MANAGEMENT SCIENCE
Vol. 55, No. 12, December 2009, pp. 2019–2027
ISSN 0025-1909 | EISSN 1526-5501 | 09 | 5512 | 2019

informs®

DOI 10.1287/mnsc.1090.1090

Conditional Monte Carlo Estimation of Quantile Sensitivities

Michael C. Fu
Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, Maryland 20742, mfu@rhsmith.umd.edu

L. Jeff Hong
Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China, hongl@ust.hk

Jian-Qiang Hu
Department of Management Science, School of Management, Fudan University, 200433 Shanghai, China, hujq@fudan.edu.cn

Estimating quantile sensitivities is important in many optimization applications, from hedging in financial engineering to service-level constraints in inventory control to more general chance constraints in stochastic programming. Recently, Hong (Hong, L. J. 2009. Estimating quantile sensitivities. *Oper. Res.* **57** 118–130) derived a batched infinitesimal perturbation analysis estimator for quantile sensitivities, and Liu and Hong (Liu, G., L. J. Hong. 2009. Kernel estimation of quantile sensitivities. *Naval Res. Logist.* **56** 511–525) derived a kernel estimator. Both of these estimators are consistent with convergence rates bounded by $n^{-1/3}$ and $n^{-2/5}$, respectively. In this paper, we use conditional Monte Carlo to derive a consistent quantile sensitivity estimator that improves upon these convergence rates and requires no batching or binning. We illustrate the new estimator using a simple but realistic portfolio credit risk example, for which the previous work is inapplicable.

Key words: quantiles; value at risk; credit risk; Monte Carlo simulation; gradient estimation
History: Received November 20, 2008; accepted August 4, 2009, by Wallace J. Hopp, stochastic models and simulation. Published online in *Articles in Advance* October 30, 2009.

1. Introduction

There are many simulation settings in which the performance measure of interest is a quantile rather than an expected value. For example, in risk management, value at risk (VaR) is a widely used measure by government regulators to specify minimal capital reserve requirements in the financial industry, and in supply chain management, inventory control using safety stocks are commonly specified. In the simulation setting, derivative estimation of performance measures is critical for optimization and sensitivity analysis. However, there has been limited work on quantile sensitivity estimation, until the recent work of Hong (2009), who derived a batched infinitesimal perturbation analysis (IPA) estimator, and Liu and Hong (2009), who derived a kernel estimator. These estimators are consistent with convergence rates bounded by $n^{-1/3}$ and $n^{-2/5}$, respectively. In this paper, we apply conditional Monte Carlo, or smoothed perturbation analysis, to derive a new estimator that has the following advantages over the previously derived estimators: (a) it doesn't require batching or binning, and (b) it has a superior convergence rate. On the other hand, applying conditional Monte Carlo requires choosing conditioning variables appropriate for the problem setting. The convergence analysis of this estimator uses a technique of handling order statistics recently developed by Hong and Liu (2009), who developed sensitivity estimators for conditional value at risk (CVaR), an alternative performance measure in risk management related to VaR and possessing some superior theoretical properties; however, the results in that paper cannot be applied to VaR, which is far more widely used in the financial services industry. Thus, this paper fills an important gap in both theory and practice.

A recent overview on estimating derivatives (or sensitivities) in simulation is given in Fu (2008), and a more in-depth technical discussion is contained in Fu (2006), which includes numerous references. Glasserman (1991) and Ho and Cao (1991) provide the background on IPA, the technique that is used in Hong (2009) and Hong and Liu (2009), and Fu and Hu (1997) provide the background on smoothed perturbation analysis, or conditional Monte Carlo, which is used in this paper. Literature on simulation estimation of quantiles can be found in Hong (2009). In the rest of this paper, we derive the general conditional

Monte Carlo quantile sensitivity estimator, illustrating it briefly using the single-server queue and then in more detail on an application from finance in portfolio credit risk, an example for which none of the results of Hong (2009), Liu and Hong (2009), and Hong and Liu (2009) are applicable.

2. Conditional Monte Carlo Form of $q'_\alpha(\theta)$

Suppose that $L(\theta)$ is a continuous random variable that depends on a parameter $\theta \in \Theta$, where $\Theta \subset \Re$ is an open set. Let $F(\cdot;\theta)$ denote the distribution function of $L(\theta)$, and $q_\alpha(\theta)$ denote the α-quantile of $L(\theta)$ for any $0 < \alpha < 1$. Then, $q_\alpha(\theta)$ satisfies

$$F(q_\alpha(\theta);\theta) = \alpha. \tag{1}$$

In this paper, we are interested in estimating $q'_\alpha(\theta) = dq_\alpha(\theta)/d\theta$ for any $\theta \in \Theta$.

If $F(t;\theta)$ and $q_\alpha(\theta)$ are both differentiable, then by differentiating with respect to θ on both sides of Equation (1), we have (where ∂ denotes the partial derivative with respect to the subscripted argument)

$$\partial_t F(t;\theta)|_{t=q_\alpha(\theta)} \cdot q'_\alpha(\theta) + \partial_\theta F(t;\theta)|_{t=q_\alpha(\theta)} = 0.$$

Therefore,

$$q'_\alpha(\theta) = -\frac{\partial_\theta F(t;\theta)}{\partial_t F(t;\theta)}\bigg|_{t=q_\alpha(\theta)}. \tag{2}$$

Because $F(t;\theta) = \Pr\{L(\theta) \le t\} = \mathrm{E}[1\{L(\theta) \le t\}]$, and the indicator function $1\{L(\theta) \le t\}$ is a discontinuous function with respect to either θ or t, we cannot interchange the order of differentiation and expectation to compute $\partial_\theta F(t;\theta)$ and $\partial_t F(t;\theta)$, so we apply conditional Monte Carlo, and summarize the results in the following theorem.

Theorem 1. *Suppose that there exist random variables $X_1(\theta)$ and $X_2(\theta)$ such that*

$$\begin{aligned} F(t;\theta) &= \mathrm{E}[\Pr\{L(\theta) \mid X_i(\theta)\}] \\ &= \mathrm{E}[G_i(t, X_i(\theta), \theta)], \quad i = 1, 2, \end{aligned}$$

$G_1(t, X_1(\theta), \theta)$ is differentiable with probability 1 (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

$$|G_1(t, X_1(\theta+\Delta\theta), \theta+\Delta\theta) - G_1(t, X_1(\theta), \theta)| \le K_1|\Delta\theta|, \tag{3}$$

$$|G_2(t+\Delta t, X_2(\theta), \theta) - G_2(t, X_2(\theta), \theta)| \le K_2|\Delta t|, \tag{4}$$

for some random variables K_1 and K_2 with $\mathrm{E}(K_1) < \infty$ and $\mathrm{E}(K_2) < \infty$. If $F(t;\theta)$ and $q_\alpha(\theta)$ are both differentiable, then

$$q'_\alpha(\theta) = -\frac{\mathrm{E}[\partial_\theta G_1(t, X_1(\theta), \theta)]}{\mathrm{E}[\partial_t G_2(t, X_2(\theta), \theta)]}\bigg|_{t=q_\alpha(\theta)}.$$

Proof. By the dominated convergence theorem (Rudin 1987) and Equations (3) and (4), we have

$$\partial_\theta \mathrm{E}[G_1(t, X_1(\theta), \theta)] = \mathrm{E}[\partial_\theta G_1(t, X_1(\theta), \theta)],$$
$$\partial_t \mathrm{E}[G_2(t, X_2(\theta), \theta)] = \mathrm{E}[\partial_t G_2(t, X_2(\theta), \theta)].$$

Then the conclusion of the theorem follows directly from Equation (2) and the definitions of G_1 and G_2. □

Remark 1. It is important to note that L is not required to be Lipschitz continuous with respect to θ, which leads to wider applicability than the results in Hong (2009), Liu and Hong (2009), and Hong and Liu (2009), as the portfolio credit risk example in §4 demonstrates.

Remark 2. Although we condition on two random variables $X_1(\theta)$ and $X_2(\theta)$ in Theorem 1, the two random variables may be the same, as shown in the single-server queue example and the example of §4.1. However, allowing the two random variables to be different provides more flexibility as shown in the example in §4.2.

As an example, we consider a first-come, first-served $G/G/1$ queue. Let A_i, S_i, T_i, and W_i denote the interarrival time, service time, system time, and waiting time of customer i, respectively. Suppose that S_i has a distribution function $G(t;\theta)$ and a density function $g(t;\theta)$, and we are interested in estimating the quantile sensitivity of T_i with respect to θ. By Lindley's equation for the $G/G/1$ queue, $W_i = (T_{i-1} - A_i)^+$. From $T_i = S_i + W_i$, and because S_i and W_i are independent,

$$\begin{aligned} \Pr\{T_i \le t\} = \Pr\{S_i \le t - W_i\} &= \mathrm{E}[\Pr\{S_i \le t - W_i \mid W_i\}] \\ &= \mathrm{E}[G(t - W_i;\theta)]. \end{aligned}$$

Note that

$$\begin{aligned} &\partial_\theta \mathrm{E}[G(t - W_i;\theta)] \\ &\quad = \mathrm{E}\left[-g(t - W_i;\theta)\frac{dW_i}{d\theta} + \partial_\theta G(t - W_i;\theta)\right], \\ &\partial_t \mathrm{E}[G(t - W_i;\theta)] = \mathrm{E}[g(t - W_i;\theta)]. \end{aligned}$$

Then, by Theorem 1,

$$q'_\alpha(\theta) = \frac{\mathrm{E}[g(t - W_i;\theta)dW_i/(d\theta) - \partial_\theta G(t - W_i;\theta)]}{\mathrm{E}[g(t - W_i;\theta)]}\bigg|_{t=q_\alpha(\theta)}. \tag{5}$$

Computation of $dW_i/d\theta$ in Equation (5) can be carried out using infinitesimal perturbation analysis (Ho and Cao 1991) as follows. Because $T_i = S_i + W_i$ and $W_i = (T_{i-1} - A_i)^+$, we have

$$\begin{aligned} \frac{dT_{i-1}}{d\theta} &= \frac{dS_{i-1}}{d\theta} + \frac{dW_{i-1}}{d\theta} \quad \text{and} \\ \frac{dW_i}{d\theta} &= \frac{dT_{i-1}}{d\theta} \cdot 1\{W_i > 0\} \quad \text{w.p.1}, \end{aligned} \tag{6}$$

where $dS_i/(d\theta) = -\partial_\theta G(S_i;\theta)/g(S_i;\theta)$. Suppose that the system starts empty. Then, $dW_1/d\theta = 0$, and Equation (6) can be used to iteratively calculate $dW_i/d\theta$ for any $i = 2, 3, \ldots$.

3. Conditional Monte Carlo Estimator of $q'_\alpha(\theta)$

Let $Y(t) = \partial_\theta G_1(t, X_1(\theta), \theta)$ and $Z(t) = \partial_t G_2 \cdot (t, X_2(\theta), \theta)$ for any $t \in \Re$. Suppose that, for any $t \in \Re$, we can observe n independent and identically distributed (i.i.d.) samples $(L_1, Y_1(t), Z_1(t)), \ldots, (L_n, Y_n(t), Z_n(t))$, e.g., from running simulation experiments. Let $L_{k:n}$ denote the kth-order statistic of $L_1, L_2, \ldots, L_n$, and let $\hat{q}_\alpha = L_{\lceil n\alpha \rceil:n}$. Note that $\hat{q}_\alpha$ is a commonly used estimator of $q_\alpha(\theta)$ (Serfling 1980). Let $\bar{Y}(t) = (1/n)\sum_{i=1}^n Y_i(t)$ and $\bar{Z}(t) = (1/n)\sum_{i=1}^n Z_i(t)$. Then, we can estimate $q'_\alpha(\theta)$ by

$$\hat{q}'_\alpha(\theta) = -\bar{Y}(\hat{q}_\alpha)/\bar{Z}(\hat{q}_\alpha), \tag{7}$$

where $\bar{Y}(\cdot)$, $\bar{Z}(\cdot)$, and $\hat{q}_\alpha$ are all computed using the same set of samples. In the rest of this section, we show that $\hat{q}'_\alpha(\theta)$ is consistent with convergence rate $n^{-1/2}$ under some regularity conditions. The major difficulty in the convergence analysis is the dependence between $(L_1, Y_1(t), Z_1(t))$ and $\hat{q}_\alpha$. We use a technique developed by Hong and Liu (2009) to circumvent this difficulty. A detailed analysis is provided in the appendix.

Let $u_1(t) = \mathrm{E}[Y_1(t)1\{L_1 < t\}]$, $u_2(t) = \mathrm{E}[Y_1(t) \cdot 1\{L_1 > t\}]$, and $u_3(t) = \mathrm{E}[Y_1(L_1)1\{L_1 < t\}]$, and let $v_1(t) = \mathrm{E}[Z_1(t)1\{L_1 < t\}]$, $v_2(t) = \mathrm{E}[Z_1(t)1\{L_1 > t\}]$, and $v_3(t) = \mathrm{E}[Z_1(L_1)1\{L_1 < t\}]$. Then, the following theorem shows that $\hat{q}'_\alpha(\theta)$ is a consistent estimator of $q'_\alpha(\theta)$ as $n \to \infty$. The proof of the theorem is provided in the appendix.

Theorem 2. *Under the assumptions in Theorem* 1, *if* $\mathrm{E}(K_j^{2+\delta}) < \infty$ *for some* $\delta > 0$ *for both* $j = 1, 2$, *where* K_1 *and* K_2 *are the random variables defined in the assumptions of Theorem* 1, *and if* $u_i(t)$ *and* $v_i(t)$, $i = 1, 2, 3$, *are all continuous at* $t = q_\alpha(\theta)$, *then* $\hat{q}'_\alpha(\theta) \xrightarrow{P} q'_\alpha(\theta)$ *as* $n \to \infty$.

Remark 3. The conditions on $u_i(t)$ and $v_i(t)$ in Theorem 2 are typically easy to verify in practice. By Equation (3), we have $|Y_1(t)1\{L_1 < t\}| \le |Y_1(t)| \le K_1$. If $Y_1(t)$ is continuous at $t = q_\alpha(\theta)$ w.p.1 and if $\Pr\{L_1 = q_\alpha(\theta)\} = 0$, then $Y_1(t)1\{L_1 < t\}$ is continuous at $t = q_\alpha(\theta)$ w.p.1. Because $\mathrm{E}(K_1) < \infty$, by the dominated convergence theorem (Rudin 1987), $u_1(t) = \mathrm{E}[Y_1(t)1\{L_1 < t\}]$ is continuous at $t = q_\alpha(\theta)$. Similarly, $u_i(t)$ and $v_i(t)$, $i = 1, 2, 3$, are all continuous at $t = q_\alpha(\theta)$ if $Y_1(t)$ is at $t = q_\alpha(\theta)$ w.p.1 and if $\Pr\{L_1 = q_\alpha(\theta)\} = 0$. For both the $G/G/1$ queue example of §2 and the credit risk example of §4, these conditions can be verified easily.

We now analyze the rate of convergence of $\hat{q}'_\alpha(\theta)$. The big O notation is often used to denote the rate of convergence of a sequence of deterministic numbers, where $a_n = O(b_n)$ as $n \to \infty$ denotes $\limsup_{n\to\infty} |a_n/b_n| < \infty$. For the rate of convergence of a sequence of random variables, we use the concept of *bounded in probability*. A sequence of random variables $\{X_n\}$ is said to be bounded in probability if for every $\epsilon > 0$ there exists $M_\epsilon > 0$ such that

$$\limsup_{n\to\infty} \Pr\{|X_n| \ge M_\epsilon\} \le \epsilon,$$

which we denote by $X_n = O_p(1)$ (see, for instance, Bhat 1985, Lehmann 1999). Furthermore, for two sequences of random variables $\{U_n\}$ and $\{V_n\}$, the notation $U_n = O_p(V_n)$ denotes that $U_n/V_n = O_p(1)$.

Let $f(t)$ denote the density of L, and further let $u_4(t) = \mathrm{E}[Y_1(L_1)Y_2(L_1)1\{L_1 > L_2\}1\{L_1 < t\}]$, $u_5(t) = \mathrm{E}[Y_1(L_1)Y_2(L_1)1\{L_1 < L_2\}1\{L_1 < t\}]$, $v_4(t) = \mathrm{E}[Z_1(L_1)Z_2(L_1)1\{L_1 > L_2\}1\{L_1 < t\}]$, and $v_5(t) = \mathrm{E}[Z_1(L_1)Z_2(L_1)1\{L_1 < L_2\}1\{L_1 < t\}]$. Then, the following theorem shows that the rate of convergence of $\hat{q}'_\alpha(\theta)$ is $n^{-1/2}$. The proof of the theorem is provided in the appendix.

Theorem 3. *Under the assumptions in Theorem* 1, *if* $\mathrm{E}(K_j^2) < \infty$ *for both* $j = 1, 2$, *where* K_1 *and* K_2 *are the random variables defined in the assumptions of Theorem* 1, *and if* $f(t)$ *is continuously differentiable at* $t = q_\alpha(\theta)$, $f(q_\alpha(\theta)) > 0$, *and if* $u_i(t)$ *and* $v_i(t)$, $i = 1, \ldots, 5$, *are all twice differentiable at* $t = q_\alpha(\theta)$, $|u'_i(t)| \le M$, $|u''_i(t)| \le M$, $|v'_i(t)| \le M$, *and* $|v''_i(t)| \le M$ *for some constant* $M > 0$ *for all* $i = 1, \ldots, 5$, *then* $\hat{q}'_\alpha(\theta) - q'_\alpha(\theta) = O_p(n^{-1/2})$.

Remark 4. Although the conditions on $u_i(t)$ and $v_i(t)$ in Theorem 3 are typically difficult to verify, we expect them to hold in practice. Similar conditions have also been used in Hong (2009) and Liu and Hong (2009) to analyze other types of estimators of quantile sensitivities. Furthermore, the numerical results in §4.3 also support the conclusion of the theorem.

By Serfling (1980, p. 52), $X_n = O_p(1)$ if $X_n \Rightarrow X$ as $n \to \infty$, where $\Rightarrow$ denotes "convergence in distribution." Hong (2009) proposes a batching estimator $\bar{D}_{mk}$ of $q'_\alpha(\theta)$. He shows that $\sqrt{k}(\bar{D}_{mk} - q'_\alpha(\theta)) \Rightarrow N(0, \sigma^2)$ as $n \to \infty$ with some $\sigma^2 > 0$, where k satisfies that $k \to \infty$ and $k^3/n^2 \to 0$ as $n \to \infty$. Therefore, $\bar{D}_{mk} = O_p(k^{-1/2}) = O_p(n^{-1/3})$. Liu and Hong (2009) propose a kernel estimator $\bar{V}_n$ of $q'_\alpha(\theta)$. They show that $\sqrt{n\delta_n}(\bar{V}_n - q'_\alpha(\theta)) \Rightarrow N(\mu, \sigma_2^2)$ as $n \to \infty$ with some μ and $\sigma_2 > 0$, where δ_n satisfies $\delta_n \to 0$ and $n\delta_n^5 \to c$ with some $c \ge 0$ as $n \to \infty$, and $\sup_{n\to\infty}(n\delta_n^3)^{-1} < \infty$. Therefore, $\bar{V}_n = O_p((n\delta_n)^{-1/2}) = O_p(n^{-2/5})$. By Theorem 3, we see that the rate of convergence of $\hat{q}'_\alpha$ is of $O_p(n^{-1/2})$. It is clear that $\hat{q}'_\alpha(\theta)$ is asymptotically more efficient than the other two estimators.

4. Sensitivity Analysis for Portfolio Credit Risk

Portfolio credit risk refers to the losses due to defaults of obligors in a portfolio. Let L denote the loss, which we model as

$$L=\sum_{i=1}^{m} l_i \cdot 1\{X_i < x_i\},$$

where m is the number of obligors in the portfolio. In this model, we use a latent random variable X_i to determine the default of obligator i. If $X_i < x_i$, where x_i is a threshold, then obligator i defaults and it will cause a (random) loss of l_i. In this section, we assume that l_i is a continuous random variable independent of X_i. Under this assumption, L is a continuous random variable, and thus our estimator can be applied. With respect to parameters that affect X_i, however, the loss function L is not Lipschitz continuous, so the estimators of Hong (2009), Liu and Hong (2009), and Hong and Liu (2009) cannot be applied.

To capture the dependence between the defaults of obligors, $(X_1, X_2, \ldots, X_m)$ are often assumed dependent. The normal (Gaussian) copula model is widely used in the industry (see, for instance, Li 2000, Glasserman 2004). It is the basis of the CreditMetrics models developed by J. P. Morgan and the Moody's KMV system. However, the model fails to capture extreme credit risk scenarios where many obligors default simultaneously, which is often observed empirically (Mashal and Zeevi 2003). Bassamboo et al. (2008) suggest using the following model:

$$X_i=\frac{\rho Z+\sqrt{1-\rho^2}\,\eta_i}{W}, \quad i=1,2,\ldots,m,$$

where Z denotes the common factor that affects all obligors, η_i denotes obligor i's idiosyncratic risk, W is a nonnegative random variable that captures a common shock to all obligors, and Z, W, and η_i are mutually independent. When Z and η_i are all independent standard normal random variables and $W=1$, the model becomes the one-factor normal copula model. When W is a random variable, a small W value will create a common shock to all obligors and cause many of them to default simultaneously. Bassamboo et al. (2008) show that the model can explain extreme credit risk when W or W^2 follows a gamma distribution.

Let q_α denote the α-VaR of the portfolio loss L. When α is close to 1, e.g., $\alpha=0.95$, q_α is an important risk measure of the portfolio credit risk. Because the loss L depends on many parameters, we are interested in finding the sensitivities of the α-VaR with respect to these parameters.

4.1. Parameter of an Individual Obligor

We first consider the situation where θ affects only a single obligor and the goal is to find $q'_\alpha(\theta)$. For instance, we consider θ as a parameter of the distribution function of η_1, and we are interested in estimating $q'_\alpha(\theta)$.

We let $H_i(\cdot)$ and $F_i(\cdot)$ denote the distribution functions of l_i and η_i, respectively, let $h_i(\cdot)$ denote the densities of l_i, and let $F_W(\cdot)$ and $F_Z(\cdot)$ denote the distribution functions of W and Z, respectively. Let $\lambda_i=(x_iW-\rho Z)/\sqrt{1-\rho^2}$ for all $i=1,\ldots,m$. Then λ_i is a random variable whose value is completely determined by (measurable with respect to) W and Z. Let $F_L(\cdot)$ denote the distribution function of L. Then

$$F_L(t)=\Pr\{L\le t\}=\mathrm{E}\left[\Pr\left\{\sum_{i=1}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\le t\,\middle|\, W,Z\right\}\right]. \quad (8)$$

Note that for any $t>0$,

$$\begin{aligned}
&\Pr\left\{\sum_{i=1}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\le t\,\middle|\, W,Z\right\}\\
&\quad=\Pr\{\eta_1<\lambda_1\mid W,Z\}\cdot\Pr\left\{l_1\le t-\sum_{i=2}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\,\middle|\, W,Z\right\}\\
&\qquad+\Pr\{\eta_1\ge\lambda_1\mid W,Z\}\cdot\Pr\left\{\sum_{i=2}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\le t\,\middle|\, W,Z\right\}\\
&\quad=F_1(\lambda_1)\cdot\mathrm{E}\left[H_1\left(t-\sum_{i=2}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\right)\,\middle|\, W,Z\right]\\
&\qquad+\bar{F}_1(\lambda_1)\cdot\Pr\left\{\sum_{i=2}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\le t\,\middle|\, W,Z\right\},
\end{aligned}$$

where $\bar{F}_i(t)=1-F_i(t)$. We may analyze $\Pr\{\sum_{i=2}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\le t\mid W,Z\}$ using the same approach. Then, after m iterations, we have

$$\begin{aligned}
&\Pr\left\{\sum_{i=1}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\le t\,\middle|\, W,Z\right\}\\
&\quad=\sum_{i=1}^{m}\left\{F_i(\lambda_i)\cdot\mathrm{E}\left[H_i\left(t-\sum_{j=i+1}^{m} l_j\cdot 1\{\eta_j<\lambda_j\}\right)\,\middle|\, W,Z\right]\cdot\prod_{j=1}^{i-1}\bar{F}_j(\lambda_j)\right\}+\prod_{i=1}^{m}\bar{F}_i(\lambda_i),
\end{aligned}$$

so by Equation (8) we have

$$\begin{aligned}
F_L(t)&=\sum_{i=1}^{m}\mathrm{E}\left[F_i(\lambda_i)\cdot H_i\left(t-\sum_{j=i+1}^{m} l_j\cdot 1\{\eta_j<\lambda_j\}\right)\cdot\prod_{j=1}^{i-1}\bar{F}_j(\lambda_j)\right]\\
&\quad+\mathrm{E}\left[\prod_{i=1}^{m}\bar{F}_i(\lambda_i)\right]. \quad (9)
\end{aligned}$$

Because θ is a parameter of $F_1(\cdot)$, differentiating Equation (9) with respect to θ yields

$$\begin{aligned}
&\partial_\theta F_L(t)\\
&\quad=\mathrm{E}\left[\partial_\theta F_1(\lambda_1)\cdot H_1\left(t-\sum_{i=2}^{m} l_i\cdot 1\{\eta_i<\lambda_i\}\right)\right]
\end{aligned}$$

$$- \sum_{i=2}^{m} \mathrm{E}\left[\partial_\theta F_1(\lambda_1)\cdot F_i(\lambda_i)\cdot H_i\left(t-\sum_{j=i+1}^{m} l_j\cdot 1\{\eta_j<\lambda_j\}\right)\cdot\prod_{j=2}^{i-1}\bar{F}_j(\lambda_j)\right]$$

$$-\mathrm{E}\left[\partial_\theta F_1(\lambda_1)\cdot\prod_{i=2}^{m}\bar{F}_i(\lambda_i)\right]. \qquad (10)$$

Furthermore, for any $t>0$, differentiating Equation (9) with respect to t yields

$$\partial_t F_L(t)=\sum_{i=1}^{m}\mathrm{E}\left[F_i(\lambda_i)\cdot h_i\left(t-\sum_{j=i+1}^{m} l_j\cdot 1\{\eta_j<\lambda_j\}\right)\cdot\prod_{j=1}^{i-1}\bar{F}_j(\lambda_j)\right]. \qquad (11)$$

Therefore, we may apply the conditional Monte Carlo estimator of §3 to estimate $q'_\alpha(\theta)$ using Equations (10) and (11).

4.2. Parameter of Common Shock

Now we consider θ as a parameter of the distribution function of W, and we are interested in estimating $q'_\alpha(\theta)$. Note that the representation of $F_L(t)$ of Equation (9) may not be used to obtain $\partial_\theta F_L(t)$ when θ is a parameter of W, because θ would occur in the indicator function. Therefore, we need another representation of $F_L(t)$ to compute $\partial_\theta F_L(t)$.

Without loss of generality, we assume that $x_i<0$ for all $i=1,\ldots,m$. Let $\xi_i=(\rho Z+\sqrt{1-\rho^2}\eta_i)/x_i$. Sort ξ_i from the smallest to the largest, and let $\xi_{(i)}$ denote the ith smallest one. We also define $\xi_{(0)}=-\infty$ and $\xi_{(m+1)}=+\infty$. Furthermore, we let $l_{(i)}$ denote the observation of $l_1,\ldots,l_m$ that corresponds to $\xi_{(i)}$ for all $i=1,\ldots,m$ (e.g., if $\xi_1=\xi_{(i)}$, then $l_{(i)}=l_1$). Then,

$$\begin{aligned}F_L(t)&=\mathrm{E}\left[\Pr\left\{\sum_{i=1}^{m} l_i\cdot 1\{W<\xi_i\}\le t\,\middle|\,\xi_i,l_i,i=1,\ldots,m\right\}\right]\\&=\mathrm{E}\left[\sum_{i=1}^{m+1}1\left\{\sum_{j=i}^{m} l_{(j)}\le t\right\}\cdot\Pr\{\xi_{(i-1)}\le W<\xi_{(i)}\,|\,\xi_1,\ldots,\xi_m\}\right]\\&=\sum_{i=1}^{m+1}\mathrm{E}\left[1\left\{\sum_{j=i}^{m} l_{(j)}\le t\right\}\cdot[F_W(\xi_{(i)})-F_W(\xi_{(i-1)})]\right]. \qquad (12)\end{aligned}$$

Because θ is a parameter of $F_W(\cdot)$, then

$$\partial_\theta F_L(t)=\sum_{i=1}^{m+1}\mathrm{E}\left[1\left\{\sum_{j=i}^{m} l_{(j)}\le t\right\}\cdot[\partial_\theta F_W(\xi_{(i)})-\partial_\theta F_W(\xi_{(i-1)})]\right], \qquad (13)$$

where we define $\partial_\theta F_W(\xi_{(0)})=\partial_\theta F_W(\xi_{(m+1)})=0$. Combining Equations (11) and (13), we can apply the conditional Monte Carlo estimator of §3 to estimate $q'_\alpha(\theta)$.

4.3. Numerical Example

We consider a simple example to illustrate the performance of the conditional Monte Carlo quantile sensitivity estimators. We suppose that there are two obligors (i.e., $m=2$), $\rho=0.6$, Z follows a standard normal distribution, η_i follows a normal distribution with mean θ_i and variance 1, $i=1,2$, and W follows an exponential distribution with rate θ. Furthermore, we suppose that $x_1=x_2=-2$ and l_i are independently uniformly distributed on $[0,1]$. We let $\theta_1=\theta_2=0$, $\theta=1/0.3$, and $\alpha=0.95$. We are interested in estimating the sensitivity of the α-quantile of the loss L with respect to θ_1 (a parameter of an individual obligor) and θ (a parameter of the common shock).

By using Equation (12), and because we only have two obligors, we can compute the quantile sensitivities by combining numerical integration with the finite-difference method. With a very large sample size (10^9), we estimate the true values of $\partial_{\theta_1}q_\alpha(L)$ and $\partial_\theta q_\alpha(L)$ to be approximately -0.2521 and 0.0628, respectively. We use these values as the benchmarks to evaluate the performance of the conditional Monte Carlo estimators. Note that we cannot compare our estimators to the batched IPA estimator of Hong (2009) and the kernel estimator of Liu and Hong (2009) for this example, because neither of them can be applied to the portfolio credit risk example.

In Tables 1 and 2, we report the estimated mean, standard deviation (SD), and root mean square error (RMSE) of our estimators under different sample sizes. All values reported in the tables are based on 1,000 independent replications. From the tables, we see that our estimators have good performance. For both $\partial_{\theta_1}q_\alpha(L)$ and $\partial_\theta q_\alpha(L)$, the RMSEs are less than 1% of the true values when the sample size is 10^5. Furthermore, the RMSEs appear to be decreasing according to the $n^{-1/2}$ rate (i.e., an additional decimal place of accuracy is obtained when the sample size

Table 1 **Performance of $\partial_{\theta_1}q_\alpha(L)$ (True Value $=-0.2521$)**

Sample size	10^3	10^4	10^5	10^6
Mean	−0.2533	−0.2525	−0.2520	−0.2521
SD	0.022	0.0067	0.0020	0.00065
RMSE	0.022	0.0067	0.0020	0.00065

Table 2 **Performance of $\partial_\theta q_\alpha(L)$ (True Value $=0.0628$)**

Sample size	10^3	10^4	10^5	10^6
Mean	0.0631	0.0629	0.0628	0.0628
SD	0.0057	0.0019	0.00060	0.00019
RMSE	0.0057	0.0019	0.00060	0.00019

is increased by a factor of 100). These results support the conclusion of Theorem 3, even though some of the conditions of the theorem are difficult to verify for this portfolio credit risk example.

5. Concluding Remarks

We have introduced a new quantile sensitivity estimator derived using conditional Monte Carlo. The new estimator is easier to apply than existing estimators in the literature because it doesn't require batching or binning; moreover, it possesses a superior asymptotic convergence rate. However, as is always the case in applying conditional Monte Carlo, the choice of appropriate conditioning variables is problem dependent. Simple examples in finance and queueing are presented to illustrate the estimators, and numerical experiments on a small portfolio credit risk example for which the previous estimators are not applicable are used to test the empirical performance of the estimators. More extensive numerical testing of the new and existing estimators in various application domains to compare empirical performance advantages and disadvantages is an area of further investigation that would be of immense benefit to the simulation practitioner, especially given the widespread use of Monte Carlo simulation in the financial services industry and the recent advances in stochastic programming.

Acknowledgments

The authors thank the associate editor and two anonymous referees for their helpful comments and suggestions. Michael C. Fu is supported in part by the National Science Foundation under Grants DMI-0540312 and DMI-0323220, and by the Air Force Office of Scientific Research under Grant FA9550-07-1-0366. L. Jeff Hong is supported in part by the Hong Kong Research Grants Council under General Research Fund 613907 and by the National Natural Science Foundation of China under Grant 70932003. Jian-Qiang Hu is supported in part by the National Natural Science Foundation of China under Grant 70832002, by the Shanghai Science and Technology PuJiang Funds under Grant 09PJ1401500, and by a grant from IBM China Research Laboratory.

Appendix

To simplify the notation, we let q_α, q'_α, $\hat{q}_\alpha$, and $\hat{q}'_\alpha$ denote $q_\alpha(\theta)$, $q'_\alpha(\theta)$, $\hat{q}_\alpha(\theta)$, and $\hat{q}'_\alpha(\theta)$, respectively, when there is no confusion. By Theorem 1, $q'_\alpha = -\mathrm{E}[Y_1(q_\alpha)]/\mathrm{E}[Z_1(q_\alpha)]$, and by Equation (7), $\hat{q}'_\alpha = -\bar{Y}(\hat{q}_\alpha)/\bar{Z}(\hat{q}_\alpha)$. In this appendix, we prove Theorems 2 and 3.

A.1. Proof of Theorem 2

Note that $\hat{q}'_\alpha = -\bar{Y}(\hat{q}_\alpha)/\bar{Z}(\hat{q}_\alpha)$. Then, by Slutsky's theorem (Serfling 1980), to prove that $\hat{q}'_\alpha \xrightarrow{P} q'_\alpha$ as $n \to \infty$, it suffices to show that $\bar{Y}(\hat{q}_\alpha) \xrightarrow{P} \mathrm{E}[Y_1(q_\alpha)]$, and $\bar{Z}(\hat{q}_\alpha) \xrightarrow{P} \mathrm{E}[Z_1(q_\alpha)]$ as $n \to \infty$. In the rest of this subsection, we only prove that $\bar{Y}(\hat{q}_\alpha) \xrightarrow{P} \mathrm{E}[Y_1(q_\alpha)]$. The proof of $\bar{Z}(\hat{q}_\alpha) \xrightarrow{P} \mathrm{E}[Z_1(q_\alpha)]$ is similar.

Because of i.i.d. sampling, $\mathrm{E}[\bar{Y}(\hat{q}_\alpha)] = \mathrm{E}[Y_1(\hat{q}_\alpha)]$. To analyze the bias of the term, the main difficulty is the dependence between $(L_1, Y_1(t))$ and $\hat{q}_\alpha$. To circumvent this difficulty, we use a technique that is also used in Hong and Liu (2009). Note that if $L_1 < \hat{q}_\alpha$, then $\hat{q}_\alpha = L_{\lceil n\alpha\rceil-1:n-1}$, where $L_{\lceil n\alpha\rceil-1:n-1}$ is the $\lceil n\alpha\rceil - 1$st-order statistic of $L_2, L_3, \ldots, L_n$ and is independent of $(L_1, Y_1(t))$. Similarly, $\hat{q}_\alpha = L_{\lceil n\alpha\rceil:n-1}$ if $L_1 > \hat{q}_\alpha$, and $L_{\lceil n\alpha\rceil-1:n-1} < L_1 < L_{\lceil n\alpha\rceil:n-1}$ if $L_1 = \hat{q}_\alpha$. Then

$$\begin{aligned}\mathrm{E}[Y_1(\hat{q}_\alpha)] &= \mathrm{E}[Y_1(\hat{q}_\alpha)\cdot 1\{L_1 < \hat{q}_\alpha\}] + \mathrm{E}[Y_1(\hat{q}_\alpha)\cdot 1\{L_1 > \hat{q}_\alpha\}] \\ &\quad + \mathrm{E}[Y_1(\hat{q}_\alpha)\cdot 1\{L_1 = \hat{q}_\alpha\}] \\ &= \mathrm{E}[Y_1(L_{\lceil n\alpha\rceil-1:n-1})\cdot 1\{L_1 < L_{\lceil n\alpha\rceil-1:n-1}\}] \\ &\quad + \mathrm{E}[Y_1(L_{\lceil n\alpha\rceil:n-1})\cdot 1\{L_1 > L_{\lceil n\alpha\rceil:n-1}\}] \\ &\quad + \mathrm{E}[Y_1(L_1)\cdot 1\{L_1 < L_{\lceil n\alpha\rceil:n-1}\}] \\ &\quad - \mathrm{E}[Y_1(L_1)\cdot 1\{L_1 < L_{\lceil n\alpha\rceil-1:n-1}\}] \\ &= \mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-1}) + u_2(L_{\lceil n\alpha\rceil:n-1}) + u_3(L_{\lceil n\alpha\rceil:n-1}) \\ &\qquad - u_3(L_{\lceil n\alpha\rceil-1:n-1})].\end{aligned}$$

Note that $\mathrm{E}[Y_1(q_\alpha)] = u_1(q_\alpha) + u_2(q_\alpha)$. Then

$$\begin{aligned}&\mathrm{E}[\bar{Y}(\hat{q}_\alpha)] - \mathrm{E}[Y_1(q_\alpha)] \\ &\quad = \mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-1}) - u_1(q_\alpha)] + \mathrm{E}[u_2(L_{\lceil n\alpha\rceil:n-1}) - u_2(q_\alpha)] \\ &\qquad + \mathrm{E}[u_3(L_{\lceil n\alpha\rceil:n-1}) - u_3(L_{\lceil n\alpha\rceil-1:n-1})]. \qquad (14)\end{aligned}$$

Then we have the following proposition on the bias of $\bar{Y}(\hat{q}_\alpha)$.

Proposition 1. *Suppose that $u_i(t)$, $i = 1, 2, 3$, are all continuous at $t = q_\alpha$. Then, $\mathrm{E}[\bar{Y}(\hat{q}_\alpha)] \to \mathrm{E}[Y_1(q_\alpha)]$ as $n \to \infty$.*

Proof. By Bahadur's theorem (Serfling 1980, p. 93), $L_{\lceil n\alpha\rceil-1:n-1} \to q_\alpha$ and $L_{\lceil n\alpha\rceil:n-1} \to q_\alpha$ w.p.1 as $n \to \infty$. Because $u_i(t)$, $i = 1, 2, 3$, are all continuous at $t = q_\alpha$, by the continuous mapping theorem (Durrett 1995), $u_1(L_{\lceil n\alpha\rceil-1:n-1}) \to u_1(q_\alpha)$, $u_2(L_{\lceil n\alpha\rceil:n-1}) \to u_2(q_\alpha)$, $u_3(L_{\lceil n\alpha\rceil:n-1}) \to u_3(q_\alpha)$, and $u_3(L_{\lceil n\alpha\rceil-1:n-1}) \to u_3(q_\alpha)$ w.p.1 as $n \to \infty$. Furthermore, note that $|u_1(t)| \le \mathrm{E}[|Y_1(t)|]$ for any t. By the definition of $Y_1(t)$ and Equation (3),

$$|u_1(t)| \le \mathrm{E}(K_1) < \infty \qquad (15)$$

for any t. Then, $|u_1(L_{\lceil n\alpha\rceil-1:n-1})| \le \mathrm{E}(K_1)$. Then by the dominated convergence theorem (Durrett 1995), $\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-1})] \to \mathrm{E}[u_1(q_\alpha)] = u_1(q_\alpha)$ as $n \to \infty$. Similarly, we can also prove that $\mathrm{E}[u_2(L_{\lceil n\alpha\rceil:n-1})] \to u_2(q_\alpha)$, $\mathrm{E}[u_3(L_{\lceil n\alpha\rceil:n-1})] \to u_3(q_\alpha)$, and $\mathrm{E}[u_3(L_{\lceil n\alpha\rceil-1:n-1})] \to u_3(q_\alpha)$ as $n \to \infty$. Therefore, by Equation (14), $\mathrm{E}[\bar{Y}(\hat{q}_\alpha)] \to \mathrm{E}[Y_1(q_\alpha)]$ as $n \to \infty$. □

By the symmetry between $Y_i(\hat{q}_\alpha)$ and $Y_j(\hat{q}_\alpha)$ for any $i \ne j$, we have

$$\begin{aligned}&\mathrm{Var}[\bar{Y}(\hat{q}_\alpha)] \\ &\quad = \frac{1}{n^2}\left\{\sum_{i=1}^{n}\mathrm{Var}[Y_i(\hat{q}_\alpha)] + \sum_{i=1}^{n}\sum_{j=1,\, j\ne i}^{n}\mathrm{Cov}[Y_i(\hat{q}_\alpha), Y_j(\hat{q}_\alpha)]\right\} \\ &\quad = \frac{1}{n}\mathrm{Var}[Y_1(\hat{q}_\alpha)] + \left(1 - \frac{1}{n}\right)\mathrm{Cov}[Y_1(\hat{q}_\alpha), Y_2(\hat{q}_\alpha)]. \qquad (16)\end{aligned}$$

Furthermore, note that

$$\begin{aligned}
&\mathrm{E}[Y_1(\hat{q}_\alpha)\cdot Y_2(\hat{q}_\alpha)]\\
&\quad = \mathrm{E}[Y_1(\hat{q}_\alpha)(1\{L_1<\hat{q}_\alpha\}+1\{L_1=\hat{q}_\alpha\}+1\{L_1>\hat{q}_\alpha\})\\
&\qquad \cdot Y_2(\hat{q}_\alpha)(1\{L_2<\hat{q}_\alpha\}+1\{L_2=\hat{q}_\alpha\}+1\{L_2>\hat{q}_\alpha\})]\\
&\quad = \mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1<\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2<\hat{q}_\alpha\}]\\
&\qquad +2\ \mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1<\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2>\hat{q}_\alpha\}]\\
&\qquad +\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1>\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2>\hat{q}_\alpha\}]\\
&\qquad +2\ \mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)]\\
&\qquad -\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2=\hat{q}_\alpha\}]. \qquad (17)
\end{aligned}$$

If $L_1<\hat{q}_\alpha$ and $L_2<\hat{q}_\alpha$, then $\hat{q}_\alpha=L_{\lceil n\alpha\rceil-2:n-2}$, where $L_{\lceil n\alpha\rceil-2:n-2}$ is an order statistic of $L_3, L_4, \ldots, L_n$ and is independent of $(L_1, Y_1(t))$ and $(L_2, Y_2(t))$. Similarly, if $L_1<\hat{q}_\alpha$ and $L_2>\hat{q}_\alpha$, $\hat{q}_\alpha=L_{\lceil n\alpha\rceil-1:n-2}$; if $L_1>\hat{q}_\alpha$ and $L_2>\hat{q}_\alpha$, $\hat{q}_\alpha=L_{\lceil n\alpha\rceil:n-2}$. Then,

$$\begin{aligned}
&\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1<\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2<\hat{q}_\alpha\}]\\
&\quad =\mathrm{E}[Y_1(L_{\lceil n\alpha\rceil-2:n-2})1\{L_1<L_{\lceil n\alpha\rceil-2:n-2}\}\\
&\qquad \cdot Y_2(L_{\lceil n\alpha\rceil-2:n-2})1\{L_2<L_{\lceil n\alpha\rceil-2:n-2}\}]\\
&\quad =\mathrm{E}[u_1^2(L_{\lceil n\alpha\rceil-2:n-2})].
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1<\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2>\hat{q}_\alpha\}]\\
&\quad =\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-2})\cdot u_2(L_{\lceil n\alpha\rceil-1:n-2})],\\
&\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1>\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2>\hat{q}_\alpha\}]=\mathrm{E}[u_2^2(L_{\lceil n\alpha\rceil:n-2})].
\end{aligned}$$

Furthermore, because L is a continuous random variable, $\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2=\hat{q}_\alpha\}]=0$. Therefore, by Equation (17),

$$\begin{aligned}
&\mathrm{E}[Y_1(\hat{q}_\alpha)\cdot Y_2(\hat{q}_\alpha)]\\
&\quad =\mathrm{E}[u_1^2(L_{\lceil n\alpha\rceil-2:n-2})]+2\ \mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-2})\cdot u_2(L_{\lceil n\alpha\rceil-1:n-2})]\\
&\qquad +\mathrm{E}[u_2^2(L_{\lceil n\alpha\rceil:n-2})]+2\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)].
\end{aligned}$$

Because $\mathrm{E}^2[Y_1(q_\alpha)]=[u_1(q_\alpha)+u_2(q_\alpha)]^2=u_1^2(q_\alpha)+2u_1(q_\alpha)\cdot u_2(q_\alpha)+u_2^2(q_\alpha)$,

$$\begin{aligned}
&\mathrm{Cov}[Y_1(\hat{q}_\alpha), Y_2(\hat{q}_\alpha)]\\
&\quad =\mathrm{E}[Y_1(\hat{q}_\alpha)\cdot Y_2(\hat{q}_\alpha)]-\mathrm{E}[Y_1(\hat{q}_\alpha)]\cdot \mathrm{E}[Y_2(\hat{q}_\alpha)]\\
&\quad =\{\mathrm{E}[Y_1(\hat{q}_\alpha)\cdot Y_2(\hat{q}_\alpha)]-\mathrm{E}^2[Y_1(q_\alpha)]\}-\{\mathrm{E}^2[Y_1(\hat{q}_\alpha)]-\mathrm{E}^2[Y_1(q_\alpha)]\}\\
&\quad =\mathrm{E}[u_1^2(L_{\lceil n\alpha\rceil-2:n-2})-u_1^2(q_\alpha)]\\
&\qquad +2\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-2})\cdot u_2(L_{\lceil n\alpha\rceil-1:n-2})-u_1(q_\alpha)u_2(q_\alpha)]\\
&\qquad +\mathrm{E}[u_2^2(L_{\lceil n\alpha\rceil:n-2})-u_2^2(q_\alpha)]+2\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}Y_2(\hat{q}_\alpha)]\\
&\qquad -\{\mathrm{E}^2[Y_1(\hat{q}_\alpha)]-\mathrm{E}^2[Y_1(q_\alpha)]\}. \qquad (18)
\end{aligned}$$

Then, we can prove the following proposition on $\mathrm{Var}[\bar{Y}(\hat{q}_\alpha)]$.

Proposition 2. *Suppose that* $\mathrm{E}(K_j^{2+\delta})<\infty$ *for some* $\delta>0$ *and for both* $j=1,2$, *and that* $u_i(t)$, $i=1,2,3$, *are all continuous at* $t=q_\alpha$. *Then,* $\mathrm{Var}[\bar{Y}(\hat{q}_\alpha)]\to 0$ *as* $n\to\infty$.

Proof. By the definition of $Y_1(t)$ and Equation (3), $|Y_1(t)|\le K_1$. Then,

$$\mathrm{Var}[Y_1(\hat{q}_\alpha)]\le \mathrm{E}[Y_1^2(\hat{q}_\alpha)]\le \mathrm{E}(K_1^2)<\infty. \qquad (19)$$

Similar to the proof of Proposition 1, by the continuous mapping theorem and the dominated convergence theorem, we have

$$\mathrm{E}[u_1^2(L_{\lceil n\alpha\rceil-2:n-2})-u_1^2(q_\alpha)]\to 0 \quad \text{as } n\to\infty, \qquad (20)$$

$$\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-2})\cdot u_2(L_{\lceil n\alpha\rceil-1:n-2})-u_1(q_\alpha)u_2(q_\alpha)]\to 0 \quad \text{as } n\to\infty, \qquad (21)$$

$$\mathrm{E}[u_2^2(L_{\lceil n\alpha\rceil:n-2})-u_2^2(q_\alpha)]\to 0 \quad \text{as } n\to\infty. \qquad (22)$$

By Hölder's inequality,

$$\begin{aligned}
&\mathrm{E}[|Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}Y_2(\hat{q}_\alpha)|]\\
&\quad \le \{\mathrm{E}[|Y_1(\hat{q}_\alpha)Y_2(\hat{q}_\alpha)|^{1+\gamma}]\}^{1/(1+\gamma)}\cdot[\Pr\{L_1=\hat{q}_\alpha\}]^{\gamma/(1+\gamma)}
\end{aligned}$$

for any $\gamma>0$. By Equation (3) and the definition of $Y_1(t)$, we have $|Y_1(\hat{q}_\alpha)Y_2(\hat{q}_\alpha)1\{L_2<\hat{q}_\alpha\}|\le K_1^2$. Also, note that $\Pr\{L_1=\hat{q}_\alpha\}=1/n$ because all n observations of L have equal probability to be the $\lceil n\alpha\rceil$th smallest value. Let $\gamma=\delta/2$. We have

$$\mathrm{E}[|Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}Y_2(\hat{q}_\alpha)|]\le\{\mathrm{E}(K_1^{2+\delta})\}^{2/(2+\delta)}\left(\frac{1}{n}\right)^{\delta/(2+\delta)}.$$

Because $\mathrm{E}(K_1^{2+\delta})<\infty$,

$$\mathrm{E}[|Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}Y_2(\hat{q}_\alpha)|]\to 0 \quad \text{as } n\to\infty. \qquad (23)$$

By Proposition 1, we also have

$$\mathrm{E}^2[Y_1(\hat{q}_\alpha)]-\mathrm{E}^2[Y_1(q_\alpha)]\to 0 \quad \text{as } n\to\infty. \qquad (24)$$

Therefore, by Equation (18) and Equations (20) to (24), we have

$$\mathrm{Cov}[Y_1(\hat{q}_\alpha), Y_2(\hat{q}_\alpha)]\to 0 \quad \text{as } n\to\infty. \qquad (25)$$

By Equations (16), (19), and (25), we have $\mathrm{Var}[\bar{Y}(\hat{q}_\alpha)]\to 0$ as $n\to\infty$. □

For any $\epsilon>0$, by Chebyshev's inequality (Durrett 1995),

$$\Pr\{|\bar{Y}(\hat{q}_\alpha)-\mathrm{E}[\bar{Y}(\hat{q}_\alpha)]|>\epsilon\}\le\frac{\mathrm{Var}[\bar{Y}(\hat{q}_\alpha)]}{\epsilon^2}.$$

By Proposition 2, $\mathrm{Var}[\bar{Y}(\hat{q}_\alpha)]\to 0$ as $n\to\infty$. Then, $\Pr\{|\bar{Y}(\hat{q}_\alpha)-\mathrm{E}[\bar{Y}(\hat{q}_\alpha)]|>\epsilon\}\to 0$ as $n\to\infty$ for any $\epsilon>0$. Therefore, $\bar{Y}(\hat{q}_\alpha)-\mathrm{E}[\bar{Y}(\hat{q}_\alpha)]\xrightarrow{P} 0$ as $n\to\infty$. By Proposition 1, we also have $\mathrm{E}[\bar{Y}(\hat{q}_\alpha)]-\mathrm{E}[Y_1(q_\alpha)]\to 0$ as $n\to\infty$. Then, $\bar{Y}(\hat{q}_\alpha)\xrightarrow{P}\mathrm{E}[Y_1(q_\alpha)]$ as $n\to\infty$. This completes the proof of Theorem 2. □

A.2. Proof of Theorem 3

We need the following lemma to prove the theorem.

LEMMA 1. *Let r and s be any fixed integers and $f(t)$ denote the density of L. Suppose that $f(t)$ is continuously differentiable at $t=q_\alpha$ and $f(q_\alpha)>0$. Then,* $\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)=O(1/n)$ *and* $\mathrm{E}[(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)^2]=O(1/n)$.

PROOF. Let $p_n=(\lceil n\alpha\rceil-r)/(n-s+1)$ and F^{-1} be the inverse cumulative distribution function of L. Then, by Equations (4.6.3) and (4.6.4) of David and Nagaraja (2003),

$$\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-s})=F^{-1}(p_n)+\frac{p_n(1-p_n)}{2(n-s+2)}(F^{-1})''(p_n)+o\left(\frac{1}{n-s}\right),$$

$$\mathrm{Var}(L_{\lceil n\alpha\rceil-r:n-s})=\frac{p_n(1-p_n)}{n-s+2}[(F^{-1})'(p_n)]^2+o\left(\frac{1}{n-s}\right).$$

Note that

$$(F^{-1})'(p_n)=\frac{1}{f(F^{-1}(p_n))},\qquad (F^{-1})''(p_n)=-\frac{f'(F^{-1}(p_n))}{f^3(F^{-1}(p_n))},$$

and $p_n\to\alpha$ as $n\to\infty$. Then, $\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-s})=F^{-1}(p_n)+O(1/n)$ and $\mathrm{Var}(L_{\lceil n\alpha\rceil-r:n-s})=O(1/n)$.

Note that

$$\begin{aligned}F^{-1}(p_n)-q_\alpha&=F^{-1}(p_n)-F^{-1}(\alpha)\\&=(F^{-1})'(\alpha)(p_n-\alpha)+o(|p_n-\alpha|),\end{aligned}$$

and

$$\begin{aligned}|p_n-\alpha|&=\left|\frac{\lceil n\alpha\rceil-n\alpha}{n-s+1}+\frac{s\alpha-\alpha-r}{n-s+1}\right|\\&\le\frac{1+|s\alpha-\alpha-r|}{n-s+1}=O(1/n).\end{aligned}$$

Then, $F^{-1}(p_n)-q_\alpha=O(1/n)$. Because

$$\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-sq})=F^{-1}(p_n)+O(1/n),$$

$\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)=O(1/n)$.

Note that

$$\mathrm{E}[(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)^2]=\mathrm{Var}(L_{\lceil n\alpha\rceil-r:n-s})+[\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)]^2.$$

Because $\mathrm{Var}(L_{\lceil n\alpha\rceil-r:n-s})=O(1/n)$ and $\mathrm{E}(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)=O(1/n)$, $\mathrm{E}[(L_{\lceil n\alpha\rceil-r:n-s}-q_\alpha)^2]=O(1/n)$. This concludes the proof of the lemma. □

We have the following proposition on the rate of convergence of $\mathrm{E}[\overline{Y}(\hat{q}_\alpha)]$ and $\mathrm{E}[\overline{Z}(\hat{q}_\alpha)]$.

PROPOSITION 3. *Suppose that $f(t)$ is continuously differentiable at $t=q_\alpha$ and $f(q_\alpha)>0$, and suppose that $u_i(t)$ and $v_i(t)$, $i=1,2,3$, are all twice differentiable functions of t in a neighborhood of $t=q_\alpha$, and $|u_i''(t)|\le M$ and $|v_i''(t)|\le M$ for some constant $M>0$ for all $i=1,2,3$. Then,* $\mathrm{E}[\overline{Y}(\hat{q}_\alpha)]-\mathrm{E}[Y_1(q_\alpha)]=O(n^{-1})$ *and* $\mathrm{E}[\overline{Z}(\hat{q}_\alpha)]-\mathrm{E}[Z_1(q_\alpha)]=O(n^{-1})$.

PROOF. By Equation (14),

$$\begin{aligned}&\mathrm{E}[\overline{Y}(\hat{q}_\alpha)]-\mathrm{E}[Y_1(q_\alpha)]\\&\quad=\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-1})-u_1(q_\alpha)]+\mathrm{E}[u_2(L_{\lceil n\alpha\rceil:n-1})-u_2(q_\alpha)]\\&\qquad+\mathrm{E}[u_3(L_{\lceil n\alpha\rceil:n-1})-u_3(L_{\lceil n\alpha\rceil-1:n-1})].\end{aligned}\tag{26}$$

By Taylor's theorem,

$$\begin{aligned}&u_1(L_{\lceil n\alpha\rceil-1:n-1})-u_1(q_\alpha)\\&\quad=u_1'(q_\alpha)(L_{\lceil n\alpha\rceil-1:n-1}-q_\alpha)+u_1''(\xi)(L_{\lceil n\alpha\rceil-1:n-1}-q_\alpha)^2,\end{aligned}$$

where ξ is a random variable between $L_{\lceil n\alpha\rceil-1:n-1}$ and q_α. Note that, by Lemma 1, $\mathrm{E}(L_{\lceil n\alpha\rceil-1:n-1}-q_\alpha)$ and $\mathrm{E}[(L_{\lceil n\alpha\rceil-1:n-1}-q_\alpha)^2]$ are of $O(n^{-1})$. Because $|u_1''(\xi)|\le M$ for a constant $M>0$, $\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-1})-u_1(q_\alpha)]=O(n^{-1})$. Similarly, all three other terms on the right-hand side of Equation (26) are also of $O(n^{-1})$. Therefore, $\mathrm{E}[\overline{Y}(\hat{q}_\alpha)]-\mathrm{E}[Y_1(q_\alpha)]=O(n^{-1})$.

Similarly, we can also prove that $\mathrm{E}[\overline{Z}(\hat{q}_\alpha)]-\mathrm{E}[Z_1(q_\alpha)]=O(n^{-1})$. □

We have the following proposition on the rate of convergence of $\mathrm{Var}[\overline{Y}(\hat{q}_\alpha)]$ and $\mathrm{Var}[\overline{Z}(\hat{q}_\alpha)]$.

PROPOSITION 4. *Suppose that $f(t)$ is continuously differentiable at $t=q_\alpha$ and $f(q_\alpha)>0$ and $\mathrm{E}(K^2)<\infty$, and suppose that $u_i(t)$ and $v_i(t)$, $i=1,\ldots,5$, are all twice differentiable functions of t in a neighborhood of $t=q_\alpha$, $|u_i'(t)|\le M$, $|u_i''(t)|\le M$, $|v_i'(t)|\le M$, and $|v_i''(t)|\le M$ for all $t\in\Re$ for some constant $M>0$ for all $i=1,\ldots,5$. Then,* $\mathrm{Var}[\overline{Y}(\hat{q}_\alpha)]=O(n^{-1})$ *and* $\mathrm{Var}[\overline{Z}(\hat{q}_\alpha)]=O(n^{-1})$.

PROOF. Note that

$$\mathrm{Var}[\overline{Y}(\hat{q}_\alpha)]=\frac{1}{n}\mathrm{Var}[Y_1(\hat{q}_\alpha)]+\left(1-\frac{1}{n}\right)\mathrm{Cov}[Y_1(\hat{q}_\alpha),Y_2(\hat{q}_\alpha)],$$

and $(1/n)\mathrm{Var}[Y_1(\hat{q}_\alpha)]=O(n^{-1})$ by Equation (19). Then, to prove $\mathrm{Var}[\overline{Y}(\hat{q}_\alpha)]=O(n^{-1})$, it is sufficient to prove that $\mathrm{Cov}[Y_1(\hat{q}_\alpha),Y_2(\hat{q}_\alpha)]=O(n^{-1})$.

Note that, by Taylor's theorem,

$$\begin{aligned}&u_1^2(L_{\lceil n\alpha\rceil-2:n-2})-u_1^2(q_\alpha)\\&\quad=2u_1(q_\alpha)u_1'(q_\alpha)(L_{\lceil n\alpha\rceil-2:n-2}-q_\alpha)\\&\qquad+2[(u_1'(\xi))^2+u_1(\xi)u_1''(\xi)](L_{\lceil n\alpha\rceil-2:n-2}-q_\alpha)^2,\end{aligned}$$

where ξ is a random variable between $L_{\lceil n\alpha\rceil-2:n-2}$ and q_α. Because both $\mathrm{E}(L_{\lceil n-2\alpha\rceil:n-2}-q_\alpha)=O(n^{-1})$ and $\mathrm{E}[(L_{\lceil n-2\alpha\rceil:n-2}-q_\alpha)^2]=O(n^{-1})$ by Lemma 1, then $\mathrm{E}[u_1^2(L_{\lceil n\alpha\rceil-2:n-2})-u_1^2(q_\alpha)]=O(n^{-1})$ if both $|u_1'(t)|$ and $|u_1''(t)|$ are bounded for any t. Similarly, we can also show that $\mathrm{E}[u_1(L_{\lceil n\alpha\rceil-1:n-2})\cdot u_2(L_{\lceil n\alpha\rceil-1:n-2})-u_1(q_\alpha)u_2(q_\alpha)]=O(n^{-1})$ and $\mathrm{E}[u_2^2(L_{\lceil n\alpha\rceil:n-2})-u_2^2(q_\alpha)]=O(n^{-1})$ under similar conditions.

By the definitions of $u_4(t)$ and $u_5(t)$, we have

$$\begin{aligned}&\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}Y_2(\hat{q}_\alpha)]\\&\quad=\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2<\hat{q}_\alpha\}]\\&\qquad+\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}\cdot Y_2(\hat{q}_\alpha)1\{L_2>\hat{q}_\alpha\}]\\&\quad=\mathrm{E}[Y_1(L_1)Y_2(L_1)1\{L_1>L_2\}1\{L_{\lceil n\alpha\rceil-2:n-2}<L_1<L_{\lceil n\alpha\rceil:n-2}\}]\\&\qquad+\mathrm{E}[Y_1(L_1)Y_2(L_1)1\{L_1<L_2\}1\{L_{\lceil n\alpha\rceil-1:n-2}<L_1<L_{\lceil n\alpha\rceil:n-2}\}]\\&\quad=\mathrm{E}[u_4(L_{\lceil n\alpha\rceil:n-2})-u_4(L_{\lceil n\alpha\rceil-2:n-2})]\\&\qquad+\mathrm{E}[u_5(L_{\lceil n\alpha\rceil:n-2})-u_5(L_{\lceil n\alpha\rceil-1:n-2})].\end{aligned}$$

If $u_4(t)$ and $u_5(t)$ have bounded second derivatives, then $\mathrm{E}[Y_1(\hat{q}_\alpha)1\{L_1=\hat{q}_\alpha\}Y_2(\hat{q}_\alpha)]=O(n^{-1})$. By Proposition 3, it is easy to see that $\mathrm{E}^2[Y_1(\hat{q}_\alpha)]-\mathrm{E}^2[Y_1(q_\alpha)]=O(n^{-1})$. Hence, all five terms on the right-hand side of Equation (18) are of $O(n^{-1})$. Then, $\mathrm{Cov}[Y_1(\hat{q}_\alpha),Y_2(\hat{q}_\alpha)]=O(n^{-1})$. Therefore, $\mathrm{Var}[\overline{Y}(\hat{q}_\alpha)]=O(n^{-1})$.

We may use the same approach to prove $\mathrm{Var}[\overline{Z}(\hat{q}_\alpha)] = O(n^{-1})$. □

Because $\mathrm{Var}\{\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[\overline{Y}(\hat{q}_\alpha]\} = O(n^{-1})$ by Proposition 4, by Chebyshev's inequality, $\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[\overline{Y}(\hat{q}_\alpha] = O_p(n^{-1/2})$. By Proposition 3, $\mathrm{E}[\overline{Y}(\hat{q}_\alpha] - \mathrm{E}[Y_1(q_\alpha] = O(n^{-1})$. Because

$$\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[Y_1(q_\alpha] = \{\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[\overline{Y}(\hat{q}_\alpha]\} + \{\mathrm{E}[\overline{Y}(\hat{q}_\alpha] - \mathrm{E}[Y_1(q_\alpha]\},$$

$$\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[Y_1(q_\alpha] = O_p(n^{-1/2}). \tag{27}$$

Similarly, we can show that

$$\overline{Z}(\hat{q}_\alpha) - \mathrm{E}[Z_1(q_\alpha] = O_p(n^{-1/2}). \tag{28}$$

By Equation (7),

$$\begin{aligned} &\hat{q}'_\alpha(\theta) - q'_\alpha(\theta) \\ &= \frac{\mathrm{E}[Y_1(q_\alpha)]}{\mathrm{E}[Z_1(q_\alpha)]} - \frac{\overline{Y}(\hat{q}_\alpha)}{\overline{Z}(\hat{q}_\alpha)} \\ &= \frac{\mathrm{E}[Y_1(q_\alpha)](\overline{Z}(\hat{q}_\alpha) - \mathrm{E}[Z_1(q_\alpha)]) - \mathrm{E}[Z_1(q_\alpha)](\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[Y_1(q_\alpha)])}{\mathrm{E}[Z_1(q_\alpha)] \cdot \overline{Z}(\hat{q}_\alpha)}. \end{aligned} \tag{29}$$

By Equations (27) and (28),

$$\mathrm{E}[Y_1(q_\alpha)](\overline{Z}(\hat{q}_\alpha) - \mathrm{E}[Z_1(q_\alpha)]) - \mathrm{E}[Z_1(q_\alpha)](\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[Y_1(q_\alpha)]) = O_p(n^{-1/2}).$$

Because $\mathrm{E}[Z_1(q_\alpha)] \cdot \overline{Z}(\hat{q}_\alpha) \xrightarrow{P} \mathrm{E}^2[Z_1(q_\alpha)]$ as $n \to \infty$ by Theorem 2, by Bhat (1985, p. 130),

$$\frac{\mathrm{E}[Y_1(q_\alpha)](\overline{Z}(\hat{q}_\alpha) - \mathrm{E}[Z_1(q_\alpha)]) - \mathrm{E}[Z_1(q_\alpha)](\overline{Y}(\hat{q}_\alpha) - \mathrm{E}[Y_1(q_\alpha)])}{\mathrm{E}[Z_1(q_\alpha)] \cdot \overline{Z}(\hat{q}_\alpha)} = O_p(n^{-1/2}).$$

Therefore, by Equation (29), $\hat{q}'_\alpha - q'_\alpha = O_p(n^{-1/2})$. This completes the proof of Theorem 3. □

References

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

Bhat, B. R. 1985. *Modern Probability Theory: An Introductory Textbook*. Wiley, New York.

David, H. A., H. N. Nagaraja. 2003. *Order Statistics*, 3rd ed. Wiley, Hoboken, NJ.

Durrett, R. 1995. *Probability: Theory and Examples*, 2nd ed. Duxbury, Belmont, CA.

Fu, M. C. 2006. Gradient estimation. S. G. Henderson, B. L. Nelson, eds. *Handbooks in Operations Research and Management Science: Simulation*, Chap. 19. Elsevier, Amsterdam, 575–616.

Fu, M. C. 2008. What you should know about simulation and derivatives. *Naval Res. Logist.* **55** 723–736.

Fu, M. C., J.-Q. Hu. 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer, Norwell, MA.

Glasserman, P. 1991. *Gradient Estimation Via Perturbation Analysis*. Kluwer, Norwell, MA.

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

Ho, Y. C., X. R. Cao. 1991. *Perturbation Analysis and Discrete Event Dynamic Systems*. Kluwer, Boston.

Hong, L. J. 2009. Estimating quantile sensitivities. *Oper. Res.* **57** 118–130.

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

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

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

Liu, G., L. J. Hong. 2009. Kernel estimation of quantile sensitivities. *Naval Res. Logist.* **56** 511–525.

Mashal, R, A. Zeevi. 2003. Beyond correlation: Extreme co-movements between financial assets. Working paper, Columbia University, New York.

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

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