Real-Time Derivative Pricing and Hedging with Consistent Metamodels

Guangxin Jiang,[a] L. Jeff Hong,[b,*] Haihui Shen[c]

[a] School of Management, Harbin Institute of Technology, Harbin 150001, China; [b] School of Management and School of Data Science, Fudan University, Shanghai 200433, China; [c] Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China
*Corresponding author

Contact: gxjiang@hit.edu.cn, https://orcid.org/0000-0002-2604-7750 (GJ); hong_liu@fudan.edu.cn, https://orcid.org/0000-0001-7011-4001 (LJH); shenhaihui@sjtu.edu.cn, https://orcid.org/0000-0002-4157-1278 (HS)

Abstract. In derivative pricing and hedging, the consistency between the price and Greek surfaces (i.e., the Greek surfaces can be obtained by differentiating the price surface) is important in stabilizing the balance sheet and reducing the hedging cost. To build consistent surfaces of the price and Greeks for real-time decisions, we propose to use the gradient-enhanced stochastic kriging method, based on the data collected through extensive simulation experiments conducted when the market is closed. In addition to the naturally guaranteed consistency, we prove that the constructed price and Greek surfaces are more accurate than those constructed separately using stochastic kriging. Besides the consistency between the price and Greeks, we show that the partial differential equation relation between the price and Greeks, implied by the famous Feynman-Kac formula, can also be used to further improve the accuracy of the constructed surfaces. The numerical studies show that our proposed metamodeling methods work well for derivative pricing and hedging.

1. Introduction

To trade financial derivatives, traders need to know the fair prices of the derivatives to determine the bid or ask prices, and at the same time, they also need to know the Greeks, that is, the sensitivities of prices with respect to market factors, to calculate and to hedge the risk of the portfolio. However, the fair prices and the Greeks change rapidly as the market factors change. Therefore, it is important to be able to calculate them in real time. As a result, closed-form formulae of the derivative prices are desired in practice, because they can be used to calculate the prices and the Greeks (by simply differentiating the prices formulae with respect to the market factors) in real time on observing the values of the market factors.

Unfortunately, closed-form formulae are in general difficult to derive for many derivatives due to their complex payment structures (e.g., exotic options) or the complex models of underlying dynamics (e.g., the stochastic volatility models). To accurately calculate the prices and Greeks in such cases, Monte Carlo simulation is often used. There are numerous studies on derivative pricing through Monte Carlo simulation, for example, Glasserman et al. (1999), Broadie and Kaya (2006), and Cai et al. (2017); see Glasserman (2003) for a comprehensive introduction. Compared with derivative pricing, estimating Greeks via Monte Carlo simulation is a more challenging problem, and several methods have been proposed to tackle this, including the likelihood ratio (LR) method (Glynn 1990), the infinitesimal perturbation analysis (IPA) method (Glasserman 1991), the smoothed perturbation analysis method (Fu and Hu 1997), the kernel method (Liu and Hong 2011), and the mixed method (Peng et al. 2018).

However, Monte Carlo simulation is typically time-consuming. Although there is a large body of research on improving simulation efficiency, for example, variance reduction techniques (Kemma and Vorst 1990, Glasserman

et al. 1999) and using high-performance computing (Zenios 1999), Monte Carlo simulation is still difficult to use in a real-time environment, especially in dealing with derivative portfolios that may contain hundreds of derivatives with hundreds to thousands of underlying assets. Therefore, for real-time pricing and hedging, practitioners still resort to simple models (e.g., the Black-Scholes model) and their closed-form price formulae (e.g., the Black-Scholes formula; Black and Scholes 1973), despite that they may be inaccurate or unreliable.

Recently, Hong and Jiang (2019) propose the offline-simulation-online-application (OSOA) framework to bridge the gap between slow simulation and fast real-time decision making. The key is to use time-consuming offline simulation experiments to learn the metamodels (i.e., surfaces, often in closed-form formulae) that may be used in real-time decision making. This OSOA idea has successful applications in, for example, real-time risk monitoring (Jiang et al. 2020) and online ranking and selection (Shen et al. 2021, Liu et al. 2022). For the problem considered in this paper, this framework would suggest to build the price and Greek metamodels based on their estimates on some design points via offline Monte Carlo simulation and to use them for real-time price quoting and risk hedging.

However, there is a difference between our problem and a typical metamodeling problem (see Barton (2015) for a review). For a derivative or a derivative portfolio, we need both the price and Greeks, and there are relations between them, that is, the Greeks are the first- or second-order partial derivatives of the price with respect to different market factors. When using Monte Carlo simulation to estimate the price and Greeks on chosen market scenarios (i.e., design points), these relations can be maintained on these design points by using appropriate Greek estimators like the LR and IPA estimators. However, when using the price and Greek estimates of the design points to build metamodels, these relations may not be necessarily maintained, and it is not clear how the violations of these relations affect risk hedging. In this paper, we first investigate this issue. We say that the metamodels are *consistent* if the price and Greek metamodels satisfy the differentiation relations. We show that consistency is important for our problem. More specifically, we prove that consistent metamodels may achieve better hedging quality and lower hedging costs. Therefore, ignoring consistency by building separate metamodels of price and Greeks is not a good idea. In this paper, our goal is to build a price metamodel that can be differentiated to produce the Greek metamodels. In this way, they are guaranteed to be consistent.

When building metamodels, we focus on the kriging-based methods (also known as Gaussian process methods), which do not require domain knowledge of the problem and provide flexible global fitting; see Cressie (1993) and Rasmussen and Williams (2006) for comprehensive surveys. In the simulation literature, Ankenman et al. (2010) propose the stochastic kriging (SK) approach to handle the random nature of simulation outputs, and Chen et al. (2013) propose the gradient-enhanced stochastic kriging (GESK), which incorporates gradient estimators into SK to improve the surface fitting accuracy. Qu and Fu (2014) also propose to incorporate the gradient information into the SK but with an extrapolation approach. In this paper, we show that the GESK can be used to achieve our goal, although it is not designed for this purpose. By building a price metamodel that incorporates both the price and the Greeks information, we prove that the price metamodel and the Greeks metamodels (obtained by differentiating the price metamodel) are more accurate than those built separately using SK in addition to the naturally guaranteed consistency.

For the price and Greeks of a derivative (or a derivative portfolio), consistency is not the only relation that they should satisfy. Theoretically, there is a dynamic equation, typically in the form of a partial differential equation (PDE), which establishes a quantitative relation between the price and the Greeks. For example, when the underlying asset follows a geometric Brownian motion, the derivative price follows the well-known Black-Scholes equation (Black and Scholes 1973). This equation specifies that the price and some Greeks (i.e., the delta, gamma, and theta) satisfy a specific relation. In the area of deep learning, the PDEs that describe physics laws (e.g., heat equation) have been effectively integrated into the neural networks to tackle supervised learning tasks, known as physics-informed neural networks (Raissi et al. 2019, Wang et al. 2021, Cuomo et al. 2022); see Karniadakis et al. (2021) for a survey. Inspired by this idea, we integrate the PDE into the GESK as a constraint and propose a PDE-constrained GESK approach to further improve the accuracy of the price and Greek metamodels.

The rest of this paper is organized as follows. Section 2 describes the setting of the metamodel-based prediction for the price and Greeks. In Section 3, we formally define consistency and demonstrate its importance in risk hedging. Then in Section 4, we propose to use the GESK to build the price and Greek metamodels, which naturally ensures consistency and achieves higher accuracy compared with the separate SK. In Section 5, we propose the PDE-constrained GESK to further improve the accuracy, where the PDE is specified by the Feynman-Kac formula. Numerical studies are presented in Section 6, followed by conclusions and discussions in Section 7. Some technical proofs and additional materials are included in the e-companion.

2. Problem Setting

Let $\mathbf{s} \in \Re^{d-1}$, where $d \geq 2$, denote a vector of $d-1$ market factors, which may include the prices and volatilities of the underlying assets, the interest rate, the exchange rates, and so on. The value of a derivative (or a derivative portfolio) depends on the market factors. Furthermore, let $t \in [0, T]$ be the time index, where $T > 0$ is the maturity time of the derivative (or the longest maturity time of all derivatives in the portfolio). For the convenience of notation, we sometimes put the time index t and the market factors $\mathbf{s}$ together as $\mathbf{x} = (\mathbf{s}^\top, t)^\top$ and call it a market scenario. Let $\mathcal{X} \subset \Re^d$ denote the range of the market scenario. In practice, $\mathcal{X}$ is the anticipated variation range of the market factors in the next trading day based on their previous closing values. It can be determined by the market mechanism or practitioners' own belief or prediction model. Let $V(\mathbf{x}) = V(\mathbf{s}, t)$ denote the value of the derivative (or the derivative portfolio), and let $G^k(\mathbf{x}) = G^k(\mathbf{s}, t) = \partial V(\mathbf{x})/\partial x_k$ denote the kth Greek (i.e., partial derivative) of the financial derivative (or the derivative portfolio), $k = 1, 2, \ldots, d$.

In this paper, we suppose that $V(\mathbf{x})$ and $G^k(\mathbf{x})$, $k = 1, 2, \ldots, d$, do not have closed-form formulae, and can only be estimated via simulation. Specifically, for a set of chosen market scenarios $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathcal{X}$, let $Y_\ell(\mathbf{x}_i)$ and $D_\ell^k(\mathbf{x}_i)$, $k = 1, 2, \ldots, d$, be the ℓth simulation observation of the price and the kth Greek at the market scenario $\mathbf{x}_i$, respectively, where $D_\ell^k(\mathbf{x}) = \partial Y_\ell(\mathbf{x})/\partial x_k$. Then, the estimated price and kth Greek at $\mathbf{x}_i$ with m_i replications are given by

$$\overline{Y}(\mathbf{x}_i) = \frac{1}{m_i} \sum_{\ell=1}^{m_i} Y_\ell(\mathbf{x}_i) \ \text{ and } \ \overline{D}^k(\mathbf{x}_i) = \frac{1}{m_i} \sum_{\ell=1}^{m_i} D_\ell^k(\mathbf{x}_i), \ k = 1, 2, \ldots, d.$$

Our goal is to build the metamodels of $V(\mathbf{x})$ and $G^k(\mathbf{x})$, denoted as $\hat{V}(\mathbf{x})$ and $\hat{G}^k(\mathbf{x})$, respectively, with the data

$$\{(\mathbf{x}_i, \overline{Y}(\mathbf{x}_i), \overline{D}^k(\mathbf{x}_i)) : \ k = 1, 2, \ldots, d, \ i = 1, 2, \ldots, n\},$$

so that the metamodels may be used like closed-form formulae for real-time evaluations of the price and Greeks once the market scenario $\mathbf{x}$ is observed.

The price and Greek surfaces satisfy $G^k(\mathbf{x}) = \partial V(\mathbf{x})/\partial x_k$ for all $k = 1, 2, \ldots, d$. However, these relations may not be necessarily satisfied by the metamodels $\hat{V}(\mathbf{x})$ and $\hat{G}^k(\mathbf{x})$. We define that the price and Greek metamodels are consistent if $\hat{G}^k(\mathbf{x}) = \partial \hat{V}(\mathbf{x})/\partial x_k$ for all $\mathbf{x} \in \mathcal{X}$ and $k = 1, 2, \ldots, d$. Intuitively, the consistency is important to ensure the quality of risk hedging. However, little theory has been established in the literature. Therefore, before discussing how to build consistent $\hat{V}(\mathbf{x})$ and $\hat{G}^k(\mathbf{x})$, we first provide a mathematical characterization of the importance of the consistency in the next section.

3. Importance of the Consistency

To analyze the importance of the consistency, we take the widely used delta hedging strategy as an example and consider only a single underlying asset while noting that the same arguments apply to other hedging strategies and multiple underlying assets, as long as the hedging instruments exist and they are included in the portfolio. The delta hedging strategy approximates $V(\mathbf{x})$, the value of the derivative portfolio at time t, by $V(S(t))$, where $S(t)$ is the price of the underlying asset. By the first-order Taylor expansion,

$$V(S(t)) \approx V(S(t_0)) + V'(S(t_0))[S(t) - S(t_0)], \tag{1}$$

for any $t \in [t_0, t_0 + \tau]$ with small $\tau > 0$, where $V'(s) \triangleq dV(s)/ds$, and s denotes a generic price of the asset. Suppose the company shorts one derivative, then the delta hedging strategy enters $V'(S(t_0))$ positions of the underlying asset to the portfolio at time t_0. Then, the profit and loss (P&L) from the time t_0 to t, denoted by $L(S(t))$, may be approximated by

$$L(S(t)) = -[V(S(t)) - V(S(t_0))] + V'(S(t_0))[S(t) - S(t_0)] \approx 0,$$

by Equation (1) for any $t \in [t_0, t_0 + \tau]$. Therefore, the portfolio is (approximately) fully hedged against the price change of the underlying asset. Because $V'(S(t_0))$ is known as delta (Δ) at time t_0, the strategy is known as the delta hedging strategy. It requires the portfolio to rebalance (i.e., change of the position of the underlying asset according to the delta) every τ units of time. In practice, however, both $V(S(t_0))$ and $V'(S(t_0))$ may be estimated using the metamodels and are not necessarily consistent, which may affect the quality of hedging. In the rest of this section, we demonstrate the importance of the consistency in achieving high hedging quality.

3.1. Known $V(s)$

When $V(s)$ is assumed to be known in the ideal situation, to perform the delta hedging, by definition one needs to calculate the Greek $\Delta(s) \triangleq dV(s)/ds$. Using such $V(s)$ and $\Delta(s)$ for delta hedging, we are of course in a case where

consistency exists. Suppose one chooses to use $V(s)$ and some wrong $\Delta^{\dagger}(s)$, then inconsistency happens. At time 0, the underlying asset price is observed as $s_0 > 0$. Then the company will hold $\Delta(s_0)$ shares of the underlying asset in the consistency case and $\Delta^{\dagger}(s_0)$ shares in the inconsistency case, respectively. Suppose that $\Delta^{\dagger}(s_0) \neq \Delta(s_0)$. At time $t \in (0, \tau_1]$, the underlying asset price changes to a random value $S(t)$, and the company will have P&L

$$L(S(t)) = -[V(S(t)) - V(s_0)] + \Delta(s_0)[S(t) - s_0], \tag{2}$$

in the consistency case, and

$$L^{\dagger}(S(t)) = -[V(S(t)) - V(s_0)] + \Delta^{\dagger}(s_0)[S(t) - s_0], \tag{3}$$

in the inconsistent case, respectively. Let $\xrightarrow{d}$ denote the convergence in distribution. The following Theorem 1 says that the fluctuation of the company's P&L will be smaller when consistency exists. The proof of Theorem 1 is in Section EC.1 of the e-companion.

Theorem 1. *Suppose that the underlying asset $S(t)$ is driven by the exponential family of stochastic process $S(t) = s_0 \exp(at + \sqrt{t} X_t)$, where $X_t \xrightarrow{d} X$ as $t \to 0^+$ with $\mathbb{E}[X^4] < \infty$ and $\mathrm{Var}[X] > 0$. Moreover, assume that there exist $h > 0$ and $t_h > 0$ such that $\sup_{0 < t \leq t_h} \mathbb{E}[e^{\theta X_t}] < \infty$ for all $|\theta| \leq h$, and the second-order derivative of $V(s)$ is bounded above. Then for the P&L $L(S(t))$ defined in (2) and $L^{\dagger}(S(t))$ defined in (3), there exists $\tau > 0$ such that $\mathrm{Var}[L(S(t))] < \mathrm{Var}[L^{\dagger}(S(t))]$ for $t < \tau$.*

Remark 1. One critical assumption of Theorem 1 is $S(t) = s_0 \exp(at + \sqrt{t} X_t)$, where $X_t \xrightarrow{d} X$ with $\mathbb{E}[X^4] < \infty$ and $\mathrm{Var}[X] > 0$. This assumption actually holds for a lot of asset models. For example, when $S(t)$ is modeled as a geometric Brownian motion, that is, $S(t) = s_0 \exp(at + \sqrt{t} Z)$, where Z is a standard normal random variable, of course it satisfies the assumption. When the Heston model is used, which is a commonly used stochastic volatility model, it can be shown that the assumption is also satisfied; see the details in Section EC.2 of the e-companion.

Remark 2. If one is willing to assume that $V(s_0)$ is approximately linear over a small range around s_0 (i.e., perfect hedging), then the result of Theorem 1 can be obtained without assuming any form of $S(t)$. Indeed, in this case, $L(S(t)) \approx 0$, whereas $L^{\dagger}(S(t)) \neq 0$ for small t.

Theorem 1 tells us that with the same hedging frequency, the existence of consistency leads to a better hedging effect, which makes the balance sheet more stable. Conversely, one could also consider achieving the same hedging effect and comparing the efforts required with and without consistency. Intuitively, the hedging cost in the consistency case will be less than that in the inconsistency case, because the position may be overadjusted and needs to be adjusted back in the latter case. We defer the rigorous analysis to Section EC.3 of the e-companion and formally state the previous result in Theorem EC.1 therein.

3.2. Estimated *V(s)*

When $V(s)$ is unknown, it needs to be estimated (constructed) in a certain way, and we denote its estimator as $\hat{V}(s)$. In such a situation, although $\hat{V}(s)$ may deviate from $V(s)$, it is used and treated as the "true price" in pricing and hedging anyway. Therefore, the P&L is calculated mainly for accounting purpose, that is, to produce the balance sheet. Another issue is that, when we want to compare the hedging quality in the consistency case and in the inconsistency case, the estimator $\hat{V}(s)$ may be different in the two cases.

Now suppose we use $\hat{V}_1(s)$ and $\Delta_1(s)$ with $\Delta_1(s) = d\hat{V}_1(s)/ds$ for delta hedging in the consistency case, whereas we use $\hat{V}_2(s)$ and $\Delta_2(s)$ with $\Delta_2(s) \neq d\hat{V}_2(s)/ds$ in the inconsistency case. Then at time 0, the company will hold $\Delta_1(s_0)$ shares of the underlying asset in the consistency case and $\Delta_2(s_0)$ shares of the underlying asset in the inconsistency case, respectively. If without successive adjustment, at time t, the bookkeeping P&L recorded in the balance sheet is

$$L_1(S(t)) = -[\hat{V}_1(S(t)) - \hat{V}_1(s_0)] + \Delta_1(s_0)[S(t) - s_0], \tag{4}$$

in the consistency case, and

$$L_2(S(t)) = -[\hat{V}_2(S(t)) - \hat{V}_2(s_0)] + \Delta_2(s_0)[S(t) - s_0], \tag{5}$$

in the inconsistency case, respectively. Theorem 2 says that the fluctuation of company's bookkeeping P&L will be smaller when consistency exists. The proof of Theorem 2 is in Section EC.4 of the e-companion.

Theorem 2. *With the same assumptions in Theorem 1 but replace the assumption for the second-order derivative of $V(s)$ with that the second-order derivatives of $\hat{V}_1(s)$ and $\hat{V}_2(s)$ are bounded above, for the P&L $L_1(S(t))$ defined in (4) and $L_2(S(t))$ defined in (5), there exists $\tau > 0$ such that $\mathrm{Var}[L_1(S(t))] < \mathrm{Var}[L_2(S(t))]$ for $t < \tau$.*

Conversely, we can also consider the comparison of hedging cost to achieve the same hedging effect for the two cases. When we additionally assume that $\hat{V}_1(s) \equiv \hat{V}_2(s)$, then the same result as in the case where $V(s)$ is known can be obtained immediately. When $\hat{V}_1(s) \neq \hat{V}_2(s)$, a similar result can be also expected, although the analysis may become mathematically intractable.

In this section, we discussed the importance of consistency in delta hedging, which makes the balance sheet more stable and leads to less hedging cost. In other words, without consistency, the delta hedging strategy may be ineffective, and the risk exposure of the portfolio may be high and sensitive to changes in market factors. The same arguments also apply to other hedging strategies. In the next section, we will provide a metamodeling method to construct surfaces of price and Greeks, which can ensure such consistency.

4. SK for Price and Greek Surfaces Construction

Recall that we require the constructed price and Greek surfaces $\hat{V}(\mathbf{x})$ and $\hat{G}^k(\mathbf{x})$ to have closed-form expressions, so they can be used in real time. The SK method introduced by Ankenman et al. (2010) and van Beers and Kleijnen (2003) is a very suitable candidate, which does not require domain knowledge of the problem context and provides a global fitting of desired metamodels with closed-form expressions. In SK, the sample means of the price and Greeks at $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are modeled as

$$\overline{Y}(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta} + \mathrm{M}(\mathbf{x}_i) + \overline{\varepsilon}(\mathbf{x}_i), \tag{6}$$

$$\overline{D}^k(\mathbf{x}_i) = \left(\frac{\partial}{\partial x_k} \mathbf{f}(\mathbf{x}_i) \right)^\top \boldsymbol{\beta} + \frac{\partial}{\partial x_k} \mathrm{M}(\mathbf{x}_i) + \overline{\epsilon}^k(\mathbf{x}_i), \quad k = 1, 2, \ldots, d, \tag{7}$$

where $\mathbf{f}(\mathbf{x}) \in \Re^p$ is the prior information (vector of basis functions), $\boldsymbol{\beta} \in \Re^p$ is a vector of unknown parameters, $\mathrm{M}(\mathbf{x})$ is a Gaussian random field with zero mean, and

$$\frac{\partial}{\partial x_k} \mathrm{M}(\mathbf{x}) \triangleq \lim_{t \to 0} \frac{\mathrm{M}(\mathbf{x} + t\mathbf{e}_k) - \mathrm{M}(\mathbf{x})}{t},$$

where $\mathbf{e}_k$ is the $d \times 1$ unit vector with the kth element being one and the others being zeros. Moreover, $\overline{\varepsilon}(\mathbf{x}_i) = \sum_{\ell=1}^{m_i} \varepsilon_\ell(\mathbf{x}_i)/m_i$, where $\varepsilon_\ell(\mathbf{x}) = Y_\ell(\mathbf{x}) - V(\mathbf{x})$ is the simulation noise of price; and $\overline{\epsilon}^k(\mathbf{x}_i) = \sum_{\ell=1}^{m_i} \epsilon_\ell^k(\mathbf{x}_i)/m_i$, where $\epsilon_\ell^k(\mathbf{x}) = D_\ell^k(\mathbf{x}) - G^k(\mathbf{x})$ is the simulation noise of the kth Greek. Assume that both $\varepsilon_\ell(\mathbf{x})$ and $\epsilon_\ell^k(\mathbf{x}_i)$ are independent of $\mathrm{M}(\mathbf{x})$, and $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \ldots$, are independent and identically distributed (i.i.d.) with zero mean and finite variance, and $\epsilon_1^k(\mathbf{x}), \epsilon_2^k(\mathbf{x}), \ldots$, are i.i.d. with zero mean and finite variance. Correlation of the noises only exists among $\{\varepsilon_\ell(\mathbf{x}), \epsilon_\ell^1(\mathbf{x}), \ldots, \epsilon_\ell^d(\mathbf{x})\}$ and not between components with different $\mathbf{x}$ or ℓ.

As commonly assumed, $\mathrm{M}(\mathbf{x})$ has covariance function

$$\mathrm{Cov}[\mathrm{M}(\mathbf{x}), \mathrm{M}(\mathbf{y})] = \tau^2 \exp \left\{ -\sum_{k=1}^{d} \theta_k (x_k - y_k)^2 \right\}, \tag{8}$$

where x_k and y_k denote the kth coordinate of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\theta_k \geq 0$ for $k = 1, 2, \ldots, d$. Given that the covariance function of $\mathrm{M}(\mathbf{x})$ has continuous second-order derivative (which is satisfied by (8)), it can be shown that $\partial \mathrm{M}(\mathbf{x})/\partial x_k$ is also a Gaussian random field (Chen et al. 2013) with zero mean and covariance function $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{x})/\partial x_k, \partial \mathrm{M}(\mathbf{y})/\partial y_k]$, whose specific form is given in Section EC.5 of the e-companion, together with the covariance functions $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{x})/\partial x_k, \mathrm{M}(\mathbf{y})]$ and $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{x})/\partial x_k, \partial \mathrm{M}(\mathbf{y})/\partial y_h]$. It is worth noting that all the previously mentioned covariance functions have common parameters τ^2 and $\boldsymbol{\theta} \triangleq (\theta_1, \ldots, \theta_d)^\top$.

4.1. Separate SK

A plain way of using SK is to build the price and Greek surfaces separately. Here we only summarize the results, and refer to Ankenman et al. (2010) for details. Let $\overline{\mathbf{Y}} = (\overline{Y}(\mathbf{x}_1), \ldots, \overline{Y}(\mathbf{x}_n))^\top$, and $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ be $n \times n$ matrices with the (i, j)-th element being $\mathrm{Cov}[\mathrm{M}(\mathbf{x}_i), \mathrm{M}(\mathbf{x}_j)]$ and $\mathrm{Cov}[\overline{\varepsilon}(\mathbf{x}_i), \overline{\varepsilon}(\mathbf{x}_j)]$, respectively. Let $\boldsymbol{\gamma}(\mathbf{z})$ be a $n \times 1$ vector with the ith element being $\mathrm{Cov}[\mathrm{M}(\mathbf{z}), \mathrm{M}(\mathbf{x}_i)]$. Also, let $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n))^\top$, which is a $n \times p$ matrix. Provided that $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\boldsymbol{\Sigma}$ are known and $(\boldsymbol{\Gamma} + \boldsymbol{\Sigma})$ is invertible, at any new point $\mathbf{z} \in \mathcal{X}$, the SK mean squared error (MSE) optimal predictor of

$V(\mathbf{z})$ is given by

$$\hat{V}(\mathbf{z}) = \mathbf{f}(\mathbf{z})^\top \boldsymbol{\beta} + \boldsymbol{\gamma}(\mathbf{z})^\top (\boldsymbol{\Gamma} + \boldsymbol{\Sigma})^{-1}(\overline{\mathbf{Y}} - \mathbf{F}\boldsymbol{\beta}), \tag{9}$$

with

$$\mathrm{MSE}_V^{\mathrm{SK}}(\mathbf{z}) \triangleq \mathbb{E}[(\hat{V}(\mathbf{z}) - V(\mathbf{z}))^2] = \mathrm{Var}[\mathrm{M}(\mathbf{z})] - \boldsymbol{\gamma}(\mathbf{z})^\top (\boldsymbol{\Gamma} + \boldsymbol{\Sigma})^{-1}\boldsymbol{\gamma}(\mathbf{z}). \tag{10}$$

For $k = 1, 2, \ldots, d$, let $\overline{\mathbf{D}}^k = (\overline{D}^k(\mathbf{x}_1), \ldots, \overline{D}^k(\mathbf{x}_n))^\top$, and $\boldsymbol{\Gamma}^{k,k}$ and $\boldsymbol{\Sigma}^{k,k}$ be $n \times n$ matrices with the (i, j)th element being $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{x}_i)/\partial x_k, \partial \mathrm{M}(\mathbf{x}_j)/\partial x_k]$ and $\mathrm{Cov}[\overline{\epsilon}^k(\mathbf{x}_i), \overline{\epsilon}^k(\mathbf{x}_j)]$, respectively. Let $\boldsymbol{\gamma}^{k,k}(\mathbf{z})$ be a $n \times 1$ vector with the ith element being $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{z})/\partial x_k, \partial \mathrm{M}(\mathbf{x}_i)/\partial x_k]$. Also, let $\mathbf{F}^k = (\partial \mathbf{f}(\mathbf{x}_1)/\partial x_k, \ldots, \partial \mathbf{f}(\mathbf{x}_n)/\partial x_k)^\top$, which is a $n \times p$ matrix. Provided that $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\boldsymbol{\Sigma}^{k,k}$ are known and $(\boldsymbol{\Gamma}^{k,k} + \boldsymbol{\Sigma}^{k,k})$ is invertible, at any new point $\mathbf{z} \in \mathcal{X}$, the SK MSE optimal predictor of $G^k(\mathbf{z})$ is given by

$$\hat{G}^k(\mathbf{z}) = \left(\frac{\partial}{\partial z_k}\mathbf{f}(\mathbf{z})\right)^\top \boldsymbol{\beta} + \boldsymbol{\gamma}^{k,k}(\mathbf{z})^\top (\boldsymbol{\Gamma}^{k,k} + \boldsymbol{\Sigma}^{k,k})^{-1}(\overline{\mathbf{D}}^k - \mathbf{F}^k\boldsymbol{\beta}), \tag{11}$$

with

$$\mathrm{MSE}_{G^k}^{\mathrm{SK}}(\mathbf{z}) \triangleq \mathbb{E}[(\hat{G}^k(\mathbf{z}) - G^k(\mathbf{z}))^2] = \mathrm{Var}\left[\frac{\partial}{\partial z_k}\mathrm{M}(\mathbf{z})\right] - \boldsymbol{\gamma}^{k,k}(\mathbf{z})^\top (\boldsymbol{\Gamma}^{k,k} + \boldsymbol{\Sigma}^{k,k})^{-1}\boldsymbol{\gamma}^{k,k}(\mathbf{z}). \tag{12}$$

In practice, $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{k,k}$ are unknown. After estimating the diagonal matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{k,k}$ with the sample variances, we can estimate $\boldsymbol{\beta}$ and $(\tau^2, \boldsymbol{\theta})$ via the maximum likelihood estimation (MLE). Then the final empirical predictor of $V(\mathbf{z})$ is $\hat{V}(\mathbf{z})$ in (9) with $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\boldsymbol{\Sigma}$ replaced with their estimators, and the final empirical predictor of $G^k(\mathbf{z})$ is $\hat{G}^k(\mathbf{z})$ in (11) with $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\boldsymbol{\Sigma}^{k,k}$ replaced with their estimators. When the prior information $\mathbf{f}(\mathbf{x})$ is unavailable, it is a convention to take $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta$ as a noninformative prior. When some stylized model for $\hat{V}(\mathbf{x})$ exists (under stronger assumptions or with more simplification), it can also be included in $\mathbf{f}(\mathbf{x})$ (Shen et al. 2018).

The price and Greek surfaces constructed separately using SK do possess closed-form expressions. However, there is no guarantee of consistency. As we have discussed in Section 3, such constructed price and Greek surfaces may lead to an unsatisfactory hedging effect, which is also demonstrated in the numerical experiments in Section 6. A naive way to ensure consistency is that we keep $\hat{V}(\mathbf{z})$ in (9) only and obtain the Greeks by $\partial \hat{V}(\mathbf{z})/\partial z_k$. However, it can be anticipated that the quality of the derived Greek surfaces is poor because no information on simulated Greeks is actually used and the derived Greeks entirely depend on the estimated price. With such Greek surfaces, the hedging effect may also be unsatisfactory, which is demonstrated as well in the numerical experiments in Section 6.

4.2. Gradient-Enhanced SK

To ensure both the consistency and the quality of the constructed price and Greek surfaces, we propose to use the GESK introduced in Chen et al. (2013), which incorporates the response surface's gradient estimators into SK to better predict the response. More specifically, it simultaneously uses the models in (6) and (7) and uses the correlation between $\{\overline{Y}(\mathbf{x}_i), \overline{D}^1(\mathbf{x}_i), \ldots, \overline{D}^d(\mathbf{x}_i)\}$ to better predict $Y(\mathbf{z})$. Let $\overline{\mathbf{Y}}_+$ be the $n(d+1) \times 1$ vector containing the averaged responses and all the averaged gradient estimators, that is, $\overline{\mathbf{Y}}_+ = (\overline{\mathbf{Y}}^\top, (\overline{\mathbf{D}}^1)^\top, \ldots, (\overline{\mathbf{D}}^d)^\top)^\top$. Let $\mathbf{F}_+$ be the $n(d+1) \times p$ matrix containing all the basis functions and their gradients, that is, $\mathbf{F}_+ = (\mathbf{F}^\top, (\mathbf{F}^1)^\top, \ldots, (\mathbf{F}^d)^\top)^\top$. Let $\boldsymbol{\gamma}_+(\mathbf{z})$ be the $n(d+1) \times 1$ vector containing the covariances between $\mathrm{M}(\mathbf{z})$ and all $\mathrm{M}(\mathbf{x}_i)$ and $\partial \mathrm{M}(\mathbf{x}_i)/\partial x_k$, that is, $\boldsymbol{\gamma}_+(\mathbf{z}) = (\boldsymbol{\gamma}(\mathbf{z})^\top, \boldsymbol{\gamma}^{0,1}(\mathbf{z})^\top, \ldots, \boldsymbol{\gamma}^{0,d}(\mathbf{z})^\top)^\top$, where $\boldsymbol{\gamma}^{0,k}(\mathbf{z})$ is a $n \times 1$ vector with the ith element being $\mathrm{Cov}[\mathrm{M}(\mathbf{z}), \partial \mathrm{M}(\mathbf{x}_i)/\partial x_k]$, for $k = 1, 2, \ldots, d$. Let $\boldsymbol{\Gamma}_+$ and $\boldsymbol{\Sigma}_+$ be $n(d+1) \times n(d+1)$ matrices including the covariances among all $\mathrm{M}(\mathbf{x}_i)$ and $\partial \mathrm{M}(\mathbf{x}_i)/\partial x_k$, and the covariances among all $\overline{\varepsilon}(\mathbf{x}_i)$ and $\overline{\epsilon}^k(\mathbf{x}_i)$, respectively. More specifically,

$$\boldsymbol{\Gamma}_+ = \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}^{0,1} & \cdots & \boldsymbol{\Gamma}^{0,d} \\ \boldsymbol{\Gamma}^{1,0} & \boldsymbol{\Gamma}^{1,1} & \cdots & \boldsymbol{\Gamma}^{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}^{d,0} & \boldsymbol{\Gamma}^{d,1} & \cdots & \boldsymbol{\Gamma}^{d,d} \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_+ = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^{0,1} & \cdots & \boldsymbol{\Sigma}^{0,d} \\ \boldsymbol{\Sigma}^{1,0} & \boldsymbol{\Sigma}^{1,1} & \cdots & \boldsymbol{\Sigma}^{1,d} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}^{d,0} & \boldsymbol{\Sigma}^{d,1} & \cdots & \boldsymbol{\Sigma}^{d,d} \end{pmatrix}, \tag{13}$$

where $\boldsymbol{\Gamma}^{0,k} = (\boldsymbol{\Gamma}^{k,0})^\top$ is a $n \times n$ matrix with the (i, j)th element being $\mathrm{Cov}[\mathrm{M}(\mathbf{x}_i), \partial \mathrm{M}(\mathbf{x}_j)/\partial x_k]$ and $\boldsymbol{\Gamma}^{k,h} = (\boldsymbol{\Gamma}^{h,k})^\top$ is a $n \times n$ matrix with the (i, j)th element being $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{x}_i)/\partial x_k, \partial \mathrm{M}(\mathbf{x}_j)/\partial x_h]$, $\boldsymbol{\Sigma}^{0,k} = \boldsymbol{\Sigma}^{k,0}$ is a $n \times n$ diagonal matrix with

the (i, i)th element being $\mathrm{Cov}[\overline{\varepsilon}(\mathbf{x}_i), \overline{\epsilon}^k(\mathbf{x}_i)]$ and $\mathbf{\Sigma}^{k,h} = \mathbf{\Sigma}^{h,k}$ is a $n \times n$ diagonal matrix with the (i, i)th element being $\mathrm{Cov}[\overline{\epsilon}^k(\mathbf{x}_i), \overline{\epsilon}^h(\mathbf{x}_i)]$, for $k, h = 1, 2, \ldots, d$. Provided that $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\mathbf{\Sigma}_+$ are known and $(\mathbf{\Gamma}_+ + \mathbf{\Sigma}_+)$ is invertible, at any new point $\mathbf{z} \in \mathcal{X}$, the GESK MSE optimal predictor of $V(\mathbf{z})$ is given by

$$\tilde{V}(\mathbf{z}) = \mathbf{f}(\mathbf{z})^\top \boldsymbol{\beta} + \boldsymbol{\gamma}_+(\mathbf{z})^\top (\mathbf{\Gamma}_+ + \mathbf{\Sigma}_+)^{-1} (\overline{\mathbf{Y}}_+ - \mathbf{F}_+ \boldsymbol{\beta}), \tag{14}$$

with

$$\mathrm{MSE}_V^{\mathrm{GESK}}(\mathbf{z}) \triangleq \mathbb{E}[(\tilde{V}(\mathbf{z}) - V(\mathbf{z}))^2] = \mathrm{Var}[\mathrm{M}(\mathbf{z})] - \boldsymbol{\gamma}_+(\mathbf{z})^\top (\mathbf{\Gamma}_+ + \mathbf{\Sigma}_+)^{-1} \boldsymbol{\gamma}_+(\mathbf{z}). \tag{15}$$

GESK does not need to incorporate all the partial derivatives on the d coordinates. When partial derivatives only on some coordinates are incorporated, it is straightforward to modify all the $\overline{\mathbf{Y}}_+$, $\mathbf{F}_+$, $\boldsymbol{\gamma}_+(\mathbf{z})$, $\mathbf{\Gamma}_+$ and $\mathbf{\Sigma}_+$ accordingly. In practice, $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\mathbf{\Sigma}_+$ are unknown. The $\mathbf{\Sigma}_+$ can be estimated with the sample variances of the response samples and gradient samples, and the sample covariances between them. For $\boldsymbol{\beta}$ and $(\tau^2, \boldsymbol{\theta})$, one can simply use the estimate in the SK model, or estimate them in the GESK model with more information. Then the final empirical predictor of $V(\mathbf{z})$ is $\tilde{V}(\mathbf{z})$ in (14) with $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, and $\mathbf{\Sigma}_+$ replaced with their estimators.

The Greek surfaces are constructed by taking partial derivatives on the constructed price surface $\tilde{V}(\mathbf{z})$ with respect to each coordinate of $\mathbf{z}$. Namely, for $k = 1, 2, \ldots, d$, the Greek surface $G^k(\mathbf{z})$ for $\mathbf{z} \in \mathcal{X}$ is constructed using $\partial \tilde{V}(\mathbf{z})/\partial z_k$, which is

$$\frac{\partial}{\partial z_k} \tilde{V}(\mathbf{z}) = \left(\frac{\partial}{\partial z_k} \mathbf{f}(\mathbf{z}) \right)^\top \boldsymbol{\beta} + \left(\frac{\partial}{\partial z_k} \boldsymbol{\gamma}_+(\mathbf{z}) \right)^\top (\mathbf{\Gamma}_+ + \mathbf{\Sigma}_+)^{-1} (\overline{\mathbf{Y}}_+ - \mathbf{F}_+ \boldsymbol{\beta}), \tag{16}$$

where $\partial \boldsymbol{\gamma}_+(\mathbf{z})/\partial z_k$ is well defined with parameters $(\tau^2, \boldsymbol{\theta})$; see the explicit forms in Section EC.5 of the e-companion. Because of the way that the Greek surfaces are constructed, the consistency between the constructed price and Greek surfaces is maintained naturally.

Before we ask what the MSE is when using $\partial \tilde{V}(\mathbf{z})/\partial z_k$ as the estimator of $G^k(\mathbf{z})$, we first point out a very important observation. The GESK is actually a specific type of the so-called cokriging (Cressie 1993), which treats $V(\mathbf{x})$ as the target variable and $\{G^k(\mathbf{x}), k = 1, 2, \ldots, d\}$ as the covariables. However, with the spirit of cokriging, we can also treat $G^k(\mathbf{x})$ as the target variable for any k and $\{V(\mathbf{x}), G^h(\mathbf{x}), h \neq k\}$ as the covariables. In this way, we can derive a cokriging predictor for $G^k(\mathbf{z})$, which is denoted as $\tilde{G}^k(\mathbf{z})$, for $k = 1, 2, \ldots, d$. Following the same derivations in GESK or cokriging, it is not difficult to see that,

$$\tilde{G}^k(\mathbf{z}) = \left(\frac{\partial}{\partial z_k} \mathbf{f}(\mathbf{z}) \right)^\top \boldsymbol{\beta} + \boldsymbol{\gamma}_+^k(\mathbf{z})^\top (\mathbf{\Gamma}_+ + \mathbf{\Sigma}_+)^{-1} (\overline{\mathbf{Y}}_+ - \mathbf{F}_+ \boldsymbol{\beta}),$$

with

$$\mathrm{MSE}_{G^k}^{\mathrm{GESK}}(\mathbf{z}) \triangleq \mathbb{E}[(\tilde{G}^k(\mathbf{z}) - G^k(\mathbf{z}))^2] = \mathrm{Var}\left[\frac{\partial}{\partial z_k} \mathrm{M}(\mathbf{z}) \right] - \boldsymbol{\gamma}_+^k(\mathbf{z})^\top (\mathbf{\Gamma}_+ + \mathbf{\Sigma}_+)^{-1} \boldsymbol{\gamma}_+^k(\mathbf{z}), \tag{17}$$

where $\boldsymbol{\gamma}_+^k(\mathbf{z}) = (\boldsymbol{\gamma}^{k,0}(\mathbf{z})^\top, \boldsymbol{\gamma}^{k,1}(\mathbf{z})^\top, \ldots, \boldsymbol{\gamma}^{k,d}(\mathbf{z})^\top)^\top$, and $\boldsymbol{\gamma}^{k,0}(\mathbf{z})$ is a $n \times 1$ vector with the ith element being $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{z})/\partial z_k, \mathrm{M}(\mathbf{x}_i)]$, and $\boldsymbol{\gamma}^{k,h}(\mathbf{z})$ is a $n \times 1$ vector with the ith element being $\mathrm{Cov}[\partial \mathrm{M}(\mathbf{z})/\partial z_k, \partial \mathrm{M}(\mathbf{x}_i)/\partial x_h]$, for $h = 1, 2, \ldots, d$.

From the explicit forms of $\partial \boldsymbol{\gamma}_+(\mathbf{z})/\partial z_k$ and $\boldsymbol{\gamma}_+^k(\mathbf{z})$ shown in Section EC.5 of the e-companion, it can be concluded that $\partial \boldsymbol{\gamma}_+(\mathbf{z})/\partial z_k = \boldsymbol{\gamma}_+^k(\mathbf{z})$, and thus $\partial \tilde{V}(\mathbf{z})/\partial z_k = \tilde{G}^k(\mathbf{z})$. Moreover,

$$\mathrm{MSE}_{G^k}^{\mathrm{GESK\text{-}D}}(\mathbf{z}) \triangleq \mathbb{E}\left[\left(\frac{\partial}{\partial z_k} \tilde{V}(\mathbf{z}) - G^k(\mathbf{z}) \right)^2 \right] = \mathrm{MSE}_{G^k}^{\mathrm{GESK}}(\mathbf{z}). \tag{18}$$

4.3. Accuracy Analysis

We just introduced how to build the price and Greek surfaces using GESK (i.e., (14) and (16)), which are obviously consistent. Next, we want to show that, even if we put the consistency aside, the price and Greek surfaces constructed using GESK are more accurate than those constructed using separate SK (i.e., (9) and (11)). Here we use the MSE at each predicted point $\mathbf{z}$ to quantify the accuracy, and a smaller MSE indicates higher accuracy. This is to say, we want to compare the MSE of $\hat{V}(\mathbf{z})$ (i.e., $\mathrm{MSE}_V^{\mathrm{SK}}(\mathbf{z})$ given in (10)) with the MSE of $\tilde{V}(\mathbf{z})$ (i.e., $\mathrm{MSE}_V^{\mathrm{GESK}}(\mathbf{z})$ given in (15)), and compare the MSE of $\hat{G}^k(\mathbf{z})$ (i.e., $\mathrm{MSE}_{G^k}^{\mathrm{SK}}(\mathbf{z})$ given in (12)) with the MSE of $\partial \tilde{V}(\mathbf{z})/\partial z_k$ (i.e., $\mathrm{MSE}_{G^k}^{\mathrm{GESK\text{-}D}}(\mathbf{z})$ given in (17) and (18)). It turns out that the constructed surfaces using GESK indeed have smaller prediction MSE, as formally stated in Theorem 3.

Theorem 3. *Suppose that* $\boldsymbol{\beta}$, $(\tau^2, \boldsymbol{\theta})$, *and* $\boldsymbol{\Sigma}_+$ *are known, then the price and Greek surfaces constructed using GESK have smaller MSE than those constructed using separate SK, that is, for any* $\mathbf{z} \in \mathcal{X}$,

$$\text{MSE}_V^{\text{GESK}}(\mathbf{z}) < \text{MSE}_V^{\text{SK}}(\mathbf{z}), \tag{19}$$

and

$$\text{MSE}_{G^k}^{\text{GESK-D}}(\mathbf{z}) < \text{MSE}_{G^k}^{\text{SK}}(\mathbf{z}), \quad \text{for} \quad k = 1, 2, \ldots, d. \tag{20}$$

Before we show the proof, it is worth noting that if we additionally assume both $\varepsilon_\ell(\mathbf{x})$ and $\epsilon_\ell^k(\mathbf{x})$ are normally distributed, for $\ell = 1, 2, \ldots$, and $k = 1, 2, \ldots, d$, then Theorem 3 can be proved with a much simpler approach. Because in such case, all the predictors (i.e., (9), (11), (14), and (16)) can be written as certain conditional expectations and their MSEs (i.e., (10), (12), (15), and (17)) can be written as certain conditional variances. Intuitively speaking, more information results in greater variance reduction.[1] The detailed analysis of this case is provided in Section EC.6 of the e-companion.

For the general case (where the simulation noises follow general distributions), the predictors are not necessarily the conditional means, hence the proof relies on the specific forms of the MSEs. Actually, a result same as (19) for a simplified problem was proved in (Chen et al. 2013, section EC.5). However, their proof relies on a sparsity assumption, that is, they assume that the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are widely spread in the design space $\mathcal{X}$ such that the spatial correlations of the observations (of both the response and the gradient) at distinct design points are approximately zero; that is, $\exp\{-\sum_{k=1}^d \theta_k (x_{ik} - x_{jk})^2\} \approx 0$ for $i \ne j$, $i, j = 1, 2, \ldots$. Such sparsity assumption is quite restrictive, and it may not be satisfied in our problem.

The proof of Theorem 3 for general simulation noises and design points uses a result of Schur complement from Horn and Zhang (2005), which is stated in the following Lemma 1.

Lemma 1 (Theorem 1.12 of Horn and Zhang 2005). *Let* $\mathbf{M}$ *be a symmetric matrix partitioned as*

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{D} \end{pmatrix},$$

where $\mathbf{A}$ *is square and nonsingular. Then* $\mathbf{M}$ *is positive (semi) definite if and only if both* $\mathbf{A}$ *is positive definite and the Schur complement* $\mathbf{M}/\mathbf{A} \triangleq \mathbf{D} - \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B}$ *is positive (semi) definite.*

Now we are ready to prove Theorem 3.

Proof of Theorem 3. We first prove (19). By (10) and (15), to show (19) is equivalent to show

$$\boldsymbol{\gamma}_+(\mathbf{z})^\top (\boldsymbol{\Gamma}_+ + \boldsymbol{\Sigma}_+)^{-1} \boldsymbol{\gamma}_+(\mathbf{z}) > \boldsymbol{\gamma}(\mathbf{z})^\top (\boldsymbol{\Gamma} + \boldsymbol{\Sigma})^{-1} \boldsymbol{\gamma}(\mathbf{z}). \tag{21}$$

Recall the forms of $\boldsymbol{\Gamma}_+$ and $\boldsymbol{\Sigma}_+$ in (13), we can partition them as

$$\boldsymbol{\Gamma}_+ = \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}_B \\ \boldsymbol{\Gamma}_B^\top & \boldsymbol{\Gamma}_D \end{pmatrix}, \quad \boldsymbol{\Sigma}_+ = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}_B \\ \boldsymbol{\Sigma}_B^\top & \boldsymbol{\Sigma}_D \end{pmatrix},$$

where $\boldsymbol{\Gamma}_B$ and $\boldsymbol{\Sigma}_B$ are $n \times nd$ matrices, and $\boldsymbol{\Gamma}_D$ and $\boldsymbol{\Sigma}_D$ are $nd \times nd$ matrices. Moreover, let

$$\boldsymbol{\Lambda}_+ \triangleq \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{\Lambda}_B \\ \boldsymbol{\Lambda}_B^\top & \boldsymbol{\Lambda}_D \end{pmatrix} \triangleq \begin{pmatrix} \boldsymbol{\Gamma} + \boldsymbol{\Sigma} & \boldsymbol{\Gamma}_B + \boldsymbol{\Sigma}_B \\ \boldsymbol{\Gamma}_B^\top + \boldsymbol{\Sigma}_B^\top & \boldsymbol{\Gamma}_D + \boldsymbol{\Sigma}_D \end{pmatrix} = \boldsymbol{\Gamma}_+ + \boldsymbol{\Sigma}_+.$$

Provided that $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_+$ are invertible as assumed in SK and GESK,

$$\boldsymbol{\Lambda}_+^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_B (\boldsymbol{\Lambda}_D - \boldsymbol{\Lambda}_B^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_B)^{-1} \boldsymbol{\Lambda}_B^\top \boldsymbol{\Lambda}^{-1} & -\boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_B (\boldsymbol{\Lambda}_D - \boldsymbol{\Lambda}_B^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_B)^{-1} \\ -(\boldsymbol{\Lambda}_D - \boldsymbol{\Lambda}_B^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_B)^{-1} \boldsymbol{\Lambda}_B^\top \boldsymbol{\Lambda}^{-1} & (\boldsymbol{\Lambda}_D - \boldsymbol{\Lambda}_B^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_B)^{-1} \end{pmatrix}.$$

Also let $\boldsymbol{\gamma}_+(\mathbf{z}) = (\boldsymbol{\gamma}(\mathbf{z})^\top, \boldsymbol{\gamma}^{0,1}(\mathbf{z})^\top, \ldots, \boldsymbol{\gamma}^{0,d}(\mathbf{z})^\top)^\top \triangleq (\boldsymbol{\gamma}(\mathbf{z})^\top, \boldsymbol{\gamma}_B(\mathbf{z})^\top)^\top$. Then,

$$\boldsymbol{\gamma}_+(\mathbf{z})^\top (\boldsymbol{\Gamma}_+ + \boldsymbol{\Sigma}_+)^{-1} \boldsymbol{\gamma}_+(\mathbf{z}) = \boldsymbol{\gamma}_+(\mathbf{z})^\top \boldsymbol{\Lambda}_+^{-1} \boldsymbol{\gamma}_+(\mathbf{z})$$

$$= \boldsymbol{\gamma}(\mathbf{z})^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\gamma}(\mathbf{z}) + \mathbf{q}^\top \mathbf{Q} \mathbf{q}$$

$$= \boldsymbol{\gamma}(\mathbf{z})^\top (\boldsymbol{\Gamma} + \boldsymbol{\Sigma})^{-1} \boldsymbol{\gamma}(\mathbf{z}) + \mathbf{q}^\top \mathbf{Q} \mathbf{q},$$

where $\mathbf{q} \triangleq \mathbf{\Lambda}_B^\top \mathbf{\Lambda}^{-1} \boldsymbol{\gamma}(\mathbf{z}) - \boldsymbol{\gamma}_B(\mathbf{z})$ and $\mathbf{Q} \triangleq (\mathbf{\Lambda}_D - \mathbf{\Lambda}_B^\top \mathbf{\Lambda}^{-1} \mathbf{\Lambda}_B)^{-1}$. By Lemma 1, $\mathbf{Q}^{-1}$ is the Schur complement of $\mathbf{\Lambda}$ in $\mathbf{\Lambda}_+$, and thus is positive definite because $\mathbf{\Lambda}_+$ is positive definite. Therefore, $\mathbf{Q}$ is positive definite, which implies that $\mathbf{q}^\top \mathbf{Q} \mathbf{q}$ is positive. Hence (21) is proved and so is (19).

The proof of (20) can be established similarly. Intuitively, one only needs to note that in this case $G^k(\mathbf{x})$ is treated as the target variable instead of $V(\mathbf{x})$, and it should make no difference in terms of the prediction if we rearrange the vector of estimators $\overline{\mathbf{Y}}_+ = (\overline{\mathbf{Y}}^\top, (\overline{\mathbf{D}}^1)^\top, \dots, (\overline{\mathbf{D}}^d)^\top)^\top$ by putting $(\overline{\mathbf{D}}^k)^\top$ in the first place and rearrange the covariance vector $\boldsymbol{\gamma}_+^k(\mathbf{z})$, covariance matrices $\mathbf{\Gamma}_+$ and $\mathbf{\Sigma}_+$, and basis vector $\mathbf{F}_+$ accordingly. Nevertheless, even if we do not do the rearrangement and keep the same $\overline{\mathbf{Y}}_+$ as before (as we do in (16) and (17)), (20) can still be proved by some matrix transformation and calculation. We leave the details to Section EC.7 of the e-companion. □

To summarize, the price and Greek surfaces constructed using GESK not only possess consistency but also have higher accuracy than those constructed using separate SK. Therefore, the GESK is indeed a desired method for the problem considered in this paper. Moreover, in the next section, we will see that, in some specific cases, the GESK method can be further improved.

5. Further Improvement with PDEs

Besides the differentiation relations between the price and the Greeks at the same design point, there is often another relation between the price and the Greeks; that is, they often satisfy a dynamic equation, typically in the form of a PDE. In this section, we investigate how to integrate the PDE into the GESK as a constraint to further improve the accuracy of the constructed metamodels of the price and the Greeks.

5.1. Feynman-Kac Formula

In the field of PDEs, the well-known Feynman-Kac formula (Øksendal 2010, theorem 8.2.1) states that the expected value with respect to some diffusion process models can be obtained as a solution of an associated PDE. In our setting, suppose the considered $(d-1)$-dimensional market scenario $\mathbf{S}$ follows general diffusion process models; that is, it changes with time t as $\mathbf{S}(t) \triangleq (S_1(t), S_2(t), \dots, S_{d-1}(t))^\top$, with

$$dS_i(t) = \mu_i(\mathbf{S}(t), t)dt + \sum_{j=1}^{d-1} \sigma_{ij}(\mathbf{S}(t), t)dB_j(t), \; i = 1, 2, \dots, d-1, \tag{22}$$

where $\mu_i(\mathbf{s}, t)$ and $\sigma_{ij}(\mathbf{s}, t)$ are known drift and volatility functions defined on $\Re^{d-1} \times [0, T]$, and $B_j(t), j = 1, 2, \dots, d-1$, are independent standard Brownian motions. Let $r(\mathbf{s}, t)$ denote the interest rate function defined on $\Re^{d-1} \times [0, T]$, and $P(\mathbf{s})$ the payoff function at the maturity T defined on $\Re^{d-1}$. Then, the Feynman-Kac formula adapted to our setting is stated in Lemma 2.

Lemma 2 (Feynman-Kac Formula, Øksendal 2010). *Suppose that the market scenario $\mathbf{S}(t)$ follows general diffusion process Models (22), and the derivative price can be expressed as*

$$V(\mathbf{s}, t) = \mathbb{E}\left[\int_t^T e^{-\int_t^\tau r(\mathbf{S}(t), t)dt} h(\mathbf{S}(\tau), \tau)d\tau + e^{-\int_t^T r(\mathbf{S}(t), t)dt} P(\mathbf{S}(T)) \bigg| \mathbf{S}(t) = \mathbf{s} \right], \tag{23}$$

where $h(\mathbf{s}, t)$ is a known residual function. Then, $V(\mathbf{s}, t)$ satisfies the following PDE:

$$\mathcal{L}V \triangleq \frac{\partial V}{\partial t} + \sum_{i=1}^{d-1} \mu_i(\mathbf{s}, t)\frac{\partial V}{\partial s_i} + \frac{1}{2}\sum_{i=1}^{d-1}\sum_{j=1}^{d-1} \gamma_{ij}(\mathbf{s}, t)\frac{\partial^2 V}{\partial s_i \partial s_j} - r(\mathbf{s}, t)V = h(\mathbf{s}, t), \tag{24}$$

with terminal condition $V(\mathbf{s}, T) = P(\mathbf{s})$, where $\gamma_{ij}(\mathbf{s}, t) = \sum_{k=1}^{d-1} \sigma_{ik}(\mathbf{s}, t)\sigma_{jk}(\mathbf{s}, t)$.

Now we present some examples of the derivatives where (22) and (23) are satisfied and thus the Feynman-Kac formula can be applied.

Example 1 (European Call Option). Consider a European call option with an underlying asset whose price $S(t)$ follows a geometric Brownian motion (which is a special case of the diffusion model), given by

$$dS(t) = rS(t)dt + \sigma S(t)dB(t), \tag{25}$$

where r is the risk-free interest rate, that is, $r(\mathbf{s}, t) \equiv r$, and σ is the constant volatility. Here the underlying asset price is the only considered market factor, which clearly satisfies (22) with $\mu(s, t) = rs$ and $\sigma(s, t) = \sigma s$. The payoff function

of the European call option is $P(S(T)) = (S(T) - K)^+$, where K is the strike price and $x^+ \triangleq \max(x, 0)$. Thus, the option price can be expressed as $V(s, t) = \mathbb{E}[e^{-r(T-t)}P(S(T))|S(t) = s]$, which satisfies (23). Then, by applying the Feynman-Kac formula, the PDE governing the option price is given by

$$\frac{\partial V(s, t)}{\partial t} + rs\frac{\partial V(s, t)}{\partial s} + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V(s, t)}{\partial s^2} - rV(s, t) = 0,$$

with terminal condition $V(s, T) = (s - K)^+$.

Example 2 (Asian Call Option). Consider an Asian call option, which is a path-dependent option. The underlying asset price follows the same model (25) in Example 1, and it is also the only market factor that affects the option price. However, to properly express the payoff function, a dummy market factor needs to be constructed. Denote the underlying asset price as $S_1(t)$. Define the dummy market factor $S_2(t) \triangleq \int_0^t S_1(u)\mathrm{d}u$. That is, $\mathrm{d}S_2(t) = S_1(t)\mathrm{d}t$. It can be seen that the two market factors satisfy (22) with $\mu_1(\mathbf{s}, t) = rs_1$, $\mu_2(\mathbf{s}, t) = s_1$, $\sigma_{11}(\mathbf{s}, t) = \sigma s_1$, and $\sigma_{12}(\mathbf{s}, t) = \sigma_{21}(\mathbf{s}, t) = \sigma_{22}(\mathbf{s}, t) = 0$. The payoff function of the Asian call option is $P(\mathbf{S}(T)) = (S_2(T) - K)^+$. Thus, the option price can be expressed as $V(\mathbf{s}, t) = \mathbb{E}[e^{-r(T-t)}P(\mathbf{S}(T))|\mathbf{S}(t) = \mathbf{s}]$, which satisfies (23). Then, by applying the Feynman-Kac formula, the PDE governing the option price is given by

$$\frac{\partial V(\mathbf{s}, t)}{\partial t} + rs_1\frac{\partial V(\mathbf{s}, t)}{\partial s_1} + s_1\frac{\partial V(\mathbf{s}, t)}{\partial s_2} + \frac{1}{2}\sigma^2 s_1^2 \frac{\partial^2 V(\mathbf{s}, t)}{\partial s_1^2} - rV(\mathbf{s}, t) = 0,$$

with terminal condition $V(\mathbf{s}, T) = (s_2 - K)^+$.

Example 3 (Option Portfolio). Consider an option portfolio with M options. The underlying assets of the options included in the portfolio are aggregated and denoted as $\mathbf{S}(t)$. Let $\Phi(\mathbf{S}(t), t)$ be the price of this portfolio at time t, and it can be expressed as $\Phi(\mathbf{S}(t), t) = \sum_{m=1}^M V_m(\mathbf{S}(t), t)$, where $V_m(\mathbf{S}(t), t)$ denotes the price of the mth option included in the portfolio. The mth option may depend on a portion or all of the underlying assets in $\mathbf{S}(t)$. Suppose each option price $V_m(\mathbf{S}(t), t), m = 1, 2, \ldots, M$, follows a PDE

$$\begin{cases} \mathcal{L}V_m(\mathbf{S}(t), t) = 0, \\ V_m(\mathbf{S}(T), T) = P_m(\mathbf{S}(T)), \end{cases} \tag{26}$$

where $P_m(\mathbf{S}(T))$ is the payoff function of the mth option. Consequently, the portfolio price is governed by PDE

$$\mathcal{L}\Phi(\mathbf{S}(t), t) = 0,$$

with terminal condition $\Phi(\mathbf{S}(T), T) = \sum_{m=1}^M P_m(\mathbf{S}(T))$. The proof is provided in Section EC.8 of the e-companion, together with a specific expression of $\mathcal{L}\Phi(\mathbf{S}(t), t)$ when $M = 2$.

For most derivatives, like those in Examples 1–3, $h(\mathbf{s}, t) \equiv 0$ and $r(\mathbf{s}, t) \equiv r$, where r is constant. In such a case, (23) can be simplified to

$$V(\mathbf{s}, t) = \mathbb{E}[e^{-r(T-t)}P(\mathbf{S}(T))|\mathbf{S}(t) = \mathbf{s}], \tag{27}$$

and the PDE (24) is simplified to $\mathcal{L}V = 0$.

These examples show the applicability of the Feynman-Kac formula to a broad category of derivatives, such as European-style financial options, Asian-style financial options, and their portfolios, thus many practical financial derivatives are governed by certain PDEs. As these PDEs also describe additional relations between the price and certain Greeks (e.g., the delta, the gamma, and the theta) in addition to consistency, it should be beneficial to use them in building metamodels for the price and Greeks.

5.2. PDE-Constrained GESK

Inspired by Theorem 3, we expect that incorporating the PDE relation may further improve the accuracy of the metamodels. However, the question is how to incorporate it. We propose to view this PDE as a *constraint* when building the metamodels of the price and Greeks using GESK, and we call it the PDE-constrained GESK, or the PDE-GESK for short.

In the simulation experiments, using appropriate estimation methods, it is possible to ensure that the PDE (24) is satisfied for $(\mathbf{s}^\top, t)^\top \in \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. That is, the estimated price and Greeks on these design points can satisfy this

PDE. However, for the entire price surfaces ($\tilde{V}(\mathbf{z})$ given in (14)) on $\mathcal{X}$, it is not necessarily satisfied. Then, the question is how to build the price surface via GESK, which can satisfy this PDE on $\mathcal{X}$.

Recall that in the original GESK, $\tilde{V}(\mathbf{z}) = \tilde{V}(\mathbf{s},t)$ is finally computed based on the estimates of $\boldsymbol{\beta}$ and $(\tau^2, \boldsymbol{\theta})$ via the MLE, which can be explicitly written as

$$\tilde{V}(\mathbf{s},t;\boldsymbol{\beta},\tau^2,\boldsymbol{\theta}) = \mathbf{f}(\mathbf{z})^\top \boldsymbol{\beta} + \boldsymbol{\gamma}_+(\tau^2,\boldsymbol{\theta})(\mathbf{z})^\top (\boldsymbol{\Gamma}_+(\tau^2,\boldsymbol{\theta}) + \boldsymbol{\Sigma}_+)^{-1} (\overline{\mathbf{Y}}_+ - \mathbf{F}_+\boldsymbol{\beta}).$$

Ideally, we want to choose proper $\boldsymbol{\beta}$, τ^2, and $\boldsymbol{\theta}$ so that the PDE (24) is strictly satisfied. However, this is impossible because $\mathcal{L}\tilde{V}(\mathbf{s},t;\boldsymbol{\beta},\tau^2,\boldsymbol{\theta}) = h(\mathbf{s},t)$ implies an infinite-dimensional equation, which means that we cannot solve the equations for all $(\mathbf{s}^\top,t)^\top \in \mathcal{X}$. Therefore, we settle for second best. The parameter $\boldsymbol{\beta}$ is less important compared with the parameters $(\tau^2, \boldsymbol{\theta})$, because the Gaussian random field $\mathsf{M}(\mathbf{x})$ has higher surface-fitting flexibility than the linear regression $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$ (that is why in practice it is common to set $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$ to zero, which is known as simple kriging, or set $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$ to an unknown constant, which is known as ordinary kriging). Moreover, the parameter τ^2 is less important than the parameter $\boldsymbol{\theta}$ because the effect of τ^2 tends to be canceled out in (14), especially when the simulation noises are small so that $\boldsymbol{\Sigma}_+$ is close to a zero matrix. Therefore, for simplicity, we keep the estimates of $\boldsymbol{\beta}$ and τ^2, which are obtained from the MLE together with the estimate of $\boldsymbol{\theta}$, and then refine the estimate of $\boldsymbol{\theta}$.

In the following analysis, without loss of generality, we consider the simple case where $h(\mathbf{s},t) \equiv 0$ and $r(\mathbf{s},t) \equiv r$ as in the case of (27). To quantify the discrepancy of the PDE constraint, one may naturally consider the integrated mean squared error (IMSE) between $\mathcal{L}\tilde{V}(\mathbf{s},t;\boldsymbol{\theta})$ and $\mathcal{L}V(\mathbf{s},t)$:

$$\text{IMSE}(\boldsymbol{\theta}) = \int_{\mathcal{S}\times[0,T]} (\mathcal{L}\tilde{V}(\mathbf{s},t;\boldsymbol{\theta}) - \mathcal{L}V(\mathbf{s},t))^2 \mathrm{d}(\mathbf{s},t) = \int_{\mathcal{S}\times[0,T]} (\mathcal{L}\tilde{V}(\mathbf{s},t;\boldsymbol{\theta}))^2 \mathrm{d}(\mathbf{s},t), \tag{28}$$

and try to minimize the IMSE. However, the previous IMSE treats all the points $(\mathbf{s}^\top, t)^\top$ in $\mathcal{X}$ equally, that is, it does not consider the distributions of underlying assets $\mathbf{S}(t)$ given that $\mathbf{S}(t) = \mathbf{s}$. Furthermore, such IMSE is not directly related to the accuracy of the price surface that we are concerned about. Therefore, we propose a more sophisticated IMSE, called the weighted IMSE (WIMSE):

$$\text{WIMSE}(\boldsymbol{\theta}) = \int_{\mathcal{S}\times[0,T]} \left(\mathbb{E}\left[\int_t^T e^{-r(T-t)} \mathcal{L}\tilde{V}(\mathbf{S}(\tau),\tau;\boldsymbol{\theta})\mathrm{d}\tau \,\middle|\, \mathbf{S}(t) = \mathbf{s} \right] \right)^2 \mathrm{d}(\mathbf{s},t). \tag{29}$$

The integrand of WIMSE (29) can be viewed as the integrand of IMSE (28) incorporated with the sample path information. Moreover, as revealed in the following Theorem 4, minimizing WIMSE (29) is equivalent to minimizing the IMSE between the price surfaces $\tilde{V}(\mathbf{s},t;\boldsymbol{\theta})$ and $V(\mathbf{s},t)$.

Theorem 4. *Let* $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \text{WIMSE}(\boldsymbol{\theta})$. *Define*

$$\text{IMSE-V}(\boldsymbol{\theta}) = \int_{\mathcal{S}\times[0,T]} (\tilde{V}(\mathbf{s},t;\boldsymbol{\theta}) - V(\mathbf{s},t))^2 \mathrm{d}(\mathbf{s},t).$$

If the derivative price $V(\mathbf{s},t)$ *satisfies* (27), *then* $\text{IMSE-V}(\boldsymbol{\theta}^*) \leq \text{IMSE-V}(\boldsymbol{\theta})$ *for all* $\boldsymbol{\theta} \in \Theta$.

Proof. By Lemma 2, $\tilde{V}(\mathbf{s},t;\boldsymbol{\theta})$ can be expressed as a conditional expected value, that is,

$$\tilde{V}(\mathbf{s},t;\boldsymbol{\theta}) = \mathbb{E}\left[\int_t^T e^{-r(T-t)} \mathcal{L}\tilde{V}(\mathbf{S}(\tau),\tau;\boldsymbol{\theta})\mathrm{d}\tau + e^{-r(T-t)} P(\mathbf{S}(T)) \,\middle|\, \mathbf{S}(t) = \mathbf{s} \right].$$

Because $V(\mathbf{s},t)$ satisfies (27), then,

$$\tilde{V}(\mathbf{s},t;\boldsymbol{\theta}) - V(\mathbf{s},t) = \mathbb{E}\left[\int_t^T e^{-r(T-t)} \mathcal{L}\tilde{V}(\mathbf{S}(\tau),\tau;\boldsymbol{\theta})\mathrm{d}\tau \,\middle|\, \mathbf{S}(t) = \mathbf{s} \right]. \tag{30}$$

Therefore, one can see that for any $\boldsymbol{\theta}$,

$$\text{IMSE-V}(\boldsymbol{\theta}) = \int_{\mathcal{S}\times[0,T]} \left(\mathbb{E}\left[\int_t^T e^{-r(T-t)} \mathcal{L}\tilde{V}(\mathbf{S}(\tau),\tau;\boldsymbol{\theta})d\tau \,\middle|\, \mathbf{S}(t) = \mathbf{s} \right] \right)^2 \mathrm{d}(\mathbf{s},t) = \text{WIMSE}(\boldsymbol{\theta}).$$

Therefore,

$$\text{IMSE-V}(\boldsymbol{\theta}^*) = \text{WIMSE}(\boldsymbol{\theta}^*) \leq \text{WIMSE}(\boldsymbol{\theta}) = \text{IMSE-V}(\boldsymbol{\theta}),$$

for all $\boldsymbol{\theta} \in \Theta$. □

Remark 3. We should notice that Theorem 4 only considers the IMSE of the price surface. It is generally difficult to derive the IMSEs of the Greek surfaces because we cannot apply the Feynman-Kac formula to the Greeks. However, we want to argue that if the price surface of the derivative is more accurate and more in line with the PDE, its Greek surfaces should also tend to be more accurate. This can be observed in the numerical experiments reported in Section 6.3.

Algorithm 1 (Determining Correlation Parameter $\boldsymbol{\theta}$)

    Input: The sampling region $\mathcal{S} \times [0, T]$, the outer sample size M_O, the inner sample size M_I, the number of the discretized points for time interval ℓ.

    1: Generate M_O sample points on $\mathcal{S} \times [0, T]$ via the Sobol sequence, which are denoted by $\{(\mathbf{s}_i, t_i), i = 1, 2, \ldots, M_O\}$;

    2: **for** $i = 1$ to M_O **do**

    3:    Let $\Delta_i = (T - t_i)/\ell$;

    4:    **for** $j = 1$ to M_I **do**

    5:        Generate sample path $\{\mathbf{S}_j(t_i + k\Delta_i), k = 1, 2, \ldots, \ell\}$ from the point $\mathbf{S}_j(t_i) = \mathbf{s}_i$ based on the underlying assets model;

    6:        Approximate the inner integral by

$$\varphi_{i,j}(\boldsymbol{\theta}) = \sum_{k=0}^{\ell} e^{-r\Delta_i \ell} \mathcal{L}\tilde{V}(\mathbf{S}_j(t_i + k\Delta_i), t_i + k\Delta_i, \boldsymbol{\theta})\Delta_i;$$

    7:    **end for**

    8: **end for**

    9: Define the SAA objective function by

$$\iota(\boldsymbol{\theta}) = \frac{1}{M_O} \sum_{i=1}^{M_O} \left(\frac{1}{M_I} \sum_{j=1}^{M_I} \varphi_{i,j}(\boldsymbol{\theta}) \right)^2;$$

    10: Solve $\min_{\boldsymbol{\theta}} \iota(\boldsymbol{\theta})$ via a deterministic optimization method, and obtain the solution $\hat{\boldsymbol{\theta}}^*$;

    Output: Return $\hat{\boldsymbol{\theta}}^*$.

Although $\mathcal{L}\tilde{V}(\mathbf{s}, t; \boldsymbol{\theta})$ has a closed-form expression, deriving closed-form expression of WIMSE$(\boldsymbol{\theta})$ given by (29) can be challenging. We propose to use the sample average approximation (SAA) approach to determine the optimal parameter $\boldsymbol{\theta}^*$. The pseudocode is provided in Algorithm 1. First, we generate samples $\{(\mathbf{s}_i, t_i), i = 1, 2, \ldots, M_O\}$ on $\mathcal{S} \times [0, T]$ to approximate the outer integral of the objective function (29), where M_O is the outer sample size. Second, we address the integrand of the outer integral, which refers to the conditional expectation in (29). Specifically, by setting $(\mathbf{s}_i, t_i)$ as an initial point (i.e., the sample path of the underlying asset begins from $\mathbf{S}(t_i) = \mathbf{s}_i$), we generate M_I (named inner sample size) sample paths $\{\mathbf{S}_j(t_i + k\Delta_i), j = 1, 2, \ldots, M_I, k = 1, 2, \ldots, \ell\}$, where $\Delta_i = (T - t_i)/\ell$ and ℓ is the number of the discretized points for time interval $[t_i, T]$. We proceed to approximate the inner integral by summing $\exp(-r\Delta_i \ell) \mathcal{L}\tilde{V}(\mathbf{S}_j(t_i + k\Delta_i), t_i + k\Delta_i, \boldsymbol{\theta}), k = 1, 2, \ldots, \ell$ along each sample path. Third, we approximate the conditional expectation within the integrand of the outer integral by using the sample means. This enables us to subsequently approximate the outer integral, thereby estimating the objective function (29). Certain dimensions of the space $\mathcal{S}$ might be theoretically unbounded. For example, the stock price changes within $[0, \infty)$. In practice, we could truncate such dimensions by applying reasonable constraints. Furthermore, in Step 1, we use the Sobol sequence for generating outer sample points, which enhances the accuracy of the outer integral estimation.

6. Numerical Experiments

In this section, we conduct numerical experiments to investigate the quality of the price and Greek surfaces constructed using the GESK and PDE-constrained GESK and make a comparison with those constructed using the separate SK. For ease of presentation, we simply denote the results obtained using the separate SK, GESK, and PDE-constrained GESK as SK, GESK, and PDE-GESK, respectively, when there is no ambiguity. In all the following experiments, we set $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0$ as an unknown constant; that is, we use ordinary kriging. According to Jones et al.

(1998) and Loeppky et al. (2009), we choose the number of design points as 10 times the number of varying market parameters (i.e., the dimensionality of **x**). Moreover, the design points are randomly generated in $\mathcal{X}$ using the Latin hypercube sampling (LHS) method. All the experiments are implemented in Matlab, and the Matlab built-in function "lhsdesign" is directly used for LHS. All the codes that produced the numerical results are available at a GitHub repository (Jiang et al. 2024).

6.1. Black-Scholes Model

We first consider a toy example, where the Black-Scholes model is applicable to evaluate and compare the performance of the GESK method and separate SK method. More specifically, we focus on the accuracy of the constructed surfaces using different methods, and the hedging effect when the consistency is or is not maintained.

6.1.1. Accuracy of Constructed Surfaces. Suppose a company sells a European call option with an underlying stock whose price is driven by a geometric Brownian motion. The analytical formula of the option price is known as the Black-Scholes formula, and the Greek formulae can be obtained by taking partial derivatives directly. Such true price and Greeks are used to evaluate the accuracy of the constructed price and Greeks.

Firstly, we consider the marginal price and Greek surfaces, that is, only change one market parameter and keep others fixed. Specifically, suppose the baseline setting is as follows: The strike price is $K = 105$, the maturity time is $T = 1$ year, and the initial stock price is $S(0) = 100$, the volatility is $\sigma = 0.2$, and the interest rate is $r = 0.02$. The option price and delta (i.e., partial derivative of price with respect to S) surfaces when S varies in the range $[80, 120]$ are shown in Figure 1; the option price and vega (i.e., partial derivative of price with respect to σ) surfaces when σ varies in the range $[0.01, 0.3]$ are shown in Figure 2; the option price and rho (i.e., partial derivative of price with respect to r) surfaces when r varies in the range $[0.001, 0.10]$ are shown in Figure 3; and the option price and theta (i.e., the negative partial derivative of price with respect to T) surfaces when T varies in the range $[0.01, 2]$ (equivalent to fixing T and varying t) are shown in Figure 4. For each case, we randomly select 10 design points (i.e., scenarios) using LHS. We run Monte Carlo simulation to estimate the prices and corresponding Greeks (using IPA), which are denoted with cross in the figures. From Figures 1–4, it is clear that the constructed price and Greek surfaces using the GESK are more accurate than those from separate SK. Especially for the Greek surfaces, the constructed Greek surfaces using GESK are almost as good as the true ones, whereas those from separate SK are not satisfying. We also investigate the Greek surfaces obtained by directly taking derivatives on the price surface constructed by SK, which are denoted as SK-D. The Greek surfaces obtained in this way may outperform those constructed using separate SK sometimes in some regions of market parameters, but overall the quality is very poor and unstable. These results clearly reveal the superiority of GESK in constructing accurate price and Greek surfaces.

Second, we consider the price and Greek surfaces when all market parameters can vary at the same time. We again use the LHS method to randomly sample 40 design points with $\mathbf{x} \triangleq (S, \sigma, r, T)^\top \in [80, 120] \times [0.01, 0.3] \times [0.001, 0.1] \times [0.01, 2]$, run simulations, and construct the price and Greek surfaces using GESK and separate SK, respectively. Because it is not convenient to visualize such high-dimensional surfaces, we randomly sample 1,000 testing points in the domain using the Sobol sequence and calculate the root mean square error (RMSE), where the true values on the testing points are given by the analytical formulae. We replicate the entire procedure for 50 times

Figure 1. (Color online) Price (Left) and delta (Right) Surfaces for $S_0 \in [80, 120]$
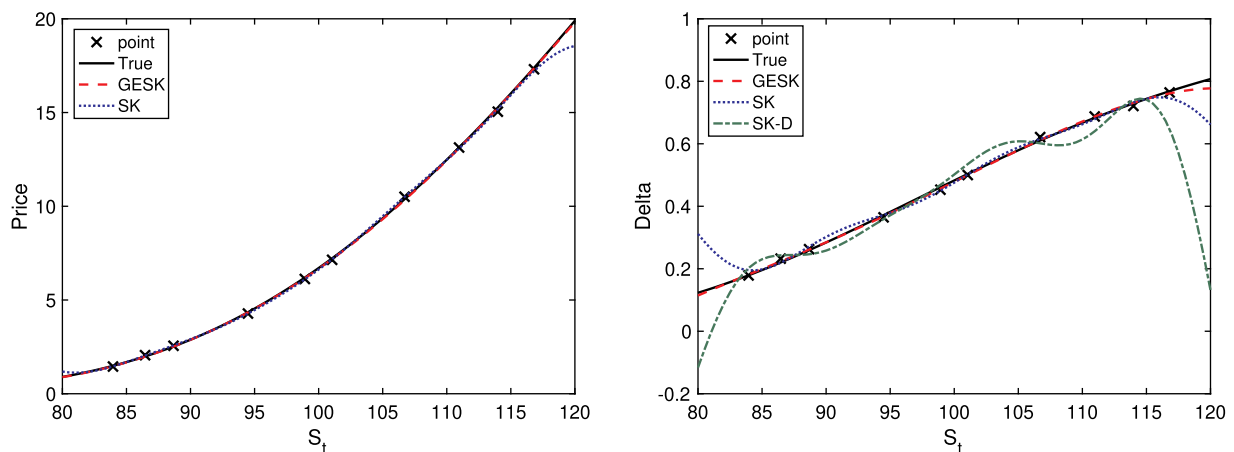
Figure 2. (Color online) Price (Left) and vega (Right) Surfaces for $\sigma \in [0.01, 0.3]$
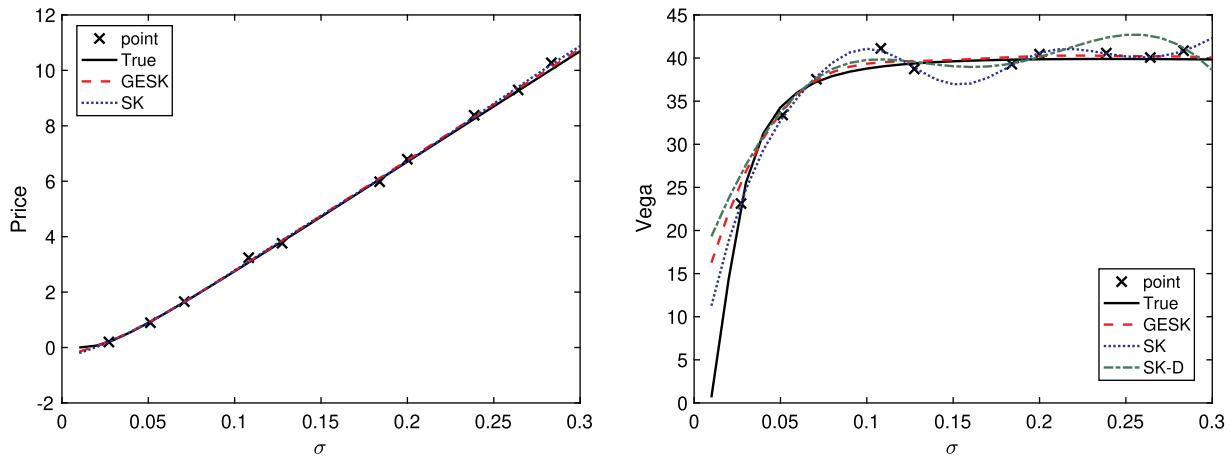

and report the boxplots of the RMSE for the price, delta, vega, rho, and theta in Figure 5. Similar to the results of marginal surfaces, the GESK method significantly outperforms the separate SK method in terms of the accuracy of constructed price and Greek surfaces.

We also try to decompose the MSE into bias and standard deviation, and their boxplots for the price, delta, vega, rho, and theta are collected in Section EC.9 of the e-companion. Numerical results show that compared with separate SK, GESK has smaller bias (in terms of amplitude) and smaller standard deviation, which are consistent to the observed MSE reduction. How to theoretically analyze and compare the bias and standard deviation (or variance) separately is still an open question and worthy of further research. Moreover, to see the effect of different LHS approaches, we try the minimax LHS, maximin LHS (Morris and Mitchell 1995, Hou and Lu 2018), and the Matlab built-in function "lhsdesign" with correlation-reducing criterion, in addition to "lhsdesign" with default setting, of which the details and results are included in Section EC.10 of the e-companion. Because no significant differences are observed, the Matlab built-in function "lhsdesign" with default setting is used throughout the following experiments. However, it is important to notice that good performance of "lhsdesign" with default setting in our examples does not necessarily guarantee good performance in other applications. As pointed by Ye (1998) and Cioppa and Lucas (2007), high correlations between each pair of components or bad spacing filling in LHS design may result in poor metamodel fitting. Therefore, in practice one should always be careful in selecting an appropriate LHS approach.

6.1.2. Hedging Effect with or Without Consistency. We now illustrate the benefit of consistency from the hedging effect and hedging cost via numerical experiments. Consider the delta hedging strategy with maturity $T = 1$ year. Specifically, the portfolio shorts 1,000 European call options and longs $1000 \times$ delta units of stock to hedge the risk.

Figure 3. (Color online) Price (Left) and rho (Right) Surfaces for $r \in [0.001, 0.1]$
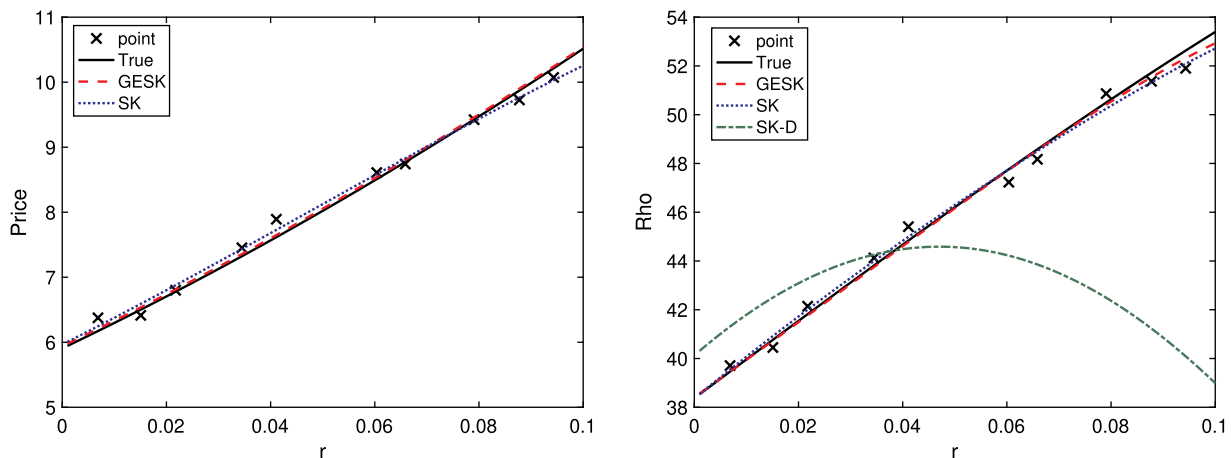
Figure 4. (Color online) Price (Left) and theta (Right) Surfaces for $T \in [0.01, 2]$
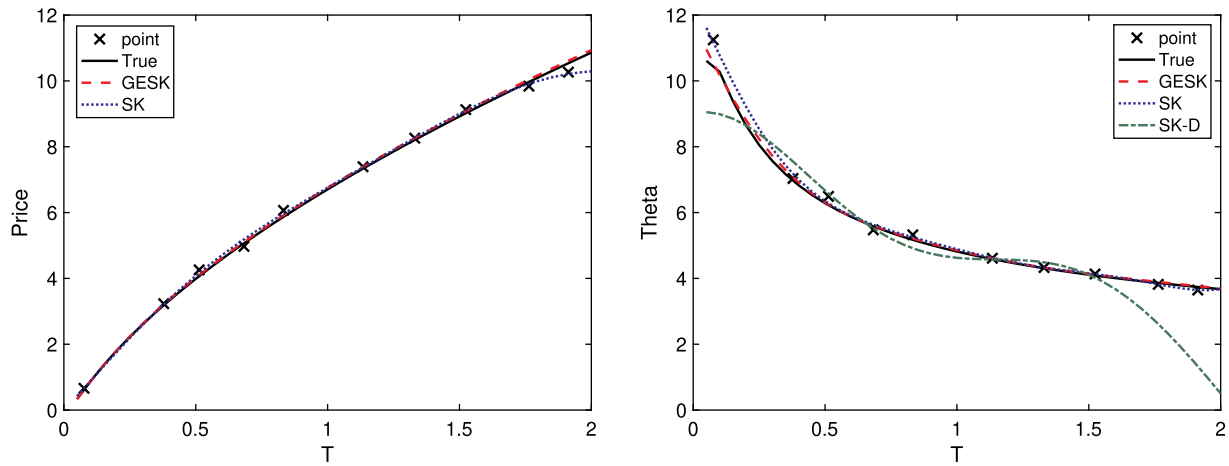

We generate three-month stock price paths while fixing (σ, r) and adjust the stock position once a day during the 21×3 trading days according to the delta under specific stock price. We record the P&L of this portfolio, for the situations where the price and delta are from the analytical Black-Scholes formula, and from the constructed surfaces using GESK or separate SK (based on the simulation on the design points of stock price according to the closing stock price on previous trading day), respectively.
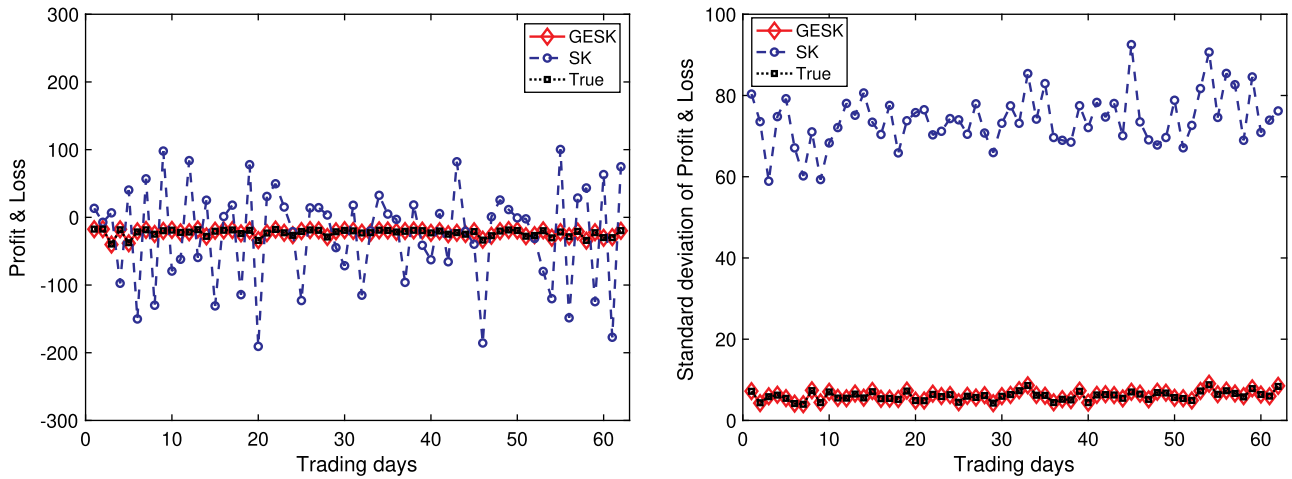
Figure 6 shows that, for different situations, the P&L under one specific stock price path and the standard deviation of P&L over 100 stock price paths. It can be seen that with consistency, the price and Greeks surfaces constructed using GESK produce a very nice hedging performance, which is almost the same as that produced by the Black-Scholes formula. The P&L is close to zero and remains stable over all stock price paths during the entire period, which will be extremely appreciated by companies in practice, because it ensures a stable balance sheet of the portfolio. As a comparison, the price and Greek surfaces constructed using separate SK produce heavily volatile P&L, whose standard deviation is high during the entire period and thus is unsatisfactory. It is worth mentioning that from Figure 1, we can see that the price and delta surfaces constructed by separate SK are actually not too bad in terms of the individual accuracy. However, because of the absence of consistency, the hedging effect can be much worse than expected.

Moreover, we consider the cumulative hedging costs when the GESK method or separate SK method is used with a transaction cost ratio $d = 1o$. Recall that in Theorem EC.1, the hedging costs are compared under the same hedging effect. Therefore, for the separate SK method, additional hedging is conducted accordingly to ensure that the resulting hedging effect is the same as that produced by the GESK method. The result is denoted as SK same effect. In addition, we also calculate the hedging cost when the separate SK method is used without additional hedging (which means the hedging effect is worse than the GESK method). The result is denoted as SK same

Figure 5. (Color online) Boxplots of RMSE for Price, delta, vega, rho, and theta Surfaces
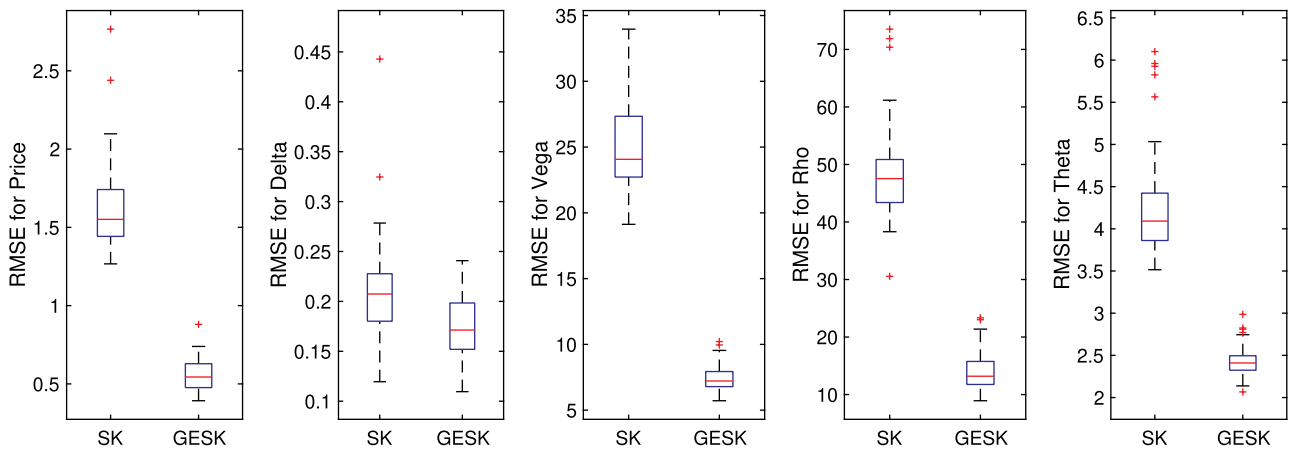
Figure 6. (Color online) P&L Under One Specific Stock Price Path (Left) and Standard Deviation over 100 Stock Price Paths (Right)

frequency. The boxplots of cumulative hedging costs for different situations over 100 stock price paths are shown in Figure 7. It can be seen that even under the same hedging frequency, with the price and Greek surfaces constructed by GESK that possess consistency, the incurred hedging cost is significantly smaller than that incurred with the surfaces constructed by separate SK. For separate SK, if additional hedging is conducted to ensure the same hedging effect as in GESK, the incurred hedging cost will be further increased, which forms a sharper contrast with the cost of GESK.

6.2. Variance-Gamma Process Model

We now consider a more realistic example that may be encountered in practice. Suppose a company holds a portfolio with five Asian options and five lookback options, based on five different stocks: (1) Apple, Inc.; (2) Facebook, Inc.; (3) Netflix, Inc.; (4) Alibaba Group; and (5) Tesla, Inc. The value of this portfolio is given by

$$\Phi = \sum_{i=1}^{5} V_i^A + V_i^L,$$

where V_i^A and V_i^L are the Asian option price and lookback option price based on stock i, respectively. The price of stock i, $S_i(t)$, is modeled by the exponential variance-gamma (VG) process:

$$S_i(t) = S_i(0) \exp((r - \phi_i)t + X_i^{VG}(t)),$$

where $X_i^{VG}(t)$ is a VG process with parameters $(\sigma_i, v_i, \theta_i)$, and $\phi_i = -1/v_i \log(1 - \theta_i v_i - \sigma_i^2 v_i/2)$ is a compensation such that $\mathbb{E}[S_i(t)] = S_i(0) \exp(rt)$, $i = 1, 2, \ldots, 5$. We consider the partial-time arithmetic average Asian call option

Figure 7. (Color online) Cumulative Hedging Costs Under the Same Hedging Effect and Same Hedging Frequency
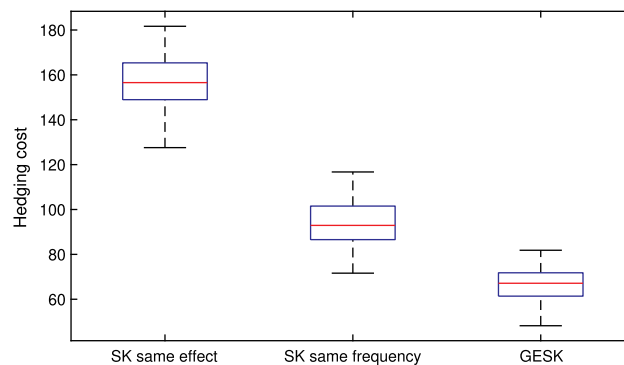
Table 1. Underlying Stocks' Closing Prices, Dividends, and Calibrated Parameters

	Apple, Inc.	Facebook, Inc.	Netflix, Inc.	Alibaba	Tesla, Inc.
$S(0)$	204.47	144.96	113.47	144.85	120.51
Yield	1.21%	0	0	0	0
σ	0.2636	0.2625	0.4012	0.2842	0.4660
ν	0.0387	0.0355	0.0394	0.0017	0.0933
θ	−0.5185	−0.8288	−1.2344	−2.6984	−1.1459

and the partial-time fixed strike lookback call option. For both the Asian option and lookback option, let the maturity be $T = 1$ year, and the partial time be from 0.5 to 1 (i.e., the last six months), during which the stock price is observed once per week.

Based on data from Yahoo Finance November 9, 2018, we observe that, for each stock, the closing price S and yield are as listed in Table 1. Moreover, the parameters (σ, ν, θ) for each stock are calibrated by the corresponding European options' trading prices in Chicago Board Options Exchange (see more details in Section EC.11 of the e-companion), and the calibrated values are also listed in Table 1. For each stock, let the strike prices of its corresponding Asian option and lookback option be the same. Specifically, the strike prices of options based on the five stocks are $(204, 140, 110, 150, 125)$. The risk-free interest rate is 2.73% (one-year U.S. treasury yield from the data).

We first consider the price and Greek surfaces of the portfolio at the initial time. For the options based on stock $i = 1, \ldots, 5$, we consider their price surfaces at the initial time over $(S_i(0), \sigma_i, \theta_i)$. Then, the price surface of the portfolio at the initial time is the summation of all options' price surfaces. We also consider the following Greek surfaces of the portfolio: $\partial \Phi / \partial S_i(0)$ (delta), $\partial \Phi / \partial \sigma_i$, $\partial \Phi / \partial \theta_i$, for $i = 1, \ldots, 5$. To compare the performance of surface construction by the GESK method and separate SK method, for each stock, we randomly sample 30 points in the ranges of $\pm 30\%$ of the current parameter values using LHS and run Monte Carlo simulations to estimate prices and corresponding Greeks for the associated options. To estimate the option price, we use the time-changed Brownian motion method (Schoutens 2003) to generate a sample path of $S_i(t)$, and the Greeks can be estimated via the IPA (see more details in Section EC.12 of the e-companion). Given all the estimates on the design points, we then construct the price and Greek surfaces for each option using GESK and separate SK, respectively, which finally gives the constructed price and Greek surfaces of the portfolio. To evaluate the accuracy, we sample 100 testing points for each stock in the aforementioned parameter ranges using a Sobol sequence, on which the "true" values of prices and Greeks of the Asian option and lookback option are obtained via extensive simulation (10^6 paths) because the analytical formulae are unavailable. Then, for the price and Greek surfaces of the portfolio constructed using GESK and separate SK, respectively, we calculate the RMSE over all testing points. After replicating the entire procedure for 50 times, the boxplots of the RMSE for Greeks are shown in Figure 8. All the results show that the GESK method significantly outperforms the separate SK method in terms of the accuracy of the Greek surfaces construction.

We also examined performance for delta hedging. Suppose the company holds one unit of the previous portfolio and shorts delta units of each stock to hedge the risk. Like in Section 6.1.2, we generate three-month stock price paths for all five stocks under the current (σ, ν, θ) and adjust the position of stock i once a day during the 21×3 trading days according to $\partial \Phi / \partial S_i(t)$, $i = 1, \ldots, 5$. We record the P&L for the situation where the price and delta surfaces are constructed using the GESK method and separate SK method, respectively, based on the simulation on the design points, which are combinations of the five stock prices that randomly sampled in the ranges of 5% of the closing prices on the previous trading day using LHS. Figure 9 shows that, for different situations, the P&L under one specific stock price path and the standard deviation of P&L over 50 stock price paths. It can be seen that with consistency, the GESK method produces a very nice hedging performance. The P&L is close to zero and remains stable over all stock price paths during the entire period. As a comparison, the separate SK method produces heavily volatile P&L, whose standard deviation is high over all stock price paths during the entire period, and thus is unsatisfactory.

6.3. Portfolio of Options

In this section, we compare the performances of GESK and PDE-GESK. Specifically, we examine two types of portfolio of options.

6.3.1. Portfolio of European Options.
We first examine a portfolio with 40 European options, where the underlying asset of each European option is driven by geometric Brownian motion. Therefore, we know the analytical formula

Figure 8. (Color online) Boxplots of RMSE for $\partial\Phi/\partial S_{0i}$ (delta), $\partial\Phi/\partial\sigma_i$, $\partial\Phi/\partial\theta_i$, $i = 1,\ldots,5$
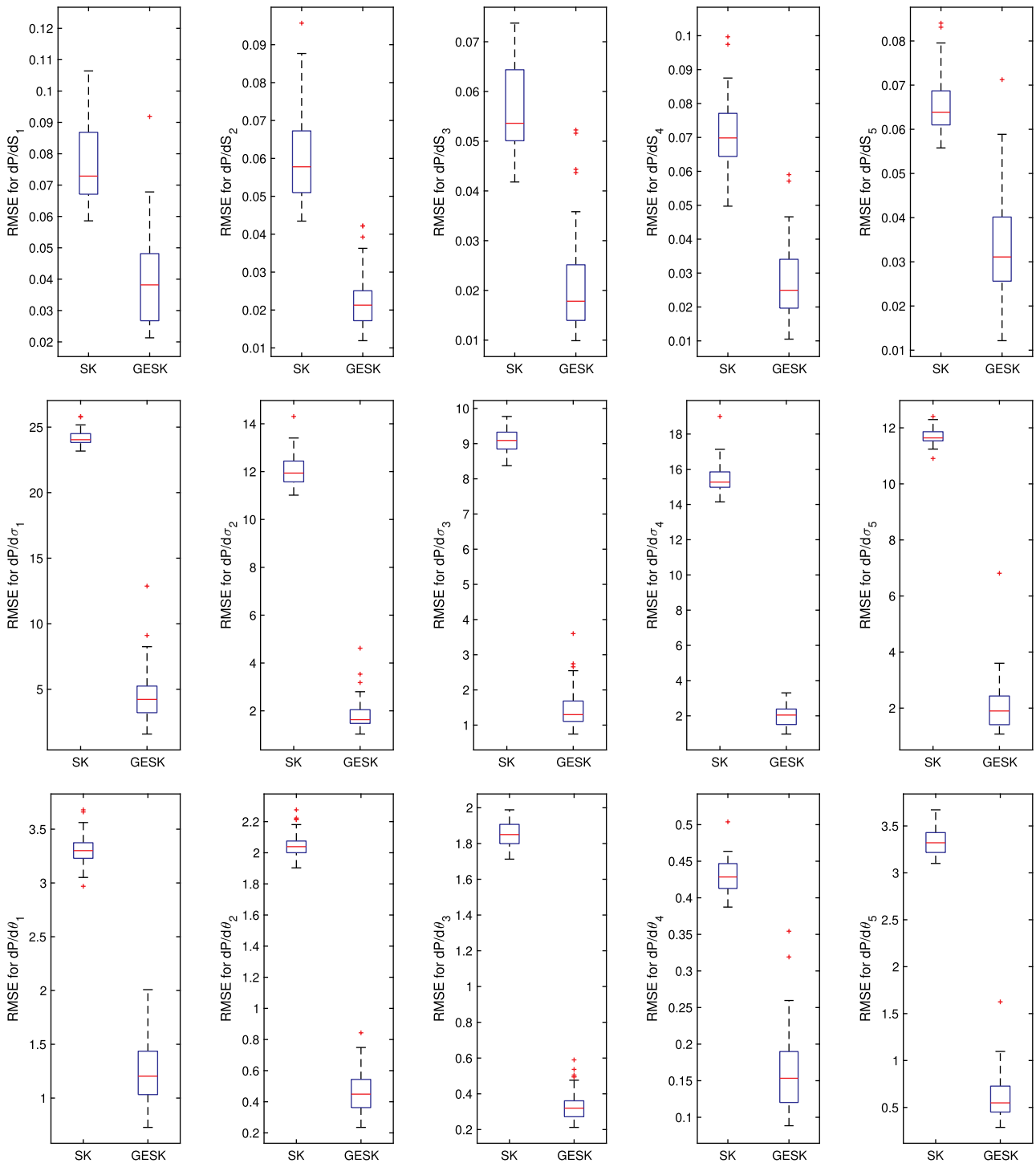
of the price and Greeks of this portfolio. Let the risk-free interest rate be $r = 0.02$ and let the maturity and volatility of all the options be $T = 1$ and $\sigma = 0.2$, respectively. For $i = 1, 2\ldots, 40$, the initial price of the underlying asset and the strike price for the ith option are $S_i(0) = 100 + i$ and $K_i = 79 + 2i$, respectively. To implement PDE-GESK, we set $M_O = 1{,}000, M_I = 100$, and $\ell = 10$ in Algorithm 1.

For each $i = 1, 2, \ldots, 40$, we randomly sample 20 design points on the space $[0.8S_i, 1.2S_i] \times [0, T]$ using LHS and use the price and the delta on these design points to conduct GESK. We also sample 500 testing points on the same

Figure 9. (Color online) P&L Under One Specific Stock Price Path (Left) and Standard Deviation over 50 Stock Price Paths (Right)
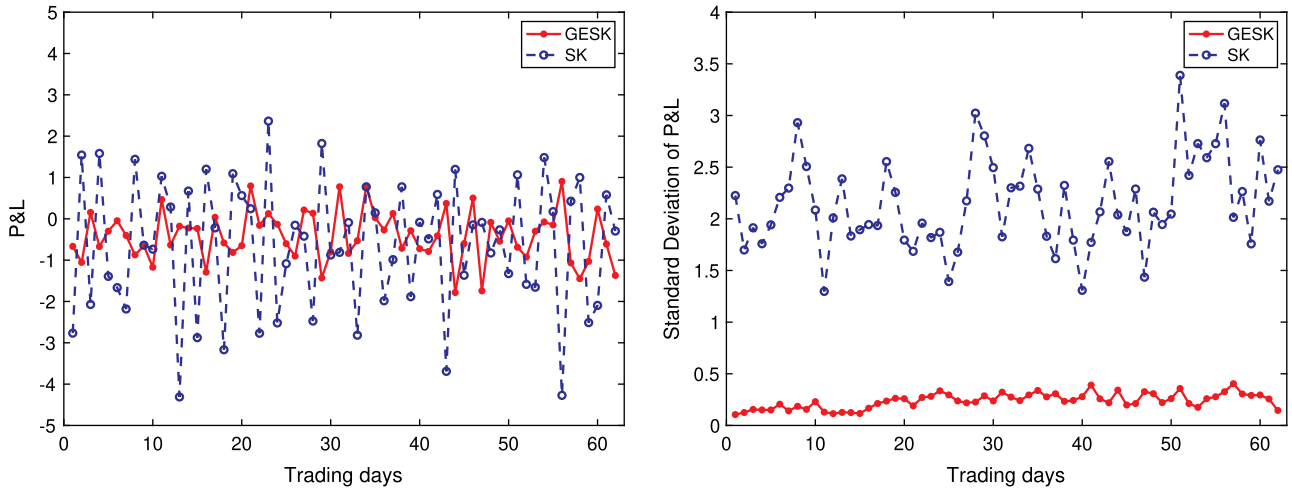

space using a Sobol sequence for each option to calculate the RMSE for price and each element of delta vector. For ease of presentation, we compute the average of the RMSEs of all elements in delta. After replicating the entire procedure for 20 times, boxplots of the RMSEs are reported in Figure 10, which indicates that the PDE-GESK is better than the GESK in constructing the price and the delta surfaces.

Furthermore, we consider another two settings of strike prices, that is, $K_i = 90 + i$ (denoted as Case 2) and $K_i = 110 + i$ (denoted as Case 3), to compare with the original setting (denoted as Case 1). The RMSE of price and averaged delta are summarized in Table 2, which also indicates the superiority of PDE-GESK in constructing the price and delta surfaces.

6.3.2. Portfolio of Basket Options. We examine a portfolio consisting of 240 basket options, each of which consists of two underlying assets driven by geometric Brownian motions. For $i = 1, 2, \ldots, 40$, let $V_i \triangleq V_i(S_{i,1}(t), S_{i,2}(t), t)$ be the price of the ith European call basket option, where $S_{i,1}(t)$ and $S_{i,2}(t)$ are the underlying assets with volatilities $\sigma_{i,1}$ and $\sigma_{i,2}$, respectively, and with correlation ρ_i. For $i = 1, 2, \ldots, 40$, the portfolio involves the following positions: (i) shorting two European call basket options V_i with strike price $K_{i,1}$; (ii) longing one European call basket option V_i with strike price $K_{i,2}$; (iii) longing one European call option V_i with strike price $K_{i,3}$; (iv) shorting two European call options V_i with strike price $K_{i,4}$; (v) longing one European call option V_i with strike price $K_{i,5}$; and (vi) longing one European call option V_i with strike price $K_{i,6}$. The values of these parameters are specified in Table vi in Section

Figure 10. (Color online) Boxplots of RMSE for Price and Averaged RMSE for delta with GESK and PDE-GESK in Portfolio of European Options
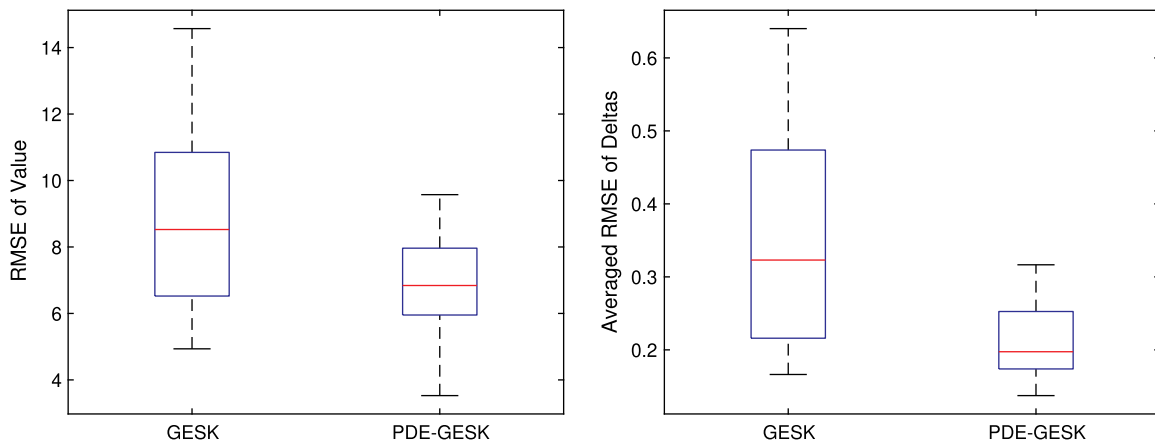
Table 2. Mean and Standard Deviation of RMSE for Price and Averaged RMSE for delta with GESK and PDE-GESK in Portfolio of European Options

| RMSE	Case 1		Case 2		Case 3	
	Price	delta	Price	delta	Price	delta
GESK						
Mean	8.8103	0.3427	13.7781	0.5657	7.2097	0.2565
Standard deviation	2.7489	0.1377	4.6751	0.2311	2.0254	0.0849
PDE-GESK						
Mean	6.8386	0.2145	9.3995	0.2690	6.4630	0.2078
Standard deviation	1.5407	0.0556	2.7119	0.0937	1.9066	0.0357

EC.13 of the e-companion. The risk-free interest rate is $r = 0.02$, and the maturity of all the options is $T = 1$ year. To implement PDE-GESK, we set $M_O = 1,000$, $M_I = 100$, and $\ell = 10$. For both GESK and PDE-GESK, we only use the price and delta for surface construction. That is, on each design point, we only have the estimated price and delta, based on which the price and Greek surfaces are constructed.

For each $i = 1, 2, \ldots, 40$, randomly sample 30 design points on the space $[0.5S_{i,1}, 1.5S_{i,1}] \times [0.5S_{i,2}, 1.5S_{i,2}] \times [0, T]$ using LHS. Randomly sample 500 testing points on the same space using a Sobol sequence for each option to calculate the RMSE for price, theta, and each element of delta vector and gamma matrix, where the true values are given by the analytical formulae. For ease of presentation, we compute the average of the RMSEs of all elements in delta and gamma, respectively. After replicating the entire procedure for 20 times, boxplots of the RMSEs can be obtained. We also consider another setting, where the only difference is that the number of design pints is 50 instead of 30. The results of GESK and PDE-GESK methods are given in Figures 11 and 12 and Table 3. We have the following two observations. (i) The PDE-GESK is better than the GESK in terms of the accuracy on the price and Greek surfaces. (ii) Although the gamma and theta information are not available (i.e., not estimated on the design points) during the surface construction, imposing the PDE constraint can still make the constructed surfaces of such Greeks more accurate.

7. Conclusion

In this paper, we introduce a kriging-based metamodeling method to build consistent surfaces of the price and Greeks for financial derivatives. We not only theoretically prove the importance of consistency in ensuring the hedging effect and maintaining a stable balance sheet but also prove that the proposed method can deliver more accurate price and Greeks prediction than some other plain methods. Furthermore, we propose to use the PDE information as a constraint to further improve the accuracy of the constructed price and Greek surfaces. These theoretical results are all well demonstrated using numerical examples.

Figure 11. (Color online) Boxplots of RMSE for Price, Averaged RMSEs for delta, gamma, and theta with GESK and PDE-GESK Under 30 Design Points in Portfolio of Basket Options
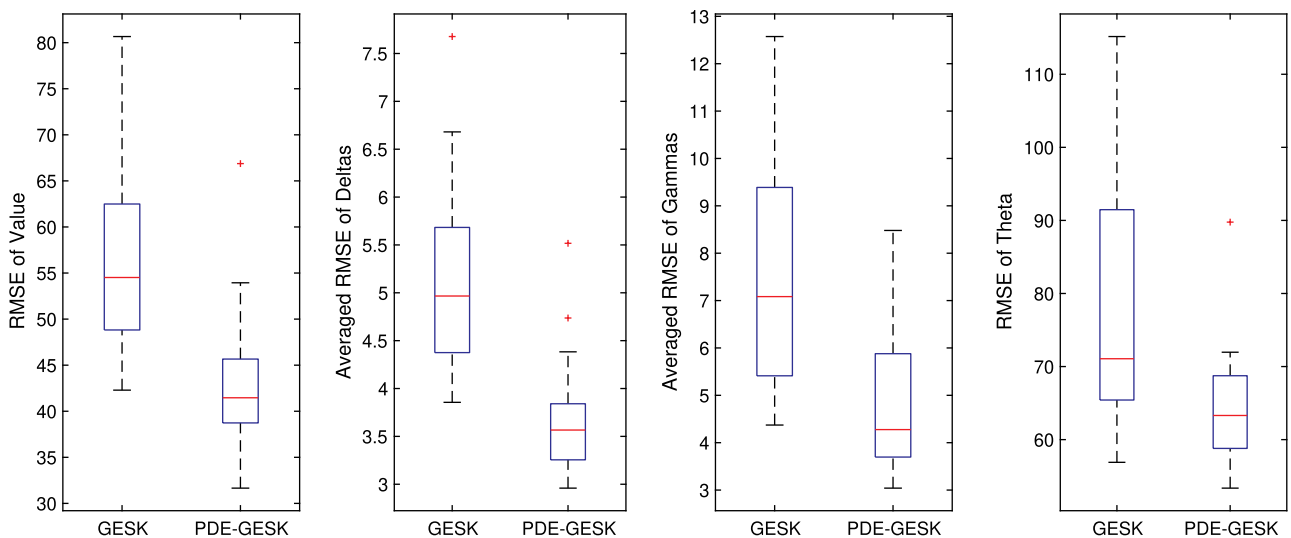
Figure 12. (Color online) Boxplots of RMSE for Price, Averaged RMSEs for delta, gamma, and theta with GESK and PDE-GESK Under 50 Design Points in Portfolio of Basket Options
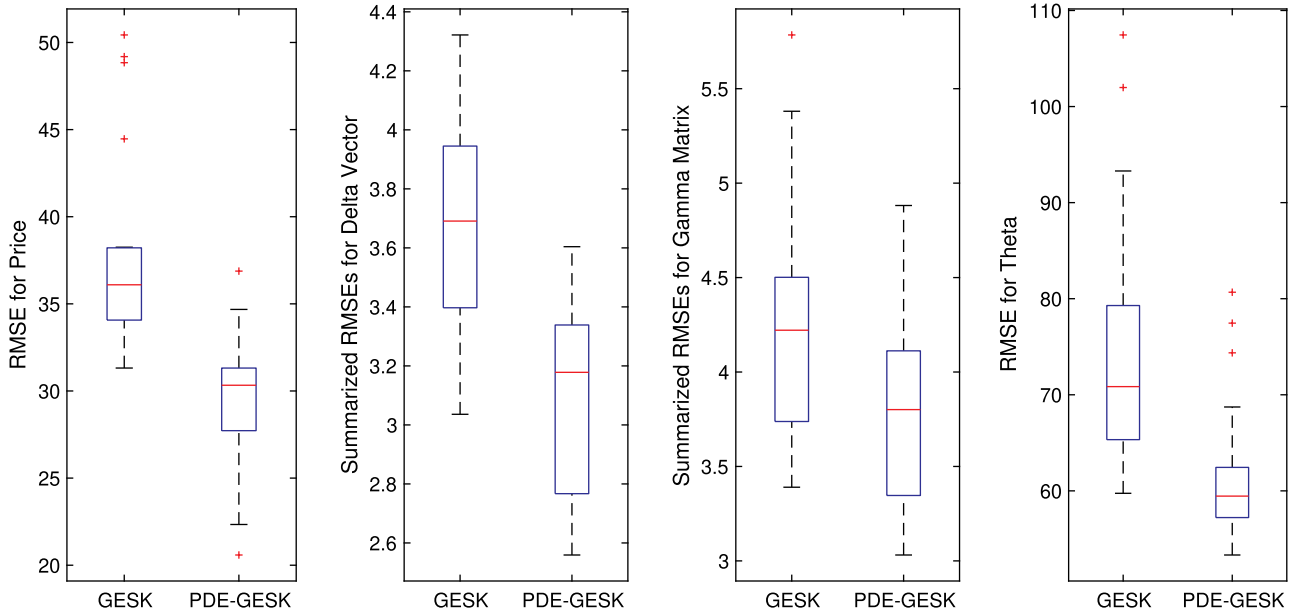
Table 3. Mean and Standard Deviation of RMSE for Price and Averaged RMSEs for delta, gamma, and theta with GESK and PDE-GESK in Portfolio of Basket Options

	30 design points				50 design points			
RMSE	Price	delta	gamma	theta	Price	delta	gamma	theta
GESK								
Mean	56.4737	5.1480	7.4114	78.4669	37.7718	3.6670	4.2461	74.7602
Standard deviation	10.4472	0.9863	2.3727	17.9559	5.8072	0.3612	0.6363	13.3145
PDE-GESK								
Mean	42.4604	3.6951	4.8113	64.5701	29.6433	3.1137	3.8571	62.0852
Standard deviation	7.9227	0.6222	1.5652	8.1093	3.7511	0.3285	0.5960	7.4612

There are several interesting directions for future research. One potential direction is to develop a sequential approach for selecting design points, which could enhance the accuracy and efficiency of surfaces. Another possible direction is to extend the PDE constrained framework to the integro-differential equation constrained framework to tackle the problem with the Lévy process underlying asset models. In addition, it may be worthwhile to consider other metamodeling techniques, such as tree-based methods or neural networks, to construct the price and Greek surfaces. These alternative techniques may offer higher accuracy or other desirable properties beyond consistency.

Endnote

[1] We acknowledge an anonymous reviewer for pointing out this fact.

References

Ankenman B, Nelson BL, Staum J (2010) Stochastic kriging for simulation metamodeling. *Oper. Res.* 58:371–382.

Barton RR (2015) Tutorial: Simulation metamodeling. Yilmaz L, Chan WKV, Moon I, Roeder TMK, Macal C, Rossetti MD, eds. *Proc. Winter Simulation Conf.* (IEEE, Piscataway, NJ), 1765–1779.

Black F, Scholes M (1973) The pricing of options and corporate liabilities. *J. Political Econom.* 81:371–382.

Broadie M, Kaya O (2006) Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.* 54(2):217–231.

Cai N, Song Y, Chen N (2017) Exact simulation of the SABR model. *Oper. Res.* 65(4):931–951.

Chen X, Ankenman B, Nelson BL (2013) Enhancing stochastic kriging metamodels with gradient estimators. *Oper. Res.* 61:512–528.

Cioppa TM, Lucas TW (2007) Efficient nearly orthogonal and space-filling latin hypercubes. *Technometrics* 49(1):45–55.

Cressie NAC (1993) *Statistics for Spatial Data* (Wiley, New York).

Cuomo S, Cola VSD, Giampaolo F, Rozza G, Raissi M, Piccialli F (2022) Scientific machine learning through physics-informed neural networks: Where we are and what's next. *J. Sci. Comput.* 92:88.

Fu MC, Hu JQ (1997) *Conditional Monte Carlo: Gradient Estimation and Optimization Applications* (Kluwer Academic Publishers, Norwell, MA).

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

Glasserman P (2003) *Monte Carlo Methods in Financial Engineering* (Springer, New York).

Glasserman P, Heidelberger P, Shahabuddin P (1999) Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* 9:117–152.

Glynn PW (1990) Likelihood ratio gradient estimation for stochastic systems. *Comm. ACM* 33(10):75–84.

Hong LJ, Jiang G (2019) Offline simulation online application: A new framework of simulation-based decision making. *Asia-Pacific J. Oper. Res.* 36(6):223–236.

Horn RA, Zhang F (2005) Basic properties of the schur complement. Zhang F, ed. *The Schur Complement and Its Applications* (Springer, Boston), 17–46.

Hou R, Lu L (2018) miMLHD: Generate the optimal Latin Hypercube Design based on the miniMax criterion. https://search.r-project.org/CRAN/refmans/MOLHD/html/MOLHD-package.html.

Jiang G, Hong LJ, Nelson BL (2020) Online risk monitoring using offline simulation. *INFORMS J. Comput.* 32(2):356–375.

Jiang G, Hong LJ, Shen H (2024) Simulation metamodel-based prediction for real-time derivative pricing and risk management. https://dx.doi.org/10.1287/ijoc.2023.0292.cd, https://github.com/INFORMSJoC/2023.0292.

Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J. Global Optim.* 13:455–492.

Karniadakis GE, Kevrekidis LG, Lu L, Perdikari P, Wang S, Yang L (2021) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Nature Rev. Physics* 3:422–440.

Kemma AGZ, Vorst ACF (1990) A pricing method for options based on average asset values. *J. Bank. Finance* 14:113–129.

Liu G, Hong LJ (2011) Kernel estimation of the Greeks for options with discontinuous payoffs. *Oper. Res.* 59(1):96–108.

Liu H, Liang L, Lee LH, Chew EP (2022) Unifying offline and online simulation for online decision-making. *IISE Trans.* 54(10):923–935.

Loeppky JL, Sachs J, Welch WJ (2009) Choosing the sample size of a computer experiment: A practical guide. *Technometrics* 51:366–376.

Morris MD, Mitchell TJ (1995) Exploratory designs for computational experiments. *J. Statist. Planning Inference* 43(3):381–402.

Øksendal B (2010) *Stochastic Differential Equations: An Introduction with Applications* (Springer, New York).

Peng YJ, Fu MC, Hu JQ, Heidergott B (2018) A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. *Oper. Res.* 66(2):487–499.

Qu H, Fu MC (2014) Gradient extrapolated stochastic kriging. *ACM Trans. Modeling Comput. Simulation* 24:23.

Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378:686–707.

Rasmussen C, Williams C (2006) *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA).

Schoutens W (2003) *Lévy Processes in Finance: Pricing Financial Derivatives* (John Wiley Sons, Chichester, UK).

Shen H, Hong LJ, Zhang X (2018) Enhancing stochastic kriging for queueing simulation with stylized models. *IISE Trans.* 50(11):943–958.

Shen H, Hong LJ, Zhang X (2021) Ranking and selection with covariates for personalized decision making. *INFORMS J. Comput.* 32(2):356–375.

van Beers WCM, Kleijnen JPC (2003) Kriging for interpolation in random simulation. *J. Oper. Res. Soc.* 54:255–262.

Wang S, Teng Y, Perdikaris P (2021) Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM J. Sci. Comput.* 43(5):3055–3081.

Ye KQ (1998) Orthogonal column latin hypercubes and their application in computer experiments. *J. Amer. Statist. Assoc.* 93(444):1430–1439.

Zenios SA (1999) High-performance computing in finance: The last 10 years and the next. *Parallel Comput.* 25:2149–2175.