MONTE CARLO ESTIMATION OF VALUE-AT-RISK, CONDITIONAL VALUE-AT-RISK AND THEIR SENSITIVITIES

L. Jeff Hong

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Guangwu Liu

Department of Management Sciences
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong

ABSTRACT

Value-at-risk and conditional value at risk are two widely used risk measures, employed in the financial industry for risk management purposes. This tutorial discusses Monte Carlo methods for estimating value-at-risk, conditional value-at-risk and their sensitivities. By relating the mathematical representation of value-at-risk to that of conditional value-at-risk, it provides a unified view of simulation methodologies for both risk measures and their sensitivities.

1 INTRODUCTION

Value-at-risk (VaR) and conditional value-at-risk (CVaR) are two widely used risk measures, employed in the financial industry for risk management purposes. For instance, VaR is used by bank regulators to set capital requirements for the banks, with the aim to ensure that banks keep enough capital to serve as a cushion against potential losses (Jorion 1997). By definition, 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 $\beta \in (\alpha, 1)$. Both VaR and CVaR provide information on the magnitude of the potential large losses that the investor may suffer.

The sensitivities of VaR and CVaR concern the derivatives of these risk measures with respect to a parameter associated with the random loss L. They measure how changes in the parameter will affect the levels of VaR and CVaR, and have a number of important applications in financial industry. In portfolio optimization problems, for instance, sensitivities of VaR (CVaR) can be incorporated into a gradient-based optimization algorithms to solve problems with VaR (CVaR) objective and/or constraints (Rockafellar and Uryasev 2000). 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).

Monte Carlo simulation has been widely used as an effective tool for calculating VaR, CVaR and their sensitivities, given that the loss models in practical use are often complicated and analytical formulas of these quantities may not be available. Over the past two decades, the problem of simulating VaR, CVaR and their sensitivities has received much attention from the simulation community. Given the existence of a large amount of work in this area, in this tutorial we do not attempt to provide a thorough and comprehensive

review of the related literature, but instead, we provide a unified view of the simulation methodologies for estimating these quantities.

The rest of the paper is organized as follows. Section 2 considers the estimation of VaR and CVaR, and discusses asymptotic properties of the estimators and variance reduction issues. 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 concludes.

2 VALUE-AT-RISK AND CONDITIONAL VALUE-AT-RISK

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) shows that c_α can also be written as the following 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). Trindade et al. (2007) show that $T = [v_\alpha, u_\alpha]$ where $u_\alpha = \sup\{t : F(t) \le \alpha\}$. 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 Equation (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.

2.1 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 $L_{i:n}$ is the ith order statistic from the n observations.

Trindade et al. (2007) suggest 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

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

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

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

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

2.2 Asymptotic Properties

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 c_α^n by using the Bahadur representations. To this end, we first make the following assumption.

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

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 is omitted here.

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

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

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

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 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 $n \to \infty$. Furthermore, if $\mathrm{E}\left(L^2\right) < \infty$, then

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

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

as $n \to \infty$, where "$\Rightarrow$" denotes "converge in distribution", $N(0,1)$ represents the standard normal random variable, and

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

2.3 Variance Reduction

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

Among various variance reduction methods proposed in the literature, importance sampling (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 this tutorial we discuss importance sampling in more details, with a focus on the asymptotic properties of the IS estimators. Specifically, suppose that L is simulated under another distribution function $G(\cdot)$, where F is absolutely continuous with respect to G, i.e., $F(dx) = 0$ if $G(dx) = 0$ for any $x \in \mathfrak{R}$. We refer to G as the IS distribution and let $S(x) = F(dx)/G(dx)$ denote the likelihood ratio function (also called score function) associated with the change of measure. Note that

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

where E_F and E_G denote taking expectations with respect to F and G respectively. Then for any $x \in \mathfrak{R}$, we may estimate $F(x)$ by

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

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

$$\hat{v}_\alpha^{n,IS} = F_{n,IS}^{-1}(\alpha) = \inf\{x : F_{n,IS} \ge \alpha\}, \quad \text{and}$$

$$\hat{c}_\alpha^{n,IS} = \hat{v}_\alpha^{n,IS} + \frac{1}{1-\alpha}\sum_{i=1}^{n}(L_i - \hat{v}_\alpha^{n,IS})^+ S(L_i).$$

Recently, Sun and Hong (2010) and Chu and Nakayama (2011) independently study 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 $\varepsilon > 0$ and $C > 0$ such that $S(x) < C$ for any $x \in (v_\alpha - \varepsilon, v_\alpha + \varepsilon)$ and there exists $p > 2$ such that $\mathrm{E}_G\left[1_{\{L \ge v_\alpha - \varepsilon\}}S^p(L)\right] < \infty$.

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

Define the notation $o_p(\cdot)$ such that a random variable $U_n = o_p(g(n))$ means $U_n/g(n) \to 0$ in probability as $n \to \infty$. Then 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 For a fixed $\alpha \in (0,1)$, suppose that Assumptions 1 and 2 are satisfied. Then,

$$\hat{v}_\alpha^{n,IS} = v_\alpha - \frac{1}{f(v_\alpha)}\left((1-\alpha) - \frac{1}{n}\sum_{i=1}^{n}1_{\{L_i \ge v_\alpha\}}S(L_i)\right) + C_n, \quad \text{and}$$

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

where $C_n = O_{a.s.}(t(n,\delta))$ with $t(n,\delta) = \max\{n^{-1+2/p+\delta}, n^{-3/4+1/(2p)+\delta}\}$ and $D_n = O_{a.s.}(n^{-1+2/p+\delta})$ for any $\delta > 0$, $C_n = o_p(n^{-1/2})$, and $D_n = o_p(n^{-1/2})$. Furthermore, if $S(x) \le C$ for all $x \in (v_\alpha - \varepsilon, \infty)$, then $C_n = O_{a.s.}(n^{-3/4}(\log n)^{3/4})$ and $D_n = O_{a.s.}(n^{-1}\log n)$.

Asymptotic normality of the estimators follows immediately from Theorem 2. In particular,

$$\sqrt{n}\left(\hat{v}_\alpha^{n,IS} - v_\alpha\right) \Rightarrow \frac{\sqrt{\text{Var}_G\left[1_{\{L \geq v_\alpha\}}S(L)\right]}}{f(v_\alpha)}N(0,1),$$

and

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

as $n \to \infty$.

If $S(x) < 1$ for all $x \geq v_\alpha$, then it can be easily verified that $\text{Var}_G\left[1_{\{L \geq v_\alpha\}}S(L)\right] < \alpha(1 - \alpha)$ and $\text{Var}_G\left[(L - v_\alpha)^+S(L)\right] < \text{Var}\left[(L - v_\alpha)^+\right]$. Then compared to Equations (7) and (8), it can be seen that the asymptotic variances of the IS estimators will be smaller than those of the estimators without IS, given a sufficient condition that $S(x) < 1$ for all $x \geq v_\alpha$. In practice, an effective IS distribution (with a density function g) often satisfies $g(x) > f(x)$ for $x > v_\alpha$, which implies the sufficient condition. This provides a guideline for selecting an appropriate IS distribution in practical implementations.

3 SENSITIVITIES OF VALUE-AT-RISK AND CONDITIONAL VALUE-AT-RISK

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 tutorial 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, and assume that $L'(\theta)$ can be evaluated for all $\theta \in \Theta$.

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 will also be discussed.

3.1 Closed-Form Expressions

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

$$F(v_\alpha(\theta), \theta) = \alpha. \tag{9}$$

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

Since a probability distribution function is simply an expectation and is easier to analyze, we start by analyzing probability sensitivity. We let $p_y(\theta) = \Pr\{L(\theta) \leq y\}$ be the probability function for some threshold y. 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 $\text{E}(K) < \infty$ and $|L(\theta + \Delta\theta) - L(\theta)| \leq K|\Delta\theta|$ 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., Broadie and Glasserman (1996). It guarantees the validity of interchanging differentiation and expectation when evaluating $d\text{E}(r(L(\theta)))/d\theta$ for any Lipschitz-continuous function $r(\cdot)$. Assumption 4 requires that $F(t, \theta)$ is continuously differentiable at (y, θ). 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) shows the following result.

Theorem 3 Suppose that Assumptions 3 and 4 are satisfied. Then,

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

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

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

Theorem 3 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, we may easily arrive at a closed-form expression of VaR sensitivity. The key is to use relation in (9). Specifically, by differentiating with respect to θ on both sides of Equation (9), we have

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

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

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

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

Theorem 4 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 4 shows that under appropriate regularity conditions, VaR sensitivity is an expectation taken in the set defined by $\{L(\theta) = v_\alpha\}$. Because the set is often a probability-zero set in many practical problems, the representation in Theorem 4 may not be directly applied for estimating VaR sensitivity. However, it does tell us a fact that only samples that lies in a neighborhood of the set play crucial roles in estimation.

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

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

The basic idea is that $h(x) = x^+$ is a Lipschitz continuous function, and hence the interchange of expectation and differentiation is valid under some mild regularity conditions. In particular, the closed-form expression of CVaR sensitivity is presented in the following theorem.

Theorem 5 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 $\theta \in \Theta$,

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

Proof. To prove the theorem, we prove that $c'_\alpha(\theta^*) = \mathrm{E}\left[L'(\theta^*)|L(\theta^*) \geq v_\alpha(\theta^*)\right]$ for any $\theta^* \in \Theta$.
Note that $L(\theta)$ has a density at $v_\alpha(\theta)$, then for any $\theta \in \Theta$,

$$
\begin{aligned}
c_\alpha(\theta) &= \mathrm{E}[L(\theta)|L(\theta) \geq v_\alpha(\theta)] = \frac{1}{1-\alpha}\mathrm{E}\left[L(\theta) \cdot 1_{\{L(\theta) \geq v_\alpha(\theta)\}}\right] \\
&= \frac{1}{1-\alpha}\mathrm{E}\left\{[L(\theta) - v_\alpha(\theta)] \cdot 1_{\{L(\theta) - v_\alpha(\theta) \geq 0\}}\right\} + v_\alpha(\theta). \quad (10)
\end{aligned}
$$

For any $\theta^* \in \Theta$, because $v_\alpha(\theta)$ is differentiable at θ^*, there exist a neighborhood of θ^*, denoted as $(a_{\theta^*}, b_{\theta^*})$, and a constant κ_{θ^*} such that $\theta^* \in (a_{\theta^*}, b_{\theta^*})$, $(a_{\theta^*}, b_{\theta^*}) \subset \Theta$, and $|v_\alpha(\theta_1) - v_\alpha(\theta_2)| \leq \kappa_{\theta^*}|\theta_1 - \theta_2|$ for all $\theta_1, \theta_2 \in (a_{\theta^*}, b_{\theta^*})$. Then by Assumption 3,

$$|[L(\theta_2) - v_\alpha(\theta_2)] - [L(\theta_1) - v_\alpha(\theta_1)]| \leq (K + \kappa_{\theta^*}) \cdot |\theta_2 - \theta_1|$$

for all $\theta_1, \theta_2 \in (a_{\theta^*}, b_{\theta^*})$, and $E(K + \kappa_{\theta^*}) = E(K) + \kappa_{\theta^*} < \infty$. Note that $h(x) = x \cdot 1_{\{x \geq 0\}}$ is a Lipschitz continuous function and $h'(x) = 1_{\{x \geq 0\}}$ when $x \neq 0$. Because $\Pr\{L(\theta) - v_\alpha(\theta) \neq 0\} = 1$, then applying the pathwise method at every $\theta \in (a_{\theta^*}, b_{\theta^*})$

$$\frac{d}{d\theta} E\left\{[L(\theta) - v_\alpha(\theta)] \cdot 1_{\{L(\theta) - v_\alpha(\theta) \geq 0\}}\right\} = \frac{d}{d\theta} E[f(L(\theta) - v_\alpha(\theta))] = E\left\{[L'(\theta) - v'_\alpha(\theta)] \cdot 1_{\{L(\theta) - v_\alpha(\theta) \geq 0\}}\right\}.$$

Then by Equation (10), at every $\theta \in (a_{\theta^*}, b_{\theta^*})$,

$$
\begin{aligned}
c'_\alpha(\theta) &= \frac{1}{1-\alpha} E\left\{[L'(\theta) - v'_\alpha(\theta)] \cdot 1_{\{L(\theta) - v_\alpha(\theta) \geq 0\}}\right\} + v'_\alpha(\theta) \\
&= \frac{1}{1-\alpha} E\left[L'(\theta) \cdot 1_{\{L(\theta) \geq v_\alpha(\theta)\}}\right] = E\left[L'(\theta) \big| L(\theta) \geq v_\alpha(\theta)\right].
\end{aligned}
$$

Since $\theta^* \in (a_{\theta^*}, b_{\theta^*})$, then $c'_\alpha(\theta^*) = E[L'(\theta^*)|L(\theta^*) \geq v_\alpha(\theta^*)]$.

This concludes the proof of the theorem. □

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) proposes 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_m^b. Then, we have k observations of D_m^b, denoted as $D_{m1}^b, \ldots, D_{m,k}^b$. A batch-mean estimator of $v'_\alpha(\theta)$ is then proposed:

$$\bar{D}_{mk} = \frac{1}{k} \sum_{l=1}^{k} D_{ml}^b. \tag{11}$$

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 $\mathfrak{R}$. 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.

$$\bar{V}_n = \frac{\sum_{i=1}^{n} D_i K\left(\frac{\hat{v}_\alpha^n - L_i}{\delta_n}\right)}{\sum_{i=1}^{n} K\left(\frac{\hat{v}_\alpha^n - L_i}{\delta_n}\right)}. \tag{12}$$

Intuitively, the kernel estimator in (12) can be viewed as a weighted average of L_i's. The weight depends on the distance of L_i to v_α. If the distance is small, typically a larger weight will be assigned. In other words, the observation 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) propose the following estimator of $c'_\alpha(\theta)$.

$$\bar{C}_n = \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} D_i 1_{\{L_i \geq \hat{v}_\alpha^n\}}. \tag{13}$$

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.

3.3 Asymptotic Properties

In this subsection we outline the asymptotic results, mainly consistency and asymptotic normality, of the estimators discussed in the previous subsection. We do not go deep into the technical conditions which are imposed in the asymptotic analysis. But instead, we aim to provide a sketch of the major results.

For the batch-mean estimator in (11), if $m \to \infty$ and $k \to \infty$ as $n \to \infty$, Hong (2009) proves its consistency under some regularity conditions. Furthermore, if $\lim_{n\to\infty} \sqrt{k}/m = 0$, then it can be shown that

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

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 (12), Liu and Hong (2009) show 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,

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

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 function, 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 is slower than $n^{-1/2}$, which is a typical rate of convergence of Monte Carlo simulation. From this perspective, there is still room for improvement. In Section 3.4, we will discuss in more details how to accelerate the rate of convergence.

The form of the CVaR sensitivity estimator in (13) 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) show that under some regularity conditions,

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

where

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

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

3.4 Accelerating the Rate of Convergence

In this subsection, we consider an issue related to the estimation of VaR sensitivity. As shown in the previous subsection, a kernel estimator of VaR sensitivity may not achieve the the $n^{-1/2}$ rate of convergence, which is the typical rate of convergence for statistical estimation. To improve the efficiency of the estimation, 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 this tutorial, we present a simulation framework of accelerating the rate of convergence by using two classical simulation techniques, importance sampling 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 importance sampling (IS) and conditional Monte Carlo in this context.

3.4.1 Importance Sampling

Before the presentation, we want to first clarify a slight difference between the IS we employ here and the classical use of IS 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. First, v_α is unknown and thus has to be estimated. Second, 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, e.g., the sample quantile $\hat{v}_\alpha^n$. Then 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

$$\bar{V}_n = \frac{\frac{1}{2n\delta_n} \sum_{i=1}^n D_i \mathbf{1}_{\{y-\delta_n \leq L_i \leq y+\delta_n\}}}{\frac{1}{2n\delta_n} \sum_{i=1}^n \mathbf{1}_{\{y-\delta_n \leq L_i \leq y+\delta_n\}}}. \tag{14}$$

Note that $\bar{V}_n$ is a ratio of two estimators. Typically, if both numerator and denominator have the same rate of convergence, then so does $\bar{V}_n$ (Law and Kelton 2000, pp. 532-533). Therefore, we focus our analysis on the numerator, while analysis of the denominator follows in a similar manner. Specifically, let

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

denote the numerator. Then it is actually an estimator of $r(y)f(y)$.

An implication of the kernel estimator in (14) is that only the observations in the region $\{y - \delta_n \leq L \leq y + \delta_n\}$ 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 the estimator is $\sqrt{n\delta_n}$. Then a reasonable guess is that having more observations that fall into the important region may accelerate the rate of convergence. To do so, IS is a natural choice.

Among many possible alternatives, one choice of IS scheme is to force all observation to fall into the important region under the IS distribution. We will show how this can be achieved under quite a general simulation framework. Compared to the typical IS schemes, we should note that such an IS scheme is atypical, in the sense that it violates absolute continuity.

Because the density of L is typically unknown, a construction of IS scheme directly from 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

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

Then A corresponds to the important region. Let A_1 be the projection of the set A to the 1st dimension, $A_2(X_1)$ be the projection of the set A to the 2nd 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

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

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

Under the IS distribution, we first simulate X_1 given that $X_1 \in A_1$, then simulate X_2 given that $X_2 \in A_2(X_1)$, and so on. Then we can compute L and D, and $L \in [y - \delta_n, y + \delta_n]$ 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

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

and

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

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

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

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

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

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

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

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

The above 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 best that can be achieved for a typical Monte Carlo estimator.

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

3.4.2 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 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, i.e., $F(t,\theta) = \Pr\{L(\theta) \leq t\} = \mathrm{E}\left[1_{\{L(\theta) \leq t\}}\right]$. Because the indicator is a discontinuous function, the interchange of differentiation and expectation is not valid. However, this problem can be circumvented by 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 6 Suppose that there exist random variables $X_1(\theta)$ and $X_2(\theta)$ such that

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

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

$$\begin{aligned}
|G_1(t, X_1(\theta + \Delta\theta), \theta + \Delta\theta) - G_1(t, X_1(\theta), \theta)| &\leq K_1|\Delta\theta|, \quad \text{and} \\
|G_2(t + \Delta t, X_2(\theta), \theta) - G_2(t, X_2(\theta), \theta)| &\leq K_2|\Delta t|,
\end{aligned}$$

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 6 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 by 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 $t \in \mathfrak{R}$. Suppose that, for any $t \in \mathfrak{R}$, we can observe n i.i.d. samples $(L_1, Y_1(t), Z_1(t)), \ldots, (L_n, Y_n(t), Z_n(t))$. 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

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

105

It has been shown in Fu et al. (2009) that under some regularity conditions, the rate of convergence of $\bar{V}_n^{cmc}$ is $n^{-1/2}$, which is the best that can be achieved.

Typically, the CMC estimators of VaR sensitivities perform very well when applicable. However, as is always the case in applying conditional Monte Carlo, the choice of appropriate conditioning variables is problem dependent. Simple examples have been presented in Fu et al. (2009) to illustrate the advantages of CMC. Further investigation on applying CMC to various practical problems is desirable, and would be of immense benefit to the simulation practitioners.

4 CONCLUSIONS

This tutorial provides a unified view of the simulation of value-at-risk, conditional value-at-risk and their sensitivities. We have focused on representations of VaR, CVaR and their sensitivities, as well as the relations among them. Asymptotic properties of the Monte Carlo estimators of these quantities have been presented. Furthermore, we discuss the issue of efficiency improvement. For VaR and CVaR, we discuss the use of importance sampling and its asymptotic representations; and for VaR sensitivity, we consider the use of importance sampling and conditional Monte Carlo to accelerate the rate of convergence of the estimators.

REFERENCES

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

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

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

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

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

Chu, F., and M. K. Nakayama. 2011. "Confidence intervals for quantiles when applying variance-reduction techniques". *Working paper. Department of Information Systems, New Jersey Institute of Technology*.

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

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

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

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

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

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

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

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

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

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

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

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

Jorion, P. 1997. *Value at risk*. New York: McGraw-Hill, Inc.

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

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

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

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

Liu, G. 2010, December. "Importance sampling for risk contributions of credit portfolios". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 2771–2781. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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

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

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

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

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

Sun, L., 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.

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

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

AUTHOR BIOGRAPHIES

L. JEFF HONG is a professor in the Department of Industrial Engineering and Logistics Management at The Hong Kong University of Science and Technology. His research interests include Monte Carlo simulation, financial risk management, and stochastic optimization. He is currently an associate editor of *Operations Research*, *Naval Research Logistics* and *ACM Transactions on Modeling and Computer Simulation*. His email address is hongl@ust.hk.

GUANGWU LIU is an assistant professor in the Department of Management Sciences at City University of Hong Kong. His research interests include financial risk management and stochastic simulation. His e-mail address is msgw.liu@cityu.edu.hk.