

Nested Simulation Procedures in Financial Engineering: A Selective Review

Tony Wirjanto

Mingbin Feng

Xintong Li

Abstract

In the task of estimating risk measures for portfolios of complex financial derivatives, nested simulation procedures are commonly required but often computationally expensive. Tremendous efforts have been made to improve the efficiency of nested simulation procedures by approximating the inner simulation model with a proxy model. In this study, we review the literature on nested simulation procedures in financial engineering and establish fair comparison of different proxy models using the same set of numerical examples. Our study shows interesting findings on the performance of different proxy models and provides useful insights for practitioners to choose proxy models for nested simulation procedures.

1 Introduction

Nested simulation procedures are commonly used in financial engineering to estimate risk measures for portfolios of complex financial derivatives. The term *nested* is referring to a nested estimation problem, in which the estimation of the risk measure requires two levels of simulations. In a typical nested simulation procedure, an outer level simulation model generates underlying risk factors, which is referred to as the *outer scenarios*. For each outer scenario, an inner level simulation model generates scenario-wise samples of the portfolio losses, which is referred to as the *inner replications*.

The nested simulation procedure is computationally expensive due to its nested structure. Given a fixed computational budget, the nested simulation procedure has to make a trade-off between the number of outer scenarios and the number of inner replications. Gordy and Juneja (2010) are the first to analyze and propose the optimal budget allocation of a standard nested simulation procedure. The term *standard* refers to using a standard Monte Carlo estimator, the sample mean of the inner replication to estimate a scenario-wise portfolio loss for an outer scenario. Gordy and Juneja (2010) investigate the optimal budget allocation for a standard nested simulation procedure with respect to the mean squared error (MSE) of the estimated risk measure.

The standard nested simulation procedure is computationally expensive with a somewhat wasteful use of the simulation budget, as only the inner replications from the same outer scenario are used in estimating the scenario-wise portfolio loss for that outer scenario. Subsequent research efforts have been made to improve the efficiency of nested simulation procedures by using the inner replications from other outer scenarios. This is referred to as pooling. Different methods pool in different ways, either by a trained proxy model or by a pre-defined likelihood ratio weights. Broadie et al. (2015) propose a regression-based nested simulation procedure, which uses a trained regression proxy model to estimate the scenario-wise portfolio loss for an outer scenario by pooling the inner replications from all outer scenarios. For risk measures in certain forms, Broadie et al. (2015) show that it is optimal to allocate all simulation budget to the outer level simulation, and the inner replication should be kept to a minimum of 1. Similarly, Hong et al. (2017), Feng and Song (2020), and Zhang et al. (2022) use a kernel smoothing model, a likelihood ratio model, and a kernel ridge regression model as proxies to pool the inner replications from all outer scenarios.

This paper presents a survey study of some popular nested simulation procedures. Many procedures are proposed in the literature, but they are not directly comparable due to different error metric, different assumptions, and different numerical examples. Within a common analytical framework, we first summarize and compare their asymptotic rate of convergence. Their asymptotic convergence results are closely examined for their assumptions that guarantee the convergence. Furthermore, our study finds that different studies propose different examples in their numerical experiments, which introduces unfair advantages for certain simulation procedures over the others. A fair comparison among popular methods

is therefore urgently needed in the literature. Our numerical experiment is the first of its kind to subject back all the aforementioned simulation procedures to a complete and unbiased comparison. Extensive numerical experiments are conducted to show, in practical examples, how well the finite-sample performance of a method matches its theoretical convergence behavior. With our numerical examples, we compare the nested simulation procedures for different payoff complexity, problem dimensions, and risk measures.

The rest of the paper is organized as follows. Section 2 introduces the nested simulation procedure and the standard Monte Carlo estimator. Section 3 summarizes the asymptotic convergence results of nested simulation procedures in the literature. Section 4 presents the numerical experiments and the comparison results. Section 6 concludes the paper.

2 Problem Formulation

In a nested estimation problem, we are interested in the quantity

$$\rho(g(X)),$$

where $X \in \Omega$. $g(X)$ can't be directly evaluated, but it is the output of

$$g(X) = \mathbb{E}[Y|X = x]_{x=X}$$

Some common risk measures are in the nested expectation form, in which

$$\rho(g(X)) = \mathbb{E}[h(g(X))]$$

where $h(\cdot)$ is a known function. Forms of h include the followings:

- Smooth functions, e.g., a quadratic tracking error with benchmark b : $h(t) = (t - b)^2$
- Lipschitz continuous functions, e.g., a mean excess loss over threshold u : $h(t) = \max\{t - u, 0\}$. Here, h is a hockey-stick function.
- Indicator functions, e.g., probability of a large loss over a threshold u : $h(t) = \mathbb{I}_{\{t \geq u\}}$

Other risk measures of interest that are not in the nested expectation form are the value at risk (VaR) and the conditional value at risk (CVaR). The α -VaR of $g(X)$ is defined as

$$\text{VaR}_\alpha(g(X)) = q_\alpha = \inf \{q : \Pr(g(X) \leq q) \geq \alpha\}.$$

The α -CVaR of $g(X)$ is defined as

$$\text{VaR}_\alpha(g(X)) = \frac{1}{1 - \alpha} \int_\alpha^1 q_v dv.$$

2.1 The Standard Nested Simulation Procedure

The standard nested simulation procedure first simulate M independent and identically distributed (iid) outer scenarios X_1, \dots, X_M from F_X , the distribution of X . For each X_i , again simulate $Y_{ij}, j = 1, \dots, N$ from $F_{Y|X_i}$, the conditional distribution of Y given X_i . Given scenario i , the Y_{ij} are conditionally iid. Let $\Gamma = M \cdot N$ denote the total simulation budget, $f_X(x)$ denote the density of X , and $\mathbf{X} = (X_1, \dots, X_M)$ denote the vector of outer scenarios.

The standard nested simulation procedure estimates $g(X_i)$ with a standard Monte Carlo estimator

$$\hat{g}_N(X_i) = \frac{1}{N} \sum_{j=1}^N Y_{ij}; \quad Y_{ij} \sim F_{Y|X_i}$$

Let $(\hat{g}_N(\mathbf{X}))_{[1]}, \dots, (\hat{g}_N(\mathbf{X}))_{[M]}$ be the order statistics of $\hat{g}_N(X_1), \dots, \hat{g}_N(X_M)$. The standard nested simulation estimators for different forms of ρ are as follows:

1. Nested expectation form:

$$\hat{\rho}_{M,N} = \frac{1}{M} \sum_{i=1}^M h(\hat{g}_N(X_i)) = \frac{1}{M} \sum_{i=1}^M h(\bar{Y}_{N,i}); \quad X_i \sim F_X$$

2. Value at risk (VaR):

$$\hat{\rho}_{M,N} = (\hat{g}_N(\mathbf{X}))_{[\alpha M]}$$

3. Conditional value at risk (CVaR):

$$\hat{\rho}_{M,N} = (\hat{g}_N(\mathbf{X}))_{[\alpha M]} + \frac{1}{(1-\alpha)M} \sum_{i=1}^M \max\{\hat{g}_N(X_i) - (\hat{g}_N(\mathbf{X}))_{[\alpha M]}, 0\}$$

2.2 Supervised Learning Models

In supervised learning, $g(\cdot)$ can be approximated by $\hat{g}_{M,N}^{\text{SL}}(\cdot)$, which is based on a chosen function family \mathcal{G} and observations from the standard nested simulation procedure. Consider the observation pairs $(X_i, \hat{g}_N(X_i))$ for $i \in \{1, \dots, M\}$ as training data, we can use supervised learning to approximate $g(\cdot)$ by $\hat{g}_{M,N}^{\text{SL}}(\cdot)$ and to pool the inner replications from all outer scenarios. Using the M training samples, a nested Monte Carlo estimator of ρ is given by

1. Nested expectation form:

$$\hat{\rho}_{M,N}^{\text{SL,Train}} = \frac{1}{M} \sum_{i=1}^M h(\hat{g}_{M,N}^{\text{SL}}(X_i)); \quad X_i \sim F_X$$

2. VaR:

$$\hat{\rho}_{M,N}^{\text{SL,Train}} = (\hat{g}_{M,N}^{\text{SL}}(\mathbf{X}))_{[\alpha M]}$$

3. CVaR:

$$\hat{\rho}_{M,N}^{\text{SL,Train}} = (\hat{g}_{M,N}^{\text{SL}}(\mathbf{X}))_{[\alpha M]} + \frac{1}{(1-\alpha)M} \sum_{i=1}^M \max\{\hat{g}_{M,N}^{\text{SL}}(X_i) - (\hat{g}_{M,N}^{\text{SL}}(\mathbf{X}))_{[\alpha M]}, 0\}$$

where $(\hat{g}_{M,N}^{\text{SL}}(\mathbf{X}))_{[\alpha M]}$ is the $[\alpha M]$ -th order statistic of $\hat{g}_{M,N}^{\text{SL}}(X_1), \dots, \hat{g}_{M,N}^{\text{SL}}(X_M)$. Similarly, with M' test samples of X , namely $\tilde{\mathbf{X}} = \tilde{X}_1, \dots, \tilde{X}_{M'}$, an estimator is given by

1. Nested expectation form:

$$\hat{\rho}_{M,N,M'}^{\text{SL,Test}} = \frac{1}{M} \sum_{i=1}^M h(\hat{g}_{M,N}^{\text{SL}}(\tilde{X}_i)); \quad \tilde{X}_i \sim F_X.$$

2. VaR:

$$\hat{\rho}_{M,N,M'}^{\text{SL,Test}} = (\hat{g}_{M,N}^{\text{SL}}(\tilde{\mathbf{X}}))_{[\alpha M]}.$$

3. CVaR:

$$\hat{\rho}_{M,N,M'}^{\text{SL,Test}} = (\hat{g}_{M,N}^{\text{SL}}(\tilde{\mathbf{X}}))_{[\alpha M]} + \frac{1}{(1-\alpha)M} \sum_{i=1}^M \max\{\hat{g}_{M,N}^{\text{SL}}(\tilde{X}_i) - (\hat{g}_{M,N}^{\text{SL}}(\tilde{\mathbf{X}}))_{[\alpha M]}, 0\},$$

where $(\hat{g}_{M,N}^{\text{SL}}(\tilde{\mathbf{X}}))_{[\alpha M]}$ is the $[\alpha M]$ -th order statistic of $\hat{g}_{M,N}^{\text{SL}}(\tilde{X}_1), \dots, \hat{g}_{M,N}^{\text{SL}}(\tilde{X}_M)$. Note that $\hat{g}_{M,N}^{\text{SL}}(\cdot)$ is derived from the training samples $(X_1, \hat{g}_N(X_1)), \dots, (X_M, \hat{g}_N(X_M))$.

We are interested in minimizing the MSE of the supervised learning-based nested simulation estimator with supervised learning $\hat{\rho}_{M,N}^{\text{SL,Train}}$ and $\hat{\rho}_{M,N,M'}^{\text{SL,Test}}$, subject to the total simulation budget Γ .

$$\begin{aligned} \min_{M,N} \quad & \text{MSE}(\hat{\rho}_{M,N}^{\text{SL}}) = \mathbb{E} \left[(\hat{\rho}_{M,N}^{\text{SL}} - \rho)^2 \right] \\ \text{subject to} \quad & M \cdot N = \Gamma \end{aligned} \tag{1}$$

Existing literature on nested simulation procedures have proposed different methods to approximate the true function $g(\cdot)$ with supervised learning.

- Regression:

$$\hat{g}_{M,N}^{\text{REG}}(X) = \Phi(X)\hat{\beta},$$

where Φ is a chosen basis, and $\hat{\beta}$ is estimated from the training samples.

- Kernel smoothing:

$$\hat{g}_{M,N}^{\text{KS}}(X) = \frac{\sum_{i=1}^M \bar{Y}_{N,i} K_w(X - X_i)}{\sum_{i=1}^M K_w(X - X_i)},$$

where K_w is the kernel function with bandwidth w .

- Kernel ridge regression:

$$\hat{g}_{M,N}^{\text{KRR}}(X) = \arg \min_{g \in \mathcal{N}_{\Psi}(\Omega)} \left(\frac{1}{M} \sum_{i=1}^M (\hat{g}_N(X_i) - g(X_i))^2 + \lambda \|g\|_{\mathcal{N}_{\Psi}(\Omega)}^2 \right),$$

where $\mathcal{N}_{\Psi}(\Omega)$ is the reproducing kernel Hilbert space (RKHS) with kernel Ψ defined domain Ω , and λ is the regularization parameter as in ridge regression. More specifically, Φ is a Matérn kernel with smoothness parameter ν and length scale parameter ℓ .

2.3 Likelihood Ratio Method

Instead of using a supervised learning model, the likelihood ratio method uses the likelihood ratio weights to pool the inner replications from all outer scenarios. Here, we restrict our attention to problems in the nested expectation form whose outer scenarios characterize the stochasticity of the inner simulation model. Specifically,

$$Y = Y(H, X),$$

where H is a random variable whose distribution is specified by the outer scenarios X . We denote the conditional distribution of $H|X$ by $f_{H|X}$. For a specific scenario X_i , we write $f_{H|X}(\cdot|X_i)$. To reconcile with previously established notations, we note that inner simulation outputs Y_{ij} can be written as

$$Y_{ij} = Y(H_{ij}, X_i),$$

where $H_{ij} \sim f_{H|X}(\cdot|X_i)$. Suppose that one can generate random variable H from some sampling distribution f_H . Then, the likelihood ratio estimator of ρ is given by

$$\hat{\rho}_{M,N}^{\text{LR}} = \frac{1}{M} \sum_{i=1}^M h(\hat{g}_N^{\text{LR}}(X_i)),$$

where the inner replications are pooled by the likelihood ratio weights with

$$\hat{g}_N^{\text{LR}}(X_i) = \frac{1}{N} \sum_{j=1}^N Y(H_j, X_i) \frac{f_{H|X}(H_j|X_i)}{f_H(H_{ij})}, \quad H_j \sim f_H, \quad i = 1, \dots, M.$$

With a total simulation budget Γ , we are interested in the order of convergence of estimators for all nested simulation procedures, which is measured by their MSE about ρ .

$$\begin{aligned} \text{MSE}(\hat{\rho}_{M,N}) &= \mathbb{E} \left[(\hat{\rho}_{M,N} - \rho)^2 \right] \\ \text{subject to } M \cdot N &= \Gamma \end{aligned} \tag{2}$$

3 Asymptotic Analysis

In this section, we summarize the asymptotic convergence results of the nested simulation procedures in the literature, and we compare their assumptions on the risk measures, the proxy models, and the simulation models.

¹A \checkmark indicates there exists an asymptotic convergence result for the corresponding estimator, a \times indicates there does not exist an asymptotic convergence result for the corresponding estimator, and a \star indicates an asymptotic result is not available in the literature, but we show them to fill in the hole.

Estimator for $g(X)$	Smooth h	Lipschitz h (hockey-stick)	Indicator h	VaR	CVaR
Standard Monte Carlo ¹	★	★(✓)	✓	✓	×
Regression	✓	✓(✓)	×	×	×
Kernel smoothing	✓	×	✓	×	×
Kernel ridge regression	×	×	×	×	×
Likelihood ratio	✓	×	✓	×	×

Table 1: Existing asymptotic convergence results of nested simulation procedures for MSE

Table 1 While most of the literature focuses on the MSE of the estimator of ρ , Wang et al. (2022) analyze the asymptotic convergence of the estimator of ρ in terms of the absolute error. Let $\hat{\rho}$ be the estimator of ρ . Its absolute error about ρ is defined as

$$\text{Absolute Error}(\hat{\rho}) = |\hat{\rho} - \rho|.$$

In Wang et al. (2022), the authors of the KRR-based nested simulation procedures claim to have bridged the gap between the cubic and square root convergence rates of nested simulation procedures. However, they analyze convergence in probabilistic order, and it is only applicable in terms of the absolute error. Instead of showing the convergence of KRR-based estimator in terms of MSE as in Gordy and Juneja (2010), we show the connections between the convergence in MSE and the convergence in probabilistic order for absolute error. Our results show that Wang et al. (2022) indeed bridge the gap, but only in terms of the absolute error.

3.1 Connections between Convergence in MSE and Absolute Error

In order to show the connections between the convergence in MSE and the convergence in probabilistic order for absolute error, we first define what it means for a sequence of random variables to converge in those two forms.

Definition 1 Let $\hat{\rho}_\Gamma$ be an estimator of ρ with a simulation budget of Γ . We write $\mathbb{E}[(\hat{\rho}_\Gamma - \rho)^2] = \mathcal{O}(\Gamma^{-k})$, that is, $\hat{\rho}_\Gamma$ converges in MSE to ρ in order k if

$$\exists C \limsup_M \frac{\mathbb{E}[(\hat{\rho}_\Gamma - \rho)^2]}{\Gamma^{-k}} \leq C$$

Definition 2 Let $\hat{\rho}_\Gamma$ be an estimator of ρ with a simulation budget of Γ . We write $|\hat{\rho}_\Gamma - \rho| = \mathcal{O}_\mathbb{P}(\Gamma^{-k})$, that is $\hat{\rho}_\Gamma$ converges in probabilistic order k to ρ if for large enough Γ ,

$$\forall \epsilon > 0 \exists C \mathbb{P}(|\hat{\rho}_\Gamma - \rho| \geq C\Gamma^{-k}) \leq \epsilon$$

We start by showing the convergence in probabilistic order from the convergence in MSE. Let $\hat{\rho}_\Gamma$ be an estimator of ρ with a simulation budget of Γ , and assume that $\mathbb{E}[(\hat{\rho}_\Gamma - \rho)^2] = \mathcal{O}(\Gamma^{-k})$. Then, from the definition of convergence in MSE, there exists a constant C such that

$$\limsup_M \frac{\mathbb{E}[(\hat{\rho}_\Gamma - \rho)^2]}{\Gamma^{-k}} \leq C.$$

Hence, there exists some Γ such that for all $\gamma \geq \Gamma$,

$$\mathbb{E}[(\hat{\rho}_\gamma - \rho)^2] \leq C\gamma^{-k}.$$

The convergence in probabilistic order can be shown by separating the expectation into two parts: tail and nontail.

$$\mathbb{E}[(\hat{\rho}_\gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\gamma - \rho| \leq d\gamma^s\}}] + \mathbb{E}[(\hat{\rho}_\gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\gamma - \rho| > d\gamma^s\}}] \leq C\gamma^{-k}.$$

The first term is always positive, and the second term can be bounded from below by the indicator function.

$$\begin{aligned} \mathbb{E}[(\hat{\rho}_\gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\gamma - \rho| > d\gamma^s\}}] &\geq \mathbb{E}[d^2\gamma^{2s} \cdot \mathbb{I}_{\{|\hat{\rho}_\gamma - \rho| > d\gamma^s\}}] \\ &= d^2\gamma^{2s} \cdot \mathbb{E}[\mathbb{I}_{\{|\hat{\rho}_\gamma - \rho| > d\gamma^s\}}] \\ &= d^2\gamma^{2s} \cdot \mathbb{P}(|\hat{\rho}_\gamma - \rho| > d\gamma^s) \end{aligned}$$

Combining bounds on the two terms, we have

$$d^2 \gamma^{2s} \mathbb{P}(|\hat{\rho}_\gamma - \rho| > d\gamma^s) \leq C\gamma^{-k}.$$

Let $s = -\frac{k}{2}$. Arranging the terms, we have

$$\mathbb{P}\left(|\hat{\rho}_\gamma - \rho| > d\gamma^{-\frac{k}{2}}\right) \leq \frac{C}{d^2}$$

Hence, for all $\epsilon > 0$, there exist $C^* = \sqrt{\frac{C}{\epsilon}}$ such that for all $\gamma \geq \Gamma$,

$$\mathbb{P}\left(|\hat{\rho}_\gamma - \rho| > C^* \gamma^{-\frac{k}{2}}\right) \leq \epsilon$$

In essence, the above is the definition of convergence in probabilistic order, that is,

$$|\hat{\rho}_\gamma - \rho| = \mathcal{O}_{\mathbb{P}}\left(\Gamma^{-\frac{k}{2}}\right)$$

Theorem 1 *Let $\hat{\rho}_\Gamma$ be an estimator of ρ with a simulation budget of Γ . If $\hat{\rho}_\Gamma$ converges in MSE to ρ in order k , then $\hat{\rho}_\Gamma$ converges in probabilistic order to ρ in order $\frac{k}{2}$.*

While the convergence in MSE implies the convergence in probabilistic order, the converse is not necessarily true. Similarly, the above argument is applied in reverse. Let $\hat{\rho}_\Gamma$ be an estimator of ρ with a simulation budget of Γ , and assume that $|\hat{\rho}_\Gamma - \rho| = \mathcal{O}_{\mathbb{P}}(\Gamma^{-k})$. The MSE of $\hat{\rho}_\Gamma$ can be separated into the same two parts.

$$\mathbb{E}\left[(\hat{\rho}_\Gamma - \rho)^2\right] = \mathbb{E}\left[(\hat{\rho}_\Gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\Gamma - \rho| \leq d\Gamma^{-2k}\}}\right] + \mathbb{E}\left[(\hat{\rho}_\Gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\Gamma - \rho| > d\Gamma^{-2k}\}}\right],$$

where the first term can be bounded from above.

$$\begin{aligned} \mathbb{E}\left[(\hat{\rho}_\Gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\Gamma - \rho| > d\Gamma^{-2k}\}}\right] &\leq d^2 \Gamma^{-2k} \cdot \mathbb{E}[\mathbb{I}_{\{|\hat{\rho}_\Gamma - \rho| > d\Gamma^{-2k}\}}] \\ &= d^2 \Gamma^{-2k} \cdot \mathbb{P}(|\hat{\rho}_\Gamma - \rho| > d\Gamma^{-2k}) \leq d^2 \Gamma^{-2k} \end{aligned}$$

However, the second term is not always bounded. If the random variable $\hat{\rho}_\Gamma$ admits a density function f , then the second term can be further decomposed.

$$\mathbb{E}\left[(\hat{\rho}_\Gamma - \rho)^2 \cdot \mathbb{I}_{\{|\hat{\rho}_\Gamma - \rho| > d\Gamma^{-2k}\}}\right] = \int_{-\infty}^{-d\Gamma^{-2k}} (x - \rho)^2 f(x) dx + \int_{d\Gamma^{-2k}}^{\infty} (x - \rho)^2 f(x) dx$$

Hence, $\hat{\rho}_\Gamma$ converges in MSE to ρ in order $2k$ if and only if both integrals converge in order higher than $2k$.

From the above analysis and 1, we conclude that convergence results for MSE in the literature can be translated into convergence results in probabilistic order for absolute error.

Estimator for $g(X)$	Smooth h	Lipschitz h (hockey-stick)	Indicator h	VaR	CVaR
Standard Monte Carlo	★	★(✓)	✓	✓	×
Regression	✓	✓(✓)	×	×	×
Kernel smoothing	✓	×	✓	×	×
Kernel ridge regression	✓	×	✓	✓	✓
Likelihood ratio	✓	×	✓	×	×

Table 2: Existing asymptotic convergence results of nested simulation procedures for absolute error

3.2 Asymptotic Analysis for the Standard Nested Simulation Procedure

In Gordy and Juneja (2010), the authors analyze the asymptotic convergence of the standard nested simulation procedure in terms of MSE. The analysis is complete for the nested expectation form where h is either an indicator function or a hockey-stick function and VaR. For the nested expectation form where h is a smooth function or a Lipschitz continuous function, the analysis is incomplete. In this section, we fill in the holes in the analysis of Gordy and Juneja (2010).

Assumption 1 $h(g(X))$ has finite second moment, i.e., $\mathbb{E}[(h(g(X)))^2] < \infty$.

Assumption 2 $\hat{g}_N(X) = g(X) + \bar{Z}_N(X)$, where the simulation noise $\bar{Z}_N(X)$ has zero mean and variance $\nu(X)/N$, where the conditional variance $\nu(X)$ is bounded, i.e., there exists $C_{\nu,1} > 0$ such that $\nu(x) \leq C_{\nu,1}$ for all $x \in \mathbb{R}$.

Let $\rho_M = \frac{1}{M} \sum_{i=1}^M h(g(X_i))$ be the nested Monte Carlo estimator with the true function g . The MSE of the standard procedure can be decomposed into two terms.

$$\begin{aligned}
& \mathbb{E}[(\hat{\rho}_{M,N} - \rho)^2] \\
& \leq 2\mathbb{E}[(\hat{\rho}_{M,N} - \rho_M)^2] + 2\mathbb{E}[(\rho_M - \rho)^2] \\
& = 2\mathbb{E}\left[\left(\frac{1}{M} \sum_{i=1}^M h(\hat{g}_N(X_i)) - \frac{1}{M} \sum_{i=1}^M h(g(X_i))\right)^2\right] + 2\mathbb{E}\left[\left(\frac{1}{M} \sum_{i=1}^M h(g(X_i)) - \mathbb{E}[h(g(X))]\right)^2\right] \\
& = 2\mathbb{E}\left[\left(\frac{1}{M} \sum_{i=1}^M h(\hat{g}_N(X_i)) - h(g(X_i))\right)^2\right] + \frac{2}{M} \text{Var}(h(g(X))) \\
& = 2\mathbb{E}\left[\left(\frac{1}{M} \sum_{i=1}^M h(\hat{g}_N(X_i)) - h(g(X_i))\right)^2\right] + \mathcal{O}(M^{-1}), \tag{3}
\end{aligned}$$

where the last equality follows from Assumption 1. The analysis of the first term is different for smooth and Lipschitz continuous h . We will analyze them separately.

Smooth h

Assumption 3 The function h has bounded first and second order derivative, i.e., there exists $C_1 > 0$ such that $|h'(x)| \leq C_1$ for all $x \in \mathbb{R}$, and there exists $C_2 > 0$ such that $|h''(x)| \leq C_2$ for all $x \in \mathbb{R}$.

Assumption 3 is similar to the smoothness assumption in Wang et al. (2022). Since h is a smooth function, Taylor expansion can be applied to the first term in Equation 3.

$$\begin{aligned}
& \mathbb{E}\left[\left(\frac{1}{M} \sum_{i=1}^M h(\hat{g}_N(X_i)) - h(g(X_i))\right)^2\right] \\
& = \mathbb{E}\left[\left(\frac{1}{M} \sum_{i=1}^M h'(g(X_i))(\hat{g}_N(X_i) - g(X_i)) + \frac{1}{2M} \sum_{i=1}^M h''(z_i)(\hat{g}_N(X_i) - g(X_i))^2\right)^2\right] \\
& \leq 2\mathbb{E}\left[\underbrace{\left(\frac{1}{M} \sum_{i=1}^M h'(g(X_i))(\hat{g}_N(X_i) - g(X_i))\right)^2}_{S_1}\right] + 2\mathbb{E}\left[\underbrace{\left(\frac{1}{2M} \sum_{i=1}^M h''(z_i)(\hat{g}_N(X_i) - g(X_i))^2\right)^2}_{S_2}\right] \tag{4}
\end{aligned}$$

where the last inequality is due to $2ab \leq a^2 + b^2$ for any $a, b \in \mathbb{R}$. For different methods of nested estimation, each of the two terms on the right-hand side of Equation 4 can be analyzed separately. We start with the first term S_1 .

$$\begin{aligned}
S_1 &= \mathbb{E} \left[\left(\frac{1}{M} \sum_{i=1}^M h'(g(X_i)) (\hat{g}_N(X_i) - g(X_i)) \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{1}{M} \sum_{i=1}^M h'(g(X_i)) \bar{Z}_N(X_i) \right)^2 \right] \\
&\leq C_1^2 \mathbb{E} \left[\left(\frac{1}{M} \sum_{i=1}^M \bar{Z}_N(X_i) \right)^2 \right] \\
&= C_1^2 \mathbb{E} \left[\frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \bar{Z}_N(X_i) \bar{Z}_N(X_j) \right] \\
&= C_1^2 \mathbb{E} \left[\frac{1}{M^2} \sum_{i=1}^M \bar{Z}_N^2(X_i) + \frac{1}{M^2} \sum_{i=1}^M \sum_{j \neq i}^M \bar{Z}_N(X_i) \bar{Z}_N(X_j) \right] \\
&\leq \frac{C_1^2 C_{\nu,1}}{MN} = \mathcal{O}(M^{-1} N^{-1})
\end{aligned} \tag{5}$$

where the last inequality in Equation 5 is due to Assumption 2, independence of X_i and X_j for $i \neq j$, and the fact that $\mathbb{E} [\bar{Z}_N(X)] = 0$. It remains to analyze the second term S_2 , where Assumption 4 is necessary to ensure the existence of the fourth moment of the simulation noise and the convergence of the second term.

Assumption 4 *The fourth moment of simulation noise $\bar{Z}_N(X)$ follows $\mathbb{E} [(\bar{Z}_N(X))^4] = \nu_2(X)/N^2$, where $\nu_2(X)$ is bounded, i.e., there exists $C_{\nu,2} > 0$ such that $\nu_2(X) \leq C_{\nu,2}$ for all $x \in \mathbb{R}$.*

$$\begin{aligned}
S_2 &= \mathbb{E} \left[\left(\frac{1}{2M} \sum_{i=1}^M h''(z_i) (\hat{g}_{M,N}^{\text{SL}}(X_i) - g(X_i))^2 \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{1}{2M} \sum_{i=1}^M h''(z_i) \bar{Z}_N^2(X_i) \right)^2 \right] \\
&\leq C_2^2 \mathbb{E} \left[\left(\frac{1}{2M} \sum_{i=1}^M \bar{Z}_N^2(X_i) \right)^2 \right] \\
&= C_2^2 \mathbb{E} \left[\frac{1}{2M^2} \sum_{i=1}^M \sum_{j=1}^M \bar{Z}_N^2(X_i) \bar{Z}_N^2(X_j) \right] \\
&= C_2^2 \mathbb{E} \left[\frac{1}{2M^2} \sum_{i=1}^M \bar{Z}_N^4(X_i) + \frac{1}{2M^2} \sum_{i=1}^M \sum_{j \neq i}^M \bar{Z}_N^2(X_i) \bar{Z}_N^2(X_j) \right] \\
&\leq C_2^2 \left(\frac{C_{\nu,2} M}{2M^2 N^2} + \frac{C_{\nu,1}^2 M(M-1)}{2M^2 N^2} \right) = \mathcal{O}(N^{-2})
\end{aligned} \tag{6}$$

where the second inequality is due to Assumption 3, and the last inequality follows from Assumption 4 and the fact that $\hat{g}_N(X)$ is a standard Monte Carlo estimator of $g(X)$. Combining Equation 3, 4, 5, and 6, we have

$$\mathbb{E} [(\hat{\rho}_{M,N} - \rho)^2] = \mathcal{O}(M^{-1}) + \mathcal{O}(N^{-2}) \tag{7}$$

Setting $M = \mathcal{O}(\Gamma^{2/3})$ and $N = \mathcal{O}(\Gamma^{1/3})$, we provide the same rate of convergence as obtained for other risk measures in Gordy and Juneja (2010). As shown in Section 3.1, the convergence in MSE automatically implies the convergence in probabilistic order.

Definition 3 Let h be a smooth function. MSE of the standard nested simulation procedure converges in order $\Gamma^{2/3}$, that is,

$$\mathbb{E} \left[(\hat{\rho}_{M,N} - \rho)^2 \right] = \mathcal{O}(\Gamma^{-2/3}).$$

The absolute error of the standard nested simulation procedure converges in probabilistic order $\Gamma^{-1/3}$, that is,

$$|\hat{\rho}_{M,N} - \rho| = \mathcal{O}_{\mathbb{P}}(\Gamma^{-1/3}).$$

It remains to analyze the convergence for Lipschitz continuous function h .

Lipschitz Continuous h

For the CVaR analysis in Gordy and Juneja (2010), the authors assume the knowing of the corresponding VaR. Hence, the analysis of CVaR is not complete. Instead, the CVaR that is being analyzed is actually the mean excess loss, which corresponds to h being a hockey-stick function and belongs to the family of the Lipschitz continuous functions. Here we present the convergence analysis for the whole family of the Lipschitz continuous functions, which includes the hockey-stick function as a special case.

Assumption 5 The function h is Lipschitz continuous. Hence, $|h(x_1) - h(x_2)| \leq K|x_1 - x_2|$ for some constant $K < \infty$.

Assumption 5 is a standard assumption for analysis involving Lipschitz continuous functions, and it is also used in Broadie et al. (2015). For the Lipschitz continuous case, the first term in Equation 3 is analyzed differently.

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{M} \sum_{i=1}^M h(\hat{g}_N(X_i)) - h(g(X)) \right)^2 \right] &\leq \mathbb{E} \left[(h(\hat{g}_N(X)) - h(g(X)))^2 \right] \\ &\leq K^2 \mathbb{E} \left[(\hat{g}_N(X) - g(X))^2 \right] \\ &= K^2 \mathbb{E} \left[(\bar{Z}_N(X))^2 \right] = \mathcal{O}(N^{-1}) \end{aligned} \quad (8)$$

where the first inequality follows from Cauchy-Schwarz inequality, and the second equality follows from Assumption 5, and the last equality follows from Assumption 2.

$$\mathbb{E} \left[(\hat{\rho}_{M,N} - \rho)^2 \right] = \mathcal{O}(M^{-1}) + \mathcal{O}(N^{-1}) \quad (9)$$

Setting $M = \mathcal{O}(\Gamma^{1/2})$ and $N = \mathcal{O}(\Gamma^{1/2})$, we provide a looser bound than the one obtained for hockey-stick h . As shown in Section 3.1, the convergence in MSE automatically implies the convergence in probabilistic order.

Theorem 2 Let h be a Lipschitz continuous function. MSE of the standard nested simulation procedure converges in order $\Gamma^{1/2}$, that is,

$$\mathbb{E} \left[(\hat{\rho}_{M,N} - \rho)^2 \right] = \mathcal{O}(\Gamma^{-1/2}).$$

The absolute error of the standard nested simulation procedure converges in probabilistic order $\Gamma^{-1/4}$, that is,

$$|\hat{\rho}_{M,N} - \rho| = \mathcal{O}_{\mathbb{P}}(\Gamma^{-1/4}).$$

3.3 Critical Assumptions in Existing Literature

In this section, we discuss the critical assumptions in the existing literature that guarantee the convergence of the nested simulation procedures.

4 Finite-Sample Experiments

The theoretical framework has allowed the comparison of the asymptotic convergence behavior for different nested simulation procedures. However, simulation budget in practice is almost always finite. So we are interested in comparing these procedures with practically relevant finite simulation budgets. Numerical experiments are conducted on portfolios that contain options on 20 underlying assets, whose dynamics follow a multidimensional geometric Brownian motion with 0.3 pairwise correlation. 5 nested simulation procedures are considered, namely the standard nested simulation procedure, the regression-based, kernel smoothing-based, likelihood ratio-based, and kernel ridge regression-based nested simulation procedures. For the standard nested simulation procedure, the bootstrap-based budget allocation strategy from Zhang et al. (2021) is implemented to estimate the optimal numbers of outer and inner simulation budget. For the regression proxy, the Laguerre polynomials up to degree 3 are used as the basis functions. The kNN proxy is implemented as the kernel smoothing proxy. The procedures are compared for 5 risk measures, namely a quadratic tracking error, a mean excess loss over threshold u , a probability of a large loss over threshold u , the 90% VaR, and the 90% CVaR, where the threshold u is set to be the 90% VaR. The procedures are compared for different portfolios.

- Portfolio 1 considers d assets. The portfolio contains 3 European call options written on each asset with strikes 90, 100, and 110, respectively.
- Portfolio 2 considers d assets. The portfolio contains 3 geometric Asian options written on each asset with strikes 90, 100, and 110, respectively.
- Portfolio 3 considers d assets. The portfolio contains 3 up-and-out barrier call options written on each asset with strikes 90, 100, and 110, respectively. They have a barrier level of 120.
- Portfolio 4 considers d assets. The portfolio contains 3 down-and-out barrier call options written on each asset with strikes 90, 100, and 110, respectively. They have a barrier level of 90.
- Portfolio 5 contains 1 asset. The portfolio longs 2 down-and-out barrier put options and shorts 1 barrier down-and-out put option.

Except for portfolio 5, 5 different asset dimensions are considered, i.e., $d = 1, 2, 5, 10, 20$, while we only consider $d = 1$ for portfolio 5. 525 experimental settings that arise from an exhaustive combination of the above simulation procedures, risk measures, portfolios, and asset dimensions are generated. For each of the scenarios, the estimation of the risk measure is repeated 150 times to obtain the empirical MSE of the corresponding estimator under a range of simulation budgets. The total simulation budget Γ ranges from 1,000 to 100,000. The empirical convergence results of each estimator are measured and recorded to illustrate their usefulness under different risk measures, option types and asset dimensions. We start by examining the empirical convergence results for the most basic case, i.e., the quadratic tracking error risk measure for European call options on Portfolio 1 with $d = 1$.

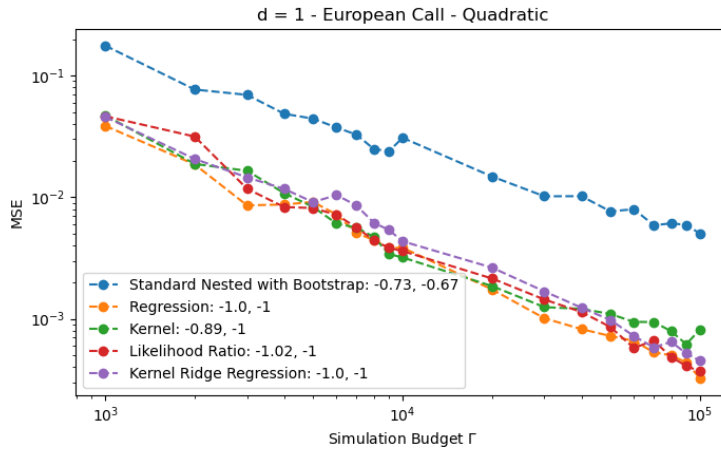


Figure 1: Empirical convergence of nested simulation procedures for quadratic tracking error on Portfolio 1 with $d = 1$

In Figure 1, the MSEs of the nested simulation procedures are plotted against the total simulation budget Γ in log-log scale, and a regression line is fitted for each procedure. In the legend, the first number denotes the slope of the fitted line, and the second number is the corresponding asymptotic rate of convergence. The slope of the fitted line can be regarded as the empirical rate of convergence of the corresponding procedure. Hence, all procedures but the kernel smoothing-based procedure have empirical rates of convergence that closely match their asymptotic rates. The kernel smoothing-based procedure has slower empirical rate due to cross-validation of the hyperparameter k for the kNN proxy at higher simulation budgets.

4.1 Sensitivity to the Asset Dimension

In this section, we examine the sensitivity of the empirical convergence of the nested simulation procedures to the number of assets.

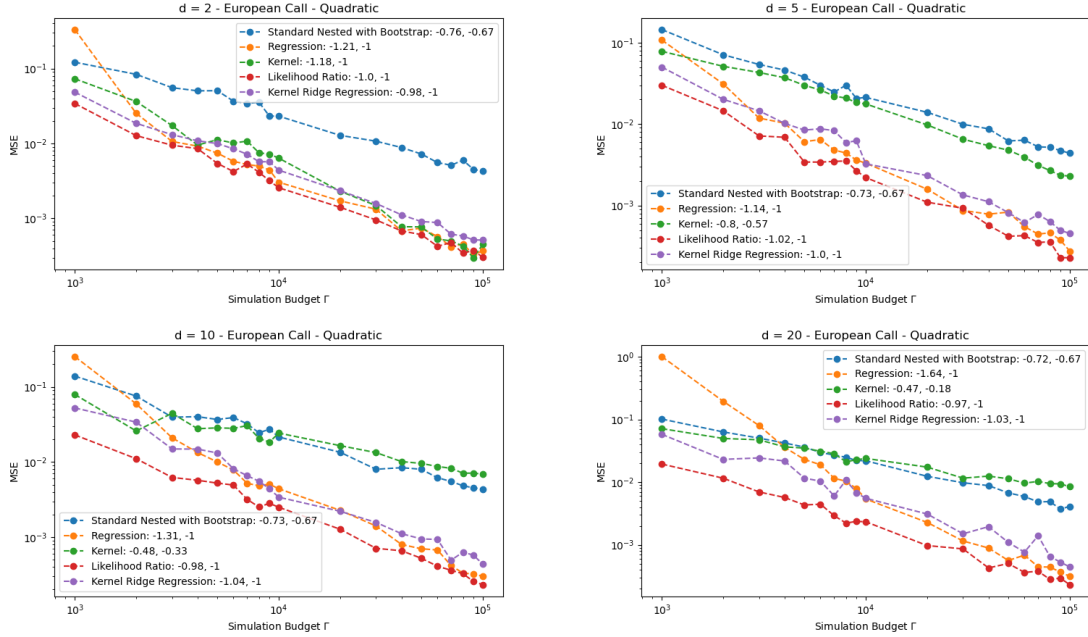


Figure 2: Empirical convergence of nested simulation procedures for quadratic tracking error on Portfolio 1 with increasing asset dimensions

In Figure 2, the empirical convergence of the nested simulation procedures for quadratic tracking error on Portfolio 1 is illustrated for increasing asset dimensions. Except for the kernel smoothing-based procedure that has a slower asymptotic rate of convergence for higher asset dimensions, all procedures have asymptotic rates of convergence that are independent of the asset dimension. While the empirical rates of convergence of the standard, the likelihood ratio-based and the KRR based procedures are close to their asymptotic rates, the empirical rates of convergence of the regression-based and the kernel smoothing-based procedures are sensitive to the asset dimension and are higher than their theoretical counterparts. For both the regression-based and the kernel smoothing-based procedures, the reason for the higher empirical rates can be explained by having poor proxy estimators for smaller simulation budgets. For smaller simulation budgets, the proxy estimators of the true inner simulation are poor, and the MSEs of the regression-based and the kernel smoothing-based procedures are dominated by the model error of the proxy estimators. In other words, they have not reached the asymptotic regime yet. In the next section, we will examine the empirical convergence of the regression-based procedure in more detail at higher simulation budgets, where asymptotic convergence is reached.

4.2 Empirical Convergence of the Regression Procedure

In our previous numerical experiments, we observe that the empirical rate of convergence of the regression procedure is much larger than its asymptotic rate of convergence. For dimensions larger than 10, the

MSE of the regression procedure decreases quickly in the beginning, and it stabilizes after a certain budget level. Here, this phenomenon is examined in more detail.

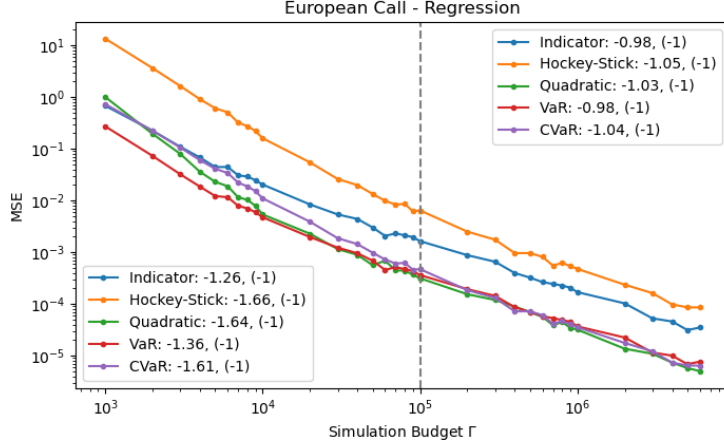


Figure 3: Empirical convergence of regression procedure for European call options and $d = 20$

On the left part of Figure 3 is the MSEs of the regression procedure for budget sizes that are smaller than 10,000. Slopes of the fitted lines on the left corresponds to the empirical rate of convergence of the regression procedure for smaller budget sizes. To investigate the convergence behavior for larger budget sizes, we conduct additional experiments for the European call option and with dimension 20. The additional experiments are summarized by the right part of Figure 3. After a certain budget level, the empirical rate of convergence for the regression procedure comes closer to the theoretical rate. For nested simulation procedures whose proxy models are biased, the proxy estimators of the true inner simulation is poor especially for smaller budget sizes. We are able to clearly observe this phenomenon for the regression proxy, and it can be explained by dividing the MSE into proxy bias and variance in simulation. For small budget sizes, the improvement of the bias of the regression proxy dominates the improvement of simulation variance. Performing poorly for extremely low simulation budgets, the regression proxy improves significantly. As the simulation budget gets larger, the improvement of the regression proxy becomes negligible comparing with that of the simulation variance. The regression proxy ceases to improve after reaching a certain level ($\Gamma = 100,000$ in our case), and the improvement of simulation variance dominates.

4.3 Sensitivity to the Option Type

In our previous numerical experiments, we have examine the empirical convergence behavior of the nested simulation procedures for European call options, which is the most basic option type. To examine the sensitivity of the empirical convergence behavior to the option type, we consider path-dependent options, namely geometric Asian options and barrier options.

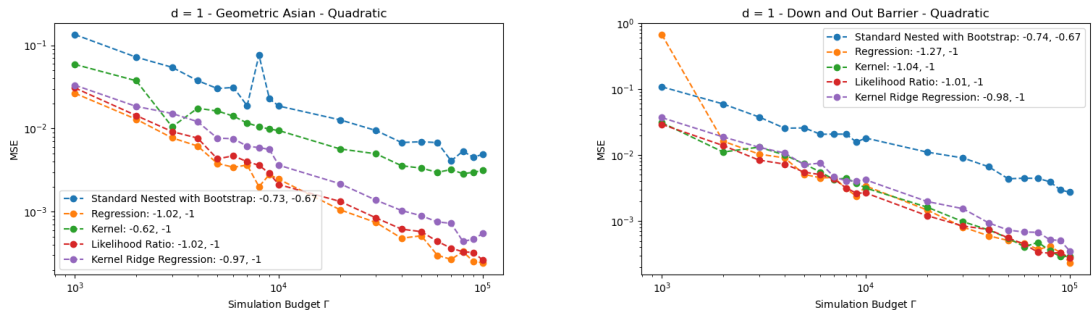


Figure 4: Empirical convergence of nested simulation procedures for quadratic tracking error on different portfolios with $d = 1$

At a lower asset dimension, the empirical rates of convergence of the standard, KRR-based, and

likelihood ratio-based procedures closely ensemble their asymptotic rates. In Figure 4, we have similar observations for these procedures as in the sensitivity analysis with respect to the asset dimension. The empirical rate of convergence of the regression-based procedure is higher than its asymptotic rate for the down-and-out barrier call options. Its poor performance is only observed at a very low simulation budget, while it reaches the asymptotic regime for any simulation budget that is larger than 1,000. The empirical rate of convergence of the kernel smoothing-based procedure is lower than its asymptotic rate for the portfolio of geometric Asian options. The cross-validation of the hyperparameter k for the kNN proxy is highly problem dependent. The optimal k is different for different combinations of option types and simulation budgets, and it is difficult to find a single k that works well under all circumstances.

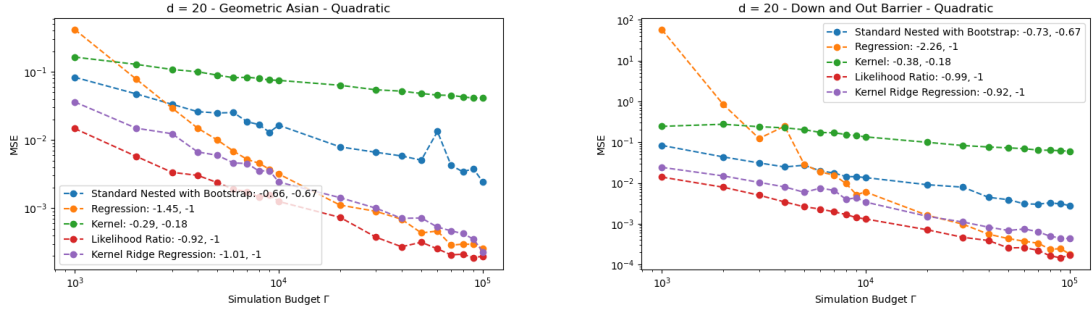


Figure 5: Empirical convergence of nested simulation procedures for quadratic tracking error on different portfolios with $d = 20$

Figure 5 illustrates similar observations for higher asset dimensions. The empirical rate of convergence of the regression-based procedure and the kernel smoothing-based procedure is much higher than their asymptotic rates at $d = 20$. Their asymptotic levels are not reached.

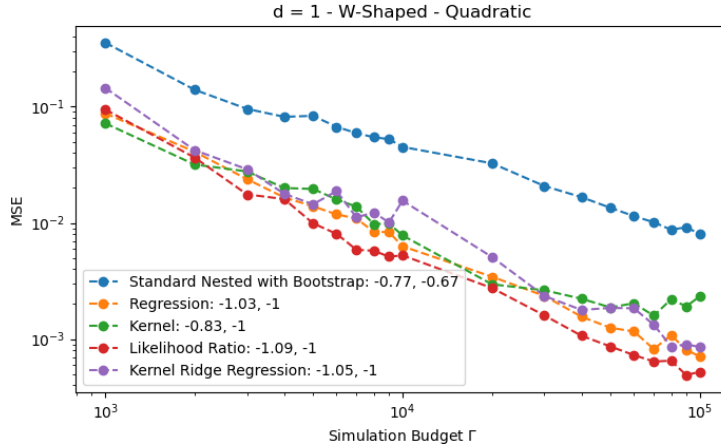


Figure 6: Empirical convergence of nested simulation procedures for quadratic tracking error on a W-shaped payoff

In Broadie et al. (2015), the authors propose a numerical example where the payoff has a W-shape with respect to the underlying asset price. We have incorporated this example as Portfolio 5. In Figure 6, the empirical convergence of the nested simulation procedures for quadratic tracking error on Portfolio 5 is illustrated. For all procedures, we observe a similar convergence behavior as in Figure 4, where the empirical rates of convergence of the standard, regression-based, KRR-based, and likelihood ratio-based procedures closely ensemble their asymptotic rates, but the empirical rate of convergence of the kernel smoothing-based procedure is lower than its asymptotic rate.

4.4 Sensitivity to the Risk Measure

In our previous numerical experiments, we have examined the empirical convergence behavior of the nested simulation procedures for the quadratic tracking error risk measure. To examine the sensitivity of

the empirical convergence behavior to the risk measure, we consider the mean excess loss over threshold u , the probability of a large loss over threshold u , the 90% VaR, and the 90% CVaR.

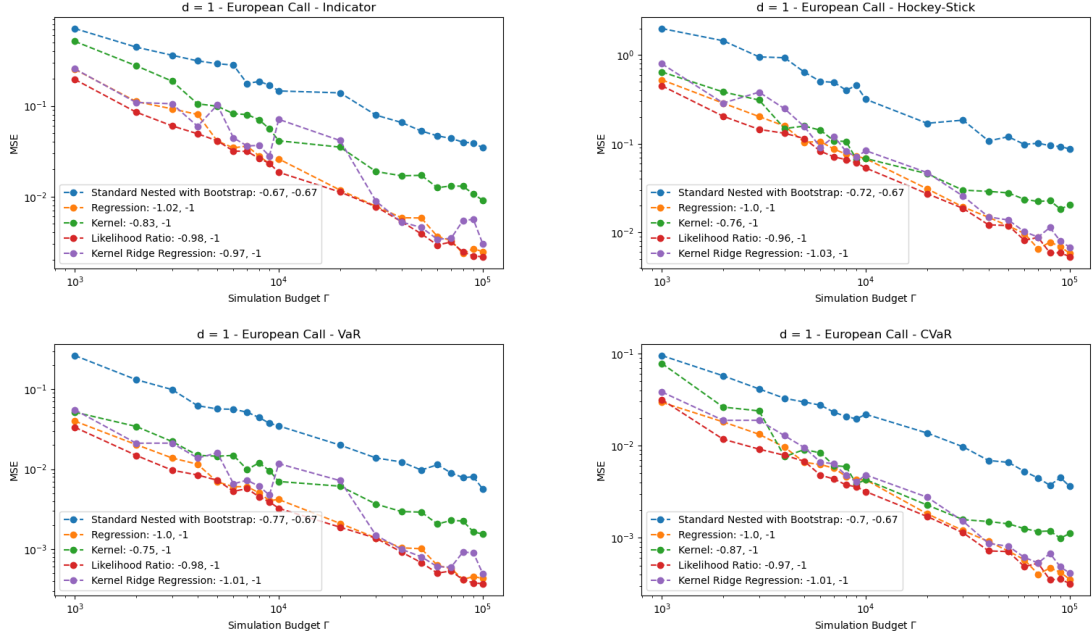


Figure 7: Empirical convergence of nested simulation procedures for different risk measures on Portfolio 1 with $d = 1$

In Figure 7, the empirical convergence of the nested simulation procedures for different risk measures on Portfolio 1 is illustrated. The empirical rates of convergence of the standard, regression-based, KRR-based, and likelihood ratio-based procedures closely ensemble their asymptotic rates for all risk measures. However, the empirical rate of convergence of the kernel smoothing-based procedure is lower than its asymptotic rate for all risk measures, for a similar reason as in Section 4.1 and Section 4.3.

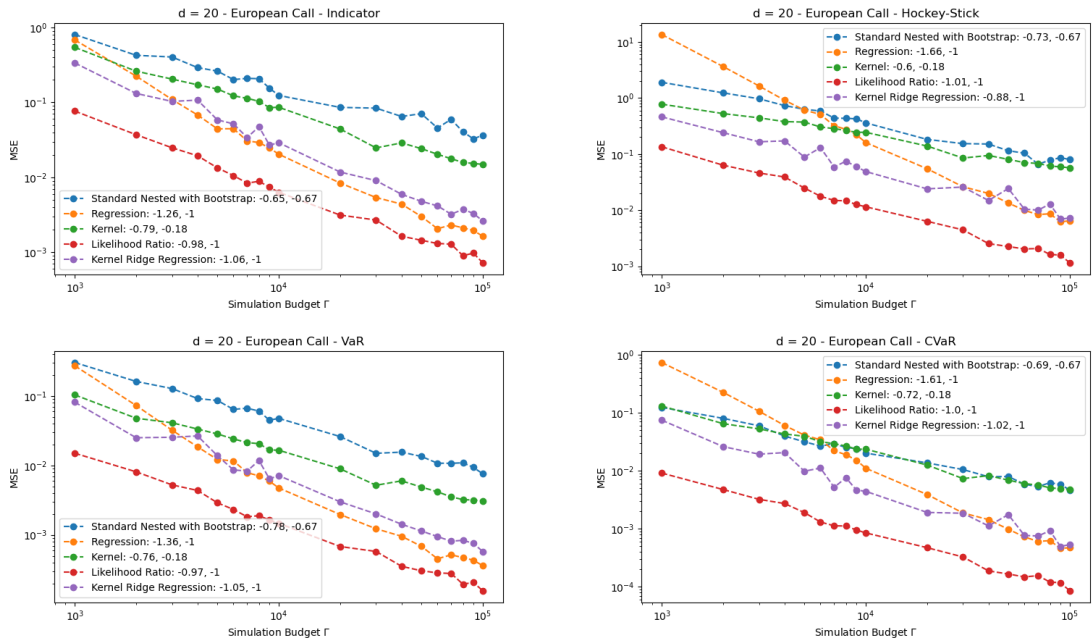


Figure 8: Empirical convergence of nested simulation procedures for different risk measures on Portfolio 1 with $d = 20$

Figure 8 illustrates the empirical results for higher asset dimensions, where the MSEs of the kernel

smoothing-based and the KRR-based procedures do not decrease monotonically as the simulation budget increases. This behavior can be attributed to the cross-validation of hyperparameters, which is a common phenomenon for any kernel-based procedures. In addition, the kernel smoothing-based procedure has not reached the asymptotic regime for any risk measure, but it is difficult to increase the simulation budget further due to its heavy computational burden.

4.5 Sensitivity to the Asset Model (TODO)

Our previous numerical experiments have been conducted under the assumption that the underlying asset dynamics follow a multidimensional geometric Brownian motion with 0.3 pairwise correlation. To examine the sensitivity of the empirical convergence behavior to the asset model, we consider the stochastic volatility model and the jump-diffusion model.

5 Computational Cost

Compared to the standard procedure, proxy-based nested simulation procedures often have faster empirical rates of convergence. The faster convergence benefits from pooling of inner samples through the proxy models. However, pooling itself comes with computational costs, which is usually ignored in most numerical comparisons. We summarize the algorithmic complexity and illustrate the computational time of different nested simulation procedures. The findings can further provide guidance on the choice of proper nested simulation procedures given a nested estimation problem. Define basic operations to be basic mathematical operators and the simulation of an outer path. The algorithmic complexity of the nested simulation procedures can be summarized and compared in the table below. The simulation cost is omitted, as all methods are given the same simulation budget.

Procedures	Training Cost	Prediction Cost
Standard procedure	0	$\mathcal{O}(M)$
Regression	$\mathcal{O}(p^3) + \mathcal{O}(p^2 M)$	$\mathcal{O}(dM) + \mathcal{O}(pM)$
Kernel smoothing	$\mathcal{O}(dM + M^3 A_{KS})$	$\mathcal{O}(dM^2 + M^3)$
Likelihood ratio	0	$\mathcal{O}(dM^2)$
Kernel ridge regression	$\mathcal{O}(dM^2 + M^3) + \mathcal{O}(dM^2 + M^4 A_{KRR_1} A_{KRR_2})$	$\mathcal{O}(dM^2)$

Table 3: Algorithmic complexity of nested simulation procedures

Table 3 shows the algorithmic complexity of the nested simulation procedures, where $p \geq d$ is the number of features for the regression proxy, $|A_{KS}|$ is the cardinality of the hyperparameter set for kernel smoothing. For the kernel ridge regression proxy, $|A_{KRR_1}|$ and $|A_{KRR_2}|$ denote the cardinalities of the hyperparameter set and the set of the penalty terms, respectively. Among the proxy-based procedures, regression is the most efficient as p is usually much smaller than M . Kernel smoothing and kernel ridge regression are kernel-based proxies, and they are more expensive due to distance calculation and cross-validation of hyperparameters. Calculating the likelihood ratio weights is inevitable for the likelihood ratio proxy, but it does not involve any training.

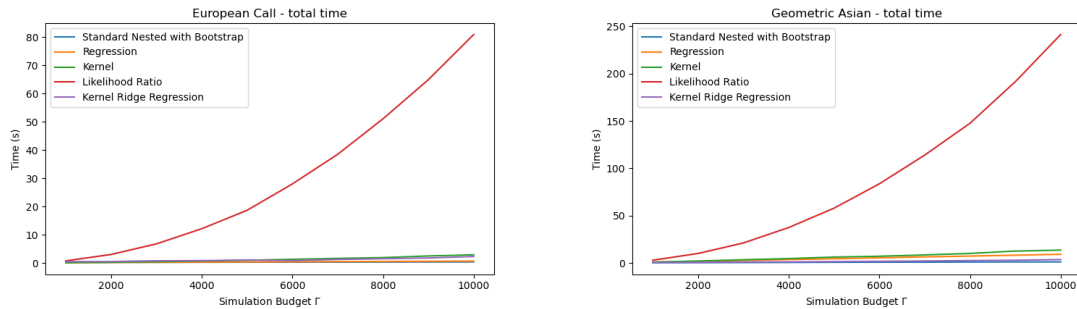


Figure 9: Total computational cost for different procedures with $d = 20$

In Figure 9, the x -axis shows the simulation budget Γ , and y -axis is the computational time for 1 replication of the numerical experiment, in seconds. We found that the computational burden of the

other procedures is negligible compared to that of the likelihood ratio. The cost of computing likelihood weights exhibits quadratic growth with the simulation budget. The computational time is higher for path-dependent options, and lower for the European call options. Other than the likelihood ratio-based procedure, the regression-based and kernel smoothing-based procedures have higher total computational cost for all option portfolios. It suffices to conclude that the likelihood ratio-based procedure is too computationally expensive to be feasible for practical applications. In addition, many proxies require validation methods to estimate their hyperparameters before the proxy models are fitted. For many simulation procedures, the validation step is time-consuming.

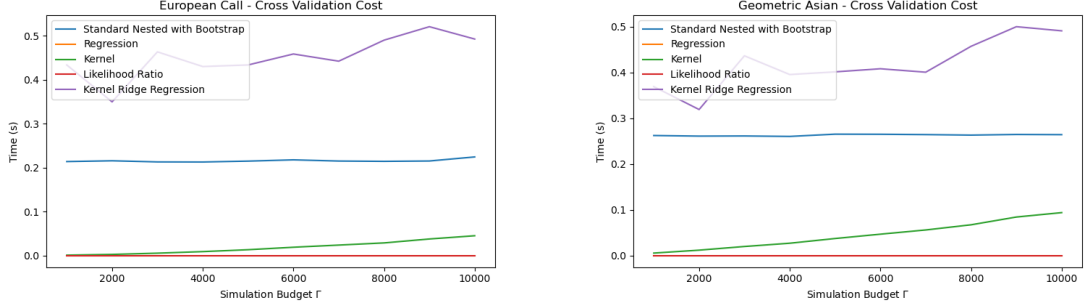


Figure 10: Validation cost for different procedures with $d = 20$

In Figure 10, we observe that the regression-based procedure and the likelihood ratio-based procedure requires model validation, while the model validation costs of the kernel smoothing-based procedure and the kernel ridge regression-based procedure come from cross-validation. The cardinality of the search space of the model hyperparameter is expected to increase in the power of the number of the hyperparameters. For the standard procedure, bootstrap-based budget allocation is used to estimate the optimal M and N . The performance of the chosen hyperparameters is robust to the increase in simulation budget, i.e., the chosen hyperparameters perform well both on lower and higher simulation budgets.

6 Conclusion

This selected review summarizes the asymptotic convergence results of different nested simulation procedures in a comprehensively theoretical framework. Asymptotic properties of estimators for different procedures are influenced by their corresponding proxy models. With extensive numerical experiments, we have found the finite-sample performance of a procedure can deviate from its theory. Proxy models provide faster convergence as simulation budget increases, but they come at the computational expense of model training and generating prediction from the trained models. In the end, the total computational budget is not necessarily the same as the simulation budget, which is usually a limiting factor for practical applications. For a nested estimation problem with a given computational budget, we suggest the use of the regression procedure when the budget size is moderate. The regression proxy is quick to fit, and its MSE decreases quickly when the simulation budget increases in a certain range. When a sufficient computational budget is given, the likelihood ratio procedure has a higher accuracy. Although expensive to compute, the likelihood ratio estimator is asymptotically unbiased, and it has a consistently fast empirical convergence for different payoff functions, risk measures, and asset dimensions. The kernel smoothing procedure requires cross-validation, and its empirical performance depends heavily on asset dimension and problem complexity. Empirically, the kernel ridge regression estimator is not guaranteed to have monotonically decreasing MSEs. The search space of its hyperparameters is prohibitively large, and the meandering behavior in its empirical MSEs is caused by errors from cross-validation.

Among all popular proxy models, we recommend the use of the regression proxy for moderate budget sizes. It is simple to implement, and it exhibits fast empirical convergence results. In addition, the choice of basis functions does not have a significant impact on the performance of the regression proxy for a problem dimension less than 20. For tasks that require a high estimation accuracy and are abundant with computational budget, the likelihood ratio proxy is recommended for its consistent convergence result and its asymptotic unbiasedness.

References

- Mark Broadie, Yiping Du, and Ciamac C Moallemi. Risk estimation via regression. *Operations Research*, 63(5):1077–1097, 2015.
- Mingbin Feng and Eunhye Song. Optimal nested simulation experiment design via likelihood ratio method. *arXiv preprint arXiv:2008.13087*, 2020.
- Michael B Gordy and Sandeep Juneja. Nested simulation in portfolio risk measurement. *Management Science*, 56(10):1833–1848, 2010.
- Jeff L Hong, Sandeep Juneja, and Guangwu Liu. Kernel smoothing for nested estimation with application to portfolio risk measurement. *Operations Research*, 65(3):657–673, 2017.
- Wenjia Wang, Yanyuan Wang, and Xiaowei Zhang. Smooth nested simulation: bridging cubic and square root convergence rates in high dimensions. *arXiv preprint arXiv:2201.02958*, 2022.
- Kun Zhang, Guangwu Liu, and Shiyu Wang. Bootstrap-based budget allocation for nested simulation. *Operations Research*, 2021.
- Kun Zhang, Mingbin Feng, Guangwu Liu, and Shiyu Wang. Sample recycling for nested simulation with application in portfolio risk measurement. *arXiv preprint arXiv:2203.15929*, 2022.