Submitted to *INFORMS Journal on Computing*
manuscript JOC-2013-02-OA-017

Estimating Sensitivities of Portfolio Credit Risk using Monte Carlo

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

Sandeep Juneja
School of Technology and Computer Science, Fundamental Research, Colaba Mumbai, India-400005
juneja@tifr.res.in

Jun Luo
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China
jluo_ms@sjtu.edu.cn

Estimating sensitivities of portfolio credit risk to the underlying model parameters is an important problem for credit risk management. In this paper, we consider performance measures that may be expressed as an expectation of a performance function of the portfolio credit loss and derive closed-form expressions of its sensitivities to the underlying parameters. Our results are applicable to both idiosyncratic and macroeconomic parameters and to performance functions that may or may not be continuous. Based on the closed-form expressions, we first develop an estimator for sensitivities, in a very general framework, that relies on the kernel method for estimation. The unified estimator allows us to further derive two general forms of the estimators by using conditioning techniques on either idiosyncratic or macroeconomic factors. We then specialize our results to develop faster estimators for three popular classes of models used for portfolio credit risk, namely, latent variable models, Bernoulli mixture models, and doubly stochastic models.

Key words: sensitivity estimation; Monte Carlo simulation; conditioning techniques

History: Accepted by Marvin Nakayama, Area Editor for Simulation; received February 2013; revised July 2013, January 2014; accepted January 2014.

---

1. Introduction

Large financial institutions are exposed to multiple sources of credit risk. A portfolio approach is then needed to accurately measure and manage the overall credit exposure. This need has become especially urgent in view of the ongoing global financial crisis triggered, amongst other reasons, by poor risk management by some of the largest financial institutions. Some of the key issues then include: coming up with an accurate model for measuring portfolio risk; selecting appropriate performance measures to capture portfolio risk; and devising effective portfolio risk hedging strategies. Computing sensitivities of the selected performance measure to underlying parameters, is crucial to risk analysis and management and to the related risk hedging problem. This is the issue that we address in this paper for a large class of portfolio risk performance measures.

Accurately modeling dependence between default events is a key issue in selecting a model for portfolio credit risk. Three types of models that are commonly used in practice are latent variable models, Bernoulli mixture models and doubly stochastic models. In latent variable models, default happens if a random variable (called a latent variable) falls below a certain threshold, and the dependence between default of different firms is modeled by the dependence between their respective latent variables. These models are motivated by the seminal firm-value work by Merton (1974). These are used in commercial products such as J.P. Morgan's CreditMetrics (Gupton et al. 1997), Moody's KMV system (Kealhofer and Bohn 2001). It can be shown that the Gaussian copula model proposed by Li (2000), which is used commonly in pricing credit derivatives, is a special instance of the latent variable models. In Bernoulli mixture models, default probabilities of individual obligors depend on each other through a common set of macroeconomic factors. Conditioned on these factors, default events are independent Bernoulli random variables. These models are used in CreditRisk$^+$, a product developed by Credit Suisse Financial Products (Credit-Suisse-Financial-Products 1997). In doubly stochastic models, also known as Cox process models, default occurs at the first jump time of a doubly stochastic process with a nonnegative intensity process. The dependence between obligors can be captured by the dependence between the intensity processes (see, e.g., Duffie and Singleton 1999).

Given a model of joint defaults, portfolio credit loss is a random variable denoting the sum of losses caused by all default obligors. In many situations, to manage credit risk or to price credit derivatives, we are interested in performance measures that can be written as expectations of some performance functions of the credit loss. For instance, the probability that the loss is beyond a certain threshold and the average loss when it is beyond the threshold are related to popular risk measures such as value-at-risk and tail conditional expectation; the expected loss when it is within a certain range (called a tranche) is the building block that defines the price of a collateralized debt obligation (CDO), a popular credit derivative.

In a model of joint defaults, there are often many parameters. Some of them affect only individual obligors, such as the parameters related to an obligor's idiosyncratic risk; some affect all obligors in the portfolio, such as the parameters related to macroeconomic factors. Any changes to these parameters may affect the portfolio credit loss and, thus, the performance measures that we are interested in. An approach to characterizing the impact of parameter changes is to calculate sensitivities, which are the first-order derivatives of the performance measures with respect to the parameters. Sensitivities provide information on how small movements of parameters affect performance measures. This information is important in risk analysis and portfolio management. For instance, delta hedging is one of the most fundamental tools in portfolio risk management, where deltas are the sensitivities of portfolio value with respect to underlying risk factors. In this paper, we consider the estimation of sensitivities of portfolio credit risk with respect to different parameters, either idiosyncratic or macroeconomic. We focus on the three different credit risk models mentioned earlier, namely, latent variable, Bernoulli mixture and doubly stochastic models. The results that we develop may be applicable to a broader set of credit risk models. However, we illustrate the key ideas through these three popular models. Fast and accurate sensitivity estimators for wide classes of default models can help analysts better measure and control portfolio credit risk, which has been becoming more and more important in the wake of the current global financial crisis.

Derivative estimation for expectations is a classic problem in Monte Carlo simulation. Besides finite-difference approximations (see, for instance, Section 7.1 of Glasserman (2004)), there are three main approaches in the simulation literature: perturbation analysis (PA), the likelihood ratio/score function (LR/SF) method, and weak derivatives (WD). PA was first proposed by Ho and Cao (1983) to study discrete-event systems. It interchanges the order of differentiation and expectation and estimates the expectation of the pathwise derivative (see Glasserman (1991) for a comprehensive introduction). To apply PA, however, the function inside the expectation needs to be stochastically Lipschitz continuous with respect to the parameter of interest. This greatly limits the applicability of PA because many functions, e.g., indicator functions, are not Lipschitz continuous. Remedies have been proposed to solve this problem. One approach is to apply conditional Monte Carlo method to smooth the discontinuous function. This method is known as smoothed PA (SPA, see, for instance, Fu and Hu (1997)). Instead of differentiating the function inside the expectation as in PA, the LR method differentiates the probability measure (see, for instance, Glynn (1987) and Rubinstein (1989)). It does not require the function inside the expectation be continuous. Therefore, it is generally more applicable than PA. However, LR estimators often have higher variances compared to PA estimators when both are applicable. The WD approach dates back to at least Pflug (1988). It is similar to the LR method, except that it represents the derivative of the measure as the difference of two (new) measures. Then, the derivative becomes the difference of two new expectations which can be estimated using sample means. Recently, WD approach has been further extended to a more general differentiation approach, known as measure-valued differentiation (see, for instance, Heidergott et al. (2010)). Like the LR method, WD approach is generally more applicable than PA, but often yields estimators with larger variance. Unlike PA and LR method, additional simulations may be necessary in order to implement WD estimators. For more comprehensive reviews of the different methods for estimating derivatives, readers are referred to L'Ecuyer (1991) and Fu (2008).

Although most derivative estimation approaches were proposed to analyze dynamic systems, such as queueing systems, some were also applied to financial applications. Fu and Hu (1995) and Broadie and Glasserman (1996) are among the early works that use Monte Carlo methods to estimate price sensitivities of financial options. The same problem has also been studied by combining Malliavin calculus and the Monte Carlo method (see, for instance, Bernis et al. (2003)). Chen and Glasserman (2007) showed that the Malliavin calculus approach can be viewed as a combination of PA and LR methods. Sensitivities of risk measures, such as value-at-risk (VaR) and conditional VaR (CVaR), have also been studied recently by Hong (2009), Hong and Liu (2009) and Fu et al. (2009). Estimating price sensitivities for portfolio credit derivatives, which is closely related to our work, has also been studied. Joshi and Kainth (2004) considered the nth-to-default credit swaps under the Gaussian copula model of Li (2000). Chan and Joshi (2012) derived finite proxy schemes, which can be viewed as a combination of pathwise method and LR method, to study the Greeks when the payoff function may be discontinuous. Chen and Glasserman (2008) generalized the problem and considered different types of portfolio credit derivatives. They used both the LR method and SPA. There is some overlap between the problems we study in this paper and those studied by Chen and Glasserman (2008). First, if we consider only idiosyncratic parameters, their LR method is applicable to our problem. Second, if we consider only idiosyncratic parameters with Lipschitz continuous functions, their SPA method is applicable to our problem. In our paper, however, we consider both idiosyncratic and macroeconomic factors and performance functions with general forms, and we focus on the use of pathwise derivatives in the estimation, and compare our method with the LR method when applicable. In addition, it is important to note that Chen and Glasserman (2008) also solved problems that do not fit into our framework, thus cannot be solved by our approach. For example, they consider the default credit swaps where the payoffs depend on the order of defaults, which cannot be written in the portfolio loss function specified in this paper. In fact, both Joshi and Kainth (2004) and Chen and Glasserman (2008) arrived the same pathwise estimator for the default credit swaps, where the first one uses the delta functions and the second uses the smoothing technique. We discuss more about the difference between SPA and our method in Section 3.4.

In this paper, we consider performance measures that can be written as the expected value of a performance function of the portfolio credit loss, and are interested in estimating sensitivities of the performance measures with respect to parameters of default models. We make the following contributions. First, we derive a closed-form expression of the sensitivities, which applies to both idiosyncratic and macroeconomic parameters and to functions that may or may not be continuous, and interestingly, we find that the differentiability of the performance measure does not depend on the continuity and differentiability of the performance function. Second, we derive fast and efficient sensitivity estimators for latent variable, Bernoulli mixture and doubly stochastic models based on the closed-form expression, and test them through a number of numerical examples. Third, we show that, to estimate a sensitivity, our method can be applied to provide multiple unbiased estimators, although a-priori it is difficult to conclude which amongst them has the smallest variance. This motivates an easy characterization of the optimal linear combination of these estimators. Empirically, the linear weights need to be estimated and as expected we see that the resultant estimator performs better than the individual estimators. Last but not least, we can easily generalize our results to estimate sensitivities of VaR and CVaR when each individual loss is a continuous random variable and some regularity conditions hold (see the e-companion EC.3 for more details).

The rest of the paper is organized as follows. In Section 2, we derive a closed-form expression of the sensitivities. In Section 3 we emphasize our method can often yield multiple sample-mean estimators using the conditioning techniques, which makes it attractive and natural to consider a linear combination of the proposed estimators to obtain further improvements with almost no additional cost. We then discuss Monte Carlo estimation of sensitivities under latent variable, Bernoulli mixture and doubly stochastic models in Section 4. The numerical results for both our method and LR method are reported in Section 5, followed by conclusions in Section 6. Some lengthy discussions and extensions of related work are presented in the electronic companion.

2. General Results

Suppose that there are m obligors in a loan portfolio. We let X_i denote a random variable that determines default of obligor i. Specifically, obligor i defaults if $X_i < 0$. Note that the dependence between any two obligors, say i and j, can be modeled through the dependence between X_i and X_j. In Section 4, we show that latent variable, Bernoulli mixture and doubly stochastic models of joint default can all be incorporated into this framework. Let l_i denote the loss due to the default of obligor i. Following the literature (e.g., Chen and Glasserman (2008)), we assume that l_i are constants for all $i = 1, 2, \ldots, m$. However, our results can be generalized easily to situations where l_i are mutually independent and bounded random variables that are also independent of X_j for all $i, j = 1, 2, \ldots, m$. Then, the portfolio credit loss L can be written as

$$L = \sum_{i=1}^{m} l_i \cdot \mathbf{1}_{\{X_i < 0\}},$$

where $\mathbf{1}_{\{A\}}$ is an indicator function which equals to 1 when A is true and 0 otherwise.

Let $p = \mathrm{E}[g(L)]$ denote the performance measure that we are interested in, where $g(\cdot)$ denotes the performance function. Note that L is a discrete random variable taking values in a finite set within $[0, \sum_{i=1}^m l_i]$. If $g(x) < \infty$ for every $x \in [0, \sum_{i=1}^m l_i]$, then $\mathrm{E}[g(L)] < \infty$. Many performance measures of portfolio credit loss can be written in this form. When $g(L) = L^2$, p is the second moment of L. It can be used to compute the variance of L which is an important measure of risk. When $g(L) = \mathbf{1}_{\{L>y\}}$, p is the probability of having a large loss beyond a given threshold y. It is also an important measure of risk, and can be used to compute the portfolio value-at-risk. When $g(L) = L \cdot \mathbf{1}_{\{L>y\}}$, p is the average loss beyond a given threshold y. It is again an important measure of risk, and is closely related to the concept of tail conditional expectation. When $g(L) = (L - S_\ell) \cdot \mathbf{1}_{\{L>S_\ell\}} - (L - S_u) \cdot \mathbf{1}_{\{L>S_u\}}$, p is the expected portfolio loss in the tranche from S_ℓ to S_u. It is used to price CDOs. When $g(L) = L^\alpha \cdot \mathbf{1}_{\{L>y\}}$ for $\alpha \geq 1$, p is αth moment of truncated random variable and it can be used to model a utility-based shortfall risk. Note that, in these examples, $g(\cdot)$ may or may not be a continuous function.

Let θ be a parameter of the model of joint defaults, i.e., $X_i = X_i(\theta)$ for all $i = 1, \ldots, m$. If θ is an idiosyncratic parameter, then it only affects one of the X_is. If θ is a macroeconomic factor, then it affects all X_i. In this paper, we do not differentiate these two situations. We consider the first situation as a special case of the second. Then, the loss $L = L(\theta)$ and $p(\theta) = \mathrm{E}[g(L(\theta))]$ both are functions of θ. Our goal is to estimate $p'(\theta)$ through a Monte Carlo method. For the work on estimating p itself, especially when p is a measure of credit risk, readers are referred to Artzner et al. (1999) and Section 9 of Glasserman (2004) for a comprehensive introduction to risk measures.

We suppose that $\nu(A) = \mathrm{E}[X; Y \in A]$ is absolutely continuous with respect to the Lebesgue measure and let $\mathrm{E}[X; Y = t]$ denote the associated density evaluated at t. Then,

$$\mathrm{E}[X; Y \in A] = \mathrm{E}\left[X \cdot \mathbf{1}_{\{Y \in A\}}\right] = \int_A \mathrm{E}[X; Y = t]\, dt \tag{1}$$

for any $A \subset \Re$, where $\Re$ is the set of all real numbers. If (X, Y) has a joint density $f_{X,Y}(\cdot, \cdot)$, recall that

$$\mathrm{E}[X; Y = t] = \int_{-\infty}^{+\infty} x\, f_{X,Y}(x, t)\, dx.$$

If Y has a density $f_Y(\cdot)$, we may write

$$\mathrm{E}[X; Y = t] = f_Y(t) \cdot \mathrm{E}[X | Y = t]. \tag{2}$$

Furthermore, if Y has a density $f_Y(\cdot)$ and Y is independent of X, then $\mathrm{E}[X; Y = t] = f_Y(t) \cdot \mathrm{E}[X]$. In this paper, we need to use the following lemma of Hong and Liu (2010) on the sensitivity of a probability function.

LEMMA 1 **(Hong and Liu 2010)**. *Suppose that $X(\theta)$ is a continuous random variable at any θ in $\mathcal{N}(\theta_0)$, an open neighborhood of θ_0, it is differentiable with probability 1 (w.p.1) at any $\theta \in \mathcal{N}(\theta_0)$, and there exists a random variable $\mathcal{K}$ with $\mathrm{E}[\mathcal{K}] < \infty$ such that $|X(\theta_0 + \Delta\theta) - X(\theta_0)| \le \mathcal{K} \cdot |\Delta\theta|$ for any $\Delta\theta$ that is close enough to 0. Let $\psi(\theta,t) = \mathrm{E}[X'(\theta); X(\theta) = t]$. If $\psi(\theta,t)$ is continuous at $(\theta_0, 0)$, then*

$$\frac{d}{d\theta} \Pr\{X(\theta_0) < 0\} = -\mathrm{E}[X'(\theta_0); X(\theta_0) = 0].$$

Given Lemma 1, we can prove the following corollary, which may be viewed as an extension or generalization of the lemma.

COROLLARY 1. *Suppose that, for any $i = 1, \ldots, k$, $X_i(\theta)$ is differentiable w.p.1 at any $\theta \in \mathcal{N}(\theta_0)$, and there exists a random variable $\mathcal{K}_i$ with $\mathrm{E}[\mathcal{K}_i] < \infty$ such that $|X_i(\theta_0 + \Delta\theta) - X_i(\theta_0)| \le \mathcal{K}_i \cdot |\Delta\theta|$ for any $\Delta\theta$ that is close enough to 0. We further suppose that $X_i(\theta)$ are continuous random variables such that $\Pr\{X_i(\theta) = X_j(\theta)\} = 0$ at any fixed $\theta \in \mathcal{N}(\theta_0)$. Let $\psi_i(\theta,t) = \mathrm{E}\left[X_i'(\theta) \prod_{j=1, j\neq i}^{k} \mathbf{1}_{\{X_j(\theta) < t\}}; X_i(\theta) = t\right]$. If $\psi_i(\theta,t), i = 1, \ldots, k$, are continuous at $(\theta_0, 0)$, then*

$$\frac{d}{d\theta}\, \mathrm{E}\left[\prod_{i=1}^{k} \mathbf{1}_{\{X_i(\theta_0) < 0\}}\right] = -\sum_{i=1}^{k} \mathrm{E}\left[X_i'(\theta_0) \prod_{j=1, j\neq i}^{k} \mathbf{1}_{\{X_j(\theta_0) < 0\}}; X_i(\theta_0) = 0\right].$$

The proof of Corollary 1 is available in the e-companion EC.1. We are ready to present the main result of this paper.

THEOREM 1. *Let $p(\theta) = \mathrm{E}[g(L)]$, where $g(\cdot)$ is a general $\Re$-valued function and $L = \sum_{i=1}^{m} l_i \cdot \mathbf{1}_{\{X_i < 0\}}$ with constant l_i and random variables X_i, $i = 1, 2, \ldots, m$. Let Θ be an open subset of $\Re$. Suppose that, for any $\theta \in \Theta$ and any $i = 1, \ldots, m$,*

1. *$X_i(\theta)$ is differentiable w.p.1 and there exists a random variable $\mathcal{K}_i$, which may depend on θ, such that $\mathrm{E}[\mathcal{K}_i] < \infty$ and $|X_i(\theta + \Delta\theta) - X_i(\theta)| \le \mathcal{K}_i \cdot |\Delta\theta|$ when $|\Delta\theta|$ is close enough to zero;*
2. *$X_i(\theta)$ is a continuous random variable and $\Pr\{X_i(\theta) = X_j(\theta)\} = 0$ for all $j \neq i$;*
3. *$\psi_i(\theta, y)$ is continuous at $(\theta, 0)$ for any $a_j = 0$ or 1, where*

$$\psi_i(\theta, y) = \mathrm{E}\left[X_i'(\theta) \prod_{j=1, j\neq i}^{m} \mathbf{1}_{\{(-1)^{a_j} X_j < y\}}; X_i = y\right].$$

Then, for any $\theta \in \Theta$,

$$p'(\theta) = -\sum_{i=1}^{m} \mathrm{E}\left\{[g(L_{-i} + l_i) - g(L_{-i})] \cdot X_i'(\theta); X_i = 0\right\}, \tag{3}$$

where $L_{-i} = \sum_{j=1, j\neq i}^{m} l_j \cdot \mathbf{1}_{\{X_j < 0\}}$.

REMARK 1. Note that the conditions in Theorem 1 are essentially used to ensure that Corollary 1 can be applied to derive Equation (6). In Sections 3–4, we show that these conditions can be verified easily and are typically satisfied by latent variable, Bernoulli mixture and doubly stochastic models.

The proof of Theorem 1 is deferred to the end of this section. It is interesting to see that the conclusion of Theorem 1 does not depend on the continuity and differentiability of $g(\cdot)$. This differs from most of the PA literature, which often requires the performance function be differentiable almost surely and Lipschitz continuous. Although the result is counter-intuitive, it can be explained by Equation (5) as later shown in the proof, which implies that the value of $g(L)$ is no longer affected by θ once the values of all indicator functions are given.

Equation (3) provides a sample-mean estimator with the $n^{-1/2}$ rate of convergence. By analyzing the structures of different default models in Sections 3–4, we can transform the conditional expectations in Equation (3) to regular expectations, which yields multiple sample-mean estimators. We show that, in Sections 3–4, this task can be done quite easily based on the conclusion of Theorem 1 for latent variable models, Bernoulli mixture models and doubly stochastic models.

From Equation (3), it is clear that the computational complexity of $p'(\theta)$ for this general form depends on whether the parameter θ is idiosyncratic or macroeconomic. If θ is an idiosyncratic parameter with respect to a particular obligor i, then $p'(\theta)$ in Equation (3) can be simplified to

$$p'(\theta) = -\mathrm{E}\left\{\left[g\left(L_{-i} + l_i\right) - g\left(L_{-i}\right)\right] \cdot X_i'(\theta);\, X_i = 0\right\},$$

which may reduce the complexity. Moreover, the computational complexity of $p'(\theta)$ also depends on that for computing X_i and $X_i'(\theta)$ in Equation (3), whose closed-form expressions are model-dependent. Therefore, we discuss more about this issue with respect to different models studied in Sections 3 and 4. Readers may be referred to Homescu (2011) for a comprehensive survey on computational complexity of sensitivity measures.

Given the closed form of Equation (3), we can use a kernel estimator to estimate $p'(\theta)$ as in Hong and Liu (2010). By Equation (1), we have

$$\mathrm{E}\left[X; Y = 0\right] \approx \frac{1}{2\delta}\mathrm{E}\left[X \cdot \mathbf{1}_{\{-\delta < Y < \delta\}}\right].$$

If we have an independent and identically distributed (i.i.d.) sample of $(X_i, X_i'(\theta))$, denoted as $\{(X_{i,1}, X_{i,1}'), \ldots, (X_{i,n}, X_{i,n}')\}$, then by Equation (3), we can estimate $p'(\theta)$ by

$$\widehat{p'(\theta)} = -\frac{1}{2n\delta_n}\sum_{j=1}^{n}\sum_{i=1}^{m}\left[g\left(L_{-i,j} + l_i\right) - g\left(L_{-i,j}\right)\right] \cdot X_{i,j}' \cdot \mathbf{1}_{\{-\delta_n < X_{i,j} < \delta_n\}}, \tag{4}$$

where $L_{-i,j} = \sum_{s=1, s\neq i}^{m} l_s \cdot \mathbf{1}_{\{X_{s,j} < 0\}}$. By Hong and Liu (2010), we can easily show that, under some mild conditions, $\widehat{p'(\theta)}$ is a consistent estimator of $p'(\theta)$ if $\delta_n \to 0$ and $n\delta_n \to \infty$ as $n \to \infty$, and

$\sqrt{2n\delta_n}\left[\widehat{p'(\theta)} - p'(\theta)\right]$ converges in distribution to a normal distribution if $n\delta_n^5 \to c$ for some constant $c \geq 0$, and the optimal rate of convergence of $\widehat{p'(\theta)}$ is $n^{-2/5}$. For instance, we can choose $\delta_n = n^{-1/5}$ to achieve the optimal rate of convergence. Interested readers may be referred to Hong and Liu (2010) for more discussion on the kernel method.

The kernel estimator has two advantages. First, it is applicable to a wide class of default models. As shown in Section 4, the conclusion of Theorem 1 holds for latent variable models, Bernoulli mixture models and doubly stochastic models, and therefore, the kernel estimator can be applied to them. Even for models that are not discussed in our paper, the kernel estimator may still be applicable. Second, the kernel estimator is generally easy to use. It requires only the samples of $(X_i, X_i'(\theta)), i = 1, \ldots, m$. It does not require the users to analyze the structures of underlying default models to derive different estimators.

However, the kernel estimator of Equation (4) has a slower rate of convergence, $(n\delta_n)^{-1/2}$, than the $n^{-1/2}$ of typical Monte Carlo estimators, because the expectations in Equation (3) depend on the occurrence of probability zero events $\{X_i = 0\}$, $i = 1, \ldots, m$.

In Section 5, we compare our method with kernel method as well as the LR method. For convenience, we derive a general formula of the LR estimator for the performance measure function considered in this paper. Recall that $p(\theta) = \mathrm{E}[g(L(\theta))]$. Suppose θ can be written as a distributional parameter. Then,

$$p(\theta) = \int_{\Re^n} g(L(y)) \cdot f(y, \theta) dy,$$

where $f(y, \theta)$ is the density function which also involves the parameter θ. Then,

$$p'(\theta) = \int_{\Re^n} g(L(y)) \cdot \frac{\partial_\theta f(y, \theta)}{f(y, \theta)} \cdot f(y, \theta) dy, \; = \; \mathrm{E}\left[g(L) \cdot \mathrm{SF}\right],$$

where $\mathrm{SF} = \frac{d}{d\theta} \log\left(f(\cdot, \theta)\right)\Big|_{\theta = \theta_0}$.

One straightforward advantage of LR estimator is that it takes a very simple closed-form expression and may be easily obtained when LR method is applicable (e.g., θ can be written into some density function as a distributional parameter). However, when θ is a structural parameter rather than a distributional parameter, we may need the assumption that $g\left(L(\theta)\right)$ is differentiable w.p.1, which may not hold in this paper. Rubinstein (1992) developed so-called "push-out" method to handle this difficulty. Readers may refer to Asmussen and Glynn (2007) for a detailed discussion about LR method.

We now prove Theorem 1.

Proof of Theorem 1. In the following analysis we suppress the dependence of X_i on θ at places for presentation convenience. To analyze $p'(\theta)$, we view $\mathbf{1}_{\{X_i < 0\}}$ as a Bernoulli random variable and consider all combinations of $\mathbf{1}_{\{X_i<0\}}$, $i=1,\ldots,m$. Let $B_i=\mathbf{1}_{\{X_i<0\}}$ and $B=(B_1,\ldots,B_m)$. Because $B_i\in\{0,1\}$, B takes value from $\mathscr{S}(m)=\{0,1\}^m$ with totally 2^m elements. For each element $s\in\mathscr{S}(m)$, we let s^1 denote the set of obligors whose $B_i=1$ and s^0 denote the set of obligors whose $B_i=0$. For instance, for $s=(1,0,1,0)\in\mathscr{S}(4)$, $s^1=\{1,3\}$ and $s^0=\{2,4\}$. Note that

$$\mathbf{1}_{\{B=s\}}=\prod_{i\in s^1}\mathbf{1}_{\{X_i<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i\geq 0\}}.$$

Then,

$$\begin{aligned}
p(\theta)=\mathrm{E}\left[g\left(L\right)\right]&=\sum_{s\in\mathscr{S}(m)}\mathrm{E}\left[g\left(\sum_{i=1}^m l_i\cdot\mathbf{1}_{\{X_i<0\}}\right)\cdot\mathbf{1}_{\{B=s\}}\right]\\
&=\sum_{s\in\mathscr{S}(m)}g\left(\sum_{i\in s^1}l_i\right)\cdot\mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i\geq 0\}}\right].
\end{aligned}\tag{5}$$

Note that when l_i are random variables that are independent of X_j for all $i,j=1,2,\ldots,m$, Equation (5) can be written as,

$$\begin{aligned}
p(\theta)&=\sum_{s\in\mathscr{S}(m)}\mathrm{E}\left\{\mathrm{E}\left[g\left(\sum_{i=1}^m l_i\cdot\mathbf{1}_{\{X_i<0\}}\right)\cdot\mathbf{1}_{\{B=s\}}\middle| l_i,\ i=1,2,\ldots,m\right]\right\}\\
&=\sum_{s\in\mathscr{S}(m)}\mathrm{E}\left[g\left(\sum_{i\in s^1}l_i\right)\right]\cdot\mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i\geq 0\}}\right].
\end{aligned}$$

The conditional expectation techniques can be applied throughout the following derivations. To simplify the presentation, however, we assume that l_i are constants for all $i=1,2,\ldots,m$ throughout Sections 2–5. Under the conditions in Theorem 1, we can apply Corollary 1 to Equation (5) and have

$$\begin{aligned}
p'(\theta)&=\sum_{s\in\mathscr{S}(m)}g\left(\sum_{i\in s^1}l_i\right)\cdot\frac{d}{d\theta}\,\mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i(\theta)<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i(\theta)\geq 0\}}\right]\\
&=-\sum_{s\in\mathscr{S}(m)}g\left(\sum_{i\in s^1}l_i\right)\cdot\left\{\sum_{i\in s^1}\mathrm{E}\left[X_i'(\theta)\prod_{j\in s^1,j\neq i}\mathbf{1}_{\{X_j<0\}}\prod_{j\in s^0}\mathbf{1}_{\{X_j\geq 0\}};X_i=0\right]\right.\\
&\qquad\left.-\sum_{i\in s^0}\mathrm{E}\left[X_i'(\theta)\prod_{j\in s^1}\mathbf{1}_{\{X_j<0\}}\prod_{j\in s^0,j\neq i}\mathbf{1}_{\{X_j\geq 0\}};X_i=0\right]\right\}.
\end{aligned}\tag{6}$$

By Equation (6), it is clear that we can write $p'(\theta)=\sum_{i=1}^m\Psi_i$, where $\Psi_i=\mathrm{E}\left[A_i\cdot X_i'(\theta);X_i=0\right]$ for some A_i. Without loss of generality, we consider Ψ_m. Note that

$$\mathscr{S}(m)=\left[\mathscr{S}(m-1)\times\{1\}\right]\,\cup\,\left[\mathscr{S}(m-1)\times\{0\}\right],$$

where $B_m = 1$ in the first set and $B_m = 0$ in the second set. Then, by Equation (6), we have

$$\begin{aligned}\Psi_m =& -\sum_{s\in\mathscr{S}(m-1)\times\{1\}} g\left(\sum_{i\in s^1} l_i\right)\cdot \mathrm{E}\left[X_m'(\theta)\prod_{j\in s^1, j\neq m}\mathbf{1}_{\{X_j(\theta)<0\}}\prod_{j\in s^0}\mathbf{1}_{\{X_j(\theta)\geq 0\}}\,;\,X_m=0\right]\\ &+\sum_{s\in\mathscr{S}(m-1)\times\{0\}} g\left(\sum_{i\in s^1} l_i\right)\cdot \mathrm{E}\left[X_m'(\theta)\prod_{j\in s^1}\mathbf{1}_{\{X_j(\theta)<0\}}\prod_{j\in s^0, j\neq m}\mathbf{1}_{\{X_j(\theta)\geq 0\}}\,;\,X_m=0\right]\\ =& -\sum_{s\in\mathscr{S}(m-1)}\left[g\left(\sum_{i\in s^1} l_i + l_m\right) - g\left(\sum_{i\in s^1} l_i\right)\right]\\ &\qquad\times \mathrm{E}\left[X_m'(\theta)\prod_{j\in s^1}\mathbf{1}_{\{X_j(\theta)<0\}}\prod_{j\in s^0}\mathbf{1}_{\{X_j(\theta)\geq 0\}}\,;\,X_m=0\right],\end{aligned}$$

where s has m elements in the first equation and s has $m-1$ elements in the second equation. Recall the definitions of s^1 and s^0 for $\mathscr{S}(m-1)$. Then, by an analog to Equation (5), we have

$$\begin{aligned}\Psi_m &= -\sum_{s\in\mathscr{S}(m-1)}\mathrm{E}\left\{\left[g\left(\sum_{i\in s^1} l_i + l_m\right) - g\left(\sum_{i\in s^1} l_i\right)\right]\cdot X_m'(\theta)\cdot\mathbf{1}_{\{B=s\}};\,X_m=0\right\}\\ &= -\mathrm{E}\left\{\left[g\left(\sum_{j=1}^{m-1} l_j\cdot\mathbf{1}_{\{X_j<0\}} + l_m\right) - g\left(\sum_{j=1}^{m-1} l_j\cdot\mathbf{1}_{\{X_j<0\}}\right)\right]X_m'(\theta)\,;\,X_m=0\right\}.\end{aligned}$$

To simplify the notation, we let

$$L_{-i} = \sum_{j=1, j\neq i}^{m} l_j\cdot\mathbf{1}_{\{X_j<0\}}$$

for all $i = 1,\ldots,m$, which is the portfolio loss without obligor i. Then,

$$\Psi_m = -\mathrm{E}\left\{\left[g\left(L_{-m}+l_m\right) - g\left(L_{-m}\right)\right]\cdot X_m'(\theta)\,;\,X_m=0\right\}.$$

By the symmetry of m and any $i = 1,\ldots,m-1$ and by Equation (6), we have

$$p'(\theta) = -\sum_{i=1}^{m}\mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right) - g\left(L_{-i}\right)\right]\cdot X_i'(\theta)\,;\,X_i=0\right\}.$$

Therefore, we conclude the proof of Theorem 1.

Remark 2. In the proof, we first transform the regular summation term into a combinatorial form, which facilitates interchanging the order between the differential operator and summation. Moreover, the final expression is written back in a regular summation form rather than a combinatorial one, which can reduce the computational complexity significantly.

3. Multiple Estimators and Optimal Linear Combination

In this section, we demonstrate the advantages of Equation (3) which for many practically important models yields multiple sample-mean estimators. Among these estimators, it may be difficult to identify the best one in advance. This then motivates us to consider an optimal linear combination of these estimators. The weights of this estimator are empirically estimated leading to some estimation bias. However, we note that the resultant estimator is often more efficient and always at least as good as the the best one. In this section we also compare and contrast the proposed method to the SPA (smoothed perturbation analysis) method.

We first provide a general framework of developing estimators for $p'(\theta)$ by applying conditioning techniques on both idiosyncratic and macroeconomic factors.

3.1. Conditioning on idiosyncratic factors

Let ϵ_i be an idiosyncratic factor that affects only obligor i and is not a function of θ, and let Υ_i denote a random variable that characterizes all other random factors of obligor i. Note that ϵ_i and Υ_i are independent of each other, and $\Upsilon_i = \Upsilon_i(\theta)$ is a function of θ. The default condition of obligor i is defined as $\{\epsilon_i < \Upsilon_i\}$. Suppose we write $X_i(\theta) = \epsilon_i - \Upsilon_i(\theta)$. Then, $X_i'(\theta) = -\Upsilon_i'(\theta)$, and obligor i defaults if $X_i < 0$. Let $f_{\epsilon_i}(\cdot)$ denote the pdf of ϵ_i. To ensure Conditions 1–3 of Theorem 1 hold, we require the following conditions on ϵ_i and Υ_i.

(a1). $\Upsilon_i(\theta)$ is continuously differentiable w.p.1 and there exists a random variable K_i, which may depend on θ, such that $\mathrm{E}[K_i] < \infty$ *and* $|\Upsilon_i(\theta+\Delta\theta) - \Upsilon_i(\theta)| \le K_i|\Delta\theta|$ *when* $|\Delta\theta|$ *is close enough to zero;*

(a2). $\Upsilon_i(\theta)$ is a continuous random variable;

(a3). $f_{\epsilon_i}(\cdot)$ is continuous a.s. and there exists a constant $B_i > 0$ such that $f_{\epsilon_i}(\cdot)$ is bounded from above by B_i.

Then by Theorem 1, Equation (3) can be further derived as

$$
\begin{aligned}
p'(\theta) &= -\sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right) - g\left(L_{-i}\right)\right] \cdot X_i'(\theta)\,;\, X_i = 0\right\} \\
&= \sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right) - g\left(L_{-i}\right)\right] \cdot \Upsilon_i'(\theta)\,;\, \epsilon_i = \Upsilon_i\right\} \\
&= \sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right) - g\left(L_{-i}\right)\right] \cdot \Upsilon_i'(\theta) \cdot f_{\epsilon_i}(\Upsilon_i)\right\}. \qquad (7)
\end{aligned}
$$

Verification of Conditions (a1)–(a3) is shown in the e-companion EC.1.

3.2. Conditioning on macroeconomic factors

Unless explicitly stated, θ refers to a macroeconomic parameter in this section. Let A be a common random factor (e.g., a macroeconomic factor) that affects all obligors and is not a function of θ, and let B denote a vector that includes all random variables in the system other than A. Note that $B = B(\theta)$ is a function of θ. Suppose that we may write $X_i(\theta) = A - \beta_i$, where $\beta_i = \beta_i(B)$ is a function of B and, thus, also a function of θ. For presentation convenience we suppress the dependence of X_i and β on θ at places where there is no ambiguity. Then, $X_i = A - \beta_i$ and $X_i'(\theta) = -\beta_i'(\theta)$. Note that $A = \beta_i$ when $X_i = 0$. Then, $X_j = \beta_i - \beta_j$ and the obligor j defaults if $\beta_i < \beta_j$. This motivates us to define $\mathcal{L}_{-i} = \sum_{j=1, j\neq i}^m l_j \cdot \mathbf{1}_{\{\beta_i < \beta_j\}}$. Then,

$$\begin{aligned}
p'(\theta) &= -\sum_{i=1}^m \mathrm{E}\left\{\left[g\left(L_{-i} + l_i\right) - g\left(L_{-i}\right)\right] \cdot X_i'(\theta)\,;\, X_i = 0\right\} \\
&= \sum_{i=1}^m \mathrm{E}\left\{\left[g\left(L_{-i} + l_i\right) - g\left(L_{-i}\right)\right] \cdot \beta_i'(\theta)\,;\, A = \beta_i\right\} \\
&= \sum_{i=1}^m \mathrm{E}\left\{\mathrm{E}\left\{\left[g\left(L_{-i} + l_i\right) - g\left(L_{-i}\right)\right] \cdot \beta_i'(\theta); A = \beta_i \,|\, B\right\}\right\} \\
&= \sum_{i=1}^m \mathrm{E}\left\{\left[g\left(\mathcal{L}_{-i} + l_i\right) - g\left(\mathcal{L}_{-i}\right)\right] \cdot \beta_i'(\theta) \cdot \mathrm{E}\left[1; A = \beta_i \,|\, B\right]\right\} \\
&= \sum_{i=1}^m \mathrm{E}\left\{\left[g\left(\mathcal{L}_{-i} + l_i\right) - g\left(\mathcal{L}_{-i}\right)\right] \cdot \beta_i'(\theta) \cdot f_{A|B}(\beta_i)\right\}, \qquad (8)
\end{aligned}$$

where $f_{A|B}(\cdot)$ is the conditional density of A conditioned on B. Furthermore, if A and B are mutually independent, then

$$p'(\theta) = \sum_{i=1}^m \mathrm{E}\left\{\left[g\left(\mathcal{L}_{-i} + l_i\right) - g\left(\mathcal{L}_{-i}\right)\right] \cdot \beta_i'(\theta) \cdot f_A(\beta_i)\right\}.$$

The difference between L_{-i} in Equation (7) and $\mathcal{L}_{-i}$ in Equation (8) may lead to different computational complexities of $p'(\theta)$. For each i, computing L_{-i} is in the same order, $O(m)$, as computing $\mathcal{L}_{-i}$. Then, the complexity of $p'(\theta)$ become $O(m^2)$. However, because $L_{-i} = L - l_i \mathbf{1}_{\{X_i < 0\}}$, the complexity of computing L_{-i} can be reduced to $O(1)$ if L is computed in advance, and that of $p'(\theta)$ using Equation (7) can be reduced to $O(m)$. On the other hand, we cannot apply this trick to $\mathcal{L}_{-i}$; but we may first sort $\beta_1, \beta_2, \ldots, \beta_m$ to achieve the order of $O(m\log(m))$ as computing $p'(\theta)$ using Equation (8). This finding suggests that conditioning on idiosyncratic factors (yielding L_{-i}) may provide better estimators compared with conditioning on macroeconomic factors (yielding $\mathcal{L}_{-i}$), in terms of the computational complexity. When estimating sensitivities with respect to an idiosyncratic parameter, the computational complexities of different estimators obtained by conditioning on various random variables seems in the same order, $O(m)$. In fact, it appears difficult to identify which estimator is the best one, which could be model-dependent, and this motivates us to consider a linear combination of all available estimators.

3.3. Optimal Combination of Multiple Estimators

As mentioned earlier, the proposed method can often provide multiple sample-mean estimators, but we may not be able to identify in advance which one of them is the best in terms of a low variance. This motivates developing an optimal minimum variance linear combination of these estimators.

Let $\{\boldsymbol{\gamma}_\ell = (\gamma_{1\ell}, \ldots, \gamma_{k\ell})', \ell = 1, 2, \ldots, n\}$ denote an i.i.d. sample of $\gamma = (\gamma_1, \ldots, \gamma_k)$. For each ℓ, $\gamma_{1\ell}, \ldots, \gamma_{k\ell}$ are computed from the same simulation run, thus they are mutually dependent. Setting $\bar{\gamma}_i = \frac{1}{n}\sum_{\ell=1}^{n} \gamma_{i\ell}$, then $\bar{\boldsymbol{\gamma}} = (\bar{\gamma}_1, \ldots, \bar{\gamma}_k)'$ are k mutually dependent unbiased sample-mean estimators of γ. Let $\mathbf{w} = (w_1, \ldots, w_k)'$ be a vector of weights and $\tilde{\gamma} = \mathbf{w}'\bar{\boldsymbol{\gamma}}$. For any constant weight vector $\mathbf{w}$, $\tilde{\gamma}$ is an unbiased estimator of γ if $\mathbf{w}'\mathbf{1} = 1$, where $\mathbf{1}$ is a k-dimension vector with all elements being 1. Our goal is to select a $\mathbf{w}$ that minimizes the variance of $\tilde{\gamma}$.

Let Σ denote the covariance matrix of $\bar{\boldsymbol{\gamma}}$. We assume that Σ is positive-definite, i.e., none of the k estimator $\bar{\gamma}_1, \ldots, \bar{\gamma}_k$ can be written as a linear combination of the other $k-1$ estimators. Then $\mathrm{Var}(\tilde{\gamma}) = \mathrm{Var}(\mathbf{w}'\bar{\boldsymbol{\gamma}}) = \mathbf{w}'\Sigma\mathbf{w}$. Therefore, we want to find a $\mathbf{w}$ that solves the following optimization problem:

$$\text{minimize} \quad \mathbf{w}'\Sigma\mathbf{w} \qquad \text{subject to} \quad \mathbf{w}'\mathbf{1} = 1. \tag{9}$$

By using the Lagrange relaxation approach, we can find the optimal solution of Problem (9) is

$$\mathbf{w}^* = \left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1} \Sigma^{-1}\mathbf{1}.$$

In practice, however, Σ is unknown. Therefore, Σ and $\mathbf{w}^*$ can only be estimated. Because $\boldsymbol{\gamma}_\ell$ across $\ell = 1, 2, \ldots, n$ are i.i.d., then an unbiased and strongly consistent estimator of Σ is

$$\hat{\Sigma} = \frac{1}{n(n-1)} \sum_{\ell=1}^{n} \left(\boldsymbol{\gamma}_\ell - \bar{\boldsymbol{\gamma}}\right)\left(\boldsymbol{\gamma}_\ell - \bar{\boldsymbol{\gamma}}\right)'.$$

Then we may estimate $\mathbf{w}^*$ by

$$\hat{\mathbf{w}}^* = \left(\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}\right)^{-1} \hat{\Sigma}^{-1}\mathbf{1}.$$

Therefore, we may use $\hat{\gamma} = \hat{\mathbf{w}}^{*'}\bar{\boldsymbol{\gamma}}$ as the estimator of γ and it is strongly consistent due to the continuous mapping theorem (Durrett 2005). However, $\hat{\gamma}$ is no longer unbiased because of the dependence between $\hat{\mathbf{w}}^*$ and $\bar{\boldsymbol{\gamma}}$. If an unbiased estimator is necessary, one may estimate $\mathbf{w}^*$ using a pilot simulation, i.e., a small number of additional simulation runs that are only used to estimate $\mathbf{w}^*$, so that $\hat{\mathbf{w}}^*$ and $\bar{\boldsymbol{\gamma}}$ can be independent.

It is worthwhile pointing out that, from the numerical results in Section 5, the resultant estimator obtained above (called “combined estimator” in Section 5) does not always perform significantly better than the best amongst all multiple estimators. In that case, instead of the optimal linear combination method, we also suggest a two-phase simulation where one quickly finds the best design using a pilot simulation and then generates sample only from that design to get an unbiased estimator. Nevertheless, the yield of multiple estimators is a natural consequence of our proposed method, and to efficiently use these estimators is the motivation behind applying either a linear combination or a two-phase simulation.

3.4. Connections to SPA Method

As is apparent, the estimators we derive above are based on conditioning techniques. We choose a particular random variable and condition on all others. In this sense, our method can be viewed as a two-step conditional Monte Carlo method for estimating $p'(\theta)$. In the first step, it differentiates $p(\theta)$ and obtains a closed-form expression as in Equation (3); and in the second step, it evaluates the expression by using conditioning techniques. Indeed, the second step of this method can also be generalized to the work of Hong and Liu (2010) and Liu and Hong (2011), which derive the closed-form expressions (similar to the result in Equation (3)) and then use kernel estimators to estimate the sensitivities of probability functions and option prices, to obtain estimators with faster rates of convergence.

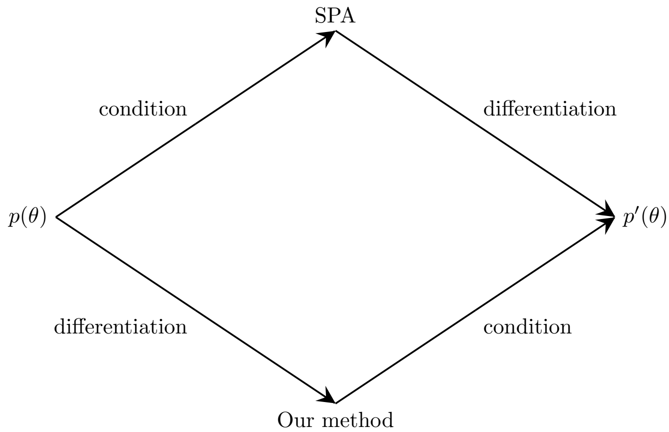

Figure 1 Comparison between SPA and our method.

SPA is another two-step conditional Monte Carlo method for estimating sensitivities of expected-value functions (see, for instance, Fu and Hu (1997) for general discussions and Chen and Glasserman (2008) for applications in credit risk management). However, different from our approach, it conditions in the first step to smooth the function inside of the expectation and differentiates the expectation in the second step (see Figure 1 for an illustration of the two approaches). We believe that our method has several advantages compared to SPA. First, it gives a closed-form expression of the sensitivity in the first step, e.g., Equation (3), which is independent of specific models. This expression provides insights on the problem itself regardless of the model, and it can also be used to develop kernel estimators, e.g., the one in Equation (4), which is model-independent and can be applied easily. Second, our approach makes conditioning on different random variables easier. Once the closed-form expression of the sensitivity is given, it is often quite straightforward to decide what to condition on and to develop multiple estimators, as demonstrated in Section 4 later. For SPA, however, one has to see through both steps (conditioning and differentiation) in order to decide what to condition on, and therefore, it is often more difficult to apply and to develop multiple estimators. Note that both our method and SPA method can be viewed as different approaches to achieving similar (possible the same) estimators under a two-step conditional Monte Carlo framework. See an example of latent variable models in the e-companion EC.2 for an illustration.

4. Applications to Three Classes of Models

In this section, we apply the results of Theorem 1 to three classes of widely used credit models, latent variable models, Bernoulli mixture models and doubly stochastic models, to derive sensitivity estimators that are in general more efficient than the kernel estimators. Specifically, we directly apply both Equation (7) and Equation (8) to develop multiple estimators of $p'(\theta)$ for all three models of joint defaults. Because the choices of A and B depend on specific models, we illustrate our idea by working on specific examples with respect to particular parameters when applying Equation (8).

4.1. Latent Variable Models

We first consider latent variable models where obligor i defaults if a latent variable Y_i is below a threshold d_i. Merton (1974) considered a one period model where Y_i denotes the value of the obligor one period later, and d_i denotes the promised debt at that time. The obligor defaults if it fails to pay the coupon, i.e., $Y_i < d_i$. By introducing dependence among Y_i, $i = 1, \ldots, m$, the model can be used to model joint defaults. We now introduce several examples of commonly used latent variable models.

Example 1 (CreditMetrics and KMV models). *As introduced in Frey and McNeil (2003), both CreditMetrics and KMV models assume that*

$$Y_i = \mathbf{a}_i \Gamma + \sigma_i \epsilon_i + \nu_i, \quad i = 1, \ldots, m, \tag{10}$$

where $\mathbf{a}_i = (a_{i,1}, \ldots, a_{i,p})$ *with* $p < m$, $\Gamma = (\Gamma_1, \ldots, \Gamma_p)^T$ *follows a multivariate normal distribution with mean vector* $\mathbf{0}$ *and covariance matrix* Ω, $\epsilon_1, \ldots, \epsilon_m$ *are independent standard normally distributed random variables, and* ν_i *is the mean value of* Y_i. *In this model, the random vector* Γ *represents the macroeconomic factors and the random variable* ϵ_i *represents obligor* i*'s idiosyncratic risk factor. Then, the dependence between* Y_i *and* Y_j *are modeled by their dependence on the common macroeconomic factors. Let* $A = (\mathbf{a}_1, \ldots, \mathbf{a}_m)^T$. *Then, the covariance matrix of* $(Y_1, \ldots, Y_m)$ *is* $A\Omega A' + \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$.

EXAMPLE 2 (THE MODEL OF LI (2000)). *Let T_i denote the default time of obligor i. We assume that the distribution function of T_i is F_i, which is typically an exponential distribution with rate λ_i, i.e., $F_i(t) = 1 - \exp(-\lambda_i t)$. Then, a loss will be incurred if the obligor defaults before the predetermined time T, i.e., $T_i < T$. Let Φ denote the standard normal distribution function. Then, $\{T_i < T\}$ is equivalent to $\{\Phi^{-1}(F_i(T_i)) \le \Phi^{-1}(F_i(T))\}$. Let $Z_i = \Phi^{-1}(F_i(T_i))$. Note that Z_i follows a standard normal distribution. In this model, Z_i is often modeled as $Z_i = [Y_i - \mathrm{E}(Y_i)]/\sqrt{\mathrm{Var}(Y_i)}$, where Y_i is defined in Equation (10), $\mathrm{E}(Y_i) = \sum_{j=1}^{p} a_{ij}\mu_j + \nu_i$, and $\mathrm{Var}(Y_i) = \mathbf{a}_i\Omega\ \mathbf{a}_i' + \sigma_i^2$. Let $d_i = \sqrt{\mathrm{Var}(Y_i)} \cdot \Phi^{-1}(F_i(T)) + \mathrm{E}(Y_i)$. Then, $\{T_i < T\}$ is equivalent to $\{Y_i < d_i\}$.*

Both Examples 1 and 2 are known as Gaussian copula models because $Y = (Y_1, \ldots, Y_m)^T$ follows a multivariate normal distribution. For Examples 1 and 2, for instance, ϵ_i is a standard normal random variable and $\Upsilon_i = (d_i - \mathbf{a}_i\Gamma - \nu_i)/\sigma_i$. Equation (7) provides an efficient approach to estimating $p'(\theta)$.

To illustrate how to apply Equation (8) to Gaussian copula models, we consider the following specific parameters. Suppose that d_i is a function of a parameter λ_i, i.e., $d_i = d_i(\lambda_i)$, (as already in Example 2), and we are interested in estimating $p'(\lambda_i)$ for some $i = 1, \ldots, m$. When p is the price of a CDO and λ_i is the default intensity of obligor i, then $p'(\lambda_i)$ is known as `delta` (Chen and Glasserman 2008). Without loss of generality, we consider $p'(\lambda_1)$.

Let $A = \Gamma_j$ for any $j = 1, \ldots, p$. Without loss of generality, we set $A = \Gamma_1$. Then, $\{Y_i < d_i\}$ is equivalent to $\{A < \beta_i\}$ where $\beta_i = \frac{1}{a_{i1}}\left[d_i - \sum_{k=2}^{p} a_{ik}\Gamma_k - \sigma_i\epsilon_i - \nu_i\right]$. Let $X_i = A - \beta_i$. Then,

$$\beta_i'(\lambda_1) = \begin{cases} \frac{d_1'(\lambda_1)}{a_{11}}, & i = 1, \\ 0, & i = 2, \ldots, m. \end{cases}$$

Note that $(\Gamma_1, \ldots, \Gamma_p)'$ follows a multivariate normal distribution with mean $\mathbf{u}$ and covariance matrix Ω. We let $\boldsymbol{\mu}_{-1}$ and Ω_{-1} denote the mean vector and covariance matrix of $\Gamma_{-1} = (\Gamma_2, \ldots, \Gamma_p)'$, and let $\sigma_1^2 = \mathrm{Var}(\Gamma_1)$ and $\omega_1 = (\mathrm{Cov}(\Gamma_1, \Gamma_2), \ldots, \mathrm{Cov}(\Gamma_1, \Gamma_p))'$. Then, by Bock (1985), $f_{\Gamma_1|\Gamma_{-1}}(\cdot)$ is the same as the density of a normal random variable with mean $\bar{\mu}_1$ and variance $\bar{\sigma}_1^2$, where $\bar{\mu}_1 = \mu_1 + \omega_1'\Omega_{-1}(\Gamma_{-1} - \mu_{-1})$ and $\bar{\sigma}_1^2 = \sigma_1^2 - \omega_1'\Omega_{-1}\omega_1$.

By Equation (8), we have

$$p'(\lambda_1) = \mathrm{E}\left\{\left[g(\mathcal{L}_{-1} + l_1) - g(\mathcal{L}_{-1})\right] \cdot \frac{d_1'(\lambda_1)}{a_{11}} \cdot f_{\Gamma_1|\Gamma_{-1}}(\beta_1)\right\}.$$

Similarly, by setting $A = \Gamma_2, \ldots, \Gamma_p$, we can also develop another $p-1$ sample-mean estimators of $p'(\lambda_1)$. Combining with the one given by Equation (7), we now have totally $p+1$ sample-mean estimators of $p'(\lambda_1)$.

EXAMPLE 3 (THE MODEL OF BASSAMBOO ET AL. (2008)). *It is known that the Gaussian copula models cannot explain well the extremal dependence among obligors that is observed empirically (Mashal and Zeevi 2003), which means that the obligors are more likely to default simultaneously than what the Gaussian copula models predict. Bassamboo et al. (2008) suggested the following single factor model:*

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

where Z denotes the common factor that affects all obligors, ϵ_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 ϵ_i are mutually independent. When Z and ϵ_i are independent normal random variables and $W = 1$, the model becomes the one-factor Gaussian 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 extremal credit risk when W or W^2 follows a Gamma distribution. Specifically, when W^2 follows a Chi-square distribution, Y_i follows a t-distribution and the model is also known as a t-copula model (Embrechts et al. 2003).

Suppose that $W = \theta\mathcal{E}$, where $\mathcal{E}$ is an exponential random variable with the mean equal to 1 and θ is the mean of W. Suppose that we are interested in estimating $p'(\theta)$, which is the sensitivity of the portfolio credit risk to the average shock size.

By by letting ϵ_i be a standard normal random variable and $\Upsilon_i = (d_i\theta\mathcal{E} - \rho Z)/\sqrt{1-\rho^2}$, Equation (7) yields the following estimator,

$$p'(\theta) = \sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(L_{-i} + l_i\right) - g\left(L_{-i}\right)\right] \cdot \Upsilon_i'(\theta) \cdot f_{\epsilon_i}(\Upsilon_i)\right\}. \tag{11}$$

Besides the estimator in Equation (11), we can also apply Equation (8) to develop two other estimators. Note that threshold d_i could be positive or negative, depending on the parameter settings in Y_i. To be consistent with numerical test in Section 5.1, we assume d_i to be negative.

First, we let $A = \mathcal{E}$ and $\beta_i = \frac{\rho Z + \sqrt{1-\rho^2}\epsilon_i}{\theta d_i}$. Then, $\{Y_i < d_i\}$ is equivalent to $\{A < \beta_i\}$. Let $X_i = A - \beta_i$. Then,

$$\beta_i'(\theta) = -\frac{\rho Z + \sqrt{1-\rho^2}\epsilon_i}{\theta^2 d_i} = -\frac{\beta_i}{\theta},$$

Let $f_{\mathcal{E}}(x) = e^{-x}$, $x \geq 0$, denote the density of $\mathcal{E}$. Then, $f_{A|B}(x) = f_A(x) = f_{\mathcal{E}}(x)$. By Equation (8), we have

$$p'(\theta) = -\sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(\mathcal{L}_{-i} + l_i\right) - g\left(\mathcal{L}_{-i}\right)\right] \cdot \frac{\beta_i}{\theta} \cdot f_{\mathcal{E}}(\beta_i)\right\}. \tag{12}$$

Second, we let $A = Z$. Similarly, we have $\beta_i = \frac{\theta d_i\mathcal{E} - \sqrt{1-\rho^2}\epsilon_i}{\rho}$ and $\beta_i'(\theta) = \frac{d_i\mathcal{E}}{\rho}$. Let $f_Z(\cdot)$ denote the density of Z. Then, by Equation (8), we have

$$p'(\theta) = \sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(\mathcal{L}_{-i} + l_i\right) - g\left(\mathcal{L}_{-i}\right)\right] \cdot \frac{d_i\mathcal{E}}{\rho} \cdot f_Z\left(\beta_i\right)\right\}. \tag{13}$$

Given Equations (11), (12) and (13), we can develop three sample-mean estimators of $p'(\theta)$ if an i.i.d. sample of $\{Z, \mathcal{E}, \epsilon_1, \ldots, \epsilon_m\}$ is available. Suppose the i.i.d. sample of $\{Z, \mathcal{E}, \epsilon_1, \ldots, \epsilon_m\}$ is denoted as $\{(Z_1, \mathcal{E}_1, \epsilon_{1,1}, \ldots, \epsilon_{m,1}), \ldots, (Z_n, \mathcal{E}_n, \epsilon_{1,n}, \ldots, \epsilon_{m,n})\}$, the three sample-mean estimators of $p'(\theta)$ given Equations (11), (12) and (13), respectively, are

$$\overline{p_1'(\theta)} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{m}\left[g\left(L_{-i,j}+l_i\right)-g\left(L_{-i,j}\right)\right]\cdot\frac{d_i\mathcal{E}}{\sqrt{1-\rho^2}}\cdot f_{\epsilon_i}\left(\Upsilon_{i,j}\right), \tag{14}$$

where $\Upsilon_{i,j} = \left(d_i\theta\mathcal{E}_j - \rho Z_j\right)/\sqrt{1-\rho^2}$ and $L_{-i,j} = \sum_{s=1,s\neq i}^{m} l_s \cdot \mathbf{1}_{\{\epsilon_{s,j}<\Upsilon_{s,j}\}}$,

$$\overline{p_2'(\theta)} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{m}\left[g\left(\mathcal{L}_{-i,j}+l_i\right)-g\left(\mathcal{L}_{-i,j}\right)\right]\cdot\frac{\beta_{i,j}}{\theta}\cdot f_{\mathcal{E}}(\beta_{i,j}), \tag{15}$$

where $\beta_{i,j} = (\rho Z_j + \sqrt{1-\rho^2}\epsilon_{i,j})/(\theta d_i)$ and $\mathcal{L}_{-i,j} = \sum_{s=1,s\neq i}^{m} l_s \cdot \mathbf{1}_{\{\beta_{i,j}<\beta_{s,j}\}}$, and

$$\overline{p_3'(\theta)} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{m}\left[g\left(\mathcal{L}_{-i,j}+l_i\right)-g\left(\mathcal{L}_{-i,j}\right)\right]\cdot\frac{d_i\mathcal{E}_j}{\rho}\cdot f_Z\left(\beta_{i,j}\right), \tag{16}$$

where $\beta_{i,j} = (\theta d_i\mathcal{E}_j - \sqrt{1-\rho^2}\epsilon_{i,j})/\rho$ and $\mathcal{L}_{-i,j} = \sum_{s=1,s\neq i}^{m} l_s \cdot \mathbf{1}_{\{\beta_{i,j}<\beta_{s,j}\}}$. Because $\overline{p_k'(\theta)}$, $k = 1, 2, 3$, is a typical sample mean estimator, it is strongly consistent by strong law of large numbers, and after normalization, it follows a central limit theorem (Durrett 2005). Furthermore, its rate of convergence is $n^{-1/2}$, which is faster than the rate of the kernel estimator $\widehat{p'(\theta)}$ of Equation (4). This conclusion holds for all the models considered throughout this paper. For the space limitation, we omit repeating this conclusion for the rest of the models, and we also omit providing the closed-form expressions of the sample-mean estimators as in Equations (14)–(16).

4.2. Bernoulli Mixture Models

Let $\Gamma = (\Gamma_1, \ldots, \Gamma_p)'$ denote a set of common economic factors, where $p < m$. In Bernoulli mixture models, the default event of obligor i follows a Bernoulli random variable with a default probability Q_i $(0 < Q_i < 1)$, and Q_i is modeled as a function of Γ, i.e., $Q_i = Q_i(\Gamma)$. Furthermore, defaults of all obligors are independent of each other once Γ is given. Therefore, in Bernoulli mixture models, the dependence among all obligors are modeled through their dependence on the common economic factors Γ. The following are a few commonly used Bernoulli mixture models.

Example 4 (CreditRisk$^+$ model). *A Bernoulli mixture model is used in CreditRisk$^+$, a financial product developed by Credit-Suisse-Financial-Products. As introduced in Frey and McNeil (2003), CreditRisk$^+$ uses $Q_i = 1 - e^{-\mathbf{w}_i'\Gamma}$, where Γ is a vector of independent gamma distributed macroeconomics factors and $\mathbf{w}_i = (w_{i,1}, \ldots, w_{i,p})'$ is a vector of positive weights.*

Example 5 (Bernoulli regression models). *In Bernoulli regression models, the individual default probability is modeled as $Q_i(\Gamma) = q(\Gamma, \mathbf{z}_i), i = 1, \ldots, m$ where $\mathbf{z}_i$ is a deterministic vector. As introduced in Frey and McNeil (2003), a particularly popular choice is*

$$q(\Gamma, \mathbf{z}_i) = h\left(\boldsymbol{\sigma}'\mathbf{z}_i\Gamma + \boldsymbol{\mu}'\mathbf{z}_i\right),$$

where $h:\Re \to (0,1)$ is a strictly increasing function, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are vectors of regression parameters and $\boldsymbol{\sigma}'\mathbf{z}_i > 0$. As shown in Frey and McNeil (2003), under some specific choices of Γ and $h(\cdot)$, an individual obligor's default may follow a probit-normal or logit-normal mixing-distribution.

We consider only the CreditRisk$^+$ model introduced in Example 4. Suppose that we are interested in estimating $p'(w_{ij})$ for some $i=1,\ldots,m$ and $j=1,\ldots,p$. Without loss of generality, we consider $p'(w_{11})$.

Let $U_i, i=1,\ldots,m$, be independent uniform$(0,1)$ random variables that are independent of Γ, and $X_i = U_i - Q_i$, where $Q_i = 1-\exp\left(-\sum_{j=1}^{p} w_{ij}\Gamma_j\right)$. Then, obligor i defaults if $X_i < 0$. Here, U_i is equivalent to ϵ_i and Q_i is equivalent to Υ_i in Equation (7), then,

$$p'(w_{11}) = \mathrm{E}\left\{\left[g\left(L_{-1}+l_1\right) - g\left(L_{-1}\right)\right]\cdot \Gamma_1 e^{w_{11}\Gamma_1}\right\}. \tag{17}$$

We next apply Equation (8) to to develop other estimators. Let $f_{\Gamma_i}(\cdot)$ denote the density of Γ_i. Note that, in this model, $\Gamma_1,\ldots,\Gamma_p$ are mutually independent.

First, we let $A=-\Gamma_1$. Then, we have $\beta_i = \frac{1}{w_{i1}}\left[\log(1-U_i) + \sum_{k=2}^{p} w_{ik}\Gamma_k\right]$ and

$$\beta_i'(w_{11}) = \begin{cases} -\frac{1}{w_{11}^2}\left[\log(1-U_1)+\sum_{k=2}^{p} w_{1k}\Gamma_k\right] = -\frac{\beta_1}{w_{11}}, & i=1, \\ 0, & i=2,\ldots,m. \end{cases}$$

By Equation (8) and similar analysis as in Section 4.1, we have

$$p'(w_{11}) = -\mathrm{E}\left\{\left[g\left(\mathcal{L}_{-1}+l_1\right) - g\left(\mathcal{L}_{-1}\right)\right]\cdot \frac{\beta_1}{w_{11}} \cdot f_{\Gamma_1}\left(-\beta_1\right)\right\}. \tag{18}$$

Second, we let $A=-\Gamma_2$ (which can be extended easily to $A=-\Gamma_j$ for any $j=2,\ldots,p$). Then, $\beta_i = \frac{1}{w_{i2}}\left[\log(1-U_i)+\sum_{k=1,k\neq 2}^{p} w_{ik}\Gamma_k\right]$ and

$$\beta_i'(w_{11}) = \begin{cases} \frac{\Gamma_1}{w_{12}}, & i=1, \\ 0, & i=2,\ldots,m. \end{cases}$$

By Equation (8) and similar analysis, we have

$$p'(w_{11}) = \mathrm{E}\left\{\left[g\left(\mathcal{L}_{-1}+l_1\right) - g\left(\mathcal{L}_{-1}\right)\right]\cdot \frac{\Gamma_1}{w_{12}} \cdot f_{\Gamma_2}\left(-\beta_1\right)\right\}. \tag{19}$$

Given Equations (17), (18) and (19), we can develop $p+1$ sample-mean estimators of $p'(w_{11})$ if an i.i.d. sample of $\{\Gamma, U_1,\ldots,U_m\}$ is available.

4.3. Doubly Stochastic Models

Let $\{N_i(t): t\geq 0\}$ denote a nonhomogeneous Poisson process with nonnegative stochastic intensity process $\lambda_i = (\lambda_i(t): t\geq 0)$. In doubly stochastic models, the default of obligor i occurs at the first jump time $\tau_i = \min\{t \geq 0 : N_i(t) = 1\}$. Then, conditioned on the intensities λ_i, $i = 1, ..., m$, the default time τ_i of obligor i are mutually independent random variables with

$$\Pr\{\tau_i > t | \lambda_i\} = \Pr\{N(t) = 0 | \lambda_i\} = \exp\left\{-\int_0^t \lambda_i(u) du\right\}.$$

Let $\Lambda_i = \int_0^T \lambda_i(u) du$ and let $E_i, i = 1, ..., m$, be independent exponential random variables with mean 1. Then $\{\tau_i < T\}$ is equivalent to $\{E_i < \Lambda_i\}$, i.e., obligor i defaults before time T if $E_i < \Lambda_i$.

To model the dependence among the obligors, the intensity process is often modeled as

$$\lambda_i(t) = S_c(t) + S_i(t), \tag{20}$$

where $\{S_c(t) \geq 0 : t \geq 0\}$ models the common part of the intensity processes of all obligors and $\{S_i(t) \geq 0 : t \geq 0\}$ models the individual part of obligor i's intensity process. In this model, $S_c(t)$ and $S_i(t)$, $i = 1, ..., m$, are often modeled as diffusion processes, e.g.,

$$dS_c(t) = \mu_c(t, S_c(t))\, dt + \sigma_c(t, S_c(t))\, dB_c(t), \tag{21}$$

$$dS_i(t) = \mu_i(t, S_i(t))\, dt + \sigma_i(t, S_i(t))\, dB_i(t), \tag{22}$$

where B_c and B_i, $i = 1, ..., m$, are mutually independent Brownian motion processes. To ensure the nonnegativity of S_c and S_i, square-root diffusion processes, e.g., CIR processes, are often used.

Example 6 (The model of Duffie and Gârleanu (2001)). *Duffie and Gârleanu (2001) modeled S_c and S_i by CIR processes with jumps, i.e.,*

$$dS_c(t) = \kappa(\mu_c - S_c(t))dt + \sigma\sqrt{S_c(t)} dB_c(t) + dJ_c(t),$$

$$dS_i(t) = \kappa(\mu_i - S_c(t))dt + \sigma\sqrt{S_i(t)} dB_i(t) + dJ_i(t),$$

where B_c and B_i are mutually independent Brownian motions, and J_c and J_i are mutually independent pure-jump processes and also independent of the Brownian motions. Their jump sizes are independent and exponentially distributed and their jump times are formulated as a series of Poisson processes (jump sizes and jump times are also independent). To simulate these processes, we may simulate the sets of jumps first and add the jump times to the set of discretized time steps and then apply the Euler scheme at the new set of time steps.

To simulate $\lambda_i(t)$ and to evaluate Λ_i in the doubly stochastic models, we often use the Euler scheme to discretize $S_c(t)$ and $S_i(t)$, $i = 1, ..., m$ (Glasserman 2004). Let k be the number of time steps in the discretization, and $\Delta t = T/k$ and $t_j = j \cdot \Delta t$, $j = 0, 1, ..., k-1$. Furthermore, let $\widehat{S}_c$ and $\widehat{S}_i$, $i = 1, ..., m$ denote time-discretized approximations to S_c and S_i. Under the Euler scheme,

$$\widehat{S}_c(t_{j+1}) = \widehat{S}_c(t_j) + \mu_c\left(t_j, \widehat{S}_c(t_j)\right)\Delta t + \sigma_c\left(t_j, \widehat{S}_c(t_j)\right)\sqrt{\Delta t} Z_{c,j+1}, \tag{23}$$

$$\widehat{S}_i(t_{j+1}) = \widehat{S}_i(t_j) + \mu_i\left(t_j, \widehat{S}_i(t_j)\right)\Delta t + \sigma_i\left(t_j, \widehat{S}_i(t_j)\right)\sqrt{\Delta t} Z_{i,j+1} \tag{24}$$

for $j=0,1,...,k-1$ and $i=1,2,...,m$, with $\widehat{S}_c(0)=S_c(0)$ and $\widehat{S}_i(0)=S_i(0)$, where $Z_{c,j+1}, Z_{i,j+1}$ are independent standard normal random variables for $j=0,...,k-1$ and $i=1,...,m$. Then, we can approximate Λ_i by

$$\widehat{\Lambda}_i = \sum_{j=0}^{k-1} \widehat{\lambda}_i(t_j)\Delta t = \sum_{j=0}^{k-1} \left[\widehat{S}_c(t_j)+\widehat{S}_i(t_j)\right]\Delta t. \tag{25}$$

Suppose we use the doubly stochastic model defined in Equations (20) to (22) and use the discretization scheme defined in Equation (25) to evaluate Λ_i. Furthermore, suppose we are interested in estimating $p'(S_i(0))$ for $i=1,...,m$. Without loss of generality, we consider $p'(S_1(0))$.

Let $X_i = E_i - \widehat{\Lambda}_i$. Here, E_i is equivalent to ϵ_i and $\widehat{\Lambda}_i$ is equivalent to Υ_i in Equation (7), then, by Equation (7), we have

$$p'(S_1(0)) = \mathrm{E}\left\{[g(L_{-1}+l_1)-g(L_{-1})]\cdot \widehat{\Lambda}_1'(S_1(0))\cdot f_{E_1}(\widehat{\Lambda}_1)\right\}, \tag{26}$$

where

$$\widehat{\Lambda}_1'(S_1(0)) = \sum_{j=0}^{k-1} \frac{d\widehat{S}_1(t_j)}{dS_1(0)}\cdot \Delta t$$

with pathwise derivative

$$\frac{d\widehat{S}_1(t_j)}{dS_1(0)} = \frac{d\widehat{S}_1(t_j)}{d\widehat{S}_1(t_{j-1})}\cdot \frac{d\widehat{S}_1(t_{j-1})}{d\widehat{S}_1(t_{j-2})}\cdot\cdots\cdot \frac{d\widehat{S}_1(t_1)}{dS_1(0)},$$

and

$$\frac{d\widehat{S}_1(t_j)}{d\widehat{S}_1(t_{j-1})} = 1 + \frac{d\mu_1\left(t_j, \widehat{S_1}(t_{j-1})\right)}{d\widehat{S}_1(t_{j-1})}\Delta t + \frac{d\sigma_1\left(t_j, \widehat{S}_1(t_{j-1})\right)}{d\widehat{S}_1(t_{j-1})}\sqrt{\Delta t}Z_{i,j}.$$

We may also use other individual random factors, $Z_{i,k-1}$ (defined in Equation (24)), to derive another set of estimators. Let $X_i = -Z_{i,k-1}-\xi_i$, where

$$\xi_i = \frac{\sum_{j=0}^{k-2}\widehat{\lambda}_i(t_j)\Delta t + \widehat{S}_c(t_{k-1})\Delta t + \widehat{S}_i(t_{k-2})\Delta t + \mu_i(t_{k-2}, \widehat{S}_i(t_{k-2}))(\Delta t)^2 - E_i}{\sigma_i(t_{k-2}, \widehat{S}_i(t_{k-2}))(\Delta t)^{3/2}}.$$

Here, $-Z_{i,k-1}$ is equivalent to ϵ_i and ξ_i is equivalent to Υ_i in Equation (7), and $X_i'(S_1(0)) = -\xi_i'(S_1(0))$ where

$$\xi_i'(S_1(0)) = \begin{cases} \frac{1}{\sigma_1\left(t_{k-2},\widehat{S}_1(t_{k-2})\right)(\Delta t)^{1/2}}\cdot\left[\sum_{j=0}^{k-2}\frac{d\widehat{S}_1(t_j)}{dS_1(0)} + \frac{d\widehat{S}_1(t_{k-2})}{dS_1(0)} + \frac{d\mu_1\left(t_{k-2},\widehat{S}_1(t_{k-2})\right)}{dS_1(0)}\Delta t\right] & \\ -\frac{\xi_1}{\sigma_1\left(t_{k-2},\widehat{S}_1(t_{k-2})\right)}\cdot\frac{d\sigma_1\left(t_{k-2},\widehat{S}_1(t_{k-2})\right)}{dS_1(0)}, & i=1, \\ 0, & i\neq 1. \end{cases}$$

Then, by Equation (7), we have

$$p'(S_1(0)) = \mathrm{E}\left\{[g(L_{-1}+l_1)-g(L_{-1})]\cdot \xi_1'(S_1(0))\cdot \phi(-\xi_1)\right\}, \tag{27}$$

where $\phi(\cdot)$ is the density of a standard normal distribution.

To apply Equation (8) to derive other estimators, we let $A = -Z_{c,k-1}$, where $Z_{c,k-1}$ is a standard normal random variable in Equation (23). The default event of obligor i, $\left\{E_i < \widehat{\Lambda}_i\right\}$ is equivalent to $\{A < \beta_i\}$, where

$$\beta_i = \frac{\sum_{j=0}^{k-2} \widehat{\lambda}_i(t_j)\Delta t + \widehat{S}_i(t_{k-1})\Delta t + \widehat{S}_c(t_{k-2})\Delta t + \mu_c\left(t_{k-2}, \widehat{S}_c(t_{k-2})\right)(\Delta t)^2 - E_i}{\sigma_c\left(t_{k-2}, \widehat{S}_c(t_{k-2})\right)(\Delta t)^{3/2}}.$$

Then,

$$\beta_i'(S_1(0)) = \begin{cases} \frac{\sum_{j=0}^{k-1} \frac{d\widehat{S}_1(t_j)}{dS_1(0)}}{\sigma_c\left(t_{k-2}, \widehat{S}_c(t_{k-2})\right)(\Delta t)^{1/2}}, & i = 1, \\ 0, & i \neq 1. \end{cases}$$

By Equation (8), we have

$$p'(S_1(0)) = \mathrm{E}\left\{[g(\mathcal{L}_{-1} + l_1) - g(\mathcal{L}_{-1})] \cdot \frac{\sum_{j=0}^{k-1} \frac{d\widehat{S}_1(t_j)}{dS_1(0)}}{\sigma_c\left(t_{k-2}, \widehat{S}_c(t_{k-2})\right)(\Delta t)^{1/2}} \cdot \phi(-\beta_1)\right\}. \tag{28}$$

Given Equations (26), (27) and (28), we can develop three sample-mean estimators of $p'(S_1(0))$ if an i.i.d. sample of $\{E_1, \ldots, E_m, Z_{c,1}, \ldots, Z_{c,k}, Z_{i,j}, i = 1, \ldots, m, j = 1, \ldots, k\}$ is available.

Under the Euler scheme, the computational complexity of $\widehat{\Lambda}_i$ is $O(k)$, but the complexity of computing $\widehat{\Lambda}_i'(\theta)$ depends on the exact form of parameter θ, so does for $p'(\theta)$. If θ is an idiosyncratic parameter, e.g., $\theta = S_1(0)$, then it takes $O(km)$ to compute $p'(\theta)$ by Equations (26), (27) or (28). If θ is a macroeconomic parameter, e.g., $\theta = S_c(0)$, then it takes $O(km^2)$ to compute $p'(\theta)$ using $\mathcal{L}_{-i}$ and $O(km)$ to compute $p'(\theta)$ using L_{-i}, which is caused by the difference between computing $\mathcal{L}_{-i}$ and L_{-i}.

5. Numerical Experiments

In this section we test the performances of our estimators through three examples, including one for latent variable models, one for Bernoulli mixture models, and one for doubly stochastic models. In each example, we consider two performance functions, $g(L) = \mathbf{1}_{\{L>y\}}$ and $g(L) = L \cdot \mathbf{1}_{\{L>y\}}$ (denoted as Cases A and B, respectively), and estimate the $d\mathrm{E}[g(L)]/d\theta$ for some parameter θ that is in the model of the joint defaults. For each example and each performance function, we consider three types of estimators, the LR estimator when it is applicable (we will derive the LR estimator later), the kernel estimator given by Equation (4) and the various sample-mean estimators developed in Section 4, and compare their performances. It is worth noting that there are multiple sample-mean estimators that can be used to estimate $p'(\theta)$ by the results in Section 4. Furthermore, these estimators can all be computed by using the sample generated in the same simulation. Therefore, this motivates us to use linear combinations of these estimators to obtain more efficient estimators.

In all three examples, without additional specifications, there are 100 obligors in the loan portfolio (i.e., $m = 100$) and the loss due to default of obligor i equals to 100 (i.e., $l_i = 100$ for all $i = 1,\ldots,100$). In both performance functions, we set $y = 2000$, i.e., we are interested in the cases where at least 20 obligors default. Other parameters of examples will be introduced according to their models.

To use the kernel estimator of Equation (4), we need to choose the bandwidth parameter δ_n. As shown in Hong and Liu (2010), to achieve the optimal rate of convergence, δ_n should be in the order of $n^{-1/5}$. Then, we set $\delta_n = cn^{-1/5}$ for some positive constant c. We test the kernel estimators with different values of c for all three examples and find that $c = 1$ is always a good choice. Therefore, we set $c = 1$ for all three examples.

5.1. A Latent Variable Model

We consider the model of Bassamboo et al. (2008) introduced in Example 3. We suppose that both the common factor Z and idiosyncratic factor ϵ_i follow a standard normal distribution, $\rho = 0.6$, $d_i = -2$, for all i, and the common shock factor $W = \theta\mathcal{E}$ where $\theta = 1$ and $\mathcal{E}$ follows an exponential distribution with mean 1. We are interested in estimating the sensitivities of the expected performances of the two performance functions with respect to the average shock size θ.

To obtain the LR estimator, we consider θ as the mean parameter of W (i.e., we transfer the structural parameter into a distributional parameter). Then, θ will not appear in other random variables, which allows us to obtain the score function (SF). The density function of W, $f_W(x) = \frac{1}{\theta} f_{\mathcal{E}}(x/\theta) = \frac{1}{\theta} e^{-x/\theta}$ for $x \geq 0$. Then,

$$p'(\theta) = \mathrm{E}\left\{g(L) \cdot \mathrm{SF}\right\} = \mathrm{E}\left\{g(L) \cdot (-1 + \mathcal{E})/\theta\right\}, \tag{29}$$

where $\mathrm{SF} = \frac{d}{d\theta} \log\left(\frac{1}{\theta} e^{-W/\theta}\right) = -\frac{1}{\theta} + \frac{W}{\theta^2} = \frac{-1+\mathcal{E}}{\theta}$. Given Equation (29), we obtain the LR estimator of $p'(\theta)$ if an i.i.d. sample of $(Z, \mathcal{E}, \epsilon_1, \ldots, \epsilon_m)$ is available.

To simulate the joint defaults, we can generate an i.i.d. sample of $(Z, \mathcal{E}, \epsilon_1, \ldots, \epsilon_m)$, denoted by $\{(Z_\ell, \mathcal{E}_\ell, \epsilon_{1,\ell}, \ldots, \epsilon_{m,\ell}) : \ell = 1, 2, \ldots, n\}$. Based on the sample, we may compute the kernel estimator of Equation (4), the LR estimator of Equation (29), the three sample-mean estimators of Equations (14)–(16) (which we denote as Estimators 1, 2, 3, respectively), and the combined estimator calculated from the three sample-mean estimators (see Section 3.3 for derivation). We report the estimates (denoted by $\bar{M}$) and the standard errors of these estimators (denoted by s.e.) in Table 1 with different sample sizes.

From Table 1, we have several findings for both performance functions. First, the kernel estimator can appropriately estimate the sensitivities. This finding demonstrates the correctness of Theorem

Table 1 **The estimates and their standard errors (s.e.) for the model of Bassamboo et al. (2008).**

Sample size n	10^4		10^5		10^6	
Case A	M	s.e. $(\times 10^{-2})$	M	s.e. $(\times 10^{-3})$	M	s.e. $(\times 10^{-3})$
Estimator 1	-0.2052	2.2	-0.2111	7.1	-0.2048	2.2
Estimator 2	-0.2041	0.12	-0.2072	0.39	-0.2068	0.12
Estimator 3	-0.2061	0.21	-0.2081	0.66	-0.2064	0.21
Combined	-0.2046	0.11	-0.2074	0.34	-0.2067	0.11
Kernel	-0.2260	2.4	-0.2091	7.5	-0.2019	2.6
LR	-0.2144	0.36	-0.2069	1.1	-0.2069	0.35
Case B	M	s.e.	M	s.e.	M	s.e.
Estimator 1	-983.8	47.5	-1003.7	15.4	-983.2	4.75
Estimator 2	-971.1	8.1	-989.9	2.6	-988.5	0.81
Estimator 3	-991.4	9.6	-994.0	3.0	-986.4	0.96
Combined	-979.5	6.2	-991.4	2.0	-987.7	0.62
Kernel	-1029.1	52.3	-1004.1	16.5	-977.8	5.6
LR	-1021.1	19.5	-989.7	6.1	-990.1	1.9

Note: Estimators 1–3 are specified by Equations (14)–(16), respectively.

1 and the usefulness of the kernel estimator. Second, the sample-mean estimators and the combined estimator appear to have a rate of convergence of $n^{-1/2}$ and the kernel estimator appears to have a rate of convergence nearly $n^{-1/2}$. This finding supports our motivation of deriving sample-mean estimators. Third, the combined estimator has a smaller standard error than the three sample-mean estimators. Fourth, conditioning on a common risk factor, e.g., Estimators 2 and 3, may yield estimators that have smaller standard error than conditioning solely on an idiosyncratic risk factor (as in Chen and Glasserman (2008)). Fifth, the LR estimator achieves the same rate of convergence of $n^{-1/2}$ as the sample-mean estimators, but has larger standard errors than the best sample-mean estimator (and also the combined estimator).

We next consider the time taken to compute each estimate. Unlike the previous experiment only using one i.i.d sample to obtain all the estimates, the numerical test for timing is carried out by using different i.i.d samples for different estimates. We run the Matlab code on a 3.40GHz Intel Quad-Core PC with 4 GB RAM for our numerical tests. Note that even though there are 4 cores, the Matlab code is always executed using a single core. The computational times for computing Estimator 1, combined estimator, kernel estimator and LR estimator with 100 independent replications are reported in Table 2. Note that we fix the number of obligors $m = 100$ on the left panel of Table 2, while we fix the sample size $n = 10^4$ on the right panel.[a]

We find that the time of computing Estimator 1 and that of computing the kernel estimator are almost the same, and they are higher than that of computing the LR estimator. From the left panel in Table 2, the time of computing the combined estimator is about one-order larger than other

[a] The threshold y also increases in proportion to the number of obligors m.

Table 2 Time (in second) taken to compute each estimator with 100 replications.

Sample size n	10^4	10^5	10^6		Number of Obligors m	10	100	1000
Estimator 1	0.09	0.79	8.10		Estimator 1	0.02	0.09	0.86
Combined	1.21	11.90	122.88		Combined	0.05	1.21	69.88
Kernel	0.08	0.69	7.02		Kernel	0.03	0.08	0.81
LR	0.05	0.43	4.31		LR	0.02	0.05	0.50

estimators (≈ 15 times greater than that of computing Estimator 1), which is reasonable based on the analysis of the computational complexities of different estimators as shown in Section 3. Recall that the computational complexity of either Estimators 2 or 3 is $O(m^2)$ and that of Estimator 1 is $O(m)$, then the computational complexity of the combined estimator, calculated based on the three sample-mean estimators, is also $O(m^2)$. This is also consistent with the result in the right panel. The time of computing Estimator 1, kernel estimator and LR estimator increases linearly as m increases while that of computing the combined estimator grows faster than $O(m)$. In this case, the benefit of the combined estimators (as well as Estimators 2 and 3) may be canceled out due to a higher computational complexity. Then we suggestion to derive estimators by conditioning only on idiosyncratic factors for macroeconomic parameters when the number of obligors is large.

5.2. A Bernoulli Mixture Model

We consider CreditRisk$^+$ model introduced in Example 4. We suppose that Γ is a 5-dimensional vector of independent gamma distributed macroeconomics factors all with shape parameter 3 and scale parameter 0.1 (i.e., $\Gamma_j \sim Gamma(3, 0.1)$ for $j = 1, \ldots, 5$), and all weights equal to 0.1 (i.e., $w_{ij} = 0.1$ for $i = 1, \ldots, 100$ and $j = 1, \ldots, 5$). We are interested in estimating the sensitivities of the expected performances of the two performance functions with respect to w_{11}.

For this example, it is not clear to us how the LR method may be applied directly since we are not able to write the parameter w_{11} as a distributional parameter of a single random variable. "Push-out" techniques may be helpful when the structural parameter w_{11} cannot be easily converted to a distributional parameter (Rubinstein 1992). However, it is model-dependent and may not be suitable for general models. Thus, we do not consider LR method for this example.

In this example we have six sample-mean estimators (denoted as Estimators 1 to 6). Estimator 1 is the one given by Equation (17), Estimator 2 is the one given by Equation (18), and Estimators 3 to 6 are the ones given by Equation (19) applied to Γ_2 to Γ_5 respectively. The combined estimator is calculated by combining Estimators 1 to 6.

We report the performances of the kernel estimators, the six sample-mean estimators and the combined estimators for different sample sizes in Table 3. From the table we see that the findings of Section 5.1 also hold in this example, except that the estimator conditioning on the idiosyncratic risk factor (i.e., Estimator 1) has a lower standard error than the estimators conditioning on the

Table 3 The estimates and their standard errors for the CreditRisk$^+$ model.

Sample size n	10^4		10^5		10^6	
Case A	M	s.e. ($\times 10^{-3}$)	M	s.e. ($\times 10^{-4}$)	M	s.e. ($\times 10^{-4}$)
Estimator 1	0.0100	0.60	0.0099	1.9	0.0099	0.61
Estimator 2	0.0091	1.9	0.0099	6.4	0.0095	2.0
Estimator 3	0.0071	2.3	0.0104	9.4	0.0093	2.8
Estimator 4	0.0071	2.1	0.0098	8.5	0.0096	2.9
Estimator 5	0.0094	2.7	0.0097	8.6	0.0093	2.8
Estimator 6	0.0075	2.0	0.0099	9.1	0.0095	2.9
Combined	0.0095	0.58	0.0099	1.9	0.0098	0.59
Kernel	0.0090	1.0	0.0100	4.4	0.0095	1.7
Case B	M	s.e.	M	s.e.	M	s.e.
Estimator 1	23.85	1.3	23.65	0.41	23.52	0.13
Estimator 2	21.97	3.9	23.49	1.35	22.74	0.42
Estimator 3	16.95	4.9	24.48	1.98	22.38	0.59
Estimator 4	17.79	4.4	23.04	1.79	23.00	0.60
Estimator 5	22.02	5.6	22.99	1.82	22.44	0.59
Estimator 6	17.78	4.1	23.56	1.93	22.84	0.61
Combined	22.87	1.2	22.86	0.40	22.84	0.12
Kernel	21.59	2.2	23.46	0.92	22.86	0.36

Note: Estimators 1,2 and 3–6 are specified by Equations (17),(18), and (19), respectively.

common risk factors (i.e., Estimators 2 to 6). That implies that it is hard to identify which one is better in advance among the estimators derived by conditioning on either idiosyncratic factors or macroeconomic factors.

We also report the computation times of different estimators in Table 4. It is interesting to find that the computational time of computing Estimator 1 and that of computing the kernel estimator are almost the same, and also in the same order of computing other estimators. This is because the computational complexities of all estimators are $O(m)$ when the parameter w_{11} is an idiosyncratic parameter. The time of computing the combined estimator, which is calculated based on the six sample-mean estimators, is less than the total time of computing each of them. This finding suggests the advantage of using a linear combination of multiple estimators for idiosyncratic parameters.

Table 4 Time (in second) taken to compute each estimator with 100 replications.

Sample size n	10^4	10^5	10^6
Estimator 1	0.05	0.40	3.47
Estimator 2	0.10	0.65	5.51
Estimators 3–6	0.20	1.68	14.79
Combined	0.26	2.09	18.43
Kernel	0.06	0.40	3.48

5.3. A Doubly Stochastic Model

We consider a doubly stochastic model where both $S_c(t)$ and $S_i(t), i=1,\dots,m$, follow CIR processes. Specifically, suppose that

$$\mu_c(t,S_c(t)) = \kappa_c\,(\mu_c - S_c(t)) \text{ and } \sigma_c(t,S_c(t)) = \sigma_c\sqrt{S_c(t)},$$
$$\mu_i(t,S_i(t)) = \kappa_i\,(\mu_i - S_i(t)) \text{ and } \sigma_i(t,S_i(t)) = \sigma_i\sqrt{S_i(t)}\,,\ i=1,...,m, \tag{30}$$

where $\kappa_c = 0.002$, $\mu_c = 0.1$, $\sigma_c = 0.02$, and $\kappa_i = 0.001$, $\mu_i = 0.07$, $\sigma_i = 0.01$ for all $i=1,\dots,m$. The initial values $S_c(0) = 0.1$ and $S_i(0) = 0.08$ for all $i=1,\dots,m$, and the time horizon $T=1$. We are interested in estimating the sensitivities of the expected performances of the two performance functions with respect to $S_1(0)$. In the numerical study we use the discretization scheme introduced in Section 4.3 to evaluate Λ_i, the integral of the default intensity of obligor i for all $i=1,\dots,m$.

The LR method cannot be directly applied to doubly stochastic models if we use Equation (25) under Euler scheme because the structural parameter $S_1(0)$ cannot be fully converted to a distributional parameter. To make LR method work, we approximate Λ_i by

$$\widehat{\Lambda}_i = \sum_{j=1}^{k} \widehat{\lambda}_i(t_j)\Delta t = \sum_{j=1}^{k}\left[\widehat{S}_c(t_j) + \widehat{S}_i(t_j)\right]\Delta t, \tag{31}$$

and use the conditional technique of Hong and Liu (2010). After some derivation (see the e-companion EC.1 for detailed derivation), we have the LR estimator

$$p'(S_1(0)) = \mathrm{E}\left\{ g\,(L)\cdot \frac{\left(\widehat{S_1}(t_1) - \kappa_1\mu_1\Delta t\right)^2 - \sigma_1^2 S_1(0)\Delta t - (1-\kappa_1\Delta t)^2\, S_1^2(0)}{2\sigma_1^2 S_1^2(0)\Delta t}\right\}. \tag{32}$$

Besides the LR estimator, we have three other sample-mean estimators. Estimator 1 is given by Equation (26), and Estimators 2 and 3 are given by Equations (27) and (28) respectively. The combined estimator is calculated by combining Estimators 1 to 3.

In this example we fix the sample size $n = 10^6$ and investigate how the number of time steps affects the accuracy of the estimators. We report the estimates ($\bar{M}$) and their standard errors (s.e.) for different numbers of time steps in Table 5 with $\widehat{\Lambda}_i$ given by Equation (25) and Table 6 with $\widehat{\Lambda}_i$ given by Equation (31). From both tables, we see that the performances of Estimators 2, 3 and the LR estimator deteriorate as the number of time steps goes up. This deterioration is typical for estimators that condition on the last time step (see, for instance, Hong and Liu (2010) for some more examples). However, in our example, both Estimator 1 (and thus the combined estimator) and the kernel estimator are not affected by the numbers of time steps. In this example. the LR estimator performs poorly when the sample size is $n = 10^6$. To make sure the LR estimator is correct, we run the experiment with time step $k = 4$ and sample size $n = 10^{10}$, and other parameters

Table 5 The estimates and their standard errors for the doubly stochastic model with $\widehat{\Lambda}_i$ given by Equation (25).

Time steps k	2		4		12	
Case A	$\bar{M}$	s.e. ($\times 10^{-3}$)	$\bar{M}$	s.e. ($\times 10^{-3}$)	$\bar{M}$	s.e. ($\times 10^{-3}$)
Estimator 1	0.0516	0.20	0.0519	0.20	0.0518	0.20
Estimator 2	0.0491	3.7	0.0587	6.9	0.0412	12.7
Estimator 3	0.0573	2.7	0.0499	4.2	0.0447	9.1
Combined	0.0516	0.20	0.0519	0.20	0.0518	0.20
Kernel	0.0518	0.64	0.0523	0.64	0.0528	0.64
Case B	$\bar{M}$	s.e.	$\bar{M}$	s.e.	$\bar{M}$	s.e.
Estimator 1	119.22	0.42	119.90	0.42	119.53	0.42
Estimator 2	113.30	7.74	133.13	14.60	99.30	26.7
Estimator 3	130.89	5.67	114.51	8.93	106.02	19.2
Combined	119.24	0.42	119.89	0.42	119.53	0.42
Kernel	119.67	1.34	120.68	1.35	121.78	1.36

Note: Estimators 1,2 and 3 are specified by Equations (26),(27), and (28), respectively.

Table 6 The estimates and their standard errors for the doubly stochastic model with $\widehat{\Lambda}_i$ given by Equation (31).

Time steps k	2		4		12	
Case A	$\bar{M}$	s.e. ($\times 10^{-3}$)	$\bar{M}$	s.e. ($\times 10^{-3}$)	$\bar{M}$	s.e. ($\times 10^{-3}$)
Estimator 1	0.0514	0.20	0.0516	0.20	0.0520	0.20
Estimator 2	0.0534	3.9	0.0462	6.1	0.0479	13.9
Estimator 3	0.0537	2.6	0.0448	4.0	0.0496	9.7
Combined	0.0514	0.20	0.0516	0.20	0.0520	0.20
Kernel	0.0513	0.64	0.0516	0.64	0.0523	0.64
LR	0.0680	187.2	0.1864	265.3	-0.4638	460.0
Case B	$\bar{M}$	s.e.	$\bar{M}$	s.e.	$\bar{M}$	s.e.
Estimator 1	118.80	0.42	119.22	0.42	120.13	0.42
Estimator 2	122.47	8.14	106.54	12.84	109.85	29.3
Estimator 3	123.59	5.53	103.99	8.42	115.23	20.5
Combined	118.80	0.42	119.21	0.42	120.13	0.42
Kernel	118.46	1.34	119.08	1.34	120.90	1.35
LR	183.57	424.31	414.97	601.39	-1186.6	1042.5

Note: Estimators 1,2 and 3 are specified by Equations (26),(27), and (28), respectively.

remaining the same. It takes around 84 hours to obtain the estimate 0.0537 with a standard error 0.0024.

We report the computational times of computing Estimator 1, combined estimator, kernel estimator and the LR estimator when sample size $n = 10^6$ in Table 7. From Table 7, we find that it almost takes the same amount of time to compute Estimator 1, kernel estimator and LR estimator. Moreover, the time for the combined estimator increases slightly compared with other estimators even though the combined estimator is obtained after computing the three sample-mean estimates.

Table 7 Time (in second) taken to compute each estimator with 100 replications.

Time steps k	2	4	12
Estimator 1	18.60	28.94	68.13
Combined	30.68	35.72	84.81
Kernel	18.67	24.96	68.48
LR	17.38	23.54	67.28

6. Conclusions

In this paper we derive a closed-form expression for the sensitivities of the expected value of a performance function of a portfolio credit loss with respect to a parameter in the model of joint defaults. We show that the differentiability does not depend on the differentiability of the performance function. Based on the closed-form expression, which is in the form of a conditional expectation, we propose two methods to estimate the sensitivities. First, we propose a kernel estimator which is typically easy to use, applicable to many models of joint default, but has a rate of convergence slower than $n^{-1/2}$. Second, we propose to use model information to further convert the conditional expectation to unconditioned expectations and use sample-mean estimators to estimate the sensitivities. We demonstrate the second method on three commonly used models of joint defaults, latent variable, Bernoulli mixture and doubly stochastic models. We show that multiple sensitivities can be derived based on the second method. This suggests to combined all sample-mean estimators to further improve the estimation performance. We test the kernel estimator, various sample-mean estimators and the combined estimator through three examples and also compare them with the estimators derived by LR method, and the numerical results show that various sample-mean estimators by our method often work well.

For future work, we will study how to estimate sensitivities when the joint defaults are modeled using frailty models or self excited models, which are closely related to doubly stochastic models but are capable of capturing default clustering effects (see, for instance Giesecke et al. (2010) for a thorough introduction of these models). In these models, the intensity function may not be continuous with respect to the parameter that we want to take derivative to. Therefore, the conditions of Theorem 1 may not hold and the sensitivities may be more difficult to estimate than under doubly stochastic models.

Acknowledgments

The authors would like to thank the associate editor and two anonymous referees for their insightful and detailed comments that have significantly improved this paper.

References

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

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

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

Bernis, G., E. Gobet and A. Kohatsu-Higa. 2003. Monte Carlo evaluation of Greeks for multidimensional barrier and lookback options. *Mathematical Finance*, **13** 99-113.

Bock, R. D. 1985. *Multivariate Statistical Methods in Behavioral Research*, Second Edition. Scientific Software Intl, Chicago.

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

Chan, J.H. and M.S. Joshi. 2012. Fast Monte-Carlo Greeks for financial products with discontinuous pay-offs. *Mathematical Finance*, forthcoming.

Chen, N. and P. Glasserman. 2007. Malliavin Greeks without Malliavin calculus. *Stochastic Processes and their Applications*, **117** 1689-1723.

Chen, Z. and P. Glasserman. 2008. Sensitivity estimates for portfolio credit derivatives using Monte Carlo. *Finance and Stochastics*, **12** 507-540.

Credit-Suisse-Finanical-Products. 1997. CreditRisk$^+$: a credit risk management framework. Technical report, available from `http://www.csfb.com/institutional/research/assets/creditrisk.pdf`.

Duffie, D. and N. Gârleanu. 2001. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, **57** 41-59.

Duffie, D and Singleton, KJ, 1999. Modeling term structures of defaultable bonds. 1999. *Review of Financial Studies*, **12**(4) 687-720.

Durrett, R. 2005. *Probability: Theory and Examples*, Third Edition. Duxbury Press, Belmont.

Embrechts, P., F. Lindskog and A. McNeil. 2003. Modeling dependence with copulas and applications to risk management. Chapter 8 in S. T. Rachev (Eds.), *Handbook of Heavy Tailed Distribution in Finance*, Elsevier/North-Holland, Amsterdam, The Netherlands.

Frey, R. and A. McNeil. 2003. Dependent defaults in models of portfolio credit risk. *Journal of Risk*, **6** 59-92.

Fu, M. C. 2008. What you should know about simulation and derivatives. *Naval Research Logistics*, **55** 723-736.

Fu, M. C., L. J. Hong and J. Q. Hu. 2009. Conditional Monte Carlo estimation of quantile sensitivities.*Management Science*, **55** 2019-2017.

Fu, M. C. and J. Q. Hu. 1995. Sensitivity analysis for Monte Carlo simulation simulation of option pricing. *Probability in the Engineering and Information Sciences*, **9** 417-446.

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

Giesecke, K., H. Kakavand, M. Mousavi and H. Takada. 2010. Exact and efficient simulation of correlated defaults. *SIAM Journal on Financial Mathematics*, **1** 868-896.

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

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

Glynn, P. W. 1987. Likelihood ratio gradient estimation: an overview. *Proceedings of the 1987 Winter Simulation Conference*, 366-375.

Gupton, G. M., C. C. Finger and M. Bhatia. 1997. *CreditMetrics^{TM} - Technical Document.* J. P. Morgan & Co., New York.

Heidergott, B., F. J. Vázquez–Abad, G. Pflug and T. F. Yuan. 2010. Gradient estimation for discrete-event systems by measure-valued differentiation. *ACM Transactions on Modeling and Computer Simulation*, **20** 5:1-5:28.

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

Homescu, Cristian. 2011. Adjoints and automatic (algorithmic) differentiation in computational finance. Available at SSRN: `http://ssrn.com/abstract=1828503`.

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

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

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

Joshi, M. S. and D.S. Kainth. 2004. Rapid computation of prices and deltas of nth to default swaps in the Li Model. *Quantitative Finance*, **4** 266-275.

Kealhofer, S. and J. Bohn. 2001. Portfolio management of default risk. KMV working paper.

L'Ecuyer, P. 1991. An overview of derivative estimation. *Proceedings of the 1991 Winter Simulation Conference*, 207-217.

Li, D. 2000. On default correlation: a Copula function approach. *Journal of Fixed Income*, **9** 43-54.

Liu, G. and L. J. Hong. 2011. Kernel estimation of the Greeks for options with discontinuous payoffs. *Operations Research*, **59** 96-108.

Mashal, R. and A. Zeevi. 2003. Beyond correlation: extreme co-movements between financial assets. Working paper.

Merton, R. 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* **29** 449-470.

Pflug, G. 1988. Derivatives of probability measures-concepts and applications to the optimization of stochastic systems. In Varaiya, P. and A. B. Kurzhanski, (Eds.), *Lecture Notes in Control and Information Sciences*, **103** 252-274.

Rubinstein, R. Y. 1989. Sensitivity analysis of computer simulation models via the score efficient. *Operations Research*, **37** 72-81.

Rubinstein, R.Y. 1992. Sensitivity analysis of discrete event systems by the “push-out” method. *Annals of Operations Research*, **39** 229-250.

Electronic Companion

EC.1. Technical Remarks

Proof of Corollary 1. Let $X(\theta) = \max_{i=1,\dots,k} X_i(\theta)$. Then $X(\theta)$ is non-differentiable when there exist some $i \neq j$ such that $X(\theta) = X_i(\theta) = X_j(\theta)$ or $X_\ell(\theta)$ is non-differentiable for $\ell = 1, 2, \dots, k$. Because $X_\ell(\theta)$ is differentiable w.p.1 and $\Pr\{X_i(\theta) = X_j(\theta)\} = 0$, $X(\theta)$ is also differentiable w.p.1 at any $\theta \in \mathcal{N}(\theta_0)$. Let $\mathcal{K} = \sum_{i=1}^k \mathcal{K}_i$. Then, $\mathrm{E}[\mathcal{K}] < \infty$ and $|X(\theta_0 + \Delta\theta) - X(\theta_0)| \leq \mathcal{K} \cdot |\Delta\theta|$ for any $\Delta\theta$ that is close enough to 0.

Let $i^* = \mathrm{argmax}_{i=1,\dots,k} X_i(\theta)$. Then, $X'(\theta) = X'_{i^*}(\theta)$ w.p.1. Note that

$$\begin{aligned}\psi(\theta,t) &= \mathrm{E}\left[X'(\theta); X(\theta) = t\right] = \mathrm{E}\left[X'_{i^*}(\theta); X_{i^*}(\theta) = t\right] = \sum_{i=1}^k \mathrm{E}\left[X'_i(\theta)\mathbf{1}_{\{i^*=i\}}; X_i(\theta) = t\right] \\ &= \sum_{i=1}^k \mathrm{E}\left[X'_i(\theta) \prod_{j=1, j\neq i}^k \mathbf{1}_{\{X_j(\theta) < X_i(\theta)\}}; X_i(\theta) = t\right] = \sum_{i=1}^k \psi_i(\theta, t).\end{aligned}$$

Therefore, $\psi(\theta,t)$ is continuous at $(\theta_0, 0)$ because $\psi_i(\theta,t)$ are continuous at $(\theta_0, 0)$. Then, the conclusion of this corollary follows directly from Lemma 1.

Verification of Conditions (a1)–(a3). We next verify that Conditions (a1)–(a3) in Section 3.1 imply the conditions of Theorem 1. Note that $X_i(\theta) = \epsilon_i - \Upsilon_i(\theta)$ and ϵ_i is a continuous random variable that is independent of all other variable factors, so Conditions 1 and 2 of Theorem 1 are satisfied. Furthermore, note that

$$\psi_i(\theta, y) = -\mathrm{E}\left[\Upsilon'_i(\theta) \prod_{j=1, j\neq i}^m \mathbf{1}_{\{(-1)^{a_j}[\epsilon_j - \Upsilon_j(\theta)] < y\}}; \epsilon_i = \Upsilon_i(\theta) + y\right].$$

Let $\boldsymbol{\xi}_i$ denote all random factors other than ϵ_i. Note that $\boldsymbol{\xi}_i$ and ϵ_i are mutually independent, and $\Upsilon_i(\theta)$ and $\Upsilon'_i(\theta)$ are fully determined when $\boldsymbol{\xi}_i$ is given. By conditioning on $\boldsymbol{\xi}_i$, we have

$$\begin{aligned}\psi_i(\theta, y) &= -\mathrm{E}\left\{\mathrm{E}\left[\Upsilon'_i(\theta) \prod_{j=1, j\neq i}^m \mathbf{1}_{\{(-1)^{a_j}[\epsilon_j - \Upsilon_j(\theta)] < y\}}; \epsilon_i = \Upsilon_i(\theta) + y \;\middle|\; \boldsymbol{\xi}_i\right]\right\} \\ &= -\mathrm{E}\left[\Upsilon'_i(\theta) \prod_{j=1, j\neq i}^m \mathbf{1}_{\{(-1)^{a_j}[\epsilon_j - \Upsilon_j(\theta)] < y\}} \cdot \mathrm{E}\left[1; \epsilon_i = \Upsilon_i(\theta) + y \;\middle|\; \Upsilon_i(\theta)\right]\right] \\ &= -\mathrm{E}\left[\Upsilon'_i(\theta) \prod_{j=1, j\neq i}^m \mathbf{1}_{\{(-1)^{a_j}[\epsilon_j - \Upsilon_j(\theta)] < y\}} \cdot f_{\epsilon_i}\left(\Upsilon_i(\theta) + y\right)\right], \qquad \text{(EC.1)}\end{aligned}$$

where Equation (EC.1) follows from Equation (2) and the independence between ϵ_i and $\Upsilon_i(\theta)$. Given the conditions of the theorem, we can easily check that the term inside of the expectation of Equation (EC.1) is continuous w.p.1 with respect to (θ, y) and is dominated by a random variable

K_iB_i with a finite first moment. Then, by the dominated convergence theorem (Durrett 2005), $\psi_i(\theta, y)$ is continuous in (θ, y), i.e., Condition 3 of Theorem 1 also holds. Therefore, the conclusion of Theorem 1 holds.

Then, by Theorem 1,

$$\begin{aligned}
p'(\theta) &= \sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right)-g\left(L_{-i}\right)\right]\cdot \Upsilon_i'(\theta)\,;\, \epsilon_i = \Upsilon_i\right\} \\
&= \sum_{i=1}^{m} \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right)-g\left(L_{-i}\right)\right]\cdot \Upsilon_i'(\theta)\cdot f_{\epsilon_i}(\Upsilon_i)\right\},
\end{aligned}$$

where the last equation can be derived similar to Equation (EC.1).

Derivation of the LR estimator in Section 5.3. Since the parameter $S_1(0)$ can be viewed as a distributional parameter under the discretization scheme in Equation (31), we now apply the conditioning technique in Hong and Liu (2010) to derive the LR estimator in Equation (32).

Let $f_i(\cdot|s_{i-1})$ denote the conditional density function of $\widehat{S_1}(t_i)$ given that $\widehat{S_1}(t_{i-1}) = s_{i-1}$, $i = 1, 2, \ldots, k$. By Equation (24),

$$f_1\left(x|S_1(0)\right) = \frac{1}{\sigma_1\left(0, S_1(0)\right)\sqrt{\Delta t}}\cdot\phi\left(\frac{x - S_1(0) - \mu_1\left(0, S_1(0)\right)\Delta t}{\sigma_1\left(0, S_1(0)\right)\sqrt{\Delta t}}\right),$$

where $\phi(\cdot)$ is the density function of standard normal distribution. Then, the SF can be expressed as

$$\begin{aligned}
\mathrm{SF} &= \frac{d}{dS_1(0)}\log\left(f_1\left(\widehat{S_1}(t_1)|S_1(0)\right)\right) \\
&= \frac{\widehat{S_1}(t_1) - S_1(0) - \mu_1\left(0, S_1(0)\right)\Delta t}{\sigma_1\left(0, S_1(0)\right)\sqrt{\Delta t}}\cdot\left[\frac{1}{\sigma_1\left(0, S_1(0)\right)\sqrt{\Delta t}}\left(1 + \frac{d\mu_1\left(0, S_1(0)\right)}{dS_1(0)}\right)\right. \\
&\quad\left. + \frac{\widehat{S_1}(t_1) - S_1(0) - \mu_1\left(0, S_1(0)\right)\Delta t}{\sigma_1^2\left(0, S_1(0)\right)\sqrt{\Delta t}}\cdot\frac{d\sigma_1\left(0, S_1(0)\right)}{dS_1(0)}\right] - \frac{1}{\sigma_1\left(0, S_1(0)\right)}\frac{d\sigma_1\left(0, S_1(0)\right)}{dS_1(0)} \\
&= \frac{\left(\widehat{S_1}(t_1) - \kappa_1\mu_1\Delta t\right)^2 - \sigma_1^2 S_1(0)\Delta t - \left(1 - \kappa_1\Delta t\right)^2 S_1^2(0)}{2\sigma_1^2 S_1^2(0)\Delta t},
\end{aligned} \tag{EC.2}$$

where Equation (EC.2) is by plugging Equation (30) with $i = 1$. Then, the LR estimator is

$$p'(S_1(0)) = \mathrm{E}\left\{g\left(L\right)\cdot\mathrm{SF}\right\} = \mathrm{E}\left\{g\left(L\right)\cdot\frac{\left(\widehat{S_1}(t_1) - \kappa_1\mu_1\Delta t\right)^2 - \sigma_1^2 S_1(0)\Delta t - \left(1 - \kappa_1\Delta t\right)^2 S_1^2(0)}{2\sigma_1^2 S_1^2(0)\Delta t}\right\}.$$

EC.2. SPA Estimators in A Latent Variable Model

We now use Example 3 (The model of Bassamboo et al. (2008)) to derive SPA estimators. We consider only the case that the performance function is $g(L) = \mathbf{1}_{\{L>y\}}$. In fact, it is not clear how to derive SPA estimators for general performance functions. Conditioning on $\{Z, \mathcal{E}\}$,

$$\mathrm{E}\left[g(L)|Z, \mathcal{E}\right] \;=\; \Pr\left\{\sum_{i=1}^{m} l_i\mathbf{1}_{\{Y_i\le d_i\}} > y\,\Big|\, Z, \mathcal{E}\right\}$$

$$
\begin{aligned}
&= \Pr\left\{\epsilon_1 \le \frac{\theta d_1\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\Big| Z,\mathcal{E}\right\}\cdot \Pr\left\{\sum_{i=2}^m l_i \mathbf{1}_{\{Y_i\le d_i\}} > y - l_1\Big| Z,\mathcal{E}\right\}\\
&\quad + \Pr\left\{\epsilon_1 > \frac{\theta d_1\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\Big| Z,\mathcal{E}\right\}\cdot \Pr\left\{\sum_{i=2}^m l_i \mathbf{1}_{\{Y_i\le d_i\}} > y \Big| Z,\mathcal{E}\right\}.
\end{aligned}
$$

We can recursively use the same approach to deal with $\Pr\left\{\sum_{i=2}^m l_i \mathbf{1}_{\{Y_i\le d_i\}} > y - l_1\Big| Z,\mathcal{E}\right\}$ and $\Pr\left\{\sum_{i=2}^m l_i \mathbf{1}_{\{Y_i\le d_i\}} > y\Big| Z,\mathcal{E}\right\}$. Then, after m iterations, we have

$$
\mathrm{E}\left[g(L)|Z,\mathcal{E}\right] = \sum_{s\in\mathscr{S}(m)} \prod_{i\in s^1} F_{\epsilon_i}\left(\frac{\theta d_i\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot \prod_{i\in s^0} \bar{F}_{\epsilon_i}\left(\frac{\theta d_i\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot \mathbf{1}_{\{\sum_{i\in s^1} l_i > y\}}, \tag{EC.3}
$$

where $F_{\epsilon_i}(\cdot)$ is the cumulative distribution function (cdf) of ϵ_i, $\bar{F}_{\epsilon_i}(\cdot) = 1 - F_{\epsilon_i}(\cdot)$, and $\mathscr{S}(m) = \{0,1\}^m$ with s^1 denoting the set of default obligors and s^0 denoting the set of non-default obligors. Because the closed-form expresson in Equation (EC.3) is Lipschitz continuous, then we can apply SPA method to obtain a SPA estimator,

$$
\begin{aligned}
p'(\theta) &= \frac{\partial}{\partial\theta}\mathrm{E}\left\{\mathrm{E}\left[g(L)|Z,\mathcal{E}\right]\right\} = \mathrm{E}\left\{\frac{\partial}{\partial\theta}\mathrm{E}\left[g(L)|Z,\mathcal{E}\right]\right\}\\
&= \mathrm{E}\left\{\sum_{s\in\mathscr{S}(m)}\left[\sum_{j\in s^1} f_{\epsilon_j}\left(\frac{\theta d_j\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot\frac{d_j\mathcal{E}}{\sqrt{1-\rho^2}}\cdot\prod_{i\in s^1, i\ne j} F_{\epsilon_i}\left(\frac{\theta d_i\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\right.\right.\\
&\quad \cdot\prod_{i\in s^0}\bar{F}_{\epsilon_i}\left(\frac{\theta d_i\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot\mathbf{1}_{\{\sum_{i\in s^1} l_i>y\}} - \sum_{j\in s^0} f_{\epsilon_j}\left(\frac{\theta d_j\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot\frac{d_j\mathcal{E}}{\sqrt{1-\rho^2}}\\
&\quad \left.\left.\cdot\prod_{i\in s^1} F_{\epsilon_i}\left(\frac{\theta d_i\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot\prod_{i\in s^0, i\ne j}\bar{F}_{\epsilon_i}\left(\frac{\theta d_i\mathcal{E}-\rho Z}{\sqrt{1-\rho^2}}\right)\cdot\mathbf{1}_{\{\sum_{i\in s^1} l_i>y\}}\right]\right\}.
\end{aligned}\tag{EC.4}
$$

Similarly, by conditioning on $\{\mathcal{E},\epsilon_1,\ldots,\epsilon_m\}$, we know that

$$
\begin{aligned}
\mathrm{E}\left[g(L)|\mathcal{E},\epsilon_1,\ldots,\epsilon_m\right] &= \sum_{s\in\mathscr{S}(m)} \Pr\left\{\bigcup_{i\in s^1}\left\{Z\le\frac{\theta d_i\mathcal{E}-\sqrt{1-\rho^2}\epsilon_i}{\rho}\right\}\right.\\
&\quad \left.\bigcup_{i\in s^0}\left\{Z>\frac{\theta d_i\mathcal{E}-\sqrt{1-\rho^2}\epsilon_i}{\rho}\right\}\Big|\mathcal{E},\epsilon_1,\ldots,\epsilon_m\right\}\cdot\mathbf{1}_{\{\sum_{i\in s^1} l_i>y\}}\\
&= \sum_{s\in\mathscr{S}(m)}\left[F_Z\left(\frac{\theta d_{i_*^1}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^1}}{\rho}\right) - F_Z\left(\frac{\theta d_{i_*^0}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^0}}{\rho}\right)\right]\\
&\quad \cdot\mathbf{1}_{\{\theta d_{i_*^1}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^1}>\theta d_{i_*^0}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^0}\}}\cdot\mathbf{1}_{\{\sum_{i\in s^1} l_i>y\}},
\end{aligned}
$$

where $i_*^1 = \mathrm{argmin}_{i\in s^1}\left\{\frac{\theta d_i\mathcal{E}-\sqrt{1-\rho^2}\epsilon_i}{\rho}\right\}$ and $i_*^0 = \mathrm{argmax}_{i\in s^0}\left\{\frac{\theta d_i\mathcal{E}-\sqrt{1-\rho^2}\epsilon_i}{\rho}\right\}$. Then,

$$
\begin{aligned}
p'(\theta) &= \mathrm{E}\left\{\sum_{s\in\mathscr{S}(m)}\left[f_Z\left(\frac{\theta d_{i_*^1}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^1}}{\rho}\right)\cdot\frac{d_{i_*^1}\mathcal{E}}{\rho} - f_Z\left(\frac{\theta d_{i_*^0}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^0}}{\rho}\right)\cdot\frac{d_{i_*^0}\mathcal{E}}{\rho}\right]\right.\\
&\quad \left.\cdot\mathbf{1}_{\{\theta d_{i_*^1}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^1}>\theta d_{i_*^0}\mathcal{E}-\sqrt{1-\rho^2}\epsilon_{i_*^0}\}}\cdot\mathbf{1}_{\{\sum_{i\in s^1} l_i>y\}}\right\}
\end{aligned}\tag{EC.5}
$$

There are several things we would like to point out. First, the reason we can easily derive multiple SPA estimators is because of the general closed-form expression in Equation (3) (also Equations (7) and (8) in Section 3), which implies the possibility of deriving SPA estimators under some transformations. In fact, if we write Equations (11) and (13) into the combinatorial form, we find that Equation (EC.4) is the same as Equation (11) and Equation (EC.5) is the same as Equation (13). Second, however, without the general form of Equation (3), it is not easy to convert Equations (EC.4) and (EC.5) into the neat form as in (11) and (13). Therefore, the computational complexity of both (EC.4) and (EC.5) are at least $O(2^m)$, which is much higher compared with their counterpart in (11) and (13) in the paper.

EC.3. Extension to Sensitivities of VaR and CVaR

Our method can be extended to compute the sensitivity of VaR and Conditional VaR (CVaR) when l_i are mutually independent continuous random variables and independent with X_j for all $i,j=1,2,\dots,m$. Then the cdf $F_L(\cdot)$ of the loss function L is continuous except at the point 0. In general, when computing the sensitivity of VaR or CVaR with respect L, we are interested in the event that the loss is beyond some large threshold y (which is typically greater than 0), so the discontinuity at 0 will not cause any problem. Note that if we restrict l_i to be constant, then L become a discrete random variable with cdf $F_L(\cdot)$ be a step function. Then the sensitivity of the VaR or the CVaR may become hard to analyze since the perturbation of θ may lead to a dramatic change or no change of the value of VaR depending on whether the level α (defined later on) is at an exact probability mass point or not. Interested readers may be referred to Rockafellar and Uryasev (2002) for representing the CVaR for general loss functions as an expectation of properly modified tail distribution.

Considering l_i, $i=1,2,\dots,m$, as continuous random variables, then L is a continuous random variable except at the point 0, where $\Pr(L=0)=\Pr(X_i\geq 0, \text{ for all } i=1,2,\dots,m)>0$. Generally, when computing the sensitivity of VaR with respect to L, we are interested in the event that the loss is beyond some large threshold y, which is typically greater than 0.

EC.3.1. The sensitivity of VaR

Let $F_L(y)=\Pr(L\leq y)$ be the cdf of L. Define the VaR at level α (α-VaR) of L as $v_\alpha=\inf\{y: F_L(y)\geq\alpha\}$. Then for $v_\alpha\neq 0$, the equality can be achieved with $F_L(v_\alpha)=\alpha$. To estimate $v'_\alpha(\theta)$, we write $F_L(v_\alpha)=\alpha$ as $F_L(v_\alpha(\theta),\theta)=\alpha$ and take derivative with respect to θ on both sides, which yields

$$\partial_y F_L(y,\theta)|_{y=v_\alpha(\theta)}\cdot v'_\alpha(\theta)+\partial_\theta F_L(y,\theta)|_{y=v_\alpha(\theta)}=0.$$

Then,

$$v'_\alpha(\theta) = \frac{\partial_\theta F_L(y,\theta)}{\partial_y F_L(y,\theta)}\bigg|_{y=v_\alpha(\theta)} = \frac{\partial_\theta F_L(y,\theta)}{f_L(y,\theta)}\bigg|_{y=v_\alpha(\theta)}, \tag{EC.6}$$

where $f_L(y,\theta)$ denote the pdf of L. Then our goal is to calculate $\partial_\theta F_L(v_\alpha(\theta),\theta)$ and $f_L(v_\alpha(\theta),\theta)$ respectively. The numerator on the right-hand-side of Equation (EC.6)

$$\partial_\theta F_L(v_\alpha(\theta),\theta) = \frac{d}{d\theta}\mathrm{E}\left[\mathbf{1}_{\{L(\theta)\le y\}}\right]\big|_{y=v_\alpha(\theta)}. \tag{EC.7}$$

Letting $g(L) = \mathbf{1}_{\{L(\theta)\le y\}}$, then Equation (EC.7) can be handled by the conditional technique in Section 2. We just summarize the result as follows,

$$\frac{d}{d\theta}\mathrm{E}\left[\mathbf{1}_{\{L(\theta)\le y\}}\right]\Big|_{y=v_\alpha(\theta)} = -\sum_{i=1}^{m} \mathrm{E}\left\{\left[\mathbf{1}_{\{L_{-i}+l_i\le y\}} - \mathbf{1}_{\{L_{-i}\le y\}}\right]\cdot X'_i(\theta)\,;\, X_i = 0\right\}\Big|_{y=v_\alpha(\theta)},$$

where $L_{-i} = \sum_{j=1,j\ne i}^{m} l_j \cdot \mathbf{1}_{\{X_j<0\}}$. The denominator on the right-hand-side of Equation (EC.6)

$$\begin{aligned} f_L(y,\theta) &= \lim_{\Delta y\to 0}\frac{F_L(y+\Delta y)-F_L(y)}{\Delta y} \\ &= \lim_{\Delta y\to 0}\frac{\mathrm{E}\left[\mathbf{1}_{\{L(\theta)\le y+\Delta y\}} - \mathbf{1}_{\{L(\theta)\le y\}}\right]}{\Delta y}. \end{aligned}$$

Then, we may choose a proper Δy to approximate $f_L(y,\theta)$ at $y = v_\alpha(\theta)$, which can be viewed as the kernel method. Another way is to apply the conditional Monte Carlo derived in Fu et al. (2009). Basically, it requires to find some random variable $Y(\theta)$ such that

$$F_L(y,\theta) = \mathrm{E}[\Pr\{L(\theta)\le y|Y(\theta)\}] = \mathrm{E}\left[G(y,Y(\theta),\theta)\right],$$

where $G(y,Y(\theta),\theta)$ is differential w.p.1 with respect to y and

$$|G(y+\Delta y, Y(\theta),\theta) - G(y,Y(\theta),\theta)| \le K|\Delta y|, \tag{EC.8}$$

for some random variable K with $\mathrm{E}[K]<\infty$. Then,

$$f_L(y,\theta)|_{y=v_\alpha(\theta)} = \mathrm{E}\left[\partial_y G(y,Y(\theta),\theta)\right]\big|_{y=v_\alpha(\theta)}. \tag{EC.9}$$

The closed-form expression of G may be complicated, but the idea behind is straightforward. By conditioning on some random variable $Y(\theta)$, we can write F_L as an expectation of a function of cdf's and pdf's with closed-form. In addition, the condition of (EC.8) can be easily verified after giving the closed-form of G (which should be satisfied due to the differentiability of F_L at $y = v_\alpha(\theta)$). Equations (EC.7) and (EC.9) together provide the estimator $v'_\alpha(\theta)$ in Equation (EC.6). Now we derive a general closed-form expression of G given the form of loss function L in our paper. Suppose we may write $X_i = \eta_i - A_i$, where η_i is an idiosyncratic random variable which is independent of all other random variables. Let $H_i(\cdot)$ and $h_i(\cdot)$ be the cdf and pdf of η_i, respectively. Let $F_i(\cdot)$ be the cdf of l_i and $\bar{F}_i(\cdot) = 1 - F_i(\cdot)$. Then,

$$F_L(y,\theta) = \mathrm{E}\left[\Pr\left\{\sum_{i=1}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \le y \,\middle|\, A_1, A_2, \ldots, A_m\right\}\right]. \tag{EC.10}$$

For any $y > 0$,

$$\begin{aligned}
&\Pr\left\{\sum_{i=1}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \le y \,\middle|\, A_1, A_2, \ldots, A_m\right\} \\
&= \Pr\{\eta_1 < A_1 | A_1\} \cdot \Pr\left\{l_1 \le y - \sum_{i=2}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \,\middle|\, A_2, A_3, \ldots, A_m\right\} \\
&\quad + \Pr\{\eta_1 \ge A_1 | A_1\} \cdot \Pr\left\{\sum_{i=2}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \le y \,\middle|\, A_2, A_3, \ldots, A_m\right\} \\
&= F_1(A_1) \cdot \mathrm{E}\left[H_1\left(y - \sum_{i=2}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}}\right) \,\middle|\, A_2, A_3, \ldots, A_m\right] \\
&\quad + \bar{F}_1(A_1) \cdot \Pr\left\{\sum_{i=2}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \le y \,\middle|\, A_2, A_3, \ldots, A_m\right\}.
\end{aligned}$$

Recursively, we can use the same approach to analyze $\Pr\left\{\sum_{i=2}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \le y \,\middle|\, A_2, A_3, \ldots, A_m\right\}$. Then, after m iterations, we obtain that

$$\begin{aligned}
&\Pr\left\{\sum_{i=1}^{m} l_i \cdot \mathbf{1}_{\{\eta_i < A_i\}} \le y \,\middle|\, A_1, A_2, \ldots, A_m\right\} \\
&= \sum_{i=1}^{m}\left\{F_i(A_i) \cdot \mathrm{E}\left[H_i\left(y - \sum_{j=i+1}^{m} l_j \cdot \mathbf{1}_{\{\eta_j < A_j\}}\right) \,\middle|\, A_{i+1}, A_{i+2}, \ldots, A_m\right] \cdot \prod_{j=1}^{i-1} \bar{F}_j(A_j)\right\} \\
&\quad + \prod_{i=1}^{m} \bar{F}_i(A_i).
\end{aligned}$$

By Equation (EC.10),

$$F_L(y) = \sum_{i=1}^{m} \mathrm{E}\left[F_i(A_i) \cdot H_i\left(y - \sum_{j=i+1}^{m} l_j \cdot \mathbf{1}_{\{\eta_j < A_j\}}\right) \cdot \prod_{j=1}^{i-1} \bar{F}_j(A_j)\right] + \mathrm{E}\left[\prod_{i=1}^{m} \bar{F}_i(A_i)\right].$$

Differentiating $F_L(y)$ with respect to y yields

$$f_L(y) = \sum_{i=1}^{m} \mathrm{E}\left[F_i(A_i) \cdot h_i\left(y - \sum_{j=i+1}^{m} l_j \cdot \mathbf{1}_{\{\eta_j < A_j\}}\right) \cdot \prod_{j=1}^{i-1} \bar{F}_j(A_j)\right].$$

EC.3.2. The sensitivity of CVaR

According to Acerbi and Tasche (2002), the CVaR is equivalent to the expected shortfall (ES) when L is a real integrable random variable (i.e., $\mathrm{E}[|L|] < \infty$). In addition, L is continuous in the neighborhood of v_α (which is the α-VaR), then CVaR at level α (denoted by α-CVaR) of L is

$$u_\alpha = v_\alpha + \frac{1}{1-\alpha}\mathrm{E}[(L - v_\alpha); L \ge v_\alpha],$$

which is also known as the tail conditional expectation. We are interested to calculate $u'_\alpha(\theta)$. Note that $\Pr\{L(\theta)=v_\alpha(\theta)\}=0$. If we want to use the Monte Carlo method in this paper, we need to re-derive $p'(\theta)$ in Equation (6) in the paper since $g(\cdot)$ is a function of θ now. Moreover, we define $g(x,\theta)=(x-v_\alpha(\theta))\cdot\mathbf{1}_{\{x\geq v_\alpha(\theta)\}}$, then $\partial_\theta g(x,\theta)=-v'_\alpha(\theta)\cdot\mathbf{1}_{\{x\geq v_\alpha(\theta)\}}$ when $v_\alpha(\theta)\neq x$. We know that

$$p(\theta) = \sum_{s\in\mathscr{S}(m)} \mathrm{E}\left[g\left(\sum_{i\in s^1} l_i\right)\right]\cdot \mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i\geq 0\}}\right].$$

Then,

$$\begin{aligned}
p'(\theta) &= \sum_{s\in\mathscr{S}(m)} \mathrm{E}\left[g'\left(\sum_{i\in s^1} l_i\right)\right]\cdot \mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i\geq 0\}}\right] \\
&\quad + \sum_{s\in\mathscr{S}(m)} \mathrm{E}\left[g\left(\sum_{i\in s^1} l_i\right)\right]\cdot \frac{d}{d\theta}\,\mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i(\theta)<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i(\theta)\geq 0\}}\right]. \\
&= -v'_\alpha(\theta) \sum_{s\in\mathscr{S}(m)} \mathrm{E}\left[\mathbf{1}_{\{\sum_{i\in s^1} l_i\geq v_\alpha(\theta)\}}\right]\cdot \mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i\geq 0\}}\right] \\
&\quad + \sum_{s\in\mathscr{S}(m)} \mathrm{E}\left[g\left(\sum_{i\in s^1} l_i\right)\right]\cdot \frac{d}{d\theta}\,\mathrm{E}\left[\prod_{i\in s^1}\mathbf{1}_{\{X_i(\theta)<0\}}\prod_{i\in s^0}\mathbf{1}_{\{X_i(\theta)\geq 0\}}\right] \\
&= -v'_\alpha(\theta)\mathrm{E}\left[\mathbf{1}_{\{\sum_{i=1}^m l_i\mathbf{1}_{\{X_i<0\}}\geq v_\alpha(\theta)\}}\right] - \sum_{i=1}^m \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right)-g\left(L_{-i}\right)\right]\cdot X'_i(\theta)\,;\,X_i=0\right\} \\
&= -(1-\alpha)v'_\alpha(\theta) - \sum_{i=1}^m \mathrm{E}\left\{\left[g\left(L_{-i}+l_i\right)-g\left(L_{-i}\right)\right]\cdot X'_i(\theta)\,;\,X_i=0\right\}.
\end{aligned}$$

The sensitivity of α-CVaR is

$$\begin{aligned}
u'_\alpha(\theta) &= v'_\alpha(\theta) + \frac{1}{1-\alpha}\cdot\frac{d}{d\theta}\mathrm{E}[(L(\theta)-v_\alpha(\theta));L(\theta)\geq v_\alpha(\theta)] \\
&= \frac{-1}{1-\alpha}\sum_{i=1}^m \mathrm{E}\left\{\left[(L_{-i}+l_i)\cdot\mathbf{1}_{\{L_{-i}+l_i\geq v_\alpha(\theta)\}} - L_{-i}\cdot\mathbf{1}_{\{L_{-i}\geq v_\alpha(\theta)\}}\right]\cdot X'_i(\theta)\,;\,X_i=0\right\}.
\end{aligned}$$

If Assumptions 1–3 in Hong and Liu (2009) hold, then we can directly apply the result in Hong and Liu (2009) to obtain that

$$\begin{aligned}
u'_\alpha(\theta) &= \frac{1}{1-\alpha}\mathrm{E}\left[L'(\theta)\cdot\mathbf{1}_{\{L(\theta)\geq v_\alpha(\theta)\}}\right] \\
&= \frac{1}{1-\alpha}\mathrm{E}\left[\left(\sum_{i=1}^m l_i\cdot\mathbf{1}_{\{X_i(\theta)<0\}}\right)'\cdot\mathbf{1}_{\{L(\theta)\geq v_\alpha(\theta)\}}\right] = 0,
\end{aligned}$$

which is obviously wrong. This is because the Assumption 1 of Hong and Liu (2009) fails with respect the loss function L considered in our paper.

References

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

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

Durrett, R. 2005. *Probability: Theory and Examples*, Third Edition. Duxbury Press, Belmont.

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

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