

## STOCHASTIC KRIGING FOR CONDITIONAL VALUE-AT-RISK AND ITS SENSITIVITIES

Xi Chen  
Barry L. Nelson

Kyoung-Kuk Kim

Dept. of Industrial Engineering &  
Management Sciences  
Northwestern University  
Evanston, IL 60208, U.S.A.

Dept. of Industrial & Systems Engineering  
KAIST  
Daejeon 305-701, SOUTH KOREA

### ABSTRACT

Measuring risks in asset portfolios has been one of the central topics in the financial industry. Since the introduction of coherent risk measures, studies on risk measurement have flourished and measures beyond value-at-risk, such as expected shortfall, have been adopted by academics and practitioners. However, the complexity of financial products makes it very difficult and time consuming to perform the numerical tasks necessary to compute these risk measures. In this paper, we introduce a stochastic kriging metamodel-based method for efficient estimation of risks and their sensitivities. In particular, this method uses gradient estimators of assets in a portfolio and gives the best linear unbiased predictor of the risk sensitivities with minimum mean squared error. Numerical comparisons of the proposed method with two other stochastic kriging based approaches demonstrate the promising role that the proposed method can play in the estimation of financial risk.

### 1 INTRODUCTION

The importance of quantification and estimation of risks in the financial industry cannot be exaggerated. From daily operations at banks to regulatory oversight, valid and efficient risk measurement practices are required in today's volatile markets. Since the publication of value-at-risk (VaR) methodology by J.P. Morgan in 1994, adequate risk measures have been the focus of academic and practical debates. Artzner et al. (1999) proposed the now widely known concept of coherent risk measures, and many others popularized related concepts and built theoretical foundations. There also has been a stream of research that focuses on computational aspects of risk measures. In particular, much research has dealt with the so-called conditional value-at-risk (CVaR) or expected shortfall (ES) of asset portfolios as well as VaR in the past decade.

Often, portfolio profits or losses are computed via Monte Carlo simulation when dimensions are high and analytic solutions are not available. Glasserman et al. (2000) is one among many papers that study efficient procedures for simulation based estimation of VaR, where the authors approximate the portfolio loss via the delta-gamma method. However, when we deal with portfolios of complex derivatives, sometimes we have to resort to two-level simulation, in which we estimate the random loss based on inner-level simulations under each outer simulation path. We refer the reader to Gordy and Juneja (2010) for details of this nested simulation scheme. In the paper, the authors apply their proposed method to estimate the probability of a large loss, VaR, and CVaR. Related works followed and they include Lan et al. (2010), Broadie et al. (2011a), and Broadie et al. (2011b), for example.

In practice, we are not only interested in estimating risk measures but also in estimating their sensitivities with respect to parameters of interest. In the literature, the knowledge of sensitivities is argued to be useful in conducting marginal analysis of portfolios, computing optimal portfolios under VaR constraint

(Gourieroux et al. 2000) or in reducing computational time to re-evaluate ES when the portfolio weight is moved slightly (Scaillet 2004). It is also helpful to assess the current portfolio position and to make re-allocation decisions. We note that in the Monte Carlo simulation context the sensitivity estimation of CVaR is studied in Hong and Liu (2009).

In this paper, we are interested in obtaining global approximations to risk measures and their sensitivities. Due to computational workloads coming from the complexity of derivatives products, it would not be practically plausible to conduct an estimation procedure for too many scenarios. We thus propose to use stochastic kriging based metamodeling techniques, in which we build a metamodel based on estimates at only a small number of design points. In addition, we utilize the information of gradients when (pathwise) gradient estimators are available. More details on stochastic kriging can be found in Ankenman et al. (2010) and Chen et al. (2011). We note that an application of stochastic kriging to the estimation of CVaR is also done in Liu and Staum (2010). Our work differs from theirs in that we build metamodels to predict CVaR and its sensitivities simultaneously, and in that we apply stochastic kriging with gradient estimators as formulated in Chen et al. (2011).

The rest of the paper is organized as follows. In Section 2, we briefly introduce VaR and CVaR. Then, the proposed metamodeling techniques are explained in Section 3, and numerical tests with three examples are reported in Section 4. Section 5 concludes.

## 2 RISK MEASUREMENT

This section presents a very brief introduction to popular risk measures VaR and CVaR. These measures have been widely studied, and there is a large literature available. The material here is mostly based on works like Acerbi and Tasche (2002) and Hong and Liu (2009). In this section,  $X$  is the random variable that represents the random profit of a financial portfolio (hence,  $-X$  is the random loss), and we assume that  $X$  is integrable in this paper. When  $X$  has a dependence on a parameter vector  $\theta$ , we write  $X(\theta)$ . For a random variable  $X$ , quantiles of  $X$  are defined as follows.

**Definition 1** (Acerbi and Tasche 2002) The *lower  $p$ -quantile* of  $X$  is  $\inf\{x : \mathbb{P}(X \leq x) \geq p\} = \sup\{x : \mathbb{P}(X \leq x) < p\}$  and denoted by  $q_p(X)$ . Similarly, the *upper  $p$ -quantile* of  $X$  is  $\inf\{x : \mathbb{P}(X \leq x) > p\} = \sup\{x : \mathbb{P}(X \leq x) \leq p\}$  and denoted by  $q^p(X)$ .

Then, the value-at-risk at level  $\alpha$ , say  $\alpha$ -VaR, is defined as the lower  $\alpha$ -quantile of the random loss  $-X$ . Mathematically,  $\alpha$ -VaR =  $q_\alpha(-X) = \inf\{x : \mathbb{P}(X + x \geq 0) \geq \alpha\}$ . Typically,  $\alpha$  is a value close to 1 such as 95% or 99%. From this, VaR can be interpreted as the minimal cash amount such that the portfolio value  $X + x$  does not become negative with probability at least  $\alpha$ . Instead of a single VaR, we can consider the average of VaR values in the right tail of the random loss  $-X$ , i.e.,  $(1 - \alpha)^{-1} \int_\alpha^1 \beta$ -VaR  $d\beta$ , and this is the conditional value-at-risk at level  $\alpha$ . See, e.g., Hong and Liu (2009).

It is well known that

$$\alpha\text{-CVaR} = \frac{1}{1 - \alpha} \mathbb{E} [(-X - t)^+] + t = \frac{1}{1 - \alpha} \left\{ \mathbb{E} [-X \mathbf{1}\{-X \geq t\}] + t(1 - \alpha - \mathbb{P}(-X \geq t)) \right\} \quad (1)$$

for any  $t \in [q_\alpha(-X), q^\alpha(-X)]$ . Furthermore, it can be shown that if  $-X(\mu)$  (with parameter  $\mu$ ) does not have an atom at  $q_\alpha(-X)$ , then  $\mathbb{P}(-X \geq q_\alpha(-X)) = 1 - \alpha$  and thus

$$\alpha\text{-CVaR}(\mu) = \mathbb{E} \left[ -X(\mu) \mid -X(\mu) \geq \alpha\text{-VaR}(\mu) \right]. \quad (2)$$

We further assume that  $X(\mu)$  is Lipschitz continuous for almost all sample point  $\omega$  with Lipschitz constant  $K(\omega)$  where  $K$  is an integrable random variable, and that  $\alpha$ -VaR is differentiable with respect to  $\mu$ . Then, Hong and Liu (2009) show that the derivative of  $\alpha$ -CVaR( $\mu$ ) with respect to  $\mu$  can be given as

$$\alpha\text{-CVaR}'(\mu) = \mathbb{E} \left[ -X'(\mu) \mid -X(\mu) \geq \alpha\text{-VaR}(\mu) \right], \quad (3)$$

where the prime is used to denote taking a derivative with respect to  $\mu$ .

Throughout this paper, we implicitly assume that the conditions above are always satisfied so that (2) and (3) are valid. Our objective is to obtain global approximations to  $\alpha$ -CVaR( $\mu$ ) and  $\alpha$ -CVaR'( $\mu$ ) over the parameter space  $\Omega_\mu$ , based on simulation observations. Even though it is straightforward to get consistent estimators for CVaR and its sensitivity at a single point  $\mu$ , it may be quite time consuming and inaccurate to perform such a task globally. Hence, we propose to apply stochastic kriging with gradient information (if available), which is described in the following section.

### 3 STOCHASTIC KRIGING

#### 3.1 Stochastic Kriging with Gradient Estimator

Kriging is an interpolation-based metamodel that uses a finite number of observations and this number is often small due to associated costs in applications. At design points  $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^d$ , the response surface is assumed to be a realization of the following stochastic process

$$Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x})$$

where  $\mathbf{f}$  is a  $\mathbb{R}^p$ -valued function,  $\boldsymbol{\beta}$  is a  $p$ -dimensional vector, and  $M$  is a mean zero stationary Gaussian random field such that  $\mathbb{E}[|M(\mathbf{x})|^2] < \infty$  for all  $\mathbf{x} \in \mathbb{R}^d$ . In addition, the correlation structure for  $M$  is given by

$$\text{Cov}[M(\mathbf{x}), M(\mathbf{y})] = \tau^2 \mathcal{R}(\mathbf{x}, \mathbf{y}), \quad \mathcal{R}(\mathbf{x}, \mathbf{y}) = \exp\left(-\sum_{r=1}^d \theta_r |x_r - y_r|^2\right).$$

Here,  $\mathbf{x} = (x_1, \dots, x_d)^\top$  and the  $\theta_r$ 's are fixed parameters in  $\mathbb{R}_+$ . Note that  $\text{Var}[M(\mathbf{x})] = \tau^2$ . Therefore, the variance of this random field remains the same at all points, but the correlation between  $M(\mathbf{x})$  and  $M(\mathbf{y})$  decreases as the distance between two points increases. Furthermore, we assume the process above has partial derivatives in the sense that

$$D^r(\mathbf{x}) := \left(\frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_r}\right)^\top \boldsymbol{\beta} + \frac{\partial M(\mathbf{x})}{\partial x_r}, \quad \frac{\partial M(\mathbf{x})}{\partial x_r} = \lim_{t \rightarrow 0} t^{-1} (M(\mathbf{x} + t\mathbf{e}_r) - M(\mathbf{x}))$$

where the limit is taken in the mean-square sense and  $\mathbf{e}_r$  is the  $r$ th canonical basis vector in  $\mathbb{R}^d$ . The limit in the mean-square sense is useful for applications. Sufficient conditions for the existence of the mean-square derivative of  $M(\mathbf{x})$  (hence  $Y(\mathbf{x})$ ) are fully characterized by conditions on  $\mathcal{R}(\cdot, \cdot)$ , thereby allowing the modeling of the dependence of the derivative processes on the original response process. See Parzen (1962) for more. Then,  $D^r(\mathbf{x})$  is again Gaussian and its correlation structure is given by

$$\text{Cov}[D^r(\mathbf{x}), D^s(\mathbf{y})] = \tau^2 \frac{\partial^2 \mathcal{R}(\mathbf{x}, \mathbf{y})}{\partial x_r \partial y_s}, \quad \text{Cov}[D^r(\mathbf{x}), Y(\mathbf{y})] = \tau^2 \frac{\partial \mathcal{R}(\mathbf{x}, \mathbf{y})}{\partial x_r}.$$

Consequently, the  $k$  observations at design points together with total  $kd$  partial derivatives have a  $k(1+d)$ -dimensional multivariate normal distribution. To simplify notation, we write  $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_k))^\top$ ,  $\mathbf{D}^r = (D^r(\mathbf{x}_1), \dots, D^r(\mathbf{x}_k))^\top$ , and  $\mathbf{Y}_+ = \left(\mathbf{Y}^\top, (D^1)^\top, \dots, (D^d)^\top\right)^\top$ . We also denote the mean and the covariance matrix of  $\mathbf{Y}_+$  by  $\mathbf{F}_+ \boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}_{\mathbf{M}_+}$ , where  $\mathbf{F}_+$  is the  $k(d+1) \times p$  matrix of functions that has  $\mathbf{f}(\mathbf{x}_1)^\top, \dots, \mathbf{f}(\mathbf{x}_k)^\top, (\partial \mathbf{f}(\mathbf{x}_1)/\partial x_1)^\top, \dots, (\partial \mathbf{f}(\mathbf{x}_k)/\partial x_1)^\top, \dots, (\partial \mathbf{f}(\mathbf{x}_1)/\partial x_d)^\top, \dots, (\partial \mathbf{f}(\mathbf{x}_k)/\partial x_d)^\top$  as its rows.

In the stochastic simulation context, however, our observations are simulation results that contain simulation errors. Suppose that we make  $n_i$  independent simulation replications at each design point  $\mathbf{x}_i$  which can be different at distinct design points. Then, we adapt the metamodel above to a stochastic simulation context:

$$\mathcal{Y}_j(\mathbf{x}_i) = Y(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i), \quad \mathcal{D}_j^r(\mathbf{x}_i) = D^r(\mathbf{x}_i) + \zeta_j^r(\mathbf{x}_i), \quad j = 1, \dots, n_i, \quad i = 1, \dots, k, \quad r = 1, \dots, d.$$

Here,  $\varepsilon_j(\mathbf{x}_i)$ 's are the i.i.d. mean zero simulation noise from  $j$ -th run at the  $i$ -th design point. The  $\zeta_j^r(\mathbf{x}_i)$ 's are similarly defined. We run simulations independently across replications and design points. However,  $\varepsilon_j(\mathbf{x}_i)$  and  $\zeta_j^r(\mathbf{x}_i)$  for the same replication at the same design point might be correlated. Hence, the correlation structure of simulation noise is assumed to be

$$\text{Var}[\varepsilon_j(\mathbf{x}_i)] = \sigma_{j0}^2, \quad \text{Var}[\zeta_j^r(\mathbf{x}_i)] = \sigma_{ir}^2, \quad \text{Corr}[\varepsilon_j(\mathbf{x}_i), \zeta_j^r(\mathbf{x}_i)] = \rho_i^{(0,r)}, \quad \text{Corr}[\zeta_j^r(\mathbf{x}_i), \zeta_j^s(\mathbf{x}_i)] = \rho_i^{(r,s)}.$$

The last term is defined only when  $r \neq s$ . We denote the sample averages of  $\mathcal{Y}_j(\mathbf{x}_i)$  and  $\mathcal{D}_j^r(\mathbf{x}_i)$  by  $\bar{\mathcal{Y}}(\mathbf{x}_i)$  and  $\bar{\mathcal{D}}^r(\mathbf{x}_i)$ . In the same manner, we define  $\bar{\varepsilon}(\mathbf{x}_i)$  and  $\bar{\zeta}^r(\mathbf{x}_i)$ . Then, again we have a  $k(1+d)$ -dimensional multivariate normal distribution for  $\bar{\mathcal{Y}}_+ = (\bar{\mathcal{Y}}(\mathbf{x}_1), \dots, \bar{\mathcal{Y}}(\mathbf{x}_k), \bar{\mathcal{D}}^1(\mathbf{x}_1), \bar{\mathcal{D}}^1(\mathbf{x}_2), \dots, \bar{\mathcal{D}}^d(\mathbf{x}_k))^\top$ . Its mean is the same as  $\mathbf{F}_+\beta$ , but its covariance matrix has the form  $\Sigma_+ := \Sigma_{M_+} + \Sigma_{\mathcal{E}_+}$  where the second term is the covariance matrix of  $\bar{\varepsilon}_+ := (\bar{\varepsilon}(\mathbf{x}_1), \dots, \bar{\varepsilon}(\mathbf{x}_k), \bar{\zeta}^1(\mathbf{x}_1), \dots, \bar{\zeta}^d(\mathbf{x}_k))^\top$ .

The objective of this metamodel is to make a prediction of  $Y(\mathbf{x})$  at a prediction point  $\mathbf{x}_0 \in \mathbb{R}^d$ . With given  $\mathbf{f}$ ,  $\beta$ , and covariance matrices, the random vector  $(Y(\mathbf{x}_0), \bar{\mathcal{Y}}_+)$  has a multivariate normal distribution. Then, it is a standard result that

$$\hat{Y}(\mathbf{x}_0) := \mathbb{E}[Y(\mathbf{x}_0) | \bar{\mathcal{Y}}_+] = \mathbf{f}(\mathbf{x}_0)^\top \beta + \Sigma_{M_+}(\mathbf{x}_0, \cdot)^\top \Sigma_+^{-1}(\bar{\mathcal{Y}}_+ - \mathbf{F}_+\beta) \quad (4)$$

where  $\Sigma_{M_+}(\mathbf{x}_0, \cdot)$  represents the covariance matrix between  $Y(\mathbf{x}_0)$  and  $\bar{\mathcal{Y}}_+$  (and thus  $Y_+$  because  $\bar{\varepsilon}_+$  is independent of  $Y(\mathbf{x}_0)$ ). It can be shown that  $\hat{Y}(\mathbf{x}_0)$  is the best linear predictor that minimizes the mean squared error (MSE) among linear predictors. Regarding parameter estimation, we start with  $\Sigma_{\mathcal{E}_+}$  estimated from the simulation replications, then calculate the maximum likelihood estimates for  $\beta$ ,  $\tau^2$ , and  $\theta_r$ . More details about stochastic kriging with gradient estimators can be found in Chen et al. (2011). Notice that response prediction with gradient information has also been studied in other research contexts. See, for instance, Morris et al. (1993) for deterministic computer experiments and Solak et al. (2003) for Gaussian process models for dynamic system identification.

### 3.2 Sensitivity Estimation with Kriging Metamodel

Often, we are not only interested in evaluating an unknown performance function  $Y(\mathbf{x})$  but also in obtaining sensitivities  $D^r(\mathbf{x})$  of such a function. There are well-known methods to obtain sensitivities via simulation such as finite difference scheme, infinitesimal perturbation analysis, or likelihood ratio method. See Chapter VII of Asmussen and Glynn (2007) for example. However, when we want to obtain global approximations to the gradients  $D^r(\mathbf{x})$ , applications of such methods might be infeasible due to high simulation costs. In this subsection, we consider three approaches based on stochastic kriging metamodels.

Suppose that we have (sample averaged) simulation outputs  $\bar{\mathcal{Y}} := (\bar{\mathcal{Y}}(\mathbf{x}_1), \dots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top = Y + \bar{\varepsilon}$  where

$$Y = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_k))^\top, \quad \bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{x}_1), \dots, \bar{\varepsilon}(\mathbf{x}_k))^\top$$

with  $n_i$  independent simulation replications at the  $i$ -th of the  $k$  design points. Further assume that we do not have any derivative information. Then, as presented above, stochastic kriging metamodel based on  $\bar{\mathcal{Y}}$  stipulates that  $(Y(\mathbf{x}_0), \bar{\mathcal{Y}})$  at a prediction point  $\mathbf{x}_0$  has a multivariate normal distribution. If the covariance matrices of  $Y$  and  $\bar{\varepsilon}$  are  $\Sigma_M$  and  $\Sigma_{\mathcal{E}}$ , respectively, then the best linear predictor (denoted by **SK**) is given by

$$\hat{Y}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \beta + \Sigma_M(\mathbf{x}_0, \cdot)^\top (\Sigma_M + \Sigma_{\mathcal{E}})^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\beta) \quad (5)$$

where  $\mathbb{E}[Y] = \mathbf{F}\beta$  and  $\Sigma_M(\mathbf{x}_0, \cdot)$  is the covariance matrix between  $Y(\mathbf{x}_0)$  and  $\bar{\mathcal{Y}}$ . Then, one simple solution for global gradient approximation is to differentiate  $\hat{Y}(\mathbf{x}_0)$  and get

$$\hat{D}^r(\mathbf{x}_0) = \left( \frac{\partial \mathbf{f}(\mathbf{x}_0)}{\partial x_r} \right)^\top \beta + \Sigma_{D^r}(\mathbf{x}_0, \cdot)^\top (\Sigma_M + \Sigma_{\mathcal{E}})^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\beta)$$

where  $\Sigma_{D^r}(\mathbf{x}_0, \cdot) = (\partial/\partial x_r)\Sigma_M(\mathbf{x}_0, \cdot)$ . It is not difficult to see that this partial derivative is nothing but the  $k$ -vector containing covariances between  $D^r(\mathbf{x}_0)$  and  $\mathcal{Y}$ . Similar methods have been studied in other contexts; see, for instance, Sakata et al. (2010) and Pardo-Igúzquiza and Chica-Olmo (2004). We denote this approach by **SK-G**.

A second method to do gradient estimation is to “map” the gradient surface by treating the gradient at a point as the response of interest. This method uses the gradient estimates obtained at the  $k$  design points to construct a stochastic kriging predictor for the gradient at  $\mathbf{x}_0$  by (5). We denote this second approach by **SK4G**.

Lastly, we present our third approach **SKG-G** that builds a metamodel for the gradients based on full information, i.e., response and gradient information. By simply differentiating (4), we obtain

$$\widehat{D}^r(\mathbf{x}_0) = \left( \frac{\partial \mathbf{f}(\mathbf{x}_0)}{\partial x_r} \right)^\top \beta + \Sigma_{D^r_+}(\mathbf{x}_0, \cdot)^\top \Sigma_+^{-1}(\bar{\mathcal{Y}}_+ - \mathbf{F}_+\beta)$$

where  $\Sigma_{D^r_+}(\mathbf{x}_0, \cdot)$  is the covariance matrix between  $D^r(\mathbf{x}_0)$  and  $Y_+$ . In particular,

$$\bar{\mathcal{Y}}_+ = Y_+ + \bar{\epsilon}_+ = \begin{pmatrix} Y \\ D \end{pmatrix} + \begin{pmatrix} \bar{\epsilon} \\ \bar{\zeta} \end{pmatrix} = \begin{pmatrix} \bar{\mathcal{Y}} \\ \bar{\mathcal{D}} \end{pmatrix}, \quad \Sigma_{D^r_+}(\mathbf{x}_0, \cdot) = \begin{pmatrix} \Sigma_{D^r}(\mathbf{x}_0, \cdot) \\ \Sigma'_{D^r}(\mathbf{x}_0, \cdot) \end{pmatrix},$$

and  $\Sigma'_{D^r}(\mathbf{x}_0, \cdot)$  is the  $kd$ -vector that contains covariances between  $D^r(\mathbf{x}_0)$  and  $D$ . We can show that this predictor is actually the best linear predictor of  $D^r(\mathbf{x}_0)$  based on  $\bar{\mathcal{Y}}_+$  among all linear predictors. When parameters are unknown, we estimate them using simulation replication results and maximum likelihood estimation.

## 4 NUMERICAL EXAMPLES

In this section, we provide three examples with increasing difficulty to demonstrate how one can apply stochastic kriging to risk measurement. Specifically, we compare the performances of the three proposed methods introduced in Section 3 in terms of the following two aspects.

1. Predicting risk measure: compare the prediction results by stochastic kriging (**SK**), stochastic kriging with gradient estimators (**SKG**) and by a naive simulation method (**NS**).
2. Sensitivity estimation: compare the estimates by differentiating stochastic kriging metamodel (**SK-G**), stochastic kriging for gradient estimators (**SK4G**), differentiating stochastic kriging metamodel with gradient estimators (**SKG-G**), and a naive simulation method (**NS**).

### 4.1 Call Option on a Single Asset

Our first example is a call option based on the Black-Scholes (BS) model. This example involves neither VaR nor CVaR but the option price  $C(S_0)$  and its first order sensitivity “delta”  $\Delta(S_0)$  where  $S_0$  is the stock price at the inception of the option contract. We choose this example because it is quite well known and  $C(S_0), \Delta(S_0)$  provide important information when managing risks in options.

The stock price dynamics of the BS model is given by the stochastic differential equation (SDE) under the risk-neutral measure,

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where  $r$  is the risk-free rate and  $\sigma$  is the volatility of  $S_t$ . This SDE admits a closed form solution  $S_t = S_0 \exp((r - \sigma^2/2)t + \sigma W_t)$ , and  $S_t$  can be easily simulated using the fact that  $W_t \sim \mathcal{N}(0, t)$ . The European call option on  $S_t$  is a right to buy a stock at option maturity  $T$  and at the pre-specified strike price  $K$ . Mathematically, this payoff discounted at rate  $r$  is expressed as  $P = e^{-rT}(S_T - K)^+$ . The price

Table 1: Parameters for the Black-Scholes European call option model.

$S_0$	$K$	$T$	$r$	$\sigma$
[85,115]	100	1/52 year	3%	40%

and the delta are, then, calculated as follows:

$$C(S_0) = E[P] = E[e^{-rT}(S_T - K)^+],$$

$$\Delta(S_0) = \frac{dC}{dS_0} = E\left[\frac{dP}{dS_T} \frac{dS_T}{dS_0}\right] = E\left[e^{-rT} \left(\frac{S_T}{S_0}\right) \cdot \mathbf{1}\{S_T \geq K\}\right].$$

In the second expression, we skipped the details in justifying the interchange of differentiation and integration. See Broadie and Glasserman (1996) or Fu and Hu (1995). Also, we used  $dS_T/dS_0 = S_T/S_0$ . Note that two functions inside the expectation on the right-hand side give us the price estimator and the gradient estimator. Parameter values used in the experiment are given in Table 1.

*Experiments.* As described above, we compare the performances of **SK** and **SKG** in predicting call option prices  $C(S_0)$  where  $S_0$  ranges over [85, 115]. Regarding  $\Delta(S_0)$ , we have three approaches **SK-G**, **SK4G**, and **SKG-G**. Recall that the second approach is based on the stochastic kriging metamodel with gradient information only. Pretending that little information about the true response surface is available, a constant trend model  $\mathbf{f}(S_0)^\top \boldsymbol{\beta} = \beta_0$  is chosen for stochastic kriging metamodels. For each type of metamodel, we fit the observational data at  $k$  design points using maximum likelihood estimation and estimate the model parameters  $\beta_0, \tau^2$  and  $\theta$ . Then  $\hat{\beta}_0, \hat{\tau}^2$  and  $\hat{\theta}$  are used for prediction at  $N = 193$  equally spaced points in [85, 115]. We select  $k \in \{13, 25\}$  equally spaced design points (including 85 and 115), and adopt an equal number of simulation replications  $n \in \{2000, 4000\}$  at each design point.

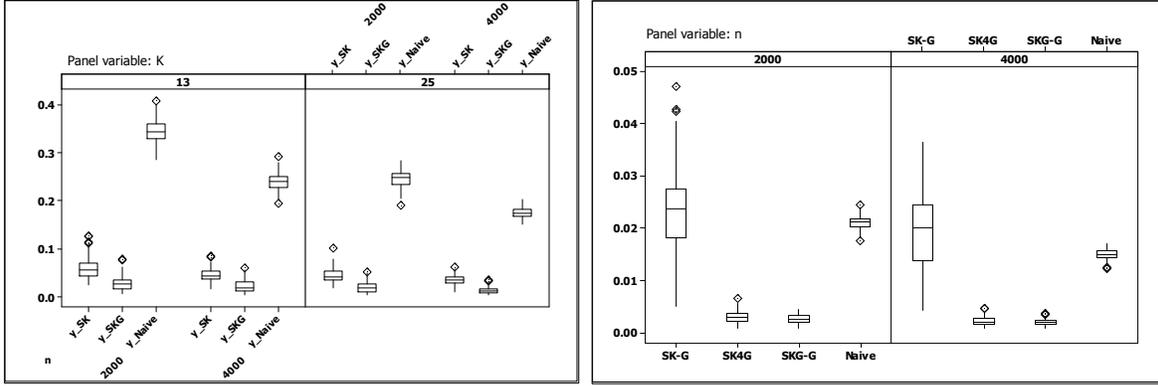
As the closed-form expressions for  $C(S_0)$  and  $\Delta(S_0)$  are available for the BS model, the estimated root mean squared error (ERMSE) over the grid of  $N = 193$  prediction points is proposed to evaluate the prediction performance,

$$\text{ERMSE}(C) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(C(S_0^i) - \hat{C}(S_0^i)\right)^2}, \quad \text{ERMSE}(\Delta) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\Delta(S_0^i) - \hat{\Delta}(S_0^i)\right)^2}, \quad (6)$$

where  $C(S_0^i), \Delta(S_0^i)$  denote the true price and delta at  $S_0^i$ , and  $\hat{C}(S_0^i), \hat{\Delta}(S_0^i)$  represent the predicted values.

One simple way of obtaining a global surface of  $C(S_0), \Delta(S_0)$  is to allocate the simulation budget on all of  $N$  prediction points and get price and delta estimates. We call this naive simulation **NS**. Then, the number of simulation trials at each point becomes  $\bar{n} = \lceil kn/N \rceil$ . This way, all methods have approximately the same computational budget. The experiment is repeated for 100 macro-replications and we summarize the resulting ERMSEs in boxplots which are available in Figures 1(a) and (b).

*Results.* Figure 1(a) shows the prediction performances of **SK**, **SKG** and **NS** in terms of  $\text{ERMSE}(C)$ . The left panel contains the boxplots when 13 design points are used while 25 design points are used for the right panel. In each panel, two groups of boxplots are ordered from left to right, respectively, for  $n = 2000$  and  $n = 4000$ . We can see that **SKG** is superior to **SK** and **NS** in predicting  $C(S_0)$ . Figure 1(b) presents the  $\text{ERMSE}(\Delta)$  by the four approaches for estimating  $\Delta(S_0)$  with 25 design points. As one can easily see, although **SK** gives a pretty good predictor for  $C(S_0)$ , **SK-G** that is obtained by differentiating the **SK**-based metamodel does not appear to be a competent gradient estimation approach. On the other hand, differentiating **SKG**, namely **SKG-G**, gives a reasonably good gradient predictor, whose performance is close to that of **SK4G**. Therefore, we see here that the **SKG** metamodel has a potential in producing good response and sensitivity predictors simultaneously.



(a) Boxplots of ERMSE(C) for 100 macro-replications with  $k = 13$  and 25 (b) Boxplots of ERMSE( $\Delta$ ) for 100 macro-replications with  $k = 25$

Figure 1: Boxplots of ERMSEs for prices and sensitivities in the Black-Scholes model.

#### 4.2 Quadratic Loss of Two Assets

In this subsection, we consider a two-dimensional problem taken from Hong and Liu (2009). Our objective is to construct a CVaR response surface and gradient surfaces of CVaR. Here, we assume that the random loss  $L(\mu)$  is given by  $L(\mu) = a_0 + a^\top \Delta S + \Delta S^\top A \Delta S$  where  $\mu = (\mu_1, \mu_2)^\top$ , and  $a_0 = 0.3$ ,  $a = (0.8, 1.5)^\top$ ,  $A = \begin{pmatrix} 1.2 & 0.6 \\ 0.6 & 1.5 \end{pmatrix}$ . Further, we assume that  $\Delta S$  is a bivariate normal random vector with mean  $\mu$  and variance-covariance matrix  $\Sigma_s = 0.02 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ . Writing  $\alpha$ -VaR,  $\alpha$ -CVaR as  $v_\alpha(\mu)$  and  $c_\alpha(\mu)$  respectively, from Section 2 we have

$$\begin{aligned} c_\alpha(\mu) &= \mathbb{E} \left[ L(\mu) \mid L(\mu) \geq v_\alpha(\mu) \right] = v_\alpha(\mu) + \frac{1}{1-\alpha} \mathbb{E} [L(\mu) - v_\alpha(\mu)]^+, \\ \frac{\partial c_\alpha(\mu)}{\partial \mu_i} &= \mathbb{E} \left[ \frac{\partial L(\mu)}{\partial \mu_i} \mid L(\mu) \geq v_\alpha(\mu) \right], \quad i = 1, 2, \end{aligned}$$

where the second equality in the first equation is obtained by setting  $t = q_\alpha(-X)$  in (1). When the gradient estimator  $D^{(i)} = \partial L / \partial \mu_i = \partial_{\mu_i} L$  exists, Hong and Liu (2009) suggests using the following consistent estimators: given  $n_s$  triples of i.i.d.  $(L_j, D_j^{(1)}, D_j^{(2)})$  for  $(L, \partial_{\mu_1} L, \partial_{\mu_2} L)$ ,

$$\hat{v}_\alpha^{n_s} = L_{[n_s \alpha]:n_s}, \quad \hat{c}_\alpha^{n_s} = \hat{v}_\alpha^{n_s} + \frac{1}{n_s(1-\alpha)} \sum_{j=1}^{n_s} [L_j - \hat{v}_\alpha^{n_s}]^+, \quad \bar{Y}_{n_s}^{(i)} = \frac{1}{n_s(1-\alpha)} \sum_{j=1}^{n_s} D_j^{(i)} \mathbf{1}\{L_j \geq \hat{v}_\alpha^{n_s}\} \quad (7)$$

where  $L_{k:n_s}$  is the  $k$ th order statistic. Clearly,  $D^{(i)} = a^\top \mathbf{e}_i + 2\mathbf{e}_i^\top A \Delta S$  and  $\mathbf{e}_i$  is the  $i$ th canonical basis vector in  $\mathbb{R}^2$ ,  $i = 1, 2$ .

*Experiments.* The experimental design is similar to the one in the previous subsection. Our design space is  $\Omega_\mu = [0.001, 0.1]^2$  from which we select  $k$  design points, i.e., choose  $k$  different  $\mu$  vectors. Then, at each design point, we obtain  $n$  observations of  $v_\alpha$ ,  $c_\alpha$ , and  $\partial_{\mu_i} c_\alpha$ . Note that each estimate involves  $n_s$  simulation trials, yielding  $(L_j, D_j^{(1)}, D_j^{(2)})$ ,  $j = 1, \dots, n_s$ . Given  $n$  observations of  $\hat{v}_{n_s}$ ,  $\hat{c}_{n_s}^\alpha$ , and  $\bar{Y}_{n_s}^{(i)}$  at each design point, we build metamodels for CVaR based on **SK**, **SKG**; and for gradient estimation, **SK-G**, **SK4G**, and **SKG-G** are used. As for design points, we use a ‘‘maxmin’’ Latin-hypercube sample of  $(k-4)$  design points from  $\Omega_\mu$  plus its four corner points with  $k \in \{16, 25\}$ , and we use in total  $N = 1601$  prediction points

Table 2: The mean ERMSEs and the corresponding standard errors for CVaR and its sensitivities over 50 macro-replications using  $n \times n_s = 10^5$  when  $\alpha = 0.99$ .

ERMSE	$k$	$n = 100, n_s = 1000$		$n = 50, n_s = 2000$		$n = 25, n_s = 4000$	
		16	25	16	25	16	25
$\widehat{c}_\alpha$	<b>SK</b>	1.21E-02 (3.25E-04)	1.22E-02 (2.94E-04)	8.10E-03 (2.91E-04)	8.90E-03 (3.22E-04)	8.40E-03 (4.75E-04)	1.02E-02 (3.15E-04)
	<b>SKG</b>	1.53E-02 (2.00E-04)	1.52E-02 (2.02E-04)	7.90E-03 (1.56E-04)	8.40E-03 (1.73E-04)	5.30E-03 (1.99E-04)	5.30E-03 (1.49E-04)
	<b>NS</b>	1.11E-01 (3.35E-05)	1.11E-01 (1.60E-05)	1.11E-01 (3.35E-05)	1.11E-01 (1.60E-05)	1.11E-01 (3.35E-05)	1.11E-01 (1.60E-05)
$\widehat{\partial_{\mu_1} c_\alpha}$	<b>SK-G</b>	1.64E-01 (1.17E-02)	1.50E-01 (9.60E-03)	2.07E-01 (2.03E-02)	2.34E-01 (2.68E-02)	2.74E-01 (3.49E-02)	4.97E-01 (3.51E-02)
	<b>SKG-G</b>	1.89E-01 (2.53E-04)	1.89E-01 (2.49E-04)	9.48E-02 (2.89E-04)	9.48E-02 (2.07E-04)	4.77E-02 (2.43E-04)	4.72E-02 (2.31E-04)
	<b>SK4G</b>	1.89E-01 (2.95E-04)	1.89E-01 (2.77E-04)	9.49E-02 (2.74E-04)	9.49E-02 (2.04E-04)	4.82E-02 (3.54E-04)	4.77E-02 (2.98E-04)
	<b>NS</b>	2.05E-01 (2.48E-04)	9.11E-02 (8.61E-05)	2.05E-01 (2.48E-04)	9.11E-02 (8.61E-05)	2.05E-01 (2.48E-04)	9.11E-02 (8.61E-05)
	<b>SK-G</b>	1.89E-01 (1.20E-02)	1.95E-01 (1.42E-02)	2.36E-01 (2.34E-02)	2.95E-01 (2.95E-02)	3.10E-01 (3.35E-02)	5.68E-01 (4.47E-02)
$\widehat{\partial_{\mu_2} c_\alpha}$	<b>SK-G</b>	2.83E-01 (2.35E-04)	2.83E-01 (2.56E-04)	1.42E-01 (2.55E-04)	1.41E-01 (2.40E-04)	7.05E-02 (3.03E-04)	7.06E-02 (2.45E-04)
	<b>SKG-G</b>	2.83E-01 (2.83E-04)	2.83E-01 (2.96E-04)	1.42E-01 (3.28E-04)	1.42E-01 (2.83E-04)	7.09E-02 (3.59E-04)	7.06E-02 (3.43E-04)
	<b>SK4G</b>	2.83E-01 (2.98E-04)	2.83E-01 (2.96E-04)	1.42E-01 (3.28E-04)	1.42E-01 (2.83E-04)	7.09E-02 (3.59E-04)	7.06E-02 (3.43E-04)
	<b>NS</b>	2.98E-01 (2.55E-04)	1.15E-01 (1.21E-04)	2.98E-01 (2.55E-04)	1.15E-01 (1.21E-04)	2.98E-01 (2.55E-04)	1.15E-01 (1.21E-04)
	<b>SK-G</b>	1.89E-01 (1.20E-02)	1.95E-01 (1.42E-02)	2.36E-01 (2.34E-02)	2.95E-01 (2.95E-02)	3.10E-01 (3.35E-02)	5.68E-01 (4.47E-02)

in  $\Omega_\mu$ . Actually, 1600 points are regularly spaced and we add the testing point in Hong and Liu (2009) for a sanity check.

The performance measures ERMSEs for predicting CVaR and its gradients are defined in a similar fashion as in (6). Since closed-form expressions of  $c_\alpha$  and  $\partial_{\mu_i} c_\alpha$  are not available, the true values at each of  $N$  points are approximated using simulations with a sample size  $n_s = 10^6$ . Experiments for  $\alpha \in \{0.95, 0.99\}$  are conducted, and only the results for  $\alpha = 0.99$  are presented below for the sake of brevity. We include ERMSEs from the naive simulation scheme NS for comparison. Recall that this method allocates the total simulation budget  $k \times n \times n_s$  equally over  $N$  prediction points, i.e.,  $\bar{n} = \lceil knn_s/N \rceil$ , and obtains global approximations to  $c_\alpha$  and  $\partial_{\mu_i} c_\alpha, i = 1, 2$  in  $\Omega_\mu$ .

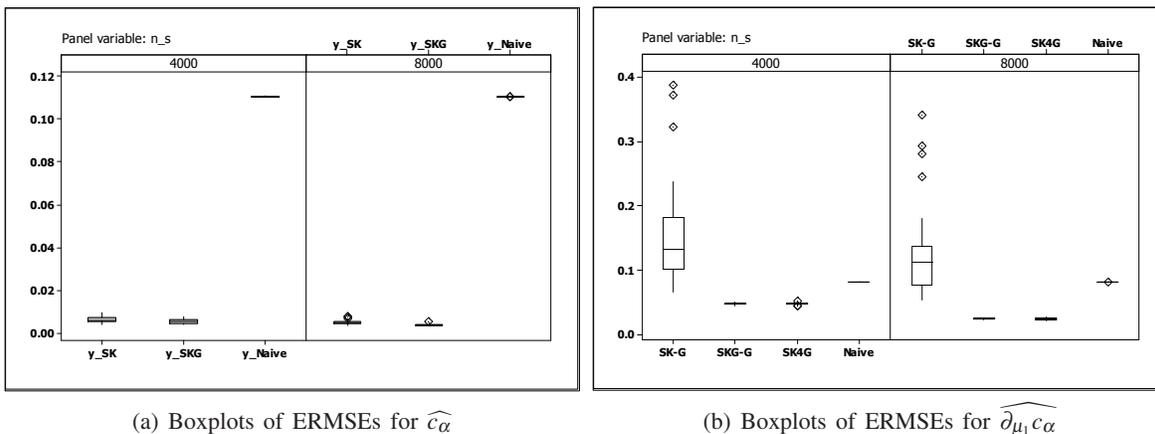


Figure 2: Boxplots of ERMSEs for  $\widehat{c}_\alpha$  and  $\widehat{\partial_{\mu_1} c_\alpha}$  over 50 macro-replications with  $n = 50$  and  $k = 16$  for  $\alpha = 0.99$ .

*Results.* The entire experiment is repeated for 50 macro-replications and the results are summarized in Table 2 and Figures 2 (a) and (b). We first focus on the impacts of the number of simulation replications  $n$  and the sample size  $n_s$  on the performances of the methods under consideration. The table shows ERMSEs

and the associated standard errors of  $\widehat{c}_\alpha, \widehat{\partial_{\mu_i} c_\alpha}$  for  $\alpha = 0.99$  with a fixed computational budget  $n \times n_s = 10^5$  allocated at each of the  $k$  design points.

For the prediction of CVaR, both **SK** and **SKG** outperform the naive simulation **NS** with their mean ERMSEs nearly one order of magnitude smaller. Furthermore, it is seen that **SKG** shows a better performance over **SK** as the sample size  $n_s$  becomes larger. We observe that the computational budget spent at each design point  $n_s \times n$  has a greater influence on the performance of metamodels than the number of design points  $k$  does; increasing  $k$  from 16 to 25 in this example does not make a noticeable difference on performances of metamodel-based approaches. When it comes to gradient estimation **SK4G** and **SKG-G** seem to play equally well; as  $n_s$  becomes moderately large, both methods outperform **NS**. It is interesting to notice that doubling the sample size  $n_s$  while reducing the number of replications  $n$  by a half nevertheless decreases the mean ERMSEs for the **SKG-G** and **SK4G** by nearly a half. Given the low simulation noise level in this particular example and the knowledge that  $n_s$  affects both the variance and bias of CVaR and its gradient estimators, it is not surprising to see the dominating role  $n_s$  plays in improving prediction and gradient estimation performances.

We further explore the impact of  $n_s$  by fixing the number of simulation replications  $n$  to 50 and the number of design points  $k$  to 16 while increasing  $n_s$  from 2000 to 8000. Figure 2 (a) and (b) show the boxplots of ERMSEs for  $\widehat{c}_\alpha$  and  $\widehat{\partial_{\mu_1} c_\alpha}$  for  $n_s = 4000$  and 8000; the results for  $\widehat{\partial_{\mu_2} c_\alpha}$  are similar in nature. In the figures, we can see the advantage of using **SKG** and **SKG-G** for global approximations to CVaR and its sensitivities. Experiments with  $k = 25$  lead to quite similar results.

### 4.3 Derivative Portfolio on a Single Asset

For this example, we consider a portfolio consisting of three derivatives on a single asset with initial price  $S_0$ :

- short one call option with payoff  $H_1(S) = -(S_T - K_c)^+$  with  $K_c = 101$ .
- long one put option with payoff  $H_2(S) = (K_p - S_T)^+$  with  $K_p = 110$ .
- short one Asian call option with payoff  $H_3(S) = -(q^{-1} \sum_{i=1}^q S_{t_i} - K_a)^+$  with  $0 \leq t_1 < \dots < t_q \leq T$  with  $K_a = 110$ .

The real-world drift is  $\mu = 8\%$ , and volatility is  $\sigma = 20\%$ . The risk-free rate is  $r = 3\%$ . The derivatives have the common maturity  $T = 1/12$  year. And the risk horizon is  $\tau = 1/52$  year and it is the time at which the random loss  $L(S_0)$  is computed. This random loss is the sum of values of three derivatives at  $\tau$ , and the calculation is done under the risk-neutral measure by  $L(S_0) = -\sum_{i=1}^3 \mathbb{E} \left[ e^{-r(T-\tau)} H_i(S) \middle| \mathcal{F}_\tau \right]$  where  $\mathcal{F}_\tau$  is the information set available at  $\tau$ . In more detail, it is given by

$$L(S_0) = e^{-r(T-\tau)} \left\{ \mathbb{E} \left[ (S_T - K_c)^+ \middle| S_\tau \right] - \mathbb{E} \left[ (K_p - S_T)^+ \middle| S_\tau \right] + \mathbb{E} \left[ \left( \frac{1}{q} \sum_{i=1}^a S_{t_i} + \frac{1}{q} \sum_{i=a+1}^q S_{t_i} - K_a \right)^+ \middle| S_{t_1}, \dots, S_{t_a}, S_\tau \right] \right\}, \quad (8)$$

where  $a$  is an integer such that  $t_1 < \dots < t_a \leq \tau < t_{a+1} < \dots < t_q \leq T$  and, in this example,  $a = 2$  and  $q = 6$ .

At this point, let us describe the dynamics of the underlying stock price  $S$ . Under the real world probability, it follows the Geometric Brownian Motion with drift  $\mu$  and volatility  $\sigma$ , and it is  $S_t = S_0 \exp((\mu - \sigma^2/2)t + \sigma W_t)$  for  $t \leq \tau$ . Recall that we calculate  $L(S_0)$  under the risk-neutral probability and in this measure  $S_t = S_\tau \exp((r - \sigma^2/2)(t - \tau) + \sigma B_{t-\tau})$  for  $t > \tau$ . Here,  $W$  and  $B$  are two independent standard Brownian motions. Knowing the history of  $S$  at time  $\tau$  is equivalent to knowing the Brownian path  $W$  up to  $\tau$ . Denoting  $q^{-1} \sum_{i=1}^q S_{t_i}$  by  $\bar{S}$ , the pathwise derivative of  $L(S_0)$  is given by

$$L'(S_0) = e^{-r(T-\tau)} \left\{ \mathbb{E} \left[ \mathbf{1}\{S_T \geq K_c\} \frac{S_T}{S_0} \middle| S_\tau \right] + \mathbb{E} \left[ \mathbf{1}\{S_T \leq K_p\} \frac{S_T}{S_0} \middle| S_\tau \right] + \mathbb{E} \left[ \mathbf{1}\{\bar{S} \geq K_a\} \frac{\bar{S}}{S_0} \middle| S_{t_1}, \dots, S_{t_a}, S_\tau \right] \right\}. \quad (9)$$

Note that two expressions inside the expectations in (8) and (9) provide us with the unbiased estimators of the random loss and its first order sensitivity with respect to  $S_0$ , respectively.

*Experiments.* For this example, we use the standard nested simulation to estimate CVaR and its sensitivity. For the outer level simulation, we generate  $n_{\text{outer}}$  scenarios up to the risk horizon  $\tau$ , say  $\omega^1, \dots, \omega^{n_{\text{outer}}}$  which represent the simulated Brownian paths of  $W$ . Conditional on each scenario  $\omega^i$ , we generate  $n_{\text{inner}}$  independent inner level sample paths, which determine the cash flows of the portfolio from time  $\tau$  to  $T$ , i.e.,  $\zeta^{i1}, \dots, \zeta^{in_{\text{inner}}}$  that are the simulated sample paths of  $B$ . For each  $\omega^i$  and  $\zeta^{ij}$ ,  $j = 1, \dots, n_{\text{inner}}$ , we obtain a pair of  $(L_i, D_i)$  where  $D_i = dL/dS_0$  under scenario  $\omega^i$ . To save some space, we do not include formulae for the pair which are quite straightforward to derive.

We continue to describe our experimental design. The space for  $S_0$  is  $\Omega_{S_0} = [80, 120]$  from which we select  $k$  design points. Recall under scenario  $\omega^i$ , we get a pair  $(L_i, D_i)$ . Hence, for  $n_{\text{outer}}$  scenarios at a fixed design point, there are  $\{(L_i, D_i)\}_{i=1}^{n_{\text{outer}}}$ , and we compute estimates for CVaR and its gradient using (7) by replacing  $n_s$  with  $n_{\text{outer}}$ . To apply stochastic kriging, we repeat this two-level simulation  $n$  times at each of  $k$  design points. Consequently, we get  $n$  estimates of  $c_\alpha$  and  $\partial_{S_0} c_\alpha$  at  $k$  design points, based on which CVaR response surface and gradient surface are obtained. In particular, the prediction points are  $N = 193$  regularly spaced points in  $\Omega_{S_0}$ . As for design points, we consider three cases  $k \in \{7, 13, 25\}$ . In all cases, the design points are also equally placed in the design space. Since the closed-form expressions of  $c_\alpha$ ,  $\partial_{S_0} c_\alpha$  are not available, the true values at  $N$  prediction points are approximated using computational budget of  $n_{\text{outer}} = 10^5$  and  $n_{\text{inner}} = 500$ . We present the results for  $\alpha = 0.95$  with  $k = 13$  and  $25$  only.

*Results.* The topic of optimal sampling rules for nested two-level simulation has been studied in the literature. See, for example, Gordy and Juneja (2010) or Broadie et al. (2011a). In this example, we make an attempt to assess the impacts of  $n_{\text{outer}}$  and  $n_{\text{inner}}$  on the performances of the metamodel-based methods under consideration. The entire experiment is repeated for 100 macro-replications and the resulting ERMSEs are summarized in Table 3.

For each two-level simulation, the 1/3 : 2/3 uniform rule uses  $n_{\text{outer}} = \lceil (n_s)^{2/3} \rceil$  and  $n_{\text{inner}} = \lceil (n_s)^{1/3} \rceil$ . In contrast, some papers propose that the sampling rule of  $n_{\text{outer}} = n_s$  and  $n_{\text{inner}} = 1$  can lead to better results. Our experience with this example tells that the 1/3 : 2/3 uniform rule works better and hence is adopted in the experiments for comparing the performances of **SK**, **SKG**, and **SK4G**. For the naive simulation scheme **NS**, the simulation budget allocation at each of the  $N = 193$  check-points is  $\bar{n}_{\text{outer}} = \lceil (k n n_s / N)^{2/3} \rceil$  and  $\bar{n}_{\text{inner}} = \lceil (k n n_s / N)^{1/3} \rceil$ , which amounts to approximately equivalent total computational budget as used in simulations for metamodel-based methods.

Table 3 shows the ERMSEs and the associated standard errors for  $\widehat{c}_\alpha$ ,  $\widehat{\partial_{S_0} c_\alpha}$  with  $\alpha = 0.95$  and the computational budget of  $n_s = n_{\text{outer}} \times n_{\text{inner}}$  at each design point. We consider three cases  $n_s \in \{3.2 \times 10^4, 10^5, 2.5 \times 10^5\}$ . For the prediction of CVaR, both **SK** and **SKG** outperform **NS** in all cases, and it seems that **SK** is slightly better than **SKG**. Regarding the prediction of CVaR sensitivity, the performances of all methods improve as the computational budget (in terms of  $k$  or  $n_s = n_{\text{outer}} \times n_{\text{inner}}$ ) increases. In particular, **SKG-G** and **SK4G** outperform **NS** as  $n_s$  increases to  $10^5$  with  $k = 25$  design points. Based on our extensive numerical tests, we emphasize the crucial role played by  $n_s$  as compared to the number of design points  $k$  and the number of simulation replications  $n$  without showing the detailed results. Although this example shows a potential of **SKG** and **SKG-G** without using advantageous experimental designs, it demonstrates that an optimal budget allocation rule is not at all obvious for building a metamodel in which the selection of design points, the number of simulation replications, outer/inner simulations are involved at the same time. This is one of several research topics that we will address in our future work.

## 5 CONCLUSION

In this paper, we proposed methods for obtaining global approximations to the so-called CVaR and its sensitivities. The methods are based on stochastic kriging using gradient estimators when available.

Table 3: The mean ERMSEs and the corresponding standard errors for  $\widehat{c}_\alpha$  and  $\widehat{\partial_{s_0}c_\alpha}$  over 100 macro-replications using  $n = 10$  and 1/3:2/3 uniform rule for  $\alpha = 0.95$ .

$n_s = n_{\text{outer}} \times n_{\text{inner}}$		$3.2 \times 10^4$		$10^5$		$2.5 \times 10^5$	
ERMSE	$k$	13	25	13	25	13	25
$\widehat{c}_\alpha$	<b>SK</b>	3.61E-01 (3.70E-03)	3.53E-01 (2.90E-03)	2.54E-01 (2.60E-03)	2.41E-01 (1.90E-03)	1.89E-01 (1.90E-03)	1.80E-01 (1.20E-03)
	<b>SKG</b>	4.38E-01 (5.60E-03)	4.50E-01 (4.40E-03)	2.51E-01 (2.50E-03)	2.48E-01 (2.10E-03)	1.98E-01 (2.20E-03)	1.95E-01 (1.40E-03)
	<b>NS</b>	5.99E-01 (3.30E-03)	4.79E-01 (2.30E-03)	4.05E-01 (2.10E-03)	3.25E-01 (1.60E-03)	2.97E-01 (1.50E-03)	2.38E-01 (1.20E-03)
$\widehat{\partial_{s_0}c_\alpha}$	<b>SK-G</b>	6.09E-02 (1.00E-03)	5.12E-02 (1.20E-03)	4.64E-02 (1.10E-03)	3.79E-02 (7.13E-04)	3.80E-02 (9.00E-04)	3.11E-02 (5.85E-04)
	<b>SKG-G</b>	2.44E-02 (6.10E-04)	2.09E-02 (9.68E-05)	1.35E-02 (3.25E-04)	8.20E-03 (4.95E-05)	1.32E-02 (2.02E-04)	6.80E-03 (3.05E-05)
	<b>SK4G</b>	2.50E-02 (1.59E-04)	2.43E-02 (1.02E-04)	1.50E-02 (1.83E-04)	9.30E-03 (8.02E-05)	1.56E-02 (1.36E-04)	7.60E-03 (5.35E-05)
	<b>NS</b>	2.42E-02 (1.41E-04)	1.67E-02 (1.06E-04)	1.45E-02 (8.59E-05)	1.68E-02 (6.24E-05)	1.14E-02 (5.71E-05)	7.90E-03 (5.10E-05)

Stochastic kriging is a metamodeling technique such that the unknown function response is modeled using Gaussian random fields and, in addition, simulation noises are taken into account. Computing risk measures of asset portfolios can be quite time-consuming due to the complexity of financial products. Thus, a good prediction model based on estimates at a small number of points can be very useful in practice. For this purpose, we build a stochastic kriging metamodel, and in particular, we do so in a way that the derivative of the metamodel for CVaR is again a metamodel for CVaR sensitivities. We demonstrated the performances of metamodels through three examples with different complexities: first, European call option, second, quadratic random loss of two assets, and lastly, a derivative portfolio of a single asset. It is observed that the methods are potentially useful, however, the optimal budget allocation for two-level simulation requires further investigations.

### ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. CMMI-0900354 and by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (No. 2012-0003203).

### REFERENCES

Acerbi, C., and D. Tasche. 2002. “On the coherence of expected shortfall”. *Journal of Banking & Finance* 26:1487–1503.

Ankenman, B., B. L. Nelson, and J. Staum. 2010. “Stochastic kriging for simulation metamodeling”. *Operations Research* 58:371–382.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath. 1999. “Coherent measures of risk”. *Mathematical Finance* 9:203–228.

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

Broadie, M., Y. Du, and C. C. Moallemi. 2011a. “Efficient risk estimation via nested sequential simulation”. *Management Science* 57:1172–1194.

Broadie, M., Y. Du, and C. C. Moallemi. 2011b. “Risk estimation via regression”. Technical report, Columbia University, New York, NY.

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

Chen, X., B. Ankenman, and B. L. Nelson. 2011. “Enhancing stochastic kriging metamodels with gradient estimators”. Technical report, Northwestern University, Chicago, IL.

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

- Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. "Variance reduction techniques for estimating value-at-risk". *Management Science* 46:1349–1364.
- Gordy, M. B., and S. Juneja. 2010. "Nested simulation in portfolio risk measurement". *Management Science* 56:1833–1848.
- Gourieroux, C., J. P. Laurent, and O. Scaillet. 2000. "Sensitivity analysis of Value at Risk". *Journal of Empirical Finance* 7:225–245.
- Hong, L. J., and G. Liu. 2009. "Simulation sensitivities of conditional value at risk". *Management Science* 55:281–293.
- Lan, H., B. L. Nelson, and J. Staum. 2010. "A confidence interval procedure for expected shortfall risk measurement via two-level simulation". *Operations Research* 58:1481–1490.
- Liu, M., and J. Staum. 2010. "Stochastic kriging for efficient nested simulation of expected shortfall". *Journal of Risk* 12:3–27.
- Morris, M. D., T. J. Mitchell, and D. Ylvisaker. 1993. "Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction". *Technometrics* 35:243–255.
- Pardo-Igúzquiza, E., and M. Chica-Olmo. 2004. "Estimation of gradients from sparse data by universal kriging". *Water Resources Research* 40:1–17 W12418.
- Parzen, E. 1962. *Stochastic Processes*. CA: Holden-Day.
- Sakata, S., F. Ashida, and M. Zako. 2010. "Comparative study on gradient and Hessian estimation using the Kriging method and neural network approximation". *Mathematical and Computer Modelling* 51:309–319.
- Scaillet, O. 2004. "Nonparametric estimation and sensitivity analysis of expected shortfall". *Mathematical Finance* 14:115–129.
- Solak, E., R. Murray-Smith, W. E. Leithead, D. J. Leith, and C. E. Rasmussen. 2003. "Derivative observations in Gaussian Process models of dynamic systems". *Advances in Neural Information Processing Systems* 15:1033–1040.

## AUTHOR BIOGRAPHIES

**XI CHEN** is a Ph.D. candidate in the Department of Industrial Engineering and Management Sciences at Northwestern University. She expects to finish her Ph.D. in August 2012 and will be an assistant professor in the Department of Statistical Sciences and Operations Research at Virginia Commonwealth University in Fall 2012. She received her M.S. degrees from North Carolina State University and Northwestern University in 2008 and 2009, respectively. Her research interests include stochastic modeling and simulation, applied probability and statistics, computer experiment design and analysis, and simulation optimization. Her e-mail address and webpage are respectively [chenxi.iems@u.northwestern.edu](mailto:chenxi.iems@u.northwestern.edu) and <http://users.iems.northwestern.edu/~phoebe08/>.

**KYOUNG-KUK KIM** is an assistant professor in the Department of Industrial and Systems Engineering at Korea Advanced Institute of Science and Technology. He received a Ph.D. from Columbia Business School. He also received his B.S. and M.S. in mathematics from Seoul National University and Stanford University, respectively. His research interests include stochastic processes and simulation, financial engineering, and risk management. He can be reached at [catenoid@kaist.ac.kr](mailto:catenoid@kaist.ac.kr).

**BARRY L. NELSON** is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. His e-mail and web addresses are [nelsonb@northwestern.edu](mailto:nelsonb@northwestern.edu) and [www.iems.northwestern.edu/nelsonb](http://www.iems.northwestern.edu/nelsonb).