times to correct the bias.
- Strong coupling between P_ℓ and $P_{\ell-1}$ ensures that the variance of $P_\ell - P_{\ell-1}$ decreases with ℓ.

Under suitable regularity conditions, this strategy reduces the overall computational complexity from $\mathcal{O}(\varepsilon^{-3})$ for crude nested Monte Carlo to $\mathcal{O}(\varepsilon^{-2})$, which corresponds to the optimal Monte Carlo rate [7].

3.3 MLMC in the Nested SCR Framework

In the nested SCR setting, the levels typically correspond to different numbers of simulations J_ℓ and K_ℓ, often chosen in geometric progression:

$$J_\ell = J_0 2^{-\ell}, \quad K_\ell = K_0\, 2^{\ell}, \qquad \ell = 0, \ldots, L.$$

At level ℓ, the conditional expectation $\mathbb{E}^{\mathbb{Q}}[Y \mid \mathcal{F}_1]$ is approximated by a Monte Carlo estimator with K_ℓ samples. The MLMC estimator then combines corrections across levels so that:

- The bias ($\mathbb{E}[P_L] - P$) induced by approximating the inner expectation is controlled by the maximum level L.
- The variance is reduced by allocating more outer samples J_ℓ to cheaper levels. In particular, the variance of the estimator is:

$$Var(P_L) = Var(P_0) + \sum_{l=1}^{L} Var(P_l - P_{l-1}).$$

This multilevel allocation results in a reduction of computational burden while preserving a fully simulation-based approach. Importantly, unlike proxy techniques such as Least Squares Monte Carlo or Replicating Portfolios, MLMC does not rely on parametric approximations of the response surface. This feature significantly simplifies model validation in an internal model context, as the methodology remains structurally close to full nested simulation.

3.4 Adaptive Refinements for Quantile Estimation

In the specific case of SCR estimation, the quantity of interest is driven by the indicator $g(u) = \mathbf{1}_{\{u \geq x\}}$, which is discontinuous at the loss threshold x. This lack of smoothness deteriorates the theoretical complexity of the standard MLMC estimator compared to the smooth-function case.
Giles and Haji-Ali (2019) [8] introduce adaptive allocation strategies in which the number

of inner simulations depends on the distance of the conditional expectation from the loss threshold.
Intuitively, outer scenarios for which the estimated loss is far from the threshold require only a coarse inner estimate, while scenarios close to the threshold demand higher precision to avoid misclassification. This stochastic allocation of inner samples further improves efficiency and leads to near-optimal complexity of order:

$$\mathcal{O}\left(\varepsilon^{-2}\log(\varepsilon)^2\right).$$

Such adaptive MLMC schemes are particularly well suited for rare-event estimation and tail quantile computation, making them highly relevant for internal model SCR calculations in insurance.

4. Application to a Participating Life Product with Minimum Guarantee

This section introduces a representative toy example to illustrate the application of the Multilevel Monte Carlo (MLMC) methodology to the computation of the Solvency Capital Requirement (SCR).

4.1 Product Outline

Consider an index-linked life insurance contract with a minimum guaranteed benefit. The policy pays a benefit either at contractual maturity T or at the end of the policy year of death of the insured, denoted by t^* (discrete time). The payoff structure embeds both financial and biometric risk components and can be interpreted as a European put-type guarantee written on an underlying equity index.
Formally, let $(S_t)_{t\geq 0}$ denote the value of the reference equity index under a risk-neutral measure, and let K be the guaranteed capital. The value at time 0 is given by

$$V_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-r\,(T\wedge t^*)}\,(K - S_{T\wedge t^*})_+\right],$$

where r is the risk-free interest rate and $T \wedge t^*$ represents the minimum between maturity and the end of year of death. The contract therefore terminates at the earlier of death or maturity, and the guarantee ensures that the policyholder (or beneficiaries) receives at least K at termination. Economically, this structure resembles a put option with random maturity driven by the lifetime of the insured.
Despite its apparent simplicity, this contract captures the essential features of a broad class of life insurance products commonly observed in practice. First, it incorporates market risk through the stochastic dynamics of the underlying index S_t, typically modeled as a diffusion process (e.g., geometric Brownian motion) or within a more general risk-factor framework. Second, it embeds mortality risk via the random time of death t^*, which interacts with financial risk by inducing a stochastic and path-dependent effective maturity. Third, the presence of the minimum guarantee introduces a non-linear payoff profile, making the liability sensitive to downside equity movements and thereby generating significant tail risk, which is central in SCR calculations under Solvency II.
From a risk management perspective, this stylized product provides a natural laboratory for studying the joint impact of financial shocks and biometric uncertainty on the distribution of Own Funds. It also highlights the computational challenges associated with nested simulation approaches traditionally used for SCR estimation, particularly when

guarantees and stochastic termination times are involved. For these reasons, it represents a parsimonious yet sufficiently rich benchmark for assessing the efficiency gains delivered by MLMC techniques in an insurance context.

4.2 SCR Framework and Numerical Implementation

In order to compute the Solvency Capital Requirement (SCR), a one-year risk horizon is adopted in accordance with the standard Solvency II framework. In particular, the relevant risk factors are simulated up to time one under the real-world probability measure $\mathbb{P}$. The two sources of uncertainty considered in this setting are: (i) the financial risk factor, represented by the underlying index (S_t), and (ii) the biometric risk factor, represented by the survival status of the insured, which are assumed to be independent.
More precisely, joint scenarios for $(S_1, \mathbf{1}_{\{1<t^*\}})$ are generated under $\mathbb{P}$. If the insured dies within the first year, the contract terminates and the liability at time one is fully determined by the payoff already realized. Conversely, if the insured survives the first year, the SCR requires the evaluation of the liability value at time one, denoted by V_1, conditional on the realized market state S_1. In this case, V_1 corresponds to the risk-neutral value (under $\mathbb{Q}$) of the remaining cash flows from 1 to $T \wedge t^*$, given survival up to the first year.
Conditionally on survival and on S_1, the continuation value can be written as

$$V_1 = \mathbb{E}^{\mathbb{Q}}\left[e^{-r((T\wedge t^*)-1)}\left(K - S_{T\wedge t^*}\right)_+ \,\middle|\, \mathcal{F}_1\right],$$

where the expectation is taken under the risk-neutral measure $\mathbb{Q}$, in accordance with market-consistent valuation principles and $\mathcal{F}_1$ represents the sigma-algebra up to time one. To assess the numerical performance of different estimation techniques, the value V_1 is computed using several alternative approaches:

- **Closed-form benchmark.** Owing to the simplified structure of the product and assuming a geometric Brownian motion dynamics for S_t, a closed-form pricing formula can be derived. This value is used as a benchmark to assess the convergence properties and bias of the simulation-based estimators.
- **Nested Monte Carlo (NMC).** For each outer real-world scenario up to time one, an inner Monte Carlo simulation under $\mathbb{Q}$ is performed to estimate the conditional expectation defining V_1. While conceptually straightforward, this approach is computationally expensive due to the nested structure.
- **Multilevel Monte Carlo (MLMC).** In order to reduce computational cost, the MLMC estimator is implemented. The telescopic decomposition of expectations across levels allows to control bias and variance simultaneously, achieving a significant reduction in complexity compared to standard nested Monte Carlo.

This setup mirrors the practical SCR computation problem: a first-stage projection of risk factors over one year under $\mathbb{P}$, followed by a second-stage market-consistent valuation under $\mathbb{Q}$ of the remaining liabilities. The toy example therefore provides a controlled environment in which the trade-off between accuracy and computational complexity of different Monte Carlo methodologies can be rigorously assessed.

4.3 Numerical Results

The financial market is modeled through a geometric Brownian motion under the real-world probability measure $\mathbb{P}$:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t^{\mathbb{P}},$$

	Classical Monte Carlo			Multilevel Monte Carlo		
N_{it}	J	K	Bias (bps)	J_0	K_0	Bias (bps)
64,000,000	32,000	2,000	6.08	16,000	1,000	5.48
128,000,000	64,000	2,000	5.77	32,000	1,000	0.46
256,000,000	128,000	2,000	1.08	64,000	1,000	0.95
512,000,000	256,000	2,000	5.61	128,000	1,000	1.73

TABLE 1: *Example results of the MLMC and classical Nested MC algorithms.*

while under the risk-neutral measure $\mathbb{Q}$ the drift is replaced by the risk-free rate r.
The baseline parameters used in the numerical experiments are:

$$S_0 = 100, \quad K = 100, \quad \sigma = 0.30,$$

$$\mu = 0.05, \quad r = 0.02, \quad T = 20.$$

The initial age of the insured is set to 50, and survival probabilities are derived from Italian population life tables published by ISTAT (latest available year). Financial and biometric risks are assumed to be independent.
Even though a closed-form pricing formula is available for the continuation value, the Solvency II framework requires simulation of the one-year distribution under the real-world measure $\mathbb{P}$.
The benchmark procedure is therefore structured as follows:

1. Simulate S_1 under $\mathbb{P}$;
2. Simulate the survival status over $[0, 1]$;
3. If death occurs within the first year, the liability is fully realized.
4. If survival occurs, compute:

$$V_1 = \mathbb{E}^{\mathbb{Q}}\left[e^{-r((T\wedge t^*)-1)}\left(K - S_{T\wedge t^*}\right)_+ \,\middle|\, \mathcal{F}_1\right].$$

Considering the independence between the mortality and the financial risk, V_1 can be computed via a closed-form formula. Hence, the benchmark still requires outer real-world simulation but avoids inner simulation by replacing it with the analytical formula. This benchmark is used to assess bias and convergence of simulation-based estimators.
As explained in Section *MLMC in the Nested SCR Framework*, the computational effort for each level is defined through the following geometric allocation:

$$J_\ell = J_0 2^{-\ell}, \quad K_\ell = K_0 2^{\ell}, \quad \ell = 0, \ldots, L.$$

This construction ensures a progressive refinement of the inner conditional expectation across levels while simultaneously reducing the number of outer simulations. The geometric structure not only stabilizes the variance of the level differences, but also allows for a convenient expression of the total computational budget. Indeed, since each level requires $J_\ell \times K_\ell = J_0 \times K_0$ iterations, the total number of iterations of the MLMC algorithm is $N_{\text{it}} = (L+1) \times J_0 \times K_0$.
In our numerical experiment, we set:

$$L = 3, \qquad J_0 \in \{2000 \cdot 2^k : k = 0, \ldots, 8\},$$

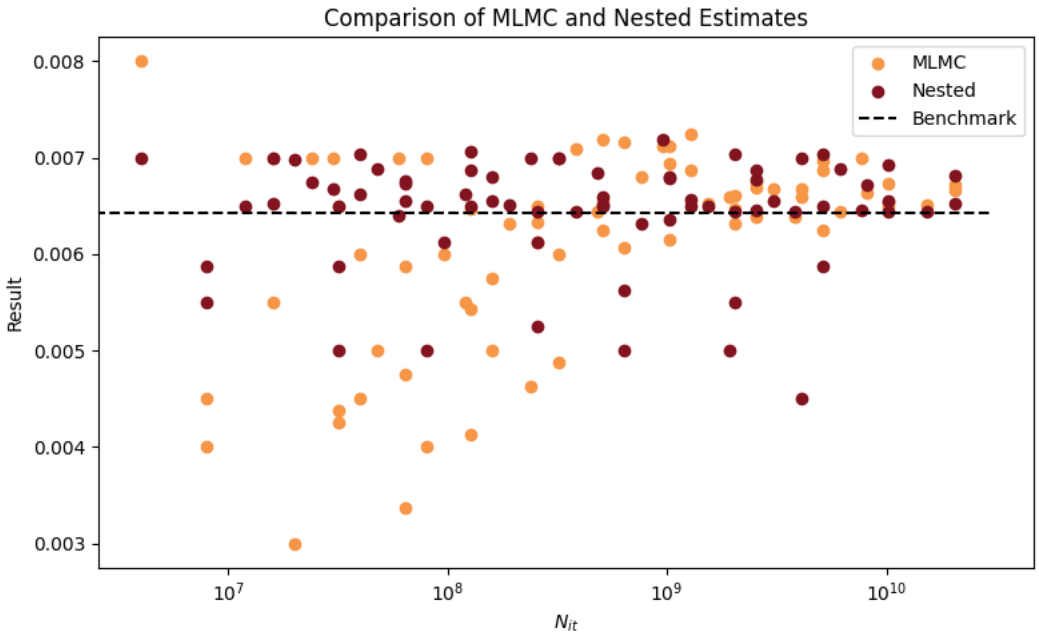

FIGURE 2: Bias as a function of the total computational budget N_{it} for MLMC and classical Nested Monte Carlo. For large N_{it}, the MLMC algorithm remains closer to the benchmark value, while the classical Nested Monte Carlo still shows significant error.

$$K_0 \in \{1000, 2000, 3000, 4000, 5000, 7500, 10000\}.$$

In order to ensure a coherent comparison with the classical Nested Monte Carlo estimator, the parameters (J, K) of the nested method are selected so that the overall computational effort coincides with that of MLMC.
More precisely,

$$J \times K = (L+1) \times J_0 \times K_0.$$

In this way, both approaches are compared under identical total iteration counts.
For illustrative purposes, Table 1 reports a small subset of representative configurations. The table provides an example of how the computational budget is allocated and the resulting biases (expressed in basis points with respect to the analytical benchmark).
A graphical comparison of the two estimators is now presented in order to better highlight their convergence behavior and stability properties.
Figure 2 displays the bias as a function of the total computational budget N_{it} for both MLMC and classical Nested Monte Carlo. This two-dimensional representation isolates the effect of the overall computational effort and allows to directly assess the relative speed of convergence of the two estimators.
Several patterns emerge from the comparison. For relatively low computational budgets, both estimators exhibit non-negligible bias, with fluctuations driven by the variance of the outer scenarios and by the accuracy of the inner conditional expectation estimates. In this region, the classical nested approach may occasionally display slightly smaller bias due to its uniform allocation of computational effort across all scenarios.
However, as the total number of iterations increases, the MLMC estimator exhibits a more stable convergence pattern. The bias decreases more regularly with respect to N_{it}, reflecting the variance reduction mechanism induced by the telescopic decomposition across levels. In contrast, the classical nested estimator shows less regular behavior, with occasional spikes in the bias even at higher computational budgets. This instability is primarily driven by the inefficient allocation of inner simulations across all outer scenarios, regardless of their

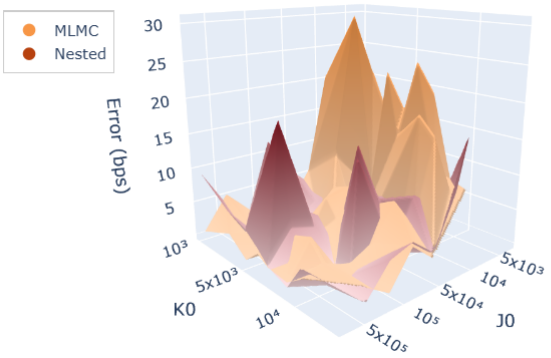

FIGURE 3: Bias surface (in basis points) as a function of outer (J_0) and inner (K_0) simulation parameters for MLMC and classical Nested Monte Carlo.

contribution to the overall variance.
Figure 3, instead, provides a three-dimensional visualization of the bias as a function of the outer and inner simulation parameters. This representation offers additional insight into how the allocation of computational resources between outer and inner simulations influences the accuracy and robustness of the estimators, thereby complementing the analysis based solely on the total budget.
For low values of J_0 (or J), corresponding to a limited number of outer real-world scenarios, the MLMC estimator exhibits higher bias. This reflects the fact that insufficient sampling of the one-year loss distribution negatively impacts the accuracy of tail estimation, which is central for SCR computation.
However, as both the number of outer scenarios and the inner simulation effort increase, the MLMC estimator demonstrates a smoother and more stable convergence surface. In particular, the bias does not display the pronounced peaks observed in the classical nested case. This highlights the efficiency of the multilevel allocation, which concentrates computational effort where it contributes most to variance reduction.
Overall, the numerical evidence confirms that, while MLMC may require a minimal scale to outperform the classical nested approach, it provides superior scalability and robustness as computational resources increase. This property is particularly relevant in realistic insurance applications, where high-dimensional portfolios and long maturities make fully nested simulation computationally prohibitive.

5. Conclusions

Internal Model SCR computation under Solvency II requires robust estimation of tail risk based on one-year changes in BOF under market-consistent valuation at the projection horizon. Although fully nested Monte Carlo provides a conceptually straightforward and model-consistent benchmark, its computational complexity (scaling with $N_{\text{out}} \times N_{\text{in}}$) together with the intrinsic difficulty of tail estimation, often limits its practical applicability at industrial scale. Proxy-based approaches reduce runtime but shift the burden toward model specification, governance, and validation, while introducing approximation risk that can be particularly material in the SCR tail.
Multilevel Monte Carlo provides a coherent alternative that preserves the nested valuation structure while improving efficiency through variance reduction and level coupling. By exploiting a telescoping decomposition across simulation resolutions, MLMC achieves faster

convergence compared to standard nested simulation, enabling accurate SCR estimation with significantly lower computational cost. The adaptive extension further strengthens this framework by reallocating computational effort, increasing precision where it matters most while keeping additional costs contained.

Overall, the proposed methodology combines statistical robustness, scalability, and transparency, offering a practical pathway toward industrial SCR computation without relying on proxy approximations. A particularly promising extension is the use of multi-index Monte Carlo methods, where refinement is performed simultaneously along several discretization dimensions rather than only across inner simulation levels. In insurance applications, these dimensions may include time discretization of stochastic processes or numerical approximations embedded in path-dependent liability valuation. By exploiting mixed-difference estimators across multiple axes, multi-index schemes can mitigate the exponential growth in computational effort that typically arises when several sources of discretization error interact. This direction, especially relevant for portfolios with strong path dependence, will be investigated in detail in forthcoming work, where a multi-index extension of the proposed framework will be introduced and analysed.

References

[1] **Alfonsi, A., Cherchali, A. and Infante Acevedo, J.A.** *Multilevel Monte-Carlo for computing the SCR with the standard formula and other stress tests*. 2021.

[2] **Boumezoued, A., Cherchali, A. and Truc, M.** *Efficient Computation of Solvency Capital Requirement using Multilevel Monte Carlo Methods*. Milliman White Paper, 2022.

[3] **Broadie, M., Du, Y. and Moallemi, C.** *Efficient Risk Estimation via Nested Sequential Simulation*. Management Science, Vol. 57, No. 6, 2011.

[4] **EIOPA.** *Guidelines on the use of Internal Models*. EIOPA Guidelines and Supervisory Statements, 2015.

[5] **Elliot, M.** *Making Proxy Functions Work in Practice*. Moody's Analytics White Paper, 2016.

[6] **European Parliament and the Council of the European Union.** *Directive 2009/138/EC (Solvency II)*. Official Journal of the European Union, 2009.

[7] **Giles, M. B.** *Multilevel Monte Carlo Path Simulation*. Operations Research, Vol. 56, No. 3, 2008.

[8] **Giles, M. B. and Haji-Ali, A.-L.** *Multilevel Nested Simulation for Efficient Risk Estimation*. SIAM/ASA Journal on Uncertainty Quantification, Vol. 7, No. 2, pp. 497-525, 2019.

[9] **Glasserman, Paul** *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

[10] **Gordy, M. B. and Juneja, S.** *Nested Simulation in Portfolio Risk Measurement*. Management Science, Vol. 56, No. 10, 2010.

[11] **Ha, H. and Bauer, D.** *A Least-Squares Monte Carlo Approach to the Calculation of Capital Requirements*. 2018.

[12] **Longstaff, F. A. and Schwartz, E. S.** *Valuing American Options by Simulation: A Simple Least-Squares Approach*. Review of Financial Studies, Vol. 14, No. 1, 2001.

Iason is an international firm that consults Financial Institutions on Risk Management. Iason is a leader in quantitative analysis and advanced risk methodology, offering a unique mix of know-how and expertise on the pricing of complex financial products and the management of financial, credit and liquidity risks. In addition Iason provides a suite of essential solutions to meet the fundamental needs of Financial Institutions.

ESSENTIAL SERVICES FOR
FINANCIAL INSTITUTIONS