Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers As Inputs

Yijie Peng, Michael Fu, Jiaqiao Hu, Pierre l'Ecuyer, Bruno Tuffin

Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers As Inputs

Yijie Peng
Department of Management Science and Information Systems, Guanghua School of Management, Peking University, Beijing, 100871, China, pengyijie@pku.edu.cn

Michael C. Fu
The Robert H. Smith School of Business, Institute for Systems Research, University of Maryland, College Park, Maryland 20742-1815, USA, mfu@umd.edu

Jiaqiao Hu
Applied Mathematics & Statistics, College of Engineering and Applied Sciences, Stony Brook University, Stony Brook, NY 11794-3600, USA, jqhu@ams.stonybrook.edu

Pierre L'Ecuyer
Department of Computer Science and Operations Research, University of Montreal, Montreal, Quebec, Canada, lecuyer@iro.umontreal.ca

Bruno Tuffin
Inria, Univ Rennes, CNRS, IRISA, Campus de Beaulieu, 35042 RENNES Cedex, France, bruno.tuffin@inria.fr

We propose a new unbiased stochastic gradient estimator for a family of stochastic models with uniform random numbers as inputs. By extending the generalized likelihood ratio (GLR) method, the proposed estimator applies to discontinuous sample performances with structural parameters without requiring that the tails of the density of the input random variables go down to zero smoothly, an assumption in Peng et al. (2018) and Peng et al. (2020a) that precludes a direct formulation in terms of uniform random numbers as inputs. By overcoming this limitation, our new estimator greatly expands the applicability of the GLR method, which we demonstrate for several general classes of uniform input random numbers, including independent inverse transform random variates and dependent input random variables governed by an Archimedean copula. We show how the new derivative estimator works in specific settings such as density estimation, distribution sensitivity for quantiles, and sensitivity analysis for Markov chain stopping time problems, which we illustrate with applications to statistical quality control, stochastic activity networks, and credit risk derivatives. Numerical experiments substantiate broad applicability and flexibility in dealing with discontinuities in sample performance.

Key words: simulation; stochastic derivative estimation; discontinuous sample performance; uniform random numbers, generalized likelihood ratio method.

History:

1. Introduction

Stochastic gradient estimation plays a central role in gradient-based optimization and sensitivity analysis (Asmussen and Glynn, 2007). The finite difference (FD) method is easily implementable, but it must balance a bias-variance tradeoff and requires extra simulations. Infinitesimal perturbation analysis (IPA) and the likelihood ratio (LR) method are two well-established unbiased derivative estimation techniques (Glynn, 1990, Ho and Cao, 1991, Glasserman, 1991, Rubinstein and Shapiro, 1993, Glynn and L'Ecuyer, 1995). IPA typically leads to lower variance than LR (L'Ecuyer, 1990; Cui et al., 2020), and the weak derivative method reduces the variance of LR at the cost of performing extra simulations (Pflug, 1988, Heidergott and Leahu, 2010). L'Ecuyer (1990) provides a general framework unifying IPA and LR, under which the resulting estimator depends on the choice of what a sample point in a probability space represents, and could in particular be a hybrid between IPA and LR. See Fu (2015) for a recent review.

Traditional applications of stochastic gradient estimation are in discrete event dynamic systems (DEDS), including queueing systems (Suri and Zazanis, 1988, Fu and Hu, 1993, L'Ecuyer and Glynn, 1994), inventory management (Fu, 1994, Bashyam and Fu, 1998), statistical quality control (Fu and Hu, 1999, Fu et al., 2009b), maintenance systems (Heidergott, 1999, Heidergott and Farenhorst-Yuan, 2010), and financial engineering and risk management, such as computing financial derivatives (Fu and Hu, 1995, Broadie and Glasserman, 1996, Liu and Hong, 2011, Wang et al., 2012, Hong et al., 2014, Chen and Liu, 2014, Lei et al., 2020), value-at-risk (VaR) and conditional VaR (CVaR) (Hong, 2009, Hong and Liu, 2009, Fu et al., 2009a, Jiang and Fu, 2015, Heidergott and Volk-Makarewicz, 2016). Recently, stochastic gradient estimation techniques have attracted attention in machine learning and artificial intelligence; see Mohamed et al. (2019) for a review paper written by a research team of Google's DeepMind. Peng et al. (2020b) show pathwise equivalence between IPA and backpropagation, and how the computational complexity for estimating the gradient is reduced by propagating the errors backwardly along the ANN. An LR-based method is then proposed to train ANNs, which can improve the robustness in classifying images under both adversarial attacks and natural noise corruptions.

IPA requires continuity in the sample performance, whereas LR does not directly apply for structural parameters (parameters directly appearing in the sample performance), which significantly limit their applicability. Smoothed perturbation analysis (SPA) deals with discontinuous sample performances by using a conditioning technique (Gong and Ho, 1987, Fu and Hu, 1997), but a good choice of conditioning is problem-dependent. Push-out LR addresses structural parameters by pushing the parameters out of the sample performance and into the density (Rubinstein and Shapiro, 1993), which can be achieved alternatively with the IPA-LR in L'Ecuyer (1990), but it requires an explicit transformation. Recently, Peng et al. (2018) proposed a generalized likelihood ratio (GLR) method that is capable of dealing with a large scope of discontinuous sample performances with structural parameters in a unified framework. The method extends the application domain of IPA and LR and does not require conditioning and transformation techniques tailored to specific problem structures.

The GLR method has the virtue of handling many applications in a uniform manner, and it has been used to deal with discontinuities in financial derivatives, statistical quality control, maintenance systems, and inventory systems (Peng et al., 2016, Peng et al., 2018). Distribution sensitivities, which mean the derivatives of the distribution function with respect to both the arguments and the parameters in the underlying stochastic model, lie at the center of many applications such as quantile sensitivity estimation, confidence interval construction for the quantile and quantile sensitivities, and statistical inference (Peng et al., 2017, Lei et al., 2018). Peng et al. (2020a) derive GLR estimators for any order of distribution sensitivities and apply them to maximum likelihood estimation for complex stochastic models without requiring analytical likelihoods. Glynn et al. (2020) apply the GLR method to estimate sensitivity of a distortion risk measure, which is a Lebesgue-Stieltjes integral of quantile sensitivities and includes Var and CVaR as special cases.

Although the existing GLR method has broad applicability, it requires that the density of the input distribution is known and that both tails of the density go down to zero smoothly and fast enough, which may not be satisfied in some applications, depending on what is interpreted as the input random variables. In this work, we relax this smoothness requirement and establish the unbiasedness of GLR gradient estimators for stochastic models whose inputs are uniform random numbers, which are the basic building blocks in generating other random variables. Unlike in Peng et al. (2018) where the surface integration part for the GLR estimator is zero, the surface integration part for the GLR estimator in the present work is not necessarily zero but can be estimated by simulation. If the surface integration part is zero, we are able to relax certain integrability conditions given in Peng et al. (2018) that are difficult to verify in practice.

We provide specific forms of the GLR estimators for two types of stochastic models and apply the GLR method to various problem settings, including distribution sensitivities, credit risk financial derivatives, and statistical quality control. The GLR estimator with independent input random variables generated from the inverse transform of uniform random numbers reduces to the classic LR estimator, which indicates that GLR is a generalization of LR from a different perspective than that of Peng et al. (2018). We also show how GLR can provide sensitivity estimators for models defined in terms of random vectors with given marginal distributions and whose dependence structures are specified by Archimedean copulas (Nelsen, 2006). The Gaussian copula has been widely used due to its simplicity (Li, 2000), and sensitivity analysis for portfolio credit risk derivatives with joint defaults governed by a Gaussian copula has been studied in Chen and Glasserman (2008) using LR and SPA. However, the Gaussian copula was widely criticized after the 2008 financial crisis, because it underestimates the probability of joint defaults. Archimedean copulas (Nelsen, 2006), not covered in Chen and Glasserman (2008), are relatively easy to simulate and can better capture the asymmetric tail dependence structure of the joint default data (Embrechts et al., 2003).

GLR can be used to estimate the distribution sensitivity functions, including the density and the quantile function together with confidence intervals, using a single batch of uniform random numbers. We provide numerical illustrations with various examples. Sample performances of control charts used in statistical quality control generally involve a stopping time defined by hitting a control limit (Fu and Hu, 1999, Fu et al., 2009b). The IPA and LR methods do not apply to this problem due to the discontinuity with respect to the control limits. We formulate a stopping time problem with uniform random numbers as inputs in a Markov chain, and estimate its sensitivity by the GLR method. We also apply GLR to estimate distribution sensitivities for stochastic activity networks (SAN) with uniform random numbers as inputs. In practice, the duration of an activity in a SAN and the output sample in control charts may be supported on a compact space, so a distribution supported on the whole space, which was assumed in Peng et al. (2018) and Peng et al. (2020a), is unsuitable for modeling input distributions in these stochastic models.

The rest of the paper is organized as follows. Section 2 sets the framework. The GLR estimator is presented in Section 3 with the specific forms of the estimators for two types of models. Applications are given in Section 4. Numerical experiments can be found in Section 5. The last section concludes. The technical proofs and additional numerical results can be found in the online appendix.

2. Problem Formulation

Consider a stochastic model of the following form:

$$\varphi(g(U;\theta)), \tag{1}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a measurable function (not necessarily continuous), $g(\cdot;\theta) = (g_1(\cdot;\theta), \ldots, g_n(\cdot;\theta))$ is a vector of functions $g_i : (0,1)^n \to \mathbb{R}$ with certain smoothness properties to be made more precise shortly, and $U = (U_1, \ldots, U_n)$ is a vector of i.i.d. $\mathcal{U}(0,1)$ random variables (i.e., uniform over $(0,1)$). For simplicity, we take θ as a scalar. When θ is a vector, each component of the gradient can be estimated separately using the method developed in this work. We consider the problem of estimating the following derivative:

$$\frac{\partial \mathbb{E}[\varphi(g(U;\theta))]}{\partial \theta}. \tag{2}$$

A straightforward pathwise derivative estimator, i.e., IPA, obtained by directly interchanging derivative and expectation, may not apply because discontinuities in the sample performance of the stochastic model could be introduced by $\varphi(\cdot)$. In Peng et al. (2018), the stochastic model considered for the derivative estimation problem is $\varphi(g(X;\theta))$, where the density of $X = (X_1, \ldots, X_n)$ is assumed to be known and both tails go down to zero smoothly and fast enough. This assumption is not satisfied by discontinuous densities, such as the uniform and exponential distributions, and we want to address this limitation. Before deriving a GLR derivative estimator, we first introduce two examples to illustrate potential applications of the stochastic model (1).

Example 1 Independent Inputs Generated via the Inverse Transform Method. Suppose $X = (X_1, \ldots, X_n)$ is a vector of independent random variables, where each X_i has cumulative distribution function (cdf) $F_i(\cdot;\theta), i = 1, \ldots, n$, and is generated by (standard) inversion:

$$X_i = F_i^{-1}(U_i;\theta), \quad i = 1, \ldots, n,$$

with i.i.d. $U_i \sim \mathcal{U}(0,1)$. A stochastic model with i.i.d. $\mathcal{U}(0,1)$ random numbers as input can be written as $\varphi(g(U;\theta)) = \varphi(F_1^{-1}(U_1;\theta), \ldots, F_n^{-1}(U_n;\theta))$, where $g(u;\theta) = (F_1^{-1}(u_1;\theta), \ldots, F_n^{-1}(u_n;\theta))$.

Example 2 Archimedean Copulas. Copulas are a general way of representing the dependence in a multivariate distribution. A *copula* is any multivariate cdf whose one-dimensional marginals are all $\mathcal{U}(0,1)$. It can be defined by a function $C(\cdot;\theta) : [0,1]^n \to [0,1]$ that satisfies certain conditions required for C to be a consistent cdf; see Nelsen (2006). For any given copula and arbitrary marginal distributions with continuous cdf's $F_1(\cdot), F_2(\cdot), ..., F_n(\cdot)$ with densities $f_i(\cdot)$, $i = 1, \ldots, n$, one can define a multivariate distribution having exactly these marginals with joint cdf F_X given by $F_X(x) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n);\theta)$ for all $x := (x_1, \ldots, x_n)$. Sklar (1959) shows that any multivariate distribution can be represented in this way. If $C(\cdot;\theta)$ is absolutely continuous, the density of the joint distribution is

$$f_X(x;\theta) = c\,(F_1(x_1), \ldots, F_n(x_n);\theta) \prod_{i=1}^{n} f_i(x_i), \qquad \text{where}$$

$$c(v;\theta) = \frac{\partial^n C(v;\theta)}{\partial v_1 \cdots \partial v_n},$$

$v = (v_1, \ldots, v_n)$, and the derivative is interpreted as a Radon-Nikodym derivative when $C(\cdot;\theta)$ is not nth-order differentiable.

To generate $X = (X_1, \ldots, X_n)$ from the joint cdf $F_X(\cdot)$, generate $V = (V_1, \ldots, V_n)$ from the copula and return $X_i = F_i^{-1}(V_i)$ for each i. Generating V from the copula is not always obvious, but there are classes of copulas for which this can be easily done, one of them being the *Archimedean copulas*. This important family of copulas can model strong forms of tail dependence using a single parameter, which makes them convenient to use. An Archimedean copula C_a is defined by

$$C_a(v;\theta) = \psi_\theta\left(\psi_\theta^{[-1]}(v_1) + \ldots + \psi_\theta^{[-1]}(v_n)\right),$$

where the generator function $\psi_\theta : [0,\infty) \to [0,1]$ is a strictly decreasing convex function such that $\lim_{x\to\infty} \psi_\theta(x) = 0$, $\theta \in [0,\infty)$ is a parameter governing the strength of dependence, and $\psi_\theta^{[-1]}$ is a pseudo-inverse defined by $\psi_\theta^{[-1]}(x) = \mathbf{1}\{0 \le x \le \psi_\theta(0)\}\psi_\theta(x)$, with the convention that $\psi_\theta^{-1}(0) = \inf\{x : \psi_\theta(x) = 0\}$. Archimedean copulas are absolutely continuous, and their densities have the form:

$$c_a(v;\theta) = \mathbf{1}\left\{0 \le \sum_{i=1}^n \psi_\theta^{-1}(v_i) \le \psi_\theta^{-1}(0)\right\} \frac{\partial^n \psi_\theta(x)}{\partial x^n}\bigg|_{x=\sum_{i=1}^n \psi_\theta^{-1}(v_i)} \prod_{i=1}^n \frac{\partial \psi_\theta^{-1}(v_i)}{\partial v_i},$$

assuming that the generator function $\psi_\theta(\cdot)$ is smooth (Nelsen, 2006). In general we do not have an analytical expression for $c_a(\cdot;\theta)$, and it can be discontinuous in θ. Therefore, the standard LR method typically does not apply to sensitivity analysis with Archimedean copulas.

Marshall and Olkin (1988) propose the following simple algorithm to generate V from an Archimedean copula with generator function $\psi_\theta(\cdot)$:

(i) Generate a random variable Y_θ from the distribution with Laplace transform $\psi_\theta(\cdot)$ (with at least one uniform random number as input).

(ii) For $i = 1,\dots,n$, let $V_i = \psi_\theta\left(-(\log U_i)/Y_\theta\right)$ with i.i.d. $U_i \sim \mathcal{U}(0,1)$.

For a given Y_θ, this gives a stochastic model with uniform random numbers U_i as inputs:

$$\varphi(g(U;\theta)) = \phi\left(F_1^{-1}\left(\psi_\theta\left(-(\log U_1)/Y_\theta\right)\right),\dots,F_n^{-1}\left(\psi_\theta\left(-(\log U_n)/Y_\theta\right)\right)\right),$$

where $\varphi(v_1,\dots,v_n) = \phi(F_1^{-1}(v_1),\dots,F_1^{-1}(v_n))$ and $g(u;\theta) = \left(\psi_\theta\left(-(\log u_1)/Y_\theta\right),\dots,\psi_\theta\left(-(\log u_n)/Y_\theta\right)\right)$.

3. A Generalized Likelihood Ratio Method

In this section, we derive the GLR estimator for the derivative (2) of the expectation of stochastic model (1). The general theory for GLR is first derived, and then it is applied to the two examples in the previous section.

3.1. General Theory

Denote the Jacobian of g by

$$J_g(u;\theta) := \begin{pmatrix} \frac{\partial g_1(u;\theta)}{\partial u_1} & \frac{\partial g_1(u;\theta)}{\partial u_2} & \cdots & \frac{\partial g_1(u;\theta)}{\partial u_n} \\ \frac{\partial g_2(u;\theta)}{\partial u_1} & \frac{\partial g_2(u;\theta)}{\partial u_2} & \cdots & \frac{\partial g_2(u;\theta)}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(u;\theta)}{\partial u_1} & \frac{\partial g_n(u;\theta)}{\partial u_2} & \cdots & \frac{\partial g_n(u;\theta)}{\partial u_n} \end{pmatrix}, \qquad \text{and}$$

$$\partial_\theta g(u;\theta) := \left(\frac{\partial g_1(u;\theta)}{\partial\theta},\dots,\frac{\partial g_n(u;\theta)}{\partial\theta}\right)^T,$$

with the superscript T indicating vector transposition. In addition, we define two weight functions in the GLR estimator:

$$r_i(u;\theta) := \left(J_g^{-1}(u;\theta)\,\partial_\theta g(u;\theta)\right)^T e_i, \quad i=1,\dots,n, \tag{3}$$

$$d(u;\theta) := \sum_{i=1}^n e_i^T J_g^{-1}(u;\theta)\,(\partial_{u_i} J_g(u;\theta)) J_g^{-1}(u;\theta)\partial_\theta g(u;\theta) - \mathrm{trace}(J_g^{-1}(u;\theta)\,\partial_\theta J_g(u;\theta)), \tag{4}$$

where e_i is the ith unit column vector and $\partial_z J_g$ is the matrix obtained by differentiating J_g with respect to z element-wise. Let x^- and x^+ be limits taken from the left-hand side and right-hand side of x, respectively, and for a function $h(\cdot)$, denote $h(x^-) := \lim_{x\to x^-} h(x)$ and $h(x^+) := \lim_{x\to x^+} h(x)$. The following conditions are introduced to guarantee the unbiasedness of the proposed GLR derivative estimator.

(A.1) There exist a smooth function $\varphi_\epsilon(\cdot) \in \mathbb{C}^\infty$ and $p > 1$ such that

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}\int_{(0,1)^n} |\varphi_\epsilon(g(u;\theta)) - \varphi(g(u;\theta))|^p\, du = 0,$$

and if $n > 1$, for a fixed $\varepsilon > 0$ and any $u_i \in (0,1)\setminus[\varepsilon, 1-\varepsilon]$, $i = 1,\dots,n$,

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}\int_{(0,1)^{n-1}} |\varphi_\epsilon(g(u;\theta)) - \varphi(g(u;\theta))|^p\, du_{-i} = 0,$$

where $u_{-i} := (u_1,\dots,u_{i-1},u_{i+1},\dots,u_n)$, and if $n = 1$, for any $u \in (0,1)\setminus[\varepsilon, 1-\varepsilon]$,

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta} |\varphi_\epsilon(g(u;\theta)) - \varphi(g(u;\theta))| = 0.$$

(A.2) The Jacobian $J_g(u;\theta)$ is invertible almost everywhere (a.e.), and the performance function $g(u;\theta)$ is twice continuously differentiable with respect to $(u,\theta) \in (0,1)^n \times \Theta$, where Θ is a bounded neighborhood of the parameter θ of interest.

(A.3) The following integrability conditions hold:

$$\int_{(0,1)^{n-1}} \sup_{\theta\in\Theta, u_i\in(0,1)} \left|\varphi(g(u;\theta))\, r_i(u;\theta)\right| du_{-i} < \infty, \quad i = 1,\dots,n, \qquad \text{and}$$

$$\int_{(0,1)^n} \sup_{\theta\in\Theta} \left|\varphi(g(u;\theta))\, d(u;\theta)\right| du < \infty.$$

(A.4) The function $g(\cdot;\theta)$ is invertible, and

$$\lim_{u_i\to 1^-}\sup_{\theta\in\Theta, u_{-i}\in(0,1)^{n-1}} |r_i(u;\theta)| = \lim_{u_i\to 0^+}\sup_{\theta\in\Theta, u_{-i}\in(0,1)^{n-1}} |r_i(u;\theta)| = 0, \quad i = 1,\dots,n.$$

Remark 1 Condition (A.1) can be checked in certain settings when $\varphi_\epsilon(\cdot)$ can be explicitly constructed; see Proposition 1. The invertibility of the Jacobian matrix in condition (A.2) justifies the local invertibility of function $g(\cdot;\theta)$, whereas global invertibility of $g(\cdot;\theta)$ in condition (A.4) is stronger, although much weaker than requiring an explicit inverse function for $g(\cdot;\theta)$ in deriving the push-out LR estimator (Rubinstein and Shapiro, 1993). In general, it is difficult to find an explicit inverse function for a nonlinear function $g(\cdot;\theta)$, but the existence of the inversion could be guaranteed by the inverse function theorem.

Unbiasedness of the new GLR estimator developed in this work is established under two sets of conditions in the following theorem.

Theorem 1 *Under conditions (A.1) – (A.3) or (A.2) – (A.4),*

$$\frac{\partial \mathbb{E}[\varphi(g(U;\theta))]}{\partial \theta} = \mathbb{E}[G(U;\theta)], \qquad \text{where} \tag{5}$$

$$G(U;\theta) := \sum_{i=1}^{n} \left[\varphi(g(\overline{U}_i;\theta)) r_i(\overline{U}_i;\theta) - \varphi(g(\underline{U}_i;\theta)) r_i(\underline{U}_i;\theta)\right] + \varphi(g(U;\theta)) d(U;\theta),$$

with $\overline{U}_i := (U_1, \ldots, \underbrace{1^-}_{i\text{th element}}, \ldots, U_n)$, $\underline{U}_i := (U_1, \ldots, \underbrace{0^+}_{i\text{th element}}, \ldots, U_n)$, *and* $r_i(\cdot)$ *and* $d(\cdot)$ *defined by (3) and (4), respectively.*

Remark 2 The proof of the theorem can be found in the online Appendix A. Even if $g(\overline{U}_i;\theta) = \infty$ or $g(\underline{U}_i;\theta) = \infty$, the GLR estimator could still be well defined; see e.g. Section 3.1. The difference between this estimator and the GLR estimator in Peng et al. (2018) is that the surface integration part in Peng et al. (2018) is shown to be zero under certain conditions including that the tails of the input densities are required to go to zero smoothly and fast enough, whereas here the surface integration part is included in the estimator and can be estimated by simulation. In the case where the surface integration part becomes zero, we can prove the result without assuming (A.1), and we also avoid the integrability condition in Peng et al. (2018) on certain intermediate quantities (the smoothed function), which is difficult to verify in practice. The proof is obtained by first truncating the support of the input random variables to a compact set and then appropriately expanding it to the whole space.

We now examine the special case where $g(u;\theta) = (g_1(u_1;\theta), \ldots, g_n(u_n;\theta))$, which covers Examples 1 and 2. For n independent uniform random numbers, we can take $g_i(u_i;\theta) = F_i^{-1}(u_i;\theta)$ for

$i = 1, \ldots, n$, while for dependent variables governed by an Archimedean copula, conditional on $Y_\theta = y$, we have $g_i(u_i;\theta) = \psi_\theta(-\log u_i/y)$ for $i = 1, \ldots, n$. In this special case, the Jacobian becomes

$$J_g(u;\theta) = \begin{pmatrix} \frac{\partial g_1(u_1;\theta)}{\partial u_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial g_2(u_2;\theta)}{\partial u_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial g_n(u_n;\theta)}{\partial u_n} \end{pmatrix}.$$

Then we have

$$r_i(u;\theta) = \frac{\partial g_i(u_i;\theta)}{\partial \theta} \Big/ \frac{\partial g_i(u_i;\theta)}{\partial u_i}, \quad i = 1, \ldots, n, \qquad \text{and}$$

$$d(u;\theta) = \sum_{i=1}^{n} \left[\frac{\partial g_i(u_i;\theta)}{\partial \theta} \frac{\partial^2 g_i(u_i;\theta)}{\partial u_i^2} \Big/ \left(\frac{\partial g_i(u_i;\theta)}{\partial u_i} \right)^2 - \frac{\partial^2 g_i(u_i;\theta)}{\partial \theta \partial u_i} \Big/ \frac{\partial g_i(u_i;\theta)}{\partial u_i} \right].$$

Moreover, condition (A.1) in Theorem 1 can be replaced by a set of a simpler assumptions when the performance function $\varphi(x)$ is a product of n indicators: $\varphi(x) = \prod_{i=1}^{n} \mathbf{1}\{x_i \leq 0\}$, in which case a smoothed function $\varphi_\epsilon(\cdot)$ can be constructed explicitly. The performance function in the distribution sensitivities discussed in Section 4.1 is an indicator function. The distribution sensitivities for the completion time in an SAN in Section 5.2 and the sensitivities of a control chart in Section 5.3 have performance functions which are products of n indicators. The proof of the following proposition can be found in the online Appendix A.

Proposition 1 *Consider the stochastic model*

$$\varphi(g(U;\theta)) = \prod_{i=1}^{n} \mathbf{1}\{g_i(U_i;\theta) \leq 0\}.$$

Condition (A.1) holds if for $i = 1, \ldots, n$ *and a fixed* $\varepsilon > 0$ *in (6),*

$$\inf_{\theta \in \Theta, u_i \in [\varepsilon, 1-\varepsilon]} \left| \frac{\partial g_i(u_i;\theta)}{\partial u_i} \right| > 0 \quad \textit{and} \quad \inf_{\theta \in \Theta, u_i \in (0,1) \setminus [\varepsilon, 1-\varepsilon]} \left| g_i(u_i;\theta) \right| > 0. \tag{6}$$

If the functions g_i can be decomposed as products of the form $g_i(u_i;\theta) = \xi_i(\theta)\eta_i(u_i)$ for $i = 1, \ldots, n$, then conditions (A.3), (A.4), and (6) can be simplified. For example, an exponential random variable with mean θ can be generated by $-\log(U_i)/\theta$ where $U_i \sim \mathcal{U}(0,1)$. When this decomposition holds, we can write

$$r_i(u;\theta) = \frac{d \log \xi_i(\theta)}{d\theta} \Big/ \frac{d \log \eta_i(u_i)}{du_i}, \quad i = 1, \ldots, n, \qquad \text{and}$$

$$d(u;\theta) = \sum_{i=1}^{n} \frac{d \log \xi_i(\theta)}{d\theta} \left[\frac{\eta_i(u_i)\eta_i''(u_i)}{\left(\eta_i'(u_i)\right)^2} - 1 \right].$$

(A.3') Boundedness and integrability conditions on functions of input random numbers:

$$\inf_{u_i\in(0,1)} \left|\frac{d\log\eta_i(u_i)}{du_i}\right| > 0 \qquad \text{and} \qquad \mathbb{E}\left[\frac{\left|\eta_i(U_i)\eta_i''(U_i)\right|}{\left(\eta_i'(U_i)\right)^2}\right] < \infty, \quad i=1,\ldots,n.$$

(A.4') Boundary condition on functions of input random numbers:

$$\lim_{u_i\to 1^-}\left|\frac{d\log\eta_i(u_i)}{du_i}\right| = \lim_{u_i\to 0^+}\left|\frac{d\log\eta_i(u_i)}{du_i}\right| = \infty, \quad i=1,\ldots,n.$$

Corollary 1 *Suppose that* $g_i(u_i;\theta)=\xi_i(\theta)\eta_i(u_i)$, $i=1,\ldots,n$, *and* $\varphi(\cdot)$ *is bounded and*

$$\max_{i=1,\ldots,n}\sup_{\theta\in\Theta}\left|\frac{\partial\log\xi_i(\theta)}{\partial\theta}\right| < \infty.$$

Then conditions (A.3) and (A.4) can be replaced by (A.3') and (A.4'), respectively, and condition (6) in Proposition 1 also simplifies to

$$\inf_{\theta\in\Theta}\left|\xi_i(\theta)\right| > 0, \qquad \inf_{u_i\in[\varepsilon,1-\varepsilon]}\left|\eta_i'(u_i)\right| > 0, \qquad \inf_{u_i\in(0,1)\setminus[\varepsilon,1-\varepsilon]}\left|\eta_i(u_i)\right| > 0, \quad i=1,\ldots,n.$$

Unbiasedness of the GLR estimators in many examples of this paper can be justified by verifying these simplified conditions.

3.2. The Independent Case

Let us return to the independent case of Example 1 and suppose that each X_i is continuous with density $f_i(\cdot;\theta)$. Our goal is to estimate

$$\frac{\partial\mathbb{E}[\varphi(F_1^{-1}(U_1;\theta),\ldots,F_n^{-1}(U_n;\theta))]}{\partial\theta}, \text{ for which the Jacobian is}$$

$$J_g(u;\theta) = \begin{pmatrix} \frac{1}{f_1(X_1(u_1;\theta);\theta)} & 0 & \cdots & 0 \\ 0 & \frac{1}{f_2(X_2(u_2;\theta);\theta)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{f_n(X_n(u_n;\theta);\theta)} \end{pmatrix}, \qquad \text{and}$$

$$\partial_\theta g(u;\theta) = \left(\frac{\partial X_1(u_1;\theta)}{\partial\theta},\ldots,\frac{\partial X_n(u_n;\theta)}{\partial\theta}\right)^T,$$

$$\text{where } X_i(u_i;\theta) := F_i^{-1}(u_i;\theta) \;\text{ and }\; \frac{\partial X_i(u_i;\theta)}{\partial\theta} := -\frac{\partial F_i(x_i;\theta)}{\partial\theta}\Big/ f_i(x_i;\theta)\Big|_{x_i=X_i(u_i;\theta)}.$$

Then the weight functions in the GLR estimator are:

$$r_i(u;\theta) = -\frac{\partial F_i(x_i;\theta)}{\partial\theta}\bigg|_{x_i=X_i(u_i;\theta)} \qquad \text{and} \qquad d(u;\theta) = \sum_{i=1}^n \frac{\partial\log f_i(x_i;\theta)}{\partial\theta}\bigg|_{x_i=X_i(u_i;\theta)}, \text{ so}$$

$$\lim_{u_i\to 1^-} r_i(u;\theta) = \lim_{u_i\to 0^+} r_i(u;\theta) = 0.$$

Therefore,
$$\frac{\partial \mathbb{E}[\varphi(F_1^{-1}(U_1;\theta),\dots,F_n^{-1}(U_n;\theta))]}{\partial\theta} = \mathbb{E}\Big[\varphi(F_1^{-1}(U_1;\theta),\dots,F_n^{-1}(U_n;\theta))\ d(U;\theta)\ \Big] = \mathbb{E}\Big[\varphi(X)\ \sum_{i=1}^{n}\frac{\partial \log f_i(X_i;\theta)}{\partial\theta}\ \Big].$$

The expression inside the last expectation coincides with the classic LR derivative estimator in the case where the LR method is applicable, i.e., when there are no structural parameters in the sample performance (Glynn, 1990). From this perspective, the GLR method generalizes the LR method by allowing the appearance of structural parameters.

3.3. Archimedean Copulas

We consider estimating
$$\frac{\partial}{\partial\theta}\mathbb{E}\Big[\varphi\Big(\psi_\theta\big(-\frac{\log U_1}{Y_\theta}\big),\dots,\psi_\theta\big(-\frac{\log U_n}{Y_\theta}\big)\Big)\Big],$$
where the expectation is with respect to both Y_θ and the independent U_i, $i=1,\dots,n$. By conditioning, we can use a mixture of LR and GLR:
$$\begin{aligned}&\frac{\partial}{\partial\theta}\mathbb{E}\Big[\varphi\Big(\psi_\theta\big(-\frac{\log U_1}{Y_\theta}\big),\dots,\psi_\theta\big(-\frac{\log U_n}{Y_\theta}\big)\Big)\Big]\\ &= \mathbb{E}\left[\varphi\Big(\psi_\theta\big(-\frac{\log U_1}{Y_\theta}\big),\dots,\psi_\theta\big(-\frac{\log U_n}{Y_\theta}\big)\Big)\frac{\partial \log f_Y(y;\theta)}{\partial\theta}\Big|_{y=Y_\theta}\right]\\ &\quad+\mathbb{E}\left[\frac{\partial\mathbb{E}\Big[\varphi\Big(\psi_\theta\big(-\frac{\log U_1}{y}\big),\dots,\psi_\theta\big(-\frac{\log U_n}{y}\big)\Big)\Big]}{\partial\theta}\Bigg|_{y=Y_\theta}\right],\end{aligned} \tag{7}$$
where $f_Y(\cdot;\theta)$ is the density function of Y_θ. This equality follows from Theorem 1 of L'Ecuyer (1990) with ω in that theorem replaced by y and $h(\theta,\omega)$ replaced by $h(y,\theta) := \mathbb{E}[\varphi(\psi_\theta(-\ln U_1/y),\dots,\psi_\theta(-\ln U_n/y))]$, under the assumption that for all y, this last expectation is continuous in θ and differentiable except perhaps on a countable set. Specifically, (7) can be rewritten as
$$\begin{aligned}\frac{\partial}{\partial\theta}\int_{\mathbb{R}} h(y,\theta)F_Y(dy;\theta) &= \frac{\partial}{\partial\theta}\int_{\mathbb{R}} h(y,\theta)L(y,\theta,\theta_0)F_Y(dy;\theta_0)\\ &= \int_{\mathbb{R}}\left(h(y,\theta)\frac{\partial L(y,\theta,\theta_0)}{\partial\theta}+\frac{\partial h(y,\theta)}{\partial\theta}L(y,\theta,\theta_0)\right)F_Y(dy;\theta_0),\end{aligned}$$
where $L(y,\theta,\theta_0) := f_Y(y;\theta)/f_Y(y;\theta_0)$. The first term on the right-hand side of (7) can be dealt with by the LR method straightforwardly if $f_Y(\cdot;\theta)$ admits an analytical form. Glasserman and Liu (2010) show how to apply the LR method with only the Laplace transform $\psi_\theta(\cdot)$.

We now show how to use GLR to handle the second term on the right-hand side of (7) with Y_θ fixed and generated from other uniform random numbers. The Archimedean copula model falls into the special case where $g(u;\theta)=(g_1(u_1;\theta),\dots,g_n(u_n;\theta))$, discussed after Theorem 1. The Jacobian in this case is

$$J_g(u;\theta,y)=\begin{pmatrix} -\frac{1}{u_1 y}\psi_\theta'\big(-\frac{\log u_1}{y}\big) & 0 & \cdots & 0 \\ 0 & -\frac{1}{u_2 y}\psi_\theta'\big(-\frac{\log u_2}{Y_\theta}\big) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{1}{u_n y}\psi_\theta'\big(-\frac{\log u_n}{y}\big) \end{pmatrix}, \qquad \text{and}$$

$$\partial_\theta g(u;\theta,y)=\left(\frac{\partial\psi_\theta(x_1)}{\partial\theta}\bigg|_{x_1=-\frac{\log u_1}{y}},\dots,\frac{\partial\psi_\theta(x_n)}{\partial\theta}\bigg|_{x_n=-\frac{\log u_n}{y}}\right)^T.$$

The weight functions in the GLR estimator are

$$r_i(u;\theta,y)=-\frac{u_i y}{\psi_\theta'(x_i)}\frac{\partial\psi_\theta(x_i)}{\partial\theta}\bigg|_{x_i=-\frac{\log u_i}{y}},$$

$$d(u;\theta,y)=\sum_{i=1}^n\left(-\frac{1}{\psi_\theta'(x_i)}\frac{\partial\psi_\theta'(x_i)}{\partial\theta}+\frac{\psi_\theta''(x_i)}{(\psi_\theta'(x_i))^2}\frac{\partial\psi_\theta(x_i)}{\partial\theta}+\frac{\partial\psi_\theta(x_i)}{\partial\theta}\frac{y}{\psi_\theta'(x_i)}\right)\bigg|_{x_i=-\frac{\log u_i}{y}}.$$

Example 3 The Clayton Copula. The generator function for the Clayton copula is

$$\psi_\theta(x)=(1+x)^{-\frac{1}{\theta}},\qquad \theta\in(0,\infty).$$

Then, $$\frac{\partial\psi_\theta(x)}{\partial\theta}=\frac{1}{\theta^2}\log(1+x)(1+x)^{-\frac{1}{\theta}},\quad \frac{\partial\psi_\theta'(x)}{\partial\theta}=\frac{1}{\theta^2}(1+x)^{-\frac{1}{\theta}-1}\left[1-\frac{1}{\theta}\log(1+x)\right],\quad \text{and}$$
$$\psi_\theta^{'}(x)=-\frac{1}{\theta}(1+x)^{-\frac{1}{\theta}-1},\quad \psi_\theta^{''}(x)=\frac{1}{\theta}\left(\frac{1}{\theta}+1\right)(1+x)^{-\frac{1}{\theta}-2}.$$

By the inverse Laplace transformation, we find that $Y_\theta\sim\Gamma(1/\theta,1)$, the gamma distribution with density $f_Y(y;\theta)=\frac{y^{\frac{1}{\theta}-1}e^{-y}}{\Gamma(1/\theta)}$, where $\Gamma(s):=\int_0^\infty t^{s-1}e^{-t}dt$, and the LR term is

$$\frac{\partial\log f_Y(y;\theta)}{\partial\theta}=-\frac{d\log\Gamma(1/\theta)}{d\theta}-\frac{1}{\theta^2}\log y.$$

The weight functions in the GLR estimator are

$$r_i(u;\theta,y)=-\frac{1}{\theta}u_i y\left(1-\frac{\log u_i}{y}\right)\log\left(1-\frac{\log u_i}{y}\right),$$

$$d(u;\theta,y)=\sum_{i=1}^n\frac{1}{\theta}\left[1+(1-(x_i+1)y)\log(1+x_i)\right]\bigg|_{x_i=-\frac{\log u_i}{y}}.$$

In addition, we have $\lim_{u_i\to 0^+}r_i(u;\theta,y)=\lim_{u_i\to 1^-}r_i(u;\theta,y)=0$.

GLR for Ali-Mikhail-Haq copulas can be found in the online Appendix A. For both the Clayton and Ali-Mikhail-Haq copulas, conditions (A.2) and (A.4) in Theorem 1 are satisfied. If $\varphi(\cdot)$ is bounded, condition (A.3) in Theorem 1 can also be verified straightforwardly for any $y=Y_\theta$.

4. Applications

We apply the GLR method to distribution sensitivity estimation, and estimate sensitivities for stopping time problems and credit risk derivatives, with specific forms for the function $\varphi(\cdot)$.

4.1. Distribution Sensitivities

For $g(\cdot;\theta):(0,1)^n\to\mathbb{R}$, we estimate the following two first-order distribution sensitivities:

$$\frac{\partial F(z;\theta)}{\partial\theta}=\frac{\partial\mathbb{E}[\mathbf{1}\{g(U;\theta)-z\le 0\}]}{\partial\theta}=\mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial\theta}\right],$$
$$f(z;\theta)=\frac{\partial\mathbb{E}[\mathbf{1}\{g(U;\theta)-z\le 0\}]}{\partial z}=\mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial z}\right],$$

where $f(\cdot;\theta)$ is the density function of $Z(\theta)=g(U;\theta)$ and $U_{-i}:=(U_1,\dots,U_{i-1},U_{i+1},\dots,U_n)$, $i=1,\dots,n$. By applying GLR, we obtain

$$\mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial\theta}\right]=\mathbb{E}[G_{1,i}(U;z,\theta)],\ \text{where}$$
$$G_{1,i}(U;z,\theta):=\mathbf{1}\{g(\overline{U}_i;\theta)-z\le 0\}r_i(\overline{U}_i;\theta)-\mathbf{1}\{g(\underline{U}_i;\theta)-z\le 0\}r_i(\underline{U}_i;\theta)+\mathbf{1}\{g(U;\theta)-z\le 0\}d(U;\theta),$$

$$r_i(u;\theta)=\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1}\frac{\partial g(u;\theta)}{\partial\theta},\qquad \text{and}$$

$$d(u;\theta)=\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1}\left[\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1}\frac{\partial g(u;\theta)}{\partial\theta}\frac{\partial^2 g(u;\theta)}{\partial u_i^2}-\frac{\partial^2 g(u;\theta)}{\partial u_i\partial\theta}\right].$$

We also obtain $\mathbb{E}\left[\frac{\partial\mathbb{E}[\mathbf{1}\{g(U_i,U_{-i};\theta)-z\le 0\}|U_{-i}]}{\partial z}\right]=\mathbb{E}[G_{2,i}(U;z,\theta)]$, where

$$\begin{aligned}G_{2,i}(U;z,\theta):=&\mathbf{1}\{g(\overline{U}_i;\theta)-z\le 0\}\tilde{r}_i(\overline{U}_i;\theta)-\mathbf{1}\{g(\underline{U}_i;\theta)-z\le 0\}\tilde{r}_i(\underline{U}_i;\theta)\\&+\mathbf{1}\{g(U;\theta)-z\le 0\}\tilde{d}(U;\theta),\qquad\text{with}\end{aligned}$$

$$\tilde{r}_i(u;\theta)=-\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-1}\qquad\text{and}\qquad\tilde{d}(u;\theta)=-\left(\frac{\partial g(u;\theta)}{\partial u_i}\right)^{-2}\frac{\partial^2 g(u;\theta)}{\partial u_i^2}.$$

To establish the unbiasedness of $G_{1,i}(U;z,\theta)$ and $G_{2,i}(U;z,\theta)$ conditional on U_{-i}, condition (A.1) in Theorem 1 can be justified by checking the conditions in Proposition 1. Given this conditional unbiasedness, the unconditional unbiasedness of $G_{1,i}(U;z,\theta)$ and $G_{2,i}(U;z,\theta)$ will follow from

$$\mathbb{E}\left[\sup_{\theta\in\Theta}|\mathbb{E}[G_{1,i}(U;z,\theta)|U_{-i}]\right]<\infty\qquad\text{and}\qquad\mathbb{E}\left[\sup_{z\in\mathcal{Z}}|\mathbb{E}[G_{2,i}(U;z,\theta)|U_{-i}]|\right]<\infty,$$

where $\mathcal{Z}$ is a neighborhood of z. Since the indicator function is bounded, the conditions of Proposition 1 follow from the integrability condition on the weight functions: for $i=1,\dots,n$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta}|r_i(\overline{U}_i;\theta)|\right]<\infty,\quad\mathbb{E}\left[\sup_{\theta\in\Theta}|r_i(\underline{U}_i;\theta)|\right]<\infty,\quad\mathbb{E}\left[\sup_{\theta\in\Theta}|d(U;\theta)|\right]<\infty,\qquad\text{and}$$

$$\mathbb{E}\left[|\tilde{r}_i(\overline{U}_i;\theta)|\right] < \infty, \quad \mathbb{E}\left[|\tilde{r}_i(\underline{U}_i;\theta)|\right] < \infty, \quad \mathbb{E}\left[|\tilde{d}(U;\theta)|\right] < \infty.$$

The GLR estimator for estimating the distribution sensitivities is not unique. We can consider the above GLR estimator for each i and construct the following linear combination of these n GLR estimators with real-values weights w_i, as in Hammersley and Handscomb (1964, p.19):

$$\sum_{i=1}^{n} w_i G_{r,i}(U;z,\theta) \qquad \text{subject to } \sum_{i=1}^{n} w_i = 1, \quad r=1,2.$$

An optimal GLR estimator, which minimizes the variance, can be obtained by solving

$$\arg\min_{(w_1,\dots,w_n)} \mathrm{Var}\left(\sum_{i=1}^{n} w_i G_{r,i}(U;z,\theta)\right) \qquad \text{subject to } \sum_{i=1}^{n} w_i = 1.$$

This leads to the optimal weights

$$w_i^* = \frac{e_i^T \Sigma^{-1} e}{e^T \Sigma^{-1} e}, \quad i=1,\dots,n, \tag{GLR-Opt}$$

where $e=(1,\dots,1)^T$, e_i is a d-dimensional unit vector in ith direction, and $\Sigma=(\Sigma_{i'i})_{n\times n}$ is the covariance matrix of $(G_{r,1}(U;z,\theta),\dots,G_{r,n}(U;z,\theta))$. In practice, w_i^*'s must be estimated, and such estimators wiil be correlated with the $G_{r,i}$. This linear combination idea is equivalent to a control variate formulation.

Example 4 Distribution Sensitivities for Quantiles. For $0\le\alpha\le1$, the α-VaR (or α-quantile) of a random variable $Z(\theta)=g(U;\theta)$ with cdf $F(\cdot;\theta)$ is defined as

$$q_\alpha(\theta) := \arg\min\{z: \ F(z;\theta)\ge\alpha\}.$$

When $F(\cdot;\theta)$ is continuous, $q_\alpha(\theta)=F^{-1}(\alpha;\theta)$. Let $U^{(j)}$, $j=1,\dots,m$, be i.i.d. realizations of $U\sim\mathcal{U}(0,1)^d$, and $\hat{F}_m(\cdot)$ be the empirical distribution of $Z_j := g(U^{(j)};\theta)$, $j=1,\dots,m$. The empirical α-quantile $\hat{F}_m^{-1}(\alpha)$, which is the inverse of the empirical distribution evaluated at α, is simply $Z_{(\lceil\alpha m\rceil)}$, where $Z_{(1)}<\cdots<Z_{(m)}$ are the realizations of $Z_1,\dots,Z_m$ sorted in increasing order (the order statistics), and $\lceil x\rceil$ denotes the smallest integer greater than or equal to x. This empirical quantile satisfies the following central limit theorem (Serfling, 1980):

$$\sqrt{m}\left(\hat{F}_m^{-1}(\alpha)-q_\alpha(\theta)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\alpha(1-\alpha)}{f(q_\alpha(\theta);\theta)}\right).$$

Traditionally, batching and sectioning techniques are used to estimate the asymptotic variance to construct a confidence interval on the empirical quantile, and these methods lead to subcanonical convergence rates (Nakayama, 2014). With the GLR density estimator, however, we can estimate the asymptotic variance by

$$\frac{m\alpha(1-\alpha)}{\sum_{j=1}^{m} G_{2,i}(U^{(j)};z,\theta)\big|_{z=\hat{F}_m^{-1}(\alpha)}},$$

using the same realizations of the uniform random variables $U^{(j)}$ as in the quantile estimator $\hat{F}_m^{-1}(\alpha)$. It follows from Peng et al. (2017) that this asymptotic variance estimator is consistent.

We can also estimate the quantile sensitivity $q'_\alpha(\theta) = \partial q_\alpha(\theta)/\partial\theta$ by estimating distribution sensitivities. By the implicit function theorem,

$$q'_\alpha(\theta) = -\left.\frac{\partial F(z;\theta)}{\partial\theta}\right|_{z=q_\alpha(\theta)} \Big/ f(q_\alpha(\theta);\theta),$$

so the quantile sensitivity can be estimated by the ratio

$$\widehat{D}(\theta) := -\left.\frac{\sum_{j=1}^m G_{1,i}(U^{(j)};z,\theta)}{\sum_{j=1}^m G_{2,i}(U^{(j)};z,\theta)}\right|_{z=\hat{F}_m^{-1}(\alpha)}.$$

This ratio estimator is biased, because the expectation of the ratio is not equal to the ratio of expectations and also because of the bias in the quantile estimator, but from Peng et al. (2017), it is consistent and obeys a central limit theorem when $m\to\infty$. Its asymptotic variance can also be estimated by the GLR estimator for higher-order distribution sensitivities (Peng et al., 2020a).

4.2. Stopping Time Problems

In this subsection, we consider estimating the derivative of the expectation of a sample performance that depends on a stopping time N:

$$\frac{\partial \mathbb{E}\left[\varphi_N(X_1,\ldots,X_N)\right]}{\partial\theta}, \tag{8}$$

where $\{X_i,\, i\geq 1\}$ is a Markov chain defined by the following stochastic recurrence:

$$X_1 = g_1(U_1;\theta), \qquad X_i = \kappa(X_{i-1}, g_i(U_i;\theta)), \tag{9}$$

and N is the first time the Markov chain hits a set Ω, i.e.,

$$N = \min\{n\in\mathbb{N}:\ X_n\in\Omega\} = \min\{n\in\mathbb{N}:\ \kappa(X_{n-1}, g_n(U_n;\theta))\in\Omega\}.$$

The expectation of this sample performance can be rewritten as

$$\mathbb{E}\left[\varphi_N(X_1,\ldots,X_N)\right] = \sum_{n=1}^{\infty}\mathbb{E}\left[\varphi_n(X_1,\ldots,X_n)\prod_{i=1}^{n-1}\mathbf{1}\{X_i\notin\Omega\}\mathbf{1}\{X_n\in\Omega\}\right]. \tag{10}$$

For each expectation in the summation on the right-hand side of the equation, the stochastic model falls into the special case $g(u;\theta) = (g_1(u_1;\theta),\ldots,g_n(u_1;\theta))$ discussed after Theorem 1, and the derivative of can be estimated by the GLR method. We have

$$\begin{aligned}\frac{\partial \mathbb{E}\left[\varphi_N(X_1,\ldots,X_N)\right]}{\partial\theta} &= \sum_{n=1}^{\infty}\sum_{i=1}^{n}\mathbb{E}\left[\varphi_n(X_1^{+i},\ldots,X_n^{+i})\prod_{j=1}^{n-1}\mathbf{1}\{X_j^{+i}\notin\Omega\}\mathbf{1}\{X_n^{+i}\in\Omega\}r_i(1^-;\theta)\right]\\ &-\sum_{n=1}^{\infty}\sum_{i=1}^{n}\mathbb{E}\left[\varphi_n(X_1^{-i},\ldots,X_n^{-i})\prod_{j=1}^{n-1}\mathbf{1}\{X_j^{-i}\notin\Omega\}\mathbf{1}\{X_n^{-i}\in\Omega\}r_i(0^+;\theta)\right]\\ &+\sum_{n=1}^{\infty}\mathbb{E}\left[\varphi_n(X_1,\ldots,X_n)\prod_{i=1}^{n-1}\mathbf{1}\{X_i\notin\Omega\}\mathbf{1}\{X_n\in\Omega\}d_n(U_1,\ldots,U_n;\theta)\right],\end{aligned}$$

where $\{X_j^{+i}, j \geq 1\}$ and $\{X_j^{-i}, j \geq 1\}$ are Markov chains generated by (9) with U_i replaced by 1 and 0, respectively, and $d_n(\cdot;\theta)$ denotes the second part in the GLR estimator for the nth term on the right-hand side of (10). We use the randomization technique of Rhee and Glynn (2015) to obtain a single-run unbiased estimator of (8), which is given by

$$\frac{1}{p(N')} S_{N'}(U_1,\ldots,U_{N'}) + D_N(U_1,\ldots,U_N), \tag{11}$$

where N' is a discrete random variable supported on $\mathbb{Z}^+$ with probability mass function $p(\cdot)$,

$$\begin{aligned} S_n(U_1,\ldots,U_n) \;:=\; & \sum_{i=1}^{n} \varphi_n(X_1^{+i},\ldots,X_n^{+i}) \prod_{j=1}^{n-1} \mathbf{1}\{X_j^{+i} \notin \Omega\} \mathbf{1}\{X_n^{+i} \in \Omega\} r_i(1^-;\theta) \\ & - \sum_{i=1}^{n} \varphi_n(X_1^{-i},\ldots,X_n^{-i}) \prod_{j=1}^{n-1} \mathbf{1}\{X_j^{-i} \notin \Omega\} \mathbf{1}\{X_n^{-i} \in \Omega\} r_i(0^+;\theta), \text{ and} \\ D_n(U_1,\ldots,U_n) \;:=\; & \varphi_n(X_1,\ldots,X_n) \prod_{i=1}^{n-1} \mathbf{1}\{X_i \notin \Omega\} \mathbf{1}\{X_n \in \Omega\} d_n(U_1,\ldots,U_n;\theta). \end{aligned}$$

This stopping time problem generalizes those in Glynn (1987) and Heidergott and Vazquez-Abad (2009) by allowing the distribution of the stopping time N to depend on parameter θ. The classical IPA and LR methods do not cover this case. In the next example, we discuss sensitivity analysis for control charts, widely used in statistical quality control.

Example 5 Controls Charts. Control charts aim to detect (statistically) when a manufacturing or business process goes out of control. The system is assumed to output samples having different statistical distributions when in control versus when out of control. The systems goes out of control at the (unobservable) random time χ with cdf $F_0(\cdot)$. The output Z_i of the ith sample has cdf $F_1(\cdot)$ when in control and $F_2(\cdot)$ when out of control. For an exponentially weighted moving average (EWMA) chart, the (observable) test statistic after the ith sample is $Y_i = \alpha Z_i + (1-\alpha) Y_{i-1}$. The stopping time N is the time when the system is declared out of control: $N = \min\{i : Y_i \notin [\theta_1,\theta_2]\} = \min\{i : X_i \notin [0,1]\}$, where θ_1 and θ_2 are fixed lower and upper control limits for the test statistic, and $X_i = \frac{Y_i-\theta_1}{\theta_2-\theta_1}$. Note that $\{X_i, i \geq 1\}$ is a Markov chain that follows the recursion

$$X_i = (1-\alpha)X_{i-1} + g_i(U_i;\theta), \qquad \text{where}$$
$$g_i(U_i;\theta) = \frac{\alpha}{\theta_2-\theta_1}\left[\mathbf{1}\{i < \chi/\Delta\} F_1^{-1}(U_i) + \mathbf{1}\{i \geq \chi/\Delta\} F_2^{-1}(U_i) - \theta_1\right],$$

and Δ is the sampling period (the time between any two successive monitoring epochs).

Here we consider sensitivity analysis with respect to $\theta = \theta_2$. This model falls into the special case of Corollary 1 where $g_i(u_i;\theta) = \xi_i(\theta)\eta_i(u_i)$, with $\xi_i(\theta) = \alpha/(\theta-\theta_1)$ and $\eta_i(u_i) = \mathbf{1}\{i < \chi/\Delta\} F_1^{-1}(u_i) + \mathbf{1}\{i \geq \chi/\Delta\} F_2^{-1}(u_i) - \theta_1$, for $i = 1,\ldots,n$. Then we have

$$r_i(u;\theta) = -\frac{1}{\theta_2-\theta_1}\left[\mathbf{1}\{i < \chi/\Delta\} F_1^{-1}(u_i) f_1(F_1^{-1}(u_i)) + \mathbf{1}\{i \geq \chi/\Delta\} F_2^{-1}(u_i) f_2(F_2^{-1}(u_i)) - \theta_1\right], \quad \text{and}$$

$$d(u;\theta) = \frac{1}{\theta_2 - \theta_1} \sum_{i=1}^{n} \left[\mathbf{1}\{i < \chi/\Delta\} \frac{F_1^{-1}(u_i)}{f_1(F_1^{-1}(u_i))} + \mathbf{1}\{i \geq \chi/\Delta\} \frac{F_2^{-1}(u_i)}{f_2(F_2^{-1}(u_i))} + 1 \right].$$

Suppose we pay a cost of c per unit of time when the system is out of control and this is not yet detected, and a one-time cost C to fix the system when it is declared out of control. This model is a regenerative process, which regenerates each time we fix the system. The expected cost over one regenerative cycle is $c\mathbb{E}[(N - \lceil \chi/\Delta \rceil)^+] + C$, and therefore the average cost per unit of time over an infinite horizon is

$$\frac{c\mathbb{E}[(N - \lceil \chi/\Delta \rceil)^+] + C}{\mathbb{E}[N]}. \tag{12}$$

The goal might be to select the control limits θ_1 and θ_2 to minimize this average cost. Widening the gap $\theta_2 - \theta_1$ would reduce the frequency of intervention, so we would pay the fixed cost C less often, but then the penalty c would be paid over longer periods of time on average. The optimal control limits achieve an optimal balance between these two types of costs. The sample performance of the expectation in the numerator of (12) is different from those treated by SPA in Fu and Hu (1999) and Fu et al. (2009b), the development of which depends on the specific structure of the problem. The GLR method in our work provides unbiased derivative estimators for the expectations in both the numerator and denominator of (12).

4.3. Credit Risk Derivatives

We consider two important types of credit risk derivatives: basket default swaps (BDSs) and collateralized debt obligations (CDOs). In a BDS contract, the buyer pays fixed premia $p_1, ..., p_k$ to the protection seller at dates $0 < T_1 < \ldots < T_k < T$, and if the ith default time $\tau_{(i)}$ occurs before T, i.e., $\tau_{(i)} < T$, these premium payments stop, and the seller undertakes the loss of the ith default and makes a payment to the buyer. Let L_i be the loss of the ith default. The discounted value of the ith default swap is the difference between the discounted payments made by the seller and those made by the buyer:

$$V_{\text{bds}}(\tau) = V_{\text{value}}(\tau) - V_{\text{prot}}(\tau),$$

where $V_{\text{prot}}(\tau)$ is the discounted premium payed by the buyer:

$$V_{\text{prot}}(\tau) = \begin{cases} \sum_{j=1}^{\ell} p_j \exp(-rT_j) + p_{\ell+1} \exp(-r\tau_{(i)}) \frac{\tau_{(i)} - T_\ell}{T_{\ell+1} - T_\ell}, & \text{if } T_\ell \leq \tau_{(i)} \leq T_{\ell+1}, \\ \sum_{j=1}^{k} p_i \exp(-rT_j), & \text{if } \tau_{(i)} > T, \end{cases}$$

and $V_{\text{value}}(X)$ the discounted payment made by the seller:

$$V_{\text{value}}(X) = L_{(i)} \exp(-r\tau_{(i)}) \mathbf{1}\{\tau_{(i)} < T\}.$$

In a CDO, the losses caused by the defaults of the assets in the portfolio are packaged together and then tranched. The tranches are ordered so that losses are absorbed sequentially following the order of the tranches. For example, a tranche of a CDO absorbs the loss above a threshold $\mathcal{L}_-$ and below a threshold $\mathcal{L}_+$, i.e.,

$$V_{\text{cdo}}(\tau) = (\mathcal{L} - \mathcal{L}_-) \cdot \mathbf{1}\{\mathcal{L} > \mathcal{L}_-\} - (\mathcal{L} - \mathcal{L}_+) \cdot \mathbf{1}\{\mathcal{L} > \mathcal{L}_+\}, \qquad \text{where}$$
$$\mathcal{L} = \sum_{i=1}^{n} L_i \cdot \mathbf{1}\{\tau_i < T\}.$$

Suppose that the vector of default times $(\tau_1, \ldots, \tau_n)$ have a joint distribution with marginal cdf's $F_i(\cdot)$, $i = 1, \ldots, n$, and a dependence structure modeled by an Archimedean copula. Then we can generate τ_i's by generating $V = (V_1, \ldots, V_n)$ from the copula, and putting $\tau_i = F_i^{-1}(V_i)$, $i = 1, \ldots, n$.

The sample performances $V_{\text{value}}(\tau)$, $V_{\text{prot}}(\tau)$, and $V_{\text{cdo}}(\tau)$ may be discontinuous with respect to the structural parameter θ in the copula model due to the presence of indicator functions and order indices. As a result, neither IPA nor LR can be applied directly for this model. On the other hand, $V_{\text{value}}(\tau)$, $V_{\text{prot}}(\tau)$, and $V_{\text{cdo}}(\tau)$ are all of the form $\varphi(g(U;\theta))$ that fits our framework, due to the generality of the measurable function $\varphi(\cdot)$. Unlike in Lei et al. (2020) where a separate CMC technique needs to be derived for each type of cash flow, the GLR method in this work can estimate the derivative of the expectation for all three types of cash flows.

5. Numerical Experiments

In this section, we present numerical examples to demonstrate broad applicability and flexibility of the proposed GLR method to estimate sensitivities in various situations. The examples include an indicator function applied to a linear combination of two exponential random variables, a stochastic activity network (SAN), control charts, and a CDO model.

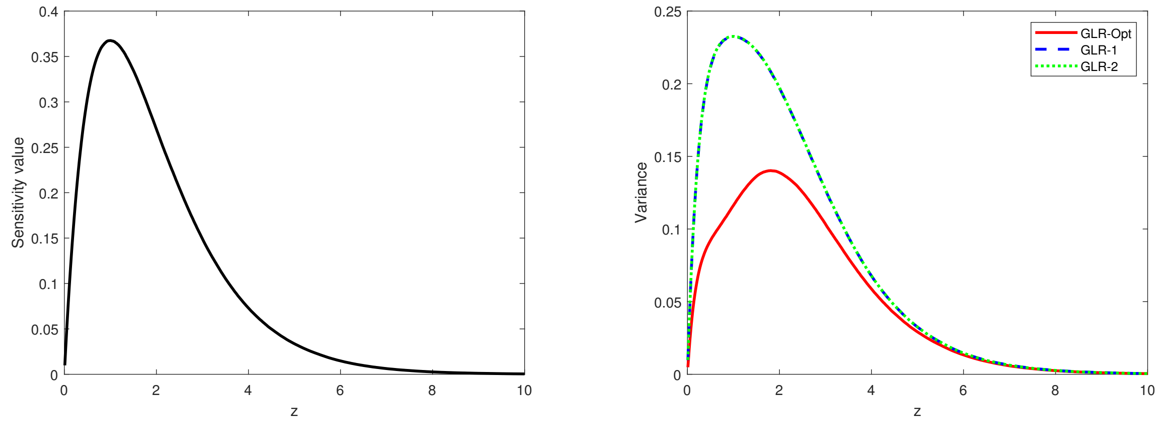
5.1. Distribution Sensitivities for a Linear Model

We estimate distribution sensitivities where $\varphi(\cdot)$ is an indicator function for a linear combination of two independent exponential random variables with means $1/\lambda_1$ and $1/\lambda_2$, i.e.,

$$\varphi(g(U;\theta)) = \mathbf{1}\{g(U;\theta) \leq z\} \quad \text{where} \quad g(U;\theta) = -\frac{\theta}{\lambda_1}\log(U_1) - \frac{1}{\lambda_2}\log(U_2).$$

This sample performance falls into the special case stated in Corollary 1, and the conditions in Proposition 1 and the integrability condition on the weight function discussed in Section 4.1 can be checked straightforwardly. We have

$$\frac{\partial g(u;\theta)}{\partial u_1} = -\frac{\theta}{\lambda_1 u_1}, \quad \frac{\partial g(u;\theta)}{\partial u_2} = -\frac{1}{\lambda_2 u_2}, \quad \frac{\partial g(u;\theta)}{\partial \theta} = -\frac{1}{\lambda_1}\log u_1, \qquad \text{and}$$
$$\frac{\partial^2 g(u;\theta)}{\partial u_1^2} = \frac{\theta}{\lambda_1 u_1^2}, \quad \frac{\partial^2 g(u;\theta)}{\partial u_2^2} = \frac{1}{\lambda_2 u_2^2}, \quad \frac{\partial^2 g(u;\theta)}{\partial \theta \partial u_1} = -\frac{1}{\lambda_1 u_1}, \quad \frac{\partial^2 g(u;\theta)}{\partial \theta \partial u_2} = 0.$$



(a) Distribution sensitivity curve.

(b) Variance curves.

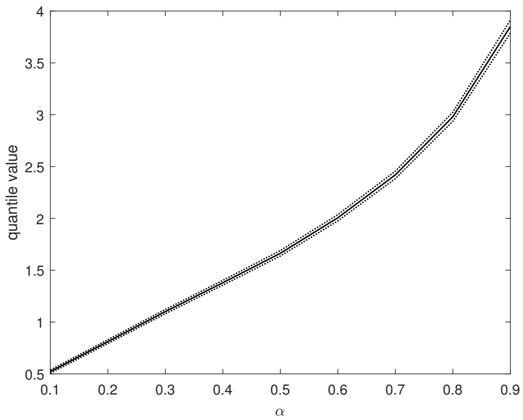
Figure 1 Density curve estimated by GLR in the example of Section 5.1.

Then the GLR estimators for distribution sensitivities in Section 4.1 are

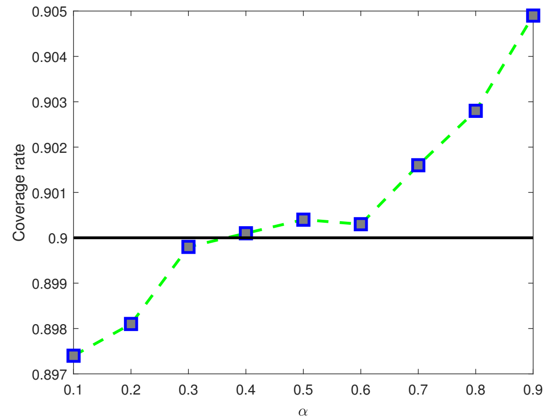
$$G_{1,1}(U;z,\theta) = -\frac{1}{\theta}\mathbf{1}\{g(U;\theta)\le z\}\,(\log U_1 + 1)\,,$$
$$G_{1,2}(U;z,\theta) = \frac{\lambda_2}{\lambda_1}\left[\mathbf{1}\left\{-\frac{\theta}{\lambda_1}\log U_1 \le z\right\}\log U_1 - \mathbf{1}\{g(U;\theta)\le z\}\log U_1\right],$$
$$G_{2,1}(U;z,\theta) = \frac{\lambda_1}{\theta}\left[\mathbf{1}\left\{-\frac{1}{\lambda_2}\log U_2 \le z\right\} - \mathbf{1}\{g(U;\theta)\le z\}\right],$$
$$G_{2,2}(U;z,\theta) = \lambda_2\left[\mathbf{1}\left\{-\frac{\theta}{\lambda_1}\log U_1 \le z\right\} - \mathbf{1}\{g(U;\theta)\le z\}\right].$$

For our numerical experiments, we take $\lambda_1 = 1$, $\lambda_2 = 1$, and $\theta = 1$, and we estimate the density of $g(U;\theta)$ at z, as a function of z from $z = 0.01$ to 10 with a step size of 0.01. We perform 10^6 independent simulation runs with GLR. The curves of the estimated density and estimated variances as a function of z are given in Figure 1. We also estimate the sensitivity of $\mathbb{E}[\varphi(g(U;\theta))]$ with respect to θ at z, as a function of z from $z = 0.01$ to 10 with a step size of 0.01. The numerical observations are similar, and the details can be found in the online Appendix B.

Figure 1(a) shows the estimated density curve, and Figure 1(b) presents the sample variance curves of three distribution sensitivity estimators: $G_{2,1}(\cdot)$ (GLR-1), $G_{2,2}(\cdot)$ (GLR-2), and a combined GLR estimator that minimizes the variance of the linear combination of GLR-1 and GLR-2 with weights given by (GLR-Opt). Due to page limit, the variance comparison between GLR and the finite difference method with common random numbers (FDC) is relegated to online Appendix B. The variance of FDC is much larger than those of the GLR estimators, and the variance of FDC(0.01) is about 10 times of that of FDC(0.1), which indicates that FDC suffers from a bias-variance tradeoff issue.



(a) Confidence intervals for quantiles.



(b) Coverage rates of confidence intervals.

Figure 2 90%-confidence intervals estimated by GLR for quantiles and coverage rates in example of Section 5.1.

In Figure 2(a), the solid line in the center is the estimate for the α-quantiles q_α of $g(U;\theta)$, $\alpha = 0.1,\ldots,0.9$, using a batch of 10^4 independent simulation replications, and the upper and lower dotted lines are calculated respectively by adding and subtracting twice the estimated standard deviation to the estimated quantile values, which are estimated by GLR using the same batch of 10^4 independent simulation replications for estimating the quantiles. Figure 2(b) presents the coverage rates of the 90%-confidence intervals for quantiles q_α, $\alpha = 0.1,\ldots,0.9$, by 10^4 independent macro experiments. The true quantile values can be calculated by inverting a hypoexponential distribution (L'Ecuyer et al., 2019, Section 4.6). We can see the coverage rates of the 90%-confidence intervals match the target value statistically. As the quantile gets closer to the tail of the distribution, the variances of the quantile estimates and confidence interval estimates are larger because there are many fewer samples in the tail.

5.2. Distribution Sensitivities for a Stochastic Activity Network

We estimate distribution sensitivities for the sample performance of a small SAN depicted in Figure 3. There are five nodes representing different stages of activity. The nodes are connected by the arcs representing the activities in each stage. The durations of activities follow independent exponential distributions, i.e., $X_i = -\frac{1}{\lambda_i}\log(U_i), \quad i = 1,\ldots,6$. Let $\theta = \lambda_6$. There are three different paths representing the tasks to reach the final stage of a project, i.e., $\pi_1 = (1,4,6), \quad \pi_2 = (2,5,6), \quad \pi_3 = (1,3,5,6)$, and the completion time for each path is additive, i.e, $\sum_{j\in\pi_i} X_j$, $i = 1,2,3$. The completion time for the entire project is $\max(X_1+X_4+X_6, X_2+X_5+X_6, X_1+X_3+X_5+X_6)$, and the sample performance for the distribution function of completion time is

$$\begin{aligned}\varphi(g(U;\theta)) =& \mathbf{1}\left\{\max(X_1+X_4+X_6, X_2+X_5+X_6, X_1+X_3+X_5+X_6) \le z\right\}\\ =& \mathbf{1}\left\{X_1+X_4+X_6 \le z\right\}\mathbf{1}\left\{X_2+X_5+X_6 \le z\right\}\mathbf{1}\left\{X_1+X_3+X_5+X_6 \le z\right\},\end{aligned}$$

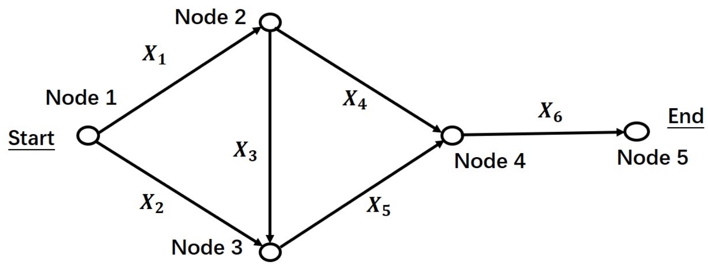


Figure 3 A SAN with six activities.

where $\varphi(v_1,v_2,v_3)=\prod_{i=1}^{3}\mathbf{1}\{v_i\le 0\}$, and we have

$$g_i(u;\theta)=-\sum_{j\in\pi_i}\frac{1}{\lambda_j}\log u_j - z,\quad i=1,2,3,$$

$$\partial_z g(u;\theta)=-\Big(1,1,1\Big)^T \quad\text{and}\quad \partial_\theta g(u;\theta)=\frac{\log u_6}{\theta^2}\Big(1,1,1\Big)^T.$$

This sample performance goes beyond the setting in Section 4.1, but it can be put under the more general stochastic model (1). In the theory of GLR, the dimension of the vector of the input uniform random numbers is assumed to be the same as that of the argument vector of function g. For the SAN model, there are six input uniform random numbers, while the dimension of the argument vector of g is three. Therefore, we can arbitrarily choose three uniform random numbers to condition on and treat the remaining three as the inputs to the stochastic models for deriving GLR. We condition on (U_4,U_5,U_6) and treat (U_1,U_2,U_3) as the input uniform random numbers for deriving GLR. The Jacobian matrix is

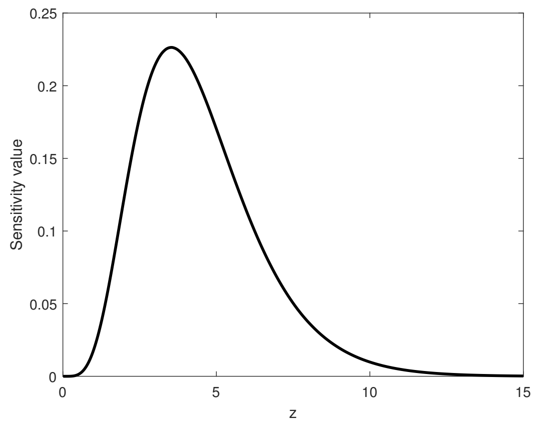
$$J_g(u;\theta,z)=-\begin{pmatrix}\frac{1}{\lambda_1 u_1} & 0 & 0\\ 0 & \frac{1}{\lambda_2 u_2} & 0\\ \frac{1}{\lambda_1 u_1} & 0 & \frac{1}{\lambda_3 u_3}\end{pmatrix}.$$

The GLR estimator for $\mathbb{E}[\varphi(g(U;\theta))]$ with respect to θ can be obtained by multiplying the GLR estimator for the density by $-\log U_6/\theta^2$. Here we only report on the performance of the GLR estimators for the density of $g(U;\theta)$. Additional results for quantiles can be found in online Appendix B. We have

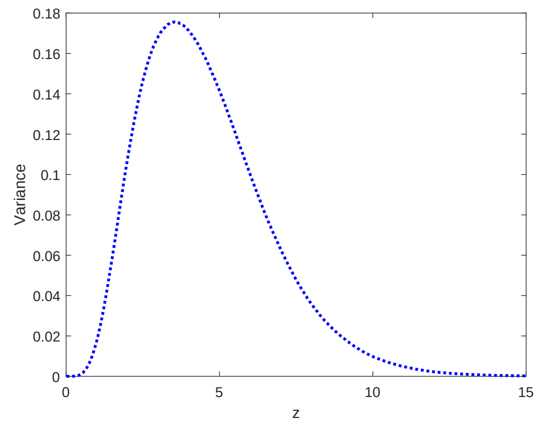
$$r_1(u;\theta,z)=\lambda_1 u_1,\quad r_2(u;\theta,z)=\lambda_2 u_2,\quad r_3(u;\theta,z)=0,\quad\text{and}\quad d(u;\theta,z)=-\lambda_1-\lambda_2.$$

In this case, condition (A.1) in Theorem 1 can be justified by checking conditions in Proposition 1. Conditions (A.2) and (A.3) can also be checked straightforwardly.

In the experiments, we set $\lambda_i=1$, $i=1,\ldots,6$, in the stochastic model, and estimate distribution sensitivity curves from $z=0.01$ to 15 with a step size of 0.01. Figure 4 presents the density and variance curves estimated by the GLR method using 10^6 independent simulation replications.



(a) Density curve.



(b) Variance curves.

Figure 4 Density estimation by GLR in the SAN example of Section 5.2.

5.3. Sensitivities of Control Charts

We estimate sensitivities of the expectations in the numerator and denominator of (12) for control charts with respect to upper control limit $\theta = \theta_2$ discussed in Section 4.2. When the system is in control, the output sample is assumed to follow a uniform distribution on $[-1, 1]$ and we define $Z_i = 2U_i - 1$. When the system is out of control, the output sample is assumed to follow a uniform distribution on $[0, 2]$ and we define $Z_i = 2U_i$. Then the transition function of the Markov chain is

$$X_i = (1-\alpha)X_{i-1} + \frac{\alpha}{\theta_2 - \theta_1}\left[\mathbf{1}\{i < \chi/\Delta\}(2U_i - 1) + \mathbf{1}\{i \geq \chi/\Delta\}2U_i - \theta_1\right],$$

and the weight functions in the GLR estimator are

$$r_i(u;\theta) = -\frac{1}{\theta_2 - \theta_1}\left[\mathbf{1}\{i < \chi/\Delta\}\left(u_i - \frac{1}{2}\right) + \mathbf{1}\{i \geq \chi/\Delta\}u_i - \frac{\theta_1}{2}\right],$$
$$d(u;\theta) = \frac{n}{\theta_2 - \theta_1}.$$

In the experiment, we set $\alpha = 1/2$, $\theta_1 = -1$, $\theta_2 = 1$, $\Delta = 1$, $X_0 = 0$, and assume $\chi = 1 + 3\log U$. The randomized horizon N' in (11) follows a geometric distribution with parameter 0.1. In the random horizon problem, it is not easy to synchronize the two sample paths for the finite difference (FD) method, because perturbing parameter θ affects the stopping time N, so it would require substantially more computational overhead to generate the sample paths using common random numbers to implement FDC. In general, decreasing the perturbation size δ would reduce the bias of FD(δ) but increase the variance. In Table 1, we can see that the sensitivity results estimated by GLR match those estimated by FD(0.01), while GLR has smaller variance than FD(0.01), whereas the sensitivity results estimated by FD(0.1) are significantly biased.

5.4. Sensitivities of Collateralized Debt Obligations

We estimate the sensitivity with respect to the parameter θ that governs the dependence in the copula model for the expectation of the loss absorbed by the tranche that covers the first 30% of the total losses for 10 assets if there are defaults, i.e., $\mathcal{L}_- = 0$ and $\mathcal{L}_+ = 0.3 \times (\sum_{i=1}^{10} L_i)$. We set $r = 0.1$ and $T = 1$. The marginal distributions of the defaults are assumed to be exponential, so $\tau_i = -\frac{1}{\lambda_i}\log(X_i),\ i = 1, \ldots, 10$. The parameters λ_i and loss L_i, $i = 1, \ldots, n$, are randomly generated from the uniform distribution over $(0, 1)$ in the experiments.

We compare the GLR estimator with FDC(δ), where δ is the perturbation size. Due to the simplicity of the weight function of GLR, the computational time of GLR barely increases relative to that required to run the simulation model itself, so the sensitivity estimate by GLR is almost a free byproduct that can be obtained simultaneously during the simulation. Table 2 shows the sensitivity estimates with sample sizes $m = 10^4$, $m = 10^5$, and $m = 10^6$ under the Clayton copula, for $\theta = 0.5$. We can see that the variances with GLR are comparable to those with FDC(0.1), but smaller than those with FDC(0.01). For sample size $m = 10^6$, the estimate with FDC(0.1) lies outside of the 90% confidence interval of the GLR estimate, whereas the estimate with GLR lies in the 90% confidence interval of the FDC(0.01) estimate with the sample size $m = 10^7$, which is -0.179 ± 0.002. This indicates that FDC suffers from the bias-variance tradeoff, while GLR is accurate under a relatively small sample size. Numerical results for sensitivities of CDOs under the Ali-Mikhail-Haq copula and sensitivities of BDS under both the Clayton and Ali-Mikhail-Haq copulas can be found in online Appendix B. The observations are similar to those in this example.

6. Conclusions

In this paper, a GLR method is proposed for a family of stochastic models with uniform random numbers as inputs. The framework studied in this work covers a large range of discontinuities, and

	GLR	FD(0.01)	FD(0.1)
$\partial\mathbb{E}[N]/\partial\theta$	8.6 ± 0.2	8.6 ± 0.4	10.5 ± 0.04
$\partial\mathbb{E}[(N - \lceil\chi/\Delta\rceil)^+]/\partial\theta$	8.8 ± 0.07	8.6 ± 0.4	10.5 ± 0.04

Table 1 Derivatives of control charts with respect to upper control limit $\theta = 1$, based on 10^6 independent replications (mean $\pm$ standard error).

	$m = 10^4$	$m = 10^5$	$m = 10^6$
GLR	-0.187 ± 0.01	-0.185 ± 0.004	-0.181 ± 0.002
FDC(0.1)	-0.171 ± 0.007	-0.177 ± 0.003	-0.176 ± 0.002
FDC(0.01)	-0.155 ± 0.07	-0.186 ± 0.02	-0.189 ± 0.006

Table 2 Sensitivity estimates of CDO with 10 assets governed by the Clayton copula with $\theta = 0.5$ based on 10^2 experiments (mean $\pm$ standard error).

it includes many applications such as density estimation, and sensitivity analysis for statistical quality control and credit risk financial derivatives. Since uniform random numbers are the basic building blocks for generating other random variables, our new method significantly relaxes the limitations on the input random variables in Peng et al. (2018) and Peng et al. (2020a). The technical conditions for justifying unbiasedness of GLR are relatively easy to satisfy in practice, and we show how to verify them on illustrative examples. Ongoing work includes combining GLR with randomized quasi-Monte Carlo methods and CMC.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 71901003, 71720107003, 71690232, 91846301, 71790615, by the National Science Foundation (NSF) under Grant CMMI-1434419, by Discovery Grant RGPIN-2018-05795 from NSERC-Canada, and by the Key Research and Development Program of Beijing Municipal Science and Technology Commission.

References

Asmussen, Søren, Peter W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*, vol. 57. Springer.

Bashyam, Sridhar, Michael C. Fu. 1998. Optimization of (s, S) inventory systems with random lead times and a service level constraint. *Management Science* **44**(12-part-2) S243–S256.

Broadie, Mark, Paul Glasserman. 1996. Estimating security price derivatives using simulation. *Management Science* **42**(2) 269–285.

Chen, Nan, Yanchu Liu. 2014. American option sensitivities estimation via a generalized infinitesimal perturbation analysis approach. *Operations Research* **62**(3) 616–632.

Chen, Zhiyong, Paul Glasserman. 2008. Sensitivity estimates for portfolio credit derivatives using Monte Carlo. *Finance and Stochastics* **12**(4) 507–540.

Cui, Zhenyu, Michael C. Fu, Jian-Qiang Hu, Yanchu Liu, Yijie Peng, Lingjiong Zhu. 2020. Technical note–On the variance of single-run unbiased stochastic derivative estimators. *INFORMS Journal on Computing* **32**(2) 390–407.

Embrechts, Paul, Filip Lindskog, Alexander McNeil. 2003. Modelling dependence with copulas and applications to risk management. S. Rachev, ed., *Handbook of Heavy Tailed Distributions in Finance*. Elsevier, 329–384. Chapter 8.

Fu, Michael C. 1994. Sample path derivatives for (s, S) inventory systems. *Operations Research* **42**(2) 351–364.

Fu, Michael C. 2015. Stochastic gradient estimation, Fu, Michael C. ed. *Handbook of Simulation Optimization*, vol. 216. Springer, 105–147. Chapter 5.

Fu, Michael C., L. Jeff Hong, Jian-Qiang Hu. 2009a. Conditional Monte Carlo estimation of quantile sensitivities. *Management Science* **55**(12) 2019–2027.

Fu, Michael C., Jian-Qiang Hu. 1993. Second derivative sample path estimators for the GI/G/m queue. *Management Science* **39**(3) 359–383.

Fu, Michael C., Jian-Qiang Hu. 1995. Sensitivity analysis for Monte Carlo simulation of option pricing. *Probability in the Engineering and Informational Sciences* **9**(3) 417–446.

Fu, Michael C., Jian-Qiang Hu. 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Kluwer Academic Publishers, Boston.

Fu, Michael C., Jian-Qiang Hu. 1999. Efficient design and sensitivity analysis of control charts using Monte Carlo simulation. *Management Science* **45**(3) 395–413.

Fu, Michael C., Shreevardhan Lele, Thomas W. M. Vossen. 2009b. Conditional Monte Carlo gradient estimation in economic design of control limits. *Production and Operations Management* **18**(1) 60–77.

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

Glasserman, Paul, Zongjian Liu. 2010. Sensitivity estimates from characteristic functions. *Operations Research* **58**(6) 1611–1623.

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

Glynn, Peter W. 1990. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM* **33**(10) 75–84.

Glynn, Peter W., Pierre L'Ecuyer. 1995. Likelihood ratio gradient estimation for stochastic recursions. *Advances in applied probability* **27**(4) 1019–1053.

Glynn, Peter W., Yijie Peng, Michael C. Fu, Jianqiang Hu. 2020. Computing sensitivities for distortion risk measures. *submitted to INFORMS Journal on Computing, under third review* .

Gong, Wei-Bo, Yu-Chi Ho. 1987. Smoothed perturbation analysis of discrete event dynamical systems. *IEEE Transactions on Automatic Control* **32**(10) 858–866.

Hammersley, John M., D. C. Handscomb. 1964. *Monte Carlo Methods*. Methuen, London.

Heidergott, Bernd. 1999. Optimisation of a single-component maintenance system: a smoothed perturbation analysis approach. *European Journal of Operational Research* **119**(1) 181–190.

Heidergott, Bernd, Taoying Farenhorst-Yuan. 2010. Gradient estimation for multicomponent maintenance systems with age-replacement policy. *Operations Research* **58**(3) 706–718.

Heidergott, Bernd, Haralambie Leahu. 2010. Weak differentiability of product measures. *Mathematics of Operations Research* **35**(1) 27–51.

Heidergott, Bernd, Felisa J. Vazquez-Abad. 2009. Gradient estimation for a class of systems with bulk services (a problem in public transportation). *ACM Transactions on Modeling and Computer Simulation* **19**(3) 13.

Heidergott, Bernd, Warren Volk-Makarewicz. 2016. A measure-valued differentiation approach to sensitivity analysis of quantiles. *Mathematics of Operations Research* **41**(1) 293–317.

Ho, Yu-Chi, Xi-Ren Cao. 1991. *Discrete Event Dynamic Systems and Perturbation Analysis*. Kluwer Academic Publishers, Boston, MA.

Hong, L. Jeff. 2009. Estimating quantile sensitivities. *Operations Research* **57**(1) 118–130.

Hong, L. Jeff, Sandeep Juneja, Jun Luo. 2014. Estimating sensitivities of portfolio credit risk using Monte Carlo. *INFORMS Journal on Computing* **26**(4) 848–865.

Hong, L Jeff, Guangwu Liu. 2009. Simulating sensitivities of conditional value at risk. *Management Science* **55**(2) 281–293.

Jiang, Guangxin, Michael C. Fu. 2015. Technical note — On estimating quantile sensitivities via infinitesimal perturbation analysis. *Operations Research* **63**(2) 435–441.

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

L'Ecuyer, Pierre, Peter W. Glynn. 1994. Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady-state. *Management Science* **40**(11) 1562–1578.

L'Ecuyer, Pierre, Florian Puchhammer, Amal Ben Abdellah. 2019. Monte Carlo and quasi-Monte Carlo density estimation via conditioning. *arXiv preprint arXiv:1906.04607* .

Lei, Lei, Yijie Peng, Michael C. Fu, Jian-Qiang Hu. 2018. Applications of generalized likelihood ratio method to distribution sensitivities and steady-state simulation. *Journal of Discrete Event Dynamic Systems* **28**(1) 109–125.

Lei, Lei, Yijie Peng, Michael C. Fu, Jian-Qiang Hu. 2020. Sensitivity analysis of portfolio credit derivatives by conditional Monte Carlo simulation. *preprint in ResearchGate with DOI: 10.13140/RG.2.2.21951.97444* .

Li, David. 2000. On default correlation: a copula approach. *Journal of Fixed Income* **9** 43–54.

Liu, Guangwu, L. Jeff Hong. 2011. Kernel estimation of the Greeks for options with discontinuous payoffs. *Operations Research* **59**(1) 96–108.

Marshall, Albert W., Ingram Olkin. 1988. Families of multivariate distributions. *Journal of the American Statistical Association* **83**(403) 834–841.

Mohamed, Shakir, Mihaela Rosca, Michael Figurnov, Andriy Mnih. 2019. Monte Carlo gradient estimation in machine learning. *arXiv preprint arXiv:1906.10652* .

Nakayama, Marvin K. 2014. Confidence intervals for quantiles using sectioning when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation* **24**(4) 1–21.

Nelsen, Roger B. 2006. *An Introduction to Copulas*, vol. 139. 2nd ed. Springer Science & Business Media.

Peng, Yijie, Michael C Fu, Peter W Glynn, Jian-Qiang Hu. 2017. On the asymptotic analysis of quantile sensitivity estimation by Monte Carlo simulation. *Proceedings of Winter Simulation Conference* 2336–2347.

Peng, Yijie, Michael C Fu, Bernd Heidergott, Henry Lam. 2020a. Maximum likelihood estimation by Monte Carlo simulation: Towards data-driven stochastic modeling. *Operations Research, forthcoming* .

Peng, Yijie, Michael C. Fu, Jian-Qiang Hu. 2016. On the regularity conditions and applications for generalized likelihood ratio method. *Proceedings of Winter Simulation Conference*. IEEE Press, 919–930.

Peng, Yijie, Michael C. Fu, Jian-Qiang Hu, Bernd Heidergott. 2018. A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. *Operations Research* **66**(2) 487–499.

Peng, Yijie, Li Xiao, Bernd Heidergott, Jeff Hong, Henry Lam. 2020b. Stochastic gradient estimation for artificial neural networks. *submitted to INFORMS Journal on Computing, under second review* .

Pflug, Georg Ch. 1988. Derivatives of probability measures-concepts and applications to the optimization of stochastic systems. Pravin Varaiya, Alexander B. Kurzhanski, eds., *Discrete Event Systems: Models and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 252–274.

Rhee, Chang-han, Peter W. Glynn. 2015. Unbiased estimation with square root convergence for SDE models. *Operations Research* **63**(5) 1026–1043.

Rubinstein, Reuven Y., Alexander Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. Wiley, New York.

Serfling, Robert J. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley.

Sklar, A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8** 229–231.

Suri, Rajan, Michael A. Zazanis. 1988. Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue. *Management Science* **34**(1) 39–64.

Wang, Yongqiang, Michael C. Fu, Steven I. Marcus. 2012. A new stochastic derivative estimator for discontinuous payoff functions with application to financial derivatives. *Operations Research* **60**(2) 447–460.

Online Supplements for "Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers As Inputs"

Yijie Peng
Department of Management Science and Information Systems, Guanghua School of Management, Peking University, Beijing, 100871, China, pengyijie@pku.edu.cn

Michael C. Fu
The Robert H. Smith School of Business, Institute for Systems Research, University of Maryland, College Park, Maryland 20742-1815, USA, mfu@umd.edu

Jiaqiao Hu
Applied Mathematics & Statistics, College of Engineering and Applied Sciences, Stony Brook University, Stony Brook, NY 11794-3600, USA, jqhu@ams.stonybrook.edu

Pierre L'Ecuyer
Department of Computer Science and Operations Research, University of Montreal, Montreal, Quebec, Canada, lecuyer@iro.umontreal.ca

Bruno Tuffin
Inria, Univ Rennes, CNRS, IRISA, Campus de Beaulieu, 35042 RENNES Cedex, France, bruno.tuffin@inria.fr

Appendix A: Theoretical Supplements

Proof of Theorem 1 under conditions (A.1)-(A.3)

Proof. We have

$$\begin{aligned}\frac{\partial}{\partial\theta}\int_{[\varepsilon,1-\varepsilon]^n}\varphi_\epsilon(g(u;\theta))\ du &= \int_{[\varepsilon,1-\varepsilon]^n}\nabla_x\varphi_\epsilon(x)|_{x=g(u;\theta)}\partial_\theta g(u;\theta)\ du\\ &= \int_{[\varepsilon,1-\varepsilon]^n}\nabla_u\varphi_\epsilon(g(u;\theta))J_g^{-1}(u;\theta)\partial_\theta g(u;\theta)\ du,\end{aligned}$$

where

$$\nabla\varphi_\epsilon(x) := \left(\frac{\partial\varphi_\epsilon(x)}{\partial x_1},\dots,\frac{\partial\varphi_\epsilon(x)}{\partial x_n}\right).$$

The interchange of the derivative and integration can be justified under condition (A.3), which implies that $\partial_\theta g(u;\theta)$ is bounded in $\Omega_\varepsilon \times \Theta$. By the Gauss-Green Theorem,

$$\begin{aligned}
&\int_{[\varepsilon,1-\varepsilon]^n} \nabla_u \varphi_\epsilon(g(u;\theta)) J_g^{-1}(u;\theta)\partial_\theta g(u;\theta)\ du_1 \cdots du_n \\
=&\sum_{i=1}^n \int_{[\varepsilon,1-\varepsilon]^{n-1}} \varphi_\epsilon(g(u;\theta)) \left(J_g^{-1}(u;\theta)\partial_\theta g(u;\theta)\right)^T e_i \ \prod_{j\neq i} du_j \Big|_{u_i=\varepsilon}^{1-\varepsilon} \\
&-\int_{[\varepsilon,1-\varepsilon]^n} \varphi_\epsilon(g(u;\theta)) \mathrm{div}\left(J_g^{-1}(u;\theta)\partial_\theta g(u;\theta)\right)\ du_1 \cdots du_n,
\end{aligned}$$

where for $n=1$, the integration in first term on the right-hand side of the equation is absent, and for $h(u) = (h_1(u), \ldots, h_n(u))$,

$$\mathrm{div}(h(u)) := \sum_{i=1}^n \frac{\partial h(u)}{\partial u_i}.$$

Then

$$\begin{aligned}
\mathrm{div}\left(J_g^{-1}(u;\theta)\partial_\theta g(u;\theta)\right) &= \sum_{i=1}^n \frac{\partial}{\partial u_i} e_i^T J_g^{-1}(u)\partial_\theta g(u;\theta) \\
&= \sum_{i=1}^n e_i^T \partial_{u_i} J_g^{-1}(u;\theta)\partial_\theta g(u;\theta) + \mathrm{trace}(J_g^{-1}(u)\partial_\theta J_g(u;\theta)).
\end{aligned}$$

By differentiating equation $J_g^{-1}(u;\theta)J_g(u;\theta) = I$ with respect to u_i on both sides, we have

$$\begin{aligned}
0 =&\partial_{u_i}\left(J_g^{-1}(u;\theta)J_g(u;\theta)\right) \\
=&\partial_{u_i} J_g^{-1}(u;\theta)J_g(u;\theta) + J_g^{-1}(u;\theta)\partial_{u_i} J_g(u;\theta),
\end{aligned}$$

which leads to

$$\partial_{u_i} J_g^{-1}(u;\theta) = -J_g^{-1}(u;\theta)\partial_{u_i} J_g(u;\theta) J_g^{-1}(u;\theta).$$

Therefore, we have

$$d(u;\theta) = -\mathrm{div}\left(J_g^{-1}(u;\theta)\partial_\theta g(u;\theta)\right).$$

With the discussion above and condition (A.2),

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int_{[\varepsilon,1-\varepsilon]^n} \varphi_\epsilon(g(u;\theta))\ du \\
=&\sum_{i=1}^n \int_{[\varepsilon,1-\varepsilon]^{n-1}} \varphi_\epsilon(g(u;\theta)) r_i(u;\theta)\ \prod_{j\neq i} du_j \Big|_{u_i=\varepsilon}^{1-\varepsilon} + \int_{[\varepsilon,1-\varepsilon]^n} \varphi_\epsilon(g(u;\theta)) d(u;\theta)\ du.
\end{aligned} \tag{1}$$

With condition (A.2), $r_i(u;\theta)$ and $d(u;\theta)$ are bounded in $\Omega_\varepsilon \times \Theta$, and with condition (A.1),

$$\begin{aligned}
&\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}\left|\int_{[\varepsilon,1-\varepsilon]^n}(\varphi_\epsilon(g(u;\theta))-\varphi(g(u;\theta)))d(u;\theta)\ du\right| \\
\leq &\sup_{\theta\in\Theta}\left(\int_{[\varepsilon,1-\varepsilon]^n}|\varphi_\epsilon(g(u;\theta))-\varphi(g(u;\theta))|^p du\right)^{1/p}\left(\int_{[\varepsilon,1-\varepsilon]^n}|d(u;\theta)|^q du\right)^{1/q} \\
\leq &C\lim_{\epsilon\to 0}\left(\int_{[\varepsilon,1-\varepsilon]^n}|\varphi_\epsilon(g(u;\theta))-\varphi(g(u;\theta))|^p du\right)^{1/p}=0,
\end{aligned}$$

where the first inequality holds by applying Hölder's inequality for $1/p+1/q=1$, and C is a positive constant. Similarly, we can show the uniform convergence of the first term on the right-hand side of (1) for $n>1$, and

$$\lim_{\epsilon\to 0}\left|\int_{[\varepsilon,1-\varepsilon]^n}\varphi_\epsilon(g(u;\theta))\ du-\int_{[\varepsilon,1-\varepsilon]^n}\varphi(g(u;\theta))\ du\right|=0.$$

For $n=1$, the first term on the right-hand side of (1) becomes

$$\varphi_\epsilon(g(1-\varepsilon;\theta))r_1(1-\varepsilon;\theta)-\varphi_\epsilon(g(\varepsilon;\theta))r_1(\varepsilon;\theta),$$

and condition (A.1) results in

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}|(\varphi_\epsilon(g(1-\varepsilon;\theta))-\varphi(g(1-\varepsilon;\theta)))r_1(1-\varepsilon;\theta)-(\varphi_\epsilon(g(\varepsilon;\theta))-\varphi(g(\varepsilon;\theta)))r_1(\varepsilon;\theta)|=0.$$

Therefore,

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int_{[\varepsilon,1-\varepsilon]^n}\varphi(g(u;\theta))\ du=\lim_{\epsilon\to 0}\frac{\partial}{\partial\theta}\int_{[\varepsilon,1-\varepsilon]^n}\varphi_\epsilon(g(u;\theta))\ du \\
&=\sum_{i=1}^n\int_{[\varepsilon,1-\varepsilon]^{n-1}}\varphi(g(u;\theta))\ r_i(u;\theta)\ \prod_{j\neq i}du_j\Big|_{u_i=\varepsilon}^{1-\varepsilon}+\int_{[\varepsilon,1-\varepsilon]^n}\varphi(g(u;\theta))\ d(u;\theta)\ du.
\end{aligned}$$

With condition (A.3), uniform convergence of the integrals on the right-hand side of equation above can be established over $\theta\in\Theta$ as $\varepsilon\to 0$, and we have

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int_{(0,1)^n}\varphi(g(u;\theta))\ du=\lim_{\varepsilon\to 0}\frac{\partial}{\partial\theta}\int_{[\varepsilon,1-\varepsilon]^n}\varphi(g(u;\theta))\ du \\
&=\sum_{i=1}^n\int_{(0,1)^{n-1}}\left[\lim_{u_i\to 1^-}\varphi(g(u;\theta))\ r_i(u;\theta)-\lim_{u_i\to 0^+}\varphi(g(u;\theta))\ r_i(u;\theta)\right]\ \prod_{j\neq i}du_j \\
&\quad+\int_{(0,1)^n}\varphi(g(u;\theta))\ d(u;\theta)\ du,
\end{aligned}$$

which completes the proof.

Proof of Theorem 1 under conditions (A.2)-(A.4)

Proof. From Peng et al. (2018), there exists a sequence of bounded functions $|\varphi_L(x)| \leq \varphi(x)$ such that $\lim_{L\to\infty} \varphi_L(x) = \varphi(x)$, and there exists a sequence of bounded and smooth functions $\varphi_{\epsilon,L}(\cdot)$ such that

$$\lim_{L\to\infty} \|\varphi_{\epsilon,L} - \varphi_L\|_p = 0,$$

where $p > 1$, and $\|h\|_p := \left(\int_{\mathbb{R}^n} |h(x)|^p dx\right)^{1/p}$. Except for replacing $\varphi_\epsilon(\cdot)$ with $\varphi_{\epsilon,L}(\cdot)$, the procedures before (1) are the same as in the proof for Theorem 1 under conditions (A.1)-(A.3). Under condition (A.4),

$$\lim_{\epsilon\to 0} \lim_{\varepsilon\to 0} \sup_{\theta\in\Theta} \left| \sum_{i=1}^{n} \int_{[\varepsilon,1-\varepsilon]^{n-1}} \varphi_{\epsilon,L}(g(u;\theta)) r_i(u;\theta) \prod_{j\neq i} du_j \Big|_{u_i=\varepsilon}^{1-\varepsilon} \right| = 0.$$

With condition (A.2), $g(u;\theta)$ and $det(J_g(u;\theta))$ are bounded in $\Omega_\varepsilon \times \Theta$, By change of variables and Hölder's inequality,

$$\begin{aligned}
&\lim_{\epsilon\to 0} \sup_{\theta\in\Theta} \left| \int_{[\varepsilon,1-\varepsilon]^n} (\varphi_{\epsilon,L}(g(u;\theta)) - \varphi_L(g(u;\theta))) d(u;\theta)\ du \right| \\
=& \lim_{\epsilon\to 0} \sup_{\theta\in\Theta} \left| \int_{\mathcal{S}^\varepsilon} (\varphi_{\epsilon,L}(x) - \varphi_L(x)) |\det(J_g(u;\theta))|\ d(u;\theta)|_{u=g^{-1}(x;\theta)}\ dx \right| \\
\leq& \lim_{\epsilon\to 0} \|\varphi_{\epsilon,L} - \varphi_L\|_p \sup_{\theta\in\Theta} \left| \int_{\mathcal{S}^\varepsilon} |\det(J_g(u;\theta))\ d(u;\theta)|^q_{u=g^{-1}(x;\theta)}\ dx \right|^{1/q} = 0,
\end{aligned}$$

where

$$\mathcal{S}^\varepsilon := \{x \in \mathbb{R}^n :\ x = g(u;\theta),\ u \in [\varepsilon, 1-\varepsilon]^n\}.$$

With condition (A.3), we have

$$\lim_{\varepsilon\to 0} \lim_{\epsilon\to 0} \sup_{\theta\in\Theta} \left| \int_{[\varepsilon,1-\varepsilon]^n} \varphi_{\epsilon,L}(g(u;\theta)) d(u;\theta)\ du - \int_{(0,1)^n} \varphi_L(g(u;\theta)) d(u;\theta)\ du \right| = 0.$$

Therefore,

$$\begin{aligned}
&\frac{\partial}{\partial\theta} \int_{(0,1)^n} \varphi_L(g(u;\theta))\ du \\
=& \lim_{\varepsilon\to 0} \lim_{\epsilon\to 0} \frac{\partial}{\partial\theta} \int_{[\varepsilon,1-\varepsilon]^n} \varphi_L(g(u;\theta))\ du = \int_{(0,1)^n} \varphi_L(g(u;\theta))\ d(u;\theta)\ du.
\end{aligned}$$

With condition (A.3),

$$\sup_{\theta\in\Theta}\left|\int_{(0,1)^n}(\varphi_L(g(u;\theta))-\varphi(g(u;\theta)))\ d(u;\theta)\ du\right|=0.$$

Then we have

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int_{(0,1)^n}\varphi_L(g(u;\theta))\ du\\ =&\lim_{L\to\infty}\frac{\partial}{\partial\theta}\int_{(0,1)^n}\varphi_L(g(u;\theta))\ du=\int_{(0,1)^n}\varphi(g(u;\theta))\ d(u;\theta)\ du,\end{aligned}$$

which proves the theorem. □

Proof of Proposition 1

Proof. Define

$$\tilde{\chi}_\epsilon(x):=\begin{cases}1 & x<-\epsilon,\\ 1-\dfrac{(x+\epsilon)}{2\epsilon} & -\epsilon\le x\le\epsilon,\\ 0 & x>\epsilon,\end{cases}$$

and

$$\chi_\epsilon(x):=\frac{1}{\epsilon}\int_{\mathbb{R}}\tilde{\varphi}_\epsilon(x-y)\phi(y/\epsilon)dz,$$

where $\phi(\cdot)$ is the density of the standard normal distribution. By construction, $\chi_\epsilon(\cdot)$ is smooth. From the condition of the proposition, $g_i(u_i;\theta)$ is strictly monotone with respect to u_i on $[\varepsilon,1-\varepsilon]$. Without loss of generality, we assume $g_i(u_i;\theta)$ is strictly increasing with respect to u_i on $[\varepsilon,1-\varepsilon]$. Then we have

$$\begin{aligned}&\int_0^1\left|\chi_\epsilon(g_i(u_i;\theta))-\mathbf{1}\{g_i(u_i;\theta)\le 0\}\right|^p\ du_i\\ \le&\frac{1}{\epsilon}\int_0^1\left(\int_{\mathbb{R}}\mathbf{1}\{\min(y-\epsilon,0)\le g_i(u_i;\theta)\le\max(y+\epsilon,0)\}\phi(y/\epsilon)dy\right)^p\ du_i\\ \le&\frac{1}{\epsilon}\int_0^1\int_{\mathbb{R}}\big(\mathbf{1}\{\min(y-\epsilon,0)\le g_i(u_i;\theta)\le\max(y+\epsilon,0)\}\big)^p\phi(y/\epsilon)dy\ du_i\\ \le&\frac{1}{\epsilon}\int_{\mathbb{R}\setminus[-\sqrt{\epsilon},\sqrt{\epsilon}]}\phi(y/\epsilon)dy+\int_0^1\mathbf{1}\{-\epsilon\le g_i(u_i;\theta)\le\epsilon+\sqrt{\epsilon}\}\ du_i\\ =&\frac{1}{\epsilon}\int_{\mathbb{R}\setminus[-\sqrt{\epsilon},\sqrt{\epsilon}]}\phi(y/\epsilon)dy+g_i^{-1}(\epsilon+\sqrt{\epsilon};\theta)-g_i^{-1}(-\epsilon;\theta),\end{aligned}$$

where the second inequality holds by applying Jensen's inequality. We have

$$g_i^{-1}(\epsilon+\sqrt{\epsilon};\theta)-g_i^{-1}(-\epsilon;\theta)=(2\epsilon+\sqrt{\epsilon})\left(\frac{\partial g_i(u_i;\theta)}{\partial u_i}\right)^{-1}\Bigg|_{u_i=g_i^{-1}(\xi_\epsilon;\theta)},$$

where $\xi_\epsilon\in(-\epsilon,\epsilon+\sqrt{\epsilon})$. Since $\inf_{\theta\in\Theta,u_i\in(0,1)\setminus[\varepsilon,1-\varepsilon]}|g_i(u_i;\theta)|>0$, $g^{-1}(\xi_\epsilon;\theta)\in[\varepsilon,1-\varepsilon]$ for a sufficiently small ϵ. Therefore,

$$\begin{aligned}
&\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}\int_0^1\left|\varphi_\epsilon(g(u;\theta)-z)-\mathbf{1}\{g(u;\theta)\le z\}\right|^p\,du_i\\
\le&\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}(2\epsilon+\sqrt{\epsilon})\left|\frac{\partial g_i(u_i;\theta)}{\partial u_i}\right|^{-1}_{u_i=g^{-1}(\xi_\epsilon;\theta)}\\
\le&\lim_{\epsilon\to 0}(2\epsilon+\sqrt{\epsilon})\left(\inf_{\theta\in\Theta,u_i\in[\varepsilon,1-\varepsilon]}\left|\frac{\partial g_i(u_i;\theta)}{\partial u_i}\right|\right)^{-1}=0.
\end{aligned}$$

From $\inf_{\theta\in\Theta,u_i\in(0,1)\setminus[\varepsilon,1-\varepsilon]}|g_i(u_i;\theta)|>0$, for $u_i\in(0,1)\setminus[\varepsilon,1-\varepsilon]$,

$$\lim_{\epsilon\to 0}\sup_{\theta\in\Theta}\left|\chi_\epsilon(g_i(u_i;\theta))-\mathbf{1}\{g_i(u_i;\theta)\le 0\}\right|=\lim_{\epsilon\to 0}\mathbf{1}\{|g_i(u_i;\theta)|\le\epsilon\}=0.$$

In addition,

$$\begin{aligned}
&\int_0^1\cdots\int_0^1\left|\prod_{j=1}^n\chi_\epsilon(g_j(u_j;\theta))-\prod_{j=1}^n\mathbf{1}\{g_j(u_j;\theta)\le 0\}\right|^p\,du_1\cdots du_n\\
=&\int_{(0,1)^n}\left|\sum_{i=1}^n\left(\prod_{j=1}^i\chi_\epsilon(g_j(u_j;\theta))\prod_{j=i+1}^n\mathbf{1}\{g_j(u_j;\theta)\le 0\}-\prod_{j=1}^{i-1}\chi_\epsilon(g_j(u_j;\theta))\prod_{j=i}^n\mathbf{1}\{g_j(u_j;\theta)\le 0\}\right)\right|^p du\\
\le&n^{p-1}\sum_{i=1}^n\int_{(0,1)^n}\left|\prod_{j=1}^i\chi_\epsilon(g_j(u_j;\theta))\prod_{j=i+1}^n\mathbf{1}\{g_j(u_j;\theta)\le 0\}-\prod_{j=1}^{i-1}\chi_\epsilon(g_j(u_j;\theta))\prod_{j=i}^n\mathbf{1}\{g_j(u_j;\theta)\le 0\}\right|du\\
\le&n^{p-1}\sum_{i=1}^n\int_{(0,1)}\left|\chi_\epsilon(g_i(u_i;\theta))-\mathbf{1}\{g_i(u_i;\theta)\le 0\}\right|du_i,
\end{aligned}$$

where the first inequality holds by applying Jensen's inequality. Then the rest of the proof is straightforward. □

GLR for the Ali-Mikhail-Haq Copula.

The generator function for this copula is

$$\psi_\theta(x)=\frac{1-\theta}{e^x-\theta},\qquad\theta\in[0,1).$$

Then, we have

$$\frac{\partial\psi_\theta(x)}{\partial\theta}=-\frac{1}{e^x-\theta}+\frac{1-\theta}{(e^x-\theta)^2},\quad\frac{\partial\psi'_\theta(x)}{\partial\theta}=\frac{e^x}{(e^x-\theta)^2}-\frac{2e^x(1-\theta)}{(e^x-\theta)^3},$$

and

$$\psi_\theta'(x) = -\frac{e^x(1-\theta)}{(e^x-\theta)^2}, \quad \psi_\theta''(x) = \frac{2e^{2x}(1-\theta)}{(e^x-\theta)^3} - \frac{e^x(1-\theta)}{(e^x-\theta)^2}.$$

Y_θ has a geometric distribution with parameter θ, with probability mass function

$$(1-\theta)\theta^{y-1} \quad \text{for } y = 1, 2, \ldots,$$

so the LR term is

$$\frac{\partial \log f_Y(y;\theta)}{\partial \theta} = -\frac{1}{1-\theta} + \frac{y-1}{\theta}.$$

The weight functions in the GLR estimator are

$$r_i(u;\theta,y) = -u_i y \left(\frac{1-u_i^{\frac{1}{y}}}{1-\theta} \right),$$

and

$$d(u;\theta,y) = \sum_{i=1}^{n} \left[\frac{1-u_i^{\frac{1}{y}}}{1-\theta} \left(y - \frac{u_i^{-\frac{1}{y}}+\theta}{u_i^{-\frac{1}{y}}-\theta} \right) - \frac{2}{u_i^{-\frac{1}{y}}-\theta} + \frac{1}{1-\theta} \right].$$

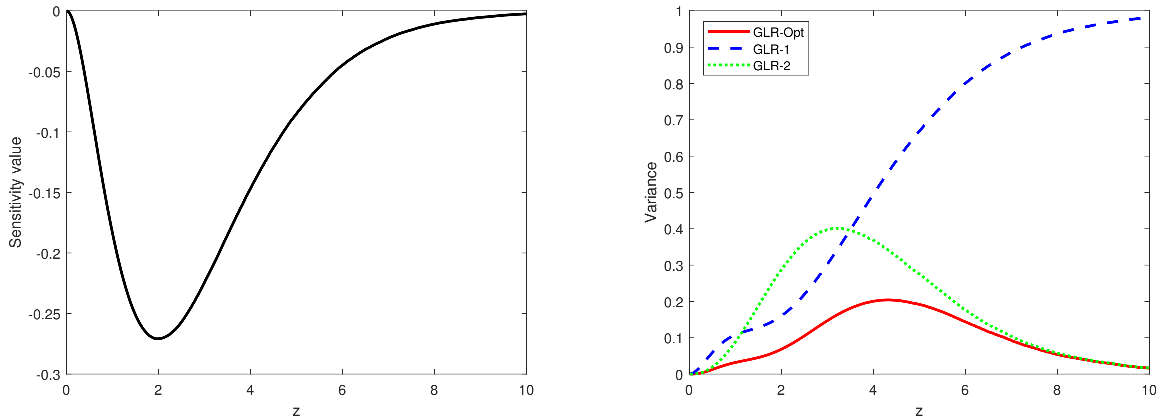
We have

$$\lim_{u_i \to 0^+} r_i(u;\theta,y) = \lim_{u_i \to 1^-} r_i(u;\theta,y) = 0.$$

Appendix B: Supplements of Section 5

Distribution Sensitivities for a Linear Model

For the numerical experiments in Section 5.1 of the main body of the paper, we estimate the sensitivity of $\mathbb{E}[\varphi(g(U;\theta))]$ with respect to θ at z, as a function of z from $z = 0.01$ to 10 with a step size of 0.01. We perform 10^6 independent simulation runs with GLR. Figure 1(a) shows the curves of the distribution sensitivities with respect to θ, and Figure 1(b) presents the sample variance curves of three distribution sensitivity estimators: $G_{1,1}(\cdot)$ (GLR-1), $G_{1,2}(\cdot)$ (GLR-2), and a combined GLR estimator minimizing variance in a family of linear combinations of GLR-1 and GLR-2 given by (GLR-Opt).



(a) Distribution sensitivity curve. (b) Variance curves.

Figure 1 Distribution sensitivities with respect to θ estimated by GLR in the example of Section 5.1 in the main body of the paper.

Figure 2 presents the variance curves for two distribution sensitivities estimated by FDC using a batch of 10^6 independent simulation replications. FDC(δ) denotes the FDC with perturbation size δ. Comparing Figure 2 with Figure 1(b) in the main body of the paper and Figure 1(b), we can see the variance of FDC is much larger than those of the GLR estimators, and the variance of FDC(0.01) is about 10 times of that of FDC(0.1), which indicates that FDC suffers from a bias-variance tradeoff issue.

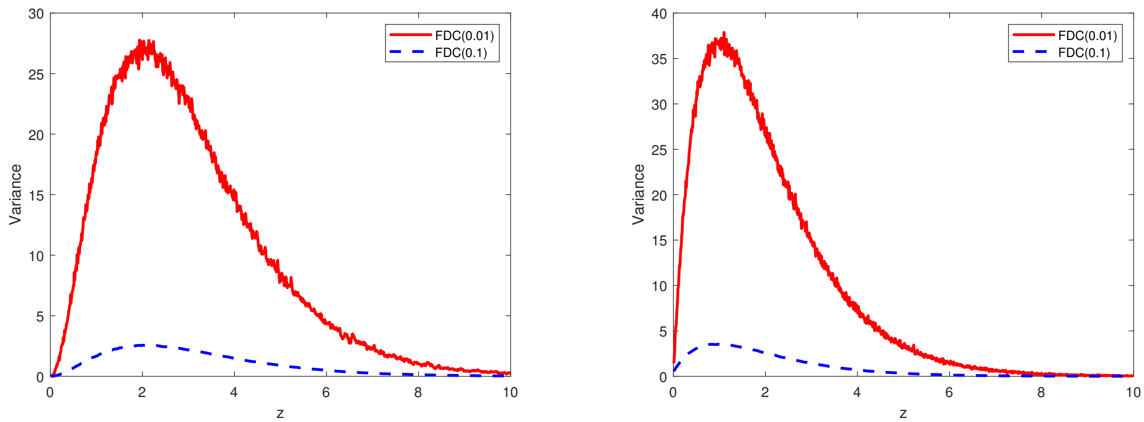
Distribution Sensitivities for a Stochastic Activity Network

An alternative GLR estimator different from that in the main body of the paper can be derived by conditioning on (U_1, U_5, U_6) and treating (U_4, U_2, U_3) as the input uniform random numbers for deriving GLR. The Jacobian matrix is

$$J_g(u;\theta,z) = -\begin{pmatrix} \frac{1}{\lambda_4 u_4} & 0 & 0 \\ 0 & \frac{1}{\lambda_2 u_2} & 0 \\ 0 & 0 & \frac{1}{\lambda_3 u_3} \end{pmatrix}.$$

Then we have

$$r_1(u;\theta,z) = \lambda_4 u_4, \quad r_2(u;\theta,z) = \lambda_2 u_2, \quad r_3(u;\theta,z) = \lambda_3 u_3,$$



(a) Variance curves for distribution sensitivity with respect to θ.

(b) Variance curves for density.

Figure 2 Variance curves for distribution sensitivities estimated by FDC in the example of Section 5.1 of the main body the paper.

and

$$d(u;\theta,z) = -\lambda_2 - \lambda_3 - \lambda_4.$$

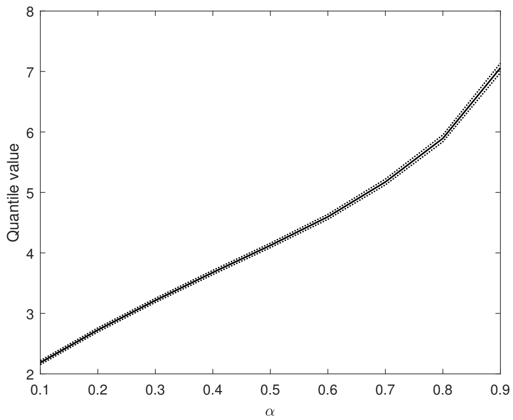
This alternative estimator leads to a comparable performance to the one in Section 4.2.

The setting of the problem is the same as that in Section 5.2 of the main body of the paper. Similar to Figure 2 in the example in Section 5.1 of the main body of the paper, Figure 3 presents confidence intervals using a batch of 10^4 independent simulation replications and the coverage rates of the 0.9-confidence intervals by 10^4 independent macro experiments for quantiles q_α, $\alpha = 0.1, \ldots, 0.9$. Again the coverage ratios of the 0.9-confidence intervals match the target value statistically.

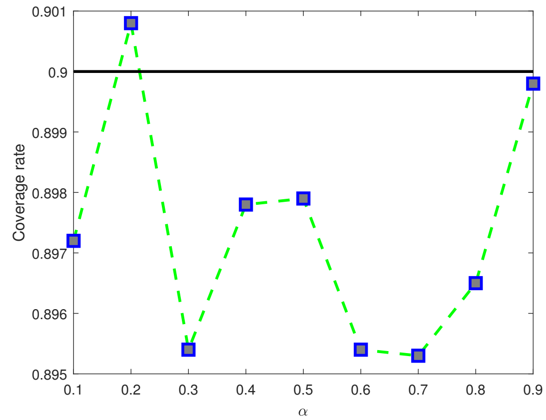
Figure 4 shows that the variances of FDC(0.01) and FDC(0.1) are much larger than those of GLR, and variances of FDC(0.01) are about 10 times of those of FDC(0.1) throughout the curve.

Sensitivities of Credit Risk Derivatives

For CDOs, Table 1 shows the sensitivity estimates for sample sizes $m = 10^4$, $m = 10^5$, and $m = 10^6$ under the Ali-Mikhail-Haq copula, with $\theta = 0.5$. The numerical observations are similar to those in Section 5.4 of the main body of the paper. Again for sample size $m = 10^6$, the estimate with FDC(0.1) falls outside of the 90% confidence interval of the GLR estimate, whereas the estimate



(a) Confidence intervals for quantiles.



(b) Coverage rates of confidence intervals.

Figure 3 90%-confidence intervals estimated by GLR for quantiles and coverage rates in example of Section 5.2 of the main body of the paper.

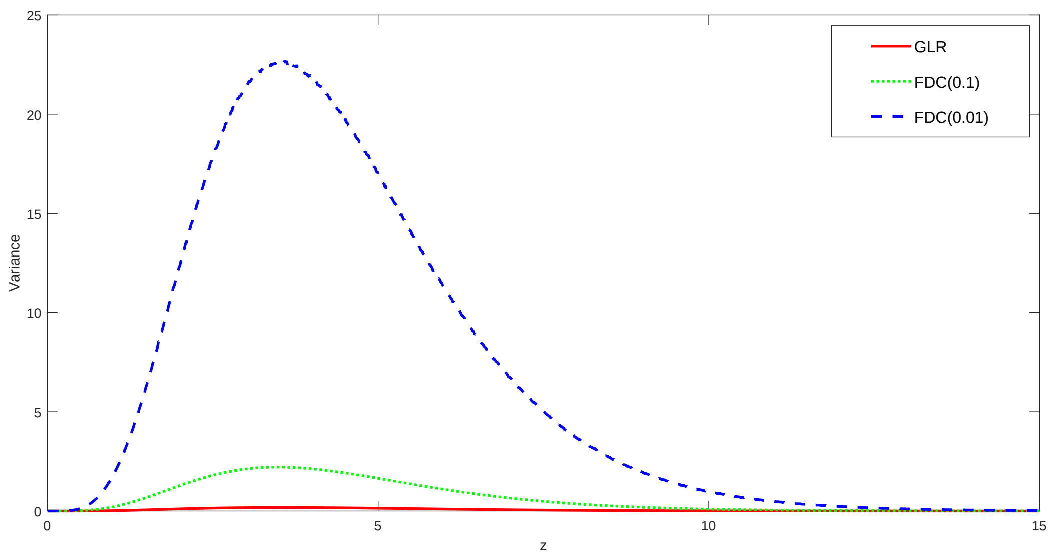


Figure 4 Variance curves for density estimates by FDC in the example of Section 5.2 of the main body of the paper.

with GLR lies in the 90% confidence interval of the FDC(0.01) estimate with sample size $m = 10^7$, which is -0.193 ± 0.002.

For BDS, we estimate the sensitivities of the expectation of the discounted payment to the fifth default of 10 assets made by the seller $V_{\text{value}}(\tau) = L_{(5)} \exp(-r\tau_{(5)})\mathbf{1}\{\tau_{(5)} < T\}$. The marginal

	$m = 10^4$	$m = 10^5$	$m = 10^6$
GLR	-0.188 ± 0.007	-0.189 ± 0.004	-0.191 ± 0.002
FDC(0.1)	-0.194 ± 0.006	-0.195 ± 0.02	-0.195 ± 0.002
FDC(0.01)	-0.110 ± 0.05	-0.202 ± 0.01	-0.192 ± 0.005

Table 1 Sensitivity estimates of CDO with 10 assets governed by the Ali-Mikhail-Haq copula with $\theta = 0.5$ based on 10^2 experiments (mean ± standard error).

	$n = 10^4$	$n = 10^5$	$n = 10^6$
GLR	0.034 ± 0.003	0.035 ± 0.001	0.035 ± 0.0007
FDC(0.1)	0.033 ± 0.004	0.031 ± 0.001	0.031 ± 0.0007
FDC(0.01)	-0.002 ± 0.03	0.021 ± 0.009	0.031 ± 0.004

Table 2 Sensitivity estimates of BDS with 10 assets governed by the Clayton copula with $\theta = 0.5$ based on 10^2 experiments (mean ± standard error).

distributions of the defaults and the parameters are set in the same way as those in Section 5.1 in the main body of the paper. Tables 2 and 3 show the respective Clayton copula and Ali-Mikhail-Haq copula sensitivity estimates with $\theta = 0.5$ for sample sizes $m = 10^4$, $m = 10^5$, and $m = 10^6$. We also implement FDC(0.01) with sample size $m = 10^7$, which leads to 0.034 ± 0.001 for the Clayton copula and 0.051 ± 0.002 for the Ali-Mikhail-Haq copula. The results are similar to those for estimating CDO sensitivities.

References

Peng, Yijie, Michael C. Fu, Jian-Qiang Hu, and Bernd Heidergott, A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. *Operations Research*, 66(2), 487–499, 2018.

	$n=10^4$	$n=10^5$	$n=10^6$
GLR	0.051 ± 0.002	0.050 ± 0.001	0.051 ± 0.0009
FDC(0.1)	0.048 ± 0.003	0.052 ± 0.001	0.051 ± 0.001
FDC(0.01)	0.074 ± 0.03	0.039 ± 0.01	0.049 ± 0.003

Table 3 Sensitivity estimates of BDS with 10 assets governed by the Ali-Mikhail-Haq copula with $\theta = 0.5$ based on 10^2 experiments (mean $\pm$ standard error).