

Nested Multilevel Monte Carlo with Preintegration for Efficient Risk Estimation

Yu Xu

Tsinghua University

Yifeng Yu

yuf22@mails.tsinghua.edu.cn

Tsinghua University

Xiaoqun Wang

Tsinghua University

Research Article

Keywords: Nested simulation, Multilevel Monte Carlo, Risk estimation, Preintegration

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11131834/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Nested Multilevel Monte Carlo with Preintegration for Efficient Risk Estimation

Yu Xu¹, Yifeng Yu^{1*}, Xiaoqun Wang¹

^{1*}Department of Mathematical Sciences, Tsinghua University, Beijing,
100084, People's Republic of China.

*Corresponding author(s). E-mail(s): yyf22@mails.tsinghua.edu.cn;
Contributing authors: yu-xu22@mails.tsinghua.edu.cn;
wangxiaoqun@mail.tsinghua.edu.cn;

Abstract

Nested Monte Carlo is widely used for risk estimation, but its efficiency is limited by the discontinuity of the indicator function and high computational cost. This paper proposes a nested Multilevel Monte Carlo (MLMC) method combined with preintegration for efficient risk estimation. We first use preintegration to integrate out one outer random variable, which effectively handles the discontinuity of the indicator function. We then construct the smoothed MLMC estimator to reduce the computational cost. We prove that the strong convergence rate of the MLMC combined with preintegration reaches -1 , compared with $-1/2$ for the standard MLMC. Consequently, a nearly optimal computational complexity is achieved. Besides, our method can also handle the high-kurtosis phenomenon caused by indicator functions. Numerical experiments verify that our MLMC with preintegration outperforms the standard MLMC and the optimal computational cost can be attained. Combining our method with quasi-Monte Carlo further improves its performance in high dimensions.

Keywords: Nested simulation, Multilevel Monte Carlo, Risk estimation, Preintegration

1 Introduction

Risk estimation is a basic and important task in financial risk management. It is used in many real financial activities, such as portfolio investment, option pricing and bank risk control. The goal of risk estimation is to calculate the probability of extreme

loss events for a financial portfolio accurately and quickly. In most practical cases, this calculation can be transformed into estimating the expectation of a function of a conditional expectation. For this kind of problem, nested Monte Carlo simulation (Broadie et al. 2011; Gordy and Juneja 2010) is the most commonly used method, because it is easy to implement and can adapt to complex financial models.

Nested Monte Carlo simulation works with two layers of sampling: the outer layer generates samples of the multidimensional random variable, and the inner layer generates samples to estimate the conditional expectation for each outer sample. After getting the conditional expectation estimate for all outer samples, we aggregate these results to get the final estimate of the target risk measure. Gordy and Juneja (2010) showed that to achieve a root mean squared error (RMSE) for given tolerance TOL, allocating proper computational effort in each level results in a total cost of $O(\text{TOL}^{-3})$. Nested simulation can be made more efficient by allocating computational effort nonuniformly across scenarios. Particularly, Broadie et al. (2011) proposed an adaptive procedure to allocate computational effort to inner simulations. The resulting nonuniform nested simulation estimator enjoys a reduced complexity of $O(\text{TOL}^{-5/2})$ under certain conditions.

To reduce the computational cost of the nested Monte Carlo simulation, Multilevel Monte Carlo (MLMC) (Giles 2008, 2015; Heinrich 1998) has been developed and widely used in recent years. MLMC is a powerful variance reduction technique. It decomposes the target expectation into a sum of differences between estimates at different accuracy levels. For high-accuracy levels with high sampling cost, MLMC uses fewer samples, and for low-accuracy levels with low cost, it uses more samples. This strategy of sample allocation greatly reduces the total computational cost compared with the standard Monte Carlo method. Recently, Giles and Haji-Ali (2019) proposed to use MLMC in nested simulation for risk estimation. They showed that in the original MLMC, the complexity is $O(\text{TOL}^{-5/2})$. By incorporating the adaptive allocations procedure of Broadie et al. (2011), Giles and Haji-Ali (2019) showed that the complexity of MLMC can be reduced to $O(\text{TOL}^{-2}(\log(\text{TOL}))^2)$ under certain conditions. For applications of nested MLMC, we refer to Bujok et al. (2015); Giles (2018); Giles and Goda (2019); Goda et al. (2020, 2018).

However, the standard MLMC cannot solve the problem of the discontinuity of the indicator function in risk estimation. This discontinuity makes the standard MLMC having poor robustness: its variance decays very slowly at deep high-accuracy levels, and the kurtosis increases sharply with the level. As a result, the standard MLMC still has limited efficiency in practical risk estimation tasks. To address the discontinuity problem of the indicator function, preintegration (Achtsis et al. 2013a,b; Bayer et al. 2018; Glasserman and Staum 2001; Griewank et al. 2018) (also called conditional MC or conditional sampling) has become a research focus in recent years. It integrates out one of the random variables analytically or numerically, which can eliminate the discontinuity caused by the indicator function. Moreover, preintegration is easy to combine with the existing MLMC framework, and it does not change the original bias of the nested simulation estimator. This means that preintegration can improve the convergence rate and robustness without increasing the bias, which makes it a perfect match for MLMC in risk estimation.

In this paper, we combine the preintegration technique with nested MLMC to develop a new efficient method for risk estimation. We start with the basic problem setting of nested simulation for risk estimation, and review the asymptotic bias property of the standard nested Monte Carlo estimator (Gordy and Juneja 2010). We then introduce the preintegration technique in detail: we define the preintegrated function for the risk measure. When a closed-form explicit expression for the function after preintegration does not exist, Bayer et al. (2023, 2024) proposed a numerical smoothing method to approximate the preintegration via Newton iteration and one-dimensional numerical quadrature schemes. They use the Newton iteration method to approximate the discontinuity point, and apply the two-sided Laguerre quadrature rule to compute the one-dimensional preintegration numerically.

After constructing the smoothed estimator with preintegration, we combine it with the MLMC method to construct the nested MLMC estimator with preintegration. Then we conduct a comprehensive and detailed theoretical analysis for the proposed method. First, we decompose the mean squared error (MSE) of the proposed estimator into three parts: the bias error of the nested simulation, the statistical error of the MLMC method, and the numerical error from preintegration. We analyze the convergence rate for each part separately. Second, we establish the strong convergence of the proposed estimator, and derive that the variance of the level differences decays at a rate of $O(m_\ell^{-1})$, where m_ℓ is the inner sample size at level ℓ . Third, we analyze the computational complexity of the method, and show that it has a complexity of $O(\text{TOL}^{-2}(\log(\text{TOL}))^2)$. Fourth, we prove that the kurtosis of the proposed method is bounded, which means the method has good robustness and avoids the high-kurtosis problem of the standard MLMC.

The numerical experiments confirm our theoretical conclusions. The smoothed MLMC with preintegration has more stable weak convergence than the standard MLMC, and its variance decays fast and nearly achieves our theoretical convergence rate. The kurtosis of the proposed method remains at a low level and almost does not change with the level, implying it has excellent robustness. In terms of computational cost, the proposed method needs much less cost to achieve the same accuracy, and its cost attained theoretical complexity $O(\text{TOL}^{-2}(\log(\text{TOL}))^2)$. We also extend the proposed method by combining it with quasi-Monte Carlo (Caffisch 1998; Caffisch et al. 1997; Dick and Pillichshammer 2010; Niederreiter 1992; Owen 2023) (QMC) to increase the accuracy of inner simulation in high-dimensional cases, and test the performance of Multilevel QMC (Goda et al. 2018; Xu et al. 2024) (MLQMC) combined with preintegration (called smoothed MLQMC) in high dimensions. The smoothed MLQMC significantly improves the computational accuracy in high-dimensional settings, and its strong convergence rate even reaches $-3/2$ in our experiments. Furthermore, combined with antithetic sampling (Bujok et al. 2015; Giles and Goda 2019; Giles and Haji-Ali 2019), the smoothed MLMC achieves further variance reduction and even higher strong convergence rates, thereby effectively reducing the computational complexity.

This paper makes three main contributions. First, we successfully combine the preintegration smoothing technique with nested MLMC, and construct a smoothed MLMC estimator for risk estimation. Second, we conduct a complete theoretical analysis for the proposed method, including error decomposition, strong convergence,

complexity and robustness, and derive clear theoretical results about convergence rate, variance decay and complexity. Third, we design detailed numerical experiments to verify the theoretical results, and show that the proposed method has obvious advantages over the standard MLMC in practical risk estimation, especially in high-dimensional cases.

The rest of the paper is organized as follows. Section 2 introduces the basic setting of nested simulation for risk estimation and the preintegration technique. Section 3 combines preintegration with MLMC to construct the proposed estimator, and performs the detailed theoretical analysis (error, convergence, complexity and robustness). Section 4 presents the numerical experiments. Section 5 presents the main conclusions of the paper and points out the directions for future research. The Appendix A provides the proofs of the lemmas used in the theoretical analysis.

2 Nested simulation and preintegration

2.1 Problem setting and nested simulation

In risk estimation, nested Monte Carlo methods are widely adopted to solve the problem of estimating expectations of functions of conditional expectations. We consider the problem of estimating

$$\theta := \mathbb{P}(\mathbb{E}[X|\boldsymbol{\omega}] > c) = \mathbb{E}[\mathbb{I}\{\mathbb{E}[X|\boldsymbol{\omega}] > c\}] := \mathbb{E}[g(\varphi(\boldsymbol{\omega}))] \quad (1)$$

via simulation for a given constant c , where $\varphi(\boldsymbol{\omega}) := \mathbb{E}[X|\boldsymbol{\omega}]$ and $g(\cdot) = \mathbb{I}\{\cdot > c\}$. The inner expectation of the one-dimensional random variable X , is conditional on the value of the outer multidimensional random variable $\boldsymbol{\omega}$. We aim to estimate the expectation of a function of a conditional expectation in (1) via nested simulation. In the nested Monte Carlo approach, we consider the estimator

$$\hat{\theta}_{n,m} := \frac{1}{n} \sum_{i=1}^n \hat{g}_m(\boldsymbol{\omega}_i),$$

where

$$\hat{g}_m(\mathbf{y}) = \mathbb{I}\{\hat{\varphi}_m(\mathbf{y}) > c\}, \quad \hat{\varphi}_m(\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m X_j(\mathbf{y}), \quad (2)$$

where $\boldsymbol{\omega}_i$ are independent and identically distributed (i.i.d.) replications of $\boldsymbol{\omega}$, and $X_j(\mathbf{y})$ are i.i.d. replications of X given $\boldsymbol{\omega} = \mathbf{y}$. The quadrature rule $\hat{\varphi}_m(\mathbf{y})$ is used to estimate the conditional expectation $\varphi(\mathbf{y}) = \mathbb{E}[X|\boldsymbol{\omega} = \mathbf{y}]$ in the inner simulation.

The bias of the nested Monte Carlo estimator is a critical factor affecting estimation accuracy, and the following theorem from Gordy and Juneja (2010) characterizes its asymptotic bias behavior.

Theorem 1 *Let $f(\cdot)$ be the probability density function of $\varphi(\boldsymbol{\omega})$. Assume the following:*

- *The joint density $p_m(x, y)$ of $\varphi(\boldsymbol{\omega})$ and $m^{1/2}[\hat{\varphi}_m(\boldsymbol{\omega}) - \varphi(\boldsymbol{\omega})]$ and the partial derivatives $(\partial/\partial x)p_m(x, y)$ and $(\partial^2/\partial x^2)p_m(x, y)$ exist for each m and (x, y) .*

- For each $m \geq 1$, there exist functions $f_{i,m}(\cdot)$ such that

$$\left| \frac{\partial^i}{\partial x^i} p_m(x, y) \right| \leq f_{i,m}(y), i = 0, 1, 2.$$

In addition,

$$\sup_{m \geq 1} \int |y|^r f_{i,m}(y) dy < \infty, \text{ for } i = 0, 1, 2 \text{ and } 0 \leq r \leq 4.$$

Then the bias of the nested estimator $\hat{\theta}_{n,m}$ asymptotically satisfies

$$|\mathbb{E}[\hat{\theta}_{n,m} - \theta]| = |\mathbb{E}[\hat{g}_m(\boldsymbol{\omega}) - \theta]| \leq \frac{|\Theta'(c)|}{m} + O(m^{-3/2}), \quad (3)$$

where

$$\Theta(c) = \frac{1}{2} f(c) \mathbb{E}[\sigma^2(\boldsymbol{\omega}) | \varphi(\boldsymbol{\omega}) = c],$$

and $\sigma^2(\mathbf{y})$ denotes the conditional variance of X (conditioned on $\boldsymbol{\omega} = \mathbf{y}$).

This theorem provides an explicit asymptotic upper bound for the bias of the nested Monte Carlo estimator $\hat{\theta}_{n,m}$. It demonstrates that the leading term of the bias decays at the rate of $1/m$.

2.2 Preintegration

The efficiency of nested Multilevel Monte Carlo (MLMC) simulation is often significantly degraded by the discontinuity of the indicator function $g(\cdot)$, which slows down the strong convergence rate in practical risk estimation. To avoid this issue and improve the accuracy of nested MLMC simulation, we introduce the preintegration technique, which integrates out one of the outer random variables analytically or numerically, and then the discontinuity can be effectively handled and the overall efficiency can be greatly enhanced. In this paper, for simplicity, we assume that the variable ω_1 has an independent probability density function ρ_1 with other variables, so we always perform preintegration over ω_1 . The specific procedure of inner simulation combined with preintegration is performed by the following formula

$$\begin{aligned} \theta &= \mathbb{E}[\mathbb{I}\{\varphi(\boldsymbol{\omega}) > c\}] = \mathbb{E}[g(\varphi(\boldsymbol{\omega}))] \\ &= \mathbb{E}\left[\int_{\mathbb{R}} g(\varphi(\omega_1, \boldsymbol{\omega}_{-1})) \rho_1(\omega_1) d\omega_1\right] := \mathbb{E}[h(\boldsymbol{\omega}_{-1})] \\ &\approx \mathbb{E}\left[\int_{\mathbb{R}} \hat{g}_m(\omega_1, \boldsymbol{\omega}_{-1}) \rho_1(\omega_1) d\omega_1\right] := \mathbb{E}[\hat{h}_m(\boldsymbol{\omega}_{-1})], \end{aligned}$$

where

$$h(\boldsymbol{\omega}_{-1}) = \int_{\mathbb{R}} g(\varphi(\omega_1, \boldsymbol{\omega}_{-1})) \rho_1(\omega_1) d\omega_1,$$

and

$$\hat{h}_m(\boldsymbol{\omega}_{-1}) = \int_{\mathbb{R}} \hat{g}_m(\omega_1, \boldsymbol{\omega}_{-1}) \rho_1(\omega_1) d\omega_1. \quad (4)$$

From the above calculations, we can observe that the preintegration does not change the bias of estimator of the nested simulation, and therefore preintegration cannot change the computational complexity of the nested simulation. It is worth noting that the preintegrated function $\hat{h}$ generally lacks a closed-form analytical expression, thus requiring numerical approximation of the one-dimensional integral. We follow the numerical smooth method of Bayer et al. (2023, 2024) and make the same assumption as Bayer et al. (2024). For fixed $\boldsymbol{x}_{-1}$, we assume that the function $\varphi(x_1, \boldsymbol{x}_{-1})$ has a simple root or is positive for all $x_1 \in \mathbb{R}$. This assumption is guaranteed by the following monotonicity condition 1 and the infinite growth condition 2. We present the monotonicity condition for an increasing function without loss of generality.

1. **Monotonicity condition.** $\frac{\partial \varphi}{\partial x_1}(\boldsymbol{x}) > 0, \forall \boldsymbol{x} \in \mathbb{R}^d$.
2. **Growth condition.** $\lim_{x_1 \rightarrow +\infty} \varphi(\boldsymbol{x}) = \lim_{x_1 \rightarrow +\infty} \varphi(x_1, \boldsymbol{x}_{-1}) = +\infty, \forall \boldsymbol{x}_{-1} \in \mathbb{R}^{d-1}$
or $\frac{\partial^2 \varphi}{\partial x_1^2}(\boldsymbol{x}) \geq 0, \forall \boldsymbol{x} \in \mathbb{R}^d$.

Under the above conditions, for any fixed $\boldsymbol{\omega}_{-1}$, we can uniquely determine the discontinuity point y_1^* by solving the equation

$$\varphi(y_1^*, \boldsymbol{\omega}_{-1}) = c.$$

We employ the Newton iteration method to determine the approximate discontinuity location $\bar{y}_1^*$.

The second step of the numerical smoothing framework is to perform numerical preintegration based on the approximate discontinuity point, which is formulated as follows:

$$\mathbb{E}[g(\varphi(\boldsymbol{\omega}))] = \mathbb{E}[h(\boldsymbol{\omega}_{-1})] \approx \mathbb{E}[\hat{h}(\boldsymbol{\omega}_{-1})] \approx \mathbb{E}[\bar{h}(\boldsymbol{\omega}_{-1})],$$

where

$$\begin{aligned} \hat{h}_m(\boldsymbol{\omega}_{-1}) &= \int_{\mathbb{R}} \hat{g}_m(\omega_1, \boldsymbol{\omega}_{-1}) \rho_1(\omega_1) d\omega_1 \\ &= \int_{-\infty}^{y_1^*} \hat{g}_m(\omega_1, \boldsymbol{\omega}_{-1}) \rho_1(\omega_1) d\omega_1 + \int_{y_1^*}^{+\infty} \hat{g}_m(\omega_1, \boldsymbol{\omega}_{-1}) \rho_1(\omega_1) d\omega_1, \end{aligned} \quad (5)$$

and $\bar{h}$ is the approximation of $\hat{h}$ obtained using the Newton iteration and a two-sided Laguerre quadrature rule, which is expressed as follows

$$\bar{h}_m(\boldsymbol{\omega}_{-1}) := \sum_{k=0}^{M_{\text{Lag}}} \eta_k \hat{g}_m(\zeta_k(\bar{y}_1^*), \boldsymbol{\omega}_{-1}), \quad (6)$$

where M_{Lag} represents the number of Laguerre quadrature points $\zeta_k \in \mathbb{R}$ with $\zeta_0 = \bar{y}_1^*$ and corresponding weights η_k . The quadrature points ζ_k must be systematically selected based on the approximate discontinuity point $\bar{y}_1^*$ to ensure numerical accuracy.

Equations (5) and (6) can be easily extended to the case of finitely many discontinuities. We refer to Remark 2.5 presented in Bayer et al. (2023) for this extension.

3 MLMC combined with preintegration

MLMC method is a powerful variance reduction technique that decomposes the target expectation into a telescoping sum of level differences, significantly reducing computational cost. In this section, we first review the fundamental idea of MLMC, then construct the MLMC estimator with preintegration technique proposed in the previous section. We recall that $h(\boldsymbol{\omega}_{-1}) = \int_{\mathbb{R}} g(\varphi(\omega_1, \boldsymbol{\omega}_{-1})) \rho_1(\omega_1) d\omega_1$. Instead of dealing with h directly, we consider a sequence of random variables $h_{m_0}, h_{m_1}, \dots$ with increasing approximation accuracy to h but with increasing cost per sample. The approximation accuracy improves with the increase of inner sample size, while the computational cost per level also rises accordingly. The ℓ th level approximation of h is $\hat{h}_{m_\ell}(\boldsymbol{\omega}_{-1}) = \int_{\mathbb{R}} \hat{g}_{m_\ell}(\omega_1, \boldsymbol{\omega}_{-1}) \rho_1(\omega_1) d\omega_1$, where $m_\ell = 2^{\ell+\ell_0}$ in our setting and $\ell_0 \geq 0$. We simplify the subscript notation m_ℓ to ℓ , like h_{m_ℓ} to h_ℓ , and L denotes the maximum level. Moreover, $M_{\text{Lag},\ell}$ and $\text{TOL}_{\text{Newton},\ell}$ denote the number of Laguerre quadrature points and the tolerance of the Newton method at the level ℓ , respectively.

Due to the linearity of expectation, we have the following telescoping representation

$$\mathbb{E}[\hat{h}_L] = \mathbb{E}[\hat{h}_0] + \sum_{\ell=1}^L \mathbb{E}[\hat{h}_\ell - \hat{h}_{\ell-1}].$$

Let $\hat{Y}_\ell = \hat{h}_\ell - \hat{h}_{\ell-1}$ for $\ell \geq 1$ and $\hat{Y}_0 = \hat{h}_0$, we have

$$\mathbb{E}[\hat{h}_L] = \sum_{\ell=0}^L \mathbb{E}[\hat{Y}_\ell]. \quad (7)$$

The MLMC method uses the above equality and estimates each term on the right hand side of (7) independently. The resulting MLMC estimator is given by

$$\hat{Q} = \sum_{\ell=0}^L \hat{Q}_\ell \quad (8)$$

with

$$\hat{Q}_\ell = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \hat{Y}_\ell^{(i)}, \quad 0 \leq \ell \leq L,$$

where $\hat{Y}_\ell^{(1)}, \dots, \hat{Y}_\ell^{(N_\ell)}$ are i.i.d. replications of $\hat{Y}_\ell$ for $\ell = 0, \dots, L$.

Similarly, let $\bar{Y}_\ell = \bar{h}_\ell - \bar{h}_{\ell-1}$ for $\ell \geq 1$ and $\bar{Y}_0 = \bar{h}_0$, we can construct the MLMC estimator with numerical preintegration

$$\bar{Q} := \sum_{\ell=0}^L \bar{Q}_\ell, \quad (9)$$

where

$$\bar{Q}_\ell := \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \bar{Y}_\ell^{(i)}, \quad 0 \leq \ell \leq L,$$

where $\bar{Y}_\ell^{(1)}, \dots, \bar{Y}_\ell^{(N_\ell)}$ are i.i.d. replications of $\bar{Y}_\ell$ for $\ell = 0, \dots, L$.

3.1 Error analysis

This subsection analyzes the different error contributions in the MLMC estimator with the numerical preintegration to approximate $\theta = \mathbb{E}[g(\varphi(\omega))]$ with g, φ given by (1). We have the following MSE decomposition

$$\begin{aligned} \mathbb{E} \left(\mathbb{E}[g(\varphi(\omega))] - \bar{Q} \right)^2 &\leq \underbrace{3 \left(\mathbb{E}[g(\varphi(\omega))] - \mathbb{E}[\hat{g}_L(\omega)] \right)^2}_{\text{Error I: bias or weak error}} \\ &+ \underbrace{3 \mathbb{E} \left(\mathbb{E}[\hat{h}_L(\omega_{-1})] - \hat{Q} \right)^2}_{\text{Error II: MLMC statistical error}} \\ &+ \underbrace{3 \mathbb{E} \left(\hat{Q} - \bar{Q} \right)^2}_{\text{Error III: numerical integration and root-finding error}} \\ &=: 3(E_1 + E_2 + E_3), \end{aligned} \quad (10)$$

where $\hat{h}_L$ is given by (4), and $\hat{Q}$ and $\bar{Q}$ are given by (8) and (9), respectively. Here E_1 , E_2 , and E_3 denote the squared bias, the statistical mean squared error, and the numerical preintegration mean squared error, respectively, without the prefactor 3.

We first analyze E_1 , which corresponds to the weak error/bias of the nested estimator. According to Theorem 1, we have

$$E_1 = O(m_L^{-2}). \quad (11)$$

For simplification, we assume that on all levels ($0 \leq \ell \leq L$) $M_{\text{Lag},\ell} = M_{\text{Lag},L}$ and $\text{TOL}_{\text{Newton},\ell} = \text{TOL}_{\text{Newton},L}$. With the convention $\hat{h}_{-1} = \bar{h}_{-1} = 0$, we estimate E_3 as follows:

$$E_3 = \mathbb{E} \left(\hat{Q} - \bar{Q} \right)^2 = \mathbb{E} \left(\sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{n_\ell=1}^{N_\ell} \left((\hat{h}_\ell - \hat{h}_{\ell-1}) - (\bar{h}_\ell - \bar{h}_{\ell-1}) \right) \right)^2$$

$$\begin{aligned}
&\leq (L+1) \sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{n_\ell=1}^{N_\ell} \mathbb{E} \left((\hat{h}_\ell - \hat{h}_{\ell-1}) - (\bar{h}_\ell - \bar{h}_{\ell-1}) \right)^2 \\
&\leq 2(L+1) \sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{n_\ell=1}^{N_\ell} \left(\mathbb{E}(\hat{h}_\ell - \bar{h}_\ell)^2 + \mathbb{E}(\hat{h}_{\ell-1} - \bar{h}_{\ell-1})^2 \right) \\
&= O(M_{\text{Lag},L}^{-s}) + O(\text{TOL}_{\text{Newton},L}^2), \tag{12}
\end{aligned}$$

where $s > 0$ is related to the regularity of the integrand. See [Bayer et al. \(2023, 2024\)](#) for details of the error bound in (12).

E_2 represents the statistical error of the MLMC estimator, which is determined by the sample variance and the number of samples at each level. From the standard multilevel error analysis, we obtain

$$E_2 = \sum_{\ell=0}^L (N_\ell)^{-1} \hat{V}_\ell, \tag{13}$$

where

$$\hat{V}_0 := \text{Var}[\hat{Y}_0], \quad \hat{V}_\ell := \text{Var}[\hat{Y}_\ell], \quad 1 \leq \ell \leq L.$$

Let C_ℓ be the cost per sample per level, given by

$$C_\ell \propto m_\ell(M_{\text{Lag},\ell} + N_{\text{iter},\ell}) \propto m_\ell \left(M_{\text{Lag},\ell} + \log \left(\text{TOL}_{\text{Newton},\ell}^{-1} \right) \right), \quad 0 \leq \ell \leq L, \tag{14}$$

where $N_{\text{iter},\ell}$ is the number of the Newton iterations at level ℓ . Under some mild conditions and by using Taylor expansion, we can show that Newton iteration has a second order convergence and we conclude that $N_{\text{iter},\ell} \propto \log \left(\text{TOL}_{\text{Newton},\ell}^{-1} \right)$ ([Bayer et al. 2024](#)). The variance sequence $\{\hat{V}_\ell\}$ is the core parameter for MLMC sample allocation, and the next section establishes its decay rate. Theorem 2 gives estimates of the variances $\{\hat{V}_\ell\}_{\ell=0}^L$, and shows that $\hat{V}_\ell = O(m_\ell^{-1})$.

Finally, using (10)–(13), the total MSE estimate of the estimator $\bar{Q}$ is

$$\begin{aligned}
&\mathbb{E}(\mathbb{E}[g(\varphi(\boldsymbol{\omega}))] - \bar{Q})^2 \\
&= O(m_L^{-2}) + O \left(\sum_{\ell=0}^L (N_\ell)^{-1} \hat{V}_\ell \right) + O(M_{\text{Lag},L}^{-s}) + O(\text{TOL}_{\text{Newton},L}^2). \tag{15}
\end{aligned}$$

3.2 Strong convergence results for MLMC with preintegration

In this subsection, we derive the strong convergence rate of our estimator based on the MLMC method with the preintegration technique. Before proving our main theorem, we require the following definition and assumptions. We first introduce the following definition for uniformly bounded random variables.

Definition 1 For sequences of random variables F_N , we write that $F_N = \mathcal{O}(1)$ if there exists a random variable C with finite moments of all orders, such that for all N , we have $|F_N| \leq C$ a.s.

We impose the following regularity assumptions on the conditional expectation function and its derivatives. We use the notation $\partial_{\omega_1}^d$ with some positive integer d to denote the d -th partial derivative with respect to ω_1 in what follows.

Assumption 1 Suppose that the function $\varphi(\omega) = \mathbb{E}[X|\omega]$ in (1) satisfies

$$\varphi(\omega_1, \omega_{-1}) = \mathcal{O}(1), \quad \partial_{\omega_1} \varphi(\omega_1, \omega_{-1}) = \mathcal{O}(1) \quad \text{and} \quad \partial_{\omega_1}^2 \varphi(\omega_1, \omega_{-1}) = \mathcal{O}(1).$$

Assume additionally that the pairs $(X_j(\omega), \partial_{\omega_1} X_j(\omega))$ are conditionally i.i.d. given ω , with derivatives taken at fixed inner random inputs independent of ω , and satisfy

$$\begin{aligned} \mathbb{E}[\partial_{\omega_1} X_1(\omega) | \omega] &= \partial_{\omega_1} \varphi(\omega), \\ \mathbb{E}[|X_1(\omega)|^q + |\partial_{\omega_1} X_1(\omega)|^q] &< \infty \quad \text{for every } q \geq 2. \end{aligned}$$

Assumption 2 Let $\hat{\varphi}_\ell = \hat{\varphi}_{m_\ell}$ be defined in (2). With the inner random inputs fixed, assume that, almost surely, each function $\hat{\varphi}_\ell(\cdot, \omega_{-1})$ is continuously differentiable on every connected component of $\{\rho_1 > 0\}$, with a locally absolutely continuous first derivative. On each component, these derivatives are nonzero and have the same sign for all $\ell \geq 0$. Furthermore, assume that

$$(\partial_{\omega_1} \hat{\varphi}_\ell(\omega))^{-1} = \mathcal{O}(1)$$

and, for some $r_* > 4$,

$$\sup_{\ell \geq 0} \mathbb{E}[|\partial_{\omega_1}^2 \hat{\varphi}_\ell(\omega)|^{r_*}] < \infty.$$

The second derivatives are interpreted in the almost-everywhere sense.

We also assume that the density of ω_1 satisfies the following assumption.

Assumption 3 Suppose that ρ_1 is locally absolutely continuous on $\mathbb{R}$. Define $s(y) := \rho_1'(y)/\rho_1(y)$ almost everywhere on $\{y : \rho_1(y) > 0\}$ and set $s(y) = 0$ where $\rho_1(y) = 0$. Assume that there exists $p_* > 1$ such that

$$\int_{\mathbb{R}} |s(y)|^{p_*} \rho_1(y) dy < \infty.$$

The following classical moment inequality (Giles and Goda 2019) for sample averages is essential for our error estimation.

Lemma 1 Let X be a random variable with zero mean, and let $\bar{X}_N$ be an average of N i.i.d. samples of X . If $\mathbb{E}[|X|^p]$ is finite for $p \geq 2$, there exists a constant C_p depending only on p such that

$$\mathbb{E}[|\bar{X}_N|^p] \leq C_p \frac{\mathbb{E}[|X|^p]}{N^{p/2}}.$$

Based on the above assumptions and the moment inequality in Lemma 1, we derive the moment bounds for the error of level difference.

Lemma 2 Suppose that Assumption 1 is satisfied and we define

$$e_\ell(\omega_1, \omega_{-1}) := \hat{\varphi}_\ell(\omega_1, \omega_{-1}) - \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}).$$

For every $p \geq 1$ and $\ell \geq 1$, we have

$$\mathbb{E}[|e_\ell(\omega)|^{2p}] = O(m_\ell^{-p}) \quad (16)$$

and

$$\mathbb{E}[|(\partial_{\omega_1} e_\ell)(\omega)|^{2p}] = O(m_\ell^{-p}). \quad (17)$$

Proof Under Assumption 1, we have $\mathbb{E}[|\varphi(\omega)|^p] < \infty$ and $\mathbb{E}[|\partial_{\omega_1} \varphi(\omega)|^p] < \infty$ for every $p \geq 1$. From the definition of e_ℓ , we have

$$\begin{aligned} \mathbb{E}[|e_\ell(\omega)|^{2p}] &= \mathbb{E}[|\hat{\varphi}_\ell(\omega) - \hat{\varphi}_{\ell-1}(\omega)|^{2p}] \\ &= \mathbb{E}[|(\hat{\varphi}_\ell(\omega) - \varphi(\omega)) - (\hat{\varphi}_{\ell-1}(\omega) - \varphi(\omega))|^{2p}] \\ &\leq 2^{2p-1} \mathbb{E}[|\hat{\varphi}_\ell(\omega) - \varphi(\omega)|^{2p}] + 2^{2p-1} \mathbb{E}[|\hat{\varphi}_{\ell-1}(\omega) - \varphi(\omega)|^{2p}]. \end{aligned}$$

Since $\hat{\varphi}_\ell(\omega)$ is conditionally unbiased given ω , applying Lemma 1 conditionally with exponent $2p \geq 2$ gives

$$\begin{aligned} \mathbb{E}[|\hat{\varphi}_\ell(\omega) - \varphi(\omega)|^{2p} | \omega] &\leq C \frac{\mathbb{E}[|X_1(\omega) - \varphi(\omega)|^{2p} | \omega]}{m_\ell^p} \\ &\leq 2^{2p} C \frac{\mathbb{E}[|X_1(\omega)|^{2p} | \omega]}{m_\ell^p}. \end{aligned}$$

Taking expectations and using the moment condition in Assumption 1, we obtain

$$\mathbb{E}[|\hat{\varphi}_\ell(\omega) - \varphi(\omega)|^{2p}] \leq \frac{2^{2p} C}{m_\ell^p} \mathbb{E}[|X_1(\omega)|^{2p}] = O(m_\ell^{-p}).$$

Thus

$$\mathbb{E}[|e_\ell(\omega)|^{2p}] = O(m_\ell^{-p}).$$

Similarly, we can obtain that

$$\mathbb{E}[|(\partial_{\omega_1} e_\ell)(\omega)|^{2p}] = O(m_\ell^{-p}).$$

□

We now present the theorem which establishes the optimal variance decay rate for the MLMC estimator with smoothing. To obtain the result of Theorem 2, we follow the proof strategy of Bayer et al. (2024). By incorporating preintegration, we improve the strong convergence rate to -1 , compared with $-1/2$ for standard MLMC (Giles and Haji-Ali 2019).

Theorem 2 Let the function g, φ be defined in (1) and $\hat{h}_\ell$ be defined in (4). Then under Assumptions 1, 2 and 3, we have

$$\hat{V}_\ell = O(m_\ell^{-1}).$$

Proof First we observe that $\hat{V}_\ell = \text{Var} [\hat{h}_\ell - \hat{h}_{\ell-1}] \leq \mathbb{E} \left[\left(\hat{h}_\ell - \hat{h}_{\ell-1} \right)^2 \right]$. Therefore, we only need to prove that $\mathbb{E} \left[\left(\hat{h}_\ell - \hat{h}_{\ell-1} \right)^2 \right] = O(m_\ell^{-1})$. For $\delta > 0$, let $\tilde{g}_\delta$ denote a C^∞ mollified version of g (i.e., obtained by convoluting g with a mollifier), then we have

$$\begin{aligned}
\Delta \hat{h}_\ell^\delta(\omega_{-1}) &:= (\hat{h}_\ell^\delta - \hat{h}_{\ell-1}^\delta)(\omega_{-1}) \\
&:= \int_{\mathbb{R}} (\tilde{g}_\delta(\hat{\varphi}_\ell(\omega_1, \omega_{-1})) - \tilde{g}_\delta(\hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}))) \rho_1(\omega_1) d\omega_1 \\
&= \int_{\mathbb{R}} \left[\int_0^1 \tilde{g}'_\delta \left(\underbrace{\hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}) + \theta e_\ell(\omega_1, \omega_{-1})}_{:= z(\theta; \omega_1, \omega_{-1})} \right) d\theta \right] e_\ell(\omega_1, \omega_{-1}) \rho_1(\omega_1) d\omega_1 \\
&= \int_{\mathbb{R}} \left[\int_0^1 \partial_{\omega_1} \tilde{g}_\delta(z(\theta; \omega_1, \omega_{-1})) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} d\theta \right] e_\ell(\omega_1, \omega_{-1}) \rho_1(\omega_1) d\omega_1 \\
&= \int_0^1 \left[\int_{\mathbb{R}} \partial_{\omega_1} \tilde{g}_\delta(z(\theta; \omega_1, \omega_{-1})) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} e_\ell(\omega_1, \omega_{-1}) \rho_1(\omega_1) d\omega_1 \right] d\theta
\end{aligned} \tag{18}$$

where the fourth equality uses $\partial_{\omega_1} \tilde{g}_\delta = \tilde{g}'_\delta \partial_{\omega_1} z$, and the last equality follows from Fubini's theorem, then we have

$$\begin{aligned}
\Delta \hat{h}_\ell^\delta(\omega_{-1}) &= - \int_0^1 \left[\int_{\mathbb{R}} \tilde{g}_\delta(z(\theta; \omega_1, \omega_{-1})) \partial_{\omega_1} \left((\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} e_\ell(\omega_1, \omega_{-1}) \rho_1(\omega_1) \right) d\omega_1 \right] d\theta \\
&= - \int_0^1 \left[\int_{\mathbb{R}} \tilde{g}_\delta(z(\theta; \omega_1, \omega_{-1})) \left(e_\ell(\omega_1, \omega_{-1}) \partial_{\omega_1} \left((\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \rho_1(\omega_1) \right) \right. \right. \\
&\quad \left. \left. + (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \rho_1(\omega_1) \partial_{\omega_1} e_\ell(\omega_1, \omega_{-1}) \right) d\omega_1 \right] d\theta \\
&= - \int_0^1 \left[\int_{\mathbb{R}} e_\ell(\omega_1, \omega_{-1}) \tilde{g}_\delta(z(\theta; \omega_1, \omega_{-1})) \right. \\
&\quad \left. \left(\partial_{\omega_1} \left((\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \right) + s(\omega_1) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \right) \rho_1(\omega_1) d\omega_1 \right] d\theta \\
&\quad - \int_0^1 \left[\int_{\mathbb{R}} \partial_{\omega_1} e_\ell(\omega_1, \omega_{-1}) \tilde{g}_\delta(z(\theta; \omega_1, \omega_{-1})) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \rho_1(\omega_1) d\omega_1 \right] d\theta,
\end{aligned} \tag{19}$$

where s is defined in Assumption 3, and the boundary terms of the first equality vanish due to Lemma 3 in Appendix A.

Taking $\delta \rightarrow 0$ and applying the dominated convergence theorem to (19), we obtain

$$\begin{aligned}
\Delta \hat{h}_\ell(\omega_{-1}) &:= (\hat{h}_\ell - \hat{h}_{\ell-1})(\omega_{-1}) \\
&= - \underbrace{\int_0^1 \left[\int_{\mathbb{R}} e_\ell(\omega_1, \omega_{-1}) g(z(\theta; \omega_1, \omega_{-1})) \right.}_{(I)} \\
&\quad \left. \underbrace{\left(\partial_{\omega_1} \left((\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \right) + s(\omega_1) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} \right) \rho_1(\omega_1) d\omega_1 \right] d\theta}_{(I)}
\end{aligned}$$

$$\underbrace{-\int_0^1 \left[\int_{\mathbb{R}} \partial_{\omega_1} e_{\ell}(\omega_1, \omega_{-1}) g(z(\cdot; \cdot, \omega_{-1})) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \rho_1(\omega_1) d\omega_1 \right] d\theta}_{(II)} \quad (20)$$

$$:= T_1 + T_2 \quad (21)$$

where we denote the first term in (20) as T_1 , and the second term in (20) as T_2 .

For the term (I) in (21), taking expectation and using Hölder's inequality twice ($p, q, p_1, q_1 \in (1, +\infty)$, $\frac{1}{p} + \frac{1}{q} = 1$ and $\frac{1}{p_1} + \frac{1}{q_1} = 1$), we have

$$\begin{aligned} \mathbb{E} [T_1^2] &\leq \mathbb{E} \left[\left\| g(z(\cdot; \cdot, \omega_{-1})) \left(\partial_{\omega_1} \left((\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) + s(\cdot) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) \right\|_{L_{\rho_1}^q([0,1] \times \mathbb{R})}^2 \right. \\ &\quad \times \left. \|e_{\ell}(\cdot, \omega_{-1})\|_{L_{\rho_1}^p(\mathbb{R})}^2 \right] \\ &\leq \left(\mathbb{E} \left[\left\| g(z(\cdot; \cdot, \omega_{-1})) \left(\partial_{\omega_1} \left((\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) + s(\cdot) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) \right\|_{L_{\rho_1}^q([0,1] \times \mathbb{R})}^{2q_1} \right] \right)^{1/q_1} \\ &\quad \times \left(\mathbb{E} \left[\|e_{\ell}(\cdot, \omega_{-1})\|_{L_{\rho_1}^p(\mathbb{R})}^{2p_1} \right] \right)^{1/p_1}. \end{aligned} \quad (22)$$

Choose the exponents above such that $q < p_*$, $p \geq 4p_1$, and $4q_1 < r_*$; such a choice is possible since $p_* > 1$ and $r_* > 4$. The first factor in the right-hand side of (22) is bounded, because we first observe that

$$(\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} = (\partial_{\omega_1} \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}))^{-1} \left((1-\theta) + \theta \frac{\partial_{\omega_1} \hat{\varphi}_{\ell}(\omega_1, \omega_{-1})}{\partial_{\omega_1} \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1})} \right)^{-1}, \quad (23)$$

and

$$\begin{aligned} &\partial_{\omega_1} \left((\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) \\ &= -\partial_{\omega_1}^2 z(\cdot; \cdot, \omega_{-1}) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-2} \\ &= - \left((1-\theta) \partial_{\omega_1}^2 \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}) + \theta \partial_{\omega_1}^2 \hat{\varphi}_{\ell}(\omega_1, \omega_{-1}) \right) (\partial_{\omega_1} \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}))^{-2} \\ &\quad \times \left((1-\theta) + \theta \frac{\partial_{\omega_1} \hat{\varphi}_{\ell}(\omega_1, \omega_{-1})}{\partial_{\omega_1} \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1})} \right)^{-2}. \end{aligned}$$

Then, using Assumption 2, we obtain that $(\partial_{\omega_1} \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}))^{-1} = \mathcal{O}(1)$ and $(\partial_{\omega_1} \hat{\varphi}_{\ell-1}(\omega_1, \omega_{-1}))^{-2} = \mathcal{O}(1)$. Moreover, Assumption 2 gives uniformly bounded r_* -th moments of $\partial_{\omega_1}^2 \hat{\varphi}_{\ell}$. Together with the common derivative sign, the inverse-derivative bound, Assumption 3, and the chosen Hölder exponents, this yields the following bound uniformly in ℓ :

$$\left(\mathbb{E} \left[\left\| g(z(\cdot; \cdot, \omega_{-1})) \left(\partial_{\omega_1} \left((\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) + s(\cdot) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right) \right\|_{L_{\rho_1}^q([0,1] \times \mathbb{R})}^{2q_1} \right] \right)^{1/q_1} < \infty. \quad (24)$$

Choosing p and p_1 such that $\frac{2p_1}{p} \leq 1$, and applying Jensen's inequality for the second factor in the right-hand side of (22), then using (16) in Lemma 2, we obtain

$$\left(\mathbb{E} \left[\|e_{\ell}(\cdot, \omega_{-1})\|_{L_{\rho_1}^p(\mathbb{R})}^{2p_1} \right] \right)^{1/p_1} = \left(\mathbb{E} \left[\left(\int_{\mathbb{R}} |e_{\ell}(\omega_1, \omega_{-1})|^p \rho_1 d\omega_1 \right)^{\frac{2p_1}{p}} \right] \right)^{1/p_1}$$

$$\begin{aligned}
&\leq \left(\mathbb{E} \left[\int_{\mathbb{R}} |e_{\ell}(\omega_1, \omega_{-1})|^p \rho_1 d\omega_1 \right] \right)^{\frac{2}{p}} \\
&= O(m_{\ell}^{-1}).
\end{aligned} \tag{25}$$

Consequently, combined with (24) and (25) we conclude that $\mathbb{E}[T_1^2] = O(m_{\ell}^{-1})$.

For the term (II) in (21), taking expectation and using Hölder's inequality twice ($p, q, p_1, q_1 \in (1, +\infty)$, $\frac{1}{p} + \frac{1}{q} = 1$ and $\frac{1}{p_1} + \frac{1}{q_1} = 1$), we have

$$\begin{aligned}
\mathbb{E}[T_2^2] &\leq \mathbb{E} \left[\left\| g(z(\cdot; \cdot, \omega_{-1})) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right\|_{L_{\rho_1}^q([0,1] \times \mathbb{R})}^2 \times \|\partial_{\omega_1} e_{\ell}(\cdot, \omega_{-1})\|_{L_{\rho_1}^{p_1}(\mathbb{R})}^2 \right] \\
&\leq \left(\mathbb{E} \left[\left\| g(z(\cdot; \cdot, \omega_{-1})) (\partial_{\omega_1} z(\cdot; \cdot, \omega_{-1}))^{-1} \right\|_{L_{\rho_1}^q([0,1] \times \mathbb{R})}^{2q_1} \right] \right)^{1/q_1} \times \left(\mathbb{E} \left[\|\partial_{\omega_1} e_{\ell}(\cdot, \omega_{-1})\|_{L_{\rho_1}^{p_1}(\mathbb{R})}^{2p_1} \right] \right)^{1/p_1}.
\end{aligned} \tag{26}$$

Similarly to (25) and using (17) in Lemma 2, we obtain that

$$\left(\mathbb{E} \left[\|\partial_{\omega_1} e_{\ell}(\cdot, \omega_{-1})\|_{L_{\rho_1}^{p_1}(\mathbb{R})}^{2p_1} \right] \right)^{1/p_1} = O(m_{\ell}^{-1}).$$

Moreover, similar to the prove of (25), we get the first term in the right-hand side of (26) to be bounded. This concludes that $\mathbb{E}[T_2^2] = O(m_{\ell}^{-1})$ and consequently finishes the proof. $\square$

Remark 1 We can reduce the variance by using antithetic sampling (Bujok et al. 2015; Giles and Goda 2019; Giles and Haji-Ali 2019), a common technique in the MLMC method. However, when the outer function is discontinuous, such as the indicator function in (1), antithetic sampling usually yields little improvement. Antithetic sampling generally achieves satisfactory variance reduction only for continuous functions, which aligns well with our proposed smoothed MLMC method. We construct the following antithetic form by replacing $\hat{Y}_{\ell} = \hat{h}_{\ell} - \hat{h}_{\ell-1}$ for coupling the consecutive levels

$$\hat{Y}_{\ell} = \hat{h}_{\ell} - \frac{1}{2}(\hat{h}_{\ell-1}^{(1)} + \hat{h}_{\ell-1}^{(2)}) \tag{27}$$

where

$$\hat{h}_{\ell-1}^{(i)}(\omega_{-1}) = \int_{\mathbb{R}} \hat{g}_{\ell-1}^{(i)}(\omega_1, \omega_{-1}) \rho_1(\omega_1) d\omega_1, \hat{g}_{\ell-1}^{(i)}(\mathbf{y}) = \mathbb{I}\{\hat{\varphi}_{\ell-1}^{(i)}(\mathbf{y}) > c\}$$

and

$$\hat{\varphi}_{\ell-1}^{(i)}(\mathbf{y}) = \frac{1}{m_{\ell-1}} \sum_{j=1+(i-1)m_{\ell-1}}^{im_{\ell-1}} X_j(\mathbf{y}), \quad i = 1, 2.$$

We can prove that the strong convergence rate of antithetic sampling remains $O(m_{\ell}^{-1})$, which can be shown by decomposing (18) into two parts. Although antithetic sampling does not alter our theoretical convergence order, it exhibits favorable performance in the numerical experiments presented in Section 4.

3.3 Complexity analysis

To achieve a root mean square error (RMSE) for certain tolerance TOL with minimal computational cost, we formulate the following constrained optimization problem for

the MLMC parameters. One needs to solve (28) using (15)

$$\begin{cases} \min_{(L, \{N_\ell\}_{\ell=0}^L, \{M_{\text{Lag}, \ell}\}_{\ell=0}^L, \{\text{TOL}_{\text{Newton}, \ell}\}_{\ell=0}^L)} \mathcal{C}(L, \{N_\ell\}_{\ell=0}^L, \{M_{\text{Lag}, \ell}\}_{\ell=0}^L, \{\text{TOL}_{\text{Newton}, \ell}\}_{\ell=0}^L) \\ \text{s.t. } \mathbb{E}[(\bar{Q} - \theta)^2] = \text{TOL}^2 \end{cases} \quad (28)$$

where $\mathcal{C}(L, \{N_\ell\}_{\ell=0}^L, \{M_{\text{Lag}, \ell}\}_{\ell=0}^L, \{\text{TOL}_{\text{Newton}, \ell}\}_{\ell=0}^L)$ is the total computational complexity of our method. From (14), we obtain the total computational complexity of our approach as follows

$$\begin{aligned} \mathcal{C}(L, \{N_\ell\}_{\ell=0}^L, \{M_{\text{Lag}, \ell}\}_{\ell=0}^L, \{\text{TOL}_{\text{Newton}, \ell}\}_{\ell=0}^L) &\propto \sum_{\ell=0}^L N_\ell C_\ell \\ &\propto \sum_{\ell=0}^L N_\ell m_\ell \left(M_{\text{Lag}, \ell} + \log \left(\text{TOL}_{\text{Newton}, \ell}^{-1} \right) \right). \end{aligned} \quad (29)$$

In this work, we do not solve (28); instead, we select the different parameters heuristically. A further investigation of optimizing (28) is left for a future study.

The following corollary summarizes the computational complexity of the MLMC with preintegration and MLMC with numerical preintegration.

Remark 2 Since we have improved the strong convergence rate from $-1/2$ to -1 by incorporating the preintegration, our MLMC with preintegration also achieves a sub-optimal computational complexity. Using numerical preintegration introduces additional computational cost, but the constant $s \gg 1$ in the $O(\text{TOL}^{-2-2/s} (\log(\text{TOL}))^2)$ usually makes this extra cost negligible (Bayer et al. 2024).

Corollary 1 *Under the Assumptions of Theorem 1 and Theorem 2, for any $0 < \text{TOL} < 1$, the complexity of MLMC with numerical preintegration is $O(\text{TOL}^{-2-2/s} (\log(\text{TOL}))^2)$. Moreover, the complexity of MLMC with preintegration is $O(\text{TOL}^{-2} (\log(\text{TOL}))^2)$.*

Proof We recall the error decomposition in (10). First, to ensure $E_2 \leq \text{TOL}^2/9$, we choose

$$N_\ell \leq C \text{TOL}^{-2} (L+1) m_\ell^{-1} + 1, \quad (30)$$

where C is a sufficiently large constant. To ensure $E_1 \leq \text{TOL}^2/9$, we obtain

$$L+1 = O(\log(\text{TOL}^{-1})),$$

and due to $0 < \text{TOL} < 1$, we have

$$\sum_{\ell=0}^L m_\ell = O(\text{TOL}^{-2}).$$

Moreover, to ensure $E_3 \leq \text{TOL}^2/9$, using (12), we obtain

$$M_{\text{Lag}, L} = O(\text{TOL}^{-2/s}),$$

$$N_{\text{iter},L} = \log \left(\text{TOL}_{\text{Newton},L}^{-1} \right) = O(\log \left(\text{TOL}^{-1} \right)).$$

For simplification, we assume that on all levels $0 \leq \ell \leq L$, $M_{\text{Lag},\ell} = M_{\text{Lag},L}$ and $\text{TOL}_{\text{Newton},\ell} = \text{TOL}_{\text{Newton},L}$. Then, by using (29) and (30), we have the computational complexity of our MLMC estimator with numerical smoothing is

$$\begin{aligned} \sum_{\ell=0}^L N_{\ell} C_{\ell} &\propto \sum_{\ell=0}^L N_{\ell} m_{\ell} \left(M_{\text{Lag},\ell} + \log \left(\text{TOL}_{\text{Newton},\ell}^{-1} \right) \right) \\ &\leq C \text{TOL}^{-2} (L+1)^2 \left(M_{\text{Lag},L} + \log \left(\text{TOL}_{\text{Newton},L}^{-1} \right) \right) \\ &\quad + \left(M_{\text{Lag},L} + \log \left(\text{TOL}_{\text{Newton},L}^{-1} \right) \right) \sum_{\ell=0}^L m_{\ell} \\ &= O(\text{TOL}^{-2-(2/s)} (\log(\text{TOL}))^2). \end{aligned}$$

□

3.4 Robustness analysis

Due to the discontinuity of the indicator function g in (1), the standard MLMC method exhibits a high kurtosis phenomenon, which adversely affects its numerical stability and robustness. We define $Y_{\ell} := \hat{g}_{m_{\ell}} - \hat{g}_{m_{\ell-1}}$. As shown in Ben Hammouda et al. (2020), the standard deviation of the sample variance for Y_{ℓ} is given by

$$\sigma_{\ell} := \frac{\text{Var}[Y_{\ell}]}{\sqrt{N_{\ell}}} \sqrt{(\kappa_{\ell} - 1) + \frac{2}{N_{\ell} - 1}},$$

where κ_{ℓ} is the kurtosis at level ℓ , given by

$$\kappa_{\ell} := \frac{\mathbb{E}(Y_{\ell} - \mathbb{E}Y_{\ell})^4}{(\text{Var}[Y_{\ell}])^2}.$$

High kurtosis makes it challenging to estimate V_{ℓ} accurately in deeper levels, since there are less samples in deeper levels under the MLMC scheme (Ben Hammouda et al. 2020). It should be noted that the high-kurtosis phenomenon is due to the structure of the indicator function. The smoothed MLMC method we proposed earlier can precisely solve this problem. We define the kurtosis of the smoothed MLMC estimator as

$$\hat{\kappa}_{\ell} := \frac{\mathbb{E}(\hat{Y}_{\ell} - \mathbb{E}\hat{Y}_{\ell})^4}{(\text{Var}[\hat{Y}_{\ell}])^2}, \quad (31)$$

where $\hat{Y}_{\ell}$ is defined in (7).

Corollary 2 *Let $\hat{\kappa}_{\ell}$ in (31) be the kurtosis of $\hat{Y}_{\ell}$, assume that $\hat{V}_{\ell}^{-1} = O(m_{\ell})$. Then under the same assumption of Theorem 2, we have*

$$\hat{\kappa}_{\ell} = O(1).$$

Proof Similar to the proof of Theorem 2, we can obtain $\mathbb{E}(\hat{Y}_\ell - \mathbb{E}\hat{Y}_\ell)^4 = O(m_\ell^{-2})$. Then, combined with $\hat{V}_\ell^{-1} = O(m_\ell)$, we get $\hat{\kappa}_\ell = O(1)$. $\square$

Remark 3 In our previous analysis, we only considered the upper bound of $\hat{V}_\ell$, for which we conducted theoretical analysis and derived the upper bound $O(m_\ell^{-1})$. However, to obtain the kurtosis of $\hat{Y}_\ell$, we need a lower bound of $\hat{V}_\ell$, which is rather difficult in theoretical analysis. In our numerical experiments, we observed that the convergence rate of $\hat{V}_\ell$ is roughly on par with -1 and does not converge faster. Therefore, in Corollary 2, we assume that $\hat{V}_\ell$ has a lower bound, i.e., $\hat{V}_\ell^{-1} = O(m_\ell)$.

4 Numerical experiments

In this section, several numerical experiments are carried out to investigate the performance of the proposed method. We focus on a portfolio composed of European options that are written on d underlying stocks, with the price dynamics governed by the Black-Scholes model. For simplicity, we assume that all stock returns are identical and are denoted by μ , whereas the risk-free interest rate is represented by μ_0 . The price dynamics of the stock $\mathbf{S}_t = (S_t^1, \dots, S_t^d)$ is described by the following stochastic differential equation

$$\frac{dS_t^i}{S_t^i} = \mu' dt + \sum_{j=1}^d \sigma_{ij} dB_t^j, \quad i = 1, \dots, d, \quad (32)$$

where $\mu' = \mu$ under the real-world probability measure, and $\mu' = \mu_0$ under the risk-neutral probability measure. Here, $\mathbf{B}_t = (B_t^1, \dots, B_t^d)$ is a d -dimensional standard Brownian motion which represents d risk factors in the model and σ_{ij} are the corresponding volatility. The solution of (32) is analytically available

$$S_t^i = S_0^i \exp\left\{(\mu' - \frac{1}{2} \sum_{j=1}^d \sigma_{ij}^2)t + \sum_{j=1}^d \sigma_{ij} B_t^j\right\}, \quad i = 1, \dots, d,$$

where $\mathbf{S}_0 = (S_0^1, \dots, S_0^d)$ are the initial prices of stocks.

We assume that the maturities of all the European options in the portfolio are the same, and are denoted by T . We want to measure the portfolio risk at a future time τ ($\tau < T$). In the simulation, we first simulate the random variable $\boldsymbol{\omega} = \mathbf{S}_\tau = (S_\tau^1, \dots, S_\tau^d)$ under real-world probability measure, which denotes the prices of stocks at the risk horizon τ as the outer sample. And then simulate $\mathbf{S}_T = (S_T^1, \dots, S_T^d)$ under the risk-neutral probability measure which is the prices of stocks at maturity T given $\boldsymbol{\omega}$ as the inner samples. Let $V_0 = \sum_{i=1}^d v_0^i$ denote the initial value of the portfolio, where each v_0^i can be calculated using the Black-Scholes pricing formula (Hull and Basu 2016). The change in the portfolio's value is

$$\varphi(\boldsymbol{\omega}) := V_0 - \mathbb{E}[V_T(\mathbf{S}_T)|\boldsymbol{\omega}],$$

where $V_T(\mathbf{S}_T)$ is the discounted payoff of the portfolio at time T which is a known function of $\mathbf{S}_T$.

The target of our numerical experiments is to estimate the probability of loss $\theta = \mathbb{P}(\varphi(\boldsymbol{\omega}) > c)$ for a given threshold c . The number of inner samples is set as $m_\ell = 32 \times 2^\ell$ in all experiments.

4.1 Multiple assets

Now we consider a portfolio consists of d European call options, which was studied in [Hong et al. \(2017\)](#). Given the outer sample $\boldsymbol{\omega} = (\omega_1, \dots, \omega_d) = \mathbf{S}_\tau = (S_\tau^1, \dots, S_\tau^d)$ which denotes the prices of stocks at the risk horizon τ under real-world measure

$$\omega_i = S_\tau^i = S_0^i \exp\left\{\left(\mu - \frac{1}{2} \sum_{j=1}^d \sigma_{ij}^2\right)\tau + \sum_{j=1}^d \sigma_{ij} \sqrt{\tau} Z_j\right\}, \quad i = 1, \dots, d,$$

where $\mathbf{Z} = (Z_1, \dots, Z_d) \sim N(\mathbf{0}, \mathbf{I}_d)$ and $\mathbf{K} = (K_1, \dots, K_d)$, the strike prices for call options, the portfolio value change is

$$\varphi(\boldsymbol{\omega}) = V_0 - \mathbb{E}[e^{-\mu_0(T-\tau)} \sum_{i=1}^d (S_T^i(\omega_i, \mathbf{W}) - K_i)^+ | \boldsymbol{\omega}],$$

the expectation is taken over the random variable $\mathbf{W} = (W^1, \dots, W^d) \sim N(\mathbf{0}, \mathbf{I}_d)$, and samples of

$$S_T^i(\omega_i, \mathbf{W}) = \omega_i \exp\left\{\left(\mu_0 - \frac{1}{2} \sum_{j=1}^d \sigma_{ij}^2\right)(T-\tau) + \sum_{j=1}^d \sigma_{ij} \sqrt{T-\tau} W^j\right\}, \quad i = 1, \dots, d,$$

are simulated under the risk-neutral measure.

Next, we provide the analytical expression for the preintegration of the loss function with respect to ω_1 . For simplicity, we suppose that $\sigma_{1j} = \sigma_{i1} = 0$ for $1 < i, j \leq d$. We rewrite the function $\varphi(\boldsymbol{\omega})$ as

$$\begin{aligned} \varphi(\boldsymbol{\omega}) &= -\mathbb{E}[e^{-\mu_0(T-\tau)} (S_T^1(\omega_1, \mathbf{W}) - K_1)^+ | \omega_1] \\ &\quad + V_0 - \mathbb{E}[e^{-\mu_0(T-\tau)} \sum_{i=2}^d (S_T^i(\omega_i, \mathbf{W}) - K_i)^+ | \boldsymbol{\omega}_{-1}] \\ &:= -\varphi_1(\omega_1) + \varphi_{-1}(\boldsymbol{\omega}_{-1}), \end{aligned}$$

where

$$\omega_1 = S_\tau^1 = S_0^1 \exp\left\{\left(\mu - \frac{1}{2} \sigma_{11}^2\right)\tau + \sigma_{11} \sqrt{\tau} Z_1\right\},$$

and

$$\omega_i = S_\tau^i = S_0^i \exp\left\{\left(\mu - \frac{1}{2} \sum_{j=2}^d \sigma_{ij}^2\right)\tau + \sum_{j=2}^d \sigma_{ij} \sqrt{\tau} Z_j\right\}, \quad i = 2, \dots, d.$$

According to the Black-Scholes model European call option formula

$$\varphi_1(\omega_1) = \omega_1 \Phi(d_1) - K_1 e^{-\mu_0(T-\tau)} \Phi(d_2),$$

where

$$d_1 = \frac{\log \frac{\omega_1}{K_1} + (\mu_0 + \frac{\sigma_{11}^2}{2})(T-\tau)}{\sigma_{11}\sqrt{T-\tau}}, d_2 = d_1 - \sigma_{11}\sqrt{T-\tau}.$$

φ_1 is the function monotonically increasing, because

$$\partial_{\omega_1} \varphi_1(\omega_1) = \Phi(d_1) > 0,$$

where $\partial_{\omega_1} \varphi_1$ is the Delta of φ_1 . Due to monotonicity, there exists a function ψ_d such that $\{\varphi(\boldsymbol{\omega}) > c\} = \{\varphi(\psi_d(\boldsymbol{\omega}_{-1}), \boldsymbol{\omega}_{-1}) > c\} = \{\omega_1 \leq \psi_d\}$. Then we have

$$\begin{aligned} \mathbb{E}[\mathbb{I}\{\varphi(\boldsymbol{\omega}) > c\} | \boldsymbol{\omega}_{-1}] &= \int_{\mathbb{R}} \mathbb{I}\{\varphi_1(\omega_1) \leq \varphi_{-1}(\boldsymbol{\omega}_{-1}) - c\} \rho_1(\omega_1) d\omega_1 \\ &= \int_{\mathbb{R}} \mathbb{I}\{\omega_1 \leq \psi_d\} \rho_1(\omega_1) d\omega_1 \\ &= \Phi\left(\frac{\log \frac{\psi_d}{S_0^1} - (\mu - \frac{\sigma_{11}^2}{2})\tau}{\sigma_{11}\sqrt{\tau}}\right). \end{aligned}$$

In our numerical experiments, we use inner simulation to approximate φ_{-1} , with the level- ℓ approximation denoted by $\hat{\varphi}_{-1}^\ell$. Then

$$\begin{aligned} \mathbb{E}[\mathbb{I}\{\hat{\varphi}_\ell(\boldsymbol{\omega}) > c\} | \boldsymbol{\omega}_{-1}] &= \int_{\mathbb{R}} \mathbb{I}\{\varphi_1(\omega_1) \leq \hat{\varphi}_{-1}^\ell(\boldsymbol{\omega}_{-1}) - c\} \rho_1(\omega_1) d\omega_1 \\ &= \int_{\mathbb{R}} \mathbb{I}\{\omega_1 \leq \hat{\psi}_d^\ell\} \rho_1(\omega_1) d\omega_1 \\ &= \Phi\left(\frac{\log \frac{\hat{\psi}_d^\ell}{S_0^1} - (\mu - \frac{\sigma_{11}^2}{2})\tau}{\sigma_{11}\sqrt{\tau}}\right). \end{aligned}$$

where $\hat{\psi}_d^\ell$ satisfies $\{\hat{\varphi}_\ell(\boldsymbol{\omega}) > c\} = \{\hat{\varphi}_\ell(\hat{\psi}_d^\ell(\boldsymbol{\omega}_{-1}), \boldsymbol{\omega}_{-1}) > c\} = \{\omega_1 \leq \hat{\psi}_d^\ell\}$.

The parameters in our experiments are set as follows: the initial prices of the assets $S_0^1 = \dots = S_0^d = 100$, $\mu = 8\%$, $\mu_0 = 5\%$, the strikes $K_1 = \dots = K_d = 95$, the maturity $T = 0.1$, and the risk horizon $\tau = 0.02$. Without loss of generality, we let $\Sigma = (\sigma_{ij})$ be a sub-triangular matrix satisfying $C = \Sigma \Sigma^T$ which corresponds with Cholesky decomposition of C , where $C_{ij} = 0.3 \cdot 0.98^{|i-j|}$ for $1 < i, j \leq d$ and $C_{11} = 0.3$.

We use the following Algorithm 1 (Giles 2015; Goda et al. 2020) to compute the number of outer simulation samples N_ℓ . Here D_ℓ denotes the level correction of the estimator under consideration: Y_ℓ , $\hat{Y}_\ell$, or $\bar{Y}_\ell$ for standard MLMC, MLMC with exact preintegration, or MLMC with numerical preintegration, respectively.

Algorithm 1 Let $\omega \in (0, 1)$ be a user-specified parameter. For a target root-mean-square accuracy ε , start with $L = L_0$ and give an initial number of samples N_* for all the levels $\ell = 0, \dots, L$. Until extra samples need to be evaluated, repeat the following:

- 1: Evaluate extra samples on each level.
- 2: Compute (or update) the empirical variances V'_ℓ for $\ell = 0, \dots, L$.
- 3: Define optimal N_ℓ for $\ell = 0, \dots, L$ according to

$$N_\ell = \left\lceil (1 - \omega)^{-1} \varepsilon^{-2} \sqrt{\frac{V'_\ell}{C_\ell}} \sum_{k=0}^L \sqrt{V'_k C_k} \right\rceil.$$

- 4: Test for the bias convergence $|\mathbb{E}[D_L]|/(2^\alpha - 1) \leq \sqrt{\omega} \varepsilon$, where we use the empirical estimates for $\mathbb{E}[D_\ell]$ and α .
 - 5: **if** the bias is not converged **then**
 - 6: Let $L = L + 1$ and give an initial number of samples N_L .
 - 7: **end if**
-

In this algorithm, the optimal allocation of N_ℓ is derived by minimizing the total cost $\sum_{\ell=0}^L N_\ell C_\ell$ for a fixed variance $\sum_{\ell=0}^L V_\ell / N_\ell = (1 - \omega) \varepsilon^2$. The bias convergence rate α is 1, due to the theoretical bias in Theorem 1.

All numerical experiments are based on the Black-Scholes model for a portfolio of European call options. For all levels ℓ , the inner sample size is set as $m_\ell = 32 \times 2^\ell$. The outer sample size N_ℓ is chosen adaptively by Algorithm 1 with parameters $\omega = 0.16$, $L_0 = 2$ and $N_* = 200,000$. We compare the standard Multilevel Monte Carlo (MLMC) and the proposed MLMC with preintegration (smoothed MLMC) in terms of the absolute value of expectation, variance, kurtosis and total computational cost, with the dimension covering $d = 4, 8, 16$. As we know, using an antithetic estimator in (27) does not change the variance convergence rate in the crude MLMC because of the indicator function. It is of interest to test whether antithetic sampling can get more benefit in the smoothed method. Therefore, we also test the strong convergence rate and computational cost of smoothed MLMC with antithetic sampling (denoted by smoothed AMLMC) in the following experiments.

Figure 1 shows the weak convergence behavior of MLMC and smoothed MLMC for dimensions $d = 4, 8, 16$. For both methods, $\log_2 |\text{mean}|$ decreases as the level increases, which means the estimation bias becomes smaller when we use more inner samples. The convergence rate of the smoothed method is almost the same as the standard MLMC method. This matches our theoretical expectation: preintegration does not change the convergence rate or the bias of the estimator. However, the curve of the smoothed MLMC is much smoother and more stable than standard MLMC. The preintegration method greatly improves the stability of the expectation estimation, even though it does not change the convergence speed.

Figure 2 shows the strong convergence of the three methods for $d = 4, 8, 16$. The variance decay of the standard MLMC is slow and its rate is near $-1/2$, while the variance of the smoothed MLMC decays fast with the increase of the level and its rate is near -1 , which verifies the conclusion of Theorem 2 that $V_\ell = O(m_\ell^{-1})$. The AMLMC

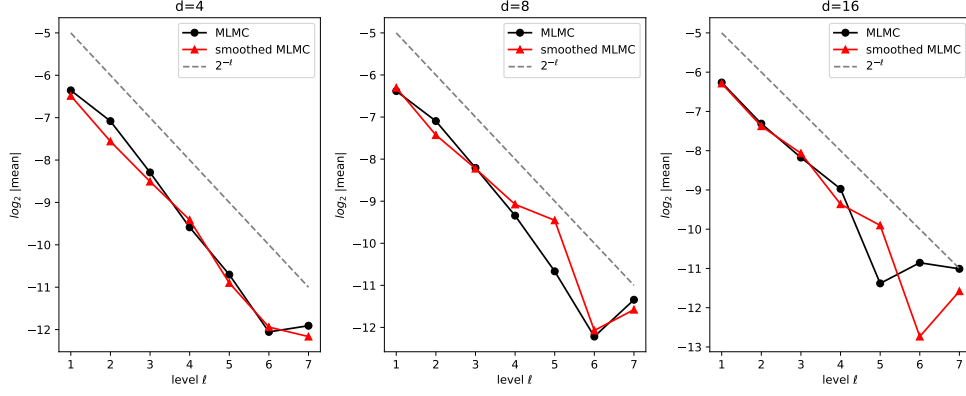


Fig. 1 Expectation for dimensions $d = 4, 8, 16$

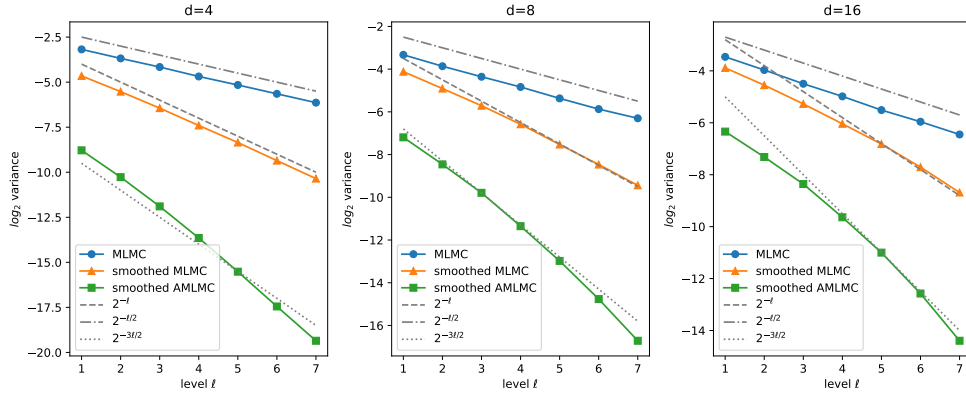


Fig. 2 Variance for dimensions $d = 4, 8, 16$

method achieves a convergence rate even close to $-3/2$. Its variance is reduced by a factor of 10,000 at the finest level compared with the standard MLMC method. This demonstrates the superiority of our smoothed MLMC method combined with antithetic sampling.

Figure 3 reflects the kurtosis of the two methods. The kurtosis of the standard MLMC increases significantly with the level, which is the high-kurtosis phenomenon caused by the discontinuity of the indicator function. In contrast, the kurtosis of the smoothed MLMC remains at a low level and gradually approaches a certain upper bound, which verifies Corollary 2 that the kurtosis of the proposed method is bounded ($\kappa_\ell = O(1)$). The preintegration technique effectively eliminates the discontinuity of the indicator function and improves the robustness of the estimation.

Figure 4 compares the total computational cost of the three methods for $d = 4, 8, 16$. Smooth MLMC achieves a near optimal computational complexity $O(\text{TOL}^{-2})$,

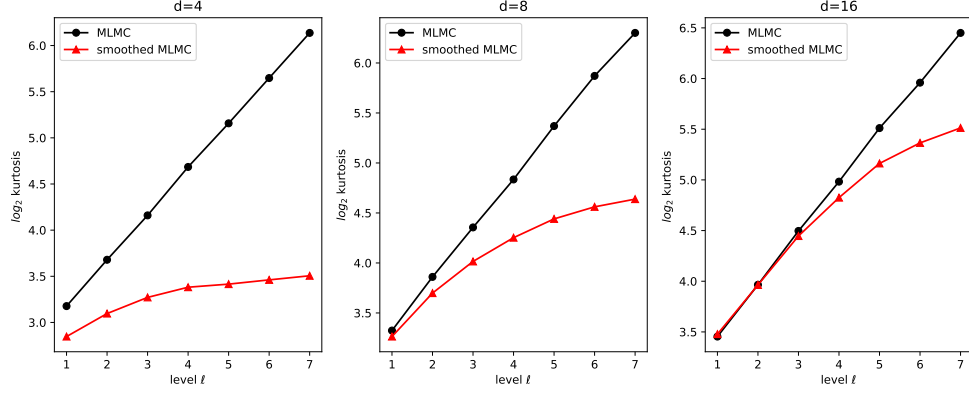


Fig. 3 Kurtosis for dimensions $d = 4, 8, 16$

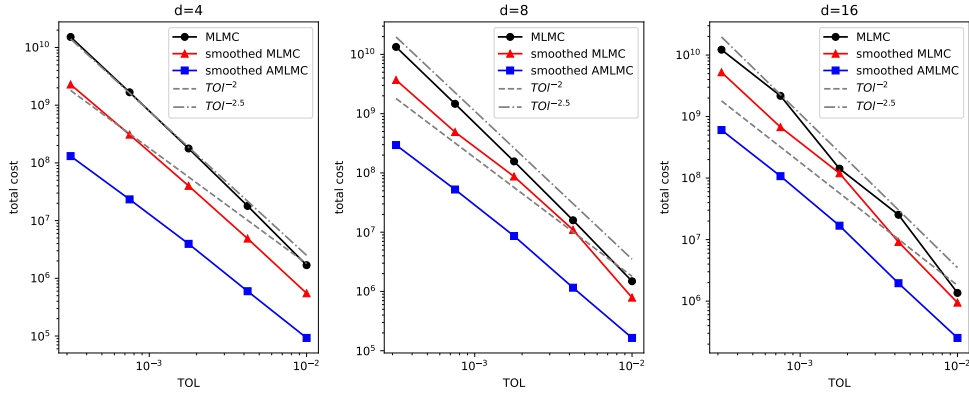


Fig. 4 Cost for dimensions $d = 4, 8, 16$

compared with the complexity $O(\text{TOL}^{-5/2})$ of standard MLMC. Smoothed AMLMC also attains the optimal complexity $O(\text{TOL}^{-2})$, while reducing the computational burden more effectively, its complexity is 100 times lower than that of the standard MLMC method.

From the variance results above, we observe that the reduction factor of the variance of the smoothed MLMC decreases as the dimension increases. This is not a flaw of the smoothed method itself. When we derive the preintegration formula, we use inner simulation to approximate the function φ . As the dimension increases, the accuracy of the inner simulation decreases, which reduces the accuracy of our preintegration approximation. To fix this accuracy issue of inner simulation in high-dimensional problems, we introduce the quasi-Monte Carlo (QMC) method and combine it with smoothed MLMC.

We incorporate QMC method within the inner simulation. To fit QMC framework, we assume that given $\omega = \mathbf{y}$, X can be generated via the mapping

$$X(\mathbf{y}) = \psi(\mathbf{u}; \mathbf{y}),$$

for some function ψ and $\mathbf{u} \sim U[0, 1]^s$ with dimension s . As a result, the inner estimator (2) is replaced by

$$\hat{\varphi}_m(\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m \psi(\mathbf{u}_j; \mathbf{y}), \quad (33)$$

and $\varphi(\mathbf{y}) = \int_{[0,1]^s} \psi(\mathbf{u}; \mathbf{y}) d\mathbf{u}$. If $\mathbf{u}_1, \dots, \mathbf{u}_m$ are QMC points (known as low discrepancy points), which are deterministic points chosen from $[0, 1]^s$ and are more uniformly distributed than random points, the approximation (33) refers to the QMC method.

The error of the QMC method can be bounded by the Koksma-Hlawka inequality (Niederreiter 1992)

$$|\varphi(\mathbf{y}) - \hat{\varphi}_m(\mathbf{y})| \leq V_{\text{HK}}(\psi(\cdot; \mathbf{y})) D^*(\mathbf{u}_1, \dots, \mathbf{u}_m), \quad (34)$$

where $V_{\text{HK}}(\psi(\cdot; \mathbf{y}))$ is the variation of the integrand $\psi(\cdot; \mathbf{y})$ for given $\mathbf{y}$ in the sense of Hardy and Krause which measures the smoothness of $\psi(\cdot; \mathbf{y})$, and $D^*(\mathbf{u}_1, \dots, \mathbf{u}_m)$ is the star discrepancy which measures the uniformity of the points set $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$. The Koksma-Hlawka inequality (34) implies that for functions of finite variation, the convergence rate of QMC approximation is determined by the star discrepancy $D^*(\mathbf{u}_1, \dots, \mathbf{u}_m)$, which is of order $O(m^{-1}(\log m)^s)$ for low discrepancy points.

There are various constructions of QMC point sets in the literature (Dick et al. 2017), such as digital nets and lattice rule point sets. For our numerical experiments, we use Sobol' sequences as QMC points and also use randomized QMC (RQMC) points. RQMC retains the essential equi-distribution structure, while allowing a statistical error estimation based on independent replications. Moreover, RQMC is able to improve the rate of convergence for smooth integrands, such as the Owen's scrambling technique (Owen 1995). For a survey on RQMC, we refer to L'Ecuyer and Lemieux (2002).

Figure 5 tests the strong convergence performance in high-dimensional cases $d = 32, 50$. Smoothed MLMC maintains a convergence rate of -1 even for high-dimensional problems. Besides, both smoothed MLQMC and smoothed AMLMC exhibit a faster convergence rate approaching $-3/2$, with smoothed MLQMC yielding a smaller variance. The smoothed MLQMC method also demonstrates superior numerical stability, as its kurtosis growth is bounded from above in Figure 6. Finally, by combining smoothed MLQMC with antithetic sampling (smoothed AMLQMC), its convergence rate even exceeds -2 , and its variance is reduced by a factor of 1,000,000 compared to MLMC at the finest level. However, owing to its high strong convergence rate, smoothed AMLQMC exhibits higher kurtosis and reduced numerical stability.

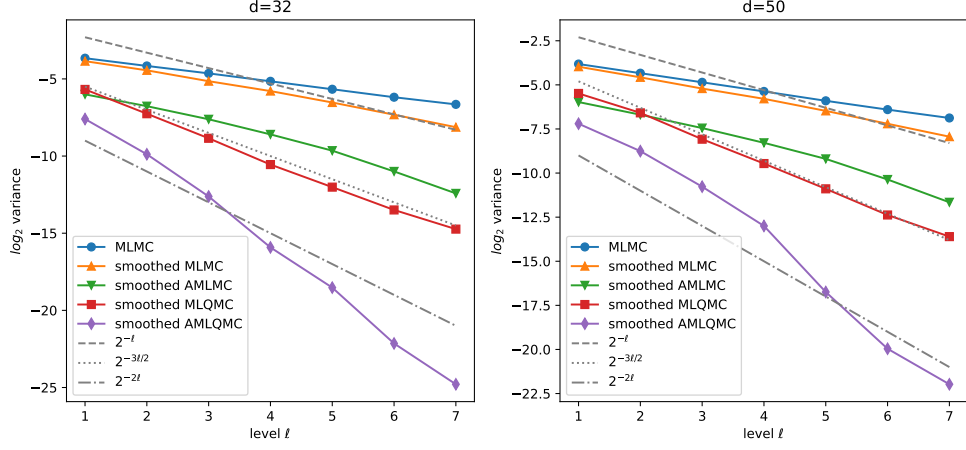


Fig. 5 Variance for dimensions $d = 32, 50$

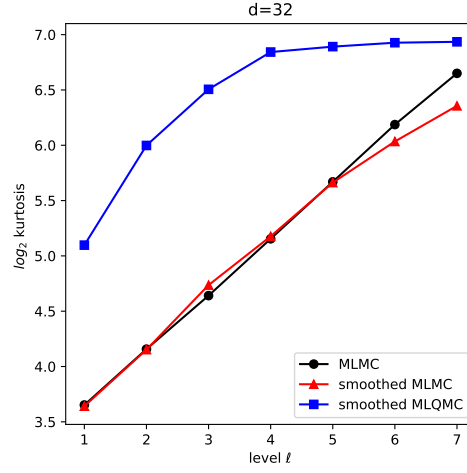


Fig. 6 Kurtosis for dimension $d = 32$

5 Conclusion

This paper develops a nested MLMC method with preintegration to solve the efficiency and robustness problems of standard nested simulation in risk estimation. The preintegration technique effectively handles the discontinuity of the indicator function in risk measures by integrating out one outer random variable. The theoretical analysis proves that the proposed method has fast variance decay $O(m_\ell^{-1})$, bounded kurtosis and lower computational complexity compared with the standard MLMC, which means better convergence and robustness.

Numerical experiments on European option portfolios under the Black-Scholes model validate the superiority of the smoothed MLMC with preintegration. It shows more stable weak convergence, faster decaying level variance and bounded kurtosis, which overcomes the high-kurtosis and slow variance decay problems of the standard MLMC. In terms of computational cost, the proposed method requires much less cost to achieve the same accuracy. For high-dimensional problems, combining the proposed method with QMC further improves the convergence performance, solving the accuracy decline of inner simulation in high dimensions.

The proposed method provides a new efficient approach for financial risk estimation, especially for high-dimensional portfolio risk measurement. Future research can focus on optimizing the parameter selection of the MLMC with preintegration, such as the optimal allocation of quadrature points and Newton iteration tolerance across different levels, and extending the method to more complex financial models (e.g., stochastic volatility models) and other risk measures (e.g., expected shortfall). Furthermore, the theory combining MLQMC and preintegration has not yet been established, which can also serve as an important direction for future research.

Declarations

Funding. This work was supported by the National Natural Science Foundation of China under Grant No. 72571153.

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate. Not applicable. This study is theoretical and computational and did not involve human participants, human data, animals, or biological materials.

Consent for publication. Not applicable.

Data availability. No external datasets were used in this study. All numerical results were generated through simulations under the Black-Scholes model using the model specifications and parameter settings described in the manuscript. The numerical data underlying the figures are available from the corresponding author upon reasonable request.

Materials availability. Not applicable.

Code availability. The source code used to generate the numerical results is available from the corresponding author upon reasonable request.

Author contributions. Yu Xu wrote the main manuscript and conducted all numerical experiments. Yifeng Yu contributed to improving the theoretical proofs and prepared it for final submission. Xiaoqun Wang revised the manuscript.

Appendix A Auxiliary Technical Result

Lemma 3 *Under Assumptions 1, 2, and 3, we have*

$$\lim_{|\omega_1| \rightarrow \infty} \mathbb{E} \left[\left(e_\ell(\omega_1, \omega_{-1}) \rho_1(\omega_1) \int_0^1 g(z(\theta; \omega_1, \omega_{-1})) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} d\theta \right)^2 \right] = 0.$$

The conclusion also holds at finite endpoints of the connected components of $\{\rho_1 > 0\}$, with one-sided limits taken from within the corresponding component. It remains valid for any measurable function g satisfying $|g| \leq 1$.

Proof We have $g(\cdot)$ is bounded and we have also $(\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1}$ is bounded in moments (similar to what we showed for (23) in the proof of Theorem 2).

Fix $\ell \geq 1$ and write $\mathcal{I} = \{y : \rho_1(y) > 0\}$. For $\omega_1 \in \mathcal{I}$, define

$$F_\ell(\theta; \omega_1, \omega_{-1}) := \rho_1(\omega_1) e_\ell(\omega_1, \omega_{-1}) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1},$$

$$b_\ell(\omega_1) := \left(\mathbb{E} \left[\int_0^1 |F_\ell(\theta; \omega_1, \omega_{-1})|^2 d\theta \right] \right)^{1/2},$$

where the expectation is over ω_{-1} and the inner random inputs. The common derivative sign bounds the inverse derivative of z by the maximum of the two endpoint inverse derivatives. Differentiating F_ℓ and applying Hölder's inequality, the assumptions and Lemma 2 therefore imply that b_ℓ is locally absolutely continuous on each connected component of $\mathcal{I}$ and satisfies

$$\int_{\mathcal{I}} (b_\ell(\omega_1) + |b'_\ell(\omega_1)|) d\omega_1 < \infty,$$

$$\int_{\mathcal{I}} \rho_1(\omega_1) \left(\frac{b_\ell(\omega_1)}{\rho_1(\omega_1)} \right)^u d\omega_1 < \infty \quad \text{for every } u \geq 2.$$

At a finite endpoint a of a component of $\mathcal{I}$, local absolute continuity of ρ_1 and Assumption 3 give $\rho_1(\omega_1) \leq C|\omega_1 - a|^{p_*-1}$ near a from within that component. Since b_ℓ has a finite endpoint limit, the second integrability bound, with $u > \max\{2, p_*/(p_* - 1)\}$, forces this limit to be zero. Extending b_ℓ by zero outside $\mathcal{I}$ thus gives an integrable absolutely continuous function with integrable derivative on $\mathbb{R}$, so $b_\ell(\omega_1) \rightarrow 0$ as $|\omega_1| \rightarrow \infty$. Finally, since $|g| \leq 1$, Cauchy–Schwarz gives

$$\mathbb{E} \left[\left| \rho_1(\omega_1) e_\ell(\omega_1, \omega_{-1}) \int_0^1 g(z(\theta; \omega_1, \omega_{-1})) (\partial_{\omega_1} z(\theta; \omega_1, \omega_{-1}))^{-1} d\theta \right|^2 \right] \leq b_\ell(\omega_1)^2.$$

The claimed limits follow. The same estimate applies to any measurable replacement of g bounded in absolute value by 1. $\square$

References

- Achtsis, N., Cools, R., Nuyens, D.: Conditional sampling for barrier option pricing under the Heston model. In: Dick, J., Kuo, F.Y., Peters, G.W., Sloan, I.H. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2012. Springer Proceedings in Mathematics & Statistics, vol. 65, pp. 253–269. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41095-6_9

- Achtsis, N., Cools, R., Nuyens, D.: Conditional sampling for barrier option pricing under the LT method. *SIAM Journal on Financial Mathematics* **4**(1), 327–352 (2013)
- Bayer, C., Ben Hammouda, C., Tempone, R.: Numerical smoothing with hierarchical adaptive sparse grids and quasi-Monte Carlo methods for efficient option pricing. *Quantitative Finance* **23**(2), 209–227 (2023)
- Bayer, C., Ben Hammouda, C., Tempone, R.: Multilevel Monte Carlo with numerical smoothing for robust and efficient computation of probabilities and densities. *SIAM Journal on Scientific Computing* **46**(3), 1514–1548 (2024) <https://doi.org/10.1137/22M1495718>
- Broadie, M., Du, Y., Moallemi, C.C.: Efficient risk estimation via nested sequential simulation. *Management Science* **57**(6), 1172–1194 (2011)
- Ben Hammouda, C., Ben Rached, N., Tempone, R.: Importance sampling for a robust and efficient multilevel Monte Carlo estimator for stochastic reaction networks. *Statistics and Computing* **30**(6), 1665–1689 (2020)
- Bujok, K., Hambly, B.M., Reisinger, C.: Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. *Methodology and Computing in Applied Probability* **17**(3), 579–604 (2015)
- Bayer, C., Siebenmorgen, M., Tempone, R.: Smoothing the payoff for efficient computation of basket option prices. *Quantitative Finance* **18**(3), 491–505 (2018)
- Caffisch, R.E.: Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* **7**, 1–49 (1998)
- Caffisch, R.E., Morokoff, W.J., Owen, A.B.: Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance* **1**(1), 27–46 (1997) <https://doi.org/10.21314/JCF.1997.005>
- Dick, J., Gantner, R.N., Le Gia, Q.T., Schwab, C.: Multilevel higher-order quasi-Monte Carlo Bayesian estimation. *Mathematical Models and Methods in Applied Sciences* **27**(5), 953–995 (2017)
- Dick, J., Pillichshammer, F.: *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, Cambridge (2010). <https://doi.org/10.1017/CBO9780511761188>
- Giles, M.B., Goda, T.: Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI. *Statistics and Computing* **29**(4), 739–751 (2019)
- Giles, M.B., Haji-Ali, A.-L.: Multilevel nested simulation for efficient risk estimation. *SIAM/ASA Journal on Uncertainty Quantification* **7**(2), 497–525 (2019)

- Goda, T., Hironaka, T., Iwamoto, T.: Multilevel Monte Carlo estimation of expected information gains. *Stochastic Analysis and Applications* **38**(4), 581–600 (2020)
- Giles, M.B.: Multilevel Monte Carlo path simulation. *Operations Research* **56**(3), 607–617 (2008)
- Giles, M.B.: Multilevel Monte Carlo methods. *Acta Numerica* **24**, 259–328 (2015)
- Giles, M.B.: MLMC for nested expectations. In: Dick, J., Kuo, F.Y., Woźniakowski, H. (eds.) *Contemporary Computational Mathematics: A Celebration of the 80th Birthday of Ian Sloan*, pp. 425–442. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-72456-0_20
- Gordy, M.B., Juneja, S.: Nested simulation in portfolio risk measurement. *Management Science* **56**(10), 1833–1848 (2010)
- Griewank, A., Kuo, F.Y., Leövey, H., Sloan, I.H.: High dimensional integration of kinks and jumps—smoothing by preintegration. *Journal of Computational and Applied Mathematics* **344**, 259–274 (2018)
- Goda, T., Murakami, D., Tanaka, K., Sato, K.: Decision-theoretic sensitivity analysis for reservoir development under uncertainty using multilevel quasi-Monte Carlo methods. *Computational Geosciences* **22**(4), 1009–1020 (2018)
- Glasserman, P., Staum, J.: Conditioning on one-step survival for barrier option simulations. *Operations Research* **49**(6), 923–937 (2001)
- Hull, J.C., Basu, S.: *Options, Futures, and Other Derivatives*. Pearson India Education Services, Noida (2016)
- Heinrich, S.: Monte Carlo complexity of global solution of integral equations. *Journal of Complexity* **14**(2), 151–175 (1998)
- Hong, L.J., Juneja, S., Liu, G.: Kernel smoothing for nested estimation with application to portfolio risk measurement. *Operations Research* **65**(3), 657–673 (2017)
- L’Ecuyer, P., Lemieux, C.: Recent advances in randomized Quasi-Monte Carlo methods. In: Dror, M., L’Ecuyer, P., Szidarovszky, F. (eds.) *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*. International Series in Operations Research & Management Science, vol. 46, pp. 419–474. Springer, New York, NY (2002). https://doi.org/10.1007/0-306-48102-2_20
- Niederreiter, H.: *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63. Society for Industrial and Applied Mathematics, Philadelphia, PA (1992). <https://doi.org/10.1137/1.9781611970081>

- Owen, A.B.: Randomly permuted (t, m, s) -nets and (t, s) -sequences. In: Niederreiter, H., Shiue, P.J.-S. (eds.) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Lecture Notes in Statistics, vol. 106, pp. 299–317. Springer, New York, NY (1995). https://doi.org/10.1007/978-1-4612-2552-2_19
- Owen, A.B.: Practical Quasi-Monte Carlo Integration. <https://artowen.su.domains/mc/practicalqmc.pdf>. Online book (2023)
- Xu, Z., He, Z., Wang, X.: Efficient risk estimation via nested multilevel quasi-Monte Carlo simulation. *Journal of Computational and Applied Mathematics* **443**, 115745 (2024)