RISK ESTIMATION VIA WEIGHTED REGRESSION

Mark Broadie

Yiping Du

Graduate School of Business
Columbia University
New York, NY 10027, USA

Industrial Engineering and Operations Research
Columbia University
New York, NY 10027, USA

Ciamac C. Moallemi

Graduate School of Business
Columbia University
New York, NY 10027, USA

ABSTRACT

In this paper we propose a method based on weighted regression for the estimation of risk in nested Monte Carlo simulation. The mean squared error (MSE) of a standard nested simulation converges at the rate $k^{-2/3}$, where k is the computational budget. Similar to the regression method proposed in Broadie, Du, and Moallemi (2011b), the MSE of the proposed weighted regression method converges at the rate k^{-1} until reaching an asymptotic bias level, which depends on the size of the regression error. However, the weighted approach further reduces MSE by emphasizing scenarios that are more important to the calculation of the risk measure. We find a globally optimal weighting strategy for general risk measures in an idealized setting. For applications, we propose and test a practically implementable two-pass method, where the first pass uses an unweighted regression and the second pass uses weights based on the first pass.

1 INTRODUCTION

Risk measurement typically has two stages: scenario generation and portfolio re-valuation. In the first scenario generation (outer) stage, risk factors are sampled by Monte Carlo simulation over a given risk time horizon. In the second portfolio re-valuation (inner) stage, the portfolio value at the risk time horizon is computed given a particular scenario of risk factors. Since a portfolio may contain a large number of securities for most of which analytical pricing formulas do not exist, the re-valuation task could also require Monte Carlo simulation. The main contributions of this paper are as follows:

1. *We propose a method via weighted regression to estimate risk measures.*
   Beginning with the unweighted regression method proposed by Broadie, Du, and Moallemi (2011b), we present a weighted regression method that assigns more weight to scenarios that are more important to the calculation of the risk measure.
2. *We provide a heuristic analysis that leads to a globally optimal weighting strategy.*
   When the computational budget is large enough, the MSE of the regression estimator will be dominated by the asymptotic bias. Based on a heuristic analysis, we give an explicit choice of weight function that leads to the global minimum of an upper bound of the squared bias.
3. *We suggest a practically implementable weighting strategy.*

Our globally optimal weight function depends on knowledge of the portfolio loss in each scenario, which is unobservable in practice. To remedy this, we propose a two-pass weighting strategy, where the first pass is an unweighted regression, and the second pass is a weighted regression with the weight function based on the first pass.

4. *We provide numerical examples to illustrate the practical benefits of our method.*
   In our numerical examples, the two-pass weighted regression method yields an estimator that is superior to standard nested simulation. It performs comparably to unweighted regression with a small computational budget, and has better performance when the computational budget is large.

All proofs are omitted and can be found in the extended version of this paper (Broadie, Du, and Moallemi 2011c).

1.1 Literature Review

Crouhy, Galai, and Mark (2000), Jorion (2006), and McNeil, Frey, and Embrechts (2006) present overviews of financial risk measurement and management. For the nested estimation of risk, Lee (1998), Lee and Glynn (2003), and Gordy and Juneja (2008, 2010) analyze the convergence properties of the standard nested Monte Carlo simulation estimator. When the underlying scenario space is continuous, under varying sets of assumptions, these author establish that given a total computational budget of k, the optimal number of outer stage samples is of order $k^{2/3}$, the optimal number of inner state samples is of order $k^{1/3}$, and the optimal asymptotic MSE decreases at rate $k^{-2/3}$. Broadie, Du, and Moallemi (2011a) illustrate how, when the risk measure is the probability of a loss, the convergence rate of the MSE can be improved to $k^{-4/5+\delta}$, for any positive δ, using non-uniform inner stage samples.

Hong and Juneja (2009) discuss the benefit of using kernel smoothing to estimate risk measures in low dimensional settings. Liu and Staum (2010) focus on expected shortfall, and explore an approach based on stochastic kriging. Stochastic kriging described, for instance, by Ankenman, Nelson, and Staum (2010), builds a metamodel that interpolates portfolio values based on nearby outer stage scenarios. The methods of Hong and Juneja (2009) and Liu and Staum (2010) are *spatial methods*; they seek to improve upon standard nested simulation by combining information from many scenarios when computing estimates of the portfolio loss. The regression methods we consider are also spatial methods. Hong and Juneja (2009) and Liu and Staum (2010) focus on the use of local information from nearby scenarios, while regression is a global spatial method. Regression is also used for American option pricing in Tsitsiklis and Roy (2001) and Longstaff and Schwartz (2001).

2 PROBLEM FORMULATION

Starting at time $t = 0$, we are interested in measuring the risk of a portfolio at the risk horizon $t = \tau$. The portfolio value at time 0, denoted by X_0, is assumed to be known, and the portfolio value at time τ, denoted by X_τ, is unknown and random. The portfolio value X_τ depends on the state of a set of financial risk factors realized by time τ. In particular, define $\Omega \subset \mathbb{R}^Q$ to be the set of possible risk factor realizations between times 0 and τ, where each scenario $\omega \in \Omega$ incorporates sufficient information to determine the portfolio value at time τ. In each scenario $\omega \in \Omega$, we have the portfolio value $X_\tau(\omega)$. The portfolio loss $L(\omega)$ at time τ is defined by $L(\omega) \triangleq X_0 - X_\tau(\omega)$. Without loss of generality, we assume the riskless rate is 0 between times 0 and τ, and no intermediate cashflows occur before time τ.

We assume that ω is distributed according to the real-world distribution of risk factors, and thus $L(\omega)$ is distributed according to the real-world distribution of portfolio losses. A risk measure is a function that

associates a loss distribution $L(\omega)$ with a scalar α. We consider risk measures of the form

$$\alpha \triangleq \mathsf{E}\big[f(L(\omega))\big], \tag{1}$$

given a function $f \colon \mathbb{R} \to \mathbb{R}$ such that the expectation in (1) exists.

Since the space of risk factors Ω may be quite large or even infinite, one approach to estimating α is using Monte Carlo sampling to approximate the distribution of $L(\omega)$ empirically. Denote by $\omega^{(1)}, \ldots, \omega^{(n)}$ the scenarios that are independent and identically distributed according to the real-world distribution of ω. We refer to $\omega^{(1)}, \ldots, \omega^{(n)}$ as the outer stage scenarios. Given a scenario $\omega^{(i)}$, denote by $\hat{Z}_{i,1}, \ldots, \hat{Z}_{i,m}$ the inner stage samples of portfolio cashflows that are independent and identically distributed. Define $\zeta^{(i)}$ to be the random variable that captures the randomness of $\hat{Z}_{i,1}, \ldots, \hat{Z}_{i,m}$. Then the portfolio loss $L(\omega^{(i)})$ can be estimated by

$$\hat{L}(\omega^{(i)}, \zeta^{(i)}) \triangleq \frac{1}{m} \sum_{j=1}^{m} \hat{Z}_{i,j}. \tag{2}$$

Notice that $\zeta^{(i)}$ is independent and identically distributed for each scenario $\omega^{(i)}$. We combine the estimates from the outer and inner stages of simulation and obtain the standard nested estimator

$$\hat{\alpha}_{\mathsf{SN}(m,n)} \triangleq \frac{1}{n} \sum_{i=1}^{n} f\big(\hat{L}(\omega^{(i)}, \zeta^{(i)})\big). \tag{3}$$

Another way to estimate the risk measure α is via regression as proposed in Broadie, Du, and Moallemi (2011b). Given samples $(\omega^{(1)}, \zeta^{(1)}), \ldots, (\omega^{(n)}, \zeta^{(n)})$ from the standard nested simulation, instead of taking a sample average as in (3), the regression method combines information from different scenarios to provide a global approximation to the portfolio loss function that is easier to evaluate than $L(\cdot)$ and may be more accurate than $\hat{L}(\cdot)$. Denote by $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ a set of real-valued basis functions on the scenario space Ω, and define the vector $\Phi(\omega) \triangleq \big(\phi_1(\omega), \ldots, \phi_d(\omega)\big) \in \mathbb{R}^d$ for each scenario ω. The portfolio loss $L(\cdot)$ is approximated by a linear combination of these basis functions: $L(\omega) \approx \Phi(\omega)r \triangleq \sum_{\ell=1}^{d} \phi_\ell(\omega)r_\ell$, for a column vector $r \in \mathbb{R}^d$. Given the basis function set Φ, we have the following optimization problem to minimize the mean squared error between $L(\omega)$ and $\Phi(\omega)r$:

$$r^* \in \underset{r \in \mathbb{R}^d}{\operatorname{argmin}} \; \mathsf{E}\left[\big(L(\omega) - \Phi(\omega)r\big)^2\right]. \tag{4}$$

We make the following assumptions.

Assumption A1 The second moment of the portfolio loss $L(\omega)$ is finite, i.e., $\mathsf{E}[L(\omega)^2] < \infty$. The estimated loss $\hat{L}(\omega, \zeta)$ satisfies $\mathsf{E}\big[\hat{L}(\omega, \zeta) \big| \omega\big] = L(\omega)$, and $\operatorname{Var}\big(\hat{L}(\omega, \zeta) \big| \omega\big) = v(\omega)/m < \infty$. Also, the conditional variance $v(\omega)$ satisfies $\mathsf{E}[v(\omega)] < \infty$.

Assumption A2 The second moments of $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are finite, i.e., $\mathsf{E}\big[\phi_\ell(\omega)^2\big] < \infty$, for $\ell = 1, \ldots, d$. Further, if $\omega^{(1)}, \ldots, \omega^{(n)}$ are i.i.d. scenarios, then the matrix with rows $\Phi(\omega^{(i)})$, for $i = 1, \ldots, n$, has full rank almost surely. Without loss of generality, we can then assume that the functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are orthonormal, i.e., we assume that $\mathsf{E}\big[\Phi(\omega)^\top \Phi(\omega)\big]$ is the identity matrix.

Under Assumptions A1 and A2, the basis functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are linearly independent. Moreover, the optimal solution to (4) exists and is unique.

In general, the portfolio loss $L(\cdot)$ may be difficult to compute analytically, and we use $\hat{L}(\omega^{(i)}, \zeta^{(i)})$ defined by (2) to approximate $L(\omega^{(i)})$. Given $\hat{L}(\omega^{(1)}, \zeta^{(1)}), \ldots, \hat{L}(\omega^{(n)}, \zeta^{(n)})$, we solve the optimization problem

$$\hat{r} \in \underset{r \in \mathbb{R}^d}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^{n} \big(\hat{L}(\omega^{(i)}, \zeta^{(i)}) - \Phi(\omega^{(i)})r\big)^2, \tag{5}$$

and estimate the risk measure by

$$\hat{\alpha}_{\mathsf{REG}(m,n)} \triangleq \mathsf{E}\left[f\left(\Phi(\omega)\hat{r}\right)\middle| \vec{\omega}, \vec{\zeta}\right], \tag{6}$$

where $\vec{\omega} \triangleq (\omega^{(1)}, \ldots, \omega^{(n)})^{\top}$ and $\vec{\zeta} \triangleq (\zeta^{(1)}, \ldots, \zeta^{(n)})^{\top}$. Given Assumptions A1 and A2, the optimal solution to (5) exists and is unique almost surely when $n \geq d$, and thus $\hat{\alpha}_{\mathsf{REG}(m,n)}$ is well defined. We refer to $\hat{\alpha}_{\mathsf{REG}(m,n)}$ as the regression estimator.

Remark 1 After obtaining $\hat{r}$, we can use Monte Carlo simulation to generate new outer stage scenarios without inner stage samples. Indeed, if we generate n' new outer stage scenarios in addition to those in nested simulation, and we estimate $\hat{\alpha}_{\mathsf{REG}(m,n)}$ with

$$\frac{1}{n'}\sum_{i=1}^{n'} f\left(\Phi(\omega_i)\hat{r}\right), \tag{7}$$

the MSE between (6) and (7) is in the order of $(n')^{-1}$, which is asymptotically negligible compared to the MSE of the estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ as analyzed in Broadie, Du, and Moallemi (2011b). Notice that generating of outer stage samples without inner stage samples is computationally cheap, and thus we can estimate $\hat{\alpha}_{\mathsf{REG}(m,n)}$ with (7) to arbitrary accuracy without excessive computational effort. Therefore, (7) is justified to be used in practice. For the purpose of discussion here, we assume $\hat{\alpha}_{\mathsf{REG}(m,n)}$ can be exactly computed.

3   THE WEIGHTED REGRESSION ALGORITHM

When using the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$, all outer stage scenarios are weighted uniformly in the sample mean squared error objective of (5). This is reasonable if we seek a globally good approximation over the entire scenario space Ω. However, when dealing with specific risk measures, the portfolio loss estimates in some scenarios deserve more precise estimation than those in other scenarios. For example, when estimating the expected excess loss, scenarios with large losses contribute substantially to the estimator. Scenarios with large profits, on the other hand, do not directly impact the risk calculation, so the accuracy of their estimates is not as important. Since scenarios are not equally important, a *weighted* regression scheme where larger weights are assigned to more important scenarios is logical.

Suppose $h\colon \Omega \to \mathbb{R}$ is a non-negative function.

Assumption A3 The second moment of $\sqrt{h(\omega)}L(\omega)$ is finite, i.e., $\mathsf{E}[h(\omega)L(\omega)^2] < \infty$. The estimated loss $\hat{L}(\omega, \zeta)$ satisfies $\mathsf{E}\left[\hat{L}(\omega, \zeta)\middle| \omega\right] = L(\omega)$, and $\mathrm{Var}\left(\hat{L}(\omega, \zeta)\middle| \omega\right) = v(\omega)/m < \infty$. Also, $h(\omega)$ and $v(\omega)$ satisfy $\mathsf{E}[h(\omega)v(\omega)] < \infty$.

Assumption A4 The second moments of $\sqrt{h(\cdot)}\phi_1(\cdot), \ldots, \sqrt{h(\cdot)}\phi_d(\cdot)$ are finite, i.e., $\mathsf{E}\left[\phi_\ell(\omega)^2\right] < \infty$, for $\ell = 1, \ldots, d$. Further, if $\omega^{(1)}, \ldots, \omega^{(n)}$ are i.i.d. scenarios, then the matrix with rows $\sqrt{h(\omega^{(i)})}\Phi(\omega^{(i)})$, for $i = 1, \ldots, n$, has full rank almost surely. Moreover, we assume that the functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are orthonormal, i.e., we assume that $\mathsf{E}\left[\Phi(\omega)^{\top}\Phi(\omega)\right]$ is the identity matrix.

Using $h(\cdot)$ as the weight function, we define

$$r^*(h) \in \underset{r \in \mathbb{R}^d}{\mathrm{argmin}}\ \mathsf{E}\left[h(\omega)\left(L(\omega) - \Phi(\omega)r\right)^2\right]. \tag{8}$$

For any specific h that satisfies Assumptions A3 and A4, the optimal solution $r^*(h)$ to (8) exists and is unique. Given outer stage scenarios $\omega^{(1)}, \ldots, \omega^{(n)}$, we assign $h(\omega^{(i)})$ to be the weight of the scenario $\omega^{(i)}$ in the regression, i.e., with $\vec{h} \triangleq \left(h(\omega^{(1)}), \ldots, h(\omega^{(n)})\right)^{\top}$. Since $L(\omega^{(1)})$ is unobservable in practice. we

use the Monte Carlo portfolio loss estimates $\hat{L}(\omega^{(1)}, \zeta^{(1)}), \ldots, \hat{L}(\omega^{(n)}, \zeta^{(n)})$, and estimate the regression coefficients by

$$\hat{r}(\vec{h}) \in \operatorname*{argmin}_{r \in \mathbb{R}^d} \ \frac{1}{n} \sum_{i=1}^{n} h(\omega^{(i)}) \big(\hat{L}(\omega^{(i)}, \zeta^{(i)}) - \Phi(\omega^{(i)}) r \big)^2. \tag{9}$$

The weighted regression estimator is

$$\hat{\alpha}_{\mathsf{REG}(m,n,h)} \triangleq \mathsf{E} \left[f\big(\Phi(\omega) \hat{r}(\vec{h}) \big) \,\Big|\, \vec{\omega}, \vec{\zeta} \right].$$

For any specific h that satisfies Assumptions A3 and A4, the optimal solution $\hat{r}(\vec{h})$ to (9) exists and is unique almost surely when $n \geq d$, so our estimator $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ is well defined. By following the same argument as in Remark 1, we assume $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ can be exactly computed.

Remark 2 Notice that the scale of $h(\cdot)$ has no impact on $\hat{r}(\vec{h})$, i.e., if we multiply $h(\cdot)$ by a constant factor, $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ will not change.

4   ANALYSIS

Consider the MSE of the weighted regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$:

$$\mathsf{E} \left[(\hat{\alpha}_{\mathsf{REG}(m,n,h)} - \alpha)^2 \right] = \mathsf{E} \left[\left(\mathsf{E} \left[f\big(\Phi(\omega) \hat{r}(\vec{h}) \big) \,\Big|\, \vec{\omega}, \vec{\zeta} \right] - \mathsf{E} \left[f\big(L(\omega) \big) \right] \right)^2 \right]. \tag{10}$$

In order to analyze (10), we need the following technical assumptions.

Assumption F1 The function $f(\cdot)$ is Lipschitz continuous, i.e., there exists a scalar U_{Lip}, such that $|f(x') - f(x'')| \leq U_{\mathrm{Lip}} |x' - x''|$, for all $x', x'' \in \mathbb{R}$.

Assumption A5 The moment generating functions of $h(\omega) \|\Phi(\omega)\|_2^2$, $h(\omega) \big(L(\omega) - \Phi(\omega) r^* \big)^2$, and $h(\omega)(\hat{L}(\omega, \zeta) - L(\omega))^2$ are finite-valued in a neighborhood of zero.

Define $G(r, \omega, \zeta, h) \triangleq h(\omega) \big(\hat{L}(\omega, \zeta) - \Phi(\omega) r \big)^2$, and $g(r, h) \triangleq \mathsf{E}[G(r, \omega, \zeta, h)]$. Given r, the function $G(r, \omega, \zeta, h)$ is the squared error of a regression estimate $\Phi(\omega) r$ versus the Monte Carlo estimate in scenario ω, and the function $g(r, h)$ is the mean over scenarios of the squared error. For any $\rho > 0$, define R_ρ to be the convex and closed set $R_\rho \triangleq \{ r \in \mathbb{R}^d : \|r - r^*\|_2^2 \leq \rho \}$. We make the following additional assumption.

Assumption A6 For any $\rho > 0$, there exists a constant $\lambda > 0$ such that for any $r', r'' \in R_\rho$, the moment generating function $\Psi_{r',r''}(t)$ of the random variable $Y_{r',r''} \triangleq (G(r', \omega, \zeta, h) - g(r', h)) - (G(r'', \omega, \zeta, h) - g(r'', h))$ satisfies $\Psi_{r',r''}(t) \leq \exp(\rho \lambda^2 t^2)$, for any $t \in \mathbb{R}$.

We have the following result, whose proof can be found be found in the extended version of this paper (Broadie, Du, and Moallemi 2011c):

Theorem 1 For a specific h, if Assumptions F1, A3, A4, A5, and A6 hold, then

$$\lim_{n \to \infty} \mathsf{E} \left[(\hat{\alpha}_{\mathsf{REG}(m,n,h)} - \alpha)^2 \right] = \big(\mathsf{E} \left[f\big(\Phi(\omega) r^*(h) \big) \right] - \mathsf{E} \left[f\big(L(\omega) \big) \right] \big)^2. \tag{11}$$

As the number of outer stage scenarios $n \to \infty$, Theorem 1 shows that the MSE of $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ diminishes until reaching the level of the asymptotic bias, and does not vanish in the limit. If the portfolio loss L is not in the span of the basis functions Φ, even with large m and n, the bias induced by the regression method exists in general, which leaves the limit of (11) non-zero. We make the following assumption.

Assumption F2 The function $f(\cdot)$ has first derivative almost everywhere.

Given Assumption F2, we applying Jensen's inequality and heuristically applying a Taylor approximation,

$$(\mathsf{E}\left[f\left(\Phi(\omega)\,r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L(\omega)\right)\right])^2 \leq \mathsf{E}\left[\left(f\left(\Phi(\omega)\,r^*\left(h\right)\right) - f\left(L(\omega)\right)\right)^2\right]$$

$$\approx \mathsf{E}\left[\left(f'\left(L(\omega)\right)\right)^2\left(\Phi(\omega)\,r^*\left(h\right) - L(\omega)\right)^2\right]. \tag{12}$$

Instead of finding the choice of weighting function that minimizes the limit of (11), for reasons of tractability, we consider optimizing the upper bound given by (12). Here, we seek to find regression approximations that are close to the true portfolio loss emphasizing scenarios ω where $f'\left(L(\omega)\right)$ is large. With (12) as our objective, we have the following optimization problem, to determine a choice of weight function:

$$\begin{aligned}
\underset{h}{\text{minimize}} \quad & \mathsf{E}\left[\left(f'\left(L(\omega)\right)\right)^2\left(\Phi(\omega)\,r^*\left(h\right) - L(\omega)\right)^2\right] \\
\text{subject to} \quad & \mathsf{E}\left[h(\omega)\right] > 0, \\
& h(\omega) \geq 0, \quad \forall \omega \in \Omega.
\end{aligned} \tag{13}$$

In the following theorem, we find a globally optimal solution to the problem (13).

Theorem 2 Assume that Assumption F2 holds. Define the weight function by

$$h_{\text{opt}}(\omega) = \left(f'\left(L(\omega)\right)\right)^2, \tag{14}$$

and assume that $\mathsf{E}\left[\left(f'\left(L(\omega)\right)\right)^2\right] > 0$. Then h_{opt} is a globally optimal solution to the optimization problem (13).

Proof. With r^* defined by (8), for any h in the domain of (13),

$$\mathsf{E}\left[\left(f'\left(L(\omega)\right)\right)^2\left(\Phi(\omega)r^*(h_{\text{opt}}) - L(\omega)\right)^2\right] \quad = \quad \mathsf{E}\left[h_{\text{opt}}(\omega)\left(\Phi(\omega)r^*(h_{\text{opt}}) - L(\omega)\right)^2\right] \tag{15}$$

$$\leq \quad \mathsf{E}\left[h_{\text{opt}}(\omega)\left(\Phi(\omega)r^*(h) - L(\omega)\right)^2\right] \tag{16}$$

$$= \quad \mathsf{E}\left[\left(f'\left(L(\omega)\right)\right)^2\left(\Phi(\omega)r^*(h) - L(\omega)\right)^2\right], \tag{17}$$

where (15) and (17) follow from (14), and (16) follows from (8).

∎

Theorem 2 suggests a specific choice of weight function, in particular, that the weight $h(\omega)$ of each scenario ω should be proportional to $\left(f'\left(L(\omega)\right)\right)^2$.

5 PRACTICAL IMPLEMENTATION

In this section and in the Section 6.2, we consider the risk measure

$$\alpha = \mathsf{E}\left[f\left(L(\omega)\right)\right] = \mathsf{E}\left[\left(L(\omega) - c\right)^+\right].$$

Given a loss threshold $c \in \mathbb{R}$, α is the expected excess loss over the level c. From (14), we have the weight function

$$h_{\text{opt}}(\omega) = \left(f'\left(L(\omega)\right)\right)^2 = \mathbb{I}_{\{L(\omega) \geq c\}}.$$

This weighting is intuitively reasonable, as scenarios that have losses larger than the threshold c are assigned with more weight. In practice, however, the portfolio loss $L(\omega)$ is unobservable, so we propose a *two-pass* procedure that does not depend on knowledge of $L(\omega)$. In particular, we will first approximate $L(\omega)$ with

an *unweighted* regression and obtain $\hat{r}$; from $\hat{r}$, we will construct an approximation to h_{opt} to be used in a *weighted* regression. In order to describe our method, define $M(\omega)$ to be the model error given the approximation $\Phi(\omega)r^*$,

$$M(\omega) \triangleq L(\omega) - \Phi(\omega)r^*. \tag{18}$$

The following assumption is a technical assumption standard in regression theory (see, e.g., White (2001)).

Assumption A7 The matrix $\mathsf{E}\big[v(\omega)\Phi(\omega)^\top\Phi(\omega)\big]$ is positive definite, and $\mathsf{E}\big[\phi_\ell(\omega)^2 M(\omega)^2\big] < \infty$, for each $\ell = 1,\dots,d$.

In what follows, denote by $N(\cdot)$ the cumulative distribution function for the normal distribution. Using standard results from the theory of regression, Broadie, Du, and Moallemi (2011b) establish the following lemma.

Lemma 1 Suppose Assumptions A1, A2, and A7 hold. As $n \to \infty$, r^* and $\hat{r}$ from the unweighted regression method defined by (4) and (5) satisfy $\sqrt{n}(\hat{r} - r^*) \xrightarrow{\text{d}} N\big(\mathbf{0}, \Sigma_M + \Sigma_v/m\big)$, where $\Sigma_M \triangleq \mathsf{E}\big[M(\omega)^2\Phi(\omega)^\top\Phi(\omega)\big]$, and $\Sigma_v \triangleq \mathsf{E}\big[v(\omega)\Phi(\omega)^\top\Phi(\omega)\big]$.

Remark 3 One implication of Lemma 1 is that, given a fixed number of inner stage samples $k = mn$, the asymptotic distribution of $\hat{r}$ has minimum variance when $m = 1$ and $n = k$. We will apply this fact in our analysis and numerical experiments.

When $m = 1$ and $n = k$, as $k \to \infty$, Lemma 1 and (18) imply that, for a fixed scenario ω,

$$\sqrt{k}\big(\Phi(\omega)\hat{r} + M(\omega) - L(\omega)\big) \xrightarrow{\text{d}} N\big(\mathbf{0}, \Phi(\omega)(\Sigma_M + \Sigma_v)\Phi(\omega)^\top\big). \tag{19}$$

Equation (19) suggests that when k is large, it is reasonable to approximate the Bayesian posterior distribution of the portfolio loss $L(\omega)$ given $\vec{\omega}$ and $\vec{\zeta}$ by a normal distribution with mean $\Phi(\omega)\hat{r} + M(\omega)$ and variance $\Phi(\omega)(\Sigma_M + \Sigma_v)\Phi(\omega)^\top/k$. Given this approximation, we can approximate h_{opt} with its posterior mean:

$$\mathsf{E}\left[h_{\text{opt}}(\omega)\,\middle|\,\vec{\omega},\vec{\zeta}\right] \approx N\left(\frac{\sqrt{k}\big(\Phi(\omega)\hat{r} + M(\omega) - c\big)}{\sqrt{\Phi(\omega)(\Sigma_M + \Sigma_v)\Phi(\omega)^\top}}\right). \tag{20}$$

(Notice that $\hat{r}$ depends on $\vec{\omega}$ and $\vec{\zeta}$.) If the basis functions are well-chosen, the model error $M(\omega)$ is small in magnitude relative to the unweighted regression approximation $\Phi(\omega)\hat{r}$. Further, in practice, we observe that the denominator of the argument in the right-hand side of (20) does not vary by more than one order of magnitude. These observations suggest an overall approximation for the globally optimal weight function h_{opt}:

$$h(\omega) = N\left(\frac{\sqrt{k}\big(\Phi(\omega)\hat{r} - c\big)}{\Gamma}\right), \tag{21}$$

for some constant $\Gamma > 0$. Using the weight function (21), we have a two-pass weighted regression procedure:

- the first pass is an unweighted regression and provides $\Phi(\omega)\hat{r}$, an approximation of $L(\omega)$ used to determine weights;
- based on $\Phi(\omega)\hat{r}$, the second pass is a weighted regression with weights defined by (21).

Notice that this two-pass weighted regression procedure assigns weights with a weight function inspired by h_{opt}, and it does not depend on information of $L(\omega)$ or any quantities that are unknown in practice. We will compare the two-pass weighted regression method with the unweighted regression method in Section 6.

6 NUMERICAL RESULTS

In this section, we use numerical examples to demonstrate the two-pass weighted regression method based on (21). The standard nested simulation method and the unweighted regression method are also implemented as competing methods.

6.1 Experimental Setting

In our examples, we consider portfolios that consist of a collection of underlying assets and derivatives defined over those assets. Underlying asset prices are assumed to follow geometric Brownian motion processes. The other assumptions of the Black-Scholes model are also assumed, so that closed-form pricing formulas are available.

Assume each Q-dimensional outer stage scenario $\omega \triangleq (\omega_1, \ldots, \omega_Q) \in \Omega \triangleq \mathbb{R}^Q$ is distributed according to a multivariate Gaussian distribution with mean zero, variance one, and a given correlation matrix. The prices of underlying assets at time τ are lognormally distributed. Define $S_\tau(\omega) \triangleq \left(S_{1,\tau}(\omega), \ldots, S_{Q,\tau}(\omega) \right)$, as the vector of asset prices, where $S_{j,\tau}(\omega) = S_{j,0} \exp \left((\mu_j - \sigma_j^2/2)\tau + \sigma_j\sqrt{\tau}\omega_j \right)$. For the jth asset, $S_{j,0}$ is the price at time 0, μ_j is the drift in the real-world distribution, and σ_j is the volatility.

In order to evaluate portfolio losses at time τ in a scenario ω, we need to sample the inner stage paths between times τ and T under the risk-neutral distribution. For the jth security and for each $p = 1, \ldots, m$, define $W_{j,t}^{(p)}$ to be a Brownian motion corresponding to the pth inner stage path for $t \in [\tau, T]$ with $W_{j,\tau}^{(p)} = 0$. Notice that the m inner stage samples of the portfolio are independent and identically distributed, i.e., $\left(W_{1,t}^{(1)}, \ldots, W_{Q,t}^{(1)} \right), \ldots, \left(W_{1,t}^{(m)}, \ldots, W_{Q,t}^{(m)} \right)$ are m i.i.d. Q-dimensional Brownian motion processes. Then in scenario ω, the inner stage randomness is summarized by

$$\zeta \triangleq \left\{ W_{j,t}^{(p)}, \text{ for } t \in [\tau, T], \ j = 1, \ldots, Q, \ p = 1, \ldots, m \right\}.$$

Conditional on ω, under the risk-neutral distribution, the jth security in the pth sample path takes the value

$$S_{j,t}^{(p)}(\omega, \zeta) = S_{j,\tau}(\omega) \exp \left((r_f - \sigma_j^2/2)(t - \tau) + \sigma_j W_{j,t}^{(p)} \right),$$

for $t \in [\tau, T]$, where r_f is the continuously compounded riskless rate of interest. The portfolio loss estimate $\hat{L}(\omega, \zeta)$ is the average of the m sample paths of expected cashflows determined by

$$\left\{ \left(S_{j,t}^{(1)}(\omega, \zeta), \ldots, S_{j,t}^{(m)}(\omega, \zeta) \right), \text{ for } t \in [\tau, T], \ j = 1, \ldots, Q \right\}.$$

6.2 Examples

6.2.1 One-Dimensional Example

- *Model.* There is one asset with initial price $S_0 = 100$. The drift under the real-world distribution is $\mu = 8\%$. The annual volatility is $\sigma = 20\%$. The continuously compounded riskless rate of interest is $r_f = 3\%$.
- *Portfolio.* The portfolio consists of a long position in a single put option with strike $K = 95$ and maturity $T = 0.25$ years. The risk horizon is $\tau = 1/52$ years. The threshold $c = 0.859$ is the 90th percentile of the portfolio loss distribution.

6.2.2 Multi-Dimensional Delta-Hedged Example

- *Model.* There are 10 i.i.d. assets each with initial price $S_0 = 100$. The drift under the real-world distribution is $\mu = 8\%$. The annual volatility is $\sigma = 30\%$. The continuously compounded riskless rate of interest is $r_f = 5\%$.

- *Portfolio.* The portfolio consists of three types of securities. All derivatives in the portfolio have strike $K = 100$ and maturity $T = 0.1$ years. The risk horizon is $\tau = 0.04$ years. The three types are:
  - Short 10 down-and-out call options on asset i with barrier $H = 95$, for $i = 1, \ldots, 10$.
  - Short 5 cash-or-nothing put options on asset i, for $i = 1, \ldots, 10$.
  - An amount of asset i so that the portfolio delta with respect to asset i is zero, for $i = 1, \ldots, 10$.

  The threshold $c = 144.007$ is the 90th percentile of the portfolio loss distribution.

Remark 4 The barrier options in this paper are assumed to be potentially knocked in or out only between times τ and T, thus are technically partial-time barrier options (see, e.g., Section 4.17.4 in Haug 2006). Along a single sample path, the cashflow of a barrier option is a function of the minimum underlying asset price and the final asset price. We exactly sample these two quantities instead of sampling the entire sample path. More details can be found in Hui (1997) and Metwally and Atiya (2002).

6.2.3 Estimators

- *Optimal standard nested estimator.* This is the estimator $\hat{\alpha}_{\mathsf{SN}(m,n)}$ with the parameters m and n chosen optimally. By testing a large number of choices of the parameter pair (m,n), we select the pair with the minimum MSE. Note that, in practice, this choice is not achievable without such an exhaustive search.
- *Unweighted quadratic regression.* The estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ with these basis functions:
  - In the one-dimensional problem, the underlying asset has price $S_\tau(\omega)$ at time τ. The basis function set Φ consists of $\phi_1(\omega) = 1$, $\phi_2(\omega) = S_\tau(\omega)$, and $\phi_3(\omega) = (S_\tau(\omega))^2$.
  - In the multi-dimensional problem, the underlying assets have price $S_{j,\tau}(\omega)$ for $j = 1, \ldots, 10$ at time τ. The basis function set Φ includes quadratic functions of each $S_{j,\tau}(\omega)$ and quadratic functions of each $(S_{j,\tau}(\omega) - H)^+$ for $j = 1, \ldots, 10$. There are 41 basis functions.
- *Two-pass weighted quadratic regression.* This is the estimator $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ with basis functions as in the unweighted quadratic regression method above and the weight function (21).

  From Remark 3, we use $m = 1$ with the estimators of unweighted quadratic regression and two-pass weighted quadratic regression.

6.3 Numerical Performance

Results for the MSE of the various estimators are given in Figures 1 and 2. We interpret these results as follows:

- *Optimal standard nested estimator.* As the total number of inner stage samples k increases, the MSE converges at the rate $k^{-2/3}$, consistent with theoretical results.
- *Unweighted quadratic regression.* The MSE of this estimator converges at the rate k^{-1} until reaching an asymptotic bias level which depends on the size of the regression model error $M(\cdot)$, or, equivalently, the quality of the basis functions. This is consistent with the analysis in Broadie, Du, and Moallemi (2011b).
- *Two-pass weighted quadratic regression.* When k is small, the MSE of the two-pass weighted quadratic regression method is close to that of the unweighted quadratic regression method. This is due to the weight function (21). When k is small, the weight function is relatively flat (i.e., nearly equal weights). When k is large, the weight function (21) becomes steep, and more weight is assigned to more important scenarios, where losses exceed the threshold c. With large k, the
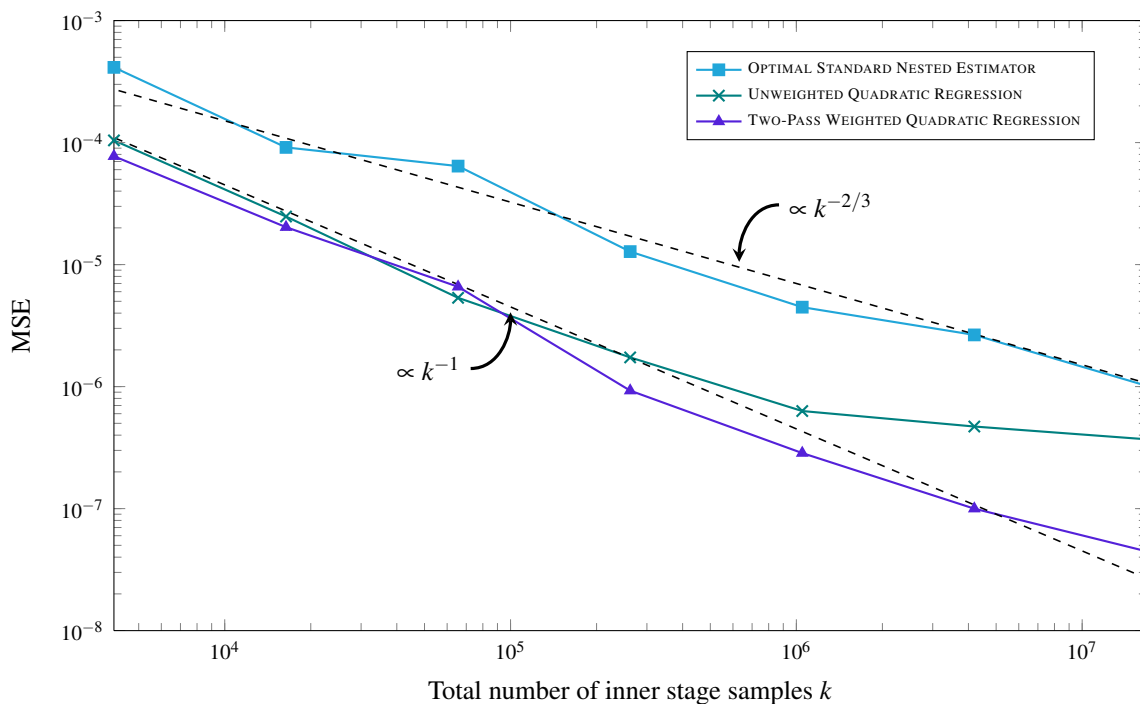
Figure 1: Illustration of the mean squared error in the one-dimensional example. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples.
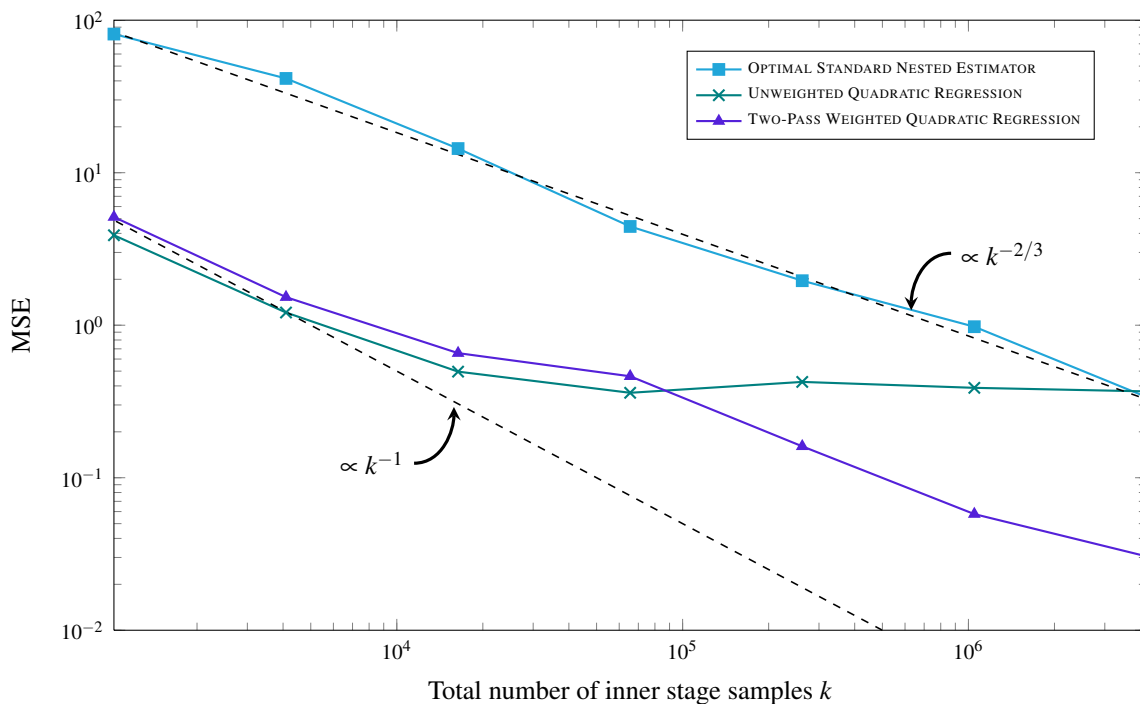

Figure 2: Illustration of the mean squared error in the 10-dimensional example. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples.
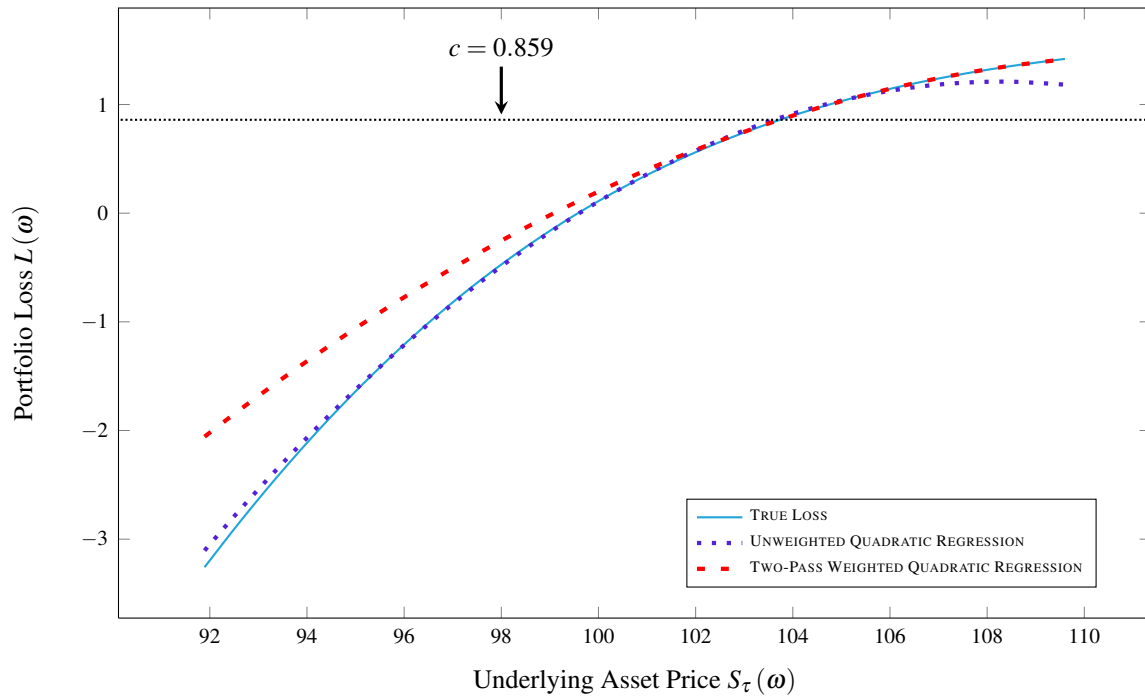
Figure 3: Illustration of approximations in the one-dimensional example. The vertical axis shows the portfolio loss, either true or estimated, and the horizontal axis represents the underlying asset price $S_\tau(\omega)$ at time τ.

> MSE of the two-pass weighted quadratic regression method works about one order of magnitude better than the unweighted quadratic regression method.

Figure 3 shows the approximations via different regression methods for large values of *n* in the one-dimensional example. Here, the unweighted quadratic regression method does not approximate the true portfolio loss in regions with large losses. On the other hand, the two-pass weighted quadratic regression method, with the same basis functions, fits the true portfolio loss much better in the large loss region.

ACKNOWLEDGMENT

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58 (2): 371–382.

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

Broadie, M., Y. Du, and C. C. Moallemi. 2011b. "Risk Estimation via Regression". Working paper, Columbia University.

Broadie, M., Y. Du, and C. C. Moallemi. 2011c. "Risk Estimation via Weighted Regression (Extended Version)". Working paper, Columbia University.

Crouhy, M., D. Galai, and R. Mark. 2000. *Risk Management*. New York: McGraw-Hill.

Gordy, M. B., and S. Juneja. 2008, April. "Nested Simulation in Portfolio Risk Measurement". FEDS 2008-21, Federal Reserve Board.

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

Haug, E. 2006. *The Complete Guide to Option Pricing Formulas*. 2nd ed. McGraw-Hill.

Hong, L. J., and S. Juneja. 2009, December. "Estimating the Mean of a Non-Linear Function of Conditional Expectation". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 1223–1236. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hui, C. H. 1997. "Time-Dependent Barrier Option Values". *The Journal of Futures Markets* 17 (6): 667–688.

Jorion, P. 2006. *Value at Risk*. New York: McGraw-Hill.

Lee, S.-H. 1998. *Monte Carlo Computation of Conditional Expectation Quantiles*. Ph.D. thesis, Stanford University.

Lee, S.-H., and P. W. Glynn. 2003, July. "Computing the Distribution Function of a Conditional Expectation via Monte Carlo: Discrete Conditioning Spaces". *ACM Transactions on Modeling and Computer Simulation* 13 (3): 238–258.

Liu, M., and J. Staum. 2010. "Stochastic Kriging for Efficient Nested Simulation of Expected Shortfall". *Journal of Risk* 12 (3): 3–27.

Longstaff, F. A., and E. S. Schwartz. 2001. "Valuing American Options by Simulation: a Simple Least-Squares Approach". *The Review of Financial Studies* 14 (1): 113–147.

McNeil, A., R. Frey, and P. Embrechts. 2006. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton, New Jersey: Princeton University Press.

Metwally, S. A. K., and A. F. Atiya. 2002. "Using Brownian Bridge for Fast Simulation of Jump-Diffusion Processes and Barrier Options". *The Journal of Derivatives* 10 (1): 43–54.

Tsitsiklis, J. N., and B. V. Roy. 2001. "Regression Methods for Pricing Complex American-Style Options". *IEEE Transactions on Neural Networks* 12 (4): 694–703.

White, H. 2001. *Asymptotic Theory for Econometricians*. Academic Press.

AUTHOR BIOGRAPHIES

MARK BROADIE is the Carson Family Professor of Business in the Graduate School of Business, Columbia University. His research focuses on issues in financial engineering, with a particular focus on the design and analysis of efficient Monte Carlo methods for pricing and risk management. His email address is mnb2@columbia.edu, and his webpage is http://www.columbia.edu/~mnb2/broadie/.

YIPING DU is a Ph.D. student in the Department of Industrial Engineering and Operations Research at Columbia University. His research focuses on efficient simulation methods for estimating risk measures. His email address is yd2166@columbia.edu.

CIAMAC C. MOALLEMI is an associate professor in the Graduate School of Business at Columbia University. His research interests are in the area of the optimization and control of large scale stochastic systems, with an emphasis on applications in financial engineering. His email address is ciamac@gsb.columbia.edu, and his webpage is http://moallemi.gsb.columbia.edu.