Monte Carlo Methods for Value-at-Risk and Conditional Value-at-Risk: A Review

L. JEFF HONG, City University of Hong Kong
ZHAOLIN HU, Tongji University
GUANGWU LIU, City University of Hong Kong

Value-at-risk (VaR) and conditional value-at-risk (CVaR) are two widely used risk measures of large losses and are employed in the financial industry for risk management purposes. In practice, loss distributions typically do not have closed-form expressions, but they can often be simulated (i.e., random observations of the loss distribution may be obtained by running a computer program). Therefore, Monte Carlo methods that design simulation experiments and utilize simulated observations are often employed in estimation, sensitivity analysis, and optimization of VaRs and CVaRs. In this article, we review some of the recent developments in these methods, provide a unified framework to understand them, and discuss their applications in financial risk management.

Categories and Subject Descriptors: I.6.3 [**Simulation and Modeling**]: Applications

General Terms: Theory, Algorithms, Management

Additional Key Words and Phrases: Financial risk management, value-at-risk, conditional value-at-risk

ACM Reference Format:
L. Jeff Hong, Zhaolin Hu, and Guangwu Liu. 2014. Monte Carlo methods for value-at-risk and conditional value-at-risk: A review. ACM Trans. Model. Comput. Simul. 24, 4, Article 22 (November 2014), 37 pages.
DOI: http://dx.doi.org/10.1145/2661631

1. INTRODUCTION

Risk is a fundamental attribute of financial activities. When investors make financial decisions, they consider not only potential returns but also potential risks. There are various kinds of risks in the financial industry. For instance, an investment bank may hold a portfolio of stocks for a period of time and the value of the portfolio may evolve at random during the period. Then, the bank faces the market risk that the value of the portfolio may fall below the initial value. Similarly, a commercial bank may hold a portfolio of loans lent to different obligors. Then, the bank faces the credit risk that some of the obligors may default. Because of the importance and ubiquity of financial risks, individual financial institutions often want to identify and understand

L. J. Hong is supported in part by the Research Grants Council of Hong Kong under Grants 613012, 613213, and N_HKUST626/10; Z. Hu is supported in part by the Natural Science Foundation of China under Grant 71201117, the Fundamental Research Funds for the Central Universities (Project No. 2013KJ028), and the Shanghai Pujiang Program; G. Liu is supported in part by CityU 145610 and 145511.
Authors' addresses: L. J. Hong, P7510, Academic 1, Department of Economics and Finance and Department of Management Sciences, College of Business, City University of Hong Kong, Hong Kong, China; email: jeffhong@cityu.edu.hk; Z. Hu, School of Economics and Management, Tongji University, No. 1239 Siping Road, Yangpu District, Shanghai, China; email: russell@tongji.edu.cn; G. Liu, G7516, Academic 1, Department of Management Sciences, City University of Hong Kong, Hong Kong, China; email: guanliu@cityu.edu.hk.

DOI: http://dx.doi.org/10.1145/2661631

the risks in their activities, based on which they can then control or manage the risks. Furthermore, because of the systematic nature of financial institutions, risks of one institution can easily spread to other institutions or even to the entire financial system, resulting in the so-called systemic risk. Such systemic risk may even affect the entire economic and social system. Therefore, consensus has been reached that regulations on financial systems and financial markets are necessary.

To understand risks, people typically look at the loss resulting from an activity. By bringing in the tool of probability measure, such a loss can be modeled as a random variable. This idea is simple but very useful. It provides a language to describe risks. However, a random variable that can take multiple values according to some probability measure (distribution) is still difficult to interpret. The underlying randomness creates difficulties in managing risks. It also creates difficulties for financial regulations. For a long time, risk was difficult to handle, until the notion of risk measures was brought in. A risk measure is a function that maps a random variable to a real number. Using risk measures, people now can focus on a single number instead of taking into consideration the whole loss distribution. Such a number is easy to interpret and understand. The introduction of risk measures paved the way for efficient and systematical study of financial risks. For a thorough discussion on the background of financial risk management, we refer readers to McNeil et al. [2005] and Jorion [2010].

There have been numerous risk measures introduced and employed in the financial industry. Value-at-risk (VaR) and conditional value-at-risk (CVaR, also known as expected shortfall or tail conditional expectation), which we review in this article, are among the most well-known and widely used ones and play dominating roles in practice. For any $\alpha \in (0, 1)$, the α-VaR of a random loss L is the α quantile of L, while the α-CVaR is the average of all β-VaR for $\beta \in (\alpha, 1)$. As we are typically interested in the risk of large losses in practice, α is typically quite close to 1, for example, $\alpha = 0.9, 0.95, 0.99$. As pointed out by Hong and Liu [2009], if we define the large losses to be the losses in the upper $(1 - \alpha)$-tail of the loss distribution, then the α-VaR is the lower bound of the large losses and the α-CVaR is the mean of the large losses. They provide information on potential large losses that an investor may suffer.

Based on Jorion [2006], VaR was promoted by J. P. Morgan and RiskMetrics Group as a tool for financial risk management in mid-1990s. In 1997, the U.S. Securities and Exchange Commission ruled that public corporations must disclose their derivative activities quantitatively, and many financial institutions chose to report the VaR information as a way to comply with the rule. In 1999, Basel II Accord was adopted internationally and it also incorporated the concept of VaR in the measurement of financial market risk and in determining regulatory capital requirements. Since then, VaR has become a standard concept and approach in financial risk measurement.

Even though VaR was widely adopted in financial practice, there is also criticism on its use as a risk measure. Artzner et al. [1999] defined four axioms and called a risk measure that satisfies these axioms a *coherent risk measure*. One of these axioms is the subadditivity axiom, which basically means that "a merger does not create extra risk." They further showed that VaR does not always satisfy the subadditivity axiom and is therefore not a coherent risk measure. Rockafellar and Uryasev [2002], on the other hand, showed that CVaR satisfies all four axioms and is therefore a coherent risk measure (see also the study of Acerbi and Tasche [2002]). Kou et al. [2013], however, argued that the subadditivity axiom is not necessary and suggested replacing it with the comonotonic subadditivity axiom, which only requires subadditivity to hold for random variables moving in the same direction. They showed that both VaR and CVaR satisfy the comonotonic subadditivity axiom. However, they argued that, compared to CVaR, VaR is often more robust to the tail behavior of the loss distribution, which is in general difficult to characterize in practice, and is therefore more suitable for regulatory purposes.

The 2008 financial crisis indicated serious flaws in the financial risk management methodologies. The use of VaR as a risk measure has also been criticized extensively. Recently, the Basel Committee of Banking Supervision of the Bank for International Settlements (2012) admitted in their consultative document that "a number of weaknesses have been identified with using VaR for determining regulatory capital requirements, including its inability to capture 'tail risk,'" and "it is questionable whether it [VaR] meets the objectives of prudential regulation which seeks to ensure that banks have sufficient capital to survive low probability, or 'tail' events." Therefore, the document suggested moving from VaR to expected shortfall (i.e., CVaR) because it "better captures tail risk." We agree that CVaR better captures tail risk than VaR, because CVaR is the average of tail losses while VaR is only a lower bound. Therefore, the use of CVaR may provide incentives for financial institutions to take into account tail risks beyond VaR. However, the use of CVaR cannot solve the problem of risk modeling entirely. In this article, we do not intend to participate in the debates of VaR and CVaR. Instead, we consider the computational issues related to both VaR and CVaR when using them in practice.

As closed-form expressions of portfolio loss distributions are typically not available in practical situations, a computer program is often built to simulate the losses under different scenarios based on a financial model (often an internal model of a financial institution). Then, we may use the simulated losses to conduct analysis on VaR and CVaR. This is what we call a "Monte Carlo method." In the past two decades, along with the wide adoption of VaR/CVaR risk management methodologies in the financial industry, the research on Monte Carlo methods for VaR and CVaR has also attracted a significant amount of attention from the academic community, and in particular, the stochastic simulation community. Given the existence of a large amount of work in this area, in this article we do not attempt to provide a thorough and comprehensive review of the related literature. Instead, we focus on some of the recent developments in these methods, especially on applications in financial risk management, and intend to provide a unified framework to understand these methods.

In particular, we review three general topics in this article: (1) VaR and CVaR estimations, (2) sensitivity analysis of VaR and CVaR, and (3) VaR and CVaR optimizations. To use VaR and CVaR, one needs to be able to estimate them to the required precision. In this part of the article, we first review the crude Monte Carlo estimators of VaR and CVaR and discuss their asymptotic properties. As both VaR and CVaR concern tail behaviors of a loss distribution, crude Monte Carlo estimators are often not efficient and a large number of observations may be necessary to achieve the required precisions. To address this issue, many variance reduction methods have been proposed to obtain estimators with smaller variance. We briefly introduce the general framework of using importance sampling for variance reduction and discuss several important financial applications in greater detail. As many financial portfolios include derivative contracts that may also need to be priced using Monte Carlo methods, to estimate VaR and CVaR, we often encounter nested estimation problems (i.e., there are two levels of expectations that need to be evaluated). This problem has attracted a significant amount of attention recently. We also provide a short description of this problem and introduce some of the recent developments for solving it.

Sensitivities of VaR and CVaR concern the derivatives of these risk measures with respect to a parameter of the loss distribution. They measure how changes in the parameter affect the values of VaR and CVaR and have a number of important applications in the financial industry. In portfolio optimization problems, for instance, sensitivities of VaR (CVaR) can be incorporated into a gradient-based optimization algorithm to solve problems with VaR (CVaR) objectives and/or constraints. In portfolio capital allocation, for instance, individual assets' marginal risk contributions, the sum

of which is exactly the total risk of the portfolio, can be expressed as sensitivities of the risk measures under appropriate conditions (see, e.g., Kalkbrener [2003], Kalkbrener et al. [2004], and Kurth and Tasche [2003]). In model validation and comparisons, for instance, sensitivities of VaR and CVaR can be used to evaluate the robustness of the financial models with respect to the uncertainty in the model parameters. In Section 3 of our article, we provide a unified approach to deriving the closed-form expressions of VaR and CVaR sensitivities, introduce Monte Carlo estimators of these sensitivities, and discuss their properties. As VaR sensitivity requires the estimation of a conditional expectation conditioning on a probability zero event, its estimator often has a rate of convergence that is slower than a typical Monte Carlo estimator of a mean. We review a number of available methods that devise estimators with faster rates of convergence.

Because VaR and CVaR are important risk measures, they are naturally used in various decision models, resulting in optimization problems that either optimize certain VaR/CVaR functions or optimize certain objectives subject to VaR/CVaR constraints. As both VaR and CVaR typically do not have closed-form expressions, these optimization problems are often solved using either deterministic approximation algorithms that yield feasible but suboptimal solutions or Monte Carlo methods that solve a sample-based optimization problem. In Section 4 of our article, we introduce various Monte Carlo methods in solving VaR and CVaR optimization problems and discuss their connections, advantages, and disadvantages. We also introduce some recent work that solves VaR optimization problems by sequentially solving CVaR optimization problems.

As presented earlier, the three topics reviewed in this article are basic concepts of financial risk management. Interestingly, the order in which we present the topics is in line with other literature. For instance, Jorion [2006] discussed in sequence "passive risk measurement," "defensive risk control," and "active risk management" using VaR. Our topic (1) can be considered as some passive risk measurement, topic (3) can be classified as certain active risk management, and topic (2) somehow falls in between. This suggests that the three topics correspond to progressive actions in the practice of risk management.

The rest of the article is organized as follows. Section 2 considers the estimation of VaR and CVaR, introduces asymptotic properties of the estimators and variance reduction issues, and discusses applications and recent advances in portfolio risk measurement. Section 3 considers the estimation of sensitivities of VaR and CVaR and discusses their closed-form expressions, asymptotic properties, and methods for efficiency improvement. Section 4 considers VaR and CVaR optimization problems, discusses properties of these problems, and introduces some recently developed Monte Carlo algorithms for solving these problems. The article concludes in Section 5 with some discussions on the topics that we think are interesting and deserve further studies.

2. ESTIMATIONS OF VAR AND CVAR

As a starting point, we define VaR and CVaR and explore their inherent connections. Let L be the random loss of interest and $F(y) = \Pr\{L \le y\}$ be the cumulative distribution function (CDF) of L. Then, the inverse CDF of L can be defined as $F^{-1}(\gamma) = \inf\{y : F(y) \ge \gamma\}$. Following the definitions of Trindade et al. [2007], for any $\alpha \in (0, 1)$, we define the α-VaR of L as

$$v_\alpha = F^{-1}(\alpha),$$

and define the α-CVaR of L as

$$c_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 v_\beta \, d\beta. \tag{1}$$

Pflug [2000] showed that c_α is also the optimal value of the stochastic program:

$$c_\alpha = \inf_{t\in\Re}\left\{t + \frac{1}{1-\alpha}\mathrm{E}[L-t]^+\right\}, \tag{2}$$

where $[a]^+ = \max\{0, a\}$. Let T be the set of optimal solutions to the stochastic program defined in Equation (2). Then it can be shown that $T = [v_\alpha, u_\alpha]$, where $u_\alpha = \sup\{t : F(t) \le \alpha\}$ (see, e.g., Rockafellar and Uryasev [2002] and Trindade et al. [2007]). In particular, note that $v_\alpha \in T$. Therefore,

$$c_\alpha = v_\alpha + \frac{1}{1-\alpha}\mathrm{E}[L-v_\alpha]^+. \tag{3}$$

When L has a positive density in the neighborhood of v_α, then $v_\alpha = u_\alpha$. Therefore, the stochastic program defined in (2) has a unique solution, and

$$c_\alpha = \mathrm{E}[L|L \ge v_\alpha], \tag{4}$$

while the right-hand side of Equation (4) is also known as expected shortfall or tail conditional expectation. To be meaningful, we assume that c_α is finite for all discussions related to CVaR in this article.

2.1. Crude Monte Carlo Estimation

Suppose that $L_1, L_2, \ldots, L_n$ are n independent and identically distributed (i.i.d.) observations from the loss L. Then, the α-VaR of L can be estimated by

$$\hat{v}_\alpha^n = L_{\lceil n\alpha\rceil:n},$$

where $\lceil a\rceil$ denotes the smallest integer larger than or equal to a, and $L_{i:n}$ is the ith order statistic from the n observations.

Trindade et al. [2007] suggested to use the estimator

$$\hat{c}_\alpha^n = \inf_{t\in\Re}\left\{t + \frac{1}{n(1-\alpha)}\sum_{i=1}^{n}[L_i - t]^+\right\} \tag{5}$$

to estimate the α-CVaR of L. Let

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n}1_{\{L_i\le y\}}$$

be the empirical CDF constructed from $L_1, L_2, \ldots, L_n$, where $1_{\{\cdot\}}$ is the indicator function. Then

$$\hat{c}_\alpha^n = \inf_{t\in\Re}\left\{t + \frac{1}{1-\alpha}\mathrm{E}[\tilde{L}-t]^+\right\},$$

where the CDF of $\tilde{L}$ is F_n. Since $\hat{v}_\alpha^n = F_n^{-1}(\alpha)$, then by Equation (3), we have

$$\hat{c}_\alpha^n = \hat{v}_\alpha^n + \frac{1}{n(1-\alpha)}\sum_{i=1}^{n}\left[L_i - \hat{v}_\alpha^n\right]^+. \tag{6}$$

Therefore, we can apply Equation (6) to directly estimate c_α instead of solving the stochastic program in Equation (5).

Consistency and asymptotic normality of the estimators $\hat{v}_\alpha^n$ and $\hat{c}_\alpha^n$ have been studied extensively in the literature (see, e.g., Serfling [1980] and Trindade et al. [2007]). Regarding the asymptotic properties, a result that is even sharper is the Bahadur representation [Bahadur 1966].

As a unified view, we present the asymptotic properties of $\hat{v}_\alpha^n$ and $\hat{c}_\alpha^n$ using the Bahadur representations. To this end, we first make the following assumption.

ASSUMPTION 1. *There exists an $\epsilon > 0$ such that L has a positive and continuously differentiable density $f(x)$ for any $x \in (v_\alpha - \epsilon, v_\alpha + \epsilon)$.*

Assumption 1 requires that L has a positive and differentiable density in a neighborhood of v_α. It implies that $F(v_\alpha) = \alpha$ and $c_\alpha = \mathrm{E}[L|L \geq v_\alpha]$.

Bahadur representations of $\hat{v}_\alpha^n$ and $\hat{c}_\alpha^n$ are summarized in the following theorem, whose proof can be found in Sun and Hong [2010].

THEOREM 2.1. *For a fixed $\alpha \in (0, 1)$, suppose that Assumption 1 is satisfied. Then*

$$\hat{v}_\alpha^n = v_\alpha + \frac{1}{f(v_\alpha)}\left(\alpha - \frac{1}{n}\sum_{i=1}^{n} 1_{\{L_i \leq v_\alpha\}}\right) + A_n, \quad \textit{and}$$

$$\hat{c}_\alpha^n = c_\alpha + \left(\frac{1}{n}\sum_{i=1}^{n}\left[v_\alpha + \frac{1}{1-\alpha}(L_i - v_\alpha)^+\right] - c_\alpha\right) + B_n,$$

where $A_n = O_{a.s.}(n^{-3/4}(\log n)^{3/4})$, $B_n = O_{a.s.}(n^{-1}\log n)$, and the statement $Y_n = O_{a.s.}(g(n))$ means that $Y_n/g(n)$ is bounded by a constant almost surely.

Consistency and asymptotic normality of $\hat{v}_\alpha^n$ and $\hat{c}_\alpha^n$ follow straightforwardly from Theorem 2.1. Specifically, if Assumption 1 is satisfied, then $\hat{v}_\alpha^n \to v_\alpha$ and $\hat{c}_\alpha^n \to c_\alpha$ with probability 1 (w.p.1) as $n \to \infty$, and

$$\sqrt{n}\left(\hat{v}_\alpha^n - v_\alpha\right) \Rightarrow \frac{\sqrt{\alpha(1-\alpha)}}{f(v_\alpha)} N(0,1), \quad \text{as} \quad n \to \infty, \tag{7}$$

where "$\Rightarrow$" denotes "converge in distribution," and $N(0, 1)$ represents the standard normal random variable. If, in addition, $\mathrm{E}[(L - v_\alpha)^2 1_{\{L \geq v_\alpha\}}] < \infty$, then

$$\sqrt{n}\left(\hat{c}_\alpha^n - c_\alpha\right) \Rightarrow \sigma_\infty \cdot N(0,1), \quad \text{as} \quad n \to \infty, \tag{8}$$

where

$$\sigma_\infty^2 = \lim_{n\to\infty} n\mathrm{Var}\left(\hat{c}_\alpha^n\right) = \frac{1}{(1-\alpha)^2} \cdot \mathrm{Var}([L - v_\alpha]^+).$$

2.2. Variance Reduction

In the simulation literature, there has been a significant amount of work on the topic of variance reduction for VaR estimation. For instance, Hsu and Nelson [1990] and Hesterberg and Nelson [1998] studied the use of control variates. Avramidis and Wilson [1998] employed correlation-induction techniques for variance reduction in quantile estimation. Glynn [1996] considered the use of importance sampling (IS) and discussed its asymptotic properties. The problem of estimating portfolio VaR has been studied in Glasserman et al. [2000] and Glasserman et al. [2002], where IS and stratified sampling are employed.

Among various variance reduction methods proposed in the literature, IS is particularly attractive, given the rare-event features of many practical problems. It has proven to be a very effective variance reduction technique in this context, and much work has been done regarding this issue.

In what follows, we discuss a general IS method for estimating VaR and CVaR, with a focus on the asymptotic properties of the IS estimators. Specifically, suppose that L is simulated under another CDF $G(\cdot)$, where F is absolutely continuous with respect to G in $[v_\alpha - \epsilon, \infty)$, with $\epsilon > 0$ being a fixed constant, that is, $F(dx) = 0$ if $G(dx) = 0$ for any

$x \in [v_\alpha - \epsilon, \infty)$. We refer to G as the IS distribution and let $l(x) = F(dx)/G(dx)$ denote the likelihood ratio function (also called score function) associated with the change of measure. Note that for $x \in [v_\alpha - \epsilon, \infty)$,

$$F(x) = \mathrm{E}_F\left[1_{\{L \le x\}}\right] = \mathrm{E}_G\left[1_{\{L \le x\}} l(L)\right],$$

where E_F and E_G denote taking expectations with respect to F and G, respectively. Then we may estimate $F(x)$ by

$$F_{n,\mathrm{IS}}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{L_i \le x\}} l(L_i).$$

Then the IS estimators of v_α and c_α, denoted by $\hat{v}_\alpha^{n,\mathrm{IS}}$ and $\hat{c}_\alpha^{n,\mathrm{IS}}$, can be defined as follows:

$$\begin{aligned}
\hat{v}_\alpha^{n,\mathrm{IS}} &= F_{n,\mathrm{IS}}^{-1}(\alpha) = \inf\{x : F_{n,\mathrm{IS}}(x) \ge \alpha\}, \quad \text{and} \\
\hat{c}_\alpha^{n,\mathrm{IS}} &= \hat{v}_\alpha^{n,\mathrm{IS}} + \frac{1}{n(1-\alpha)}\sum_{i=1}^{n}\left(L_i - \hat{v}_\alpha^{n,\mathrm{IS}}\right)^+ l(L_i).
\end{aligned}$$

Recently, Sun and Hong [2010] and Chu and Nakayama [2012] independently studied the Bahadur representations of the IS estimators. To present this result, we follow the framework of Sun and Hong [2010] and make a further assumption.

ASSUMPTION 2. *There exist $\epsilon > 0$ and $C > 0$ such that $l(x) \le C$ for any $x \in (v_\alpha - \epsilon,\ v_\alpha + \epsilon)$, and there exists $p > 2$ such that $\mathrm{E}_G\left[l^p(L)\right] < \infty$.*

Assumption 2 requires that the likelihood ratio is bounded from above in a neighborhood of v_α and has a finite $p > 2$ moment on the right tail of the loss.

The Bahadur representations of the IS estimators of v_α and c_α are summarized in the following theorem. Interested readers may refer to Sun and Hong [2010] for its proof.

THEOREM 2.2. *For a fixed $\alpha \in (0, 1)$, suppose that Assumptions 1 and 2 are satisfied. Then,*

$$\begin{aligned}
\hat{v}_\alpha^{n,\mathrm{IS}} &= v_\alpha + \frac{1}{f(v_\alpha)}\left(\alpha - \frac{1}{n}\sum_{i=1}^{n} 1_{\{L_i \le v_\alpha\}} l(L_i)\right) + C_n, \quad \textit{and} \\
\hat{c}_\alpha^{n,\mathrm{IS}} &= c_\alpha + \left(\frac{1}{n}\sum_{i=1}^{n}\left[v_\alpha + \frac{1}{1-\alpha}(L_i - v_\alpha)^+ l(L_i)\right] - c_\alpha\right) + D_n,
\end{aligned}$$

where $C_n = O_{a.s.}(\max\{n^{-1+2/p+\delta}, n^{-3/4+1/(2p)+\delta}\})$ and $D_n = O_{a.s.}(n^{-1+2/p+\delta})$ for any $\delta > 0$.

Asymptotic normality of the estimators follows immediately from Theorem 2.2. In particular, under Assumptions 1 and 2,

$$\sqrt{n}\left(\hat{v}_\alpha^{n,\mathrm{IS}} - v_\alpha\right) \Rightarrow \frac{\sqrt{\mathrm{Var}_G\left[1_{\{L \ge v_\alpha\}} l(L)\right]}}{f(v_\alpha)} N(0,1), \quad \text{as} \quad n \to \infty.$$

If, in addition, $\mathrm{E}_G[(L - v_\alpha)^2 l^2(L) 1_{\{L \ge v_\alpha\}}] < \infty$, then

$$\sqrt{n}\left(\hat{c}_\alpha^{n,\mathrm{IS}} - c_\alpha\right) \Rightarrow \frac{\sqrt{\mathrm{Var}_G\left[(L - v_\alpha)^+ l(L)\right]}}{1-\alpha} N(0,1), \quad \text{as} \quad n \to \infty.$$

If $l(x) \le 1$ for all $x \ge v_\alpha$, then it can be easily verified that $\mathrm{Var}_G[1_{\{L \ge v_\alpha\}} l(L)] \le \alpha(1-\alpha)$ and $\mathrm{Var}_G[(L - v_\alpha)^+ l(L)] \le \mathrm{Var}[(L - v_\alpha)^+]$. Then, compared to Equations (7) and (8),

it can be seen that the asymptotic variances of the IS estimators are smaller than those of the estimators without IS, given that $l(x) \leq 1$ for all $x \geq v_\alpha$. In practice, an effective IS distribution (with a density function g) often satisfies $g(x) \geq f(x)$ for $x \geq v_\alpha$. This provides a guideline for selecting an appropriate IS distribution during practical implementation.

2.3. Applications in Portfolio Risk Measurement

In financial applications such as portfolio risk measurement, the random variable L discussed in previous sections typically represents the loss of a portfolio up to a future time horizon. When the portfolio risk is measured by α-VaR or α-CVaR, it becomes a rare-event simulation problem, because the confidence level α is usually close to 1 and thus the upper tail of the loss distribution is of major concern in risk measurement. In this setting, IS has been successfully applied to enhance estimation efficiency. In what follows, we briefly review the use of IS in such contexts. For greater detail, interested readers may refer to Glasserman et al. [2000, 2002], Glasserman [2004], Glasserman and Li [2005], Glasserman et al. [2007, 2008], Bassamboo et al. [2008], and references therein.

2.3.1. Equity Portfolios. Consider a portfolio consisting of a number of financial instruments (e.g., stocks, options, etc.). The value of each instrument may depend on several risk factors (e.g., stock prices, interest rates, exchange rates, etc.). Suppose there are m risk factors on which the portfolio value depends. Let $S(t) = (S_1(t), \ldots, S_m(t))^T$ denote the values of the m risk factors at t, and $V(S(t), t)$ denote the value of the portfolio at time t. Over a future time interval $[t, t + \Delta t]$, the loss of the portfolio is

$$L = V(S(t), t) - V(S(t + \Delta t), t + \Delta t).$$

At time t, the loss of the portfolio relies on the change in risk factors during the time interval $[t, t + \Delta t]$, which is denoted by $\Delta S = S(t + \Delta t) - S(t)$. In a financial simulation, one typically simulates a number of scenarios of ΔS and then evaluates the corresponding loss L for each scenario. Then, based on the sample of L, estimates of the portfolio risk can be obtained. In practical problems, a large number of scenarios is usually required to obtain an accurate estimate because the portfolio risk is usually associated with the upper tail of the loss distribution. For instance, when estimating α-VaR with $\alpha = 99.9\%$, a very large sample may be required to obtain an accurate estimate because of the large estimation variance. In the literature, a considerable amount of work has been done to reduce the estimation variance. This subsection focuses on the use of IS and reviews the related work.

Estimating the risk associated with L (e.g., VaR and CVaR) is closely related to the problem of estimating $\Pr(L > x)$ for a given x. To see why it is so, we note that estimating α-VaR of L is equivalent to finding x such that $\Pr(L > x) = 1 - \alpha$ when L is a continuous random variable. In other words, if the probability $\Pr(L > x)$ can be estimated efficiently for a sequence of x values, then an estimate of α-VaR can be easily obtained by taking the inverse of the CDF. When the objective is to estimate CVaR, one typically employs a two-step procedure, in which an estimate of α-VaR, $\hat{v}_\alpha$, is obtained in the first step, and then an estimate of $\mathrm{E}[L|L \geq \hat{v}_\alpha]$ is obtained in the second step. It turns out that IS techniques that work for estimating $\Pr(L > \hat{v}_\alpha)$ usually perform well for estimating $\mathrm{E}[L|L \geq \hat{v}_\alpha]$ in the second step. Therefore, when studying IS techniques, much of the work in the literature starts with the problem of estimating the probability $\Pr(L > x)$. For instance, Glynn [1996] proposed several IS-based estimators for quantiles and showed that inverting the sample average (in terms of G) of $1 - 1_{\{L>x\}} l(L)$ sometimes leads to a preferred quantile estimator when

α is close to 1. Glasserman et al. [2000] mainly focused on $\Pr(L > x)$ when using IS techniques to estimate VaR.

To design an IS method for estimating $\Pr(L > x)$, the basic idea of Glasserman et al. [2000] is to approximate the portfolio loss by a quadratic function of the risk factors and to use this quadratic approximation to guide the selection of an appropriate IS distribution. More specifically, the portfolio loss can be approximated by

$$L \approx a_0 + a^T \Delta S + (\Delta S)^T A \Delta S \equiv a_0 + Q,$$

where a_0 is a scalar, a is a vector, and A is a matrix. As noted in Glasserman et al. [2000], such an approximation becomes more accurate when Δt becomes smaller.

Based on the delta-gamma approximation, $a_0 = -\Theta \Delta t$, $a = -\delta$, and $A = -\frac{1}{2}\Gamma$, where $\Theta = \partial V/\partial t$, $\delta_i = \partial V/\partial S_i$, and $\Gamma_{ij} = \partial^2 V/\partial S_i \partial S_j$ (all partial derivatives are evaluated at $(S(t), t)$). In financial industries, Θ, δ, and Γ are known as the Greek letters and are often routinely computed for other purposes.

In Glasserman et al. [2000], ΔS is assumed to follow a multivariate normal distribution with mean zero and covariance matrix Σ. If we let C be such that $CC^T = \Sigma$, then ΔS can be generated by $\Delta S = CZ$, where Z is a vector of independent standard normals. For the covariance matrix Σ, there are different choices for matrix C. Glasserman et al. [2000] show how to find such a C so that $C^T AC$ is a diagonal matrix. Then,

$$Q = a^T CZ + Z^T C^T ACZ = a^T CZ + Z^T \Lambda Z = b^T Z + Z^T \Lambda Z = \sum_{i=1}^{m} \left(\lambda_i Z_i^2 + b_i Z_i\right),$$

where $b^T = a^T C$, Λ is a diagonal matrix with λ_is in the diagonal, and it is assumed that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ for ease of discussion.

Intuitively, a large number of scenarios of L may be required to obtain an accurate estimate of $\Pr(L > x)$ because very few scenarios are obtained in the important region $L \approx x$ when x is large. To address this issue, Glasserman et al. [2000] suggested using exponential twisting to change the mean of Z from 0 to μ and its covariance from the identity matrix I to B such that the mean of L is equal to x under the IS distribution, thus generating more scenarios in the important region. More specifically, the changes in covariance B and in mean μ are restricted to the forms

$$B(\theta) = (I - 2\theta\Lambda)^{-1}, \quad \text{and} \quad \mu(\theta) = \theta B(\theta) b,$$

respectively, where θ is the twisting parameter satisfying $1 - 2\theta\lambda_i > 0$ for all i.

Exponential twisting requires a finite moment generating function of Q, and the likelihood ratio associated with the IS distribution can be represented as

$$l(Z) = \exp(-\theta Q + \psi(\theta)),$$

where $\psi(\theta)$ is the logarithm of the moment generating function of Q and is given by

$$\psi(\theta) = \sum_{i=1}^{m} \frac{1}{2}\left(\frac{(\theta b_i)^2}{1 - 2\theta\lambda_i} - \log(1 - 2\theta\lambda_i)\right).$$

Let E_θ denote expectation under the IS distribution with twisting parameter θ. Then,

$$\Pr(L > x) = \mathrm{E}\left[1_{\{L>x\}}\right] = \mathrm{E}_\theta\left[1_{\{L>x\}} l(Z)\right],$$

where Z follows a multivariate normal distribution with mean $\mu(\theta)$ and covariance $B(\theta)$.

To select a good IS distribution, the basic principle is to choose a twisting parameter such that the variance of $1_{\{L>x\}} l(Z)$ is minimized, which is equivalent to minimizing its

second moment

$$m_2(x,\theta) = \mathrm{E}_\theta\big[1_{\{L>x\}} l^2(Z)\big].$$

However, directly minimizing $m_2(x,\theta)$ is difficult. By noticing that

$$m_2(x,\theta) \le \exp(2\psi(\theta) - 2\theta(x-a_0)), \tag{9}$$

Glasserman et al. [2000] suggested minimizing the upper bound in Equation (9), which is equivalent to finding θ_x that solves the nonlinear equation

$$\psi'(\theta_x) = x - a_0.$$

As illustrated in Glasserman et al. [1999], this nonlinear equation is easy to solve numerically. An observation that is worth mentioning is that although θ_x is derived for the threshold value x, numerical evidence suggests that it is also an appropriate twisting parameter for nearby threshold values. This property is of great use when one estimates VaR by taking the inverse CDF.

The exponential twisting method in Glasserman et al. [2000] builds on the assumption that the risk factors ΔS follow a multivariate normal distribution. However, empirical evidence in finance literature suggests that the risk factors may be heavy tailed. When the risk factors are modeled by heavy-tailed distributions, such as a multivariate t distribution, the exponential twisting method is inapplicable because of the nonexistence of the moment generating function. To fix this issue, Glasserman et al. [2002] transformed the heavy-tailed problem into a light-tailed one and then applied exponential twisting.

Glasserman et al. [2002] developed efficient computational procedures based on modeling the joint distribution of risk factors using a multivariate t distribution and some extensions of it. We briefly describe their main idea as follows. Consider the case where ΔS follows a multivariate t distribution with mean zero. It can be represented by

$$\frac{Z}{\sqrt{Y/\nu}},$$

where Z is a multivariate normal random variable with mean zero and covariance Σ, and Y is a chi-square random variable with ν degrees of freedom. Moreover, Z and Y are independent.

Recall that because $L \approx a_0 + Q$, the problem becomes one of selecting an efficient change of measure for estimating $\Pr(Q > x - a_0)$ for large x. Define

$$Q_x = (Y/\nu)(Q - x + a_0).$$

Then

$$\Pr(Q > x - a_0) = \Pr(Q_x > 0).$$

While Q is heavy tailed, the newly defined Q_x is light tailed, and thus exponential twisting can be applied when estimating $\Pr(Q_x > 0)$. Via an argument similar to the multivariate normal case, Glasserman et al. [2002] determine a good twisting parameter and provide explicit IS distributions of Y and Z. Interested readers may refer to Glasserman et al. [2002] for more details.

2.3.2. Credit Portfolios. The IS methods reviewed in the previous subsection rely on quadratic approximations of portfolio loss, and the underlying model framework typically applies to equity portfolios. In the financial industry, another important problem is the risk measurement of credit portfolios that consist of defaultable instruments such as corporate bonds, bank loans, and credit derivatives. The model frameworks for credit portfolios are quite different from those for equity portfolios. The major modeling

challenge in portfolio credit risk is how to capture the dependence among the defaults of different borrowers or obligors. A widely used modeling framework assumes that the defaults of individual obligors are conditionally independent given a set of underlying factors, which may represent the common macroeconomic or industry-wide environment. In what follows, we focus on this conditional-independence framework and review the IS techniques tailored for this framework.

Consider a credit portfolio associated with m obligors. For $k = 1, \ldots, m$, we let Y_k denote the default indicator for the kth obligor, that is, $Y_k = 1$ if the obligor defaults over a fixed time horizon $[0, T]$, and $Y_k = 0$ otherwise. Marginal default probability of the kth obligor is denoted by p_k, which is an input to the model and is assumed to be a known constant. The loss resulting from the default of the kth obligor is assumed to be a continuous random variable C_k with a density $f_k(\cdot)$, where C_ks are assumed to be independent. Then the loss of the credit portfolio is

$$L = \sum_{k=1}^{m} C_k Y_k. \tag{10}$$

One of the most widely used models in capturing the dependence among Y_ks is the Gaussian copula model; see Gupton et al. [1997] and Li [2000]. In this model, the default indicator Y_k is specified as $Y_k = 1_{\{X_k > x_k\}}$ for $k = 1, \ldots, m$, where latent variables $X_1, \ldots, X_m$ follow standard normal distributions, and the constant x_k is chosen to match the marginal default probability p_k. Dependence among Y_ks is captured by the covariance of $X = (X_1, \ldots, X_m)$. More specifically, the covariance structure of X is determined by the following factor model:

$$X_k = a_{k1} Z_1 + \cdots + a_{kd} Z_d + b_k \epsilon_k, \quad k = 1, \ldots, m,$$

where $Z = (Z_1, \ldots, Z_d)^T$ denotes the systematic risk factors, $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^T$ denotes the idiosyncratic risk factors of individual obligors, $A = (a_{ij})$ is the factor loading matrix, and $b_i = \sqrt{1 - \sum_{j=1}^{d} a_{ij}^2}$, where we assume that $\sum_{j=1}^{d} a_{ij}^2 \le 1$ for all $i = 1, \ldots, m$. Here Z and ϵ are assumed to be independent of each other and both of them follow standard multivariate normal distributions.

Note that in the Gaussian copula model, Y_ks are conditionally independent given the systematic risk factors Z. Specifically, conditional on Z, Y_k is a Bernoulli random variable with a success probability

$$p_k(Z) = \Pr(Y_k = 1 | Z) = \Phi\left(\frac{a_k Z + \Phi^{-1}(p_k)}{b_k}\right), \tag{11}$$

with $a_k = (a_{k1}, \ldots, a_{kd})$, and Φ denoting the standard normal CDF.

Similar to the argument in the previous subsection, study of IS for portfolio credit risk starts with the problem of estimating $\Pr(L > x)$ for a given x. Glasserman and Li [2005] suggested a two-step IS such that more scenarios of L can be generated in the important region $L \approx x$. In the first step, the mean of the independent standard normal random vector Z is changed to a vector μ. Then, in the second step, the default probabilities $p_k(Z)$s and losses given defaults C_ks are exponentially twisted conditional on Z. We adopt the notation in Glasserman [2008] and illustrate the idea in greater detail.

We first consider the second step. Note that Y_ks are conditionally independent Bernoulli random variables given Z. Given a twisting parameter $\theta \in \Re$, define the cumulant generating functions of C_k and L by

$$\Lambda_k(\theta) = \log \mathrm{E}[\exp(\theta C_k)]$$

and

$$\psi_L(\theta, Z) = \sum_{k=1}^{m} \log(1 + p_k(Z)(\exp(\Lambda_k(\theta)) - 1)).$$

Exponential twisting suggests changing the conditional default probability $p_k(Z)$ to

$$p_{k,\theta}(Z) = \frac{p_k(Z)e^{\Lambda_k(\theta)}}{1 + p_k(Z)(e^{\Lambda_k(\theta)} - 1)}$$

and changing the density of C_k from $f_k(\cdot)$ to $f_{k,\theta}(v) = e^{\theta v - \Lambda_k(\theta)} f_k(v)$ for each $k = 1, \ldots, m$.

By the definitions of Λ_k and ψ_L, it can be verified that $\partial_\theta \psi_L(\theta, Z)$ is equal to the mean of L after changing $p_k(Z)$ and $f_k(v)$ to $p_{k,\theta}(Z)$ and $f_{k,\theta}(v)$, respectively. In other words, $\partial_\theta \psi_L(\theta, Z)$ is the mean of L after exponential twisting with the parameter θ.

To select a good twisting parameter conditional on Z, a natural choice is to find a θ such that the mean of L is equal to x, that is, to solve the equation

$$\frac{\partial}{\partial \theta} \psi_L(\theta_x(Z), Z) = x.$$

A subtle issue arises when the solution $\theta_x(Z) < 0$ for some Z, which corresponds to the case $\mathrm{E}[L|Z] > x$. In this case, one may prefer to replace $\theta_x(Z)$ with 0, because conditional on Z, $\{L > x\}$ is not a rare event and thus no exponential twisting is required. In summary, one may choose the twisting parameter as

$$\theta_x^+(Z) = \max(\theta_x(Z), 0),$$

and the corresponding conditional likelihood for this change of measure is

$$l_2(Z, L) = \exp\big(-\theta_x^+(Z)L + \psi_L\big(\theta_x^+(Z), Z\big)\big).$$

In the first step of the IS method, the mean of Z is changed from zero to μ, and thus the corresponding likelihood ratio is $l_1(Z) = \exp(-\mu^T Z + \mu^T \mu/2)$. To select an appropriate IS distribution of Z, one may choose μ so that it reduces variance in estimating the integral of $\Pr(L > x|Z)$ against the density of Z. An ideal IS distribution of Z for this problem has the well-known zero-variance density that is proportional to $\Pr(L > x|Z = z)\exp(-z^T z/2)$. However, sampling from this zero-variance density is in general impossible, because the normalization constant in this density is $\Pr(L > x)$, which is the quantity we seek. To circumvent this difficulty, Glasserman and Li [2005] suggested choosing μ such that the IS distribution of Z and the zero-variance distribution have the same mode, which corresponds to setting μ as the value of z that maximizes

$$\Pr(L > x|Z = z)\exp(-z^T z/2).$$

However, directly solving this optimization problem is difficult. To obtain an approximation, Glasserman and Li [2005] noticed that

$$\Pr(L > x|Z = z)\exp(-z^T z/2) \le \exp\big(-\theta_x^+(z)x + \psi_L\big(\theta_x^+(z), z\big) - z^T z/2\big) \tag{12}$$

and suggested minimizing the upper bound in Equation (12).

To summarize, the likelihood ratio of the two-step IS method is

$$l(Z, L) = l_1(Z)l_2(L, Z) = \exp\big(-\theta_x^+(Z)L + \psi_L\big(\theta_x^+(Z), Z\big)\big)\exp(-\mu^T Z + \mu^T \mu/2),$$

and thus $\Pr(L > x)$ can be estimated based on

$$\Pr(L > x) = \widetilde{\mathrm{E}}\big[1_{\{L>x\}} l(Z, L)\big],$$

where Z and L are generated from the two-step IS measure, and $\widetilde{\mathrm{E}}$ denotes the expectation under this IS measure.

On the theoretical side, Glasserman and Li [2005] showed the asymptotic optimality of the previous IS method for single-factor homogeneous models where Z is one-dimensional. For a multifactor model where Z is multidimensional, Glasserman et al. [2008] studied the asymptotic optimality of an IS method that is based on a mixture of mean shifts for Z.

The previous IS method works for a Gaussian copula model. For other credit risk models, it is not applicable. Bassamboo et al. [2008] studied a class of extremal dependence models, including the t copula as a special case. In contrast to the Gaussian model where the latent variable X_k is represented by $X_k = a_k Z + b_k \epsilon_k$, they assumed that

$$X_k = \frac{a_k Z + b_k \epsilon_k}{W}, \tag{13}$$

where W is a nonnegative random variable independent of Z and ϵ_ks. When W takes values close to zero, all X_ks are likely to be large, leading to many simultaneous defaults. Thus, W helps to capture the effect of common shocks.

For the extremal dependence models, Bassamboo et al. [2008] proposed two IS algorithms to efficiently estimate the portfolio risk. The first algorithm is based on exponential twisting of the common shock random variable W and conditional default probabilities, while the second algorithm is based on hazard-rate twisting on $1/W$ and exponential twisting on conditional default probabilities. For a detailed discussion of these algorithms, we refer readers to Bassamboo et al. [2008]. Another line of research on modeling of portfolio credit risk focuses on dynamic intensity-based point process models. For such models, Deng et al. [2012] recently proposed a sequential importance sampling and resampling scheme for estimating rare-event probabilities. They showed that a logarithmically efficient estimator of the probability of large loss can be obtained by selecting appropriate resampling weights.

2.4. Recent Advances in Portfolio Risk Measurement

Recall that in a financial simulation, one typically simulates a number of scenarios of the risk factors and then evaluates the corresponding loss L for each scenario. Then, based on the samples of L, estimates of the portfolio risk can be obtained. This simulation is often computationally costly because of two issues. The first issue is that a large number of scenarios for the risk factors is required because the portfolio risk is usually associated with the upper tail of the loss distribution. As reviewed in the previous subsection, IS has been successfully applied to address this issue.

The second issue is that given a scenario of the risk factors, evaluation of the corresponding portfolio loss L may be computationally costly. In many practical problems, closed-form formulas of L are not available, especially when complex pricing models are used. In this case, one may need to resort to numerical methods (e.g., Monte Carlo simulation and numerical integration) to approximate L. In recent years, increasing attention has been paid to this nested setting where both the risk factors and the corresponding portfolio loss require Monte Carlo simulation. Sometimes it is also referred to as a two-level simulation, in which the risk factors are simulated in the outer level, while the portfolio loss given each scenario of risk factors is simulated in the inner level. Straightforward implementation of the two-level (or nested) simulation scheme is computationally intensive. A key issue in this regard is to develop efficient ways to allocate computational budget to both outer and inner levels such that the resulting estimator achieves an optimal rate of convergence in a certain sense. Other

work toward addressing this issue focuses on the development of efficient simulation methods that are not limited to the nested simulation framework.

For ease of presentation, we simplify the notation. Denote the risk factors at time t by a random vector X and the discounted payoff of the portfolio at a maturity time T $(T > t)$ by a random variable P. Then the portfolio loss L is a function of X. By standard option pricing theory [Duffie 2001],

$$L \equiv L(X) = V_0 - \mathrm{E}_M[P|X] = \mathrm{E}_M[V_0 - P|X] = \mathrm{E}_M[Y|X],$$

where the constant V_0 denotes the value of the portfolio at time 0, $Y \triangleq V_0 - P$, and E_M denotes that the expectation is taken under the pricing martingale measure. Then the problem of interest becomes estimating a risk measure associated with L, for example, VaR or CVaR of L.

The standard nested simulation approach to the previous problem can be described as follows. We first simulate n scenarios of X, denoted by $\{X_i, 1 \le i \le n\}$. Then conditional on each X_i for $i = 1, \ldots, n$, we simulate K observations of Y, denoted by $\{Y_{ik}, 1 \le k \le K\}$. Let $L_i = \frac{1}{K}\sum_{k=1}^{K} Y_{ik}$. Then the samples $\{L_i, 1 \le i \le n\}$ can be used to estimate the portfolio risk. To ensure the convergence of the estimate, both n and K need to go to infinity. In many practical situations, the main computational effort is spent on simulating the inner-level sample Y_{ik}s. Thus, the total computational effort required for the standard nested simulation scheme is approximately proportional to nK.

Traditionally, the computational burden of the standard nested simulation scheme is often perceived to be unacceptable for practical problems. However, Gordy and Juneja [2010] showed that this perception is not necessarily true in the context of portfolio risk measurement. On the theoretical side, they analyzed how a fixed computational budget can be allocated to the inner and outer levels to minimize the mean square error (MSE) of the resulting estimator and established an asymptotic result that MSE of the estimator converges to zero at a rate of $\Gamma^{-2/3}$, where Γ represents the computational budget; see also Lee [1998] for a similar analysis. On the practical side, they showed that a relatively small number of inner-level samples could yield accurate estimates when measuring portfolio risk. It should also be noted that when Y is linear in X, the standard nested simulation scheme leads to an unbiased estimator even with $K = 1$.

The standard nested simulation scheme can be improved by exploiting structural information of the risk measures being considered. When the risk measure of interest is CVaR, an intuition is that the outer-level scenarios that lead to large portfolio losses play more important roles in estimation. To make use of this property, Lan et al. [2010] proposed using ranking-and-selection techniques to screen out some less important scenarios and thus allocate more of the computational effort to the remaining scenarios. For scenarios that survive screening, they set the sample size of the inner-level simulation proportional to the sample variance of the inner estimate, with the goal of equalizing the standard errors of the inner estimates. By doing so, they developed an efficient simulation scheme to construct a confidence interval for CVaR. Liu and Staum [2010] used stochastic kriging, a metamodeling technique studied in Ankenman et al. [2010], to speed up the nested simulation of CVaR by noticing that inner-level simulation is not necessary for every outer-level scenario when a metamodel has been constructed. When the risk measure of interest is the probability of a large loss, Broadie et al. [2010] proposed a sequential simulation scheme to allocate computational effort nonuniformly across outer-level scenarios. Their main idea is to allocate additional computational effort to scenarios with greater expected marginal changes to the risk measure. They used a Chebyshev bound to approximate the marginal change, which leads to implementable allocation rules. They further showed that the

nonuniform inner sampling approach results in an estimator with a faster rate of convergence than its uniform counterpart.

Different from the nested simulation schemes, Broadie et al. [2011] recently proposed estimating the inner expectation $\mathrm{E}[Y|X]$ using the least-squares method (LSM) and provided a theoretical analysis of the method. LSM is related to a vast literature in the area of approximate dynamic programming, in particular, on the least-squares approximations of value function in solving Bellman's equation; see Bertsekas [2007] for a good overview. In the context of optimal stopping, LSM was first investigated by Longstaff and Schwartz [2001] and Tsitsiklis and Van Roy [2001]; see also the previous work of Carrière [1996]. Note that LSM can also be interpreted from a metamodeling perspective. The method in Broadie et al. [2011] requires only one inner sample for each outer scenario and can be efficiently implemented. Its major drawback is that the estimate does not converge to the true value of the risk measure due to the bias introduced in the selection of basis functions. Under the same setting where only one inner sample is required for each outer scenario, Hong et al. [2012b] studied a kernel smoothing approach to estimating $\mathrm{E}[Y|X]$ and analyzed its asymptotic properties, especially the impact of dimensionality (e.g., number of risk factors) on the rate of convergence. While their asymptotic results suggest that the kernel smoothing approach is preferable over nested simulation only for low-dimensional problems, they illustrated through numerical experiments that kernel smoothing may have superior finite-sample performance for quite high dimensions, at least as high as 20. By noticing that each individual instrument in a portfolio often depends on only a relatively small number of risk factors, they further proposed a dimension reduction technique, through which a high-dimensional portfolio risk measurement problem can be decomposed into low-dimensional ones. With this dimension reduction technique, the kernel smoothing approach may become a viable tool for portfolio risk measurement.

Efficient simulation of portfolio risk under the nested setting is an important and challenging simulation problem. It is still an evolving area of research, in which further investigation is desirable and would be of benefit to simulation practitioners, especially given the widespread use of simulation in risk management practice.

2.5. Estimation Based on Stochastic Approximation

So far we have mainly discussed estimations of VaR/CVaR using a sample counterpart approach; that is, one first generates a set of observations and then uses the whole sample to construct estimates. This is also the main stream of study on estimations of VaR/CVaR. Besides this, there also exist other approaches to estimating VaR and CVaR. An interesting one is based on stochastic approximation, a classical procedure used to solve stochastic optimization problems (we call it a stochastic approximation (SA) approach). The main idea of the SA approach is to convert the estimation problem as a stochastic root finding problem and then apply corresponding procedures for example, a Robbins-Monro algorithm, combined with variance reduction techniques, for example, importance sampling to solve the root finding problem (see, e.g., Bardou et al. [2009]). As noted in Bardou et al. [2009], there is no reason to expect that a pure stochastic approximation procedure can do better than directly using the sample to conduct estimation. Nevertheless, the merit of the stochastic approximation approach is that it can incorporate certain variance reduction techniques to achieve efficient estimation.

3. SENSITIVITY ANALYSIS FOR VAR AND CVAR

To analyze the sensitivities of VaR and CVaR, we let θ denote the parameter with respect to which we calculate sensitivities. Without loss of generality, in this article, we assume that θ is one-dimensional and $\theta \in \Theta$, where $\Theta \subset \Re$ is an open set. If θ is

multidimensional, we may treat each dimension as a one-dimensional parameter while fixing other dimensions constants.

The random loss of interest may depend on the parameter, and thus we write it as $L(\theta)$ to incorporate the dependence. Correspondingly, we let $F(\cdot, \theta)$ and $f(\cdot, \theta)$ denote its CDF and probability density function (PDF), respectively, and $v_\alpha(\theta)$ and $c_\alpha(\theta)$ denote the α-VaR and α-CVaR, respectively. Furthermore, the subsequent sensitivity analysis may involve the derivative of $L'(\theta)$. In particular, we let $L'(\theta) = dL(\theta)/d\theta$ be the pathwise derivative, representing the sensitivity of a sample path of $L(\theta)$ with respect to an infinitesimal perturbation on θ (see, e.g., Ho and Cao [1983] for early work on perturbation analysis and Glasserman [1991] and Fu and Hu [1997] for detailed treatments and applications). We assume that $L'(\theta)$ can be evaluated for all $\theta \in \Theta$.

We give simple examples to help understand the notations and the framework. Consider first the equity portfolio example discussed in Section 2.3.1. Suppose we use the delta-gamma approximation of Glasserman [2004] to model the random loss:

$$L = a_0 + a^T \Delta S + \Delta S^T A \Delta S,$$

where a_0, a, and A are specified as in Section 2.3.1. Suppose further that ΔS follows a multivariate normal distribution with mean vector μ and covariance matrix Σ, where μ and Σ are estimated using historical data. The investor may want to know how estimation errors in μ and Σ, especially in μ, affect the portfolio risk of interest, as he or she realizes that the historical data may not be sufficient and the mean μ is especially difficult to estimate. Thus, he or she may want to know the sensitivities of the risk (measured by VaR and CVaR) of the portfolio with respect to μ. Without loss of generality, let us consider sensitivity to the mean μ_1 of the first element of ΔS. In this case, we can let $\theta = \mu_1$ and write L as $L(\theta)$, and VaR and CVaR of $L(\theta)$ become functions of θ. Sensitivities of VaR and CVaR with respect to θ can then be used to quantify the impact of estimation errors on the portfolio risk.

Consider next the credit risk model of Bassamboo et al. [2008] introduced in Section 2.3.2. In the model, the loss function L is given by Equation (10) and X_k is defined by Equation (13), where Z is a common risk factor, ϵ_k represents idiosyncratic risk factor of obligor k, and W is a nonnegative random variable that models the common shock to all obligors. Suppose that Z, X_k, and W are independent of each other. Suppose further that Z follows a standard normal distribution, X_k follows a normal distribution with mean μ_k and variance σ_k^2, and W follows an exponential distribution with rate μ. As in the equity portfolio example, the modeler does not fully trust the estimate of the individual obligor parameter μ_k as well as the common shock parameter μ. Then the modeler can study the sensitivities of the credit portfolio risk to these parameters. In this case, we can set $\theta = \mu_k$ or $\theta = \mu$, treat L as a function of the parameters, and estimate the derivatives of $v_\alpha(\theta)$ and $c_\alpha(\theta)$ accordingly.

In this section, we attempt to provide a unified view of sensitivities of VaR and CVaR by summarizing the results presented in several papers of the authors. We first derive the closed-form expressions of sensitivities of VaR and CVaR, then discuss the issues of Monte Carlo estimation, including asymptotic properties. Methods to accelerate the rates of convergence of VaR sensitivity estimators are also discussed.

3.1. Closed-Form Expressions

As a starting point toward deriving the closed-from expressions, we note that VaR (quantile) and CDF are closely related. Specifically, for a fixed $\alpha \in (0, 1)$,

$$F(v_\alpha(\theta), \theta) = \alpha. \tag{14}$$

We explore this relation to link together VaR sensitivity and probability sensitivity.

Since a CDF is simply an expectation and is easier to analyze, we start by analyzing probability sensitivity. We let $p_y(\theta) = F(y, \theta) = \Pr\{L(\theta) \le y\}$ for some threshold y. We treat $p_y(\theta)$ as a function of θ with y being a parameter. To facilitate the analysis, we first make the following assumptions.

ASSUMPTION 3. *For any $\theta \in \Theta$, $L'(\theta)$ exists w.p.1 and there exists a random variable K, which may depend on θ, such that $\mathrm{E}(K) < \infty$ and $|L(\theta + \Delta\theta) - L(\theta)| \le K|\Delta\theta|$ for any $\Delta\theta$ that is close enough to 0.*

ASSUMPTION 4. *For any $\theta \in \Theta$, $F(t, \theta)$ is continuously differentiable at (y, θ).*

Assumption 3 is a typical assumption used in pathwise derivative estimation (see, e.g., L'Ecuyer [1990], Glasserman [1991], and Broadie and Glasserman [1996]). It is essentially a form of the Lebesgue dominated convergence theorem and guarantees the validity of interchanging differentiation and expectation when evaluating $d\mathrm{E}(r(L(\theta)))/d\theta$ for any Lipschitz-continuous function $r(\cdot)$. Assumption 4 requires that $F(t, \theta)$ is continuously differentiable in the subregion of the two-dimensional region $\Re \times \Theta$ (i.e., fix the first dimension at y). Typically this is satisfied for practical problems. More discussions on this assumption can be found in Hong and Liu [2010].

Given the assumptions, Hong [2009] showed the following result.

THEOREM 3.1. *Suppose that Assumptions 3 and 4 are satisfied. Then,*

$$p'_y(\theta) = -\partial_y \mathrm{E}\big[L'(\theta) \cdot 1_{\{L(\theta) \le y\}}\big],$$

where ∂_y denotes the operator of taking partial derivative with respect to y. Furthermore, if $\mathrm{E}[L'(\theta)|L(\theta) = t]$ is continuous at $t = y$, then

$$p'_y(\theta) = -f(y, \theta)\mathrm{E}[L'(\theta)|L(\theta) = y].$$

Theorem 3.1 shows that probability sensitivity can be written as the product of a density and a conditional expectation. This is an important step toward developing estimation methods.

Starting from the result of Theorem 3.1, we may easily arrive at a closed-form expression of VaR sensitivity. The key is to use the identity in Equation (14). Specifically, by differentiating with respect to θ on both sides of Equation (14), we have

$$\partial_t F(v_\alpha(\theta), \theta)v'_\alpha(\theta) + \partial_\theta F(a, \theta)\big|_{a=v_\alpha(\theta)} = 0. \tag{15}$$

Noting that $\partial_t F(v_\alpha(\theta), \theta) = f(v_\alpha(\theta), \theta)$ and $\partial_\theta F(a, \theta) = p'_a(\theta)$, we immediately have

$$v'_\alpha(\theta) = -\frac{1}{f(v_\alpha(\theta), \theta)} p'_a(\theta)|_{a=v_\alpha(\theta)} = \mathrm{E}[L'(\theta)|L(\theta) = v_\alpha(\theta)]. \tag{16}$$

In other words, VaR sensitivity can be written as a conditional expectation, which is summarized in the following theorem.

THEOREM 3.2. *Suppose that Assumptions 3 and 4 are satisfied at $y = v_\alpha(\theta)$, and $\mathrm{E}[L'(\theta)|L(\theta) = t]$ is continuous at $t = v_\alpha(\theta)$. Then,*

$$v'_\alpha(\theta) = \mathrm{E}[L'(\theta)|L(\theta) = v_\alpha(\theta)].$$

Theorem 3.2 shows that under appropriate regularity conditions, VaR sensitivity is an expectation taken in the set defined by $\{L(\theta) = v_\alpha(\theta)\}$. Because the set is often a probability-zero set in many practical problems, the representation in Theorem 3.2 may not be directly applied for estimating VaR sensitivity. However, it does tell us a fact that only samples that lie in a neighborhood of $\{L(\theta) = v_\alpha(\theta)\}$ play crucial roles in estimation.

To derive a closed-from expression for CVaR sensitivity from VaR sensitivity, we can use the relation

$$c_\alpha(\theta) = v_\alpha(\theta) + \frac{1}{1-\alpha}\mathrm{E}[L(\theta) - v_\alpha(\theta)]^+.$$

The basic idea is to differentiate (with respect to θ) the previous equation on both sides. Since $h(x) = x^+$ is a Lipschitz continuous function, the interchange of expectation and differentiation is valid under some mild regularity conditions. We present the closed-form expression of CVaR sensitivity in the following theorem. The detailed proof of the result can be found in Hong and Liu [2009].

THEOREM 3.3. *Suppose that Assumptions 3 and 4 are satisfied at $y = v_\alpha(\theta)$, and $v_\alpha(\theta)$ is differentiable for any $\theta \in \Theta$. Then, for any $\theta \in \Theta$,*

$$c'_\alpha(\theta) = \mathrm{E}[L'(\theta) \big| L(\theta) \geq v_\alpha(\theta)].$$

3.2. Monte Carlo Estimation

To simplify the notation, we let $D(\theta)$ denote $L'(\theta)$ and suppress the dependence of L and D on θ when there is no confusion. Let $(L_1, D_1), \ldots, (L_n, D_n)$ be n i.i.d. observations of (L, D). Moreover, we let $L_{(i)}$ denote the ith order statistic of $\{L_1, \ldots, L_n\}$, and $D_{(i)}$ the corresponding pathwise derivative. We are interested in estimating $v'_\alpha(\theta)$ and $c'_\alpha(\theta)$ using these observations.

Hong [2009] proposed a batch-mean estimator for $v'_\alpha(\theta)$. The key idea is that $D_{(\lceil n\alpha \rceil)}$ can be viewed as an IPA (infinitesimal perturbation analysis) estimator of $v'_\alpha(\theta)$. However, it has been shown that this IPA estimator is not consistent [Hong 2009]. To circumvent the problem, a batch-mean estimator can be devised, which is consistent. Specifically, suppose that there exist positive integers m and k such that $m \times k = n$. Then, we divide the n i.i.d. observations into k batches and each batch has m observations. For each batch, we calculate the IPA estimator, denoted by D^b_m. Then, we have k observations of D^b_m, denoted as $D^b_{m,1}, \ldots, D^b_{m,k}$. A batch-mean estimator of $v'_\alpha(\theta)$ is then proposed:

$$\bar{D}_{mk} = \frac{1}{k}\sum_{l=1}^{k} D^b_{m,l}. \tag{17}$$

Another estimator of VaR sensitivity stems from kernel regression estimation, given the observation that $v'_\alpha(\theta)$ can be viewed as a regression function in nonparametric statistics [Bosq 1998]. Suppose that K is a bounded symmetric density such that $yK(y) \to 0$ as $|y| \to \infty$ and $\int_{-\infty}^{\infty} y^2 K(y)\,dy < \infty$. Then K is called a kernel on $\Re$. For instance, the standard normal density is a kernel. Moreover, we define the so-called bandwidth parameter δ_n, which satisfies $\delta_n \to 0$ and $n\delta_n \to \infty$ as $n \to \infty$. Then, a kernel estimator of $v'_\alpha(\theta)$ can be proposed as follows:

$$\bar{V}_n = \frac{\sum_{i=1}^{n} D_i K\left(\frac{\hat{v}^n_\alpha - L_i}{\delta_n}\right)}{\sum_{i=1}^{n} K\left(\frac{\hat{v}^n_\alpha - L_i}{\delta_n}\right)}. \tag{18}$$

Intuitively, the kernel estimator in Equation (18) can be viewed as a weighted average of L_is. The weight depends on the distance between L_i and v_α. If the distance is small, typically a larger weight is assigned. In other words, the observations closer to VaR play a more important role in the estimation.

Compared to VaR sensitivity, estimation of CVaR sensitivity is relatively simpler, in the sense that a typical sample-mean-type estimator can be proposed. Based on the closed-form expression for CVaR sensitivity, Hong and Liu [2009] proposed the

following estimator of $c'_\alpha(\theta)$:

$$\bar{C}_n = \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} D_i 1_{\{L_i \geq \hat{v}_\alpha^n\}}. \tag{19}$$

Compared to the kernel estimator proposed in Scaillet [2004] for linear loss functions, this estimator is more intuitive and does not require the selection of kernel function and bandwidth function.

We now outline the asymptotic results, mainly consistency and asymptotic normality, of the sensitivity estimators discussed earlier in this subsection. We do not go deep into the technical conditions that are imposed in the asymptotic analysis. Instead, we aim to provide a sketch of the major results.

For the batch-mean estimator in Equation (17), if $m \to \infty$ and $k \to \infty$ as $n \to \infty$, Hong [2009] proved its consistency under some regularity conditions. Furthermore, if $\lim_{n\to\infty} \sqrt{k}/m = 0$, it can be shown that

$$\sqrt{k}\big(\bar{D}_{mk} - v'_\alpha(\theta)\big) \Rightarrow \sigma_m N(0,1)$$

for some $\sigma_m > 0$. Therefore, the rate of convergence of the batch-mean estimator is $k^{-1/2}$, which is always strictly slower than $n^{-1/3}$ due to the constraint that $\lim_{n\to\infty} \sqrt{k}/m = 0$ and $n = m \times k$.

For the kernel estimator in Equation (18), Liu and Hong [2009] showed that it is consistent and follows asymptotic normality. Specifically, if $\delta_n \to 0$ and $n\delta_n^5 \to c$ as $n \to \infty$ for some constant $c > 0$, and $\sup_n (n\delta_n^3)^{-1} < \infty$, then under some regularity conditions,

$$\sqrt{n\delta_n}\big(\bar{V}_n - v'_\alpha(\theta)\big) \Rightarrow \mu + \sigma N(0,1)$$

for some constants μ and σ. It can be verified that the optimal bandwidth δ_n should be of order $n^{-1/5}$ if the criterion is to minimize the asymptotic mean square error. Under this optimal choice of bandwidth, the rate of convergence of the kernel estimator is $n^{-2/5}$.

For both the batch-mean and kernel estimators, we can see that the rates of convergence are slower than $n^{-1/2}$, which is the typical rate of convergence of sample-mean estimators. From this perspective, there is still room for improvement. In Section 3.3, we will discuss in more detail how to accelerate the rate of convergence.

The form of the CVaR sensitivity estimator in Equation (19) is very much like a sample-mean estimator, except that it involves a VaR estimator $\hat{v}_\alpha^n$. Although $\hat{v}_\alpha^n$ induces some technical complications in the analysis of $\bar{C}_n$, it does not affect the rate of convergence of the estimator. Hong and Liu [2009] showed that under some regularity conditions,

$$\sqrt{n}\big(\bar{C}_n - c'_\alpha(\theta)\big) \Rightarrow \sigma_1 N(0,1),$$

where

$$\sigma_1 = \frac{1}{(1-\alpha)^2} \mathrm{Var}\big[(L - v'_\alpha(\theta)) 1_{\{L \geq v_\alpha(\theta)\}}\big].$$

In other words, $\bar{C}_n$ does achieve the typical $n^{-1/2}$ rate of convergence.

Recently, Heidergott and Volk-Makarewicz [2012] applied the concept of measure-valued differentiation (MVD) to the estimation of VaR sensitivities. Note that, by Equation (15), we have

$$v'_\alpha(\theta) = -\frac{\partial_\theta F(v_\alpha(\theta), \theta)}{\partial_t F(v_\alpha(\theta), \theta)} = -\frac{\partial_\theta F(v_\alpha(\theta), \theta)}{f(v_\alpha(\theta), \theta)}. \tag{20}$$

By the theory of spacings of order statistics of i.i.d. data,

$$\lim_{m\to\infty} m\left(L_{\lceil m\alpha\rceil:m} - L_{\lceil m\alpha\rceil-1:m}\right) \Rightarrow \frac{\mathcal{E}}{f(v_\alpha(\theta),\theta)},$$

where $\mathcal{E}$ denotes an exponential random variable with mean one. Then, combined with Equation (20) and the VaR estimator $L_{\lceil m\alpha\rceil:m}$, Heidergott and Volk-Makarewicz [2012] obtained that

$$v'_\alpha(\theta) = \lim_{m\to\infty} \mathrm{E}\left[-m\left(L_{\lceil m\alpha\rceil:m} - L_{\lceil m\alpha\rceil-1:m}\right)\partial_\theta F\left(L_{\lceil m\alpha\rceil:m},\theta\right)\right].$$

Furthermore, by applying MVD, we can often write

$$\partial_\theta F(t,\theta) = c(\theta)[F^+(t,\theta) - F^-(t,\theta)]$$

for some constant $c(\theta)$ and $F^{\pm}(t,\theta)$ distribution functions. Then, one may use a batching idea to estimate $v'_\alpha(\theta)$. Suppose that there are n observations of L. One may divide them into k batches and each batch consists of m independent observations of L. Then, for each batch, an estimator in the form of

$$-mc(\theta)\left(L_{\lceil m\alpha\rceil:m} - L_{\lceil m\alpha\rceil-1:m}\right)\left[F^+\left(L_{\lceil m\alpha\rceil:m},\theta\right) - F^-\left(L_{\lceil m\alpha\rceil:m},\theta\right)\right]$$

can be computed. Then, the sample mean of all k estimators is the MVD estimator of $v'_\alpha(\theta)$. Heidergott and Volk-Makarewicz [2012] further studied cases where an explicit form of $F(t,\theta)$ is not available and also proved the consistency and asymptotic normality of the MVD estimator. It is worthwhile to note that the MVD estimator has the same optimal rate of convergence as the batching estimator of Hong [2009].

3.3. Accelerating the Rate of Convergence of VaR Sensitivity Estimators

As shown in the previous subsection, a kernel estimator of VaR sensitivity may not achieve the $n^{-1/2}$ rate of convergence, which is the typical rate of convergence for statistical estimation. To improve estimation efficiency, it is desirable to develop methods that may accelerate the rate of convergence. This is still an evolving area of research that is of great interest to both researchers and practitioners. In general, successful adoption of such methods depends significantly on the problem being tackled, and it often comes with an exploration of the specific features of the problem itself. In what follows, we present a simulation framework of accelerating the rate of convergence using two classical simulation techniques, IS and conditional Monte Carlo, which has proven to be effective in some practical problems. Interested readers may refer to Hong and Liu [2010] and Fu et al. [2009] for more details of using IS and conditional Monte Carlo in this context.

3.3.1. Importance Sampling. Before the presentation, we want to first clarify a slight difference between the IS we employ here and the classical IS used in the context of variance reduction. Our primary goal is to accelerate the asymptotic rate of convergence, while classical IS has a different focus and may not accelerate the rate of convergence. Related work on the classical use of IS for estimating VaR sensitivity includes Glasserman [2005] and Tasche [2009], which apply the IS method developed in Glasserman and Li [2005] to reduce the variance in estimating risk contributions. However, their methods do not accelerate the asymptotic rate of convergence.

Recall that VaR sensitivity can be written as $\mathrm{E}[D|L = v_\alpha]$, the estimation of which involves two key issues. The first is that v_α is unknown and thus has to be estimated. The second is how to estimate the conditional expectation given an estimate of v_α. It turns out that the first issue can be easily resolved by replacing v_α with a known estimator in the literature, for example, the sample quantile $\hat{v}^n_\alpha$. The second issue is of major concern, and thus we focus exclusively on this issue during the discussions of

accelerating the rate of convergence. In particular, the problem of interest now is how to estimate $r(y) \equiv \mathrm{E}[D|L = y]$ for a given y.

As a starting point, we first consider the kernel estimator of $r(y)$. If we choose the kernel function to be a uniform density over $[-1, 1]$, then it can be seen that a kernel estimator of $r(y)$ is

$$\bar{V}_n = \frac{\frac{1}{2n\delta_n}\sum_{i=1}^{n} D_i 1_{\{y-\delta_n \le L_i \le y+\delta_n\}}}{\frac{1}{2n\delta_n}\sum_{i=1}^{n} 1_{\{y-\delta_n \le L_i \le y+\delta_n\}}}. \tag{21}$$

Note that $\bar{V}_n$ is a ratio of two estimators, where both numerator and denominator are kernel estimators. We may first analyze the numerator, while analysis of the denominator follows in a similar manner. Specifically, let

$$\bar{M}_n = \frac{1}{2n\delta_n}\sum_{i=1}^{n} D_i 1_{\{y-\delta_n \le L_i \le y+\delta_n\}}$$

denote the numerator. Because the denominator in Equation (21) is a kernel estimator of $f(y)$, $\bar{M}_n$ is actually an estimator of $r(y)f(y)$.

An implication of the kernel estimator in Equation (21) is that only the observations in the region $\{y - \delta_n \le L \le y + \delta_n\}$ play roles in the estimation. Thus, we call it an important region. Intuitively, the total number of observations that lie in the important region is of order $n\delta_n$. This explains why the rate of convergence of both the numerator and denominator of Equation (21) is of order $\sqrt{n\delta_n}$. Based on an analysis of ratio estimators (see, e.g., Law and Kelton [2000], pp. 532–533), it can be shown that the rate of convergence of $\bar{V}_n$ is also of order $\sqrt{n\delta_n}$ (see, e.g., Liu and Hong [2009] for a rigorous proof).

To accelerate the rate of convergence of $\bar{V}_n$, a reasonable direction is to generate more observations that fall into the important region. To do so, IS is a natural choice. Among many possible alternatives, one choice of IS scheme is to force all observations to fall into the important region under the IS distribution. We will show how this can be achieved under quite a general simulation framework. It should be emphasized that in this setting, the original distribution is not absolutely continuous with respect to the IS distribution. Instead, absolute continuity is satisfied only within the important region $\{y-\delta_n \le L \le y+\delta_n\}$, which is sufficient for IS to work in our setting; Glynn and Iglehart [1989] also noticed such a sufficient condition in a discussion on general IS theory.

Because the density of L is typically unknown, a construction of IS scheme directly from the density function of L may not be feasible. However, in simulation models, L can often be expressed as a function of a sequence of random variables with known densities. Suppose that $L = L(X_1, X_2, \ldots, X_k)$ and $D = D(X_1, X_2, \ldots, X_k)$, where $X_1, X_2, \ldots, X_k$ are independent random variables with known densities $f_1, f_2, \ldots, f_k$, respectively. Then the joint density of $(X_1, X_2, \ldots, X_k)$ is $f(x_1, x_2, \ldots, x_k) = f_1(x_1)f_2(x_2)\cdots f_k(x_k)$.

Suppose that $X_1, X_2, \ldots, X_k$ are generated sequentially in the simulation to obtain observations of (L, D). Let

$$A = \left\{(x_1, x_2, \ldots, x_k) \in \Re^k : L(x_1, x_2, \ldots, x_k) \in [y - \delta_n, y + \delta_n] \text{ and } f_1(x_1)\cdots f_k(x_k) > 0\right\}.$$

Then A corresponds to the important region. Let A_1 be the projection of the set A to the first dimension, $A_2(X_1)$ be the projection of the set A to the second dimension given $X_1, \ldots,$ and $A_k(X_1, X_2, \ldots, X_{k-1})$ be the projection of the set A to the kth dimension given $X_1, X_2, \ldots, X_{k-1}$. Then we may define the IS distribution as

$$\tilde{f}(x_1, x_2, \ldots, x_k) = \frac{f_1(x_1)}{\Pr\{X_1 \in A_1\}} \cdot \frac{f_2(x_2)}{\Pr\{X_2 \in A_2(x_1)\}} \cdots \frac{f_k(x_k)}{\Pr\{X_k \in A_k(x_1, \ldots, x_{k-1})\}}$$

for all $(x_1, x_2, \ldots, x_k) \in A$, and $\tilde{f}(x_1, x_2, \ldots, x_k) = 0$ otherwise.

In an ideal case, sampling from this IS distribution can be accomplished by first simulating X_1 given that $X_1 \in A_1$, then simulating X_2 given that $X_2 \in A_2(X_1)$, and so on. (Whether such a sampling procedure is practically feasible is problem dependent; see Hong and Liu [2010] for several examples on successful design of such sampling procedures.) Then we can compute L and D and ensure that $L \in [y - \delta_n, y + \delta_n]$ w.p.1. Since the IS distribution $\widetilde{f}$ is absolutely continuous with respect to f in the set A, the likelihood ratio

$$\begin{aligned} P_n &= \frac{f(X_1, X_2, \ldots, X_k)}{\widetilde{f}(X_1, X_2, \ldots, X_k)} \\ &= \Pr\{X_1 \in A_1\} \cdot \Pr\{X_2 \in A_2(X_1)|X_1\} \cdots \Pr\{X_k \in A_k(X_1, \ldots, X_{k-1})|X_1, \ldots, X_{k-1}\}. \end{aligned}$$

Since $(X_1, X_2, \ldots, X_k)$ is a random vector, P_n is also a random variable and $P_n \leq 1$. Then IS estimators of $r(y)f(y)$ and $r(y)$ are

$$\bar{M}_n^{\mathrm{IS}} = \frac{1}{2n\delta_n} \sum_{i=1}^{n} P_{n,i} \cdot D_i 1_{\{y-\delta_n \leq L \leq y+\delta_n\}} = \frac{1}{2n\delta_n} \sum_{i=1}^{n} P_{n,i} \cdot D_i,$$

and

$$\bar{V}_n^{\mathrm{IS}} = \frac{\frac{1}{2n\delta_n} \sum_{i=1}^{n} P_{n,i} \cdot D_i}{\frac{1}{2n\delta_n} \sum_{i=1}^{n} P_{n,i}},$$

respectively, where the observations $(L_i, D_i, P_{n,i})$ are generated under the IS distribution.

Let $\widetilde{\mathrm{E}}$ and $\widetilde{\mathrm{Var}}$ denote taking expectation and variance under the IS distribution. Then it can be easily shown that $\widetilde{\mathrm{E}}(\bar{M}_n^{\mathrm{IS}}) = \mathrm{E}(\bar{M}_n)$ and

$$\begin{aligned} \widetilde{\mathrm{Var}}\big(\bar{M}_n^{\mathrm{IS}}\big) &= \frac{1}{4n\delta_n^2}\big[\mathrm{E}\big(D^2 \cdot P_n \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big) - \mathrm{E}^2\big(D \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big)\big] \\ &\leq \frac{1}{4n\delta_n^2}\big[\mathrm{E}\big(D^2 \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big) - \mathrm{E}^2\big(D \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big)\big] = \mathrm{Var}(\bar{M}_n). \end{aligned}$$

Therefore, $\bar{M}_n^{\mathrm{IS}}$ has a smaller variance than $\bar{M}_n$ when both estimators use the same δ_n.

Note that $\widetilde{\mathrm{E}}(P_n) = \mathrm{E}(1_{\{y-\delta_n \leq L \leq y+\delta_n\}}) = 2f(y)\delta_n + o(\delta_n)$. In many situations, we can prove that $P_n = K_n\delta_n$, where K_n is some random variable for which $\mathrm{E}(K_n \cdot D^2 \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}})$ is often of $O(\delta_n)$. Then,

$$\begin{aligned} \widetilde{\mathrm{Var}}\big(\bar{M}_n^{\mathrm{IS}}\big) &= \frac{1}{4\delta_n^2}\big[\widetilde{\mathrm{E}}\big(P_n^2 \cdot D^2\big) - \widetilde{\mathrm{E}}^2\big(P_n \cdot D\big)\big] \\ &= \frac{1}{4n\delta_n^2}\big[\delta_n\widetilde{\mathrm{E}}\big(P_nK_n \cdot D^2 \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big) - \widetilde{\mathrm{E}}^2\big(P_n \cdot D \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big)\big] \\ &= \frac{1}{4n\delta_n^2}\big[\delta_n\mathrm{E}\big(K_n \cdot D^2 \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big) - \mathrm{E}^2\big(D \cdot 1_{\{y-\delta_n \leq L \leq y+\delta_n\}}\big)\big], \end{aligned}$$

which is of $O(n^{-1})$. Then the rate of convergence of $\bar{M}_n^{\mathrm{IS}}$ is $n^{-1/2}$, and so is $\bar{V}_n^{\mathrm{IS}}$ [Law and Kelton 2000, pp. 532–533], if $n\delta_n^4 \to a$ with $a \geq 0$ as $n \to \infty$.

The previous framework shows that IS can be employed to improve the performance of the VaR sensitivity estimators and may achieve the $n^{-1/2}$ rate of convergence, which is the canonical rate that can be achieved for a typical Monte Carlo estimator.

We close this section by emphasizing that successful adoption of this framework is problem dependent. Other than the examples illustrated in Hong and Liu [2010] and

Liu [2012] considered the problem of simulating capital allocations for credit portfolio where the portfolio risk is measured by VaR. The use of the IS framework in other potential applications is interesting and deserves further investigation.

3.3.2. Conditional Monte Carlo. Conditional Monte Carlo (CMC) is an important simulation technique that can be used for both variance reduction and sensitivity analysis (see, e.g., Asmussen and Glynn [2007] and Fu and Hu [1997]). It turns out that CMC may also be employed in the context of estimating VaR sensitivities and may lead to estimators that achieve the $n^{-1/2}$ rate of convergence.

Recall that

$$F(v_\alpha(\theta), \theta) = \alpha,$$

and thus differentiating with respect to θ on both sides we have

$$v'_\alpha(\theta) = -\left.\frac{\partial_\theta F(t,\theta)}{\partial_t F(t,\theta)}\right|_{t=v_\alpha(\theta)}.$$

Note that the CDF $F(t,\theta)$ is essentially the expectation of an indicator, that is, $F(t,\theta) = \Pr\{L(\theta) \le t\} = \mathrm{E}[1_{\{L(\theta)\le t\}}]$. Because the indicator is a discontinuous function, the interchange of differentiation and expectation is not valid. However, this problem can be circumvented using conditional Monte Carlo, which cancels the discontinuity and thus makes the interchange valid. Specifically, we have the following theorem, whose proof can be found in Fu et al. [2009].

THEOREM 3.4. *Suppose that there exist random variables $X_1(\theta)$ and $X_2(\theta)$ such that*

$$F(t,\theta) = \mathrm{E}[\Pr\{L(\theta) \le t | X_i(\theta)\}] = \mathrm{E}[G_i(t, X_i(\theta), \theta)], \quad i = 1, 2,$$

where $G_1(t, X_1(\theta), \theta)$ is differentiable w.p.1 with respect to θ and $G_2(t, X_2(\theta), \theta)$ is differentiable w.p.1 with respect to t, and there exist random variables K_1 and K_2 with $\mathrm{E}(K_1) < \infty$ and $\mathrm{E}(K_2) < \infty$ such that

$$|G_1(t, X_1(\theta+\Delta\theta), \theta+\Delta\theta) - G_1(t, X_1(\theta), \theta)| \le K_1|\Delta\theta|, \quad \textit{and}$$
$$|G_2(t+\Delta t, X_2(\theta), \theta) - G_2(t, X_2(\theta), \theta)| \le K_2|\Delta t|,$$

for all small enough $\Delta\theta$ and Δt. If $F(t,\theta)$ and $v_\alpha(\theta)$ are both differentiable, then

$$v'_\alpha(\theta) = -\left.\frac{\mathrm{E}[\partial_\theta G_1(t, X_1(\theta), \theta)]}{\mathrm{E}[\partial_t G_2(t, X(\theta), \theta)]}\right|_{t=v_\alpha(\theta)}.$$

Theorem 3.4 shows that under appropriate conditions, VaR sensitivity can be written as a ratio of two expectations. In some practical problems, by carefully selecting appropriate $X_1(\theta)$ and $X_2(\theta)$, we may be able to calculate $\partial_\theta G_1(t, X_1(\theta), \theta)$ and $\partial_t G_2(t, X(\theta), \theta)$ analytically. In such cases, we may estimate VaR sensitivity using a ratio of two sample means.

Specifically, let $Y(t)$ and $Z(t)$ denote $\partial_\theta G_1(t, X_1(\theta), \theta)$ and $\partial_t G_2(t, X(\theta), \theta)$, respectively, for any $t \in \Re$. Suppose that, for any $t \in \Re$, we have n i.i.d. observations $(L_1, Y_1(t), Z_1(t)), \ldots, (L_n, Y_n(t), Z_n(t))$. Let $\bar{Y}_n(t) = (1/n)\sum_{i=1}^n Y_i(t)$ and $\bar{Z}_n(t) = (1/n)\sum_{i=1}^n Z_i(t)$. Then we can estimate $v'_\alpha(\theta)$ by

$$\bar{V}_n^{\mathrm{CMC}} = -\frac{\bar{Y}_n(\hat{v}_\alpha^n)}{\bar{Z}_n(\hat{v}_\alpha^n)}.$$

It has been shown in Fu et al. [2009] that under some regularity conditions, the rate of convergence of $\bar{V}_n^{\mathrm{CMC}}$ is $n^{-1/2}$, which is the canonical rate that a typical Monte Carlo estimator can achieve.

Typically, the CMC estimators of VaR sensitivities perform very well when applicable. Fu et al. [2009] studied sensitivity analysis for the portfolio credit risk model of Bassamboo et al. [2008]. Specifically, they considered the sensitivity to some parameter of an individual obligor and the sensitivity to some parameter of the common shock, and discussed how to calculate the corresponding functions $\partial_\theta G_1(t, X_1(\theta), \theta)$ and $\partial_t G_2(t, X(\theta), \theta)$. Furthermore, some numerical results have been presented in Fu et al. [2009] to illustrate the advantages of CMC. However, as is always the case in applying conditional Monte Carlo, the choice of appropriate conditioning variables is problem dependent. Further investigation on applying CMC to various practical problems is desirable and would be of benefit to simulation practitioners.

4. OPTIMIZATION OF VAR AND CVAR

Evaluating values and sensitivities of risk measures is the foundation of risk assessment and risk control. They are often considered as passive behaviors in the practice of risk analysis. Perhaps a more active and important practice for decision makers is to efficiently manage risks in financial activities. Quantification of risk via risk measures allows decision makers to build various decision models. As important risk measures, VaR and CVaR naturally enter these decision models, resulting in optimization problems that either optimize (say, minimize) certain VaR/CVaR functions, which we call the VaR/CVaR minimization problem, or optimize certain objectives subject to VaR/CVaR constraints, which we call the VaR/CVaR constrained program. In the optimization models, the random loss L that we are interested in becomes a random variable that depends on our decisions. More specifically, we assume that $L = L(x, \xi)$; that is, L is a function of the decision vector x and the random vector ξ. Furthermore, to explicitly show the dependence of risk measures to the loss function, we denote by $v_\alpha(L(x, \xi))$ and $c_\alpha(L(x, \xi))$ the VaR and CVaR of $L(x, \xi)$.

Let us see the loss functions in simple examples of equity portfolio and credit portfolio. Consider a one-period equity portfolio selection problem. Suppose the portfolio consists of d assets and the investor would like to invest his or her wealth W among these assets. The value of these assets will evolve during the period. Suppose the initial value is $\xi_0 = (\xi_0^1, \xi_0^2, \ldots, \xi_0^d)^T$, which is given, and the value vector at maturity is $\xi = (\xi^1, \xi^2, \ldots, \xi^d)^T$, which is a random vector. Let x_j denote the wealth invested on asset j. Then $x = (x_1, x_2, \ldots, x_d)^T$ is the decision vector and the loss of the portfolio at maturity takes the expression $L(x, \xi) = (\xi_0 - \xi)^T x$. Consider next a credit portfolio selection problem, in which the investor wants to lend his or her wealth W to d obligors. Similarly, let $x = (x_1, x_2, \ldots, x_d)^T$, where x_j is the wealth lending to obligor j. As in Section 2.3.2, suppose Y_j is the default indicator of obligor j. Then Y_j depends on some random vector ξ. To indicate the dependence, we write Y_j as $Y_j(\xi)$. Then the loss function for the credit portfolio is $L(x, \xi) = \sum_{j=1}^d Y_j(\xi) x_j$. In both examples, using VaR and CVaR as risk measures, the investor can build different optimization models according to his or her purpose. By solving the corresponding model, the investor can obtain the desired investment decision.

In this section, we briefly review VaR and CVaR optimization models and the techniques and approaches developed to handle them.

4.1. VaR Optimization

The VaR minimization problem takes the following representation

$$\underset{x \in X}{\text{minimize}} \; v_\alpha(L(x, \xi)), \tag{22}$$

where X is the feasible region, which may be defined by a set of deterministic constraints. Throughout this article, we assume that X is a convex and compact subset of the d-dimensional space $\Re^d$. The VaR constrained program can be formulated as

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \quad v_\alpha(L(x,\xi)) \le 0, \end{aligned} \tag{23}$$

where $h(x)$ is the objective that we want to optimize. Note that in Problem (23), we assume the right-hand side of the VaR constraint is 0. If it is another constant c, we can redefine $L = L - c$ and convert the problem into Problem (23) using the property that $v_\alpha(L + c) = v_\alpha(L) + c$ for any constant c.

Clearly, the difficulty of solving these models depends critically on the expression of h and/or L. Furthermore, as has been mentioned in Section 1, VaR lacks the subadditivity and is not a coherent risk measure. This implies the VaR optimization problems may not be a convex optimization problem even when $h(x)$ is convex in x and $L(x,\xi)$ is convex in x for every ξ. Therefore, models involving VaR are in general difficult to solve. Due to the difficulty, in the following discussion we mainly focus on the simple linear equity portfolio model introduced earlier. We also mention if the methodologies can be applied to more complicated cases such as the credit portfolio model. Also, due to the nonconvexity, we typically try to find a good feasible solution or a locally optimal solution for the problems instead of guaranteeing to find a globally optimal solution. On the other hand, there exist some cases where VaR optimization problems turn out to be convex, which we may or may not know a priori.

A basic fact is that $v_\alpha(L(x,\xi))$ itself is a nonlinear function of the decision vector x. Therefore, we can view VaR optimization problems as general nonlinear optimization problems. Suppose that $v_\alpha(L(x,\xi))$ is smooth in x. We can implement gradient-based nonlinear optimization algorithms to solve them. The key to this approach is to estimate the function values as well as the gradients of VaR at different feasible points. We have reviewed a number of techniques in Sections 2 and 3 that can do this. Using these estimates, we can implement the gradient-based approach with the estimated function values and gradient values to solve the problems. Clearly, to apply the gradient-based approach, one first step is to check the smoothness (in x) of the $v_\alpha(L(x,\xi))$ function. In the equity portfolio problem, if ξ is some continuous random vector, the smoothness can typically be guaranteed. However, in the credit portfolio problem, the indicator $Y_j(\xi)$ has a discrete distribution. Thus, the smoothness may be difficult to ensure.

Another, or perhaps a more conventional, way to handle VaR optimization problems is to treat them as chance-constrained programs (CCPs). Specifically, Problem (22) can be reformulated as the following CCP:

$$\begin{aligned} &\underset{x\in X, t\in\Re}{\text{minimize}} \quad t \\ &\text{subject to} \quad \Pr\{L(x,\xi) - t \le 0\} \ge \alpha, \end{aligned} \tag{24}$$

whereas Problem (23) is equivalent to the following CCP:

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \quad \Pr\{L(x,\xi) \le 0\} \ge \alpha. \end{aligned} \tag{25}$$

CCPs are a large class of stochastic programs. They were first introduced and studied by Charnes et al. [1958], Miller and Wagner [1965], and Prékopa [1970]. Since then, there have been various approaches developed to handling the CCPs. In this section, we mainly focus on those approaches that rely on Monte Carlo simulations.

4.1.1. Scenario Approach. An important method to handle the CCPs is called the scenario approach. This approach was studied first by Calafiore and Campi [2005, 2006] and De Farias and Van Roy [2004], independently, and developed further by Nemirovski and Shapiro [2005]. The basic idea of the scenario approach is to require the constraint within the probability function to be satisfied for all the scenarios generated from the random vector ξ. When applied to Equation (25), for n independent realizations of ξ denoted by $\xi_1, \xi_2, \ldots, \xi_n$, this gives the *scenario counterpart* of the CCP:

$$\begin{aligned} &\underset{x \in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \quad L(x, \xi_j) \le 0, \quad j = 1, 2, \ldots, n. \end{aligned} \tag{26}$$

If h and L are convex, then Equation (26) is a convex problem and there are efficient techniques to solve it. The fundamental question of the scenario approach is how to determine the sample size n so that the optimal solution of the scenario counterpart is in the feasible region of the original CCP with a specified probability. Calafiore and Campi [2006] showed that, if $h(x)$ is convex in x and $L(x, \xi)$ is convex in x for every ξ, and if

$$n \ge \left\lceil \inf_{v \in (0,1)} \frac{1}{1-v} \left(\frac{1}{1-\alpha} \log \frac{1}{\delta} + d + \frac{d}{1-\alpha} \log \frac{1}{v(1-\alpha)} + \frac{1}{1-\alpha} \log \frac{\left(\frac{d}{e}\right)^d}{d!} \right) \right\rceil, \tag{27}$$

then under some additional regularity conditions, the optimal solution of Problem (26) is feasible to Problem (25) with probability at least $1 - \delta$, where $\delta \in (0, 1)$ is typically a small positive constant. Calafiore and Campi [2006] further pointed out that a special case of the bound Equation (27) is

$$n \ge \left\lceil \frac{2}{1-\alpha} \log \frac{1}{\delta} + 2d + \frac{2d}{1-\alpha} \log \frac{2}{1-\alpha} \right\rceil.$$

The elegant result of Calafiore and Campi [2006] shows that the lower bound on the sample size n is independent of the functions $h(x)$ and $L(x, \xi)$. The merit of the scenario approach is that it is very general and easy to apply. Furthermore, according to our experience, solving the scenario counterpart is often computationally easy even when the sample size n is large. However, there are also some potential concerns about the scenario approach. It is not difficult to see that the scenario approach does not really solve the original CCPs. A consequence is that the solutions are often very conservative. Furthermore, the solutions obtained may vary significantly across simulation replications and do not converge as the sample size n goes to infinity. Nemirovski and Shapiro [2006] showed that the scenario approach is often more conservative than a convex conservative approximation called Bernstein approximation that they proposed. The numerical experiments in Hong et al. [2011] also revealed some numerical issues of the scenario approach.

To further understand the scenario approach for VaR optimization, we now apply the approach to the VaR minimization Problem (24), which yields the following optimization problem:

$$\begin{aligned} &\underset{x \in X}{\text{minimize}} \quad t \\ &\text{subject to} \quad L(x, \xi_j) - t \le 0, \quad j = 1, \ldots, n. \end{aligned} \tag{28}$$

The optimal value of Problem (28) is equal to that of the following problem:

$$\underset{x \in X}{\text{minimize}} \ \underset{j=1,\ldots,n}{\text{maximize}} \ L(x, \xi_j). \tag{29}$$

Problem (29) explicitly shows that using the scenario approach to solve the VaR minimization problem is equivalent to minimizing the worst-case loss function when the random vector takes values in the set of scenarios that have been observed. Consequently, when the sample size n becomes larger and larger, the optimal value of Problem (29) that is used to approximate the optimal VaR value typically becomes larger and larger. In the extreme case where ξ takes all values in its support, Problem (29) becomes the framework of robust optimization. Understanding this relation, we think the scenario approach may not be an ideal technique to solving VaR minimization problems.

4.1.2. Sample Average Approximation. Note that CCPs are stochastic programs of a special category, in that the probability function can be expressed as the expectation of an indicator function, that is,

$$\Pr\{L(x,\xi) > 0\} = \mathrm{E}\left[1_{\{L(x,\xi)>0\}}\right].$$

There are a number of Monte Carlo methods that can be used to solve stochastic programs. Sample average approximation (SAA) is a very popular one. However, the indicator function inside of the expectation is discontinuous. This may create both theoretical and computational issues for the use of SAA. Using SAA to handle CCPs has been studied in the literature (see, e.g., Luedtke and Ahmed [2008] and Pagnoncelli et al. [2009]). In contrast to the scenario approach, which requires the constraints to be satisfied for all the realizations of the random vector, the SAA approach uses the sample mean estimator to replace the expectation and suggests solving the following problem:

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \quad \frac{1}{n}\sum_{j=1}^{n} 1_{\{L(x,\xi_j)>0\}} \le 1-\alpha. \end{aligned} \tag{30}$$

Luedtke and Ahmed [2008] considered a more general form of Problem (30) by replacing $1-\alpha$ with an arbitrary γ (e.g., γ could be 0). They studied how to use the SAA approach to obtain optimality bounds and feasible solutions for the original CCP in some probability sense. They also demonstrated that when $\gamma = 0$, the SAA is actually the scenario counterpart in the scenario approach. On the other hand, it is worthwhile to note that in the scenario approach (i.e., in Equation (26)), it is required that the constraint holds for all $j = 1, 2, \ldots, n$. If one requires it only for the $\lceil n\alpha \rceil$ order statistic, the resulting problem will coincide with Equation (30). Pagnoncelli et al. [2009] studied both theoretical and computational aspects of the SAA for CCPs. They showed that under some regularity conditions, the optimal value and the set of optimal solutions of Problem (30) converge to those of Problem (25), respectively, as n goes to infinity. Note that Problem (30) becomes a deterministic combinatorial optimization problem. To practically solve Problem (30), Luedtke and Ahmed [2008] and Pagnoncelli et al. [2009] suggested reformulating it into some mixed integer program (MIP) and then applying certain MIP techniques. Specifically, by introducing binary decision variables z_j, $j = 1, \ldots, n$, Problem (30) can be reformulated as the following MIP:

$$\begin{aligned} &\text{minimize} \quad h(x) \\ &\text{subject to} \quad L(x,\xi_j) \le M_j z_j, \ \ j = 1,\ldots,n, \\ &\qquad \frac{1}{n}\sum_{j=1}^{n} z_j \le 1-\alpha \\ &\qquad x \in X, z_j \in \{0,1\}, \ \ j = 1,\ldots,n, \end{aligned}$$

where M_j is a large positive number satisfying $M_j \geq \sup_{x \in X} L(x, \xi_j)$. Theoretically, to obtain a good approximation, the sample size n in the SAA should be set as large as possible. However, the resulting MIP may become excessively computationally difficult when n becomes large. In current implementations of the SAA, the sample size can only be moderate, for example, a few hundreds. How to efficiently solve the MIPs with a large sample size, say, $n = 10{,}000$, especially by incorporating special structures of the problems, is an interesting research topic.

4.1.3. DC Approach. Recently, Hong et al. [2011] took a different approach and proposed an ε-approximation approach to solving CCPs. Recall that $[a]^+ = \max\{0, a\}$. Hong et al. [2011] showed that

$$\inf_{t>0} \frac{1}{t}\{\mathrm{E}[[L(x,\xi)+t]^+] - \mathrm{E}[[L(x,\xi)]^+]\} = \Pr\{L(x,\xi) \geq 0\}$$

and suggested solving

$$\begin{aligned} &\underset{x \in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \quad \inf_{t>0} \frac{1}{t}\{\mathrm{E}[[L(x,\xi)+t]^+] - \mathrm{E}[[L(x,\xi)]^+]\} \leq 1-\alpha \end{aligned} \tag{31}$$

instead of the CCP. To handle Problem (31), Hong et al. [2011] suggested fixing $t = \varepsilon$ where $\varepsilon > 0$ is a small constant, and then solving the following ε-approximation:

$$\begin{aligned} &\underset{x \in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \quad \mathrm{E}[[L(x,\xi)+\varepsilon]^+] - (1-\alpha)\varepsilon - \mathrm{E}[[L(x,\xi)]^+] \leq 0. \end{aligned} \tag{32}$$

Note that by assuming convexity of h and L, the constraint function in Problem (32) is a DC function (i.e., a difference of two convex functions). Problem (32) is known as a DC program. To solve it, Hong et al. [2011] proposed a DC algorithm. They showed that the algorithm converges to the stationary points of the ε-approximation. They also showed that the optimal value, the set of optimal solutions, and the set of stationary points of the ε-approximation converge to those of the true CCP, respectively, as ε tends to 0 under a certain set of conditions. The DC approach can be viewed as a nonlinear optimization technique and enjoys nice properties. To practically solve the optimization problem in each iteration of the DC algorithm, Hong et al. [2011] used a gradient-based Monte Carlo method.

Although the ε-approximation approach has the desired convergence property, it also has some limitations. Especially, the ε-approximation approach was developed for smooth joint chance constrained programs (JCCPs). Note that in a JCCP, the chance constraint takes the following form:

$$\Pr\{L_i(x,\xi) \leq 0, i = 1, 2, \ldots, m\} \geq \alpha. \tag{33}$$

However, in many real applications, the smoothness of a JCCP could be violated very easily. To overcome such a difficulty, Hu et al. [2013b] suggested a smooth Monte Carlo (SMC) approach to solving general JCCPs. The basic idea of the SMC approach is to conservatively approximate the ε-approximation using the difference of two logarithm-sum-exponential functions. Specifically, it can be shown that for any $a_i, i = 1, 2, \ldots, m$ and $\mu > 0$,

$$[\max\{a_1, \ldots, a_m\}]^+ \leq \mu \log\left[1 + \sum_{i=1}^{m} \exp(\mu^{-1} a_i)\right] \leq [\max\{a_1, \ldots, a_m\}]^+ + \mu \log(m+1),$$

which together with the ε-approximation shows that

$$\mu \log\left[1+\sum_{i=1}^{m}\exp\{\mu^{-1}(L_i(x,\xi)+\varepsilon)\}\right] - (1-\alpha)\varepsilon - \mu\log\left[1+\sum_{i=1}^{m}\exp\{\mu^{-1}L_i(x,\xi)\}\right] + \mu\log(m+1) \le 0 \tag{34}$$

is a conservative approximation to Equation (33). Hu et al. [2013b] suggested solving constraint (34) instead of the original joint chance constraint (33). They further demonstrated that the parameter ε in Equation (34) can be relaxed as a decision variable. The SMC approach is capable of handling smooth and nonsmooth JCCPs where the distribution of ξ can be either continuous, discrete, or mixed.

4.2. CVaR Optimization

Parallel to VaR optimization formulations, a CVaR minimization problem takes the following expression:

$$\underset{x\in X}{\text{minimize}}\ c_\alpha(L(x,\xi)), \tag{35}$$

whereas a CVaR constrained program has the following form:

$$\begin{aligned}&\underset{x\in X}{\text{minimize}}\ \ h(x)\\ &\text{subject to}\ \ c_\alpha(L(x,\xi)) \le 0.\end{aligned} \tag{36}$$

Different from VaR, CVaR satisfies the subadditivity axiom and is a coherent risk measure; see Artzner et al. [1999], Rockafellar and Uryasev [2000], and Pflug [2000]. This implies that $c_\alpha(L(x,\xi))$ is convex in x if the loss function $L(x,\xi)$ is convex in x for every ξ. Therefore, CVaR optimization problems are typically technically significantly more tractable than VaR optimization problems. Indeed, there is a vast literature on CVaR optimization, which is still an evolving area of research. A full review of this topic is clearly beyond the scope of this article. In what follows, we review some popular approaches that rely on Monte Carlo simulations.

4.2.1. Linear Approach. Along with introducing the notion of CVaR, Rockafellar and Uryasev [2000] proposed a linear approach to solving CVaR minimization problems. This approach has significantly stimulated the use of CVaR in the field of decision under uncertainty. Subsequently, Andersson et al. [2001] generalized the linear approach to handle credit risk optimization problems with the CVaR objective. Krokhmal et al. [2002] extended the approach to solving portfolio optimization problems with the CVaR objective and constraints. More recently, Rockafellar and Royset [2010] used the linear approach to solve the buffered failure probability constrained problems in reliability engineering design.

One critical step of the linear approach of Rockafellar and Uryasev [2000] is to reformulate the CVaR minimization problem as a standard stochastic program. Specifically, by the stochastic program representation (2) of CVaR, Problem (35) is equivalent to the following optimization problem:

$$\underset{x\in X,\ t\in\Re}{\text{minimize}}\ \frac{1}{1-\alpha}\mathrm{E}[[L(x,\xi)+t]^+] - t. \tag{37}$$

Similarly, Problem (36) can be equivalently formulated as the following problem:

$$\begin{aligned}&\underset{x\in X,\ t\in\Re}{\text{minimize}}\ \ h(x)\\ &\text{subject to}\ \ \mathrm{E}[[L(x,\xi)+t]^+] - (1-\alpha)t \le 0.\end{aligned} \tag{38}$$

We review the linear approach from the perspective of implementing the SAA for the stochastic programs. Suppose that $\xi_1, \xi_2, \ldots, \xi_n$ are n i.i.d. observations from the random vector ξ. As mentioned in Section 4.1, SAA suggests using the sample mean $\frac{1}{n}\sum_{j=1}^{n}[L(x,\xi_j)+t]^+$ to estimate the expectation function $\mathrm{E}[[L(x,\xi)+t]^+]$. Consequently, Problems (37) and (38) can be approximated by

$$\underset{x\in X,\ t\in\Re}{\text{minimize}}\ \frac{1}{n}\sum_{j=1}^{n}[L(x,\xi_j)+t]^+ - (1-\alpha)t \tag{39}$$

and

$$\begin{aligned}\underset{x\in X,\ t\in\Re}{\text{minimize}}\ & h(x) \\ \text{subject to}\ & \frac{1}{n}\sum_{j=1}^{n}[L(x,\xi_j)+t]^+ - (1-\alpha)t \le 0,\end{aligned} \tag{40}$$

respectively. The rationale for solving Problems (39) and (40) instead of the original CVaR optimization problems is guaranteed by the theory of SAA (see, e.g., Shapiro et al. [2009]). In particular, it can be shown that, under some regularity conditions, the optimal values and the set of optimal solutions of Problems (39) and (40) converge to those of the true optimization problems, respectively, as the sample size n goes to infinity. Therefore, when selecting a reasonably large sample size, solving the SAA provides a good approximation to the true stochastic program.

By introducing a set of auxiliary decision variables z_j, $j = 0, 1, \ldots, n$, Problem (39) can be reformulated as

$$\begin{aligned}\text{minimize}\ & z_0 \\ \text{subject to}\ & L(x,\xi_j)+t \le z_j,\ \ j=1,\ldots,n, \\ & \frac{1}{n}\sum_{j=1}^{n} z_j - (1-\alpha)t - z_0 \le 0, \\ & x\in X, t\in\Re, z_0\in\Re, z_j\ge 0,\ \ j=1,\ldots,n.\end{aligned} \tag{41}$$

Similarly, introducing auxiliary decision variables z_j, $j = 1, 2, \ldots, n$, one can reformulate Problem (40) as the following problem:

$$\begin{aligned}\text{minimize}\ & h(x) \\ \text{subject to}\ & L(x,\xi_j)+t \le z_j,\ \ j=1,\ldots,n, \\ & \frac{1}{n}\sum_{j=1}^{n} z_j - (1-\alpha)t \le 0, \\ & x\in X, t\in\Re, z_j\ge 0,\ \ j=1,\ldots,n.\end{aligned} \tag{42}$$

When the set X is defined by linear constraints, the function $h(x)$ is linear, and $L(x,\xi)$ is linear in x for every $\xi \in \Xi$, Problems (41) and (42) are linear programs. That is why this approach is called the linear approach. Linear programs can be solved using standard optimization techniques such as simplex methods or interior point methods. When $h(x)$ and/or $L(x,\xi_j)$ are nonlinear but convex, the linear approach typically results in standard convex optimization problems, which can be solved using conventional convex optimization techniques.

The linear approach needs to introduce auxiliary variables, and both the decision variables and the number of constraints are proportional to the sample size n. Consequently, solving the resulting sample problems, that is, Problems (41) and (42), could

become quite slow, especially when n is large, say, $n \geq 10{,}000$. To partially address this issue, a dual-program approach was proposed when the sample problems are linear. Ogryczak and Śliwiński [2011] formulated the dual problem of the sample CVaR optimization problem of Rockafellar and Uryasev [2000] and found that the dual problem is much easier to solve. Specifically, they considered instances with 50,000 scenarios and 100 and 200 instruments and compared the computational efforts of solving the primal and dual problems. They found that the increase in efficiency is encouraging. In the linear approach, although there exist numerous constraints, a significant proportion of these constraints may be inactive for a specific solution searched. This structure was explored by Basova et al. [2011], who implemented an active-set method to solve Problem (41). The basic idea of the active-set method is to construct at each iteration a set of subscripts, of which the constraints at the currently searched solution are essentially tight, and solve the corresponding optimization problem, which only has constraints contained in the active set. Basova et al. [2011] compared the linear approach and the active-set method. They found that the active-set method is significantly faster than the linear approach for their reliability-based optimal design problems.

4.2.2. Other Approaches. Note that the operator $[\cdot]^+$ in Problem (40) makes the SAAs nonsmooth. To remove the nonsmoothness, the linear approach introduces extra decision variables with the price of increasing the dimension of decision vector and the scale of the problem. An alternative approach is to treat the SAA as a nonsmooth optimization problem and implement some nonsmooth optimization techniques to solve it. Lim et al. [2010] proposed a three-phase subgradient approach to solve Problem (40). They showed that the approach is computationally much more efficient than the linear approach for a set of test examples. Another idea for removing the nonsmoothness is to smooth the maximum operator [Alexander et al. 2006; Xu and Zhang 2009]. Alexander et al. [2006] compared the smoothing method and the linear approach and found that the smoothing method is much more efficient for their set of test problems. Iyengar and Ma [2013] suggested a fast gradient descent method to solving the linear portfolio CVaR optimization problem. The idea of Iyengar and Ma [2013] is to reformulate their problem as a min-max problem using a dual representation of CVaR and then use a fast iterative procedure developed in Nesterov [2005] to solve the resulting nonsmooth problem.

Similar to the VaR optimization, we can view the CVaR optimization problems as standard nonlinear optimization problems. When the optimization problems are smooth, we can also solve them using a gradient-based approach. Hong and Liu [2009] studied the gradients of CVaR and showed how to use the gradient estimates to solve CVaR constrained optimization problems. They compared their approach with the linear approach and showed that the gradient-based method is at least an order of magnitude faster than the linear approach for the portfolio optimization problem that they considered. However, it is worthwhile to note that the gradient-based method requires more conditions on the loss function $L(x, \xi)$ and the distribution of the random vector ξ to ensure the differentiability of the CVaR function, while other approaches, such as the linear approach, can often handle more general types of CVaR optimization problems.

4.2.3. Links Between VaR and CVaR Optimizations. We have briefly reviewed optimization problems of VaR and CVaR. Besides being considered as popular decision models, CVaR optimization problems are also often used as approximations of VaR optimization problems. From the definitions, it is clear that CVaR is an upper bound of VaR. Therefore, minimizing CVaR can often help achieve some small value of VaR. This idea is due to Rockafellar and Uryasev [2000]. Also, the bounding relation between VaR and CVaR shows that Problem (36) is a conservative approximation of Problem (23). In

the literature, Problem (36) is called *CVaR approximation* to Problem (23). Nemirovski and Shapiro [2006] demonstrated that the CVaR approximation is the best convex conservative approximation of the VaR constrained program under their generating function scheme.

Recently, Hong et al. [2012a] discovered some interesting connections between VaR constrained programs and their CVaR approximations. They showed that a VaR constrained program may be solved by solving a sequence of CVaR-like approximations. We briefly review their approach in the following. Consider Problem (31), which is a reformulation of Problem (23). The DC structure in Problem (31) is the source of nonconvexity. Thus, a natural idea is to convexify the DC constraint. To do this, we can find some concave function g satisfying $g(x) \le \mathrm{E}[[L(x,\xi)]^+]$ and then use g to approximate $\mathrm{E}[[L(x,\xi)]^+]$. In particular, replacing $\mathrm{E}[[L(x,\xi)]^+]$ with 0 in Problem (31), one can obtain the following problem:

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \ \inf_{t>0} \frac{1}{t}\mathrm{E}[[L(x,\xi)+t]^+] \le \alpha. \end{aligned} \tag{43}$$

It can be shown that under some minor conditions Problem (43) is equivalent to the CVaR approximation, that is, Problem (36). Therefore, roughly speaking, the CVaR approximation loses a term $\mathrm{E}[[L(x,\xi)]^+]$ and this term determines the conservatism of the CVaR approximation. However, using 0 to approximate $\mathrm{E}[[L(x,\xi)]^+]$ may be quite conservative. It is desirable to have other less conservative approximations. Suppose a feasible solution $y \in X$ of Problem (31) has been obtained. For instance, y could be an optimal solution of the CVaR approximation. Suppose y satisfies $\Pr\{L(y,\xi) > 0\} < 1-\alpha$. Let

$$C(x,y) = \mathrm{E}[[L(y,\xi)]^+] + \nabla_x \mathrm{E}[[L(y,\xi)]^+]^{\mathrm{T}}(x-y)$$

be the first-order Taylor approximation of $\mathrm{E}[[L(x,\xi)]^+]$ at point y. Then, $C(x,y) \le \mathrm{E}[[L(x,\xi)]^+]$. We can then convexify the DC function in Problem (31) using $C(x,y)$ and propose solving the following problem:

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \ \inf_{t>0} \frac{1}{t}\{\mathrm{E}[[L(x,\xi)+t]^+] - C(x,y)\} \le \alpha. \end{aligned} \tag{44}$$

It can be shown that under some conditions, Problem (44) is equivalent to

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to} \ \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) - \alpha^{-1}C(x,y) \le 0. \end{aligned} \tag{45}$$

Note that $C(x,y)$ is affine in x. Problem (45) takes a similar expression as the CVaR approximation. It is called a *CVaR-like approximation* in Hong et al. [2012a]. Because y is a feasible solution of Problem (45), by solving Problem (45), one can obtain an improved solution. Once an improved solution is obtained, one can construct another CVaR-like approximation. This suggests that we can solve a VaR constrained program by iteratively solving CVaR-like approximations. Based on this, Hong et al. [2012a] proposed some procedure for VaR constrained programs and studied properties and performances of such procedure.

4.3. More Discussions on Optimization

Optimization problems involving VaR/CVaR are typically difficult to solve. As discussed in this article, we often have to resort to sampling-based techniques. When the sample

size is small, the sampling approximations may not be accurate. On the other hand, when the sample size is large, solving the sampling problems may become computationally expensive. Deriving a promising tradeoff is often problem dependent and thus needs dedicated investigation. In preceding sections, we have discussed the asymptotic properties for various estimates of values and derivatives of VaR and CVaR. It would be desirable to build asymptotic results (e.g., central limit theorems) for the optimal solutions and optimal values in VaR/CVaR optimization problems when using sampling-based procedures. Such asymptotic analysis can provide certain guides for setting the sample size. Also, to improve the efficiency of the sampling-based procedures, it is good to implement variance reduction techniques in the optimization algorithms. In the stochastic optimization literature, asymptotic analysis for expectation-based stochastic programs has been studied for a long time (see, e.g., King and Rockafellar [1993] and Shapiro [1993]). There also exists much study on the application of variance reduction techniques for expectation-based stochastic programs. The recent paper by Homem-de-Mello and Bayraksan [2013] provided an excellent review on Monte Carlo sampling-based stochastic optimization. Readers can refer to it, especially Section 2 and Section 7 therein, for a detailed introduction of these topics. However, VaR and CVaR functions are often more complicated and more difficult to handle than an expectation function. Therefore, building asymptotic results and applying variance reduction techniques for VaR and CVaR optimization should be more challenging. These could be interesting future research topics.

5. CONCLUSIONS AND FURTHER DISCUSSIONS

This article provides a unified view of the simulation of VaR, CVaR, and their sensitivities. It also gives a brief review on VaR and CVaR optimization. These topics are inherently related and are important content of financial risk management. We believe the methodologies and techniques covered in this article are very important for financial risk management practice.

However, the context of this article is far from sufficient for the practice of risk management. In this article, we have mainly focused on research for dealing with VaR and CVaR. We did not study in depth the properties of VaR and CVaR risk measures. Every risk measure has its properties, advantages, and disadvantages. Understanding these properties is important and could be beneficial from a risk management perspective. For instance, one important feature of using VaR optimization is that the model may result in very skewed loss distribution, and consequently, the risk may hide in the tail of the distribution (see, e.g., Natarajan et al. [2008]). This issue is very important for risk management practice. Similarly, we think using the CVaR optimization model may also bring in important issues. For instance, Lim et al. [2011] showed that CVaR is fragile in portfolio optimization; that is, estimation errors in CVaR may affect optimization results and thus decisions significantly. Also, we did not include any empirical study on VaR and CVaR, which is very important. It is of great meaning to analyze VaR/CVaR-based models and to study the pros and cons of these models in practice using data and information available.

Another important theoretical question is the specification of distributions of random variables in risk management models. In the context of this article, we have assumed that an input distribution is predetermined and is given to modelers. However, in practice, it is often difficult to specify the input distribution precisely. A considerable amount of research has been devoted to the issue of uncertainty in models of VaR/CVaR (see, e.g., El Ghaoui et al. [2003], Zymler et al. [2013], Hu and Hong [2012], Hu et al. [2013a], and many others). However, it is far from sufficient and more study on input uncertainty is necessary in the context of financial risk management. Modeling input uncertainty should incorporate information available and should reflect the practice.

ACKNOWLEDGMENTS

The authors would like to thank Pierre L'Ecuyer, the associate editor, and two anonymous referees for their valuable comments and suggestions that significantly improved the article.

REFERENCES

C. Acerbi and D. Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26, 1487–1503.

S. Alexander, T. F. Coleman, and Y. Li. 2006. Minimizing CVaR and VaR for a portfolio of derivatives. *Journal of Banking and Finance* 30, 583–605.

F. Andersson, H. Mausser, D. Rosen, and S. Uryasev. 2001. Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming* 89, 273–291.

B. Ankenman, B. L. Nelson, and J. Staum. 2010. Stochastic kriging for simulation metamodeling. *Operations Research* 58, 2, 371–382.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9, 3, 203–228.

S. Asmussen and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.

A. N. Avramidis and J. R. Wilson. 1998. Correlation-induction techniques for estimating quantiles in simulation experiments. *Operations Research* 46, 574–591.

R. Bahadur. 1966. A note on quantiles in large samples. *Annals of Mathematical Statistics*, 37, 577–580.

O. A. Bardou, N. Frikha, and G. Pagès. 2009. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications* 15, 3, 173–210.

Basel Committee of Banking Supervision of the Bank for International Settlements. 2012. Fundamental review of the trading book. *Consultative Document*.

H. G. Basova, R. T. Rockafellar, and J. O. Royset. 2011. A computational study of the buffered failure probability in reliability-based design optimization. In *Proceedings of the 11th Conference on Application of Statistics and Probability in Civil Engineering*.

A. Bassamboo, S. Juneja, and A. Zeevi. 2008. Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research* 56, 593–606.

D. P. Bertsekas. 2007. *Dynamic Programming and Optimal Control* (Vol. 2). Athena Scientific, Belmont, MA.

D. Bosq. 1998. *Nonparametric Statistics for Stochastic Processes*. New York: Springer.

M. Broadie, Y. Du, and C. C. Moallemi. 2010. Efficient risk estimation via nested sequential simulation. *Management Science*, 57, 1172–1194.

M. Broadie, Y. Du, and C. C. Moallemi. 2011. Risk estimation via regression. Working paper, Graduate School of Business, Columbia University.

M. Broadie and P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Science*, 42, 269–285.

G. Calafiore and M. C. Campi. 2005. Uncertain convex programs: Randomized solutions and confidence levels. *Mathematical Programming* 102, 25–46.

G. Calafiore and M. C. Campi. 2006. The scenario approach to robust control design. *IEEE Transactions on Automatic Control* 51, 742–753.

J. F. Carrière. 1996. Valuation of early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics* 19, 19–30.

A. Charnes, W. W. Cooper, and G. H. Symonds. 1958. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science* 4, 235–263.

F. Chu and M. K. Nakayama. 2012. Confidence intervals for quantiles when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation* 22, 2, Article 10.

S. Deng, K. Giesecke, and T. L. Lai. 2012. Sequential importance sampling and resampling for dynamic portfolio credit risk. *Operations Research* 60, 78–91.

D. P. De Farias and B. Van Roy. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research* 29, 3, 462–478.

D. Duffie. 2001. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, NJ.

L. El Ghaoui, M. Oks, and F. Oustry. 2003. Worst case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research* 51, 543–556.

M. C. Fu, L. J. Hong, and J.-Q. Hu. 2009. Conditional Monte Carlo estimation of quantile sensitivities. *Management Science* 55, 2019–2027.

M. C. Fu and J.-Q. Hu. 1997. *Conditional Monte Carlo, Gradient Estimation and Optimization Applications*. Kluwer Academic Publishers, Boston.

P. Glasserman. 1991. *Gradient Estimation via Perturbation Analysis*. Kluwer Academic Publishers, Boston, MA.

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

P. Glasserman. 2005. Measuring marginal risk contributions of credit portfolios. *Journal of Computational Finance* 9, 1–41.

P. Glasserman. 2008. Calculating portfolio credit risk. In *Financial Engineering: Handbooks in Operations Research and Management Science*, J. Birge and V. Linetsky (Eds.). Vol. 15. Elsevier, North-Holland, The Netherlands.

P. Glasserman, P. Heidelberger, and P. Shahabuddin. 1999. Importance sampling and stratification for Value-at-Risk. *Computational Finance 1999*, Y. S. Abu-Mostafa, B. Le Baron, A. W. Lo, and A. S. Weigend (Eds.). MIT Press, Cambridge, MA.

P. Glasserman, P. Heidelberger, and P. Shahabuddin. 2000. Variance reduction techniques for estimating value-at-risk. *Management Science* 46, 1349–1364.

P. Glasserman, P. Heidelberger, and P. Shahabuddin. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12, 239–269.

P. Glasserman, W. Kang, and P. Shahabuddin. 2007. Large deviations in multifactor portfolio credit risk. *Mathematical Finance* 17, 345–379.

P. Glasserman, W. Kang, and P. Shahabuddin. 2008. Fast simulation of multifactor portfolio credit risk. *Operations Research* 56, 1200–1217.

P. Glasserman and J. Li. 2005. Importance sampling for portfolio credit risk. *Management Science* 51, 1643–1656.

P. W. Glynn. 1996. Importance sampling for Monte Carlo estimation of quantiles. In *Proceedings of 1996 Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design, St. Petersburg, Russia,*180–185.

P. W. Glynn and D. L. Iglehart. 1989. Importance sampling for stochastic simulation. *Management Science* 35, 1367–1392.

M. Gordy and S. Juneja. 2010. Nested simulation in portfolio risk measurement. *Management Science* 56(10), 1833–1848.

G. Gupton, C. Finger, and M. Bhatia. 1997. CreditMetrics technical document. New York: J. P. Morgan & Co.

B. Heidergott and W. Volk-Makarewicz. 2012. Sensitivity analysis of quantiles. Working paper.

T. C. Hesterberg and B. L. Nelson. 1998. Control variates for probability and quantile estimation. *Management Science* 44, 1295–1312.

Y. C. Ho and X. R. Cao. 1983. Perturbation analysis and optimization of queueing networks. *Journal of Optimization Theory and Applications* 40, 4, 559–582.

T. Homem-de-Mello and G. Bayraksan. 2013. Monte Carlo sampling-based methods for stochastic optimization. Available at http://www.optimization-online.org/DB_FILE/2013/06/3920.pdf.

L. J. Hong. 2009. Estimating quantile sensitivities. *Operations Research* 57, 118–130.

L. J. Hong and G. Liu. 2009. Simulating sensitivities of conditional value-at-risk. *Management Science* 55, 281–293.

L. J. Hong and G. Liu. 2010. Pathwise estimation of probability sensitivities through terminating or steady-state simulations. *Operations Research* 58, 357–370.

L. J. Hong, Y. Yang, and L. Zhang. 2011. Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Operations Research* 59, 617–630.

L. J. Hong, Z. Hu, and L. Zhang. 2012a. Conditional value-at-risk approximation to value-at-risk constrained programs: A remedy via Monte Carlo. *INFORMS Journal on Computing* 26, 2, 385–400.

L. J. Hong, S. Juneja, and G. Liu. 2012b. Kernel smoothing for nested estimation with application to portfolio risk measurement. Working paper.

J. C. Hsu and B. L. Nelson. 1990. Control variates for quantile estimation. *Management Science* 36, 835–851.

Z. Hu and L. J. Hong. 2012. Kullback-Leibler divergence constrained distributionally robust optimization. Available at http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf.

Z. Hu, L. J. Hong, and A. M. C. So. 2013a. Distributionally robust probabilistic programs. Working paper. Available at http://www.optimization-online.org/DB_FILE/2013/09/4039.pdf.

Z. Hu, L. J. Hong, and L. Zhang. 2013b. A smooth Monte Carlo approach to joint chance constrained program. *IIE Transactions* 45, 7, 716–735.

G. Iyengar and A. K. C. Ma. 2013. Fast gradient descent method for mean-CVaR optimization. *Annals of Operations Research* 205, 203–212.

P. Jorion. 2006. *Value at Risk*, (2nd ed.). McGraw-Hill, New York.

P. Jorion. 2010. *Financial Risk Manager Handbook* (6th ed.). Wiley, New York.

M. Kalkbrener. 2003. An axiomatic approach to capital allocation. *Mathematical Finance* 15, 425–437.

M. Kalkbrener, H. Lotter, and L. Overbeck. 2004. Sensible and efficient capital allocation for credit portfolios. *Risk* 17, 19–24.

A. J. King and R. T. Rockafellar. 1993. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research* 18, 1, 148–162.

S. G. Kou, X. Peng, and C. C. Heyde. 2013. External risk measures and Basel accords. *Mathematics of Operations Research* 38, 3, 393–417.

P. Krokhmal, J. Palmquist, and S. Uryasev. 2002. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* 4, 11–27.

A. Kurth and D. Tasche. 2003. Contributions to credit risk. *Risk* 16, 84–88.

P. L'Ecuyer. 1990. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science* 36, 11, 1364–1383.

H. Lan, B. L. Nelson, and J. Staum. 2010. A confidence interval procedure for expected shortfall risk measurement via two-level simulation. *Operations Research* 58, 5, 1481–1490.

A. M. Law and W. D. Kelton. 2000. *Simulation Modeling & Analysis* (3rd ed.). McGraw-Hill, New York.

S. H. Lee. 1998. Monte Carlo expectation of conditional quantiles. Ph.D. Thesis, Department of Operations Research, Stanford University.

D. Li. 2000. On default correlation: A copula function approach. *Journal of Fixed Income* 9, 43–54.

C. Lim, H. D. Sherali, and S. Uryasev. 2010. Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization. *Computational Optimization and Applications* 46, 391–415.

A. E. Lim, J. G. Shanthikumar, and G. Y. Vahn. 2011. Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Operations Research Letters* 39, 3, 163–171.

G. Liu. 2012. Simulating risk contributions of credit portfolios. Working paper, Department of Management Sciences, City University of Hong Kong.

G. Liu and L. J. Hong. 2009. Kernel estimation of quantile sensitivities. *Naval Research Logistics* 56, 511–525.

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

F. A. Longstaff and S. Schwartz. 2001. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies* 14, 113–147.

J. Luedtke and S. Ahmed. 2008. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization* 19, 2, 674–699.

A. J. McNeil, R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.

L. B. Miller and H. Wagner. 1965. Chance-constrained programming with joint constraints. *Operations Research* 13, 930–945.

K. Natarajan, D. Pachamanova, and M. Sim. 2008. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science* 54, 3, 573–585.

A. Nemirovski and A. Shapiro. 2005. Scenario approximation of chance constraints. *Probabilistic and Randomized Methods for Design Under Uncertainty*, G. Calafiore and F. Dabbene (Eds.). Springer, London, 3–48.

A. Nemirovski and A. Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* 17, 969–996.

Y. Nesterov. 2005. Smooth minimization of nonsmooth functions. *Mathematical Programming* 103, 127–152.

W. Ogryczak and T. Śliwiński. 2011. On solving the dual for portfolio selection by optimizing conditional value at risk. *Computational Optimization and Applications* 50, 591–595.

B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. 2009. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications* 142, 399–416.

G. Pflug. 2000. Some remarks on the value-at-risk and conditional value-at-risk. In *Probabilistic Constrained Optimization: Methodology and Applications,* S. Uryasev (Ed.). Kluwer, Dordrecht, The Netherlands, 272–281.

A. Prékopa. 1970. On probabilistic constrained programming. *Proceedings of the Princeton Symposium on Mathematical Programming*, 113–138.

R. T. Rockafellar and J. O. Royset. 2010. On buffered failure probability in design and optimization of structures. *Reliability Engineering & System Safety* 95, 499–510.

R. T. Rockafellar and S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk*, 2, 21–41.

R. T. Rockafellar and S. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26, 1443–1471.

O. Scaillet. 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14, 115–129.

R. J. Serfling. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

A. Shapiro. 1993. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research* 18, 4, 829–845.

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

L. Sun and L. J. Hong. 2010. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters* 38, 246–251.

D. Tasche. 2009. Capital allocation for credit portfolios with kernel estimators. *Quantitative Finance* 9, 581–595.

A. A. Trindade, S. Uryasev, A. Shapiro, and G. Zrazhevsky. 2007. Financial prediction with constrained tail risk. *Journal of Banking and Finance* 31, 3524–3538.

J. N. Tsitsiklis and B. van Roy. 2001. Regression method for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12, 694–703.

H. Xu and D. Zhang. 2009. Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Mathematical Programming* 119, 371–401.

S. Zymler, D. Kuhn, and B. Rustem. 2013. Worst-case value-at-risk of nonlinear portfolios. *Management Science*, 59(1), 172–188.

Received January 2013; revised August 2013; accepted August 2014