OPERATIONS RESEARCH
Vol. 63, No. 5, September–October 2015, pp. 1077–1097
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.2015.1419

Risk Estimation via Regression

Mark Broadie
Graduate School of Business, Columbia University, New York, New York 10027, mnb2@columbia.edu

Yiping Du
Industrial Engineering and Operations Research, Columbia University, New York, New York 10027, yd2166@columbia.edu

Ciamac C. Moallemi
Graduate School of Business, Columbia University, New York, New York 10027,
ciamac@gsb.columbia.edu

We introduce a regression-based nested Monte Carlo simulation method for the estimation of financial risk. An outer simulation level is used to generate financial risk factors and an inner simulation level is used to price securities and compute portfolio losses given risk factor outcomes. The mean squared error (MSE) of standard nested simulation converges at the rate $k^{-2/3}$, where k measures computational effort. The proposed regression method combines information from different risk factor realizations to provide a better estimate of the portfolio loss function. The MSE of the regression method converges at the rate k^{-1} until reaching an asymptotic bias level which depends on the magnitude of the regression error. Numerical results consistent with our theoretical analysis are provided and numerical comparisons with other methods are also given.

Subject classifications: statistics: estimation; decision analysis: risk; simulation.
Area of review: Financial Engineering.
History: Received January 2013; revision(s) received May 2014; accepted July 2015.

1. Introduction

Financial risk measurement is an important tool for monitoring the financial stability of banks and other financial institutions. When risks are found to be too large, banks may need to hold additional capital to reduce the chance of financial distress. Risk measurement is also used to recognize and evaluate the financial risks in portfolios of securities at mutual funds, hedge funds, endowments, corporations, and other nonfinancial institutions. Portfolio risk arises because the values of the constituent securities change over time in response to changes in risk factors, e.g., interest rates, exchange rates, stock prices, commodity prices, etc. In a large bank, risk assessment requires revaluing portfolios consisting of potentially hundreds of thousands of securities at a future date (called the risk horizon) under possibly thousands of realizations of risk factors. In many cases, portfolios contain derivative securities (e.g., options, swaps, mortgage-backed instruments, etc.) whose valuation can entail the simulation of cash flows and risk factors over time horizons extending to 30 years. The need to revalue a large number of securities, especially derivative securities, over a large number of risk factor realizations makes the risk measurement problem extremely computationally challenging.

Previous approaches have addressed the computational challenges of risk measurement by sacrificing either accuracy or speed. For example, the delta-gamma method (see, e.g., Rouvinez 1997, Britten-Jones and Schaefer 1999, Duffie and Pan 2001) approximates the portfolio loss function using first and second derivative information and is often combined with normal distribution assumptions for risk factor changes. This method allows faster risk calculations at the expense of accuracy and convergence. At the other extreme is nested simulation (see, e.g., Lee 1998; Lee and Glynn 2003; Gordy and Juneja 2008, 2010), which proceeds in two stages. The first, or outer, scenario generation stage uses Monte Carlo simulation to generate possible scenarios of financial risk factors until the given risk horizon. The second, or inner, portfolio revaluation stage also uses Monte Carlo simulation in order to generate financial risk factors and security cash flows until the maturity of the securities, conditioned on an outer scenario. The nesting of Monte Carlo simulation is computationally burdensome but has the advantage of converging to the true risk measure (where "true" refers to the exact value given a financial model for risk factors and security valuation). The mean squared error (MSE) of standard nested simulation converges at the rate $k^{-2/3}$, where k measures computational effort.

Several approaches to speed the convergence of the nested simulation method have been proposed. The adaptive method of Broadie et al. (2011) exploits outer stage information to generate a nonuniform number of inner stage samples. Computational effort is saved by spending less revaluation effort at outer stage scenarios that represent less portfolio risk. This adaptive method converges at the improved rate $k^{-4/5}$. Gordy and Juneja (2010) achieve a similar improvement through the jackknife method. Local spatial methods combine information from nearby scenarios to better approximate the portfolio loss function. Examples include the kernel regression method of Hong and Juneja (2009) and the

stochastic kriging method of Liu and Staum (2010). Although these methods lead to better convergence in low dimensions, they can suffer from the curse of dimensionality in higher dimensions.

In this paper we propose a global spatial method based on regression. Similar to kernel regression and stochastic kriging, the global regression method combines information from different outer stage scenarios to better approximate the portfolio loss function. We provide a specific theoretical analysis of the regression approach in the context of nested simulation for risk estimation. The MSE of the regression method converges at the rate k^{-1} until reaching an asymptotic bias level, which depends on the size of the regression error. Theoretical analysis is provided to highlight and quantify the effect of model error; i.e., we analyze the case when the regression basis functions do not span the true portfolio loss function and contrast the results with a complete set of basis functions. In this regression approach, the convergence rate is independent of the problem dimension, so the curse of dimensionality can be bypassed if well-chosen regression basis functions can be identified. Another advantage of the regression approach is that it can be easily and directly applied to a wide range of risk measures (e.g., expected shortfall or conditional value at risk, value at risk, spectral risk measures, and others). The regression method is practically implementable, and we provide numerical results that illustrate the computational savings on a range of examples. Specifically, we make the following main contributions in this paper:

1. *We propose a method for nested simulation based on regression.*

The method proceeds as in standard nested simulation, but it uses a linear combination of regression basis functions, instead of sample averages, to more accurately approximate portfolio losses. Since the entire distribution of portfolio losses is estimated by this method, it can be easily applied to compute a large range of risk measures.

2. *We provide a theoretical analysis of the performance of our method.*

We characterize the asymptotic behavior of the regression-based risk estimator. A novel analysis shows that the MSE of the regression-based risk estimator converges at the rate $n^{-1+\delta}$, for all positive δ, where n is the number of outer stage scenarios, plus a nondiminishing bias term that is determined by the magnitude of the regression error. The performance of the regression estimator does not depend on the dimension of a problem except through the quality of the regression basis functions. With well-chosen basis functions, the regression method performs significantly better than the standard nested simulation method.

3. *The theoretical analysis shows the asymptotically optimal trade-off between inner and outer sampling in the regression method.*

Given a fixed computational budget of inner stage samples $k = mn$, where n is the number of outer stage scenarios and m is the number of inner stage samples per scenario, our analysis shows that the asymptotically minimum mean squared error is achieved when $m = 1$ and $n = k$ in the regression method. That is, the regression method works best with one inner stage sample per outer stage scenario. With this asymptotically optimal choice, the regression method recovers the k^{-1} convergence rate of non-nested simulation until the MSE reaches an asymptotic bias level determined by the regression model error.

Note that this result is surprising and counterintuitive when compared to other risk estimation methods. For example, in nested simulation, the asymptotically optimal choice of samples is $n \propto k^{-2/3}$ and $m \propto k^{-1/3}$. Our intuition for this is as follows: in nested simulation, each outer stage scenario is considered independently. The nonlinearity of the risk function creates a bias in each scenario that can only be eliminated by taking an asymptotically increasing number of inner stage samples in that scenario. With regression, on the other hand, information from other scenarios can be used to reduce error in a particular scenario, and additional inner stage sampling is ultimately not necessary.

4. *We provide numerical examples that compare the performance of the regression method with other methods.*

We compare the mean squared error of the regression estimator with other methods used in the literature and in practice. Numerical results are provided that are consistent with the theoretical analysis and demonstrate the advantage of the regression method over other approaches.

5. *We propose a weighted variation of the regression method that offers improved asymptotic bias.*

Weighted regression emphasizes certain scenarios as more important to the calculation of the resulting risk measure via a weight function. We describe and analyze a weighted regression algorithm for risk estimation. We establish that the asymptotic bias of this algorithm is determined by the choice of weight function. We describe an idealized "optimal" choice of weight function, along with a practically implementable variation. We provide numerical results that demonstrate an improvement consistent with theory.

One persistent challenge in the nested estimation of risk is the presence of bias. So far as we know, *all* methods for risk estimation in the present setting are biased. For example, the delta-gamma method is biased, and there is no way to estimate or bound the bias. The situation is somewhat better for nested simulation. It is also biased, but Gordy and Juneja (2008, 2010) develop a theoretical bound on the bias that scales with the number of samples. This is analogous to the bound we establish that bias scales with the model error. Note that in both of these cases, it is not possible to practically estimate bias in a particular problem instance. However, the bounds are nevertheless useful—the Gordy and Juneja (2008, 2010) result illustrates how bias can be reduced by adding samples, whereas our result illustrates how bias can be reduced by adding basis functions. Further, we show that bias can be reduced by allowing for regression weights. Estimation of the bias in a individual problem instance remains an interesting direction for future research;

we highlight the work of Lan et al. (2010) as an important early contribution.

We also note that although the idea of using regression in risk estimation may not seem novel, our paper provides the first thorough analysis to our knowledge. Our analysis provides intuition on how regression should be applied, for example, by demonstrating that the regression method works best with one inner stage sample for each outer stage sample or by suggesting a choice for regression weights. Moreover, our results illustrate how the method offers faster convergence and how the ultimate accuracy depends on problem parameters such as the quality and number of basis functions, and the smoothness of the risk functional.

To our knowledge, among practitioners, the dominant method of choice is the delta-gamma method. This is because there are already so many modeling assumptions in risk estimation that precise accuracy of numerical methods is considered to be less important than computational speed. We provide compelling theoretical and numerical evidence that with a small number of samples and a quadratic basis, our regression method dominates delta-gamma. Because of the k^{-1} convergence rate, on a wide range of financial models and risk measures, our method outperforms other methods and is state of the art except when a very precise risk estimate is desired and a very large computational budget is available (which is not the case in practical applications).

Though we did not test it, we note that the global regression method can be combined with other local methods and further improved using standard simulation variance reduction techniques.

1.1. Literature Review

The standard nested simulation estimator with uniform inner stage sampling has been analyzed by Lee (1998); Lee and Glynn (2003); and Gordy and Juneja (2008, 2010). If the underlying scenario space is continuous, the optimal asymptotic MSE of the standard nested estimator diminishes at rate $k^{-2/3}$. Convergence is slower than the k^{-1} rate typical of non-nested simulation estimators. The reduced convergence rate occurs because risk measures are typically nonlinear functions of the portfolio loss, and noisy estimates of portfolio losses introduce bias in the estimator, which slows convergence. Sun et al. (2011) consider nested simulation in the context of variance estimation. As in our setting, they establish that the number of inner stage scenarios need not go to infinity.

Overviews of the delta-gamma approximation method are given in Rouvinez (1997), Britten-Jones and Schaefer (1999), and Duffie and Pan (2001). This method uses quadratic approximation to model the portfolio loss with its first (i.e., delta) and second (i.e., gamma) derivative information. Importance sampling and stratification techniques are used in Glasserman et al. (2000) to accelerate the computation in the delta-gamma method. More details are provided in §2.2.

Recently, nested simulation has been combined with kernel smoothing and stochastic kriging methods. Hong and Juneja (2009) use kernel smoothing with nested simulation and show the resulting estimator converges at rate $k^{-\min(1,4/(d+2))}$, where d is the problem dimension. This method improves on standard nested simulation for low dimensional problems, where $d \leqslant 3$. Liu and Staum (2010) use stochastic kriging, an interpolation-based meta-modeling approach, to estimate expected shortfall.

When the dimension (i.e., the number of risk factors) is large, local methods that rely on combining information from nearby outer stage scenarios to get accurate estimates of portfolio loss become inefficient. This curse of dimensionality arises because, in high dimension, any particular scenario will have very few neighboring scenarios. The regression approach we propose uses spatial information globally instead of locally, and thus it does not directly depend on the dimension of a problem. With a good choice of basis functions, the regression method initially converges at the same k^{-1} rate as non-nested simulation. Regression has also been used by Tsitsiklis and Van Roy (2001) and Longstaff and Schwartz (2001) in the context of simulation-based estimation of American option prices. Although the precise formulations and theory are different, analogous to our proposed algorithm, these least squares Monte Carlo (LSM) methods also involve estimating security values via a regression over basis functions. In practice, such LSM methods sample in a way that is consistent with what we present here: there, for the purposes of regression, continuation values in a state are typically estimated with a single point estimate. This is analogous to using a single inner stage sample ($m = 1$) in our setting. Moreover, LSM methods face many of the same issues as in the present paper, for example, the selection of good basis functions for regression. We refer the reader to the monograph of Glasserman (2004) for a survey on this topic.

2. Problem Formulation: Nested Simulation

We are interested in the loss of a portfolio at a future time τ, relative to its value at time 0. The time τ is called the risk horizon. The portfolio loss at time τ depends on a collection of financial risk factors that are realized between times 0 and τ. Examples of such risk factors include stock prices, commodity prices, interest rates, and exchange rates. Denote by $\Omega \subset \mathbb{R}^N$ the set of all possible realizations of the risk factors at time τ. We assume that each $\omega \in \Omega$ is a sufficient statistic[1] to determine the values of all securities in the portfolio at time τ. Hence, we will refer to ω as a *state* or *scenario*. The portfolio loss, which we will denote by $L(\omega)$, is thus a function of ω.

We assume that there exists a probability space such that ω is distributed according to the real-world distribution of the risk factors over the state space Ω and that $L(\cdot)$ is a measurable function so that $L(\omega)$ is distributed according to the real-world distribution of portfolio losses. A *risk measure* associates the real-world distribution of the portfolio

loss L with a single scalar α. Specifically, we consider risk measures of the form

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

for some function $f\colon \mathbb{R} \to \mathbb{R}$, where we assume that $L(\cdot)$ and $f(\cdot)$ are functions such that the expectation in (1) exists. Examples of such risk measures include the following:

- *Probability of a large loss.* Given a loss threshold $c \in \mathbb{R}$, define

$$f(L) \triangleq \mathbb{I}_{\{L \geqslant c\}}.$$

In this case, the risk measure α is the probability that the portfolio loss exceeds the level c.

- *Expected excess loss.* Given a loss threshold $c \in \mathbb{R}$, define

$$f(L) \triangleq (L-c)^+.$$

In this case, α is the expected value of losses in excess of the level c.

- *Squared tracking error.* Given a target level $c \in \mathbb{R}$, define

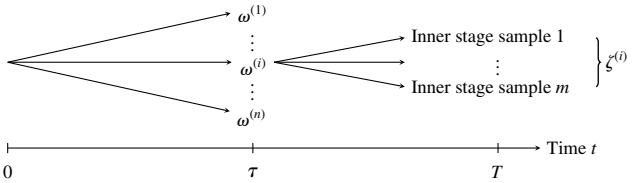
$$f(L) \triangleq (L-c)^2.$$

In this case, α is the squared error of the portfolio loss relative to the target c.

Note that many common risk measures are not explicitly in the class defined by (1), for example, value at risk, conditional value at risk, expected shortfall, etc. However, all of risk measures are functionals of the distribution of portfolio losses, and expectations of the form (1) are the most basic questions to ask about a loss distribution. In many cases, more complex risk measures can be computed indirectly from expectations of the form (1). For example, the value at risk is the threshold level c such that the probability of a loss exceeding c reaches a specified level. The ability to accurately compute the probability of a loss exceeding an arbitrary threshold can be an important step toward computing the value at risk. Similarly, the ability to accurately compute the expected excess loss can be an important subroutine in the computation of the conditional value at risk or expected shortfall. Hence, although our methods are broadly useful, in order to simplify the analysis, we will focus on this particular class where risk measures can be expressed as in (1).

Our challenge in computing the risk measure α is that, given a scenario ω, often the portfolio loss $L(\omega)$ is not directly computable. In many financial applications, the portfolio may include complex securities whose values cannot be determined analytically. Hence, the loss $L(\omega)$ may also need to be numerically computed. A common way to compute the portfolio loss in a given scenario is via Monte Carlo simulation. In what follows, we will describe two common approaches for estimating the risk of portfolios that are valued via Monte Carlo simulation.

Figure 1. Illustration of two-stage sampling.



Notes. The outer stage generates n scenarios $\omega^{(1)}, \ldots, \omega^{(n)}$. Conditional on each scenario $\omega^{(i)}$, m inner stage samples are generated that determine the cash flows of the portfolio from time τ to time T. Notice that the outer scenarios are generated according to the real-world distribution and the inner stage samples are generated according to the risk-neutral distribution.

2.1. Standard Nested Simulation

Standard nested simulation is the method of estimating the risk measure α by first approximating the expectation in (1) via a Monte Carlo simulation. In particular, consider a set of n scenarios $\omega^{(1)}, \ldots, \omega^{(n)}$ that are independent and identically distributed (i.i.d.) according to the real-world distribution of the risk factors ω. These samples are referred to as the *outer stage* of the simulation.

Given a scenario $\omega^{(i)}$, the portfolio loss $L(\omega^{(i)})$ will be estimated via an *inner stage* of Monte Carlo simulation. Specifically, suppose that T is the longest maturity time of all of the securities in the portfolio. The value of the portfolio at the risk horizon τ is equal to the expected discounted cash flows of the portfolio over the interval $[\tau, T]$ under the risk-neutral distribution conditioned on the scenario $\omega^{(i)}$. This can be approximated by the sample average of m independent and identically distributed samples of the discounted cash flows. In other words, the portfolio loss $L(\omega^{(i)})$ is estimated by a quantity $\hat{L}(\omega^{(i)}, \zeta^{(i)})$. Here, $\zeta^{(1)}, \ldots, \zeta^{(n)}$ are independent random variables that capture the randomness of the inner stage simulation, and are identically distributed.[2] This procedure is illustrated in Figure 1.

We make the following standard assumption:

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*

$$\mathsf{E}[\hat{L}(\omega, \zeta) \mid \omega] = L(\omega), \tag{2}$$

and

$$\mathrm{Var}(\hat{L}(\omega, \zeta) \mid \omega) = \frac{v(\omega)}{m} < \infty. \tag{3}$$

Also, the conditional variance $v(\omega)$ satisfies

$$\mathsf{E}[v(\omega)] < \infty. \tag{4}$$

Notice that (2) states that the portfolio loss estimate is unbiased. Equation (3) implies that the conditional variance of the portfolio loss estimate decays at the rate m^{-1} as a function of the number of inner stage samples m, with a

scenario-dependent constant $v(\omega)$. Equation (4) implies that the portfolio loss estimate has finite second moment.

We approximate the risk measure α according to the empirical distribution of the portfolio loss estimates that arise from this two-stage, nested simulation procedure. Specifically, given loss estimates

$$\hat{L}(\omega^{(1)}, \zeta^{(1)}), \ldots, \hat{L}(\omega^{(n)}, \zeta^{(n)}),$$

the *standard nested* estimator is defined by

$$\hat{\alpha}_{\mathsf{SN}(m,n)} \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\hat{L}(\omega^{(i)}, \zeta^{(i)})). \tag{5}$$

Standard nested simulation has been studied by a number of authors, e.g., Lee (1998), Lee and Glynn (2003), Gordy and Juneja (2008), Gordy and Juneja (2010), and Hong and Juneja (2009). The analysis is based on two criteria: computational effort and accuracy. We summarize these results as follows.

The computational effort to compute the estimator $\hat{\alpha}_{\mathsf{SN}(m,n)}$ is determined by the choice of the two parameters: n, the number of scenarios, and m, the number of inner stage samples per scenario. Given a choice of (m, n), a total of n outer stage scenarios and mn inner stage samples is required. Generally, the computational effort is dominated by the time required for inner stage samples. This is because, in practice, the time horizon $(0, \tau]$ corresponding to the outer stage of simulation is much shorter than the time horizon $(\tau, T]$ corresponding to the inner stage. For example, if we are interested in a daily risk measure of a mortgage portfolio, the outer stage time horizon is a single day, whereas the inner stage time horizon may be as long as 30 years. Moreover, the inner stage involves not only the simulation of future risk factors but also the computation of security payoffs. The computation of discounted payoffs may involve the evaluation of complicated rules or models, for example, as in the case of a mortgage portfolio. This may also require significant computational effort. Therefore, as a proxy for the total computational effort, we use the total number of inner stage samples,[3] denoted by $k \triangleq mn$.

The accuracy of the estimator also depends on the choice of (m, n). We measure the accuracy according to the mean squared error of the estimator. The MSE $\hat{\alpha}_{\mathsf{SN}(m,n)}$ can be decomposed into the variance and the squared bias; i.e.,

$$\mathsf{E}[(\hat{\alpha}_{\mathsf{SN}(m,n)} - \alpha)^2] = \underbrace{\mathsf{E}[(\hat{\alpha}_{\mathsf{SN}(m,n)} - \mathsf{E}[\hat{\alpha}_{\mathsf{SN}(m,n)}])^2]}_{\text{variance}} + \underbrace{(\mathsf{E}[\hat{\alpha}_{\mathsf{SN}(m,n)} - \alpha])^2}_{\text{bias}^2}. \tag{6}$$

Under appropriate technical assumptions, the asymptotic variance depends only on the number of outer stage scenarios n and decays as n^{-1}, and the asymptotic bias depends only on the number of inner stage samples per scenario m and decays as m^{-1} (see, e.g., Hong and Juneja 2009).

Given a fixed computational budget k, we can choose the parameters (m, n) so as to minimize the MSE of the estimator. Thus, an optimal estimator can be found by solving the optimization problem

$$\begin{aligned} \underset{m,\,n}{\text{minimize}} \quad & \mathsf{E}[(\hat{\alpha}_{\mathsf{SN}(m,n)} - \alpha)^2] \\ \text{subject to} \quad & mn = k, \\ & m, n \geqslant 0. \end{aligned} \tag{7}$$

Using the decomposition (6) and the asymptotic rates of decay of variance and bias, it follows that the asymptotically optimal choice m^* is of order $k^{1/3}$, and the optimal choice of n^* of order $k^{2/3}$. With these choices, the asymptotically optimal MSE of $\hat{\alpha}_{\mathsf{SN}(m,n)}$ decays as a function of the total number of inner stage samples k at rate $k^{-2/3}$.

To interpret this result, it is instructive to compare it with traditional Monte Carlo simulation, which only needs a single stage of simulation. There, the MSE decays as a function of the number of scenarios k at rate k^{-1}. The rate for nested simulation is slower because of the inner stage Monte Carlo estimates of portfolio loss. Although these estimators are unbiased relative to the true portfolio loss L, the risk measure α may not be a linear function of L, so an additional bias is introduced. This slows down the convergence.

Remark 1. Although we know the asymptotic orders of magnitudes of m^* and n^*, their asymptotic coefficients are difficult to derive. Therefore, in general cases, it is not clear how to solve (7) given a finite k, which makes the optimal performance of $\hat{\alpha}_{\mathsf{SN}(m,n)}$ unachievable. To address this, Broadie et al. (2011) suggest a method to adaptively estimate (m^*, n^*) in the context of a sequential nested simulation.

2.2. Delta-Gamma Approximation

The delta-gamma approximation takes an alternative approach to estimate the portfolio loss $L(\omega)$ in a scenario ω. Here, $L(\omega)$ is approximated by a quadratic function of underlying risk factors ω, using a second order local approximation. This method is discussed, for example, by Rouvinez (1997) and Glasserman et al. (2000). Consider a single representative scenario ω^*. For example, the scenario ω^* can be selected to be the expected value of ω under the real-world distribution at time τ. If $L(\cdot)$ is a twice differentiable function in a neighborhood of ω^*, then a quadratic approximation can be made as

$$\tilde{L}_{\mathsf{DG}}(\omega) \triangleq L(\omega^*) + \nabla L(\omega^*)^\top(\omega - \omega^*) + \frac{1}{2}(\omega - \omega^*)^\top \nabla^2 L(\omega^*)(\omega - \omega^*). \tag{8}$$

Given $\tilde{L}_{\mathsf{DG}}$, the risk measure α can be estimated by

$$\hat{\alpha}_{\mathsf{DG}} \triangleq \mathsf{E}[f(\tilde{L}_{\mathsf{DG}}(\omega))]. \tag{9}$$

This is referred to as the *delta-gamma* estimator.

Note that this quadratic approximation requires the knowledge of the portfolio loss $L(\omega^*)$ as well as the *delta* (i.e., the gradient) $\nabla L(\omega^*)$ and the *gamma* (i.e., the Hessian matrix) $\nabla^2 L(\omega^*)$ at the representative scenario ω^*. In general, these quantities may be analytically unavailable in closed form. However, since they are only needed in a single scenario, they can be computed to arbitrary accuracy without excessive computational effort. For the purposes of our discussion, we assume they are known exactly.

Further, given the quadratic approximation $\tilde{L}_{\mathrm{DG}}$, we need to evaluate the expectation in (9). Observe that for any scenario ω, evaluating $\tilde{L}_{\mathrm{DG}}(\omega)$ is easy since it involves only basic vector operations and, in particular, requires no simulation. Thus, with a modest computational effort, we can perform a single stage Monte Carlo simulation to approximate the estimator $\hat{\alpha}_{\mathrm{DG}}$ to a high degree of accuracy. For the purpose of discussion, we assume the expectation in (9) can be exactly evaluated.

3. The Regression Algorithm

We now introduce a method that is based on regression. The idea is to approximate the portfolio loss $L(\cdot)$ by an approximation that is easy to evaluate. This is reminiscent of delta-gamma approximation. However, delta-gamma approximation has two major restrictions. First, it uses only a quadratic approximation. We allow higher order approximations and allow approximations that can be tailored based on knowledge of the portfolio. Second, the delta-gamma approximation computes a *local* approximation around some representative scenario. There is no reason to expect that such an approximation will accurately describe the portfolio loss across a broad set of scenarios. On the other hand, we will attempt to find an approximation that is *globally* good.

In particular, consider a set of d real-valued functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ on the state space Ω, which we will call basis functions. The basis functions can be written as a row vector

$$\Phi(\omega) \triangleq (\phi_1(\omega), \ldots, \phi_d(\omega)) \in \mathbb{R}^d,$$

for each scenario ω. We seek to approximate the portfolio loss function $L(\cdot)$ by a linear combination of these basis functions. In other words, we would like to find a column vector $r \in \mathbb{R}^d$ so that for each scenario ω,

$$L(\omega) \approx \Phi(\omega) r \triangleq \sum_{l=1}^{d} \phi_l(\omega) r_l.$$

We will then estimate the risk measure α using this approximation.

There are two requirements for this procedure to be effective:

(i) The basis functions should incorporate features of the state space relevant to determining the portfolio loss so that a linear combination of these functions can accurately approximate the portfolio loss.

(ii) The basis functions should be fast to evaluate. Then when using an approximation defined by the basis functions, the outer stage expectation in the risk measure can be computed quickly.

In general, a generic choice of basis functions can always be made (e.g., all low order polynomials). However, the intelligent selection of basis functions is problem dependent. For example, given a portfolio of exotic derivatives, the values of corresponding plain vanilla derivatives (which can be obtained via closed form analytical expressions) might be used to construct basis functions. We will see examples of this in the numerical case studies of §5.

Given the basis functions Φ, a global approximation to the portfolio loss L can be found by solving a minimum mean squared error problem

$$r^* \in \operatorname*{arg\,min}_{r \in \mathbb{R}^d} \ \mathsf{E}[(L(\omega) - \Phi(\omega) r)^2]. \tag{10}$$

We can then approximate the risk measure by $\mathsf{E}[f(\Phi(\omega) r^*)]$. Given the optimal regression coefficients r^*, for each scenario ω, define $M(\omega)$ to be the *model error* of the approximation $\Phi(\omega) r^*$,

$$M(\omega) \triangleq L(\omega) - \Phi(\omega) r^*. \tag{11}$$

Here, $M(\cdot)$ represents the residual error under the best approximation afforded by the basis functions Φ. Further, define $\varepsilon(\omega, \zeta)$ by

$$\begin{aligned} \varepsilon(\omega, \zeta) &\triangleq \hat{L}(\omega, \zeta) - L(\omega) \\ &= \hat{L}(\omega, \zeta) - \Phi(\omega) r^* - M(\omega). \end{aligned} \tag{12}$$

This quantity measures the discrepancy between the Monte Carlo estimate of portfolio loss in the scenario ω and the true portfolio loss.

In order for this regression procedure to be well defined, we make the following assumption:

Assumption A2. *The second moments of $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are finite; i.e., $\mathsf{E}[\phi_l(\omega)^2] < \infty$ for each $l = 1, \ldots, d$. Further, $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are linearly independent; i.e., when $n \geqslant d$,*

$$\mathsf{P}\left(\operatorname{rank} \begin{pmatrix} \Phi(\omega^{(1)}) \\ \vdots \\ \Phi(\omega^{(n)}) \end{pmatrix} = d \right) = 1.$$

Without loss of generality, we can assume that the functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are orthonormal, i.e., we assume that $\mathsf{E}[\Phi(\omega)^\top \Phi(\omega)]$ is the identity matrix.

Assumption A2 ensures that the basis functions are linearly independent. If this is the case, given finite second moments, there is no loss of generality in assuming that they are orthonormal. Otherwise, one could construct an equivalent orthonormal basis through the Gram-Schmidt procedure. We will assume the orthonormality for the rest of this paper because it greatly simplifies the exposition.

Given Assumptions A1 and A2, the optimal solution r^* to the optimization problem (10) exists and is unique. However, it is not possible to directly compute r^* since we cannot evaluate $L(\cdot)$ in general. Instead, our method seeks to solve an analog of the optimization problem (10) that is obtained by nested simulation. In particular, in order to get a tractable problem, we will first replace the expectation in (10) with a sample average over scenarios; then the portfolio loss in each scenario can be estimated by inner stage Monte Carlo simulation. As in §2.1, suppose there are n scenarios $\vec{\omega} \triangleq (\omega^{(1)}, \ldots, \omega^{(n)})^\top$. In each scenario $\omega^{(i)}$, let $\zeta^{(i)}$ be an i.i.d. random variable that captures the randomness of the corresponding m inner stage samples so that $\hat{L}(\omega^{(1)}, \zeta^{(1)}), \ldots, \hat{L}(\omega^{(n)}, \zeta^{(n)})$ are the nested Monte Carlo portfolio loss estimates across scenarios. Define the vector $\vec{\zeta} \triangleq (\zeta^{(1)}, \ldots, \zeta^{(n)})^\top$. Given $(\vec{\omega}, \vec{\zeta})$, we solve the optimization problem

$$\hat{r} \in \arg\min_{r \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} (\hat{L}(\omega^{(i)}, \zeta^{(i)}) - \Phi(\omega^{(i)}) r)^2. \tag{13}$$

Note that this is a standard ordinary least squares problem. Given the coefficient vector $\hat{r}$, we estimate the risk measure α by

$$\hat{\alpha}_{\mathsf{REG}(m,n)} \triangleq \mathsf{E}[f(\Phi(\omega)\hat{r}) \mid \vec{\omega}, \vec{\zeta}]. \tag{14}$$

We define $\hat{\alpha}_{\mathsf{REG}(m,n)}$ to be the *regression* estimator. Given Assumptions A1 and A2, the optimal solution $\hat{r}$ in (13) exists and is unique almost surely. Hence, our estimator is well defined.

Remark 2. Observe that the expectation in (14) can be estimated via a single stage Monte Carlo simulation that only requires evaluation of the basis functions. Since we assume that the basis functions are fast to compute, it will be possible to approximate $\hat{\alpha}_{\mathsf{REG}(m,n)}$ to a high degree of accuracy, given modest computational effort. Indeed, if we sample n' additional outer stage scenarios, and we estimate (14) with

$$\frac{1}{n'} \sum_{i=1}^{n'} f(\Phi(\omega_i')\hat{r}), \tag{15}$$

the MSE between (14) and (15) is in the order of $(n')^{-1}$. We will see that this is asymptotically negligible compared to the MSE of the estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ (e.g., Corollary 2). Therefore, although the estimator (15) would typically be employed in practice,[4] for the purposes of discussion and analysis, we assume (14) can be exactly computed.[5]

Remark 3. The regression model has been specified with portfolio loss as the dependent variable. An alternative is to specify and estimate a distinct linear model for the value of each of the instruments that constitute the portfolio. This alternative[6] could have important practical advantages. For example, portfolios can change rapidly, but the set of instruments in which an entity trades tends to persist for long periods. For each instrument, we can estimate a linear model just once and possibly use the same model in risk-measurement exercises in the future. By contrast, because of changes in portfolio composition, a portfolio-level model calibrated to the current portfolio will not (in general) be useful in the future. Note that if the same set of basis functions is used, there is no difference between the models obtained by regression at the portfolio level and by regression at the instrument level; hence our present formulation is without loss of generality.

The regression of (13) is *unweighted* in that each scenario is equally weighted. One can imagine *weighted* variations as well; e.g.,

$$\hat{r} \in \arg\min_{r \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} h(\omega^{(i)})(\hat{L}(\omega^{(i)}, \zeta^{(i)}) - \Phi(\omega^{(i)}) r)^2,$$

where $h(\cdot)$ is a weight function. Note that this is a standard weighted least squares problem. Flexibility in choosing a weight function can allow emphasis in fitting the loss accurately in scenarios that have the greatest impact on the overall risk calculation. We will describe and analyze weighted generalizations of our regression estimator in §6. More generally, there is no reason the outer stage scenarios need to be generated by simulation at all. Instead, they could be considered as points specified in a designed experiment. We will not explore this alternative further, but it is an interesting direction for future research.

4. Analysis

In this section, we provide theoretical analysis of the regression method presented in §3. Here we are interested in the asymptotic squared error of the regression method as the number of scenarios n tends to infinity. Our analysis consists of two separate cases with different assumptions on the loss function $f(\cdot)$.

In §4.1 we consider the first case, where the function $f(\cdot)$ is assumed to be twice differentiable. Our analysis here relies on existing results in econometrics for the asymptotic analysis of linear regression (White 2001), including the asymptotic normality of regression estimates and the delta method. We establish that the squared error converges in probability at the rate n^{-1} until it reaches an asymptotic bias level at which the error ceases to improve. In §4.2 we consider the second case, where the function $f(\cdot)$ is assumed to be Lipschitz continuous. This line of analysis is quite different and builds on theory for the asymptotic analysis of sample average approximations in optimization

(Shapiro et al. 2009). Here, under different probabilistic assumptions, we establish that, in fact, the squared error converges in mean at the rate $n^{-1+\delta}$, for any $\delta > 0$, until it reaches an asymptotic bias level. Moreover, we establish bounds on the mean squared error for finite n.

In both cases, the asymptotic level of bias is bounded by the model error associated with the basis functions. Although the exact assumptions and conclusions differ in the two analyses, taken together, the spirit of these results is to suggest that our regression approach will converge at the same rate as traditional non-nested Monte Carlo simulation over a large range, given a suitably good choice of basis functions.

All proofs for this section are provided in the online supplement (available as supplemental material at http://dx.doi.org/10.1287/opre.2015.1419).

4.1. Differentiable Case

In the first case, we make the following differentiability assumption:

Assumption F1. *The function $f(\cdot)$ is twice differentiable with bounded second derivative so that there exists a scalar U_{diff} with*

$$|f''(L)| \leqslant U_{\text{diff}} \quad \text{for any } L \in \mathbb{R}. \tag{16}$$

In order to characterize the asymptotic distribution of the regression estimator, we make the following technical assumption:

Assumption A3. *The matrix $\mathsf{E}[v(\omega)\Phi(\omega)^\top\Phi(\omega)]$ is positive definite, and*

$$\mathsf{E}[\phi_l(\omega)^2 M(\omega)^2] < \infty,$$

for each $l = 1, \ldots, d$.

Assumption A3 is a technical assumption standard in regression theory (see, e.g., White 2001). To begin our analysis, we have the following lemma that characterizes the convergence[7] of coefficients of the regression estimator. This lemma is based on the asymptotic normality of ordinary least squares estimators, applied to the present setting. In what follows, denote by $N(\cdot)$ the cumulative distribution function for the normal distribution.

Lemma 1. *Suppose Assumptions A1–A3 hold. As the number of scenarios $n \to \infty$,*

$$\sqrt{n}(\hat{r} - r^*) \xrightarrow{\text{d}} N\left(\mathbf{0}, \Sigma_M + \frac{\Sigma_v}{m}\right),$$

where

$$\begin{aligned} \Sigma_M &\triangleq \mathsf{E}[M^2(\omega)\Phi(\omega)^\top\Phi(\omega)], \\ \Sigma_v &\triangleq \mathsf{E}[v(\omega)\Phi(\omega)^\top\Phi(\omega)]. \end{aligned} \tag{17}$$

Therefore, as $n \to \infty$,

$$\|\hat{r} - r^*\|_2 = \frac{O_{\text{P}}(1)}{\sqrt{n}}.$$

According to Lemma 1, as the number of scenarios $n \to \infty$, the estimated coefficients $\hat{r}$ converge to the optimal coefficients r^* at the rate $n^{-1/2}$ in probability. However, we are interested not only in the convergence of regression coefficients but also in the convergence of the resulting estimated risk measure. To this end, we have the following result, which is based on the multivariate delta method of establishing asymptotic normality applied to the present setting:

Theorem 1. *Suppose that Assumptions F1, A1–A3 hold. Then there exists a sequence of random variables $\{B_{M,n}\}$, for $n = 1, 2, \ldots,$ satisfying*

$$B_{M,n} \xrightarrow{\text{P}} B_M^* \triangleq \mathsf{E}[f(\Phi(\omega)r^*)] - \alpha,$$

so that

$$\begin{aligned} &\sqrt{n}(\hat{\alpha}_{\text{REG}(m,n)} - \alpha - B_{M,n}) \\ &\quad \xrightarrow{\text{d}} N\Bigg(0, \mathsf{E}[f'(L(\omega))\Phi(\omega)]\left(\Sigma_M + \frac{\Sigma_v}{m}\right) \\ &\qquad \cdot (\mathsf{E}[f'(L(\omega))\Phi(\omega)])^\top\Bigg), \end{aligned} \tag{18}$$

where Σ_M and Σ_v are defined by (17). Further, the asymptotic bias B_M^* satisfies

$$|B_M^* - \mathsf{E}[f'(L(\omega))M(\omega)]| \leqslant \frac{U_{\text{diff}}}{2}\mathsf{E}[(M(\omega))^2]. \tag{19}$$

Theorem 1 establishes three points: first, observe that the quantity B_M^* is the asymptotic bias of the regression estimator. The bounds in (19) indicate that the asymptotic bias B_M^* is controlled by the model error $M(\cdot)$, i.e., the quality of the best approximation under the basis functions. In particular, if the basis functions are chosen so that the model error is small, the asymptotic bias will also be small. Second, (18) suggests that the error of the regression estimator decreases at the rate $n^{-1/2}$ in probability until it reaches a level term that converges to a level that is dominated by the asymptotic bias, at which point the estimator ceases to improve. Third, for large n, the quantity $\hat{\alpha}_{\text{REG}(m,n)} + B_{M,n}$, which is the regression estimate with bias corrected, is approximately normal with mean α and variance

$$\frac{1}{n}\mathsf{E}[f'(L(\omega))\Phi(\omega)]\left(\Sigma_M + \frac{\Sigma_v}{m}\right)(\mathsf{E}[f'(L(\omega))\Phi(\omega)])^\top.$$

Given a fixed number of inner stage samples $k = mn$, Theorem 1 suggests that the choice of m and n do not impact the asymptotic bias but do affect the asymptotic variance. Hence, it is clear that the asymptotically minimum squared error is achieved when $m = 1$ and $n = k$.

In order to further interpret Theorem 1, consider, as a special case, the following corollary:

COROLLARY 1. *Suppose that Assumptions F1, A1–A3 hold. When $m = 1$, $n = k$, and the portfolio loss L is in the span of the basis functions Φ, then $M(\omega) \equiv 0$, and we have*

$$\sqrt{k}(\hat{\alpha}_{\mathrm{REG}(m,n)} - \alpha) \xrightarrow{\mathrm{d}} N(0, \mathsf{E}[f'(L(\omega))\Phi(\omega)]\Sigma_v(\mathsf{E}[f'(L(\omega))\Phi(\omega)])^\top), \quad (20)$$

as $k \to \infty$.

Moreover, suppose that the conditional variance of inner samples does not depend on the scenario; i.e., $v(\omega) \equiv v$ for every scenario ω. Then there exists a constant scalar v^* defined by

$$v^* \triangleq v\mathsf{E}[f'(L(\omega))\Phi(\omega)](\mathsf{E}[f'(L(\omega))\Phi(\omega)])^\top \leqslant v\mathsf{E}[(f'(L(\omega)))^2],$$

such that, as $k \to \infty$,

$$\sqrt{k}(\hat{\alpha}_{\mathrm{REG}(m,n)} - \alpha) \xrightarrow{\mathrm{d}} N(0, v^*).$$

According to Corollary 1, the error scales as $k^{-1/2}$ in probability as a function of the total number of inner samples k. By applying the continuous mapping theorem, we have that, as $k \to \infty$, the quantity $k(\hat{\alpha}_{\mathrm{REG}(m,n)} - \alpha)^2$ converges to a random variable. This is the same rate of convergence as the convergence rate in the case of non-nested Monte Carlo simulation.

4.2. Lipschitz Continuous Case

Section 4.1 established theoretical results when $f(\cdot)$ has bounded second derivative everywhere. Since many risk measures of interest arise when $f(\cdot)$ is not twice differentiable, in this section, we investigate the convergence of the regression estimator under the alternative assumption of Lipschitz continuity:

ASSUMPTION F2. *The function $f(\cdot)$ is Lipschitz continuous; i.e., there exists a scalar U_{Lip}, such that*

$$|f(L') - f(L'')| \leqslant U_{\mathrm{Lip}}|L' - L''|, \quad \text{for any } L', L'' \in \mathbb{R}. \quad (21)$$

Under this assumption, we can bound the asymptotic squared error of the regression estimator $\hat{\alpha}_{\mathrm{REG}(m,n)}$ as follows:

THEOREM 2. *Suppose that Assumptions F2, A1–A3 hold. Then as the number of scenarios $n \to \infty$,*

$$(\hat{\alpha}_{\mathrm{REG}(m,n)} - \alpha)^2 \leqslant U_{\mathrm{Lip}}^2\mathsf{E}[(M(\omega))^2] + O_{\mathsf{P}}\left(\frac{1}{n}\right).$$

According to Theorem 2, the squared error is bounded above by a random variable that decays at the rate n^{-1} in probability plus a constant that is a function of the model error, i.e., the quality of the basis functions. This is analogous to the conclusion of Theorem 1, as discussed earlier. As in that case, given a computational budget $k = mn$ on the total number of inner stage samples, the bound on the asymptotic squared error is minimized when $m = 1$ and $n = k$.

The result of Theorem 2 provides an asymptotic bound on the convergence of the squared error in probability. This can be strengthened to bounding the *mean squared error* of the regression estimator $\hat{\alpha}_{\mathrm{REG}(m,n)}$, both asymptotically and for finite n. In order to do so, we will apply the methodology of Shapiro et al. (2009). To this end, we make the following technical assumptions:

ASSUMPTION A4. *The moment generating functions of $\|\Phi(\omega)\|_2^2$, $(M(\omega))^2$, and $(\varepsilon(\omega, \zeta))^2$ are finite valued in a neighborhood of zero.*

In our problem, define

$$G(r, \omega, \zeta) \triangleq (\hat{L}(\omega, \zeta) - \Phi(\omega)r)^2$$

and

$$g(r) \triangleq \mathsf{E}[G(r, \omega, \zeta)].$$

In other words, $G(r, \omega, \zeta)$ is the squared error of the regression estimate with coefficient vector r versus the standard nested estimate in a single scenario, and $g(r)$ is the mean squared error across all scenarios. For any $\rho > 0$, define the set

$$\mathcal{R}_\rho \triangleq \{r \in \mathbb{R}^d : \|r - r^*\|_2^2 \leqslant \rho\}, \quad (22)$$

which is a compact and convex neighborhood of r^*.

ASSUMPTION A5. *For any $\rho > 0$, there exists a constant $\lambda > 0$ such that for any $r', r'' \in \mathcal{R}_\rho$, the moment generating function $\Psi_{r',r''}(t)$ of the random variable*

$$(G(r', \omega, \zeta) - g(r')) - (G(r'', \omega, \zeta) - g(r''))$$

satisfies, for any $t \in \mathbb{R}$,

$$\Psi_{r',r''}(t) \leqslant \exp(\rho\lambda^2 t^2).$$

Informally, Assumption A5 requires that $(G(r', \omega, \zeta) - g(r')) - (G(r'', \omega, \zeta) - g(r''))$ has sub-Gaussian tails. Given the assumptions above, we have the following lemma:

LEMMA 2. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold. Let $\rho > 0$ be an arbitrary constant. Then for any positive integer n,*

$$\mathsf{P}(\hat{r} \notin \mathcal{R}_\rho) \leqslant \left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right),$$

where λ is defined in Assumptions A5; C' and C'' are universal constants (i.e., constants that do not depend on the problem); and

$$\Lambda_\rho \triangleq (2\sqrt{\rho} + 1)d + 2\mathsf{E}[(M(\omega))^2] + 2\mathsf{E}[(\varepsilon(\omega, \zeta))^2].$$

Lemma 2 bounds the probability that the estimated regression coefficients $\hat{r}$ are not in the fixed neighborhood $\mathcal{R}_\rho$ of the optimal coefficients r^* and demonstrates that this probability decays exponentially as $n \to \infty$. Lemma 2 is not only an asymptotic result but a finite-sample result that holds for every n. Given Lemma 2, we can establish the following theorem:

THEOREM 3. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold, and let $\delta > 0$ be an arbitrary positive constant. Then for any positive integer n,*

$$\begin{aligned}
&\mathsf{E}[(\Phi(\omega)(\hat{r}-r^*))^2] = \mathsf{E}[\|\hat{r}-r^*\|_2^2] \\
&\leqslant \frac{1}{n^{1-\delta}} + 2^{(3d)/2} C'(C'')^d (\Lambda_2)^d \lambda^2 n^{((1-\delta)d/2)-1} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right) \\
&\quad + \frac{2^d C'(C'')^d (\Lambda_2)^d \lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right) \\
&= O(n^{-1+\delta}).
\end{aligned}$$

Theorem 3 establishes that the squared error between our approximation of the loss function and the best possible approximation using the same basis functions decays at the rate $n^{-1+\delta}$ for any $\delta > 0$. A corollary of Theorem 3 is our main result, which establishes the rate of convergence of the MSE of the risk estimator:

COROLLARY 2. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold, and let $\delta > 0$ be an arbitrary positive constant. Then for any positive integer n,*

$$\begin{aligned}
&\mathsf{E}[(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha)^2] \\
&\leqslant U_{\mathrm{Lip}}^2 \left(2^{(3d)/2} C'(C'')^d (\Lambda_2)^d \lambda^2 n^{((1-\delta)d)/2-1} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right) \right. \\
&\qquad \left. + \frac{2^d C'(C'')^d (\Lambda_2)^d \lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right)\right) \\
&\quad + U_{\mathrm{Lip}}^2 (\mathsf{E}[(M(\omega))^2] + n^{-1+\delta}) \\
&= U_{\mathrm{Lip}}^2 \mathsf{E}[(M(\omega))^2] + O(n^{-1+\delta}).
\end{aligned}$$

According to Corollary 2, first, the MSE of the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ decays at the rate $n^{-1+\delta}$ for any $\delta > 0$ until it hits an asymptotic bias level, which is a function of the model error; this is reminiscent of Theorem 2 except that we have δ in the exponent of $n^{-1+\delta}$ here. Second, Corollary 2 holds for any arbitrary n, which is a finite-sample result rather than an asymptotic result in Theorem 2. Third, Corollary 2 implies the convergence of the MSE, which is stronger than the convergence of the squared error in probability in Theorem 2.

5. Numerical Results

In this section we use several examples to compare the relative performance of standard nested simulation, the delta-gamma approximation, and our regression method. In §5.2, we present a simple example where there is only one underlying asset. In §5.3, we have four examples with portfolios of multiple assets.

5.1. Experimental Setting

Our examples involve portfolios consisting of one or more underlying assets as well as derivatives based on them. We assume that all underlying asset prices follow geometric Brownian motion processes and that option prices are determined according to the standard single-asset Black-Scholes model and its multi-asset generalization.

Specifically, assume risk factors $\omega \triangleq (\omega_1, \ldots, \omega_Q) \in \Omega \subset \mathbb{R}^Q$ are distributed according to a multivariate Gaussian distribution with mean zero, variance one, and correlations specified by a given correlation matrix. Given ω and the risk horizon τ, define $S_\tau(\omega)$ to be the prices of underlying assets at time τ, with

$$S_\tau(\omega) \triangleq (S_{1,\tau}(\omega), \ldots, S_{Q,\tau}(\omega)),$$

where

$$S_{j,\tau}(\omega) = S_{j,0} \exp((\mu_j - \sigma_j^2/2)\tau + \sigma_j \sqrt{\tau}\omega_j).$$

Here $S_{j,0}$ is the price of the jth asset at time 0, μ_j is the drift of the jth asset under the real-world distribution, and σ_j is the annual volatility of the jth asset. In this setting, asset prices are lognormally distributed and there is exactly one risk factor per asset.

To estimate the portfolio loss at time τ we use Monte Carlo simulation of the inner stage sample paths under the risk-neutral distribution to generate asset cash flows between times τ and T. For the jth security and for each inner stage sample path $p = 1, \ldots, m$, define $W_{j,t}^{(p)}$ to be a Brownian motion for $t \in [\tau, T]$ with $W_{j,\tau}^{(p)} = 0$. Notice that there is no correlation between $W_{j',t}^{(p')}$ and $W_{j'',t}^{(p'')}$ if $p' \neq p''$, i.e., the m inner stage sample paths are independent; on the other hand, there could exist correlation between securities on the same path. The set

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

represents all of the inner stage uncertainty given the outer stage scenario ω. Specifically, conditional on ω, we assume that the risk-neutral asset prices on the pth sample path are given by

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

for $t \in [\tau, T]$ and $j = 1, \ldots, Q$, where r_f is the continuously compounded riskless rate of interest.

For each $p = 1, \ldots, m$, the discounted portfolio cash flows along the pth sample path may, in general, depend on asset prices at all times between τ and T. Typically, however, portfolio cash flows only depend on prices at a finite number of times, so only these essential points are simulated. The portfolio loss estimate $\hat{L}(\omega, \zeta)$ is determined by the average over the m sample paths. We focus on the expected excess loss risk measure:

$$\mathsf{E}[(L(\omega) - c)^+],$$

where c is a threshold that will be specified in each example.

In each of the following examples, we compare the accuracy of a number of methods that estimate the risk measure. The closed form expression for the portfolio losses $L(\omega)$ given a risk factor scenario ω is known in all of the examples except example $\mathsf{EX}_{5\mathsf{H}}$. In these cases, we will precisely compute the risk measure α by using non-nested Monte Carlo simulation in the outer stage. Similarly, the delta-gamma estimator[8] $\hat{\alpha}_{\mathsf{DG}}$ and the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ provide approximations of portfolio losses and also require non-nested simulation in the outer stage to estimate the risk measure. For all of these non-nested simulations, we employ either stratified sampling or a Sobol sequence (see, e.g., Sobol 1967, Glasserman 2004).

Note that for the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$, consistent with the theory we have developed, a single inner stage sample is used for each outer stage scenario ($m = 1$), and the outer stage scenarios that used to estimate risk are independent of those used to estimate regression coefficients. For standard nested simulation estimation, the asymptotically optimal choices of the numbers of outer stage scenarios and the inner stage scenarios are given by $n = \beta k^{2/3}$ and $m = k^{1/3}/\beta$, where β is a constant. In all cases except Example $\mathsf{EX}_{5\mathsf{H}}$, the optimal standard nested estimator corresponds to the choice of β that minimizes the asymptotic MSE; this is determined by trying many different values of β. Doing this would not usually be feasible in practice, and hence these results for the "optimal" standard nested estimator may be optimistic. Example $\mathsf{EX}_{5\mathsf{H}}$ was meant to be a real-world example, and hence the standard nested estimator in that case is computed with $\beta = 1$; i.e., $n = k^{2/3}$ and $m = k^{1/3}$, without further search for the optimal β.

Details are given in what follows.

5.2. Single Asset Example

• *Model*: There is one asset with initial price $S_0 = 100$, real-world drift $\mu = 8\%$, and volatility $\sigma = 20\%$. The interest rate is $r_f = 3\%$.

• *Portfolio*: The portfolio consists of three barrier options[9] that mature at time $T = 1/12$ year. The risk horizon is $\tau = 1/52$ year.[10]

1. Long one down-and-out put option with strike $K_1 = 101$ and barrier $H_1 = 91$.

2. Long one down-and-out put option with strike $K_2 = 110$ and barrier $H_2 = 100$.

3. Short one down-and-out put option with strike $K_3 = 114.5$ and barrier $H_3 = 104.5$.

The loss threshold is set to $c = 0.3608$, the 95th percentile of $L(\omega)$.

• *Basis functions*: We test several sets of basis functions. For simplicity, denote the asset price at time τ by $S_\tau(\omega)$.

$\Phi^{(1)}$: This basis set includes quadratic functions of the asset price $S_\tau(\omega)$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_1(\omega) = 1$, $\phi_2(\omega) = S_\tau(\omega)$, and $\phi_3(\omega) = (S_\tau(\omega))^2$.

$\Phi^{(2)}$: This basis set includes all elements in $\Phi^{(1)}$ as well as $\phi_4(\omega) = (S_\tau(\omega) - H_1)^+$, $\phi_5(\omega) = (S_\tau(\omega) - H_2)^+$, $\phi_6(\omega) = (S_\tau(\omega) - H_3)^+$, and $\phi_7(\omega)$, $\phi_8(\omega)$, $\phi_9(\omega)$ defined as the squares of $\phi_4(\omega)$, $\phi_5(\omega)$, and $\phi_6(\omega)$, respectively.

$\Phi^{(3)}$: This basis set includes all elements of $\Phi^{(2)}$ as well as $\phi_{10}(\omega) = L(\omega)$, the exact expression for the loss determined using analytical formulas to value the derivatives.

• *Risk estimation*: We estimate the true risk measure α with a non-nested simulation using 10^6 stratified samples. The delta-gamma estimator $\hat{\alpha}_{\mathsf{DG}}$ and the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ are estimated with non-nested simulations using 10^4 samples.

The first basis set $\Phi^{(1)}$ was chosen because the regression method with this set is directly comparable to the delta-gamma approximation—the approximation employed in the delta-gamma method is within the span of this basis. The second basis set $\Phi^{(2)}$ was chosen to illustrate the benefit of introducing basis functions chosen based on problem specific knowledge. Here, the portfolio loss is known to depend on the barrier levels, and the additional basis functions can capture such a dependence in a piecewise linear way. The third basis set $\Phi^{(3)}$ represents the ideal case where the basis functions span the true portfolio loss function and there is no model error.

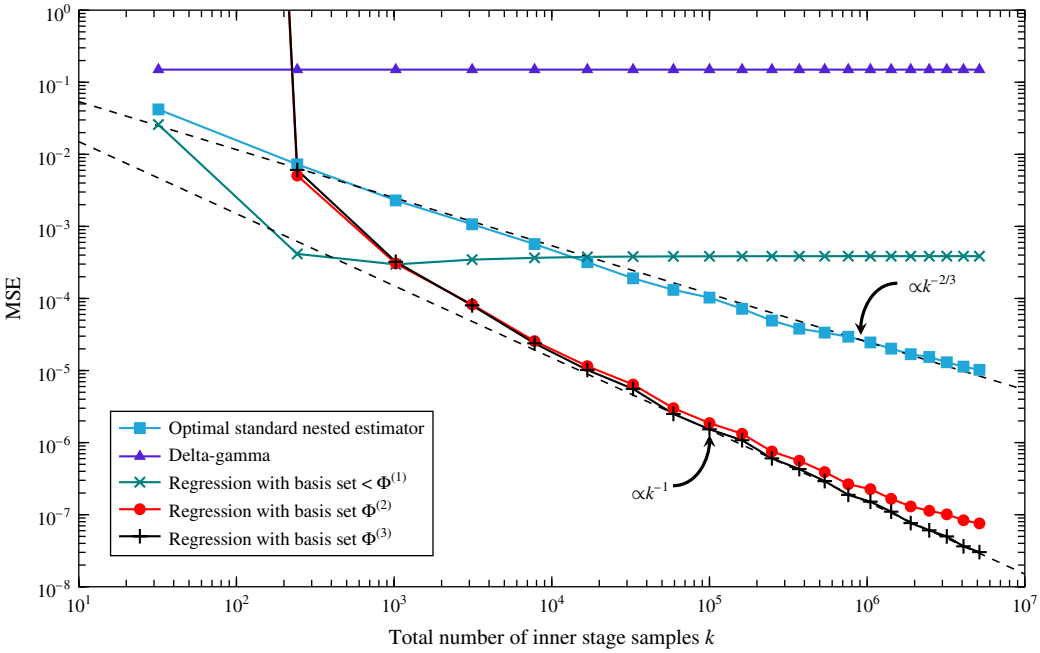
For each method, we simulate 1,000 independent trials to estimate the MSE of the estimator. Results are given in Figure 2. The MSE of the standard nested estimator decays at $k^{-2/3}$, which is consistent with the discussion in §2.1. Observe that the delta-gamma estimator is the least accurate of the five estimators. Figure 3 shows the local delta-gamma approximation around $S_\tau(\omega^*)$ does not provide a good global approximation. The regression method with basis set $\Phi^{(1)}$ fits a quadratic function globally and leads to a better risk estimate than the local approximation used in the delta-gamma approach.

In Figure 2, the MSE of the regression method with basis set $\Phi^{(1)}$ initially decays at the rate k^{-1} but then stops decreasing when the MSE reaches the asymptotic bias level determined by the regression model error. Basis sets $\Phi^{(2)}$ and $\Phi^{(3)}$ perform much better, with MSEs decaying at the rate k^{-1} over the current range in Figure 2. In fact, over the current range of simulation, the regression method with basis set $\Phi^{(2)}$, which is the basis set chosen more intelligently than $\Phi^{(1)}$, does about as well as the regression method with $\Phi^{(3)}$, whereas $\Phi^{(3)}$ contains the actual loss function in the basis set. Although the MSE of the regression method with $\Phi^{(2)}$ decays over a larger range than the one with $\Phi^{(1)}$, eventually, according to our theoretical results, the MSE of the regression method with $\Phi^{(2)}$ will also flatten out.

5.3. Multiple Asset Examples

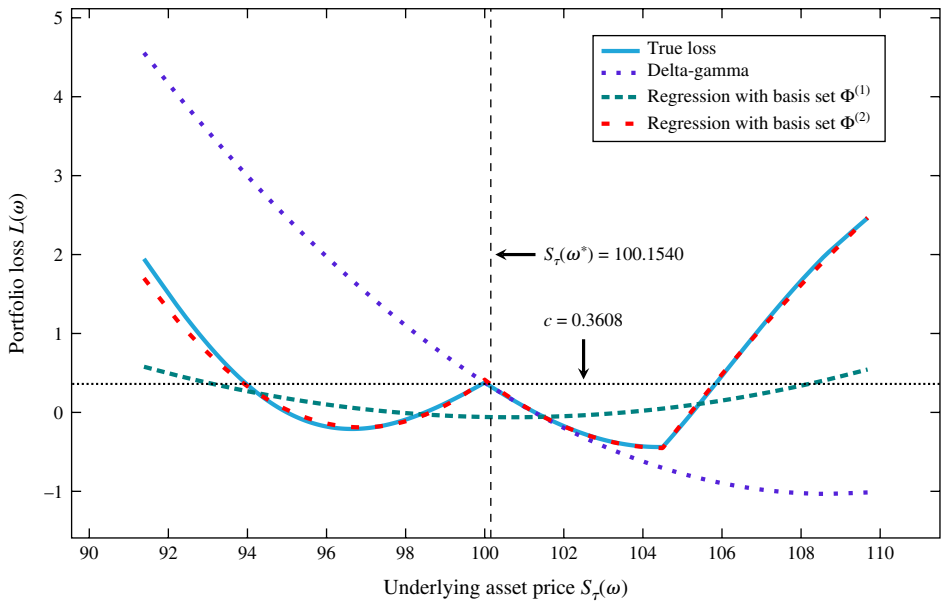
In this section we test four multi-asset examples. Common properties of these four examples are summarized here. Asset prices, denoted by $S_{j,t}(\omega)$, follow independent geometric Brownian motion processes unless otherwise stated. Initial asset prices are all $S_{j,0} = 100$. Assets share common real-world drifts of $\mu = 8\%$ and annual volatilities of $\sigma = 30\%$. The interest rate is $r_f = 5\%$. Derivatives all have strikes

Figure 2. (Color online) Illustration of the mean squared error in the single asset example.



Notes. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples. For the standard nested estimator, the optimized parameter value $\beta = 0.076$ was used.

Figure 3. (Color online) Illustration of approximations in the single asset example.



Note. 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.

$K = 100$ and a common maturity $T = 0.1$ year unless otherwise stated. The risk horizon is $\tau = 0.04$ year. We consider the following four examples:

Example EX_{10}. Ten underlying assets and a portfolio of derivatives with discontinuous payoff functions.

- *Portfolio*: The portfolio consists of three derivatives on each of the 10 assets: (a) short 10 down-and-out call options with barrier $H = 95$; (b) short 5 cash-or-nothing put options; and (c) an amount of underlying asset that delta hedges the two derivatives above.

The loss threshold is $c = 306.8763$, the 99th percentile of $L(\omega)$.

- *Basis functions*:

$\Phi^{(1)}$: This basis set includes quadratic functions of $S_{j,\tau}(\omega)$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l = 1, \ldots, 21$, where $\phi_1(\omega) = 1$, $\phi_{1+\iota}(\omega) = S_{\iota,\tau}(\omega)$ and $\phi_{11+\iota}(\omega) = (\phi_{1+\iota}(\omega))^2$ for $\iota = 1, \ldots, 10$.

$\Phi^{(2)}$: This basis set includes fifth order polynomials of $S_{j,\tau}(\omega)$'s and fifth order polynomials of $(S_{j,\tau}(\omega)-H)^+$ for $j=1,\ldots,10$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l=1,\ldots,101$, where $\phi_1(\omega)=1$, $\phi_{1+\iota}(\omega)=S_{\iota,\tau}(\omega)$ and $\phi_{11+\iota}(\omega)=(S_{\iota,\tau}(\omega)-H)^+$ for $\iota=1,\ldots,10$, and $\phi_{21+\iota}(\omega)=(\phi_{1+\iota}(\omega))^2$, $\phi_{41+\iota}(\omega)=(\phi_{1+\iota}(\omega))^3$, $\phi_{61+\iota}(\omega)=(\phi_{1+\iota}(\omega))^4$, and $\phi_{81+\iota}(\omega)=(\phi_{1+\iota}(\omega))^5$ for $\iota=1,\ldots,20$.

$\Phi^{(3)}$: This basis set includes all elements in $\Phi^{(2)}$ as well $\phi_{102}(\omega)=L(\omega)$, the exact expression for the loss determined using the analytical formulas to value the derivatives.

- *Risk estimation*: We estimate the true risk measure α, the delta-gamma estimator $\hat{\alpha}_{\mathsf{DG}}$, and the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ using non-nested simulations with 2^{15} Sobol points.

Example $\mathsf{EX}_{10\mathsf{E}}$. Ten underlying assets with a portfolio consisting of exchange options whose payoffs depend on pairs of underlying assets.

- *Portfolio*: Short 10 exchange options on each of five asset pairs j and $j+5$, for $j=1,\ldots,5$. The payoff of the jth exchange option is $\max(0, S_{j,T}-S_{j+5,T})$.

The loss threshold is $c=278.8783$, the 99th percentile of $L(\omega)$.

- *Basis functions*:

$\Phi^{(1)}$: This basis set consists of bivariate quadratic functions of $S_{j,\tau}(\omega)$ and $S_{j+5,\tau}(\omega)$ for $j=1,\ldots,5$. Specifically, $\Phi^{(1)}(\omega)$ contains $\phi_l(\omega)$ for $l=1,\ldots,26$, where $\phi_1(\omega)=1$, $\phi_{1+\iota}(\omega)=S_{\iota,\tau}(\omega)$, $\phi_{6+\iota}(\omega)=S_{\iota+5,\tau}(\omega)$, $\phi_{11+\iota}(\omega)=(S_{\iota,\tau}(\omega))^2$, $\phi_{16+\iota}(\omega)=S_{\iota,\tau}(\omega)\cdot S_{\iota+5,\tau}(\omega)$, and $\phi_{21+\iota}(\omega)=(S_{\iota+5,\tau}(\omega))^2$ for $\iota=1,\ldots,5$.

$\Phi^{(2)}$: This basis set includes bivariate fifth order polynomials of $S_{j,\tau}(\omega)$ and $S_{j+5,\tau}(\omega)$ for $j=1,\ldots,5$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l=1,\ldots,101$, where $\phi_l(\omega)$ for $l=1,\ldots,26$ are defined as in $\Phi^{(1)}(\omega)$, and for the third-, fourth- and fifth-order terms, we have the corresponding powers of $S_{j,\tau}(\omega)$ for $j=1,\ldots,10$ as well as the cross terms of $S_{j,\tau}(\omega)$ and $S_{j+5,\tau}(\omega)$ for $j=1,\ldots,5$.

$\Phi^{(3)}$: This basis set includes all elements in $\Phi^{(2)}$ as well as $\phi_{102}(\omega)=L(\omega)$, the exact expression for the loss determined using the analytical formulas to value the derivatives.

- *Risk estimation*: We estimate the true risk measure α, the delta-gamma estimator $\hat{\alpha}_{\mathsf{DG}}$, and the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ using non-nested simulations with 2^{15} Sobol points.

Example EX_{100}. There are 100 underlying assets with nonzero correlations.

- *Model*: The assets are divided into 10 groups of 10 assets each. The correlation is 20% between assets in the same group and is 0% otherwise. The assets in the first three groups have a volatility of 50%, those in the next four groups have a volatility of 30%, and those in the last three groups have a volatility of 10%.
- *Portfolio*: Short 10 at-the-money calls and 10 at-the-money puts on each of the 100 underlying assets.

The loss threshold is $c=876.8636$, the 99th percentile of $L(\omega)$.

- *Basis functions*:

$\Phi^{(1)}$: This basis set includes quadratic functions of $S_{j,\tau}(\omega)$ for $j=1,\ldots,100$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l=1,\ldots,201$, where $\phi_1(\omega)=1$, $\phi_{1+\iota}(\omega)=S_{\iota,\tau}(\omega)$ and $\phi_{101+\iota}(\omega)=(\phi_{1+\iota}(\omega))^2$ for $\iota=1,\ldots,100$.

$\Phi^{(2)}$: This basis set includes fifth order polynomials of $S_{j,\tau}(\omega)$ for $j=1,\ldots,100$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l=1,\ldots,501$, where $\phi_1(\omega)=1$, $\phi_{1+\iota}(\omega)=S_{\iota,\tau}(\omega)$, $\phi_{101+\iota}(\omega)=(\phi_{1+\iota}(\omega))^2$, $\phi_{201+\iota}(\omega)=(\phi_{1+\iota}(\omega))^3$, $\phi_{301+\iota}(\omega)=(\phi_{1+\iota}(\omega))^4$, and $\phi_{401+\iota}(\omega)=(\phi_{1+\iota}(\omega))^5$ for $\iota=1,\ldots,100$.

$\Phi^{(3)}$: This basis set includes all elements in $\Phi^{(2)}$ as well as $\phi_{502}(\omega)=L(\omega)$, the exact expression for the loss determined using the analytical formulas to value the derivatives.

- *Risk estimation*: We estimate the true risk measure α, the delta-gamma estimator $\hat{\alpha}_{\mathsf{DG}}$, and the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ using non-nested simulations with 2^{15} Sobol points.

Example $\mathsf{EX}_{5\mathsf{H}}$. Five underlying assets with a portfolio consisting of one Himalayan option (see, e.g., Gharavi 2010) whose payoffs depend on these five underlying assets.

- *Portfolio*: Long one Himalayan option based on a basket of five underlying assets. The maturity is $T=0.5$ year. The risk horizon is $\tau=0.04$ year. The payoff times are $t_1=0.1$, $t_2=0.2$, $t_3=0.3$, $t_4=0.4$, and $t_5=0.5$ year. At payoff time t_j, for $j=1,\ldots,5$, the underlying asset with the largest return in the basket, denoted by $S_{m_j,t_j}(\omega)$, is noted and removed from the basket. At the maturity, the basket is emptied and the payoff is

$$\sum_{j=1}^{5}\max\left(\frac{S_{m_j,t_j}(\omega)-S_{m_j,0}(\omega)}{S_{m_j,0}(\omega)},0\right),$$

where $S_{m_j,0}(\omega)=100$, for $j=1,\ldots,5$, as assumed at the beginning of §5.3.

The loss threshold is set to be $c=1.622\times10^{-1}$, which is estimated as the 99th percentile of $L(\omega)$.

- *Basis functions*:[11]

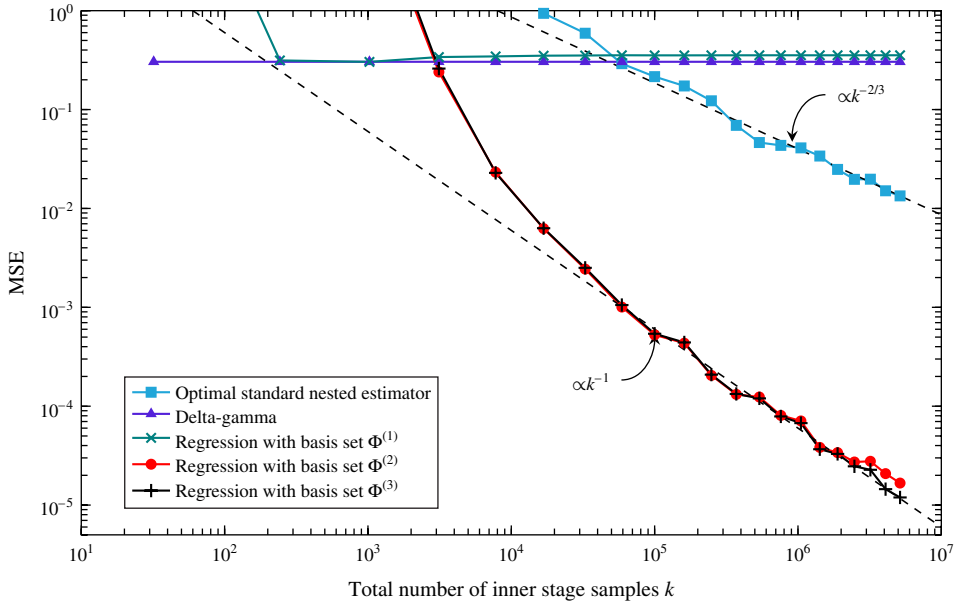
$\Phi^{(1)}$: This basis set includes quadratic functions of $S_{m_j,\tau}(\omega)$ for $j=1,\ldots,5$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l=1,\ldots,11$, where $\phi_1(\omega)=1$, $\phi_{1+\iota}(\omega)=S_{m_\iota,\tau}(\omega)$ and $\phi_{6+\iota}(\omega)=(\phi_{1+\iota}(\omega))^2$ for $\iota=1,\ldots,5$.

$\Phi^{(2)}$: This basis set includes fifth order polynomials of $S_{m_j,\tau}(\omega)$ for $j=1,\ldots,5$. Specifically, $\Phi^{(1)}(\omega)$ consists of $\phi_l(\omega)$ for $l=1,\ldots,26$, where $\phi_1(\omega)=1$, $\phi_{1+\iota}(\omega)=S_{m_\iota,\tau}(\omega)$, $\phi_{6+\iota}(\omega)=(\phi_{1+\iota}(\omega))^2$, $\phi_{11+\iota}(\omega)=(\phi_{1+\iota}(\omega))^3$, $\phi_{16+\iota}(\omega)=(\phi_{1+\iota}(\omega))^4$, and $\phi_{21+\iota}(\omega)=(\phi_{1+\iota}(\omega))^5$ for $\iota=1,\ldots,5$.

- *Risk estimation*: Since analytic expressions for the value of Himalayan options are not known, we estimate the true risk measure α using an exhaustive nested simulation with n outer stage scenarios generated using Sobol points and mn overall inner stage samples where $n=2^{24}$ and $m=2^{12}$. We estimate the delta-gamma estimator $\hat{\alpha}_{\mathsf{DG}}$ and the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$ using non-nested simulations with 2^{15} Sobol points.

The first three examples are from the paper of Glasserman et al. (2000). The last example, $\mathsf{EX}_{5\mathsf{H}}$, is meant to illustrate the performance of the methods in a more realistic setting where

Figure 4. (Color online) Illustration of the mean squared error in Example EX_{10}.



Notes. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples. For the standard nested estimator, the optimized parameter value $\beta = 0.357$ was used.

analytic pricing formulas are unavailable and simulation-based methods are the only option.

As in the single asset case, in each of these examples, the first basis set $\Phi^{(1)}$ was chosen because the regression method with this set is directly comparable to the delta-gamma approximation. The second basis set $\Phi^{(2)}$ was chosen to illustrate the potential benefit of introducing basis functions chosen based on problem specific knowledge. Better choices of basis functions, e.g., using the analytical expressions for vanilla derivative prices and powers of these expressions, are clearly possible and may lead to smaller asymptotic bias levels. In Examples EX_{10}, $\mathsf{EX}_{10\mathsf{E}}$, and EX_{100}, the third basis set $\Phi^{(3)}$ represents the ideal case where the basis functions span the true portfolio loss function and there is no model error.

For each method, we simulate 100 independent trials to estimate the MSE of the estimator. Results are given in Figures 4–7. In each example the MSE of the standard nested estimator decays at $k^{-2/3}$, as expected. The delta-gamma estimator is fast to compute but not very accurate since the local approximation does not capture important features of the portfolio loss functions. The regression method with basis set $\Phi^{(3)}$ works the best in each applicable example, with the MSE decaying at rate k^{-1}. With more realistic basis sets $\Phi^{(1)}$ and $\Phi^{(2)}$, the MSEs decay at the rate k^{-1} until hitting the asymptotic bias levels, which are often beyond the scale of the figures. MSE results for the $k = 5{,}000{,}000$ case are summarized in Table 1.

6. The Weighted Regression Method

This section follows the ideas of Section 3 and develops a risk estimation procedure via a *weighted regression* algorithm. Weighted regression uses spatial information across scenarios to build a global approximation to the loss function but also emphasizes some scenarios that are more important to the calculation of the resulting risk measure via a weight function.

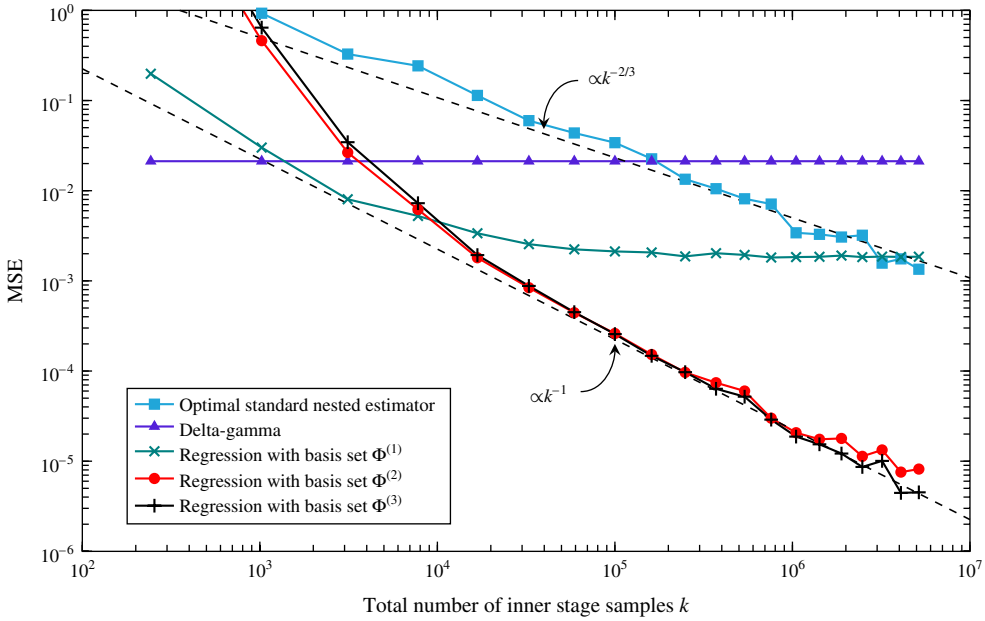
In what follows, we will describe and analyze a weighted regression algorithm. We will establish that, as in the unweighted case, the weighted regression estimation has an asymptotically nondiminishing bias term. The bias term is determined in part by the choice of weight function. We will describe an idealized "optimal" choice of weight function and then numerically demonstrate a practically implementable variation in order to illustrate the benefits of weighted regression.

6.1. The Weighted Regression Algorithm

When computing the regression estimator $\hat{\alpha}_{\mathsf{REG}(m,n)}$, all outer stage scenarios are weighted uniformly in the sample mean squared error objective of (13). This is reasonable if we seek a globally good approximation over the entire scenario space Ω. However, when dealing with the estimation of a specific risk measure, the portfolio loss estimates in some scenarios may 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.

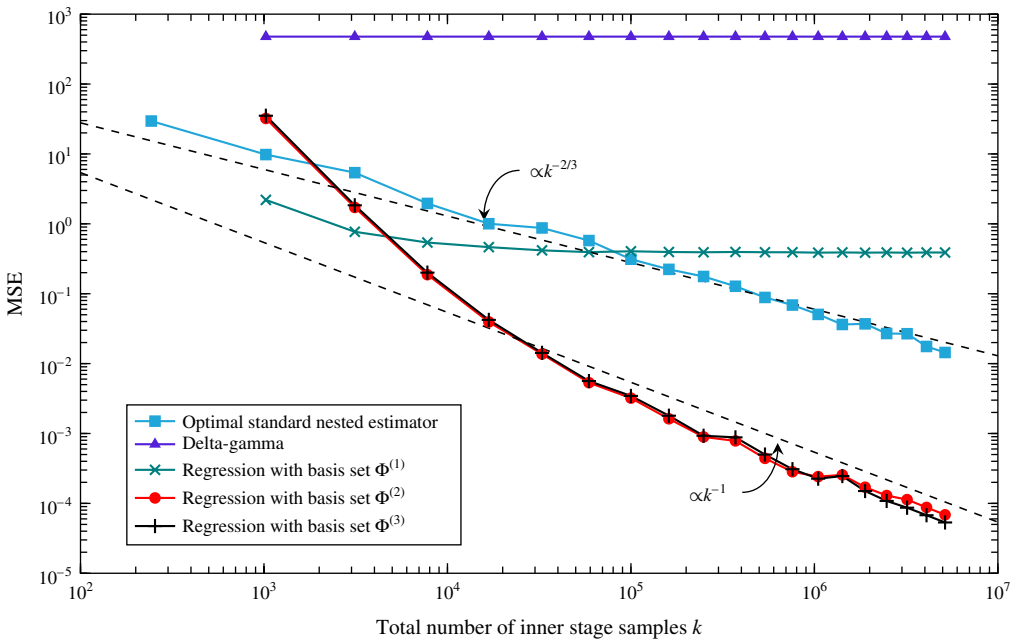
Suppose $h\colon \Omega \to \mathbb{R}_+$ is a non-negative weight function on the set of scenarios. In order to define the weighted regression

Figure 5. (Color online) Illustration of the mean squared error in Example $\mathsf{EX}_{10\mathsf{E}}$.



Notes. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples. For the standard nested estimator, the optimized parameter value $\beta = 1.674$ was used.

Figure 6. (Color online) Illustration of the mean squared error in Example EX_{100}.



Notes. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples. For the standard nested estimator, the optimized parameter value $\beta = 1.674$ was used.

algorithm, we will make the following two assumptions, which are analogs of Assumptions A1 and A2, respectively.

ASSUMPTION A6. *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*

$$\mathsf{E}[\hat{L}(\omega, \zeta) \mid \omega] = L(\omega)$$

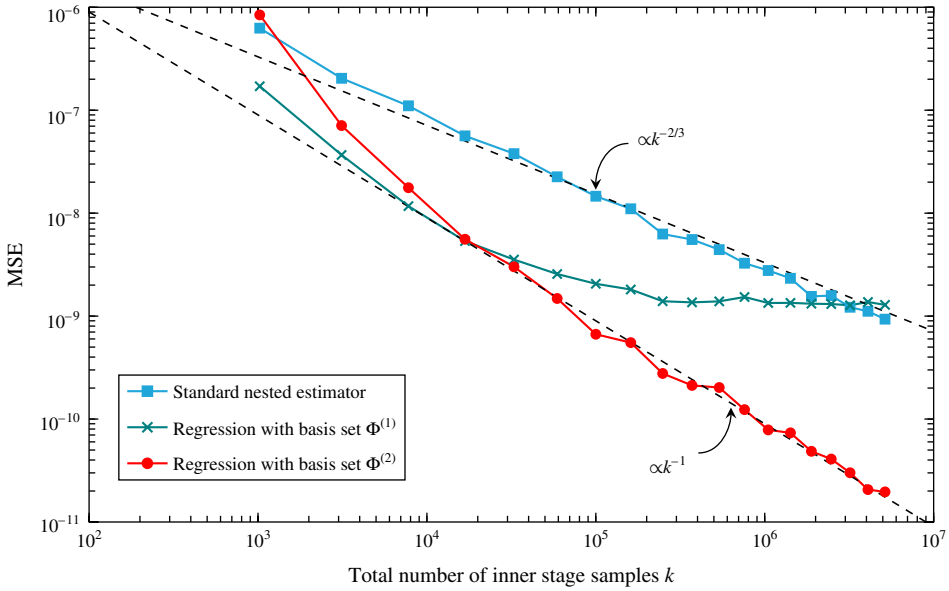
and

$$\mathrm{Var}(\hat{L}(\omega, \zeta) \mid \omega) = \frac{v(\omega)}{m} < \infty.$$

Also, $h(\omega)$ *and* $v(\omega)$ *satisfy*

$$\mathsf{E}[h(\omega)v(\omega)] < \infty.$$

Figure 7. (Color online) Illustration of the mean squared error in Example EX_{5H}.



Notes. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples. For the standard nested estimator, the unoptimized parameter value $\beta = 1$ was used.

Assumption A7. *The second moments of* $\sqrt{h(\cdot)}\phi_1(\cdot), \ldots, \sqrt{h(\cdot)}\phi_d(\cdot)$ *are finite; i.e.,*

$$\mathsf{E}[h(\omega)\phi_l(\omega)^2] < \infty$$

for $l = 1, \ldots, d$. *Further,* $\sqrt{h(\cdot)}\phi_1(\cdot), \ldots, \sqrt{h(\cdot)}\phi_d(\cdot)$ *are linearly independent; i.e., when* $n \geqslant d$,

$$\mathsf{P}\left(\operatorname{rank}\begin{pmatrix}\sqrt{h(\omega^{(1)})}\Phi(\omega^{(1)})\\ \vdots \\ \sqrt{h(\omega^{(n)})}\Phi(\omega^{(n)})\end{pmatrix} = d\right) = 1.$$

Moreover, we assume that the functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ *are orthonormal; i.e., we assume that* $\mathsf{E}[\Phi(\omega)^\top\Phi(\omega)]$ *is the identity matrix.*

Using $h(\cdot)$ as the weight function, we define the optimal weighted regression coefficients according to

$$r^*(h) \in \operatorname*{arg\,min}_{r\in\mathbb{R}^d} \mathsf{E}[h(\omega)(L(\omega) - \Phi(\omega)r)^2]. \tag{23}$$

For any $h(\cdot)$ that satisfies Assumptions A6 and A7, the optimal solution $r^*(h)$ to (23) exists and is unique.

As before, the solution to (23) is not directly computable since the loss function $L(\cdot)$ cannot be exactly evaluated. Instead, we will seek to approximate a solution to (23) by nested simulation. Given n outer stage scenarios $\vec{\omega} \triangleq (\omega^{(1)}, \ldots, \omega^{(n)})^\top$, we will assign weight $h(\omega^{(i)}))$ to the ith scenario and define the vector of weights $\vec{h} \triangleq (h(\omega^{(1)}), \ldots, h(\omega^{(n)}))^\top$. As in (13), we use the Monte Carlo portfolio loss estimates $\hat{L}(\omega^{(1)}, \zeta^{(1)}), \ldots, \hat{L}(\omega^{(n)}, \zeta^{(n)})$ based on m inner stage samples in each scenario and estimate the regression coefficients according to the weighted least squares problem

$$\hat{r}(\vec{h}) \in \operatorname*{arg\,min}_{r\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^{n} h(\omega^{(i)})(\hat{L}(\omega^{(i)}, \zeta^{(i)}) - \Phi(\omega^{(i)})r)^2. \tag{24}$$

Given the estimated regression coefficients $\hat{r}(\vec{h})$, weighted regression estimator of the risk measure α is given by

$$\hat{\alpha}_{\mathsf{REG}(m,n,h)} \triangleq \mathsf{E}[f(\Phi(\omega)\hat{r}(\vec{h})) \mid \vec{\omega}, \vec{\zeta}]. \tag{25}$$

For any $h(\cdot)$ that satisfies Assumptions A6 and A7, the optimal solution $\hat{r}(\vec{h})$ to (24) exists and is unique almost surely when $n \geqslant d$, so our estimator $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ is well defined. By the same reasoning as in Remark 2, we assume the expectation in (25) and hence the estimate $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ can be exactly computed.

Remark 4. 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 positive factor, $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$ will not change.

Remark 5. Note that an alternative way to interpret the objective in (23) is as the mean squared error under a new probability distribution over outer stage scenarios ω, namely, the probability distribution which Radon-Nikodym derivative given by a normalized version of the weight function $h(\cdot)$. In this way, instead of estimating the regression coefficients through the sampled weighted least squares problem (24), one could solve an ordinary least squares problem where the outer stage scenarios are drawn according from the new distribution. Analogous to importance sampling, if efficient sampling from this new distribution is possible, it may lead to more efficient algorithms.

Table 1. MSE results for the five examples.

Example	Estimator	MSE	MSE (normalized)
Single asset example	Optimal standard nested estimator	$1.0 \cdot 10^{-5}$	333.9
	Delta-gamma	$1.5 \cdot 10^{-1}$	4,807,400.0
	Regression with basis set $\Phi^{(1)}$	$3.9 \cdot 10^{-4}$	12,414.1
	Regression with basis set $\Phi^{(2)}$	$7.7 \cdot 10^{-8}$	2.5
	Regression with basis set $\Phi^{(3)}$	$3.1 \cdot 10^{-8}$	1.0
Example EX_{10}	Optimal standard nested estimator	$1.4 \cdot 10^{-2}$	1,107.9
	Delta-gamma	$3.0 \cdot 10^{-1}$	24,704.7
	Regression with basis set $\Phi^{(1)}$	$3.5 \cdot 10^{-1}$	28,761.4
	Regression with basis set $\Phi^{(2)}$	$1.7 \cdot 10^{-5}$	1.4
	Regression with basis set $\Phi^{(3)}$	$1.2 \cdot 10^{-5}$	1.0
Example $\mathsf{EX}_{10\mathsf{E}}$	Optimal standard nested estimator	$1.4 \cdot 10^{-3}$	311.7
	Delta-gamma	$2.1 \cdot 10^{-2}$	4,713.3
	Regression with basis set $\Phi^{(1)}$	$1.9 \cdot 10^{-3}$	411.2
	Regression with basis set $\Phi^{(2)}$	$8.1 \cdot 10^{-6}$	1.8
	Regression with basis set $\Phi^{(3)}$	$4.5 \cdot 10^{-6}$	1.0
Example EX_{100}	Optimal standard nested estimator	$1.5 \cdot 10^{-2}$	268.5
	Delta-gamma	$4.8 \cdot 10^{2}$	8,605,300.0
	Regression with basis set $\Phi^{(1)}$	$3.9 \cdot 10^{-1}$	7,016.0
	Regression with basis set $\Phi^{(2)}$	$7.1 \cdot 10^{-5}$	1.3
	Regression with basis set $\Phi^{(3)}$	$5.5 \cdot 10^{-5}$	1.0
Example $\mathsf{EX}_{5\mathsf{H}}$	Standard nested estimator	$9.6 \cdot 10^{-10}$	48.6
	Regression with basis set $\Phi^{(1)}$	$1.3 \cdot 10^{-9}$	65.7
	Regression with basis set $\Phi^{(2)}$	$2.0 \cdot 10^{-11}$	1.0

Notes. The results are computed over independent trials (1,000 in the single asset example and 100 in the other three), each with a total simulation budget of $k = 5{,}000{,}000$. The last column contains MSE results normalized relative to the regression estimator with the most sophisticated basis set ($\Phi^{(2)}$ or $\Phi^{(3)}$).

Remark 6. 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.

6.2. Analysis

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

$$\mathsf{E}[(\hat{\alpha}_{\mathsf{REG}(m,n,h)} - \alpha)^2] = \mathsf{E}[(\mathsf{E}[f(\Phi(\omega)\hat{r}(\vec{h})) \mid \vec{\omega}, \vec{\zeta}] - \mathsf{E}[f(L(\omega))])^2]. \quad (26)$$

We will analyze the performance of the weighted regression estimator following the analysis of the unweighted Lipschitz continuity case of §4.2. We start with the following technical assumption, which is an analog of Assumption A4:

Assumption A8. *The moment generating functions of $h(\omega)\|\Phi(\omega)\|_2^2$, $h(\omega)(M(\omega))^2$, and $h(\omega)(\varepsilon(\omega,\zeta))^2$ are finite valued in a neighborhood of zero.*

Define

$$G(r,\omega,\zeta,h) \triangleq h(\omega)(\hat{L}(\omega,\zeta) - \Phi(\omega)r)^2, \quad (27)$$

and

$$g(r,h) \triangleq \mathsf{E}[G(r,\omega,\zeta,h)]. \quad (28)$$

Given r, the function $G(r,\omega,\zeta,h)$ is the weighted 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 weighted squared error. For any $\rho > 0$, recall the neighborhood $\mathscr{R}_\rho$ defined by (22). We make the following additional assumption, which is an analog of Assumption A5:

Assumption A9. *For any $\rho > 0$, there exists a constant $\lambda > 0$ such that for any $r', r'' \in \mathscr{R}_\rho$, the moment generating function $\Psi_{r',r''}(t)$ of the random variable*

$$(G(r',\omega,\zeta,h) - g(r',h)) - (G(r'',\omega,\zeta,h) - g(r'',h))$$

satisfies

$$\Psi_{r',r''}(t) \leqslant \exp(\rho\lambda^2 t^2),$$

for any $t \in \mathbb{R}$.

We have the following result, whose proof is given in Section C:

Theorem 4. *Given a weight function $h(\cdot)$, if Assumptions F2, A6–A9 hold, then*

$$\lim_{n\to\infty} \mathsf{E}[(\hat{\alpha}_{\mathsf{REG}(m,n,h)} - \alpha)^2] = (\mathsf{E}[f(\Phi(\omega)r^*(h))] - \mathsf{E}[f(L(\omega))])^2. \quad (29)$$

As the number of outer stage scenarios $n \to \infty$, Theorem 4 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 (29) nonzero.

We will seek to minimize the asymptotic limit of (29), i.e., the squared bias term, by picking a good weight function $h(\cdot)$. We make the following assumption:

Assumption F3. *The function $f(\cdot)$ has first derivative for almost every value of $L(\omega)$.*

Given Assumption F3, applying Jensen's inequality and heuristically applying a Taylor approximation,

$$\begin{aligned}&(\mathsf{E}[f(\Phi(\omega)r^*(h))] - \mathsf{E}[f(L(\omega))])^2\\ &\quad\leqslant \mathsf{E}[(f(\Phi(\omega)r^*(h)) - f(L(\omega)))^2]\\ &\quad\approx \mathsf{E}[(f'(L(\omega)))^2(\Phi(\omega)r^*(h) - L(\omega))^2]. \qquad (30)\end{aligned}$$

Instead of finding the choice of weighting function that minimizes the limit of (29), for reasons of tractability, we consider optimizing the upper bound given by (30). Here, we seek to find regression approximations that are close to the true portfolio loss, emphasizing scenarios ω where $f'(L(\omega))$ is large. With (30) as our objective, we have the following optimization problem to determine a choice of weight function:

$$\begin{aligned}\underset{h}{\text{minimize}}\quad & \mathsf{E}[(f'(L(\omega)))^2(\Phi(\omega)r^*(h) - L(\omega))^2]\\ \text{subject to}\quad & \mathsf{E}[h(\omega)] > 0, \qquad (31)\\ & h(\omega) \geqslant 0, \quad \forall\, \omega \in \Omega.\end{aligned}$$

In the following theorem, we find a globally optimal solution to the problem (31).

Theorem 5. *Suppose that Assumption F3 holds. Define the weight function $h_{opt}(\cdot)$ by*

$$h_{opt}(\omega) = (f'(L(\omega)))^2 \qquad (32)$$

and assume that $\mathsf{E}[(f'(L(\omega)))^2] > 0$. Then $h_{opt}(\cdot)$ is a globally optimal solution to the optimization problem (31).

Proof. With r^* defined by (23), for any h in the domain of (31),

$$\begin{aligned}&\mathsf{E}[(f'(L(\omega)))^2(\Phi(\omega)r^*(h_{opt}) - L(\omega))^2]\\ &\quad= \mathsf{E}[h_{opt}(\omega)(\Phi(\omega)r^*(h_{opt}) - L(\omega))^2] \qquad (33)\\ &\quad\leqslant \mathsf{E}[h_{opt}(\omega)(\Phi(\omega)r^*(h) - L(\omega))^2] \qquad (34)\\ &\quad= \mathsf{E}[(f'(L(\omega)))^2(\Phi(\omega)r^*(h) - L(\omega))^2], \qquad (35)\end{aligned}$$

where (33) and (35) follow from (32) and (34) follows from (23). □

Theorem 5 suggests a specific good choice of weight function, in particular, that the weight $h(\omega)$ of each scenario ω should be proportional to $(f'(L(\omega)))^2$.

6.3. Numerical Results

In this section, we use numerical examples to demonstrate the benefits of the regression method. The standard nested simulation method and the unweighted regression method are also implemented as competing methods. We follow the same experimental setting as in §5.1.

In particular, given a loss threshold $c \in \mathbb{R}$, consider the expected excess loss risk measure

$$\alpha = \mathsf{E}[f(L(\omega))] = \mathsf{E}[(L(\omega) - c)^+]. \qquad (36)$$

From (32), the optimal weight function is given by

$$h_{opt}(\omega) = (f'(L(\omega)))^2 = \mathbb{I}_{\{L(\omega)\geqslant c\}}. \qquad (37)$$

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 practically implementable *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 $h_{opt}(\cdot)$ to be used in a *weighted* regression. The details of this two-pass procedure are described in Section A of the online supplement.

We consider two examples. First, we consider a one-dimensional problem as follows:

- *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.

Our second example is a multidimensional, delta-hedged example as follows:

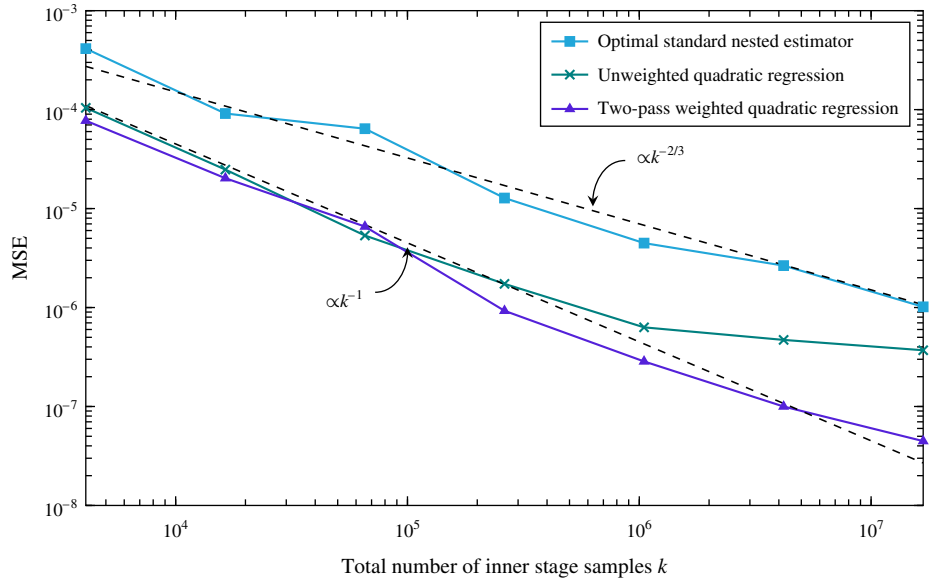
- *Model.* There are 10 i.i.d. assets each with initial price $S_0 = 100$. The drift under the real-world distribution is $\mu = 8\%$. The annual volatility is $\sigma = 30\%$. The continuously compounded riskless rate of interest is $r_f = 5\%$.
- *Portfolio.* The portfolio consists of three types of 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 (1) short 10 down-and-out call options on asset i with barrier $H = 95$, for $i = 1, \ldots, 10$; (2) short 5 cash-or-nothing put options on asset i, for $i = 1, \ldots, 10$; and (3) 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.

We test the following estimators[12] for both examples:

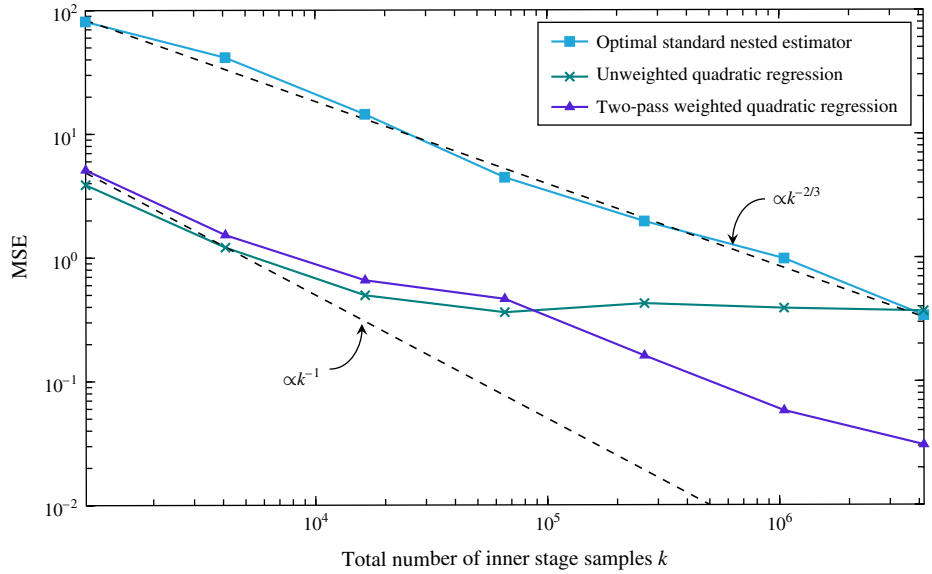
- *Optimal standard nested estimator.* This is the estimator $\hat{\alpha}_{\mathsf{SN}(m,n)}$ with the parameters m and n chosen optimally according to (7). As described in Remark 1, we test a large number of choices of the parameter pair (m, n) and choose the one with the minimum MSE among them. Note that in practice, this optimal choice is not achievable.

Figure 8. (Color online) Illustration of the mean squared error in the one-dimensional example.



Note. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples.

Figure 9. (Color online) Illustration of the mean squared error in the 10-dimensional example.



Note. The vertical axis shows the mean squared error, and the horizontal axis represents the total number of inner stage samples.

• *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 $\phi_3(\omega) = (S_\tau(\omega))^2$.

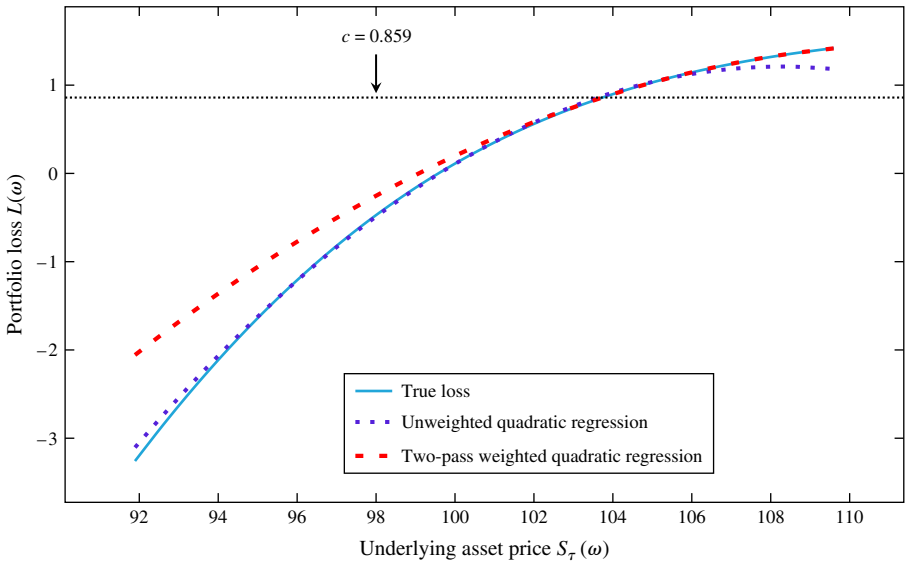
—In the multidimensional problem, the underlying assets have price $S_{j,\tau}(\omega)$ for $j = 1, \ldots, 10$ at time τ. The basis function set Φ includes quadratic functions of each $S_{j,\tau}(\omega)$ and quadratic functions of each $(S_{j,\tau}(\omega) - H)^+$ for $j = 1, \ldots, 10$. Specifically, Φ consists of $\phi_l(\omega)$ for $l = 1, \ldots, 41$, where $\phi_1(\omega) = 1$, $\phi_{1+\iota}(\omega) = S_{\iota,\tau}(\omega)$ and $\phi_{11+\iota}(\omega) = (S_{\iota,\tau}(\omega) - H)^+$ for $\iota = 1, \ldots, 10$ and $\phi_{21+\iota}(\omega) = (\phi_{1+\iota}(\omega))^2$ for $\iota = 1, \ldots, 20$. 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 a weight function that is determined through a first-pass unweighted regression, are described in Section A of the online supplement.

Results for the MSE of the various estimators are given in Figures 8 and 9. 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.

Figure 10. (Color online) Illustration of approximations in the one-dimensional example.



Note. 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.

- *Unweighted quadratic regression.* The MSE of this estimator converges at the rate k^{-1} until reaching an asymptotic bias level that depends on the size of the regression model error $M(\cdot)$ or, equivalently, the quality of the basis functions.
- *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. With large k, the MSE of the two-pass weighted quadratic regression method works about one order of magnitude better than the unweighted quadratic regression method.

Figure 10 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.

7. Conclusion

Risk computation for realistic portfolios of derivatives securities presents an important and challenging computational problem. In addition to the usual estimator error due to random sampling, the nonlinearity of portfolio loss functions introduces bias into the estimation of risk measures. Nested simulation can be used to minimize both estimator bias and variance, but the additional computation time of the "simulation within a simulation" leads to a convergence rate of $k^{-2/3}$, where k represents the computation time.

In this paper we propose a new risk estimation method based on Monte Carlo simulation and regression. Given n outer stage scenarios and mn inner stage samples, we show that the optimal choice of m and n are $m^* = 1$ and $n^* = k$. With these choices, our theoretical results show that the mean squared error diminishes at a rate close to k^{-1} until a nondiminishing bias level is reached. The proposed regression method outperforms standard nested simulation and the delta-gamma method when used with a "good" set of basis functions as long as k is not too large. More importantly, before hitting the bias level, the proposed method recovers the standard k^{-1} convergence rate of non-nested unbiased simulation estimators. Further, weighted variations of our method offer improved asymptotic bias. Numerical results show that the regression method significantly outperforms other methods and illustrated consistency with the theoretical results.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2015.1419.

Acknowledgments

This work was supported by NSF grants DMS–0914539 and CMMI–1235023.

Endnotes

1. Our formulation includes path dependent portfolios and portfolios with cash flows in the interval $[0, \tau]$ as special cases. In such instances, the state ω can be augmented with additional information so that the portfolio value at time τ is uniquely determined as a function of ω.
2. Our assumption that $\{\zeta^{(i)}\}$ are identically distributed over outer stage scenarios is without loss of generality. In particular, this does *not* imply that the estimated losses are identically distributed across scenarios. This is because any heteroscedasticity in $\zeta^{(i)}$ can be absorbed into the functional form of $L(\omega, \zeta)$. For example,

in §5 we take each $\zeta^{(i)}$ to be a collection of independent Brownian motion processes. These processes drive the Itō processes that determine asset prices. The overall loss estimate in each scenario is a function of the scenario $\omega^{(i)}$ in combination with $\zeta^{(i)}$, and the loss estimates have different distributions across scenarios.

3. This line of analysis can be easily extended to also account for the computational effort required for the n outer stage scenarios, but we will not do so here.

4. One might consider using both the initial n outer state scenarios as well as the additional n' scenarios to estimate (14) via $(1/(n+n'))(\sum_{i=1}^{n}\hat{L}(\omega^{(i)},\zeta^{(i)})+\sum_{i=1}^{n'}f(\Phi(\omega'_{(i)}\hat{r}))$. However, for reasons discussed in §2.1, we assume that it is much easier computationally to sample outer stage scenarios than inner stage scenarios. In this case, $n' \gg n$, and hence we will not consider such a hybrid estimator.

5. In addition to the nested sampling, additional computation is required for the regression step, i.e., the computation of (13). From standard results on least squares optimization (e.g., Golub and Van Loan 2012), the computational effort required to compute the regression coefficients $\hat{r}$ is a linear function of the number of outer stage scenarios n; i.e., the extra computational burden is $O(n)$. This burden is asymptotically proportional to (if m is $O(1)$) or dominated by (if $m \to \infty$) the total required computation requirement $O(k)$ for the generation of nested samples. Therefore, the additional computational requirement for computing regression weights (13) is ignored in our analysis.

6. We thank an anonymous referee for this suggestion.

7. Let $\xi_1, \xi_2, \ldots$ be a sequence of random vectors. If there exists a vector ξ^* such that for every $b > 0$, $\mathsf{P}(\|\xi_n - \xi^*\|_2 < b) \to 1$ as $n \to \infty$, then ξ_n converges to ξ^* in probability. We write this as $\xi_n \xrightarrow{\mathsf{P}} \xi^*$ or $\|\xi_n - \xi^*\|_2 = O_{\mathsf{P}}(1)$. If we denote by F_{ξ_n} and F_{ξ^*} the cumulative distribution functions of random variables ξ_n and ξ^*, and if $\lim_{n\to\infty} F_{\xi_n} = F_{\xi^*}$ at all continuity points of F_{ξ^*}, then ξ_n converges to ξ^* in distribution. We write this as $\xi_n \xrightarrow{\mathsf{d}} \xi^*$.

8. In the delta-gamma approximation, we will approximate the loss as a quadratic function of $S_\tau(\omega)$ instead of a quadratic function of ω.

9. All barrier options are partial time barrier options (see, e.g., Haug 2006, §4.17.4) that can be knocked in or out only between times τ and T.

10. Along each sample path, the cash flow of a barrier option depends only on minimum underlying asset price and the final asset price. We simulate these two quantities instead of sampling the entire sample path. Details can be found in Hui (1997) and Metwally and Atiya (2002).

11. In previous examples, $\Phi^{(3)}$ is the basis set including the exact expression for the loss determined with the analytical formulas, whereas in Example $\mathsf{EX}_{5\mathsf{H}}$, there is no analytical formula for Himalayan options, and thus only two sets of basis functions $\Phi^{(1)}$ and $\Phi^{(2)}$ are presented.

12. Following Remark 6, we use $m = 1$ with the estimators of unweighted quadratic regression and two-pass weighted quadratic regression.

References

Britten-Jones M, Schaefer SM (1999) Non-linear value-at-risk. *Eur. Finance Rev.* 2(2):161–187.

Broadie M, Du Y, Moallemi CC (2011) Efficient risk estimation via nested sequential simulation. *Management Sci.* 57(6):1172–1194.

Duffie D, Pan J (2001) Analytical value-at-risk with jumps and credit risk. *Finance Stochastics* 5(2):155–180.

Gharavi RK (2010) Himalayan option. *Encyclopedia of Quantitative Finance* (John Wiley & Sons, New York).

Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (Springer, New York).

Glasserman P, Heidelberger P, Shahabuddin P (2000) Variance reduction techniques for estimating value-at-risk. *Management Sci.* 46(10): 1349–1364.

Golub GH, Van Loan CF (2012) *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore).

Gordy MB, Juneja S (2008) Nested simulation in portfolio risk measurement. FEDS 2008-21, Federal Reserve Board, Washington, DC.

Gordy MB, Juneja S (2010) Nested simulation in portfolio risk management. *Management Sci.* 56(10):1833–1848.

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

Hong LJ, Juneja S (2009) Estimating the mean of a non-linear function of conditional expectation. *Proceedings of the 2009 Winter Simulation Conference* (IEEE Press, Piscataway, NJ), 1223–1236.

Hui CH (1997) Time-dependent barrier option values. *J. Futures Markets* 17(6):667–688.

Lan H, Nelson BL, Staum J (2010) A confidence interval procedure for expected shortfall risk measurement via two-level simulation. *Oper. Res.* 58(5):1481–1490.

Lee S-H (1998) *Monte Carlo Computation of Conditional Expectation Quantiles*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Lee S-H, Glynn PW (2003) Computing the distribution function of a conditional expectation via Monte Carlo: Discrete conditioning spaces. *ACM Trans. Modeling Comput. Simulation* 13(3):238–258.

Liu M, Staum J (2010) Stochastic kriging for efficient nested simulation of expected shortfall. *J. Risk* 12(3):3–27.

Longstaff FA, Schwartz ES (2001) Valuing American options by simulation: A simple least-squares approach. *Rev. Financial Stud.* 14(1):113–147.

Metwally SAK, Atiya AF (2002) Using Brownian bridge for fast simulation of jump-diffusion processes and barrier options. *J. Derivatives* 10(1): 43–54.

Rouvinez C (1997) Going Greek with VaR. *Risk* 10(2):57–63.

Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (SIAM, Philadelphia).

Sobol IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Physics* 7(4): 86–112.

Sun Y, Apley DW, Staum J (2011) Efficient nested simulation for estimating the variance of a conditional expectation. *Oper. Res.* 59(4):998–1007.

Tsitsiklis JN, Van Roy B (2001) Regression methods for pricing complex American-style options. *IEEE Trans. Neural Networks* 12(4):694–703.

White H (2001) *Asymptotic Theory for Econometricians* (Academic Press, Orlando, FL).

Mark Broadie is the Carson Family Professor of Business at Columbia Business School. His research addresses issues in quantitative finance and sports analytics and, more generally, methods for decision making under uncertainty.

Yiping Du is a Ph.D. graduate of the Department of Industrial Engineering and Operations Research at Columbia University. His research focuses on efficient simulation methods for estimating risk measures.

Ciamac C. Moallemi is an associate professor of Business in the Decision, Risk, and Operations Division of the Graduate School of Business at Columbia University, where he has been since 2007. He is a past recipient of the British Marshall Scholarship (1996), the Benchmark Stanford Graduate Fellowship (2003), first place in the INFORMS Junior Faculty Paper Competition (2011), and the Best Simulation Publication Award of the INFORMS Simulation Society (2014). His research interests are in the area of the optimization and control of large-scale stochastic systems and decision-making under uncertainty, with an emphasis on applications in financial engineering.