Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

CONSTRUCTING CONFIDENCE INTERVALS FOR VALUE-AT-RISK VIA NESTED SIMULATION

Qianwen Zhu[1], Guangwu Liu[1], and Xianyu Kuang[1]

[1]Dept. of Decision Analytics and Operations, City University of Hong Kong, Kowloon, Hong Kong

ABSTRACT

Nested simulation is a powerful tool for estimating widely-used risk measures, such as Value-at-Risk (VaR). While point estimation of VaR has been extensively studied in the literature, the topic of interval estimation remains comparatively underexplored. In this paper, we present a novel nested simulation procedure for constructing confidence intervals (CIs) for VaR with statistical guarantees. The proposed procedure begins by generating a set of outer scenarios, followed by a screening process that retains only a small subset of scenarios likely to result in significant portfolio losses. For each of these retained scenarios, inner samples are drawn, and the minimum and maximum means from these scenarios are used to construct the CI. Theoretical analysis confirms the asymptotic coverage probability of the resulting CI, ensuring its reliability. Numerical experiments validate the method, demonstrating its high effectiveness in practice.

1 INTRODUCTION

Measurement of portfolio risk is fundamental to risk management for financial institutions. It serves as a critical component of the current regulatory framework, providing essential information for risk managers to take timely actions to ensure risks stay within acceptable limits. In practice, however, institutions often manage large-scale portfolios comprising hundreds of thousands of financial instruments—including equities, currencies, and derivatives such as futures and options. Accurate valuation of these instruments often requires intensive simulation, particularly when sophisticated pricing models are employed. This makes the measurement of portfolio risk computationally challenging, especially for large-scale portfolios.

To measure portfolio risk, various risk measures have been proposed. Among them, one of the most widely-used risk measures is Value-at-Risk (VaR), which is defined as the quantile of the probability distribution of the portfolio loss at a given level. The main difficulty in estimating risk measures such as VaR stems from the fact that the loss function, which is expressed as a conditional expectation, cannot be evaluated analytically and requires simulation. In this case, estimation of risk measures usually calls for nested simulation (also referred to as two-level simulation), where risk scenarios are simulated in the outer level, while inner simulation is employed to estimate portfolio loss given each outer-level scenario.

In the early years, practitioners often perceived the computational burden of nested simulation procedures to be prohibitive, leading them to adopt highly simplified models to avoid the need for inner simulations. However, Gordy and Juneja (2010) showed that a relatively small number of inner samples could suffice to achieve estimates with reasonable accuracy, sparking renewed interest in nested simulation techniques. To enhance the efficiency of nested simulation, Broadie et al. (2011) proposed a method that sequentially allocates the computational budget in the inner simulation based on marginal impact of each scenario on the risk estimator, thereby improving the estimation of large-loss probabilities. Furthermore, in contrast to earlier work requiring both the inner- and outer-level samples to grow indefinitely, Sun et al. (2011) showed that a "one-and-a-half-level" simulation approach is sufficient to estimate the variance of a conditional expectation with arbitrarily high precision with fixed inner-level computational effort. To accelerate the convergence of nested simulations, Liang et al. (2024) introduced a jackknife-based nested simulation method. Additionally, Feng and Song (2025) proposed a nested simulation experiment design that fully

utilizes inner samples by pooling them across outer scenarios and employing the likelihood ratio method to estimate the conditional mean of other outer scenarios, thereby improving computationally efficiency.

While much of the existing literature focuses on point estimation of risk measures, interval estimation is equally important for risk management. Confidence interval (CI) provides a range of plausible values for risk measures, enabling risk managers to better understand the uncertainty associated with their estimates. For constructing CIs for conditional Value-at-Risk (CVaR), Lan et al. (2010) proposed a nested simulation procedure that integrates the theory of empirical likelihood with tools from ranking and selection literature, significantly improving estimation efficiency. More recently, Zhang et al. (2022) proposed a sample-driven budget allocation rule by bootstrap sampling, and further established a central limit theorem (CLT) for point estimators of risk measures, which provides a theoretical foundation for constructing CIs.

In contrast to nested simulation, another stream of literature focuses on leveraging metamodeling techniques to approximate the unknown functional form of the portfolio loss function. This approach is motivated by the observation that, in real-world applications, loss functions are usually smooth and continuous, making them well-suited for approximation using metamodels. Common metamodeling techniques in this area include stochastic kriging (Liu and Staum 2010), least-squares method (Broadie et al. 2015), kernel smoothing (Hong et al. 2017), and kernel ridge regression (Wang et al. 2024). Beyond point estimates, Lai et al. (2024) proposed a novel least-squares method to construct CIs for CVaR. Specifically, their method first develops lower and upper bounds of CVaR, constructs CIs for these bounds, and combines the lower end of the CI for the lower bound and the upper end of the CI for the upper bound to form a CI of CVaR with statistical guarantees. For readers interested in a comprehensive overview of nested simulation problems and methods, Liu and Zhang (2024) provided an in-depth tutorial and review of the relevant literature.

The construction of CIs for VaR remains relatively underexplored. A notable work in this area is Zhang et al. (2022), who established a CLT for VaR point estimators and proposed a budget allocation rule for constructing CIs, by requiring the inner sample size to grow faster than the square root of the outer sample size to ensure asymptotic validity of the CIs. Recently, Yi and Xie (2017) constructed percentile-based CIs to quantify the impact of the input uncertainty on the system performance estimates. Specifically, they developed a CI for the percentile itself, which effectively corresponds to constructing a CI for VaR. Notably, they highlighted the challenges of constructing metamodels to propagate input uncertainty to the output could be challenging, which is why direct inner-level simulation is employed in their article. In addition, Zhu et al. (2020) provided inference for extreme scenarios of mean responses across all possible input models by establishing asymptotic normality for the proposed nested risk estimators, allowing the construction of CIs.

Constructing CIs for VaR is particularly challenging due to three reasons. First, VaR focuses on extreme events, which are inherently rare and thus difficult to capture accurately in simulations. Second, the nested simulation structure introduces inner-level simulation errors that can significantly reduce the accuracy of VaR estimates. Third, providing rigorous theoretical guarantees for these CIs, such as ensuring coverage probabilities, is nontrivial under the nested framework. In this paper, we address these challenges and propose a procedure for constructing CIs for VaR using nested simulation. The key components of our procedure include a screening process that retains only a small subset of scenarios likely to result in high portfolio losses, and estimation procesures of the minimum and maximum portfolio losses within these selected scenarios.

The remainder of the paper is organized as follows. Section 2 presents the CI construction procedure for VaR. Section 3 provides theoretical guarantees for the proposed procedure. Section 4 reports numerical results, and Section 5 concludes the paper.

2 A PROCEDURE FOR VAR CONFIDENCE INTERVALS

Let the risk scenario $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top \in \mathbb{R}^d$ be generated independently from the distribution of $\mathbf{Z}$. Further, the portfolio loss $X \in \mathbb{R}$ is generated independently from conditional distribution of $X \mid \mathbf{Z}$. Define

the conditional expectation as

$$V(\mathbf{Z}) = \mathbb{E}[X \mid \mathbf{Z}],$$

which represents the portfolio loss function given a risk scenario $\mathbf{Z}$ with the expectation taken under the risk-neutral pricing measure.

Our objective is to construct a CI for VaR at $1-p$ level of $V(\mathbf{Z})$ with $p \in (0,1)$, defined as

$$\mathrm{VaR}_{1-p}(V(\mathbf{Z})) = \inf\{x \in \mathbb{R} : \mathbb{P}(V(\mathbf{Z}) \le x) \ge 1-p\}.$$

The key idea behind the proposed procedure is to screen out scenarios that are unlikely to result in portfolio losses close to the true VaR, and then allocate the majority of the simulation budget to these retained scenarios. CIs are subsequently established by estimating the minimum and maximum losses within these retained scenarios.

Without loss of generality, assume that the cost of simulating an inner sample is normalized to 1. The proposed procedure takes as input the total simulation budget Γ, the VaR confidence level $1-p$, and the CI significance level $1-\alpha$, where $\alpha \in (0,1)$, and returns the confidence interval $\left[\hat{L}, \hat{U}\right]$. The proposed procedure is outlined as follows.

1. **Outer Scenario Generation.** Generate n scenarios $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$.

2. **First-Stage Inner Sampling.** For each scenario $\mathbf{Z}_i$, $i = 1, \cdots, n$, simulate m' loss samples $X_{i,1}, X_{i,2}, \ldots, X_{i,m'}$ using Common Random Number techniques, and calculate the sample mean $\overline{X}_{i,m'} = \sum_{j=1}^{m'} X_{i,j}/m'$ and the sample variance $S^2_{i,m'} = \sum_{j=1}^{m'} \left(X_{i,j} - \overline{X}_{i,m'}\right)^2/(m'-1)$.

3. **Scenario Screening.**
   (1) For each scenario pair $(\mathbf{Z}_i, \mathbf{Z}_j)$, $i, j = 1, \cdots, n$, the pair difference is defined as

   $$D_{i,j} = V(\mathbf{Z}_i) - V(\mathbf{Z}_j).$$

   Samples of the paired differences are computed as $D_{i,j,l} = X_{i,l} - X_{j,l}$, $l = 1, \cdots, m'$. Calculate the sample mean $\overline{D}_{i,j} = \sum_{l=1}^{m'} D_{i,j,l}/m'$ and the sample variance $S^2_{D_{i,j}} = \sum_{l=1}^{m'} \left(D_{i,j,l} - \overline{D}_{i,j}\right)^2/(m'-1)$.

   (2) Let $k_{\min}$ and $k_{\max}$ be the minimum and maximum elements of the set

   $$\left\{k : n^n \left(\frac{1-p}{k}\right)^k \left(\frac{p}{n-k}\right)^{n-k} \ge \exp\left(-\frac{1}{2}\chi^2_{(1),1-\alpha_{\text{out}}}\right)\right\}. \tag{1}$$

   For $i, j = 1, \cdots, n$, define the test statistic as $T_{i,j} = \sqrt{m'}\,\overline{D}_{i,j}/S_{D_{i,j}}$, and

   $$B_i = \sum_{j=1, j\neq i}^{n} \mathbf{1}\{T_{i,j} > d_1\} \text{ and } B'_i = \sum_{j=1, j\neq i}^{n} \mathbf{1}\{T_{i,j} < -d_2\}, \tag{2}$$

   where $\mathbf{1}\{\cdot\}$ is an indicator function, which equals 1 if the condition is true and 0 otherwise. Define $l_1 = (k_{\max}+1)(n-k_{\max}-1)$, $l_2 = (k_{\min}-1)(n-k_{\min}+1)$, $\alpha_{\text{screen1}} = \alpha_{\text{screen}}\, l_1/(l_1+l_2)$, and $\alpha_{\text{screen2}} = \alpha_{\text{screen}}\, l_2/(l_1+l_2)$, then $d_1 = t_{m'-1,1-\alpha_{\text{screen1}}/l_1}$ and $d_2 = t_{m'-1,1-\alpha_{\text{screen2}}/l_2}$, where $t_{m'-1,1-\beta}$ is the $1-\beta$ quantile of the t-distribution with $m'-1$ degrees of freedom.

   (3) For $i = 1, \cdots, n$, if

   $$B_i < k_{\max}+1 \text{ and } B'_i < n - k_{\min} + 1,$$

then include index i in the set of surviving scenario indices, denoted by $\mathcal{I}$. Additionally, we denote $c = \text{card}(\mathcal{I})$, where $\text{card}(\cdot)$ is the cardinality of a set.

4. **Restarting and Second-Stage Inner Sampling.** Define $\mathbf{m} = (m_i)^\top \in \mathbb{R}^c$, $i \in \mathcal{I}$, where m_i denotes the number of second-stage inner samples for scenario i as

$$m_i = \left\lfloor (\Gamma - nm') \frac{S_{i,m'}^2}{\sum_{j\in\mathcal{I}} S_{j,m'}^2} \right\rfloor.$$

For each surviving scenario $\mathbf{Z}_i$, $i \in \mathcal{I}$, simulate m_i new loss samples $X_{i,1}, \ldots, X_{i,m_i}$, and calculate the sample mean $\overline{X}_{i,m_i}$ and the sample variance $S_{i,m_i}^2 = \sum_{j=1}^{m_i} \left(X_{i,j} - \overline{X}_{i,m_i}\right)^2 / (m_i - 1)$.

5. **Constructing VaR Confidence Interval.** For convenience, we introduce a permutation π_1 mapping $\{1, \cdots, c\}$ to $\mathcal{I}$, which orders indices according to ascending values of their sample means. Specifically, the permutation satisfies $\overline{X}_{\pi_1(1),m_{\pi_1(1)}} \leq \cdots \leq \overline{X}_{\pi_1(c),m_{\pi_1(c)}}$. Moreover, for instance, if $\overline{X}_{c,m_c} = \min_{i\in\mathcal{I}} \left\{\overline{X}_{i,m_i}\right\}$, then $\pi_1(1) = c$. Calculate

$$\hat{L} = \overline{X}_{\pi_1(1),m_{\pi_1(1)}} - z_{1-\frac{\alpha_{\text{est}}}{2}} \frac{S_{\pi_1(1),m_{\pi_1(1)}}}{\sqrt{m_{\pi_1(1)}}} \text{ and } \hat{U} = \overline{X}_{\pi_1(c),m_{\pi_1(c)}} + z_{1-\frac{\alpha_{\text{est}}}{2}} \frac{S_{\pi_1(c),m_{\pi_1(c)}}}{\sqrt{m_{\pi_1(c)}}},$$

where $z_{1-\alpha_{\text{est}}/2}$ is the $1-\alpha_{\text{est}}/2$ quantile of the standard normal distribution. Finally, the confidence interval is $\left[\hat{L}, \hat{U}\right]$.

To provide intuition, we offer a detailed explanation of the scenario screening step. The goal of this step is to retain extreme scenarios that are crucial for VaR estimation by focusing on those ranked ascendingly between $k_{\min}$ and $k_{\max}+1$. The range $[k_{\min}, k_{\max}+1]$ comes from (1). In the equation, k indicates the smallest number of scenarios whose cumulative probability reaches $1-p$. A likelihood-ratio test applied to the empirical data then retains statistically plausible k values. To produce the range, we employ the pairwise comparison tool, a method commonly used in the ranking and selection literature; see Hong et al. (2021) for a detailed introduction. Specifically, for each pair of scenarios, we compute a statistic $T_{i,j}$. If $T_{i,j} > d_1$, it indicates $V_i > V_j$ with statistical significance, while $T_{i,j} < -d_2$ suggests that $V_i < V_j$ statistically. Then, we introduce two counters. The counter B_i records the number of scenario i is statistically larger. The criterion $B_i < k_{\max}+1$ identifies scenarios that belong to the smallest $k_{\max}+1$ scenarios statistically. Similarly, B_i' counts the number of scenarios for which scenario i is statistically smaller. The criterion $B_i' < n - k_{\min} + 1$ ensures that scenario i is among the largest $n - k_{\min} + 1$ scenarios statistically. By combining two criteria, the scenario screening step retains the essential scenarios needed to capture the VaR accurately. Furthermore, when the number of scenarios is large, we provide enhancement strategies to implement large-scale pairwise scenario comparisons. The details of these strategies are deferred to Appendix A. It is worth noting that an important issue requiring dedicated study is the selection of key parameters, such as n and m', which we leave as a direction for future research.

3 THEORETICAL GUARANTEES

For notational convenience, we denote $V(\mathbf{Z})$ as V, $V(\mathbf{Z}_i)$ as V_i, $\text{VaR}_{1-p}(V(\mathbf{Z}))$ as VaR_{1-p}. Let F_V be the distribution of V, and $V_1, \cdots, V_n$ are independent and identically distributed from F_V. Let F_n be the empirical distribution supported on $V_1, \cdots, V_n$. $F_V(\cdot)$ is the cumulative probability function of distribution F_V. To facilitate analysis for the coverage probability of constructed CI, we make the following assumptions.

Assumption 1 F_V is continuous and strictly increasing at VaR_{1-p}.

Assumption 2 For any scenario $\mathbf{Z}_i, i = 1, \cdots, n$, loss samples $X_{i,j}$, $j = 1, \cdots, m'$, are normally distributed.

Assumption 1 is a standard regularity condition in analyzing the properties of VaR. Assumption 2 is commonly used in the ranking and selection literature, and has also been used in works such as Lan et al. (2010) and Yi and Xie (2017). Although payoffs may not be normally distributed in practice, sample averages of payoffs are approximately normal due to CLT, provided that the sample size m' is sufficienty large.

Lemma 1 (*M*-estimate for VaR, Proposition 5.1 in Baysal and Staum (2008))**.** *Let $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a measurable function. An M-estimate is a statistical functional defined as a solution $q = T_\psi(F_V)$ of*

$$\int \psi(V,q)\mathrm{d}F_V = 0. \tag{3}$$

Letting

$$\psi(V,q) = (1-p) - \mathbf{1}\{V \le q\}, \tag{4}$$

under Assumption 1, $T_\psi(F_V)$, the solution to (3)*, is an M-estimate for* VaR_{1-p}.

Proof. Plugging (4) into (3), we have $(1-p) - F_V(q) = 0$, which implies $F_V(q) = 1-p$. Assumption 1 indicates that F_V has a unique $1-p$ quantile that $F_V(\mathrm{VaR}_{1-p}) = 1-p$, which implies that a unique solution of (3) is VaR_{1-p}. □

Lemma 1 characterizes VaR as an *M*-estimate by linking it explicitly to a statistical estimating equation. Specifically, it identifies VaR as a unique solution that equates the probability to desired risk level, captured by the chosen function ψ. This perspective reveals that estimating VaR is basically about finding a distributional threshold satisfying the balance condition, bridging VaR estimation with general statistical inference methods.

Lemma 2. *Consider the set of discrete distributions absolutely continuous with respect to F_n, denoted by $\{F : F \ll F_n\}$. Each distribution F in the set assigns weights $w_1, \ldots, w_n$ to the points $V_1, \ldots, V_n$, satisfying $w_i \ge 0, \sum_{i=1}^n w_i = 1$. Denote the empirical likelihood ratio (ELR) of F as $R(F) = \prod_{i=1}^n (nw_i)$. Suppose Assumption 1 holds. Then, as $n \to \infty$, a confidence interval for* VaR_{1-p} *with* $1-\alpha_{\text{out}}$ *asymptotic coverage probability is given by*

$$\left\{ T_\psi(F) : F \ll F_n, R(F) \ge \exp\left(-\frac{1}{2}\chi^2_{(1),1-\alpha_{\text{out}}}\right)\right\}. \tag{5}$$

Proof. Under Assumption 1 and by Lemma 1, $T_\psi(F_V)$ exists and is unique. Note that $\mathrm{var}[\psi(V,t)] = F_V(t)(1-F_V(t)) = F_V(\mathrm{VaR}_{1-p})(1-F_V(\mathrm{VaR}_{1-p}))$, where $\mathrm{var}[\cdot]$ denotes the variance operator. Then, $p \in (0,1)$ implies that $\mathrm{var}[\psi(V,t)] \neq 0$, and the rank of $\mathrm{var}[\psi(V,t)]$ is one. Therefore, we apply Theorem 3 in Owen (1990) to show that if we pick $r = \exp\left(-\chi^2_{(1),1-\alpha_{\text{out}}}/2\right)$, then $\mathbb{P}\left(T_\psi(F_V) \notin \left\{T_\psi(F) : F \ll F_n, R(F) \ge r\right\}\right) \to \alpha_{\text{out}}$, as $n \to \infty$. □

Lemma 2 constructs a confidence interval for VaR based on empirical likelihood. Intuitively, it contains all VaR estimates from distributions sufficiently close to the empirical data (i.e., with large ELR values). Based on Lemma 2, we derive an explicit CI expression in the following proposition. For convenience, we introduce a permutation π_2, which maps indices $\{1, \cdots, n\}$ to a permuted order where scenarios are sorted by their true means in ascending order, i.e., $V_{\pi_2(1)} \le V_{\pi_2(2)} \le \cdots \le V_{\pi_2(n)}$.

Proposition 1. *Under Assumption 1, given the outer-level error probability $\alpha_{\text{out}} \in (0,1)$ for finite outer scenarios,*

$$\mathbb{P}\left(\mathrm{VaR}_{1-p} \in [V_{\pi_2(k_{\min})}, V_{\pi_2(k_{\max}+1)}]\right) \to 1-\alpha_{\text{out}}, \text{ as } n \to \infty.$$

Proof.

$$\begin{aligned}
&\text{Equation (5)} = \left\{ q : \int \psi(V,q)\,\mathrm{d}F_V = 0, F \ll F_n, R(F) \geq \exp\left(-\frac{1}{2}\chi^2_{(1),1-\alpha_{\text{out}}}\right)\right\} \\
&= \left\{ q : \sum_{i=1}^{n} w_i \mathbf{1}\{V_{\pi_2(i)} \leq q\} = 1-p, w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \prod_{i=1}^{n}(nw_i) \geq \exp\left(-\frac{1}{2}\chi^2_{(1),1-\alpha_{\text{out}}}\right)\right\} \\
&= \bigcup_{k=1}^{n-1} \left\{ q : \sum_{i=1}^{n} w_i \mathbf{1}\{V_{\pi_2(i)} \leq q\} = 1-p, \sum_{i=1}^{k} w_i = 1-p, w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \prod_{i=1}^{n}(nw_i) \geq \exp\left(-\frac{1}{2}\chi^2_{(1),1-\alpha_{\text{out}}}\right)\right\} \\
&= \bigcup_{k=k_{\min}}^{k_{\max}} \left\{ q : \sum_{i=1}^{n} w_i \mathbf{1}\{V_{\pi_2(i)} \leq q\} = 1-p, \sum_{i=1}^{k} w_i = 1-p, w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \prod_{i=1}^{n}(nw_i) \geq \exp\left(-\frac{1}{2}\chi^2_{(1),1-\alpha_{\text{out}}}\right)\right\} \\
&= \bigcup_{k=k_{\min}}^{k_{\max}} [V_{\pi_2(k)}, V_{\pi_2(k+1)}) = [V_{\pi_2(k_{\min})}, V_{\pi_2(k_{\max}+1)}).
\end{aligned}$$

□

Proposition 1 suggests that, as the number of outer-level scenarios is sufficiently large, the empirical likelihood method guarantees that true VaR lies within the constructed CI with the desired coverage probability. Specifically, this result highlights a critical insight that empirical likelihood inherently generates quantile-based intervals. Consequently, the quantile-based characteristic motivates the subsequent methodological developments.

Proposition 2. *Let* $\gamma = \{\pi_2(k_{\min}), \cdots, \pi_2(k_{\max}+1)\}$. *Under Assumption 2, given the screening error probability* $\alpha_{\text{screen}} \in (0,1)$*,*

$$\mathbb{P}(\gamma \subseteq \mathcal{I}) \geq 1 - \alpha_{\text{screen}}.$$

Proof. Let $\gamma_1 = \{\pi_2(1), \pi_2(2), \cdots, \pi_2(k_{\max}+1)\}$, $\gamma_2 = \{\pi_2(k_{\min}), \pi_2(k_{\min}+1), \cdots, \pi_2(n)\}$, and thus $\gamma = \gamma_1 \cap \gamma_2$. Let $\mathcal{I}_1 = \{i : B_i < k_{\max}+1\}$, $\mathcal{I}_2 = \{i : B'_i < n - k_{\min}+1\}$, and thus $\mathcal{I} = \mathcal{I}_1 \cap \mathcal{I}_2$. In Lan et al. (2010), they have proven that, under Assumption 2, $\mathbb{P}(\gamma_1 \subseteq \mathcal{I}_1) \geq 1 - \alpha_{\text{screen1}}$. Similarly, we have $\mathbb{P}(\gamma_2 \subseteq \mathcal{I}_2) \geq 1 - \alpha_{\text{screen2}}$. Therefore,

$$\mathbb{P}(\gamma \subseteq \mathcal{I}) = \mathbb{P}(\gamma_1 \cap \gamma_2 \subseteq \mathcal{I}_1 \cap \mathcal{I}_2) \geq \mathbb{P}(\gamma_1 \subseteq \mathcal{I}_1, \gamma_2 \subseteq \mathcal{I}_2) \geq 1 - (\alpha_{\text{screen1}} + \alpha_{\text{screen2}}) = 1 - \alpha_{\text{screen}}.$$

□

Proposition 2 confirms the effectiveness of screening process for retaining crucial scenarios stated in Proposition 1, and provides the probability guarantee of not missing any elements in γ after screening.

For convenience of presentation, we introduce another permutation π_3 mapping $\{1, \cdots, c\}$ to $\mathcal{I}$, which orders indices according to ascending values of their true means, i.e., $V_{\pi_3(1)} \leq V_{\pi_3(2)} \leq \cdots \leq V_{\pi_3(c)}$. Typically, the difference between permutations π_1 and π_3 is that π_1 ranks surviving scenarios based on ascending sample means, whereas π_3 ranks scenarios based on ascending true means.

Proposition 3. *Suppose V is a continuous random variable, given the estimation error probability* $\alpha_{\text{est}} \in (0,1)$ *at the second-stage estimation,*

$$\mathbb{P}\left(V_i \in \left[\hat{L}, \hat{U}\right], \forall i \in \mathcal{I}\right) \geq 1 - \alpha_{\text{est}}, \text{ as } \mathbf{m} \to \infty,$$

where $\mathbf{m} \to \infty$ *means that each component of the vector* $\mathbf{m}$ *tends to infinity.*

Proof. Suppose that

$$\lim_{\mathbf{m}\to\infty} \mathbb{P}\left(V_{\pi_3(1)} < \hat{L}\right) \leq \frac{\alpha_{\text{est}}}{2} \text{ and } \lim_{\mathbf{m}\to\infty} \mathbb{P}\left(V_{\pi_3(c)} > \hat{U}\right) \leq \frac{\alpha_{\text{est}}}{2}.$$

Then, $\forall i \in \mathcal{I}$, we have

$$\begin{aligned}\lim_{\mathbf{m}\to\infty} \mathbb{P}\left(\hat{L} \leq V_i \leq \hat{U}\right) &\geq \lim_{\mathbf{m}\to\infty} \mathbb{P}\left(\hat{L} \leq V_{\pi_3(1)} \leq V_i \leq V_{\pi_3(c)} \leq \hat{U}\right) = \lim_{\mathbf{m}\to\infty} \mathbb{P}\left(\hat{L} \leq V_{\pi_3(1)} \cap V_{\pi_3(c)} \leq \hat{U}\right) \\ &\geq \lim_{\mathbf{m}\to\infty} \mathbb{P}\left(\hat{L} \leq V_{\pi_3(1)}\right) + \lim_{\mathbf{m}\to\infty} \mathbb{P}\left(V_{\pi_3(c)} \leq \hat{U}\right) - 1 \\ &\geq (1 - \frac{\alpha_{\text{est}}}{2}) + (1 - \frac{\alpha_{\text{est}}}{2}) - 1 = 1 - \alpha_{\text{est}}.\end{aligned}$$

Therefore, our goal is to construct a lower confidence limit $\hat{L}$ for $V_{\pi_3(1)}$ and an upper confidence limit $\hat{U}$ for $V_{\pi_3(c)}$. To achieve this, we analyze the asymptotic distribution

$$\begin{aligned}&\mathbb{P}\left(\sqrt{m_{\pi_1(1)}}\left(\overline{X}_{\pi_1(1),m_{\pi_1(1)}} - V_{\pi_3(1)}\right) > y\right) \leq \mathbb{P}\left(\sqrt{m_{\pi_1(1)}}\left(\overline{X}_{\pi_3(1),m_{\pi_3(1)}} - V_{\pi_3(1)}\right) > y\right) \\ =&\mathbb{P}\left(\frac{\sqrt{m_{\pi_3(1)}}\left(\overline{X}_{\pi_3(1),m_{\pi_3(1)}} - V_{\pi_3(1)}\right)}{\sigma_{\pi_3(1)}} > \frac{\sqrt{N_{2,\pi_3(1)}}}{\sqrt{m_{\pi_1(1)}}}\frac{y}{\sigma_{\pi_3(1)}}\right) \to 1 - \Phi\left(\frac{\sqrt{N_{2,\pi_3(1)}}}{\sqrt{m_{\pi_1(1)}}}\frac{y}{\sigma_{\pi_3(1)}}\right), \text{ as } \mathbf{m}\to\infty,\end{aligned}$$

where $\sigma_{\pi_3(1)}$ is the unknown standard deviation of scenario $\pi_3(1)$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The convergence in the last step follows from CLT,

$$\frac{\sqrt{m_{\pi_3(1)}}\left(\overline{X}_{\pi_3(1),m_{\pi_3(1)}} - V_{\pi_3(1)}\right)}{\sigma_{\pi_3(1)}} \xrightarrow{d} N(0,1), \text{ as } \mathbf{m}\to\infty.$$

Let $y = z_{1-\alpha_{\text{est}}/2}\sqrt{m_{\pi_1(1)}}\sigma_{\pi_3(1)}/\sqrt{m_{\pi_3(1)}}$, then $1 - \Phi\left(y\sqrt{N_{2,\pi_3(1)}}/\left(\sqrt{m_{\pi_1(1)}}\sigma_{\pi_3(1)}\right)\right) = \alpha_{\text{est}}/2$. Denote $S_{\pi_3(1),m_{\pi_3(1)}}$ as the sample standard deviation of scenario $\pi_3(1)$, and $S_{\pi_3(1),m_{\pi_3(1)}} \to \sigma_{\pi_3(1)}$ as $\mathbf{m}\to\infty$. Therefore, by Slutsky's theorem,

$$\lim_{\mathbf{m}\to\infty} \mathbb{P}\left(\sqrt{m_{\pi_1(1)}}\left(\overline{X}_{\pi_1(1),m_{\pi_1(1)}} - V_{\pi_3(1)}\right) > z_{1-\frac{\alpha_{\text{est}}}{2}}\frac{\sqrt{m_{\pi_1(1)}}}{\sqrt{m_{\pi_3(1)}}}S_{\pi_3(1),m_{\pi_3(1)}}\right) \leq \frac{\alpha_{\text{est}}}{2}.$$

Rearranging the terms inside the probability, it becomes

$$\lim_{\mathbf{m}\to\infty} \mathbb{P}\left(V_{\pi_3(1)} < \overline{X}_{\pi_1(1),m_{\pi_3(1)}} - z_{1-\frac{\alpha_{\text{est}}}{2}}\frac{S_{\pi_3(1),m_{\pi_3(1)}}}{\sqrt{m_{\pi_3(1)}}}\right) \leq \frac{\alpha_{\text{est}}}{2}.$$

Noting that $\overline{X}_{i,m_i} \to V_i$ for $\forall i \in \mathcal{I}$ as $\mathbf{m}\to\infty$, it is easy to show that $\pi_1(1) \to \pi_3(1)$. Thus, by Slutsky's theorem, we show that the lower confidence limit for $V_{\pi_3(1)}$ can be constructed as $\hat{L}$. Similarly, we construct an upper confidence limit for $V_{\pi_3(c)}$ as $\hat{U}$. □

Proposition 3 ensures that, given sufficient samples at the second-stage estimation, the constructed CI simultaneously covers true means of surviving scenarios with statistical guarantees.

Theorem 1 Under Assumptions 1 and 2, suppose V is continuous, given a total error probability α,

$$\mathbb{P}(\text{VaR}_{1-p} \in \left[\hat{L}, \hat{U}\right]) \geq 1 - \alpha, \text{ as } n\to\infty, \mathbf{m}\to\infty.$$

Proof. The overall error probability can be decomposed as $\alpha = \alpha_{\text{out}} + \alpha_{\text{screen}} + \alpha_{\text{est}}$.

$$\begin{aligned}
\mathbb{P}\left(\text{VaR}_{1-p} \in [\hat{L}, \hat{U}]\right) &\geq \mathbb{P}\left(\text{VaR}_{1-p} \in [V_{\pi_2(k_{\min})}, V_{\pi_2(k_{\max}+1)}] \cap \gamma \subseteq \mathcal{I} \cap V_i \in [\hat{L}, \hat{U}], \forall i \in \mathcal{I}\right) \\
&\geq \mathbb{P}\left(\text{VaR}_{1-p} \in [V_{\pi_2(k_{\min})}, V_{\pi_2(k_{\max}+1)}]\right) + \mathbb{P}\left(\gamma \subseteq \mathcal{I}\right) + \mathbb{P}\left(V_i \in [\hat{L}, \hat{U}], \forall i \in \mathcal{I}\right) - 2 \\
&\geq (1 - \alpha_{\text{out}}) + (1 - \alpha_{\text{screen}}) + (1 - \alpha_{\text{est}}) - 2 = 1 - \alpha.
\end{aligned}$$

□

Theorem 1 shows that the proposed CI procedure achieves the asymptotic confidence level.

4 NUMERICAL EXPERIMENTS

Consider a portfolio risk measurement example to study the performance of the proposed procedure for constructing CI for VaR, using a portfolio comprised of multiple options written on a single underlying asset. The underlying asset price dynamics, denoted by $S(t) \in \mathbb{R}$, is modeled as a geometric Brownian motion with drift μ' and volatility σ, given by

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu' \mathrm{d}t + \sigma \mathrm{d}B(t),$$

where $B(t)$ is a standard Brownian motion.

In this model, the asset price at time t is explicitly given by

$$S(t) = S(0) \exp\left\{\left(\mu' - \frac{1}{2}\sigma^2\right)t + \sigma B(t)\right\}.$$

Note that μ' is chosen to be the real-world return μ under the real-world probability measure during the risk horizon $[0, \tau]$, and the risk-free rate r under the risk-neutral measure for the interval $(\tau, T]$, with τ and T denoting the risk horizon and the maturity of the options, respectively. We set $S(0) = 100$, $\mu = 0.08$, $r = 0.05$, and $\sigma = 30\%$.

The portfolio comprises five European call options, all maturing at a common time T, with distinct strike prices $K_1 = 80$, $K_2 = 90$, $K_3 = 100$, $K_4 = 110$, and $K_5 = 120$. We denote the current time by 0 and are interested in measuring the portfolio risk at a given risk horizon $\tau < T$. We set $T = 1/12$ year, that is, 1 month, and $\tau = 1/52$ year, that is, 1 week. The current portfolio value, denoted by $C(0)$, is explicitly computed using the Black-Scholes formula. At time τ, the portfolio value $C(\tau)$ is defined as

$$C(\tau) = \mathbb{E}\left[\sum_{l=1}^{5} \mathrm{e}^{-r(T-\tau)} \left(S(T) - K_l\right)^+ \middle| S(\tau)\right].$$

Accordingly, the portfolio loss at the risk horizon τ is given by

$$V(S(\tau)) = \mathbb{E}\left[C(0) - \sum_{l=1}^{5} \mathrm{e}^{-r(T-\tau)} \left(S(T) - K_l\right)^+ \middle| S(\tau)\right].$$

In our numerical example, we set $p = 1\%$ and $\alpha = 10\%$. To evaluate the performance of the procedure, we need an accurate benchmark value for VaR_{1-p}. To achieve this, we generate a large number (10^8) of simulated values for $S(\tau)$, compute the corresponding portfolio losses using the explicit Black-Scholes formula, and accurately estimate VaR_{1-p} as 20.615.

When implementing the procedure, we heuristically set $n = 1.5\Gamma^{2/3}$, initialize m' at 10, and increment it by 5 until either $c - c' < 0.1\% c'$ or $(\Gamma - m'n)/c < 30$, where $c' = \text{card}(\gamma)$. The rationale behind this allocation

is twofold. First, our goal is to generate a sufficient number of outer scenarios to satisfy the coverage probability requirement established in Proposition 1. Second, we aim to screen out as many unpromising scenarios as possible to reduce the computational burden during the second-stage estimation. Specifically, the first stopping criterion ensures that the difference between c and c' does not exceed $0.1\%\,c'$, thereby limiting the number of scenarios entering the second-stage estimation. The second criterion guarantees an adequate sample size for the second-stage estimation, with a threshold of 30 samples aligns well with a commonly accepted minimum sample size justified by CLT, and is consistent with Proposition 3. Additionally, the initial value for m' (set at 10) and the incremental step size (chosen as 5) are selected arbitrarily. These parameters primarily influence the number of iterations required but have negligible impact on the final results, which are primarily governed by the stopping criteria. Furthermore, following the guidance of Lan et al. (2010) and Yi and Xie (2017), which indicate that the performance of the procedure is relatively insensitive to the choice of α_{out}, α_{screen}, and α_{est}, we set these parameters as 6%, 1%, and 3%, respectively.

To examine the coverage probability of the constructed CI and verify consistency with the theoretical result provided in Theorem 1, we estimate the coverage probabilities (CPs), as the ratio of the number of times that the true value is covered by the CI relative to the number of macro replications. The experiments are conducted through 500 independent macro replications. To comprehensively evaluate the performance of the CIs, we assess metrics beyond the coverage probabilities, including the average width ratio and its standard deviation (AWR and SD in shorthand notation), the average lower endpoint and its standard deviation (AL and SD), and the average upper endpoint and its standard deviation (AU and SD). The width ratio is calculated as the ratio of the CI width to the true value, providing a relative measure of interval length.

To further investigate the effectiveness of our proposed CI, we compare its performance metrics to those obtained from the bootstrap method introduced by Zhang et al. (2022). In their method, a user-specified parameter $\varepsilon \in (0, 2/3)$ is required. Typically, bootstrap methods with larger ε allocate more budget to inner simulations, while smaller ε allocates more to outer simulations. To evaluate the performance of the bootstrap-based CIs under different budget allocations, we set $\varepsilon = 1/24, 1/12$, and $1/6$ in our experiments. Morerover, to assess the computational efficiency of different methods, we record the average computation time per macro replication in minutes. All computer programs are coded in Python, and all experiments are run on a Linux server of Ubuntu 18.04 with two Intel Xeon E5-2650v4 processors (each with 12 physical cores, 24 logical threads, 2.20 GHz) and 128 GB RAM. In the implementation, we parallelize the screening step in our proposed method using the strategy described in Appendix A.

Since the theoretical result describes the asymptotic coverage probability of CIs, it is essential to evaluate their performance under different budget sizes to examine their validity. To this end, we conduct experiments using budgets $\Gamma = 5 \times 10^5$ and 5×10^6, and report the results in Table 1. In the table, metric values obtained from our procedure are highlighted in bold, as are bootstrap results with coverage probabilities closest to the nominal 90% level.

Table 1: Performance comparison of our method and bootstrap method.

Budget	Method	CP	AWR (SD)	AL (SD)	AU (SD)	Time
5×10^5	Ours	**1.000**	**0.093 (0.010)**	**19.668 (0.258)**	**21.594 (0.256)**	**0.133**
	Bootstrap ($\varepsilon = 1/24$)	0.872	0.089 (0.032)	19.726 (0.873)	21.561 (0.696)	0.249
	Bootstrap ($\varepsilon = 1/12$)	**0.894**	**0.123 (0.046)**	**19.425 (0.911)**	**21.951 (1.026)**	**0.229**
	Bootstrap ($\varepsilon = 1/6$)	0.853	0.197 (0.085)	18.622 (1.348)	22.680 (1.784)	0.223
5×10^6	Ours	**1.000**	**0.040 (0.003)**	**20.209 (0.114)**	**21.032 (0.122)**	**1.751**
	Bootstrap ($\varepsilon = 1/24$)	0.856	0.045 (0.013)	20.225 (0.309)	21.156 (0.365)	0.708
	Bootstrap ($\varepsilon = 1/12$)	**0.884**	**0.063 (0.019)**	**20.012 (0.439)**	**21.303 (0.461)**	**0.673**
	Bootstrap ($\varepsilon = 1/6$)	0.862	0.117 (0.037)	19.515 (0.818)	21.920 (0.817)	0.634

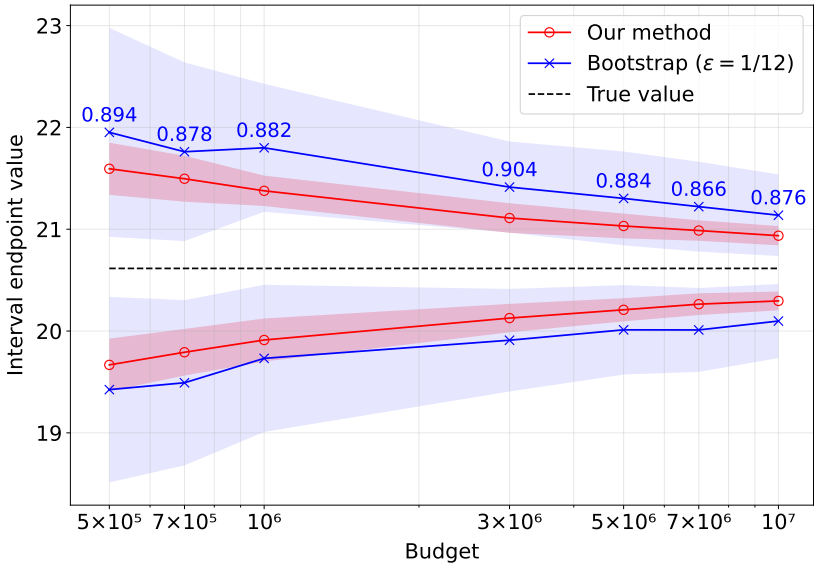

Figure 1: Constructed CIs using our method and bootstrap ($\varepsilon = 1/12$) with varying budgets.

From Table 1, it can be observed that our procedure achieves 100% coverage probability. This clearly aligns with the theoretical guarantee ($\geq 90\%$), but significantly exceeds the nominal level. Such a high coverage suggests the CI constructed is conservative. The conservativeness primarily arises from the use of Bonferroni inequality in our analysis. Surprisingly, despite the higher coverage, our procedure produces narrower and more stable CIs compared to the bootstrap method. The stability of CI width results from large values of $m_{\pi_1(1)}$ and $m_{\pi_1(c)}$, because a sufficient computational budget remains available for the inner-level estimation after screening. Moreover, it can be seen that our method is faster than bootstrap methods at $\Gamma = 5 \times 10^5$ but slower at $\Gamma = 5 \times 10^6$. To better understand the empirical results of computational efficiency, we analyze the algorithmic time complexity. The total computation time for each method consists of two components. Specifically, the first is the sampling time, and the second is the time to calculate the auxiliary quantities according to the specific method design. Since the sampling time equals the time for one simulation multiplied by the budget size and thus is same for all methods, we analyze the second component of each method. In bootstrap methods, the leading term is $In\log(n)$, where I is the number of bootstrap replications and $n\log(n)$ comes from the standard time complexity of finding a quantile. In our procedure, the leading term is n^2m'/P, where P is the number of cores used in parallel screening. With sufficient cores, the wall-clock time reduces to N^2m', where N is a given block size parameter. To sum up briefly, we characterize the total computation time using $\Theta(\cdot)$; see Cormen et al. (2022) for a formal definition. Suppose the sampling time is C_0 times the duration of a machine instruction, where C_0 tends to be a large constant. The bootstrap methods run in $C_0\Gamma + In\log(n) = \Theta(\Gamma + n\log(n))$. Our method's runtime is $C_0\Gamma + n^2m'/P = \Theta(\Gamma + n^2)$ with P parallel cores. Furthermore, under full parallelization it becomes $C_0\Gamma + N^2m' = \Theta(\Gamma)$. In this example, C_0 is small, so the auxiliary time cost is considerable. When Γ is small, n is small, whereas $I = 500$ remains considerable, making the bootstrap methods slower. By contrast, for large Γ, n^2m'/P soon exceeds $In\log(n)$, so our method becomes slower.

Additionally, we observe several trends in bootstrap methods with varying values of ε. First, regarding interval width, we find that under a fixed computational budget, increasing ε results in wider width. This occurs because a larger ε reduces n, which increases the variance and consequently enlarges the CI. As the budget increases, the interval widths decrease significantly. Second, we find two distinct patterns in the coverage probabilities, depending on the size of ε. For smaller values of $\varepsilon = 1/24, 1/12$, increasing ε increases the inner-level sample size and decreases the outer-level sample size. According to Theorem 1 in Zhang et al. (2022), this reduces bias and increases variance, thereby increasing coverage probability.

However, for larger values of $\varepsilon = 1/6$, the coverage probability decreases despite the wider interval. This anomaly arises because, for large ε, the outer- and inner-level sample sizes become comparable in order. In this case, Proposition 2 in Gordy and Juneja (2010) implies that the bias does not vanish due to the non-negligible term $O(1/n)$. This residual bias reduces coverage probability. Third, we find that for large ε, the computation time declines slightly. This is because increasing ε decreases n and thus the term $In\log(n)$. Overall, the effectiveness of the bootstrap method depends on an appropriate choice of ε. An improper choice may significantly degrade the accuracy and reliability of the CIs.

We vary the budget sample size to examine how the quality of CIs improves as the budget increases. Following Lan et al. (2010), we consider CIs with width ratios smaller than 10% to be meaningful. Accordingly, we select budget sizes ranging from 5×10^5 to 10^7 to ensure that the CIs presented in the figure are meaningful. Moreover, it is observed from Table 1 that the bootstrap method with $\varepsilon = 1/12$ often achieves coverage probabilities closest to the nominal level. For better readability, Figure 1 plots the CIs constructed from our procedure and the bootstrap method with $\varepsilon = 1/12$. In this figure, the blue numbers above the lines represent the coverage probabilities of the bootstrap method under the corresponding budgets. Our procedure is not annotated, as it always achieves 100% coverage probabilities. To show the uncertainty of each interval endpoint, the shaded bands indicate $\pm$ empirical standard deviation around that endpoint. Similarly to the findings from Table 1, our procedure produces CIs characterized by overcoverage, narrower interval widths, and reduced variability compared to the bootstrap method.

5 CONCLUSIONS

In this paper, we have developed a procedure for constructing CIs for VaR using nested simulation. We conduct a theoretical analysis to confirm the asymptotic validity of the constructed CIs, ensuring that they achieve the desired coverage probabilities as the simulation budget increases. Numerical experiments demonstrate that our proposed procedure achieves robust coverage probabilities, often exceeding nominal levels, while producing narrower and less noisy CIs compared to the existing bootstrap-based method. These results highlight the reliability and computational efficiency of the proposed procedure in practice.

ACKNOWLEDGMENTS

The research of the second author is supported partially by the National Natural Science Foundation of China and the Research Grants Council of Hong Kong (RGC-HK), under the RGC-HK General Research Fund Projects 11508620 and 11508523 and Theme-based Research Project T32-615/24-R, and NSFC/RGC-HK Joint Research Scheme under project N_CityU 105/21.

A PROCEDURE IMPLEMENTATION DETAILS

First, the term n^n in (1) grows so rapidly that it easily overflows double precision (64-bit) floating-point numbers (hereafter 'double precision') as soon as $n \geq 200$. Yet many practical applications require n to be tens of thousands to achieve desired estimation precision. Therefore, we take natural logarithms on both sides of (1). After transformation, the term becomes $n\log(n)$, avoiding numerical overflow.

Second, when dealing with paired scenario comparisions, a naïve implementation would explicitly form a three-dimensional array $D_{i,j,l}$, $i, j = 1, \cdots, n$, $l = 1, \cdots, m'$, then build two $n \times n$ arrays of means and variances, and finally compute B_i and B'_i. However, these arrays require far more memory than the 16–32 GB typically available on a standard workstation. To illustrate, when $n = 30,000$ and $m' = 30$, $D_{i,j,l}$ in double precision already consumes roughly 200 GB memory. This memory explosion problem raises the need of enhancements for large-scale problems. Therefore, we design a block-decomposition strategy compatible with parallel processing paradigms. Specifically, we decompose the $n \times n \times m'$ full array into blocks of size $N \times N \times m'$, where N is called block size parameter. For instance, with $n = 30,000$ and $N = 1,000$, we get 900 blocks. Each block holds $1,000 \times 1,000 \times 30$ doubles, about 0.224 GB, instead of the 200 GB needed for the full huge array. Furthermore, by leveraging the antisymmetry of the pairwise

differences $D_{j,i,l} = -D_{i,j,l}$, we only process the upper-triangle blocks, reducing the total number from 900 to 465. Most importantly, because each block can be computed independently, this scheme parallelizes naturally. Moreover, the fourth step is easy to be parallelized by allocating independent scenarios to CPU cores or cluster nodes, running each inner sampling task in parallel, and merging the individual outputs to obtain the final estimates.

REFERENCES

Baysal, R. E., and J. Staum. 2008. "Empirical Likelihood for Value-at-Risk and Expected Shortfall". *Journal of Risk* 11(1):3.

Broadie, M., Y. Du, and C. C. Moallemi. 2011. "Efficient Risk Estimation via Nested Sequential Simulation". *Management Science* 57(6):1172–1194.

Broadie, M., Y. Du, and C. C. Moallemi. 2015. "Risk Estimation via Regression". *Operations Research* 63(5):1077–1097.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein. 2022. *Introduction to Algorithms*. Cambridge, MA: MIT press.

Feng, B. M., and E. Song. 2025. "Efficient Nested Simulation Experiment Design via the Likelihood Ratio Method". *INFORMS Journal on Computing* 37(3):723–742.

Gordy, M. B., and S. Juneja. 2010. "Nested Simulation in Portfolio Risk Measurement". *Management Science* 56(10):1833–1848.

Hong, L. J., W. Fan, and J. Luo. 2021. "Review on Ranking and Selection: A New Perspective". *Frontiers of Engineering Management* 8(3):321–343.

Hong, L. J., S. Juneja, and G. Liu. 2017. "Kernel Smoothing for Nested Estimation with Application to Portfolio Risk Measurement". *Operations Research* 65(3):657–673.

Lai, Q., G. Liu, B. Zhang, and K. Zhang. 2024. "Simulating Confidence Intervals for Conditional Value-at-Risk via Least-Squares Metamodels". *INFORMS Journal on Computing* 37(4):1087–1105.

Lan, H., B. L. Nelson, and J. Staum. 2010. "A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-Level Simulation". *Operations Research* 58(5):1481–1490.

Liang, G., K. Zhang, and J. Luo. 2024. "A FAST Method for Nested Estimation". *INFORMS Journal on Computing* 36(6):1481–1500.

Liu, G., and K. Zhang. 2024. "A Tutorial on Nested Simulation". In *2024 Winter Simulation Conference (WSC)*, 1–15 https://doi.org/10.1109/WSC63780.2024.10838912.

Liu, M., and J. Staum. 2010. "Stochastic Kriging for Efficient Nested Simulation of Expected Shortfall". *Journal of Risk* 12(3):3.

Owen, A. 1990. "Empirical Likelihood Ratio Confidence Regions". *The Annals of Statistics* 18(1):90–120.

Sun, Y., D. W. Apley, and J. Staum. 2011. "Efficient Nested Simulation for Estimating the Variance of a Conditional Expectation". *Operations Research* 59(4):998–1007.

Wang, W., Y. Wang, and X. Zhang. 2024. "Smooth Nested Simulation: Bridging Cubic and Square Root Convergence Rates in High Dimensions". *Management Science* 70(12):9031–9057.

Yi, Y., and W. Xie. 2017. "An Efficient Budget Allocation Approach for Quantifying the Impact of Input Uncertainty in Stochastic Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(4):1–23.

Zhang, K., G. Liu, and S. Wang. 2022. "Bootstrap-Based Budget Allocation for Nested Simulation". *Operations Research* 70(2):1128–1142.

Zhu, H., T. Liu, and E. Zhou. 2020. "Risk Quantification in Stochastic Simulation under Input Uncertainty". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 30(1):1–24.

AUTHOR BIOGRAPHIES

QIANWEN ZHU is a Ph.D. candidate in the Department of Decision Analytics and Operations, College of Business at City University of Hong Kong. Her research focuses on stochastic simulation and machine learning, and their applications. Her email address is qianwen.zhu@my.cityu.edu.hk.

GUANGWU LIU is a Professor in the Department of Decision Analytics and Operations, College of Business at City University of Hong Kong. His research interests include stochastic simulation, business analytics, financial engineering, and risk management. He serves as an Associate Editor of Naval Research Logistic. His e-mail address is msgw.liu@cityu.edu.hk and his website is https://www.cb.cityu.edu.hk/staff/guanliu.

XIANYU KUANG earned his Ph.D. in the Department of Decision Analytics and Operations, College of Business at City University of Hong Kong, and was a Research Assistant in the same department. His research interests include stochastic simulation, machine learning, financial engineering and risk management. His email address is dwyanekuang1996@gmail.com.