MANAGEMENT SCIENCE
Vol. 55, No. 2, February 2009, pp. 281–293
ISSN 0025-1909 | EISSN 1526-5501 | 09 | 5502 | 0281

DOI 10.1287/mnsc.1080.0901

Simulating Sensitivities of Conditional Value at Risk

L. Jeff Hong, Guangwu Liu
Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong {hongl@ust.hk, liugw@ust.hk}

Conditional value at risk (CVaR) is both a coherent risk measure and a natural risk statistic. It is often used to measure the risk associated with large losses. In this paper, we study how to estimate the sensitivities of CVaR using Monte Carlo simulation. We first prove that the CVaR sensitivity can be written as a conditional expectation for general loss distributions. We then propose an estimator of the CVaR sensitivity and analyze its asymptotic properties. The numerical results show that the estimator works well. Furthermore, we demonstrate how to use the estimator to solve optimization problems with CVaR objective and/or constraints, and compare it to a popular linear programming-based algorithm.

Key words: simulation; statistical analysis; applications; portfolio
History: Accepted by Michael Fu, stochastic models and simulation; received August 19, 2007. This paper was with the authors 3 weeks for 4 revisions. Published online in *Articles in Advance* September 5, 2008.

1. Introduction

Value at risk (VaR) and conditional value at risk (CVaR) are two widely used risk measures. The α-VaR of a random loss L is the α quantile of L, and the α-CVaR of L is the average of all β-VaR of L with $\beta \in (\alpha, 1)$, where $0 < \alpha < 1$ is typically close to 1. If we define the large losses to be the losses in the upper $(1-\alpha)$-tail of the loss distribution, then α-VaR is the lower bound of the large losses, α-CVaR is the mean of the large losses, and $(\alpha+1)/2$-VaR, also known as the tail conditional median (Heyde et al. 2007), is the median of the large losses. They provide information on the potential large losses that an investor may suffer.

Both VaR and CVaR have been used in practice and have also been studied extensively in the literature. There has been debate on which one is a better risk measure. Artzner et al. (1999) define a set of axioms and call the risk measures that satisfy these axioms *coherent risk measures*. One of the axioms is the subadditivity axiom, which requires a risk measure ρ to satisfy $\rho(L_1 + L_2) \le \rho(L_1) + \rho(L_2)$. The subadditivity axiom basically means that "a merger does not create extra risk" (Artzner et al. 1999). They show that VaR does not always satisfy the subadditivity and, therefore, is not a coherent risk measure. CVaR, on the other hand, always satisfies the subadditivity and is a coherent risk measure (Rockafellar and Uryasev 2002).

Heyde et al. (2007), however, argue that the requirement of subadditivity may lead to risk measures that are not robust with respect to the underlying models and data, and thus are not suitable for regulatory purposes. They also provide other evidence that supports the relaxation of the subadditivity, including evidence from utility theory, bankruptcy risk related to merger, psychology theory, and the study of the tail subadditivity of VaR (see also Daníelsson et al. 2005). Heyde et al. (2007) suggest replacing the subadditivity axiom by the comonotonic subadditivity axiom, which requires the subadditivity to hold only for the random variables that always move in the same direction. They call risk measures that satisfy their new set of axioms *natural risk statistics*. They show that both VaR and CVaR are natural risk statistics but that VaR is more robust, i.e., it is less sensitive to the tail distribution, which is often difficult to characterize in practice (see also Heyde and Kou 2004). In this paper, we focus on CVaR. Though it may not be suitable as a regulatory or external risk measure, it is nevertheless widely used in practice and in academic research, and it may also be an excellent internal risk measure for financial institutions.

VaR and CVaR have also been used in stochastic optimization, where the stochastic objective or constraint functions may be substituted by their VaRs or CVaRs to obtain robust solutions. Because a coherent risk measure of a stochastic convex function is also convex (Ruszczyński and Shapiro 2006), CVaR is more popular in stochastic convex optimization. For instance, Rockafellar and Uryasev (2000) and Ruszczyński and Shapiro (2006) study the use of CVaR in stochastic linear programming and stochastic convex optimization, respectively. VaR has also been used in this context. For instance, Hong and Qi (2007) suggest using VaR in the stochastic linear programming.

Suppose that the loss is a function of some parameters, e.g., the loss of a portfolio is often a random function of the interest rate and the volatilities of the assets. Then, the VaR and CVaR of the loss are both functions of these parameters. The partial derivatives of these functions are called VaR and CVaR sensitivities, which provide information on how changes in these parameters affect the output risk measures. These sensitivities are useful in managing risk, verifying model adequacy, and solving stochastic optimization problems.

If the parameters of the loss model are controllable, i.e., the risk managers can adjust these parameters, then their sensitivities can be used for risk management. For instance, if the parameters are the percentages of the total portfolio value allocated to different financial assets, then their sensitivities can be used to adjust the portfolio to ensure that the VaR or CVaR of the loss is below certain level.

If the parameters are uncontrollable, e.g., they are modeled as constants in the loss model, then their sensitivities are measures of model adequacy. In any loss models, there are constants. Some of them may be estimated through historical data, e.g., the volatilities of the financial assets. They are subject to estimation errors. Some of them may be modeled as constants for convenience, e.g., the interest rate may be modeled as a constant in a short period. They are subject to modeling errors. In both situations, if the sensitivity with respect to a certain parameter is high, then we can conclude that the information about this parameter is valuable and it may be necessary to reconsider the model or estimate the parameter accurately, because there are often errors in the specification of the parameter and the VaR or CVaR may change significantly with respect to a small error in the specification.

When the VaR and CVaR are used in stochastic optimization, their sensitivities are the partial derivatives which form the gradients. If we know how to calculate the gradients, we can then apply gradient-based nonlinear optimization algorithms, e.g., the quasi-Newton method (Nocedal and Wright 1999), to solve the optimization problems. The rates of convergence of these algorithms are generally much faster than gradient-free algorithms.

In this paper, we assume that there exists a model of the loss function and that we may simulate the model to obtain independent and identically distributed observations of the loss. Under this assumption, the estimation of VaR sensitivities for general loss functions has been studied in Hong (2008). Hong (2008) shows that the VaR sensitivity can be written as a conditional expectation and also provides a batching estimator to consistently estimate the sensitivity. The estimation of CVaR sensitivities for linear loss functions has been studied by Scaillet (2004). Scaillet (2004) provides a kernel estimator of the sensitivity and shows that the estimator is consistent and follows an asymptotic normal distribution.

In practice, however, many loss functions are nonlinear. For instance, the loss functions of portfolios with financial derivatives are often nonlinear functions of the interest rate and volatilities, and the objective and constraints functions of stochastic convex optimization problems may also be nonlinear. Therefore, we need to have CVaR sensitivity estimators that can be applied to nonlinear loss functions. Furthermore, the performances of the kernel estimator proposed in Scaillet (2004) are sensitive to the choices of kernel function and bandwidth, which may be difficult to choose well in practice. In this paper, we study the estimation of CVaR sensitivities for general loss functions and we propose an estimator that is easier to compute than the kernel estimator. We prove that the estimator is consistent and follows an asymptotic normal distribution.

The general problem of sensitivity estimation has been studied extensively in the simulation literature (see, for instance, Fu 2006 for a recent review). Many approaches have been proposed, including finite-difference approximations, perturbation analysis, and the likelihood ratio method. The key problem of the finite-difference approximation is the tradeoff between the bias and variance of the estimator (e.g., Fox and Glynn 1989). The perturbation analysis uses the pathwise derivatives of the random loss function in the estimation (e.g., Glasserman 1991). It often results in estimators with lower variances when it is applicable. In the likelihood ratio method, the probability density of the loss function is differentiated (e.g., Glynn 1990). It is more widely applicable than the perturbation analysis. However, its estimators often have large variances. The CVaR sensitivity estimator proposed in this paper uses pathwise derivatives. It belongs to the category of the perturbation analysis.

The rest of this paper is organized as follows: In §2, we introduce the background of CVaR and discuss its estimation. In §3, we derive a closed-form expression of the CVaR sensitivity for general loss functions. Then, in §4, we study the asymptotic properties of the estimator, including consistency, asymptotic bias and variance, and asymptotic normality. Numerical results are reported in §5, followed by the conclusions in §6. Some of the lengthy proofs are included in the online appendix (provided in the e-companion).[1]

[1] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

2. Background and Estimation of CVaR

Let L be the random loss and $F_L(y) = \Pr\{L \le y\}$ be the cumulative distribution function (c.d.f.) of L. Then, the inverse c.d.f. of L can be defined as $F_L^{-1}(\gamma) = \inf\{y\colon F_L(y) \ge \gamma\}$. Following the definitions of Trindade et al. (2007), for any $\alpha \in (0,1)$, we define the α-VaR of L as $v_\alpha = F_L^{-1}(\alpha)$ and define the α-CVaR of L as $c_\alpha = 1/(1-\alpha)\int_\alpha^1 v_\beta\, d\beta$. Pflug (2000) shows that c_α can also be written as the following stochastic program:

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

where $[a]^+ = \max\{0, a\}$. Let T be the set of optimal solutions to the stochastic program defined in Equation (1). Trindade et al. (2007) show that $T = [u_\alpha, v_\alpha]$, where $u_\alpha = \sup\{t\colon F_L(t) \le \alpha\}$. In particular, note that $v_\alpha \in T$. Therefore,

$$c_\alpha = v_\alpha + \frac{1}{1-\alpha}\mathrm{E}[L - v_\alpha]^+. \tag{2}$$

When L has a density in the neighborhood of v_α, then $v_\alpha = u_\alpha$. Therefore, the stochastic program defined in Equation (1) has a unique solution. Then, $c_\alpha = \mathrm{E}[L \mid L \ge v_\alpha]$, where the right-hand side of the equation is also known as the expected shortfall or the tail conditional expectation.

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. Serfling (1980) shows that $\hat{v}_\alpha^n \to v_\alpha$ with probability 1 (w.p.1) as $n \to \infty$. If L has a density $f_L(\cdot)$ in the neighborhood of v_α and $f_L(v_\alpha) > 0$, Serfling (1980) also shows that

$$\sqrt{n}(\hat{v}_\alpha^n - v_\alpha) \;\Rightarrow\; \frac{\sqrt{\alpha(1-\alpha)}}{f_L(v_\alpha)}N(0,1) \quad \text{as } n \to \infty, \tag{3}$$

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

Trindade et al. (2007) suggest using the estimator

$$\hat{c}_\alpha^n = \inf_{t\in\Re}\left\{t + \frac{1}{n(1-\alpha)}\sum_{i=1}^n [L_i - t]^+\right\} \tag{4}$$

to estimate the α-CVaR of L. Let $F_n(y) = (1/n)\cdot\sum_{i=1}^n 1_{\{L_i \le y\}}$ be the empirical c.d.f. 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 c.d.f. of $\tilde{L}$ is F_n. Because $\hat{v}_\alpha^n = F_n^{-1}(\alpha)$, then by Equation (2), we have

$$\hat{c}_\alpha^n = \hat{v}_\alpha^n + \frac{1}{n(1-\alpha)}\sum_{i=1}^n [L_i - \hat{v}_\alpha^n]^+. \tag{5}$$

Therefore, we can apply Equation (5) to directly estimate c_α instead of solving the stochastic program in Equation (4).

By the stochastic-program form of $\hat{c}_\alpha^n$, Trindade et al. (2007) show that when $\mathrm{E}(L^2) < \infty$, $\hat{c}_\alpha^n$ is a consistent estimator of c_α, i.e., $\hat{c}_\alpha^n \to c_\alpha$ in probability as $n \to \infty$. Furthermore, if $v_\alpha = u_\alpha$, then

$$\sqrt{n}(\hat{c}_\alpha^n - c_\alpha) \;\Rightarrow\; \tilde{\sigma}_\infty \cdot N(0,1) \quad \text{as } n \to \infty,$$

where

$$\tilde{\sigma}_\infty^2 = \lim_{n\to\infty} n\mathrm{Var}(\hat{c}_\alpha^n) = \frac{1}{(1-\alpha)^2}\cdot\mathrm{Var}([L - v_\alpha]\cdot 1_{\{L \ge v_\alpha\}}). \tag{6}$$

3. A Closed-Form Expression of CVaR Sensitivity

Suppose that the random loss can be modeled as a function $L(\theta)$, where θ is the parameter with respect to which we differentiate. In this paper, we assume that θ is one-dimensional and that $\theta \in \Theta$, 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. Let $L'(\theta) = dL(\theta)/d\theta$ be the sample path derivative of $L(\theta)$. If $L(\theta) = h(\theta, X)$ with some function h and random variable (or vector) X, then $L'(\theta) = \partial h(\theta, X)/\partial\theta$. For instance, $L(\theta)$ may be the random annual loss of a financial portfolio which contains θ shares of a stock with an annual loss X, i.e., $L(\theta) = \theta X + Y$, where Y denotes the annual loss of other assets. Then, $L'(\theta) = X$. When $L(\theta)$ cannot be represented by a closed-form function, $L'(\theta)$ may still be evaluated numerically through perturbation analysis (PA) in many situations (Glasserman 1991). For instance, $L(\theta)$ may be the negative of the value of an asset that is modeled as a diffusion process and θ may be the volatility of the asset. Though the closed form of $L(\theta)$ is often not available, $L'(\theta)$ may be computed through PA (see, for instance, Broadie and Glasserman 1996). In this paper, we assume that $L'(\theta)$ is available for any $\theta \in \Theta$.

Let $v_\alpha(\theta)$ and $c_\alpha(\theta)$ be the α-VaR and α-CVaR of $L(\theta)$ for any θ, where $0 < \alpha < 1$. They are both functions of θ. We make the following assumptions on $L(\theta)$.

Assumption 1. *There exists a random variable K with* $\mathrm{E}(K) < \infty$ *such that* $|L(\theta_2) - L(\theta_1)| \le K|\theta_2 - \theta_1|$ *for all* $\theta_1, \theta_2 \in \Theta$, *and* $L'(\theta)$ *exists w.p.1 for all* $\theta \in \Theta$.

Assumption 2. *The VaR function* $v_\alpha(\theta)$ *is differentiable for any* $\theta \in \Theta$.

Assumption 3. *For any* $\theta \in \Theta$, $\Pr\{L(\theta) = v_\alpha(\theta)\} = 0$.

Assumption 1 is a typical assumption used in pathwise derivative estimation (e.g., Broadie and Glasserman 1996). Assumption 2 implies that there exists some constant $\kappa_\theta > 0$ such that $|v_\alpha(\theta + \Delta\theta) - v_\alpha(\theta)| \le \kappa_\theta|\Delta\theta|$ when $\Delta\theta$ is small enough for any

$\theta \in \Theta$, i.e., $v_\alpha(\theta)$ is locally Lipschitz continuous. Assumption 3 implies that $\Pr\{L(\theta) \geq v_\alpha(\theta)\} = 1 - \alpha$. Then, by Equation (2),

$$\begin{aligned} c_\alpha(\theta) &= v_\alpha(\theta) + \frac{1}{1-\alpha}\mathrm{E}[L(\theta) - v_\alpha(\theta)]^+ \\ &= \mathrm{E}[L(\theta) \mid L(\theta) \geq v_\alpha(\theta)]. \end{aligned}$$

To establish a closed-from expression of the CVaR sensitivity, we need the following lemma that is often used to analyze pathwise derivatives.

Lemma 3.1 (Broadie and Glasserman 1996). *Let f denote a Lipschitz continuous function and D_f denote the set of points at which f is differentiable. Suppose that Assumption 1 is satisfied on an open set $\mathscr{A} \subset \Theta$, and $\Pr\{L(\theta) \in D_f\} = 1$ for all $\theta \in \mathscr{A}$. Then, at every $\theta \in \mathscr{A}$, $d\mathrm{E}[f(L(\theta))]/d\theta = \mathrm{E}[f'(L(\theta)) \cdot L'(\theta)]$.*

Then, we have the following theorem that gives a closed-form expression of $c'_\alpha(\theta)$.

Theorem 3.1. *Suppose that Assumptions 1–3 are satisfied. Then, for any $\theta \in \Theta$,*

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

Proof. To prove the theorem, we prove that $c'_\alpha(\theta^*) = \mathrm{E}[L'(\theta^*) \mid L(\theta^*) \geq v_\alpha(\theta^*)]$ for any $\theta^* \in \Theta$.

Note that, by Assumption 3, for any $\theta \in \Theta$,

$$\begin{aligned} c_\alpha(\theta) &= \mathrm{E}[L(\theta) \mid L(\theta) \geq v_\alpha(\theta)] = \frac{1}{1-\alpha}\mathrm{E}[L(\theta)\cdot 1_{\{L(\theta)\geq v_\alpha(\theta)\}}] \\ &= \frac{1}{1-\alpha}\mathrm{E}\big\{[L(\theta)-v_\alpha(\theta)]\cdot 1_{\{L(\theta)-v_\alpha(\theta)\geq 0\}}\big\} + v_\alpha(\theta). \qquad (7) \end{aligned}$$

For any $\theta^* \in \Theta$, by Assumption 2, there exists 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 1,

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

for all $\theta_1, \theta_2 \in (a_{\theta^*}, b_{\theta^*})$, and $\mathrm{E}(K + \kappa_{\theta^*}) = \mathrm{E}(K) + \kappa_{\theta^*} < \infty$. Note that $f(x) = x \cdot 1_{\{x\geq 0\}}$ is a Lipschitz continuous function and $f'(x) = 1_{\{x\geq 0\}}$ when $x \neq 0$. Because $\Pr\{L(\theta) - v_\alpha(\theta) \neq 0\} = 1$ by Assumption 3, then by Lemma 3.1 with $\mathscr{A} = (a_{\theta^*}, b_{\theta^*})$, at every $\theta \in (a_{\theta^*}, b_{\theta^*})$,

$$\begin{aligned} \frac{d}{d\theta}\mathrm{E}\big\{[L(\theta)-v_\alpha(\theta)]\cdot 1_{\{L(\theta)-v_\alpha(\theta)\geq 0\}}\big\} &= \frac{d}{d\theta}\mathrm{E}[f(L(\theta)-v_\alpha(\theta))] \\ &= \mathrm{E}\big\{[L'(\theta) - v'_\alpha(\theta)] \cdot 1_{\{L(\theta)-v_\alpha(\theta)\geq 0\}}\big\}. \end{aligned}$$

Then, by Equation (7), at every $\theta \in (a_{\theta^*}, b_{\theta^*})$,

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

Because $\theta^* \in (a_{\theta^*}, b_{\theta^*})$, then $c'_\alpha(\theta^*) = \mathrm{E}[L'(\theta^*) \mid L(\theta^*) \geq v_\alpha(\theta^*)]$. This concludes the proof of the theorem. □

Remark 3.1. The conclusion of Theorem 3.1 can also be obtained by differentiating the stochastic program representation of $c_\alpha(\theta)$ (see Equation (1)), where

$$c_\alpha(\theta) = \inf_{t\in\Re}\left\{t + \frac{1}{1-\alpha}\mathrm{E}[L(\theta) - t]^+\right\}.$$

Because $(d/d\theta)\mathrm{E}[L(\theta) - t]^+ = \mathrm{E}[L'(\theta) \cdot 1_{\{L(\theta)\geq t\}}]$ by Lemma 3.1 and $t^* = v_\alpha(\theta)$, then by Danskin's theorem (see Bernhard and Rapaport 1995),

$$c'_\alpha(\theta) = \frac{1}{1-\alpha}\mathrm{E}[L'(\theta)\cdot 1_{\{L(\theta)\geq t\}}]\bigg|_{t=t^*} = \mathrm{E}[L'(\theta) \mid L(\theta) \geq v_\alpha(\theta)]$$

under some technical conditions.

Remark 3.2. Note that $(1-\alpha)c_\alpha(\theta) = \int_\alpha^1 v_\beta(\theta)\,d\beta$. Then,

$$(1-\alpha)c'_\alpha(\theta) = \int_\alpha^1 v'_\beta(\theta)\,d\beta \qquad (8)$$

under some technical conditions. By Theorem 3.1,

$$\begin{aligned} (1-\alpha)c'_\alpha(\theta) &= \mathrm{E}[L'(\theta)\cdot 1_{\{L(\theta)\geq v_\alpha(\theta)\}}] \\ &= \int_\alpha^1 \mathrm{E}[L'(\theta) \mid L(\theta) = v_\beta(\theta)]\,d\beta. \qquad (9) \end{aligned}$$

By Equations (8) and (9), $v'_\alpha(\theta) = \mathrm{E}[L'(\theta) \mid L(\theta) = v_\alpha(\theta)]$ under some technical conditions. This result is the same as the result derived in Hong (2008).

Theorem 3.1 gives a closed-form expression of the CVaR sensitivity. When the closed-form expression can be evaluated, the theorem provides an approach to calculating CVaR sensitivity directly. When the closed-form expression cannot be evaluated, the theorem may be used to derive estimators of CVaR sensitivity, as we do in the next section.

4. Estimation of CVaR Sensitivity

To simplify the notation, we let L and D denote $L(\theta)$ and $L'(\theta)$, respectively, and let c_α and v_α denote $c_\alpha(\theta)$ and $v_\alpha(\theta)$, respectively, when there is no ambiguity. Suppose that we have n i.i.d. observations of (L, D), denoted as $(L_1, D_1), (L_2, D_2), \ldots, (L_n, D_n)$. In this section, we are interested in estimating CVaR sensitivity $c'_\alpha(\theta)$ using the observations.

By Theorem 3.1,

$$c'_\alpha(\theta) = \mathrm{E}[D \mid L \geq v_\alpha] = \frac{1}{1-\alpha}\mathrm{E}[D \cdot 1_{\{L\geq v_\alpha\}}]. \qquad (10)$$

Then, we propose the following estimator of $c'_\alpha(\theta)$:

$$\bar{Y}_n = \frac{1}{n(1-\alpha)}\sum_{j=1}^n D_j \cdot 1_{\{L_j\geq \hat{v}_\alpha^n\}}, \qquad (11)$$

where $\hat{v}_\alpha^n = L_{\lceil n\alpha\rceil : n}$ is a strongly consistent estimator of v_α. 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.

4.1. Strong Consistency

In this subsection, we prove that $\bar{Y}_n$ is a strongly consistent estimator of $c'_\alpha(\theta)$. We need the following proposition, which is proved in the online appendix:

PROPOSITION 4.1. *If* $\Pr\{L = v_\alpha\} = 0$, *then* $(1/n) \cdot \sum_{j=1}^n (1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}) \to 0$ *w.p.*1 *as* $n \to \infty$.

Now, we can state and prove the strong consistency of $\bar{Y}_n$ in estimating $c'_\alpha(\theta)$.

THEOREM 4.1. *Suppose that Assumptions* 1–3 *are satisfied and* $\mathrm{E}(|D|^{1+\gamma}) < \infty$ *for some* $\gamma > 0$. *Then,* $\bar{Y}_n \to c'_\alpha(\theta)$ *w.p.*1 *as* $n \to \infty$.

PROOF. By the strong law of large numbers (Durrett 1996),

$$\frac{1}{n}\sum_{j=1}^n D_j \cdot 1_{\{L_j \ge v_\alpha\}} \to \mathrm{E}(D \cdot 1_{\{L \ge v_\alpha\}}) \quad \text{w.p.1}$$

as $n \to \infty$. Then, it suffices to prove that

$$\frac{1}{n}\sum_{j=1}^n D_j \cdot (1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}) \to 0 \quad \text{w.p.1.}$$

By Hölder's inequality (Rudin 1987),

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^n D_j \cdot (1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}})\right| \\
&\quad \le \left[\frac{1}{n}\sum_{j=1}^n |D_j|^{1+\gamma}\right]^{1/(1+\gamma)} \\
&\quad\quad \cdot \left[\frac{1}{n}\sum_{j=1}^n \left|1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}\right|^{1+1/\gamma}\right]^{\gamma/(1+\gamma)}. \qquad (12)
\end{aligned}$$

Because $\mathrm{E}(|D|^{1+\gamma}) < \infty$, then by the strong law of large numbers and the continuous mapping theorem (Durrett 1996),

$$\begin{aligned}
&\left[\frac{1}{n}\sum_{j=1}^n |D_j|^{1+\gamma}\right]^{1/(1+\gamma)} \\
&\quad \to [\mathrm{E}(|D|^{1+\gamma})]^{1/(1+\gamma)} < \infty \quad \text{w.p.1.} \qquad (13)
\end{aligned}$$

Furthermore, note that

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^n \left|1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}\right|^{1+1/\gamma} \\
&\quad = \frac{1}{n}\sum_{j=1}^n \left|1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}\right| \\
&\quad = (1_{\{\hat{v}_\alpha^n \le v_\alpha\}} - 1_{\{\hat{v}_\alpha^n > v_\alpha\}}) \cdot \frac{1}{n}\sum_{j=1}^n (1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}). \qquad (14)
\end{aligned}$$

By Assumption 3, $\Pr\{L = v_\alpha\} = 0$. Then, $(1/n) \cdot \sum_{j=1}^n (1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}) \to 0$ w.p.1 by Proposition 4.1. Then, Equation (14) goes to 0 w.p.1 as $n \to \infty$. By the continuous mapping theorem,

$$\left[\frac{1}{n}\sum_{j=1}^n \left|1_{\{L_j \ge \hat{v}_\alpha^n\}} - 1_{\{L_j \ge v_\alpha\}}\right|^{1+1/\gamma}\right]^{\gamma/(1+\gamma)} \to 0 \quad \text{w.p.1.} \qquad (15)$$

Therefore, by Equations (12), (13), and (15), the conclusion of the theorem holds. □

4.2. Asymptotic Bias

Let L be a continuous random variable with density $f_L(y)$ and define $g(y) = \mathrm{E}(D \mid L = y)$. Hong (2008) proves that $v'_\alpha(\theta) = g(v_\alpha)$. In the following two subsections, we make the following assumptions.

ASSUMPTION 4. *For all* $\theta \in \Theta$, L *is a continuous random variable with a density function* $f_L(y)$. *Furthermore,* $f_L(y)$ *and* $g(y)$ *are continuous at* $y = v_\alpha$, *and* $f_L(v_\alpha) > 0$.

ASSUMPTION 5. *For all* $\theta \in \Theta$, $\mathrm{E}(L^2) < \infty$ *and* $\mathrm{E}(D^2) < \infty$.

Given Assumptions 4 and 5, we have the following lemma, which is proved in the online appendix:

LEMMA 4.1. *Suppose that Assumptions* 4 *and* 5 *are satisfied. Let* $b_n = 2(\log n)^{1/2}/f_L(v_\alpha)n^{1/2}$ *and* $S_n = \{\omega: |L_{\lceil n\alpha\rceil + k:n} - v_\alpha| > b_n\}$, *where* ω *denotes the realization of the random variable. Then, for any* $0 < \alpha < 1$ *and any fixed constant* $J > 0$, *there exists a constant* $B > 0$ *such that* $\Pr(S_n) \le 2/n^2$ *and* $\mathrm{E}(L^2_{\lceil n\alpha\rceil + k:n}) < B$ *for all sufficiently large* n *and all integers* k *such that* $|k| \le J$.

Let $h(y) = \mathrm{E}(D \cdot 1_{\{L \ge y\}})$, then $h(y) = \int_y^{+\infty} g(t) f_L(t)\, dt$. If $f_L(\cdot)$ and $g(\cdot)$ are continuous at y, then $h(\cdot)$ is differentiable at y and

$$h'(y) = -g(y) f_L(y). \qquad (16)$$

Therefore, by Assumption 4, $h'(v_\alpha) = -g(v_\alpha) f_L(v_\alpha)$. Furthermore, by Equation (10),

$$c'_\alpha(\theta) = \frac{1}{1-\alpha} h(v_\alpha). \qquad (17)$$

Because L is a continuous random variable by Assumption 4, then

$$1_{\{L_1 \ge \hat{v}_\alpha^n\}} = 1_{\{L_1 \ge L_{\lceil n\alpha\rceil:n}\}} = 1_{\{L_1 > L_{\lceil n\alpha\rceil - 1:n}\}} = 1_{\{L_1 > L_{\lceil n\alpha\rceil - 1:n-1}\}} \quad \text{w.p.1}, \qquad (18)$$

where $\hat{v}_\alpha^n$ is calculated from $L_1, L_2, \ldots, L_n$ and $L_{\lceil n\alpha\rceil - 1:n-1}$ is calculated from $L_2, \ldots, L_n$. Therefore, L_1 is independent of $L_{\lceil n\alpha\rceil - 1:n-1}$.

Note that

$$\begin{aligned}
\mathrm{E}(\bar{Y}_n) &= \frac{1}{1-\alpha}\mathrm{E}(D_1 \cdot 1_{\{L_1 \ge \hat{v}_\alpha^n\}}) = \frac{1}{1-\alpha}\mathrm{E}(D_1 \cdot 1_{\{L_1 > L_{\lceil n\alpha\rceil - 1:n-1}\}}) \\
&= \frac{1}{1-\alpha}\mathrm{E}[h(L_{\lceil n\alpha\rceil - 1:n-1})].
\end{aligned}$$

By Equation (17),

$$\mathrm{E}(\bar{Y}_n) - c'_\alpha(\theta) = \frac{1}{1-\alpha}\cdot \mathrm{E}[h(L_{\lceil n\alpha\rceil -1:n-1}) - h(v_\alpha)]. \quad (19)$$

In the next theorem, we show that the bias of $\bar{Y}_n$ is of $o(n^{-1/2})$.

THEOREM 4.2. *Suppose that Assumptions* 1–5 *are satisfied. Then,* $\mathrm{E}(\bar{Y}_n) - c'_\alpha(\theta)$ *is of* $o(n^{-1/2})$.

PROOF. By Equation (19), we only need to show that

$$\sqrt{n}\cdot \mathrm{E}[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)] \to 0,$$

where $U_{k:n-1}$ is the kth order statistic of $n-1$ standard uniform random variables.

Note that

$$(h\circ F_L^{-1})'(\alpha) = h'(v_\alpha)/f_L(v_\alpha) = -g(v_\alpha)$$

by Equation (16), and

$$\sqrt{n}(U_{\lceil n\alpha\rceil -1:n-1} - \alpha) \;\Rightarrow\; \sqrt{\alpha(1-\alpha)}\cdot N(0,1)$$

by Equation (3). Then, by the delta method (Lehmann 1999), we have

$$\sqrt{n}\cdot[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)] \Rightarrow -g(v_\alpha)\sqrt{\alpha(1-\alpha)}\cdot N(0,1).$$

Then, it suffices to show that the sequence $\{\sqrt{n}[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)],\, n=2,3,\ldots\}$ is uniformly integrable.

Define b_n and S_n as those in Lemma 4.1. Then, $\Pr(S_{n-1})$ is of $O(n^{-2})$ by Lemma 4.1. Because $|h(y)| = |\mathrm{E}(D\cdot 1_{\{L\ge y\}})| \le \mathrm{E}(|D|\cdot 1_{\{L\ge y\}}) \le \mathrm{E}(|D|) < \infty$, then

$$n\mathrm{E}\{[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)]^2\cdot 1_{\{\omega\in S_{n-1}\}}\} \le 4n\mathrm{E}^2(|D|)\Pr(S_{n-1}) = O(n^{-1}). \quad (20)$$

Because $f_L(y)$ and $g(y)$ are continuous at $y = v_\alpha$ with $f_L(v_\alpha) > 0$ and $(h\circ F_L^{-1})'(\beta) = -g(F_L^{-1}(\beta))$, by Taylor's theorem (Lehmann 1999), we have

$$n\mathrm{E}\big\{[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)]^2\cdot 1_{\{\omega\notin S_{n-1}\}}\big\} = n\mathrm{E}\big[g^2(\xi_n)(U_{\lceil n\alpha\rceil -1:n-1} - \alpha)^2\cdot 1_{\{\omega\notin S_{n-1}\}}\big],$$

where ξ_n is a random variable between v_α and $L_{\lceil n\alpha\rceil -1:n-1}$. By continuity of $g(y)$ at $y = v_\alpha$, there exists $\epsilon > 0$ and $B > 0$ such that $|g(y)| \le B$ for all $y \in (v_\alpha - \epsilon, v_\alpha + \epsilon)$. When $\omega \notin S_{n-1}$, $|L_{\lceil n\alpha\rceil -1:n-1} - v_\alpha| \le b_{n-1}$, where $b_n \to 0$ as $n \to 0$. Therefore, when n is large enough, $b_{n-1} < \epsilon$ and $|g(\xi_n)| \le B$. Hence,

$$n\mathrm{E}\big\{\big[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)\big]^2\cdot 1_{\{\omega\notin S_{n-1}\}}\big\} \le nB^2\mathrm{E}\big[(U_{\lceil n\alpha\rceil -1:n-1} - \alpha)^2\big]. \quad (21)$$

It is known that $n\mathrm{E}[(U_{\lceil n\alpha\rceil -1:n-1} - \alpha)^2]$ converges to a constant (David 1981). Then, by Equations (20) and (21),

$$\sup_n n\mathrm{E}\{[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)]^2\} < \infty.$$

Therefore, the sequence $\{\sqrt{n}[h\circ F_L^{-1}(U_{\lceil n\alpha\rceil -1:n-1}) - h\circ F_L^{-1}(\alpha)],\, n=2,3,\ldots\}$ is uniformly integrable. $\square$

4.3. Asymptotic Variance

Let

$$\bar{M}_n = (1-\alpha)\bar{Y}_n = \frac{1}{n}\sum_{j=1}^n D_j\cdot 1_{\{L_j\ge \hat{v}_\alpha^n\}}. \quad (22)$$

In this subsection, we first study the asymptotic variance of $\bar{M}_n$. Then, the asymptotic variance of $\bar{Y}_n$ can be easily derived using Equation (22).

Note that

$$\begin{aligned}\mathrm{Var}(\bar{M}_n) &= \frac{1}{n}\mathrm{Var}(D_1\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}) + \left(1-\frac{1}{n}\right)\cdot\mathrm{Cov}(D_1\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}, D_2\cdot 1_{\{L_2\ge\hat{v}_\alpha^n\}})\\ &= \frac{1}{n}\mathrm{Var}(D_1\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}) + \left(1-\frac{1}{n}\right)\cdot[\mathrm{E}(D_1D_2\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}1_{\{L_2\ge\hat{v}_\alpha^n\}}) - \mathrm{E}^2(D_1\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}})].\end{aligned}$$

Furthermore, note that

$$\begin{aligned}1_{\{L_1\ge\hat{v}_\alpha^n\}}\cdot 1_{\{L_2\ge\hat{v}_\alpha^n\}} &= 1_{\{L_1\ge\hat{v}_\alpha^n,\, L_2\ge\hat{v}_\alpha^n\}} = 1_{\{L_1>L_{\lceil n\alpha\rceil -1:n},\, L_2>L_{\lceil n\alpha\rceil -1:n}\}}\\ &= 1_{\{L_1>L_{\lceil n\alpha\rceil -1:n-2},\, L_2>L_{\lceil n\alpha\rceil -1:n-2}\}}\\ &= 1_{\{L_1>L_{\lceil n\alpha\rceil -1:n-2}\}}\cdot 1_{\{L_2>L_{\lceil n\alpha\rceil -1:n-2}\}} \quad \text{w.p.1},\end{aligned}$$

where $L_{\lceil n\alpha\rceil -1:n-2}$ is formed by $L_3, L_4, \ldots, L_n$. It is independent of L_1 and L_2. Then,

$$\begin{aligned}&\mathrm{E}(D_1D_2\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}1_{\{L_2\ge\hat{v}_\alpha^n\}})\\ &\quad= \mathrm{E}(D_1D_2\cdot 1_{\{L_1>L_{\lceil n\alpha\rceil -1:n-2}\}}\cdot 1_{\{L_2>L_{\lceil n\alpha\rceil -1:n-2}\}})\\ &\quad= \mathrm{E}[\mathrm{E}^2(D_1\cdot 1_{\{L_1>L_{\lceil n\alpha\rceil -1:n-2}\}}\mid L_{\lceil n\alpha\rceil -1:n-2})]\\ &\quad= \mathrm{E}[h^2(L_{\lceil n\alpha\rceil -1:n-2})].\end{aligned}$$

Similarly, by Equation (18), $\mathrm{E}^2(D_1\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}) = \mathrm{E}^2[h(L_{\lceil n\alpha\rceil -1:n-1})]$. Therefore,

$$\begin{aligned}&n\mathrm{Var}(\bar{M}_n)\\ &= \mathrm{Var}(D_1\cdot 1_{\{L_1\ge\hat{v}_\alpha^n\}}) + (n-1)\mathrm{Var}[h(L_{\lceil n\alpha\rceil -1:n-2})] + (n-1)\\ &\quad\cdot\{\mathrm{E}^2[h(L_{\lceil n\alpha\rceil -1:n-2})] - \mathrm{E}^2[h(L_{\lceil n\alpha\rceil -1:n-1})]\}. \quad (23)\end{aligned}$$

In the following three propositions, we show that all three terms on the right-hand side of Equation (23) converge to constants as $n \to \infty$. The proofs of the propositions are included in the online appendix.

Proposition 4.2. *Suppose that Assumptions* 1–5 *are satisfied. Then,*

$$\lim_{n\to\infty} \mathrm{Var}(D_1 \cdot 1_{\{L_1 \geq \hat{v}_\alpha^n\}}) = \mathrm{Var}(D \cdot 1_{\{L \geq v_\alpha\}}).$$

Proposition 4.3. *Suppose that Assumptions* 1–5 *are satisfied. Then,*

$$\lim_{n\to\infty} (n-1)\mathrm{Var}[h(L_{\lceil n\alpha\rceil -1:n-2})] = \alpha(1-\alpha)g^2(v_\alpha).$$

Proposition 4.4. *Suppose that Assumptions* 1–5 *are satisfied. Then,*

$$\lim_{n\to\infty} (n-1)\{\mathrm{E}^2[h(L_{\lceil n\alpha\rceil -1:n-2})] - \mathrm{E}^2[h(L_{\lceil n\alpha\rceil -1:n-1})]\} = -2\alpha h(v_\alpha)g(v_\alpha).$$

Combining Propositions 4.2, 4.3, and 4.4, we have the following theorem that characterizes the asymptotic variance of $\bar{Y}_n$.

Theorem 4.3. *Suppose that Assumptions* 1–5 *are satisfied. Then,*

$$\lim_{n\to\infty} n\mathrm{Var}(\bar{Y}_n) = \frac{1}{(1-\alpha)^2}\mathrm{Var}([D - g(v_\alpha)] \cdot 1_{\{L\geq v_\alpha\}}).$$

Proof. By Propositions 4.2–4.4, we have

$$\lim_{n\to\infty} n\mathrm{Var}(\bar{M}_n) = \mathrm{Var}[D\cdot 1_{\{L\geq v_\alpha\}}] + \alpha(1-\alpha)g^2(v_\alpha) - 2\alpha h(v_\alpha)g(v_\alpha).$$

Because $h(v_\alpha) = \mathrm{E}[D\cdot 1_{\{L\geq v_\alpha\}}]$, then

$$\begin{aligned}
\lim_{n\to\infty} n\mathrm{Var}(\bar{M}_n)
&= \{\mathrm{E}[D^2\cdot 1_{\{L\geq v_\alpha\}}] - \mathrm{E}^2[D\cdot 1_{\{L\geq v_\alpha\}}]\} \\
&\quad + \{(1-\alpha)g^2(v_\alpha) - (1-\alpha)^2 g^2(v_\alpha)\} \\
&\quad - \{2g(v_\alpha)\mathrm{E}[D\cdot 1_{\{L\geq v_\alpha\}}] - 2(1-\alpha)g(v_\alpha)\mathrm{E}[D\cdot 1_{\{L\geq v_\alpha\}}]\} \\
&= \{\mathrm{E}[D^2\cdot 1_{\{L\geq v_\alpha\}}] + (1-\alpha)g^2(v_\alpha) - 2g(v_\alpha)\mathrm{E}[D\cdot 1_{\{L\geq v_\alpha\}}]\} \\
&\quad - \{\mathrm{E}^2[D\cdot 1_{\{L\geq v_\alpha\}}] + (1-\alpha)^2 g^2(v_\alpha) \\
&\qquad - 2(1-\alpha)g(v_\alpha)\mathrm{E}[D\cdot 1_{\{L\geq v_\alpha\}}]\} \\
&= \mathrm{E}\{[D-g(v_\alpha)]^2\cdot 1_{\{L\geq v_\alpha\}}\} - \mathrm{E}^2\{[D-g(v_\alpha)]\cdot 1_{\{L\geq v_\alpha\}}\} \\
&= \mathrm{Var}([D-g(v_\alpha)]\cdot 1_{\{L\geq v_\alpha\}}).
\end{aligned}$$

Because $\bar{Y}_n = (1-\alpha)^{-1}\cdot \bar{M}_n$ by Equation (22), then the conclusion of the theorem holds. □

Hong (2008) shows that $v'_\alpha(\theta) = g(v_\alpha)$. Then, the conclusion of Theorem 4.3 becomes

$$\lim_{n\to\infty} n\mathrm{Var}(\bar{Y}_n) = \frac{1}{(1-\alpha)^2}\mathrm{Var}([D - v'_\alpha(\theta)] \cdot 1_{\{L\geq v_\alpha\}}). \quad (24)$$

Comparing Equation (6) to Equation (24), we see clear resemblance between the two.

4.4. Asymptotic Normality

Let $g_2(y) = \mathrm{E}(D^2 \mid L = y)$ and $q(y) = f_L(y)\mathrm{E}(|D| \mid L = y)$. To analyze the asymptotic normality of $\bar{Y}_n$, we make the following additional assumption.

Assumption 6. *The functions $g'(y)$, $f'_L(y)$, $q(y)$, and $g_2(y)$ are continuous at $y = v_\alpha$.*

Remark 4.1. Assumption 6 ensures that L has good mathematical properties in the neighborhood of v_α. Let $r(x) = f_L(y)g(y)$. The continuity of $g'(y)$ and $f'_L(y)$ at $y = v_\alpha$ implies that there exists $\epsilon > 0$ and $B > 0$ such that $|r'(y)| \leq B$ and $|r(y)| \leq B$, for any $y \in (v_\alpha - \epsilon, v_\alpha + \epsilon)$.

Analyzing the asymptotic normality of $\bar{Y}_n$ is equivalent to analyzing the asymptotic normality of $\bar{M}_n$. Note that

$$\bar{M}_n = \frac{1}{n}\sum_{j=1}^n D_j \cdot 1_{\{L_j \geq \hat{v}_\alpha^n\}} = \frac{1}{n}\sum_{j=1}^n D_j \cdot 1_{\{\hat{v}_\alpha^n - L_j \leq 0\}}.$$

There are two major difficulties in analyzing the asymptotic normality of $\bar{M}_n$. The first difficulty is that $D_j \cdot 1_{\{\hat{v}_\alpha^n - L_j\leq 0\}}$, $j = 1, 2, \ldots, n$, are not independent because $\hat{v}_\alpha^n$ is estimated from $L_1, L_2, \ldots, L_n$. To solve this problem, we let

$$\bar{M}_n(v_\alpha) = \frac{1}{n}\sum_{j=1}^n D_j \cdot 1_{\{v_\alpha - L_j \leq 0\}}.$$

Then, the asymptotic normality of $\bar{M}_n(v_\alpha)$ can be analyzed by the classical central limit theorem (Durrett 1996). The second difficulty is that $\bar{M}_n - \bar{M}_n(v_\alpha)$ is difficult to analyze because it involves an indicator function $1_{\{x\leq 0\}}$ that is nonsmooth. To solve this problem, we define

$$\psi_n(x) = \begin{cases} 1 & \text{if } x \leq -\delta_n, \\ \dfrac{1}{2} - \dfrac{1}{2}\sin\left(\dfrac{x}{2\delta_n}\pi\right) & \text{if } -\delta_n < x < \delta_n, \\ 0 & \text{if } x \geq \delta_n, \end{cases}$$

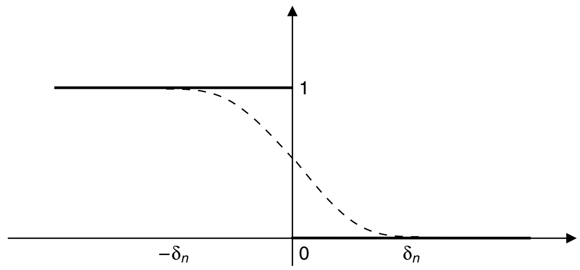
where we require that δ_n satisfies that $n\delta_n^3 \to \infty$ and $n\delta_n^4 \to 0$, e.g., $\delta_n = n^{-7/24}$. Note that $\psi_n(\cdot)$ is a smooth approximation of $1_{\{x\leq 0\}}$ (Figure 1), and $\psi_n(x) \to 1_{\{x\leq 0\}}$ as $n \to \infty$ for any $x \neq 0$. Furthermore, we let

$$\bar{Z}_n = \frac{1}{n}\sum_{j=1}^n D_j \cdot \psi_n(\hat{v}_\alpha^n - L_j),$$

$$\bar{Z}_n(v_\alpha) = \frac{1}{n}\sum_{j=1}^n D_j \cdot \psi_n(v_\alpha - L_j).$$

They are approximations of $\bar{M}_n$ and $\bar{M}_n(v_\alpha)$, respectively.

Figure 1 A Smooth Approximation of $1_{\{x\le 0\}}$



Then, we may write

$$\begin{aligned}\sqrt{n}[\bar{Y}_n - c'_\alpha(\theta)] &= \frac{1}{1-\alpha}\cdot\sqrt{n}[\bar{M}_n-(1-\alpha)c'_\alpha(\theta)]\\ &= \frac{1}{1-\alpha}\cdot\sqrt{n}[\bar{M}_n-\bar{Z}_n]+\frac{1}{1-\alpha}\cdot\sqrt{n}[\bar{Z}_n-\bar{Z}_n(v_\alpha)]\\ &\quad+\frac{1}{1-\alpha}\cdot\sqrt{n}[\bar{Z}_n(v_\alpha)-\bar{M}_n(v_\alpha)]\\ &\quad+\frac{1}{1-\alpha}\cdot\sqrt{n}[\bar{M}_n(v_\alpha)-(1-\alpha)c'_\alpha(\theta)].\end{aligned}$$

In the online appendix, we prove that both $\sqrt{n}[\bar{M}_n-\bar{Z}_n]$ and $\sqrt{n}[\bar{Z}_n(v_\alpha)-\bar{M}_n(v_\alpha)]$ converge to zero in probability as $n\to\infty$, and that $\sqrt{n}[\bar{Z}_n-\bar{Z}_n(v_\alpha)]$ has the same asymptotic distribution as $T_n = -g(v_\alpha)\sqrt{n}[(1/n)\sum_{j=1}^n 1_{\{L_j\ge v_\alpha\}} - (1-\alpha)]$. Then, $\sqrt{n}[\bar{Y}_n-c'_\alpha(\theta)]$ has the same asymptotic distribution as

$$\begin{aligned}&\frac{1}{1-\alpha}\{T_n+\sqrt{n}[\bar{M}_n(v_\alpha)-(1-\alpha)c'_\alpha(\theta)]\}\\ &\quad=\frac{1}{1-\alpha}\cdot\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^n[D_j-g(v_\alpha)]\cdot 1_{\{L_j\ge v_\alpha\}}\right.\\ &\qquad\qquad\left.-(1-\alpha)[c'_\alpha(\theta)-g(v_\alpha)]\right\},\end{aligned}$$

which converges in distribution to $\sigma_\infty\cdot N(0,1)$ by the classical central limit theorem, where

$$\sigma_\infty^2=\frac{1}{(1-\alpha)^2}\mathrm{Var}([D-g(v_\alpha)]\cdot 1_{\{L\ge v_\alpha\}}).$$

Note that σ_∞^2 is the same as the asymptotic variance we obtain in Theorem 4.3. Therefore, we have the following theorem on the asymptotic normality of $\bar{Y}_n$. The detailed proof of the theorem can be found in the online appendix.

Theorem 4.4. *Suppose that Assumptions* 1–6 *are satisfied. Then,*

$$\sqrt{n}[\bar{Y}_n-c'_\alpha(\theta)] \;\Rightarrow\; \sigma_\infty\cdot N(0,1)$$

as $n\to\infty$.

4.5. Constructing Confidence Intervals

In some situations, we may need to know the precision of the estimator or determine the sample size to achieve a certain level of precision. Then, confidence intervals of the estimator can be used (Law and Kelton 2000).

To construct asymptotically valid confidence intervals of $c'_\alpha(\theta)$, we may use Theorem 4.4. However, σ_∞^2 is typically unknown. To estimate σ_∞^2, we first use the following procedure to estimate $g(v_\alpha)$.

1. Divide (L_j, D_j), $j=1,2,\ldots,n$, into k groups. Each group has m observations, where k and m satisfy that $k\to\infty$ and $m\to\infty$ as $n\to\infty$. Denote the observations as (L_{il}, D_{il}), where $i=1,2,\ldots,k$ and $l=1,2,\ldots,m$.
2. For each group, sort L_{il} such that $L_{i(1)}\le L_{i(2)}\le\cdots\le L_{i(m)}$, and denote the corresponding D_{il} as $D_{i(1)}, D_{i(2)},\ldots, D_{i(m)}$. Let $Q_i = D_{i(\lceil m\alpha\rceil)}$.
3. Let $\bar{Q}_n=(1/k)\sum_{i=1}^k Q_i$.

Hong (2008) shows that $\bar{Q}_n$ converges in probability to $g(v_\alpha)$ as $n\to\infty$. Let

$$W_j=\frac{1}{1-\alpha}(D_j-\bar{Q}_n)\cdot 1_{\{L_j\ge\hat{v}_\alpha^n\}}$$

and

$$S_n^2=\frac{1}{n-1}\sum_{j=1}^n(W_j-\bar{W})^2,$$

where $\bar{W}=(1/n)\sum_{j=1}^n W_j$. Then, we have the following lemma, which is proved in the online appendix:

Lemma 4.2. *Suppose that Assumptions* 1–6 *are satisfied and* $\mathrm{E}(|D|^{2+\gamma})<\infty$ *for some* $\gamma>0$. *Then,* $S_n^2\to\sigma_\infty^2$ *in probability as* $n\to\infty$.

Because $S_n^2\to\sigma_\infty^2$ in probability as $n\to\infty$ by Lemma 4.2, then by Theorem 4.4, we have

$$S_n^{-1}\cdot\sqrt{n}[\bar{Y}_n-c'_\alpha(\theta)] \;\Rightarrow\; N(0,1).$$

Then, an asymptotically valid $100(1-\beta)\%$ confidence interval of $c'_\alpha(\theta)$ is

$$(\bar{Y}_n-z_{1-\beta/2}S_n/\sqrt{n},\ \bar{Y}_n+z_{1-\beta/2}S_n/\sqrt{n}),$$

where $z_{1-\beta/2}$ is the $1-\beta/2$ quantile of the standard normal distribution. The numerical results reported in §5 show that the confidence intervals have appropriate coverage probabilities.

5. Numerical Study

In this section, we study the performances of the CVaR sensitivity estimators through numerical experiments. We first consider an example where the portfolio CVaR depends on some input parameters that may have estimation errors. Therefore, we may be interested in their sensitivities. The numerical results show that the estimators and confidence intervals we proposed in this paper can appropriately estimate the sensitivities. We then consider an example on CVaR sensitivity with respect to the tail parameter for a heavy-tailed distribution, to study the impact of the tail parameter on the CVaR. Lastly, we consider a portfolio optimization problem that is subject to a CVaR constraint. We show how to use the CVaR sensitivity estimator to conduct optimization. Compared to the linear programming method of Krokhmal et al. (2002), which is often used to solve this type of problems, our method generally achieves the same level of accuracy but requires significantly lower computational effort.

5.1. Portfolio CVaR Sensitivities with Respect to Input Parameters

Suppose there is a portfolio of many assets, e.g., stocks, options, and securities, that may depend on a number of risk factors. We are interested in estimating the CVaR of the portfolio loss in a given time period. In such a case, we often have a simulation model that can simulate the changes of risk factors and then calculate the future portfolio losses. The portfolio CVaR can then be computed based on the simulated losses. We often assume that the risk factors follow some known distributions, and the parameters of the distributions are often estimated from historical or simulated data. These parameters may have estimation errors. The CVaR sensitivities with respect to these parameters provide measures of model adequacy. If the sensitivities are high, a small error in the parameters may result in a large change in the estimated CVaR value. We can then conclude that the information about these parameters is valuable and it may be necessary to reconsider the model or estimate the parameters more accurately.

In the rest of this subsection, we consider a very simple loss model and illustrate the performances of our CVaR sensitivity estimators. Let ΔS denote the changes in the risk factors in the time period. It is a random vector. Suppose that we model the loss as the following quadratic function of ΔS:

$$L = a_0 + a'\Delta S + \Delta S' A \Delta S,$$

where a_0, a, and A are known. For instance, the loss L is often approximated by the delta-gamma approximation (Glasserman 2004), which follows this quadratic model. In this example, we suppose that $a_0 = 0.3$, $a = (0.8, 1.5)'$, and

$$A = \begin{pmatrix} 1.2 & 0.6 \\ 0.6 & 1.5 \end{pmatrix}.$$

Furthermore, we suppose that ΔS follows a multivariate normal distribution with mean $\mu = (\mu_1, \mu_2)' = (0.01, 0.03)'$ and covariance

$$\Sigma = 0.02 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix},$$

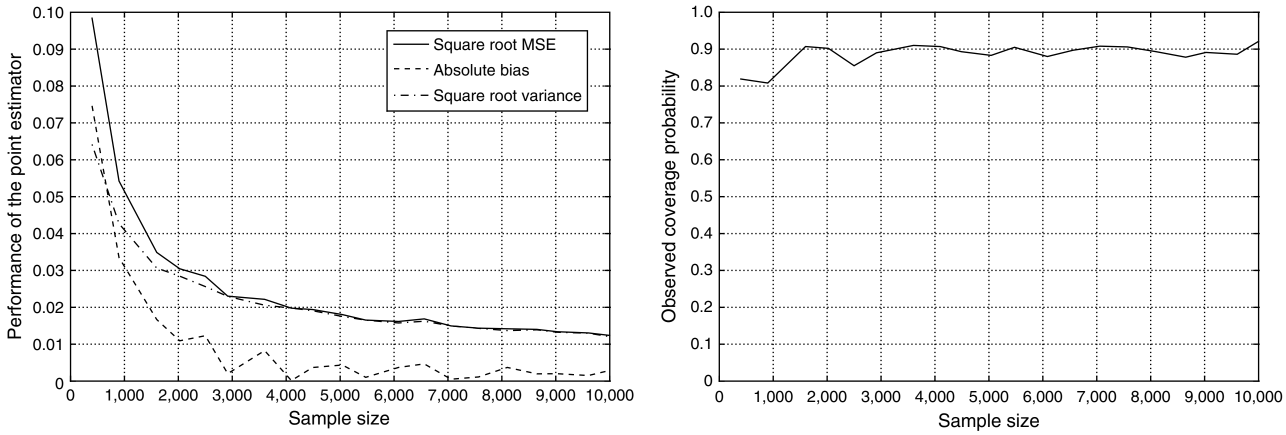
where μ and Σ are estimated from historical data. Let $c_\alpha(\mu)$ denote the α-CVaR of L, where $\alpha = 95\%$. Suppose that the estimation error of μ_1 may be large, e.g., as large as 50% of the estimated value, then we are interested in estimating $\partial c_\alpha(\mu)/\partial \mu_1$ to see if it is sensitive to changes in μ_1.

To analyze the performances of our CVaR sensitivity estimator, we need to know the theoretical value of the sensitivity. Because it is difficult to compute the sensitivity analytically, we use finite difference approximation with 1,000 replications, each using 10^6 independent samples, to remove the estimation error. We estimate $\partial c_\alpha(\mu)/\partial \mu_1 = 1.7391$. We then use this value as a benchmark to test the performances of our estimator. Specifically, we treat the benchmark as the theoretical value and consider the bias, variance, and mean square error (MSE) of our estimator of the CVaR sensitivity as well as the coverage probability of the 90% confidence intervals. To compute the CVaR sensitivity estimator $\bar{Y}_n$, we also need $\partial L/\partial \mu_1$, which we compute through the infinitesimal perturbation analysis (see, for instance, Glasserman 1991).

The performances of the estimator and confidence intervals are reported in Figure 2. The plots are the average performance based on 1,000 independent replications. In the left panel of Figure 2, we plot the estimated absolute biases, the standard deviations, and the square roots of the MSEs relative to the sensitivity value with respect to different sample sizes. From the plot, we see that the errors in the estimator decrease as the sample size increases. The square root of the MSE is below 2% of the sensitivity value when the sample size is 5,000. In the right panel of Figure 2, we plot the observed coverage probabilities of the confidence intervals with respect to different sample sizes. We see that the coverage probabilities are close to 90%, which is the nominal coverage probability, when the sample sizes are larger than 2,000. From the plots, we see that both the point estimator and the confidence interval have the desired properties.

As explained in the example, the estimation error of μ_1 may be large, e.g., as large as 50% of the estimated value. We study the effect of the estimation error to the CVaR of the portfolio through a simulation study. We obtain 1,000 observations of (L, D), and find out that the estimated CVaR is 1.26 with a 90% confidence interval $(1.18, 1.33)$ and the estimated CVaR sensitivity with respect to μ_1 is 1.73 with a 90% confidence interval $(1.66, 1.80)$. Because the estimated $\mu_1 = 0.01$,

Figure 2 **Estimated Absolute Bias, Square Root Variance, and Square Root MSE of the Point Estimator (Left) and Observed Coverage Probabilities of the 90% Confidence Intervals (Right) with Respect to Different Sample Sizes for the Portfolio CVaR Sensitivity Example**



then a 50% error in the estimation of μ_1, i.e., $\Delta\mu_1 = 0.005$, may change the CVaR value by approximately $\partial c_\alpha(\mu)/\partial\mu_1 \cdot \Delta\mu_1$. Considering the worst case in the confidence intervals, a 50% error in μ_1 may change the estimated CVaR value by approximately 0.76%. Therefore, we conclude that the estimated CVaR value is insensitive to the estimation error in μ_1 even when the error may be as large as 50%.

5.2. CVaR Sensitivity with Respect to Tail Parameter for Heavy-Tailed Distribution

Financial data such as daily returns of stock prices often possess heavy tails. Many heavy-tailed distributions have been adopted to model this phenomena in financial and economic applications. In this section, we consider an example where the loss L follows a stable distribution (Nolan 2007) with characteristic function $\mathrm{E}[\exp(iuL)] = \exp(-|u|^{\xi})$ where $\xi \in (0, 2]$. When $\xi = 2$, the stable distribution is a normal distribution. When $\xi \in (0, 2)$, however, the distribution is heavy-tailed, e.g., a Cauchy distribution when $\xi = 1$. In this example, we are interested in estimating the CVaR sensitivity with respect to the tail parameter ξ.

By Chambers et al. (1976), when $\xi \neq 1$, L can be simulated by

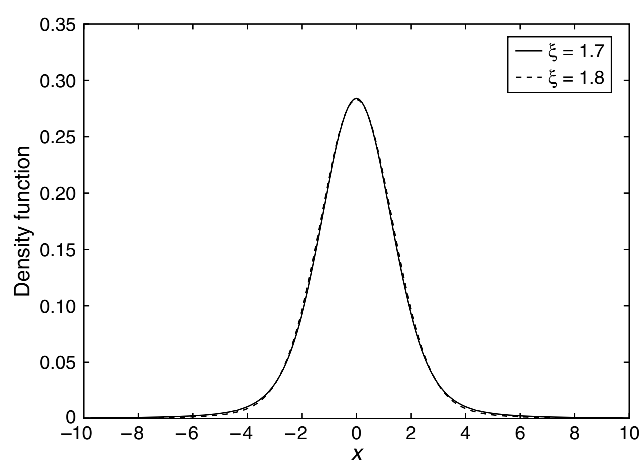
$$L = \frac{\sin(\xi M)}{[\cos(M)]^{1/\xi}} \left[\frac{\cos[(1-\xi)M]}{W}\right]^{(1-\xi)/\xi},$$

where random variables M and W follow a uniform distribution on $(-\pi/2, \pi/2)$ and an exponential distribution with mean 1, respectively. The sample-path derivative $\partial L/\partial\xi$ can also be obtained. We can apply the method developed in this paper to estimate $\partial c_\alpha(\xi)/\partial\xi$. In the numerical example, we let $\alpha = 95\%$ and $\xi = 1.7$.

With 10^6 independent samples, we find out the estimated $c_\alpha(\xi)$ value is 5.0 and the estimated $\partial c_\alpha(\xi)/\partial\xi$ value is -10.3. As pointed out by Heyde and Kou (2004), the tail behaviors of distributions are often difficult to distinguish using practical data. For instance, we plot the densities of the stable distributions with $\xi = 1.7$ and $\xi = 1.8$ in Figure 3. In practice, it is very difficult to distinguish them. However, by Taylor's approximation, changing ξ from 1.7 to 1.8, i.e., $\Delta\xi = 0.1$, will cause $c_\alpha(\xi)$ to change approximately $\partial c_\alpha(\xi)/\partial\xi \cdot \Delta\xi \approx -1.03$, which is about 20.6% of the estimated $c_\alpha(\xi)$ value. This shows that the CVaR is difficult to estimate, which is consistent with the message in Heyde and Kou (2004).

To compare the robustness of VaR and CVaR with respect to the tail parameter, we also estimate $v_\alpha(\xi)$ and $\partial v_\alpha(\xi)/\partial\xi$ with $\alpha = 95\%$ and $\xi = 1.7$ using the method of Hong (2008). With 10^6 independent samples, we find that the estimated $v_\alpha(\xi)$ and $\partial v_\alpha(\xi)/\partial\xi$ values are 2.6 and -1.7, respectively. Then, changing ξ by 0.1 will cause $v_\alpha(\xi)$ to change about 6.5%. This shows that the VaR is significantly more robust to the tail parameter than the CVaR, which is consistent with the message in Heyde et al. (2007).

Figure 3 **Densities of Stable Distributions with** $\xi = 1.7$ **and** $\xi = 1.8$



5.3. Portfolio Optimization with CVaR Constraint

The portfolio optimization problem often tries to find a portfolio that has the maximum expected return and limited downside risk. In this subsection, we consider the problem where the downside risk is measured by the portfolio CVaR. The problem can be formulated as follows:

$$\begin{aligned} &\text{maximize} && \mathrm{E}[\mathbf{p}(t)'\mathbf{x} - \mathbf{p}(0)'\mathbf{x}] \\ &\text{subject to} && c_\alpha[\mathbf{p}(0)'\mathbf{x} - \mathbf{p}(t)'\mathbf{x}] \le K \\ &&& \mathbf{p}(0)'\mathbf{x} \le W \\ &&& x_i \ge 0 \quad \forall\, i = 1, 2, \ldots, k, \end{aligned} \tag{25}$$

where $\mathbf{p}(0)$ is the (deterministic) price vector of all securities at the current time and $\mathbf{p}(t)$ is the (random) price vector at time t in the future, K is the allowed α-CVaR limit at time t, and W is the upper limit of the total wealth of the portfolio at time 0. In problem (25), α is often set as 0.9 or 0.95. To simplify the problem, we assume that short sales are not allowed for all securities, i.e., $x_i \ge 0$ for all $i = 1, 2, \ldots, k$.

Because CVaR is a coherent risk measure, problem (25) is a convex programming problem (Rockafellar and Uryasev 2000). Then, we can use nonlinear optimization algorithms, e.g., sequential quadratic programming (Nocedal and Wright 1999), to solve the problem. Based on the model of $\mathbf{p}(t)$, closed-from expressions for the objective function and the constraint function may be difficult to obtain. In this section, we suggest using the estimates of the objective function and its gradient and the estimates of the constraint function and its gradient to conduct optimization. We first generate n observations of $\mathbf{p}(t) - \mathbf{p}(0)$. Then, the objective function and constraint function and their gradients may be estimated for any given $\mathbf{x}$ and used to solve the optimization problem.

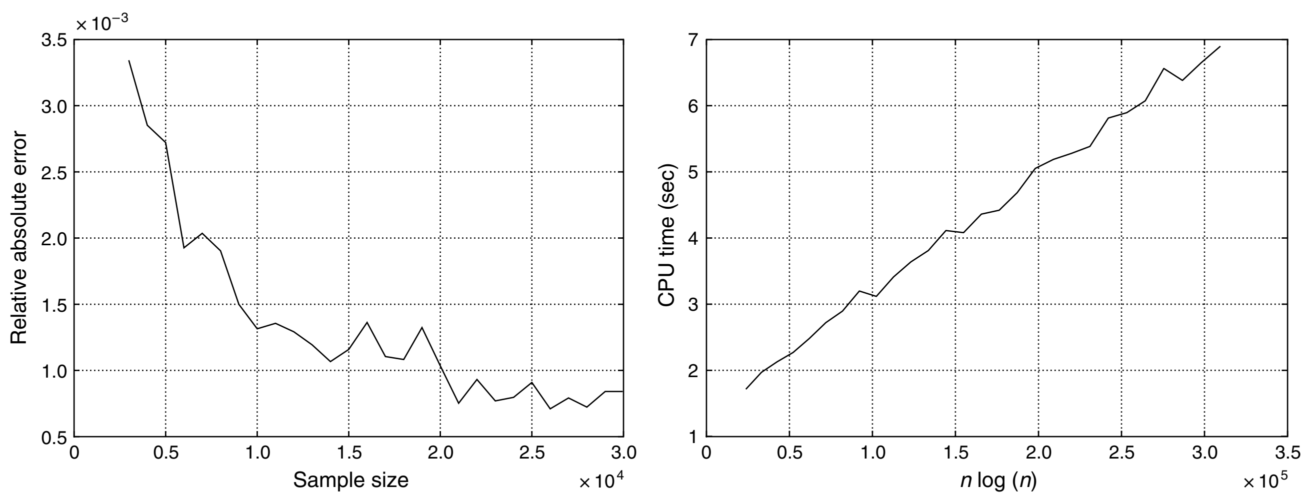
To test the performance of the algorithm, we let $\mathbf{p}(t) - \mathbf{p}(0)$ take a multivariate normal distribution. Then, closed-form expressions for the objective and constraint functions can be derived and the optimal solution to problem (25) can be obtained. This solution enables us to study the performance of our algorithm. When $\mathbf{p}(t) - \mathbf{p}(0)$ follows more complicated models, closed-form expressions for the objective and constraint may not be available but our algorithm can still be applied.

We let $k = 100$, and let $\mathrm{E}[p_i(t) - p_i(0)]$ evenly spread between 0.04 and 0.50 and the standard deviation $\mathrm{Std}[p_i(t) - p_i(0)] = \mathrm{E}[p_i(t) - p_i(0)] - 0.03$ for all $i = 1, 2, \ldots, k$. We let the coefficients of correlation between $p_i(t) - p_i(0)$ and $p_j(t) - p_j(0)$ be 0.35 for any $i \ne j$. We also let $p_i(0) = 1$ for all $i = 1, 2, \ldots, k$, $W = 1$, $K = 0.2$, and $\alpha = 0.95$. Based on closed-form expressions for the objective and constraints, we find that the optimal objective value is 0.4901.

To solve the above problem using simulation, we use Matlab 7.3 function `fmincon` to conduct the optimization with estimated objective function values, constraint function values and their gradients, and randomly generated starting points. The function `fmincon` implements a version of sequential quadratic programming to solve constrained nonlinear optimization problems (http://www.mathworks.com). After `fmincon` reports the optimal solution, we compute the actual objective value of the solution and the 0.95 CVaR of the loss at the solution. We can calculate them because $\mathbf{p}(t) - \mathbf{p}(0)$ takes a multivariate normal distribution. These values can be used to analyze the performance of the algorithm.

In the left panel of Figure 4, we plot the relative differences between the optimal objective value from our method and the actual optimal with respect to different sample sizes. We see that the difference

Figure 4 **Absolute Error of Objective Value Relative to the True Optimal (Left) and CPU Time `Gradient Method` With Respect to $n \log n$ Where n Is the Sample Size (Right) for the Portfolio Optimization Problem**



decreases as the sample size increases. The differences are below 0.1% of the actual optimal when the sample sizes are larger than 20,000. In the right panel of Figure 4, we plot the computational time with respect to $n \log n$, where n is the sample size. We see that the computational time increases linearly with respect to $n \log n$. Note that to solve the problem using our method, the only computational time that is related to the sample size is the estimation of the CVaR and CVaR sensitivities, which is an $O(n \log n)$ operation due to the sorting of the samples. This explains why the total computational time increases in a scale of $n \log n$. Both plots of Figure 4 are averages of 100 independent replications conducted on a PC with Pentium(R)4 3 GHz CPU and 1 GB of RAM.

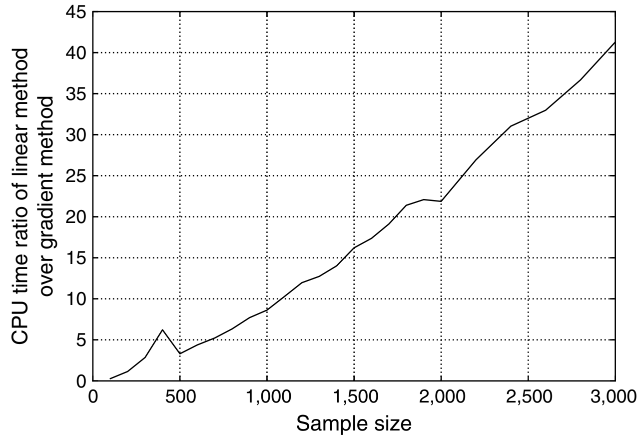
Krokhmal et al. (2002) also propose a method to solve this problem based on the stochastic program representation of CVaR. Let L_j denote the loss of the jth sample path, i.e., $L_j = (\mathbf{p}(0) - \mathbf{p}(t))_j$. Then, the problem can be formulated as the following linear programming problem:

$$\begin{aligned} \text{maximize} \quad & \mathrm{E}[\mathbf{p}(t)'\mathbf{x} - \mathbf{p}(0)'\mathbf{x}] \qquad (26) \\ \text{subject to} \quad & v + \frac{1}{1-\alpha}\frac{1}{n}\sum_{j=1}^{n} z_j \le K \\ & z_j \ge L_j'\mathbf{x} - v, \quad z_j \ge 0, \quad \forall\, j = 1, \ldots, n \\ & \mathbf{p}(0)'\mathbf{x} \le W \\ & x_i \ge 0, \quad \forall\, i = 1, 2, \ldots, k. \end{aligned}$$

The problem can be solved by the Matlab 7.3 function `linprog`, which solves linear programming problems.

We compare the performances of Krokhmal et al. (2002), denoted as the `linear method`, with our method, denoted as the `gradient method`, through numerical experiments. The comparison results are reported in Figures 5 and 6. In Figure 5, we plot the average relative errors of the objective values

Figure 6 CPU Time Ratio of the `Linear Method` to the `Gradient Method`

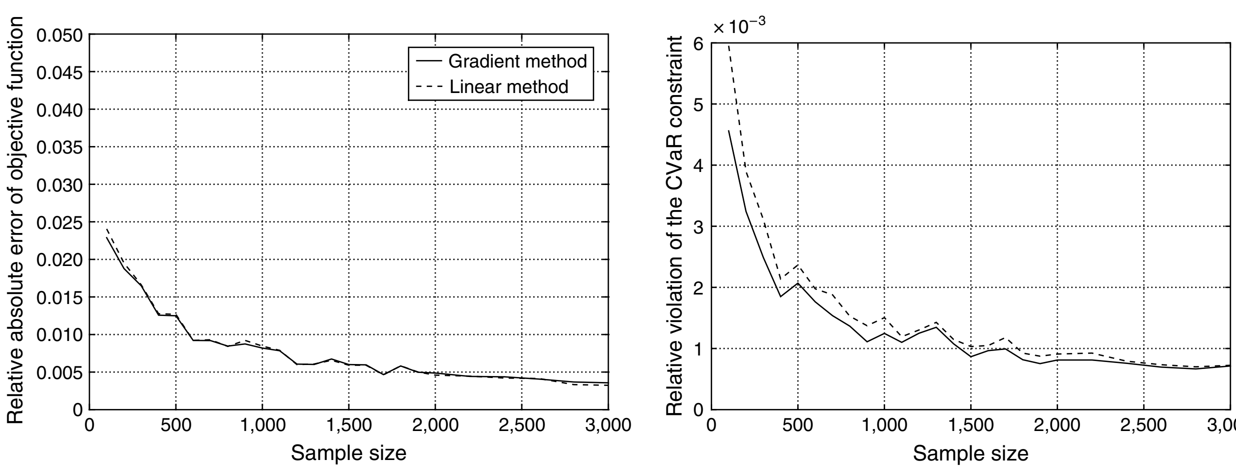


(left panel) and the average violations of the CVaR constraint (right panel) for both methods. We see that both methods have excellent performances and the `gradient method` has smaller constraint violations when the sample size is small. In Figure 6, we plot the ratios of the CPU times of the `linear method` to the `gradient method`. We see that this ratio grows fast as the sample size grows. When the sample size is 3,000, the `linear method` uses 40 times more computational time than the `gradient method`. Note that the number of constraints in the linear programming formulation grows linearly with respect o the sample size n. Therefore, when n becomes larger, the linear program becomes much more difficult to solve and the computational time grows faster than $n \log(n)$. This explains why the `gradient method` is more efficient than the `linear method` when the sample size becomes large.

6. Conclusions

In this paper, we study the estimation of CVaR sensitivity using simulation. We first show that the CVaR

Figure 5 Performances of the Optimal Solutions Found by the `Linear Method` and the `Gradient Method`



sensitivity can be written as a conditional expectation. Based on this result, we propose an estimator of CVaR sensitivity and study its asymptotic properties. The numerical results show that the estimators and confidence intervals we propose in the paper work well.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The authors thank the department editor, associate editor, and referees for their insightful comments that improved the paper in numerous ways, and Steven G. Kou of Columbia University for sharing his working paper (Heyde et al. 2007). Remark 3.1 was pointed out by one of the referees. This research was partially supported by Hong Kong Research Grants Council Grant CERG 613907.

References

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

Bernhard, P., A. Rapaport. 1995. On a theorem of Danskin with an application to a theorem of Von Neumann-Sion. *Nonlinear Anal., Theory, Methods Appl.* **24** 1163–1181.

Broadie, M., P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Sci.* **42**(2) 269–285.

Chambers, J. M., C. L. Mallows, B. W. Stuck. 1976. A method for simulating stable random variables. *J. Amer. Statist. Assoc.* **71** 340–344.

Daníelsson, J., B. N. Jorgensen, G. Samorodnitsky, M. Sarma, C. G. de Vries. 2005. Subadditivity re-examined: The case of value-at-risk. Preprint, London School of Economics, London.

David, H. 1981. *Order Statistics*, 2nd ed. Wiley, New York.

Durrett, R. 1996. *Probability*: *Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA.

Fox, B. L., P. W. Glynn. 1989. Replication schemes for limiting expectations. *Probab. Engrg. Inform. Sci.* **3** 299–318.

Fu, M. C. 2006. Gradient estimation. S. G. Henderson, B. L. Nelson, eds. *Handbooks in Operations Research and Management Science*: *Simulation*. Elsevier, Amsterdam, 575–612.

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

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

Glynn, P. W. 1990. Likelihood ratio gradient estimation for stochastic systems. *Comm. ACM* **33** 75–84.

Heyde, C. C., S. G. Kou. 2004. On the controversy over tailweight of distributions. *Oper. Res. Lett.* **32** 399–408.

Heyde, C. C., S. G. Kou, X. H. Peng. 2007. What is a good external risk measure: Bridging the gaps between robustness, subadditivity, and insurance risk measures. Working paper, Department of Industrial Engineering and Operations Research, Columbia University, New York.

Hong, L. J. 2008. Estimating quantile sensitivities. *Oper. Res.*, ePub ahead of print September 17, http://or.journal.informs.org/cgi/content/abstract/opre.1080.0531v1.

Hong, L. J., X. Qi. 2007. Stochastic linear programming and value-at-risk: An efficient Monte-Carlo approach. Technical report, Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Hong Kong.

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

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, Boston.

Lehmann, E. L. 1999. *Elements of Large-Sample Theory*. Springer-Verlag, New York.

Nocedal, J., S. J. Wright. 1999. *Numerical Optimization*. Springer-Verlag, New York.

Nolan, J. P. 2007. *Stable Distributions—Models for Heavy Tailed Data*. Birkhäuser, Boston.

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

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

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

Rudin, W. 1987. *Real and Complex Analysis*, 3rd ed. McGraw-Hill, Singapore.

Ruszczyński, A., A. Shapiro. 2006. Optimization of convex risk functions. *Math. Oper. Res.* **31** 433–452.

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

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

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