Submitted to *INFORMS Journal on Computing*

Real-time Derivative Pricing and Hedging with Consistent Metamodels

Guangxin Jiang
School of Management, Harbin Institute of Technology, Harbin 150001 China, gxjiang@hit.edu.cn,

L. Jeff Hong*
School of Management and School of Data Science, Fudan University, Shanghai 200433 China, hong_liu@fudan.edu.cn,

Haihui Shen
Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China, shenhaihui@sjtu.edu.cn,

---

* Corresponding author

E-Companion

EC.1. Proof of Theorem 1

Proof: Let $V^{(1)}(s)$ and $V^{(2)}(s)$ denote the first- and second-order derivatives of $V(s)$. By Taylor's Theorem,

$$V(S(t)) = V(s_0) + V^{(1)}(s_0)(S(t)-s_0) + \frac{1}{2}V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2,$$

where $\tilde{S}_t$ is between s_0 and $S(t)$, which is also a random variable. Then, $L(S(t))$ in (2) can be written as

$$L(S(t)) = -\frac{1}{2}V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2,$$

and $L^\dagger(S(t))$ in (3) can be written as

$$L^\dagger(S(t)) = \left(\Delta^\dagger(s_0) - V^{(1)}(s_0)\right)(S(t)-s_0) - \frac{1}{2}V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2.$$

Before we consider the variances of $L(S(t))$ and $L^\dagger(S(t))$, we first establish some convergence results that will be used later. Under assumption that $\sup_{0<t\le t_h}\mathbb{E}\left[e^{\theta X_t}\right] < \infty$ for all $|\theta| \le h$, for some $\varepsilon > 0$ and $t_1 = \min\left\{t_h, \left(\frac{h}{4+\varepsilon}\right)^2\right\}$,

$$\begin{aligned}\sup_{0<t\le t_1}\mathbb{E}\left[|S(t)|^{4+\varepsilon}\right] &= \sup_{0<t\le t_1}\mathbb{E}\left[\left(s_0 e^{at+\sqrt{t}X_t}\right)^{4+\varepsilon}\right] \\ &= \sup_{0<t\le t_1} s_0^{4+\varepsilon}e^{(4+\varepsilon)at}\mathbb{E}\left[e^{(4+\varepsilon)\sqrt{t}X_t}\right] < \infty. \end{aligned} \tag{EC.1}$$

Then from Hölder's inequality, we have $\sup_{0<t\le t_1}\mathbb{E}\left[|S(t)|^{c+\varepsilon}\right] < \infty$, for $c = 2,3,4$. Since $X_t \xrightarrow{d} X$ as $t \to 0^+$, then $at + \sqrt{t}X_t \xrightarrow{d} 0$ thus $S(t) \xrightarrow{d} s_0$. Then by the Theorem 25.12 (convergence of expectation) of Billingsley (1995) and its corollary, we immediately have

$$\mathbb{E}\left[S(t)^c\right] \to \mathbb{E}\left[s_0^c\right] = s_0^c \text{ as } t \to 0^+, \text{ for } c = 1,2,3,4. \tag{EC.2}$$

Besides, notice that

$$\begin{aligned}\frac{S(t)-s_0}{\sqrt{t}} &= s_0\frac{\exp\left(at+\sqrt{t}X_t\right)-1}{\sqrt{t}} \\ &= s_0\frac{e^{\xi_t}(at+\sqrt{t}X_t)}{\sqrt{t}} \qquad \text{(by Taylor's theorem)} \\ &= s_0 e^{\xi_t}(a\sqrt{t}+X_t),\end{aligned}$$

where ξ_t is between 0 and $at+\sqrt{t}X_t$. Since $at+\sqrt{t}X_t \xrightarrow{d} 0$ as $t \to 0^+$, then $\xi_t \xrightarrow{d} 0$ thus $e^{\xi_t} \xrightarrow{d} 1$. Hence

$$\frac{S(t)-s_0}{\sqrt{t}} \xrightarrow{d} s_0 X \text{ as } t \to 0^+. \tag{EC.3}$$

Moreover, for $p = \frac{4+\varepsilon}{3+\varepsilon} > 1$ and $q = 4+\varepsilon$ such that $\frac{1}{p}+\frac{1}{q} = 1$,

$$\begin{aligned}\mathbb{E}\left[\left|\frac{S(t)-s_0}{\sqrt{t}}\right|^{3+\varepsilon}\right] &= \mathbb{E}\left[\left|s_0 e^{\xi_t}\right|^{3+\varepsilon}\left|a\sqrt{t}+X_t\right|^{3+\varepsilon}\right] \\ &\le \left(\mathbb{E}\left[\left|s_0 e^{\xi_t}\right|^{4+\varepsilon}\right]\right)^{\frac{1}{p}}\left(\mathbb{E}\left[\left|a\sqrt{t}+X_t\right|^{(3+\varepsilon)q}\right]\right)^{\frac{1}{q}}. \quad \text{(by Hölder's inequality)}\end{aligned}$$

Since ξ_t is between 0 and $at+\sqrt{t}X_t$, $s_0e^{\xi_t}$ is between $s_0(>0)$ and $S(t)(\geq 0)$. Therefore $s_0e^{\xi_t}<s_0+S(t)$. By (EC.1) and Minkowski inequality, it is easy to see $\sup_{0<t\leq t_1}\mathbb{E}\left[|s_0e^{\xi_t}|^{4+\varepsilon}\right]<\infty$. On the other hand, the assumption that $\sup_{0<t\leq t_h}\mathbb{E}\left[e^{\theta X_t}\right]<\infty$ for all $|\theta|\leq h$ implies that $\sup_{0<t\leq t_h}\mathbb{E}\left[\left|X_t\right|^c\right]<\infty$ for any finite c. So again by Minkowski inequality, $\sup_{0<t\leq t_h}\mathbb{E}\left[\left|a\sqrt{t}+X_t\right|^{(3+\varepsilon)q}\right]<\infty$. Thus, we can see that

$$\sup_{0<t\leq t_1}\mathbb{E}\left[\left|\frac{S(t)-s_0}{\sqrt{t}}\right|^{3+\varepsilon}\right]<\infty. \tag{EC.4}$$

Then, with the same arguments used for (EC.2), from (EC.3) we can have

$$\mathbb{E}\left[\left(\frac{S(t)-s_0}{\sqrt{t}}\right)^c\right]\to\mathbb{E}\left[(s_0X)^c\right]=s_0^c\mathbb{E}\left[X^c\right]\ \text{ as } t\to 0^+, \text{ for } c=1,2,3. \tag{EC.5}$$

Now we consider the variance of $L(S(t))$ and $L^\dagger(S(t))$. Due to (EC.2), these variances are well defined at least for small t. And

$$\mathrm{Var}[L^\dagger(S(t))]-\mathrm{Var}[L(S(t))]=b^2\mathrm{Var}\left[S(t)-s_0\right]-b\mathrm{Cov}\left[S(t)-s_0,V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right], \tag{EC.6}$$

where $b\triangleq\Delta^\dagger(s_0)-V^{(1)}(s_0)$ is a nonzero constant. Next, we will show that $\mathrm{Var}[S(t)-s_0]=\Theta(t)$¶ and $\mathrm{Cov}[S(t)-s_0,V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2]=O(t^{3/2})$ as $t\to 0^+$. The first result holds because the convergence in (EC.5) implies that

$$\frac{\mathrm{Var}\left[S(t)-s_0\right]}{t}=\mathbb{E}\left[\left(\frac{S(t)-s_0}{\sqrt{t}}\right)^2\right]-\mathbb{E}^2\left[\frac{S(t)-s_0}{\sqrt{t}}\right]\to s_0^2\left(\mathbb{E}\left[X^2\right]-\mathbb{E}^2\left[X\right]\right)=s_0^2\mathrm{Var}[X]\ \text{as } t\to 0^+,$$

where $0<\mathrm{Var}[X]<\infty$. To see the second result, first notice that

$$\frac{\mathrm{Cov}\left[S(t)-s_0,V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right]}{t^{3/2}}=\mathbb{E}\left[V^{(2)}(\tilde{S}_t)\left(\frac{S(t)-s_0}{\sqrt{t}}\right)^3\right]-\mathbb{E}\left[\frac{S(t)-s_0}{\sqrt{t}}\right]\mathbb{E}\left[V^{(2)}(\tilde{S}_t)\left(\frac{S(t)-s_0}{\sqrt{t}}\right)^2\right]. \tag{EC.7}$$

Since $\tilde{S}_t$ is between s_0 and $S(t)$ and $S(t)\xrightarrow{d}s_0$ as $t\to 0^+$, then $\tilde{S}_t\xrightarrow{d}s_0$ thus $V^{(2)}(\tilde{S}_t)\xrightarrow{d}V^{(2)}(s_0)$. So, together with (EC.3),

$$V^{(2)}(\tilde{S}_t)\left(\frac{S(t)-s_0}{\sqrt{t}}\right)^c\xrightarrow{d}V^{(2)}(s_0)s_0^cX^c\ \text{as } t\to 0^+, \text{ for } c=2,3.$$

Then due to the assumption that $V(s)$ is bounded above and the boundedness result of $\frac{S(t)-s_0}{\sqrt{t}}$ in (EC.4), with arguments similar as before we will have

$$\mathbb{E}\left[V^{(2)}(\tilde{S}_t)\left(\frac{S(t)-s_0}{\sqrt{t}}\right)^c\right]\to\mathbb{E}\left[V^{(2)}(s_0)s_0^cX^c\right]=V^{(2)}(s_0)s_0^c\mathbb{E}\left[X^c\right],\ \text{for } c=2,3. \tag{EC.8}$$

By (EC.5), (EC.7) and (EC.8),

$$\frac{\mathrm{Cov}\left[S(t)-s_0,V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right]}{t^{3/2}}\to V^{(2)}(s_0)s_0^3\left(\mathbb{E}\left[X^3\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[X^2\right]\right)\ \text{as } t\to 0^+,$$

where $\mathbb{E}\left[X^c\right]<\infty$ for $c=1,2,3$. So, we have

$$\frac{b\mathrm{Cov}\left[S(t)-s_0,V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right]}{b^2\mathrm{Var}\left[S(t)-s_0\right]}\to 0\ \text{as } t\to 0^+,$$

which implies that there exists a $\tau>0$ such that when $t<\tau$, $b\mathrm{Cov}[S(t)-s_0,V^{(2)}(\tilde{S}_t)(S(t)-s_0)^2]<b^2\mathrm{Var}[S(t)-s_0]$. Therefore, by (EC.6), $\mathrm{Var}\left(L(S(t))\right)<\mathrm{Var}\left(L^\dagger(S(t))\right)$ for $t<\tau$. □

¶ We say $f(t)=O(g(t))$ as $t\to 0$ if and only if there exists positive numbers δ and M such that $|f(t)/g(t)|\leq M$ when $0<|t|<\delta$; We say $f(t)=\Theta(g(t))$ as $t\to 0$ if and only if there exists positive numbers δ, M and m such that $m\leq|f(t)/g(t)|\leq M$ when $0<|t|<\delta$.

EC.2. Verification for Heston Model

We verify that the assumption in Theorem 1 that $S(t) = s_0 \exp(at + \sqrt{t}X_t)$, where $X_t \xrightarrow{d} X$ with $\mathbb{E}[X^4] < \infty$ and $\mathrm{Var}[X] > 0$ holds for the following Heston model:

$$\begin{aligned} \mathrm{d}S(t) &= rS(t)\mathrm{d}t + \sqrt{\nu_t}S(t)\mathrm{d}W_t^S, \\ \mathrm{d}\nu_t &= \kappa[\theta - \nu_t]\mathrm{d}t + \sigma\sqrt{\nu_t}\mathrm{d}W_t^\nu, \end{aligned}$$

where W_t^S and W_t^ν are Brownian motions with correlation ρ. The initial values s_0 and ν_0 are positive. The underlying asset price at time t, given the values of s_0 and ν_0, can be written as

$$S(t) = s_0 \exp\left(rt - \frac{1}{2}\int_0^t \nu_s \mathrm{d}s + \rho\int_0^t \sqrt{\nu_t}\mathrm{d}W_s^S + \sqrt{1-\rho^2}\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^\nu\right).$$

In this case, the X_t in that assumption is as follows:

$$X_t = -\frac{\int_0^t \nu_s \mathrm{d}s}{2\sqrt{t}} + \frac{\rho\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^S}{\sqrt{t}} + \frac{\sqrt{1-\rho^2}\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^\nu}{\sqrt{t}}. \tag{EC.9}$$

Since ν_s is pathwise continuous, the mean value theorem, $\int_0^t \nu_s \mathrm{d}s = t\tilde{\nu}_t$ where $\tilde{\nu}_t$ is between ν_0 and ν_t. Since $\nu_t \xrightarrow{p} \nu_0$ as $t \to 0^+$, then $\tilde{\nu}_t \xrightarrow{p} \nu_0$. Hence,

$$\frac{\int_0^t \nu_s \mathrm{d}s}{2\sqrt{t}} = \frac{t\tilde{\nu}_t}{2\sqrt{t}} \xrightarrow{p} 0 \text{ as } t \to 0^+.$$

Next, we consider the second and third terms of X_t in (EC.9). By the fundamental theorem of stochastic calculus (see Theorem 1.1 in Isaacson (1969)), we have

$$\frac{1}{W_t^S}\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^S \xrightarrow{p} \sqrt{\nu_0} \text{ as } t \to 0^+.$$

So, as $t \to 0^+$,

$$\frac{\rho\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^S}{\sqrt{t}} = \rho Z^S \frac{1}{W^S(t)}\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^S \xrightarrow{p} \rho\sqrt{\nu_0}Z^S,$$

and

$$\frac{\sqrt{1-\rho^2}\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^\nu}{\sqrt{t}} = \sqrt{1-\rho^2}Z^\nu \frac{1}{W^S(t)}\int_0^t \sqrt{\nu_s}\mathrm{d}W_s^S \xrightarrow{p} \sqrt{1-\rho^2}\sqrt{\nu_0}Z^\nu,$$

where Z^S and Z^ν are standard normal random variables with correlation ρ.

Finally, $X_t \xrightarrow{p} \rho\sqrt{\nu_0}Z^S + \sqrt{1-\rho^2}\sqrt{\nu_0}Z^\nu \triangleq X$ as $t \to 0^+$ with $\mathbb{E}[X^4] < \infty$ and $\mathrm{Var}[X] > 0$. □

EC.3. Hedging Costs in Consistency and Inconsistency Cases

Recall that the delta hedging will adjust the underlying asset's position to $\Delta(s_0)$ in the consistency case and $\Delta^\dagger(s_0)$ in the inconsistency case, respectively. However, in the inconsistency case, to achieve the same hedging effect, i.e., to make $L^\dagger(S(t))$ have the same variance as $L(S(t))$, one needs to additionally conduct a series of hedging at time $t_1, \ldots, t_m$ with $0 < t_1 < \cdots < t_m < t$, which successively adjusts the position to $\Delta_1^\dagger, \ldots, \Delta_m^\dagger$ such that $\Delta_m^\dagger = \Delta(s_0)$, for some $m \geq 1$. Now we consider the hedging cost, which is usually a percentage d of the total trading volume. Suppose the previous position of the underlying asset is $\tilde{\Delta}$, then the hedging cost is

$$C = \left|\Delta(s_0) - \tilde{\Delta}\right| s_0 d, \tag{EC.10}$$

in the consistency case, and

$$C^\dagger = \left|\Delta^\dagger(s_0) - \tilde{\Delta}\right| s_0 d + \sum_{i=1}^{m} \left|\Delta_i^\dagger - \Delta_{i-1}^\dagger\right| S(t_i) d, \tag{EC.11}$$

in the inconsistency case, respectively, where $\Delta_0^\dagger \triangleq \Delta^\dagger(s_0)$. The following Theorem EC.1 says that to achieve the same hedging effect, the hedging cost in the consistency case is less than or equal to that in the inconsistency case.

THEOREM EC.1. *Suppose in the inconsistency case the risk manager needs to conduct a series of hedging at time $t_1, \ldots, t_m$ with $0 < t_1 < \cdots < t_m < t$, which successively adjusts the position to $\Delta_1^\dagger, \ldots, \Delta_m^\dagger$ such that $\Delta_m^\dagger = \Delta(s_0)$, for some $m \geq 1$, in order to achieve the same hedging effect in the consistency case. Moreover, assume that $\mathbb{E}[S(t_i)] = s_0$, for $i = 1, \ldots, m$. Then for the hedging cost C defined in* (EC.10) *and $C^\dagger$ defined in* (EC.11)*, $C \leq \mathbb{E}[C^\dagger]$.*

Proof: It is easy to see that

$$\begin{aligned}
\mathbb{E}[C^\dagger] &= \mathbb{E}\left[\left|\Delta^\dagger(s_0) - \tilde{\Delta}\right| s_0 d + \sum_{i=1}^{m} \left|\Delta_i^\dagger - \Delta_{i-1}^\dagger\right| S_{t_i} d\right] \\
&= \left|\Delta^\dagger(s_0) - \tilde{\Delta}\right| s_0 d + \sum_{i=1}^{m} \left|\Delta_i^\dagger - \Delta_{i-1}^\dagger\right| s_0 d \\
&\geq \left|\Delta^\dagger(s_0) - \tilde{\Delta} + \sum_{i=1}^{m} \left(\Delta_i^\dagger - \Delta_{i-1}^\dagger\right)\right| s_0 d \\
&= \left|\Delta^\dagger(s_0) - \tilde{\Delta} - \Delta_0^\dagger + \Delta_m^\dagger\right| s_0 d \\
&= \left|\Delta(s_0) - \tilde{\Delta}\right| s_0 d \qquad (\text{recall } \Delta_0^\dagger = \Delta^\dagger(s_0) \text{ and } \Delta_m^\dagger = \Delta(s_0)) \\
&= C,
\end{aligned}$$

which finishes the proof. □

REMARK EC.1. (i) The assumption that $\mathbb{E}[S(t_i)] = s_0$ in Theorem EC.1 makes sense for $S(t)$ considered in Theorem 1. Because $\mathbb{E}[S(t)] \to s_0$ as $t \to 0^+$, so for small t and $0 < t_1 < \cdots < t_m < t$, $\mathbb{E}[S(t_i)] \approx s_0$. (ii) The result in Theorem EC.1 can be seen intuitively. In consistency case we can adjust the position to cancel the first order effect of $V(s)$ in one step, while in inconsistency case we need to adjust the position more than once. And it is possible that the position is over adjusted in some step and needs to be adjusted back, which causes extra hedging cost.

EC.4. Proof of Theorem 2

Proof: Let $\hat{V}_i^{(1)}(s)$ and $\hat{V}_i^{(2)}(s)$ denote the first- and second-order derivatives of $\hat{V}_i(s)$, for $i = 1, 2$. By Taylor's Theorem,

$$\hat{V}_i(S(t)) = \hat{V}_i(s_0) + \hat{V}_i^{(1)}(s_0)(S(t) - s_0) + \frac{1}{2}\hat{V}_i^{(2)}(\tilde{S}(t))(S(t) - s_0)^2, \text{ for } i = 1, 2,$$

where $\tilde{S}_t$ is between s_0 and $S(t)$, which is also a random variable. Then, $L_1(S(t))$ in (4) can be written as

$$L_1(S(t)) = -\frac{1}{2}\hat{V}_1^{(2)}(\tilde{S}_t)(S(t) - s_0)^2,$$

and $L_2(S(t))$ in (5) can be written as

$$L_2(S(t)) = \left(\Delta_2(s_0) - \hat{V}_2^{(1)}(s_0)\right)(S(t)-s_0) - \frac{1}{2}\hat{V}_2^{(2)}(\tilde{S}_t)(S(t)-s_0)^2.$$

So, the variances of $L_1(S(t))$ and $L_2(S(t))$ are

$$\mathrm{Var}[L_1(S(t))] = \frac{1}{4}\mathrm{Var}\left[\hat{V}_1^{(2)}(\tilde{S}(t))(S(t)-s_0)^2\right],$$

and

$$\mathrm{Var}[L_2(S(t))] = b^2\mathrm{Var}\left[S(t)-s_0\right] + \frac{1}{4}\mathrm{Var}\left[\hat{V}_2^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right] - b\mathrm{Cov}\left[S(t)-s_0, \hat{V}_2^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right],$$

where $b \triangleq \Delta_2(s_0) - \hat{V}_2^{(1)}(s_0)$. Their existence is due to (EC.2) shown in the proof of Theorem 1. With the same arguments used to prove Theorem 1, we can have that, as $t \to 0^+$, $\mathrm{Var}[S(t)-s_0] = \Theta(t)$, $\mathrm{Cov}[S(t)-s_0, \hat{V}_2^{(2)}(\tilde{S}_t)(S(t)-s_0)^2] = O(t^{3/2})$ and $\mathrm{Var}\left[\hat{V}_i^{(2)}(\tilde{S}_t)(S(t)-s_0)^2\right] = O(t^2)$, for $i=1,2$. These results imply that $\mathrm{Var}[L_1(S(t))] = O(t^2)$ and $\mathrm{Var}[L_2(S(t))] = O(t)$ as $t \to 0^+$, which implies that $\mathrm{Var}[L_1(S(t))]/\mathrm{Var}[L_2(S(t))] \to 0$. Hence, there exists a $\tau > 0$ such that when $t < \tau$, $\mathrm{Var}[L_1(S(t))] < \mathrm{Var}[L_2(S(t))]$. □

EC.5. Explicit Forms of Some Covariances

Recall that

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

By the definition of $\partial\mathsf{M}(\mathbf{x})/\partial x_k$, we have

$$\begin{aligned}
&\mathrm{Cov}\left[\frac{\partial}{\partial x_k}\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{y})\right] \\
&= \mathbb{E}\left[\lim_{t\to 0}\frac{\mathsf{M}(\mathbf{x}+t\mathbf{e}_k)-\mathsf{M}(\mathbf{x})}{t}\mathsf{M}(\mathbf{y})\right] - \mathbb{E}\left[\lim_{t\to 0}\frac{\mathsf{M}(\mathbf{x}+t\mathbf{e}_k)-\mathsf{M}(\mathbf{x})}{t}\right]\mathbb{E}\left[\mathsf{M}(\mathbf{y})\right] \\
&= \lim_{t\to 0}\frac{1}{t}\left\{\mathbb{E}\left[(\mathsf{M}(\mathbf{x}+t\mathbf{e}_k)-\mathsf{M}(\mathbf{x}))\mathsf{M}(\mathbf{y})\right] - \mathbb{E}\left[\mathsf{M}(\mathbf{x}+t\mathbf{e}_k)-\mathsf{M}(\mathbf{x})\right]\mathbb{E}\left[\mathsf{M}(\mathbf{y})\right]\right\} \\
&= \lim_{t\to 0}\frac{1}{t}\left\{\mathbb{E}\left[\mathsf{M}(\mathbf{x}+t\mathbf{e}_k)\mathsf{M}(\mathbf{y})\right] - \mathbb{E}\left[\mathsf{M}(\mathbf{x}+t\mathbf{e}_k)\right]\mathbb{E}\left[\mathsf{M}(\mathbf{y})\right] - \left(\mathbb{E}\left[\mathsf{M}(\mathbf{x})\mathsf{M}(\mathbf{y})\right] - \mathbb{E}\left[\mathsf{M}(\mathbf{x})\right]\mathbb{E}\left[\mathsf{M}(\mathbf{y})\right]\right)\right\} \\
&= \lim_{t\to 0}\frac{1}{t}\left\{\mathrm{Cov}\left[\mathsf{M}(\mathbf{x}+t\mathbf{e}_k), \mathsf{M}(\mathbf{y})\right] - \mathrm{Cov}\left[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{y})\right]\right\} \\
&= \frac{\partial}{\partial x_k}\mathrm{Cov}\left[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{y})\right] \\
&= 2\theta_k(y_k-x_k)C(\mathbf{x},\mathbf{y}),
\end{aligned}$$

where the second equality holds since the Gaussian random field $\mathsf{M}(\mathbf{x})$ is differentiable and its covariance function is twice differentiable (Stein 1999, §2.4). Similarly, we also have

$$\mathrm{Cov}\left[\frac{\partial}{\partial x_k}\mathsf{M}(\mathbf{x}), \frac{\partial}{\partial y_h}\mathsf{M}(\mathbf{y})\right] = \frac{\partial^2}{\partial x_k \partial y_h}\mathrm{Cov}\left[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{y})\right] = \begin{cases} -4\theta_k\theta_h(x_k-y_k)(x_h-y_h)C(\mathbf{x},\mathbf{y}), & k \neq h, \\ 2\theta_k[1-2\theta_k(x_k-y_k)^2]C(\mathbf{x},\mathbf{y}), & k = h. \end{cases}$$

Recall that $\boldsymbol{\gamma}_+(\mathbf{z}) = (\boldsymbol{\gamma}(\mathbf{z})^\top, \boldsymbol{\gamma}^{0,1}(\mathbf{z})^\top, \ldots, \boldsymbol{\gamma}^{0,d}(\mathbf{z})^\top)^\top$, where $\boldsymbol{\gamma}(\mathbf{z})$ is a $n \times 1$ vector with the i-th element being $\mathrm{Cov}\left[\mathsf{M}(\mathbf{z}), \mathsf{M}(\mathbf{x}_i)\right]$, $\boldsymbol{\gamma}^{0,k}(\mathbf{z})$ is a $n \times 1$ vector with the i-th element being $\mathrm{Cov}\left[\mathsf{M}(\mathbf{z}), \frac{\partial}{\partial x_k}\mathsf{M}(\mathbf{x}_i)\right]$, for $k = 1, 2, \ldots, d$. Also notice that

$$\frac{\partial}{\partial z_k}\boldsymbol{\gamma}_+(\mathbf{z}) = \left(\frac{\partial}{\partial z_k}\boldsymbol{\gamma}(\mathbf{z})^\top, \frac{\partial}{\partial z_k}\boldsymbol{\gamma}^{0,1}(\mathbf{z})^\top, \ldots, \frac{\partial}{\partial z_k}\boldsymbol{\gamma}^{0,d}(\mathbf{z})^\top\right)^\top.$$

It is easy to see that $\frac{\partial}{\partial z_k}\boldsymbol{\gamma}(\mathbf{z})$ is a $n\times 1$ vector with the i-th element being

$$\frac{\partial}{\partial z_k}\mathrm{Cov}\left[\mathsf{M}(\mathbf{z}),\mathsf{M}(\mathbf{x}_i)\right]=\mathrm{Cov}\left[\frac{\partial}{\partial z_k}\mathsf{M}(\mathbf{z}),\mathsf{M}(\mathbf{x}_i)\right],$$

and, for $h=1,\dots,d$, $\frac{\partial}{\partial z_k}\boldsymbol{\gamma}^{0,h}(\mathbf{z})$ is a $n\times 1$ vector with the i-th element being

$$\frac{\partial}{\partial z_k}\mathrm{Cov}\left[\mathsf{M}(\mathbf{z}),\frac{\partial}{\partial x_h}\mathsf{M}(\mathbf{x}_i)\right]=\frac{\partial^2}{\partial x_h\partial z_k}\mathrm{Cov}\left[\mathsf{M}(\mathbf{z}),\mathsf{M}(\mathbf{x}_i)\right]=\mathrm{Cov}\left[\frac{\partial}{\partial z_k}\mathsf{M}(\mathbf{z}),\frac{\partial}{\partial x_h}\mathsf{M}(\mathbf{x}_i)\right].$$

EC.6. Proof of Theorem 3 with Normality Assumption

Proof: Additionally assume $\varepsilon_\ell(\mathbf{x})\sim\mathcal{N}(0,\sigma^2(\mathbf{x}))$ and $\epsilon_\ell^k(\mathbf{x})\sim\mathcal{N}(0,\sigma_k^2(\mathbf{x}))$ for $k=1,2,\dots,d$ and $\ell=1,2,\dots$. Then under the SK formulation, $(V(\mathbf{z}),\bar{\mathbf{Y}}^\top)^\top$ follows a multivariate normal distribution. It is not difficult to see that

$$V(\mathbf{z})|\bar{\mathbf{Y}}\sim\mathcal{N}\left(\mathbf{f}(\mathbf{z})^\top\boldsymbol{\beta}+\boldsymbol{\gamma}(\mathbf{z})^\top(\boldsymbol{\Gamma}+\boldsymbol{\Sigma})^{-1}(\bar{\mathbf{Y}}-\mathbf{F}\boldsymbol{\beta}),\mathrm{Var}\left[\mathsf{M}(\mathbf{z})\right]-\boldsymbol{\gamma}(\mathbf{z})^\top(\boldsymbol{\Gamma}+\boldsymbol{\Sigma})^{-1}\boldsymbol{\gamma}(\mathbf{z})\right).$$

So, the SK predictor (9) is the conditional expectation, i.e., $\hat{V}(\mathbf{z})=\mathbb{E}\left[V(\mathbf{z})|\bar{\mathbf{Y}}\right]$. Then,

$$\begin{aligned}\mathrm{MSE}_V^{\mathrm{SK}}(\mathbf{z})&=\mathbb{E}\left[\left(\hat{V}(\mathbf{z})-V(\mathbf{z})\right)^2\right]=\mathbb{E}\left[\mathbb{E}\left[\left(\hat{V}(\mathbf{z})-V(\mathbf{z})\right)^2\Big|\bar{\mathbf{Y}}\right]\right]=\mathbb{E}\left[\mathrm{Var}\left(V(\mathbf{z})|\bar{\mathbf{Y}}\right)\right]\\&=\mathrm{Var}(V(\mathbf{z}))-\mathrm{Var}\left(\mathbb{E}\left[V(\mathbf{z})|\bar{\mathbf{Y}}\right]\right)=\mathrm{Var}(V(\mathbf{z}))-\mathrm{Var}\left(\hat{V}(\mathbf{z})\right).\end{aligned}\qquad\text{(EC.12)}$$

Similarly, the GESK predictor (14) is also the conditional expectation, i.e., $\widetilde{V}(\mathbf{z})=\mathbb{E}\left[V(\mathbf{z})|\bar{\mathbf{Y}}_+\right]$, and

$$\mathrm{MSE}_V^{\mathrm{GESK}}(\mathbf{z})=\mathbb{E}\left[\left(\widetilde{V}(\mathbf{z})-V(\mathbf{z})\right)^2\right]=\mathrm{Var}(V(\mathbf{z}))-\mathrm{Var}\left(\widetilde{V}(\mathbf{z})\right).\qquad\text{(EC.13)}$$

Partition $\bar{\mathbf{Y}}_+$ as $\bar{\mathbf{Y}}_+\triangleq\left(\bar{\mathbf{Y}}^\top,\bar{\mathbf{D}}^\top\right)^\top$, then $\widetilde{V}(\mathbf{z})=\mathbb{E}\left[V(\mathbf{z})|\bar{\mathbf{Y}},\bar{\mathbf{D}}\right]$, and it is clear that $\hat{V}(\mathbf{z})=\mathbb{E}\left[\widetilde{V}(\mathbf{z})|\bar{\mathbf{Y}}\right]$. So,

$$\mathrm{Var}\left(\hat{V}(\mathbf{z})\right)=\mathrm{Var}\left(\mathbb{E}\left[\widetilde{V}(\mathbf{z})|\bar{\mathbf{Y}}\right]\right)=\mathrm{Var}\left(\widetilde{V}(\mathbf{z})\right)-\mathbb{E}\left(\mathrm{Var}\left[\widetilde{V}(\mathbf{z})|\bar{\mathbf{Y}}\right]\right)\le\mathrm{Var}\left(\widetilde{V}(\mathbf{z})\right),\qquad\text{(EC.14)}$$

where the exact equality will hold if and only if $\bar{\mathbf{D}}$ is perfectly determined by $\bar{\mathbf{Y}}$, which will not happen in practice. Combing (EC.12)-(EC.14) finishes the proof of (19).

The proof of (20) can be established in the same way, hence we omit the details. □

EC.7. Proof of (20) in Theorem 3

Proof: By (12), (17) and (18), to show (20) is equivalent to show

$$\boldsymbol{\gamma}_+^k(\mathbf{z})^\top(\boldsymbol{\Gamma}_++\boldsymbol{\Sigma}_+)^{-1}\boldsymbol{\gamma}_+^k(\mathbf{z})>\boldsymbol{\gamma}^{k,k}(\mathbf{z})^\top(\boldsymbol{\Gamma}^{k,k}+\boldsymbol{\Sigma}^{k,k})^{-1}\boldsymbol{\gamma}^{k,k}(\mathbf{z}).\qquad\text{(EC.15)}$$

We define a new matrix $\boldsymbol{\Lambda}_+^*$ from $\boldsymbol{\Lambda}_+$ by swapping the block matrices $\boldsymbol{\Gamma}+\boldsymbol{\Sigma}$ with $\boldsymbol{\Gamma}^{k,k}+\boldsymbol{\Sigma}^{k,k}$, that is,

$$\boldsymbol{\Lambda}_+^*=\begin{pmatrix}\boldsymbol{\Gamma}^{k,k}+\boldsymbol{\Sigma}^{k,k}&\boldsymbol{\Gamma}^{k,1}+\boldsymbol{\Sigma}^{k,1}&\cdots&\boldsymbol{\Gamma}^{k,k-1}+\boldsymbol{\Sigma}^{k,k-1}&\boldsymbol{\Gamma}^{k,0}+\boldsymbol{\Sigma}^{k,0}&\boldsymbol{\Gamma}^{k,k+1}+\boldsymbol{\Sigma}^{k,k+1}&\cdots&\boldsymbol{\Gamma}^{k,d}+\boldsymbol{\Sigma}^{k,d}\\\boldsymbol{\Gamma}^{1,k}+\boldsymbol{\Sigma}^{1,k}&\boldsymbol{\Gamma}^{1,1}+\boldsymbol{\Sigma}^{1,1}&\cdots&\boldsymbol{\Gamma}^{1,k-1}+\boldsymbol{\Sigma}^{1,k-1}&\boldsymbol{\Gamma}^{1,0}+\boldsymbol{\Sigma}^{1,0}&\boldsymbol{\Gamma}^{1,k+1}+\boldsymbol{\Sigma}^{1,k+1}&\cdots&\boldsymbol{\Gamma}^{1,d}+\boldsymbol{\Sigma}^{1,d}\\\vdots&\vdots&\ddots&\vdots&\vdots&\ddots&\vdots&\vdots\\\boldsymbol{\Gamma}^{k-1,k}+\boldsymbol{\Sigma}^{k-1,k}&\boldsymbol{\Gamma}^{k-1,1}+\boldsymbol{\Sigma}^{k-1,1}&\cdots&\boldsymbol{\Gamma}^{k-1,k-1}+\boldsymbol{\Sigma}^{k-1,k-1}&\boldsymbol{\Gamma}^{k-1,0}+\boldsymbol{\Sigma}^{k-1,0}&\boldsymbol{\Gamma}^{k-1,k+1}+\boldsymbol{\Sigma}^{k-1,k+1}&\cdots&\boldsymbol{\Gamma}^{k-1,d}+\boldsymbol{\Sigma}^{k-1,d}\\\boldsymbol{\Gamma}^{0,k}+\boldsymbol{\Sigma}^{0,k}&\boldsymbol{\Gamma}^{0,1}+\boldsymbol{\Sigma}^{0,1}&\cdots&\boldsymbol{\Gamma}^{0,k-1}+\boldsymbol{\Sigma}^{0,k-1}&\boldsymbol{\Gamma}+\boldsymbol{\Sigma}&\boldsymbol{\Gamma}^{0,k+1}+\boldsymbol{\Sigma}^{0,k+1}&\cdots&\boldsymbol{\Gamma}^{0,d}+\boldsymbol{\Sigma}^{0,d}\\\boldsymbol{\Gamma}^{k+1,k}+\boldsymbol{\Sigma}^{k+1,k}&\boldsymbol{\Gamma}^{k+1,1}+\boldsymbol{\Sigma}^{k+1,1}&\cdots&\boldsymbol{\Gamma}^{k+1,k-1}+\boldsymbol{\Sigma}^{k+1,k-1}&\boldsymbol{\Gamma}^{k+1,0}+\boldsymbol{\Sigma}^{k+1,0}&\boldsymbol{\Gamma}^{k+1,k+1}+\boldsymbol{\Sigma}^{k+1,k+1}&\cdots&\boldsymbol{\Gamma}^{k+1,d}+\boldsymbol{\Sigma}^{k+1,d}\\\vdots&\vdots&\ddots&\vdots&\vdots&\ddots&\vdots&\vdots\\\boldsymbol{\Gamma}^{d,k}+\boldsymbol{\Sigma}^{d,k}&\boldsymbol{\Gamma}^{d,1}+\boldsymbol{\Sigma}^{d,1}&\cdots&\boldsymbol{\Gamma}^{d,k-1}+\boldsymbol{\Sigma}^{d,k-1}&\boldsymbol{\Gamma}^{d,0}+\boldsymbol{\Sigma}^{d,0}&\boldsymbol{\Gamma}^{d,k+1}+\boldsymbol{\Sigma}^{d,k+1}&\cdots&\boldsymbol{\Gamma}^{d,d}+\boldsymbol{\Sigma}^{d,d}\end{pmatrix}.$$

We also define

$$\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})=\left(\boldsymbol{\gamma}^{k,k}(\mathbf{z})^{\top},\boldsymbol{\gamma}^{k,1}(\mathbf{z})^{\top},\ldots,\boldsymbol{\gamma}^{k,k-1}(\mathbf{z})^{\top},\boldsymbol{\gamma}^{k,0}(\mathbf{z})^{\top},\boldsymbol{\gamma}^{k,k+1}(\mathbf{z})^{\top},\ldots,\boldsymbol{\gamma}^{k,d}(\mathbf{z})^{\top}\right)^{\top}.$$

First, it is easy to see that the positive definiteness of $\boldsymbol{\Lambda}_{+}$ implies the positive definiteness of $\boldsymbol{\Lambda}_{+}^{*}$. Then, we observe the following result

$$\boldsymbol{\gamma}_{+}^{k}(\mathbf{z})^{\top}\boldsymbol{\Lambda}_{+}^{-1}\boldsymbol{\gamma}_{+}^{k}(\mathbf{z})=\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})^{\top}\boldsymbol{\Lambda}_{+}^{*\,-1}\boldsymbol{\gamma}_{+}^{*}(\mathbf{z}). \tag{EC.16}$$

There are many ways to show (EC.16), here we do it in virtue of the elementary probability. First notice that $\boldsymbol{\Lambda}_{+}$ is actually the covariance matrix of $\bar{\mathbf{Y}}_{+}$. So if we let $\boldsymbol{\xi}=(\boldsymbol{\xi}_0^{\top},\boldsymbol{\xi}_1^{\top},\ldots,\boldsymbol{\xi}_d^{\top})^{\top}$ be a $n(d+1)\times 1$ random vector with multivariate normal distribution $\mathcal{N}(\mathbf{0},\boldsymbol{\Lambda}_{+})$, then its density function is $f_{\boldsymbol{\xi}}(\mathbf{x})=\frac{1}{\sqrt{(2\pi)^{n(d+1)}\det\boldsymbol{\Lambda}_{+}}}\exp\left\{-\frac{1}{2}\mathbf{x}^{\top}\boldsymbol{\Lambda}_{+}^{-1}\mathbf{x}\right\}$ for $\mathbf{x}\in\Re^{n(d+1)}$. If we let $\boldsymbol{\xi}^{*}=(\boldsymbol{\xi}_k^{\top},\boldsymbol{\xi}_1^{\top},\ldots,\boldsymbol{\xi}_{k-1}^{\top},\boldsymbol{\xi}_0^{\top},\boldsymbol{\xi}_{k+1}^{\top},\ldots,\boldsymbol{\xi}_d^{\top})^{\top}$, then due to the way $\boldsymbol{\Lambda}_{+}^{*}$ is constructed, $\boldsymbol{\xi}^{*}$ follows the multivariate normal distribution $\mathcal{N}(\mathbf{0},\boldsymbol{\Lambda}_{+}^{*})$, and its density function is $f_{\boldsymbol{\xi}^{*}}(\mathbf{x})=\frac{1}{\sqrt{(2\pi)^{n(d+1)}\det\boldsymbol{\Lambda}_{+}^{*}}}\exp\left\{-\frac{1}{2}\mathbf{x}^{\top}\boldsymbol{\Lambda}_{+}^{*\,-1}\mathbf{x}\right\}$. Since $f_{\boldsymbol{\xi}}(\mathbf{0})=f_{\boldsymbol{\xi}^{*}}(\mathbf{0})$, we can have $\det\boldsymbol{\Lambda}_{+}=\det\boldsymbol{\Lambda}_{+}^{*}$. Moreover, due to the way $\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})$ is constructed, $f_{\boldsymbol{\xi}}(\boldsymbol{\gamma}_{+}(\mathbf{z}))=f_{\boldsymbol{\xi}^{*}}(\boldsymbol{\gamma}_{+}^{*}(\mathbf{z}))$, which implies (EC.16).

With (EC.16), it suffices to prove

$$\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})^{\top}\boldsymbol{\Lambda}_{+}^{*\,-1}\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})>\boldsymbol{\gamma}^{k,k}(\mathbf{z})^{\top}(\boldsymbol{\Gamma}^{k,k}+\boldsymbol{\Sigma}^{k,k})^{-1}\boldsymbol{\gamma}^{k,k}(\mathbf{z}). \tag{EC.17}$$

If we partition $\boldsymbol{\Lambda}_{+}^{*}$ as

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

and partition $\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})$ as $\boldsymbol{\gamma}_{+}^{*}(\mathbf{z})=\left(\boldsymbol{\gamma}^{k,k}(\mathbf{z})^{\top},\boldsymbol{\gamma}_B^{*}(\mathbf{z})^{\top}\right)^{\top}$, then with the same arguments used in proving (21), we can prove (EC.17). Thus (EC.15) is proved, so is (20). □

EC.8. PDE for Option Portfolio

Proposition EC.1. *Suppose that a portfolio consists M derivatives, whose PDEs are given by* (26). *Then, the price of this portfolio $\Phi(\mathbf{S}(t),t)$ follows the PDE*

$$\begin{cases}\mathcal{L}\Phi(\mathbf{S}(t),t)=0,\\ \Phi(\mathbf{S}(T),T)=\displaystyle\sum_{m=1}^{M}P_m(\mathbf{S}(T)).\end{cases} \tag{EC.18}$$

Proof: Recall the notations in Lemma 2, and define $h(\mathbf{s},t)=0$, and let $P(\mathbf{s})=\sum_{m=1}^{M}P_m(\mathbf{s})$. Then,

$$\begin{aligned}&\mathbb{E}\left[\int_t^T e^{-\int_t^{\tau}r(\mathbf{S}(\iota),\iota)\mathrm{d}\iota}h(\mathbf{S}(\tau),\tau)\mathrm{d}\tau+e^{-\int_t^{T}r(\mathbf{S}(\iota),\iota)\mathrm{d}\iota}P(\mathbf{S}(T))\middle|\mathbf{S}(t)=\mathbf{s}\right]\\&=\mathbb{E}\left[e^{-\int_t^{T}r(\mathbf{S}(\iota),\iota)\mathrm{d}\iota}P(\mathbf{S}(T))\middle|\mathbf{S}(t)=\mathbf{s}\right]=\mathbb{E}\left[e^{-\int_t^{T}r(\mathbf{S}(\iota),\iota)\mathrm{d}\iota}\sum_{m=1}^{M}P_m(\mathbf{S}(T))\middle|\mathbf{S}(t)=\mathbf{s}\right]\\&=\sum_{m=1}^{M}\mathbb{E}\left[e^{-\int_t^{T}r(\mathbf{S}(\iota),\iota)\mathrm{d}\iota}P_m(\mathbf{S}(T))\middle|\mathbf{S}(t)=\mathbf{s}\right]=\sum_{m=1}^{M}V_m(\mathbf{s},t)=\Phi(\mathbf{s},t),\end{aligned} \tag{EC.19}$$

where the second to last equality is due to (26) and Lemma 2. Then due to (EC.19) and Lemma 2, (EC.18) is proved. □

For example, we consider a portfolio consists of two options $V_1(S_1(t),t)$ and $V_1(S_2(t),t)$, which are based on underlying assets $S_1(t)$ and $S_2(t)$, respectively, and $S_1(t)$ and $S_2(t)$ are independent. Specifically, let

$$\frac{\mathrm{d}S_1(t)}{S_1(t)} = (r-q_1)\mathrm{d}t + \sigma_1 \mathrm{d}B_1(t) \quad \text{and} \quad \frac{\mathrm{d}S_2(t)}{S_2(t)} = (r-q_2)\mathrm{d}t + \sigma_2 \mathrm{d}B_2(t),$$

and the corresponding PDEs are

$$\mathcal{L}V_1 = \frac{\partial V_1}{\partial t} + (r-q_1)S_1\frac{\partial V_1}{\partial S_1} + \frac{1}{2}\sigma_1^2 S_1^2 \frac{\partial^2 V_1}{\partial S_1^2} - rV_1 = 0,$$
$$\mathcal{L}V_2 = \frac{\partial V_2}{\partial t} + (r-q_2)S_2\frac{\partial V_2}{\partial S_2} + \frac{1}{2}\sigma_2^2 S_2^2 \frac{\partial^2 V_2}{\partial S_2^2} - rV_2 = 0.$$

Rewrite

$$\frac{\mathrm{d}S_1(t)}{S_1(t)} = (r-q_1)\mathrm{d}t + \sigma_{11}\mathrm{d}B_1(t) + \sigma_{12}\mathrm{d}B_2(t),$$
$$\frac{\mathrm{d}S_2(t)}{S_2(t)} = (r-q_2)\mathrm{d}t + \sigma_{21}\mathrm{d}B_1(t) + \sigma_{22}\mathrm{d}B_2(t),$$

where $\sigma_{11}=\sigma_1$, $\sigma_{22}=\sigma_2$, and $\sigma_{12}=\sigma_{21}=0$. So, by Proposition EC.1, the portfolio price $\Phi = V_1 + V_2$ follows the PDE

$$\begin{aligned}
\mathcal{L}\Phi &= \frac{\partial \Phi}{\partial t} + \sum_{i=1}^{2}(r-q_i)S_i\frac{\partial \Phi}{\partial S_i} + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} S_i S_j \gamma_{ij}\frac{\partial^2 \Phi}{\partial S_i \partial S_j} - r\Phi \\
&= \frac{\partial \Phi}{\partial t} + (r-q_1)S_1\frac{\partial \Phi}{\partial S_1} + (r-q_2)S_2\frac{\partial \Phi}{\partial S_2} + \frac{1}{2}S_1^2\sigma_1^2\frac{\partial^2\Phi}{\partial S_1^2} + \frac{1}{2}S_2^2\sigma_2^2\frac{\partial^2\Phi}{\partial S_2^2} - r\Phi \\
&= \frac{\partial V_1}{\partial t} + (r-q_1)S_1\frac{\partial V_1}{\partial S_1} + (r-q_2)S_2\frac{\partial V_2}{\partial S_2} + \frac{1}{2}S_1^2\sigma_1^2\frac{\partial^2 V_1}{\partial S_1^2} + \frac{1}{2}S_2^2\sigma_2^2\frac{\partial^2 V_2}{\partial S_2^2} - r(V_1+V_2) \\
&= \left(\frac{\partial V_1}{\partial t} + (r-q_1)S_1\frac{\partial V_1}{\partial S_1} + \frac{1}{2}\sigma_1^2 S_1^2\frac{\partial^2 V_1}{\partial S_1^2} - rV_1\right) + \left(\frac{\partial V_2}{\partial t} + (r-q_2)S_2\frac{\partial V_2}{\partial S_2} + \frac{1}{2}\sigma_2^2 S_2^2\frac{\partial^2 V_2}{\partial S_2^2} - rV_2\right) \\
&= \mathcal{L}V_1 + \mathcal{L}V_2 = 0.
\end{aligned}$$ □

EC.9. Boxplots of Bias and Standard Deviation

For the second part of numerical experiments in Section 6.1.1, the boxplots of the bias and standard deviation for the price, delta, vega, rho and theta using GESK and separate SK are shown in Figures EC.1-EC.2.

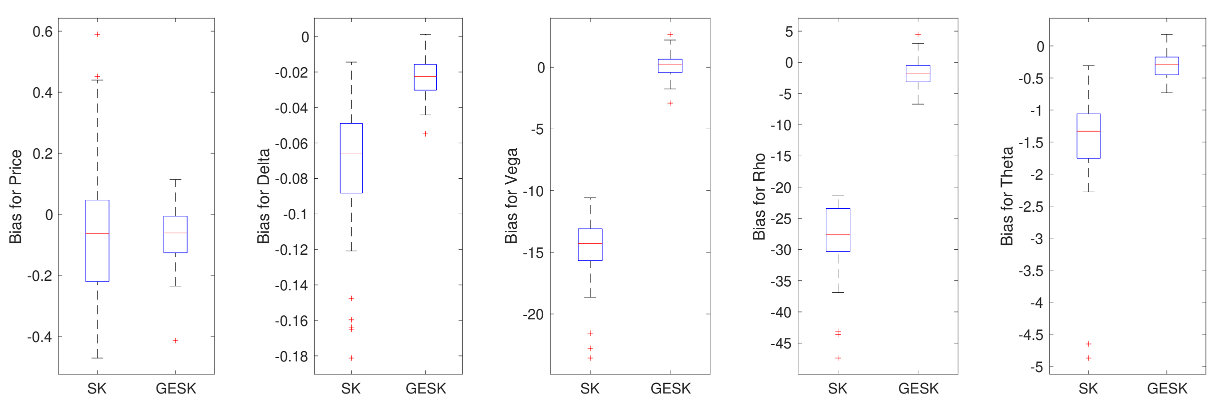

Figure EC.1 Boxplots of bias for price, delta, vega, rho, and theta surfaces.

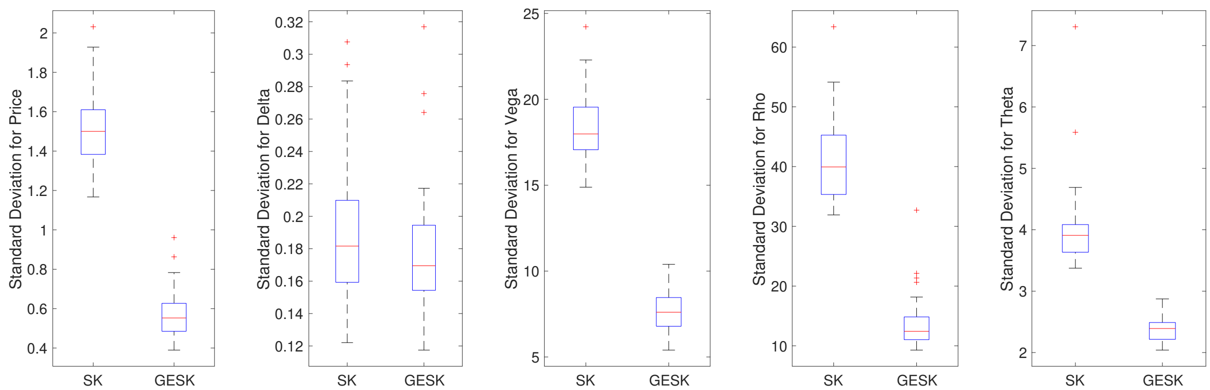

Figure EC.2 Boxplots of standard deviation for price, delta, vega, rho, and theta surfaces.

EC.10. Effects of Different LHS Approaches

For the second part of numerical experiments in Section 6.1.1, we investigate the effects when different LHS approaches are used to choose the design points. In particular, we try the minimax LHS (denoted as "-mM") and maximin LHS (denoted as "-Mm") in the R package "MOLHD" (see Morris and Mitchell 1995 and Hou and Lu 2018 for more details), to compare with the built-in lhsdesign function in Matlab. Since our numerical experiments are implemented in Matlab, we first generate the design points in R and then import them into Matlab. In addition to the Matlab built-in function lhsdesgin(n,p) with default setting (denoted as "-Matlab"), we also try lhsdesgin(n,p,'Criterion','correlation') that minimizes the sum of between-column squared correlations (denoted as "-Matlab-CR"). Figures EC.3-EC.7 show the comparison of different LHS approaches, where no significant differences are observed.

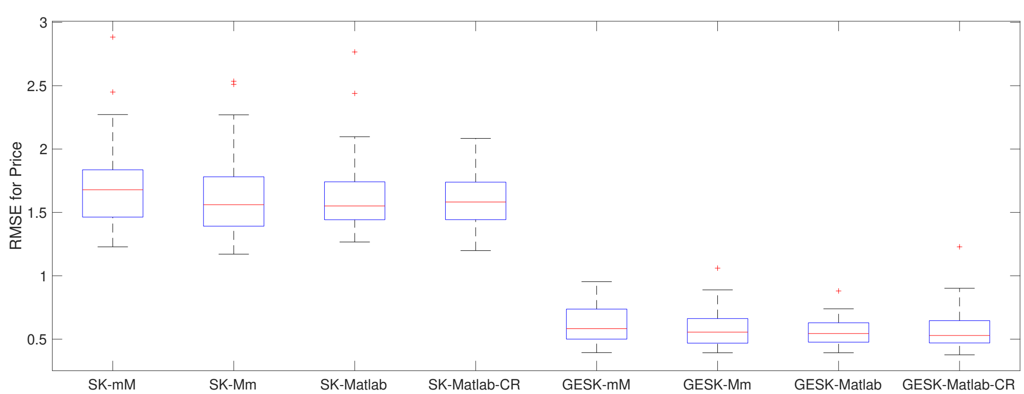

Figure EC.3 Boxplots of RMSE for price surface with different LH methods

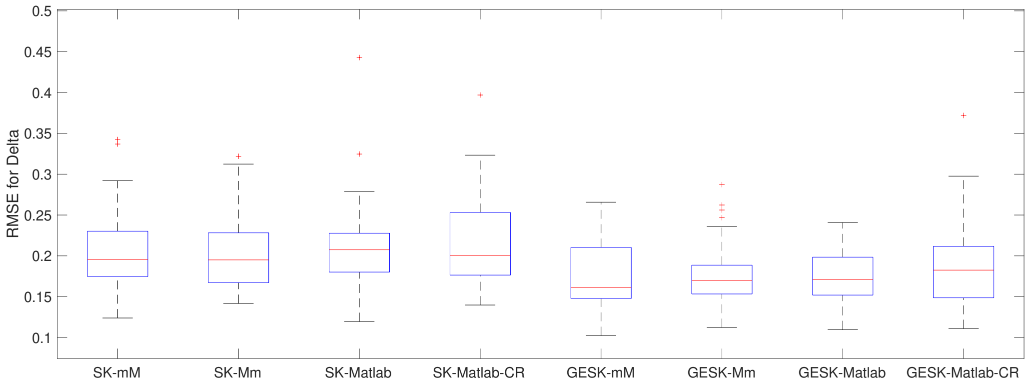

Figure EC.4 Boxplots of RMSE for delta surface with different LH methods

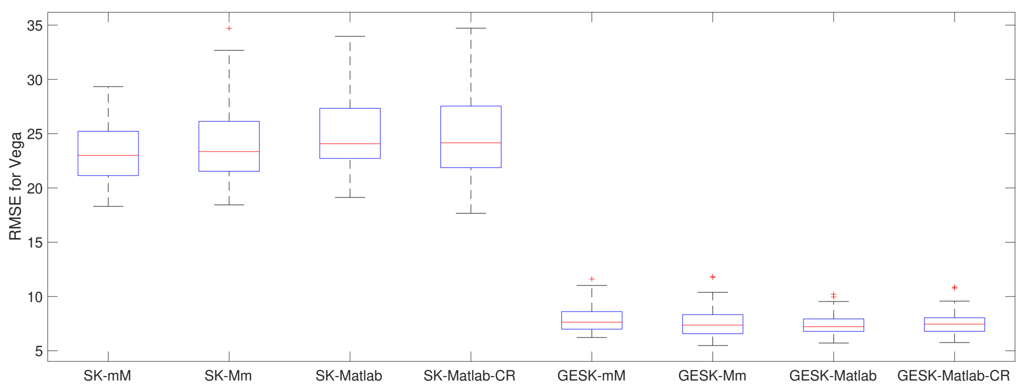

Figure EC.5 Boxplots of RMSE for vega surface with different LH methods

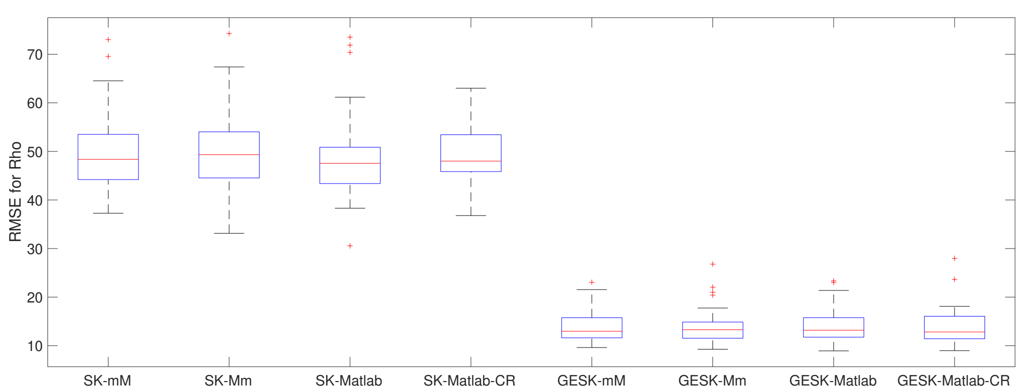

Figure EC.6 Boxplots of RMSE for rho surface with different LH methods

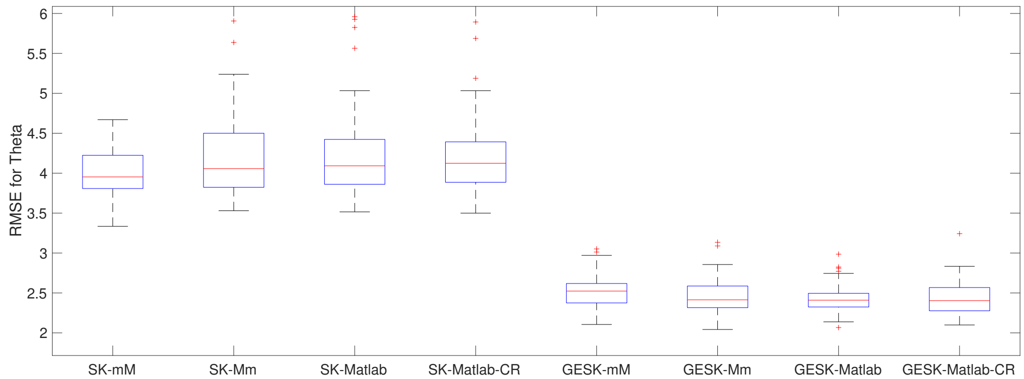

Figure EC.7 Boxplots of RMSE for theta surface with different LH methods

EC.11. Parameters Calibration of VG Process

Based on the data from Yahoo Finance on 9th November 2018, we collect the corresponding European options' trading prices in Chicago Board Options Exchange with high liquidity to calibrate the parameters (σ, ν, θ) of VG process for each considered stock. The considered trading prices for each option under specific strike prices and maturities are listed in Tables i – v. Here we use the method proposed in Carr and Madan (1999), which applies fast Fourier transform (FFT) based on the characteristic function to price options driven by Lévy processes. Given the real European option prices, least squares criterion is used to calibrate the parameters (σ, ν, θ).

Table i Apple, Inc. option chain, $S_0 = 204.47$, yield 1.21%.

strike \ maturity	20181116	20181130	20181221	20190118	20190215
180	24.53		24.90	27.25	
185			21.50		
190	14.90		17.31	19.00	20.75
195	10.07		13.45	15.49	
197.5	7.90				
200	5.92	7.70	10.00	12.20	15.10
202.5	4.10	6.15			
205	2.71	4.80	7.15	9.36	12.25
207.5	1.62	3.50			
210	0.90	2.56	4.84	7.00	9.84
212.5	0.48	1.87			
215	0.25	1.28	3.05	5.02	7.90
220	0.08	0.60	1.87	3.59	6.10
225	0.05	0.26	1.10	2.43	4.70
230	0.03	0.15	0.63	1.65	3.60
235	0.02		0.38	1.13	2.74
240	0.01		0.26	0.77	2.02
245		0.03	0.17	0.52	
250	0.01		0.12	0.40	1.14
255				0.27	
260				0.22	
265				0.18	

Table ii Facebook, Inc. option chain, $S_0 = 144.96$, yield 0.

strike \ maturity	20181123	20181214	20181221	20181228	20190118
130					18.20
135					14.35
140	6.50	8.50	9.07	9.50	10.70
143	4.35				
144	3.80				
145	3.21	5.45	5.94		7.74
146	2.71	4.90		5.81	
147	2.25	4.35		5.35	
148	1.81	3.90		4.85	
149	1.46			4.26	
150	1.19		3.60	4.01	5.25
155	0.33	1.54	1.99	2.25	3.42
160	0.10		1.00	1.19	2.11
165			0.50		1.26
170			0.26		0.74
175	0.02	0.18	0.16		0.45
180			0.10		0.27
185					0.17
190			0.04		0.12
195			0.02		0.08
200			0.01		0.07
205					0.04
210					0.03
215					0.02
220					0.01

Table iii Netflix, Inc. option chain, $S_0 = 113.47$, yield 0.

strike \ maturity	20181214	20181221	20190118	20190215	20190315
80		41.85			
100		27.20		37.80	
105	21.40	23.55		34.47	
110	19.50	20.60	25.84	31.74	34.66
115	16.69	17.84	23.45	29.37	32.23
120		15.65	20.80	27.14	29.90
125		13.45		24.30	27.53
130		11.15	16.31	22.43	24.65
135		9.53	14.45	20.95	
140	6.90	8.00	12.61	18.50	22.00
145		6.62		16.85	20.00
150		5.49		13.78	17.50
155	3.48	4.51			
160	2.95	3.65	7.16	12.70	14.60
165		2.97		10.40	13.36
170		2.42		10.00	11.75
175		1.90		8.75	11.02
180	1.04	1.59	3.90	7.90	10.05
185	0.83	1.25		6.50	
190	0.75	1.03	2.90	6.30	8.35
195	0.64	0.85			7.47
200	0.49	0.72			6.98
205			1.75		
210		0.49	1.57	4.10	5.55
215	0.35	0.48			
220		0.37		3.04	4.02
230			0.88	2.56	3.30

Table iv Alibaba Group option chain, $S_0 = 144.85$, yield 0.

strike \ maturity	20181116	20181123	20181214	20181221	20190118
139	6.9				
140	6.14			10.75	12.50
141	5.50		9.20		
142	4.80				
143	4.17	5.60	7.90		
144	4.00				
145	3.40	4.10	7.50	7.95	9.90
146	2.93	3.63			
147	2.38	3.37			
148	2.06	2.99			
149	1.71	2.45			
150	1.43	2.31	4.75	5.48	7.45
152.5	0.85	1.52			
155	0.49	1.02		3.82	5.55
157.5	0.29	0.60			
160	0.17	0.43		2.54	4.05
162.5	0.12	0.28	1.65		
165	0.10			1.57	4.05
170	0.07			1.00	1.85
180	0.06			0.39	0.98
190				0.20	0.49
200					0.29
220	0.01	0.05			

Table v Tesla, Inc. option chain, $S_0 = 120.51$, yield 0.

strike \ maturity	20170407	20170428	20170616	201710915	20180119
100	23.74			52.39	
105	19.85	32.95			54.00
107.5	17.25		31.70		
110	16.06				51.25
120	9.35	22.51		41.85	
122.5	8.60		23.00		
130	5.05			36.55	41.30
140	2.55		17.78	31.60	36.30
150	1.17	9.15	12.32	26.94	31.35
155	0.78		11.03		
160	0.56				27.93
165	0.40		9.40		
170	0.24	4.84			23.53
180	0.11		6.30		21.43
185		2.81	4.76		
190	0.07				16.90
200	0.04			11.20	14.45
210				9.24	12.62
270				2.78	4.30

EC.12. IPA Estimators for Greeks

The stock price under exponential variance gamma (VG) process model is given by

$$S^{VG}(t) = S(0)\exp\left((r-\phi)t + X_t^{VG}\right),$$

where X_t^{VG} is a VG process with parameter (σ,ν,θ), and $\phi = -1/\nu \log(1-\theta\nu-\sigma^2\nu/2)$ is a compensation such that $\mathbb{E}[S^{VG}(t)] = S(0)\exp(rt)$. At time 0, the European call option price is given by

$$V = \exp(-rT)\mathbb{E}\left[\left(S^{VG}(T)-K\right)^+\right].$$

To estimate V via Monte Carlo simulation, we use time-changed Brownian motion method to simulate X_t^{VG}, i.e., $X_t^{VG} = \theta G_t + \sigma W_{G_t}$, where G_t is a gamma process with distribution $\Gamma(t/\nu,\nu)$, and W_{G_t} is a Brownian motion replacing t by G_t (see Schoutens 2003).

Notice that the payoff function in V is Lipschitz continuous, so we can apply IPA to estimate the Greeks. The details are as follows.

1. $\partial V/\partial S(0)$ (Delta). The IPA estimator is given by

$$\begin{aligned}\frac{\partial\left(e^{-rT}\left(S^{VG}(T)-K\right)^+\right)}{\partial S(0)} &= \frac{\partial\left(e^{-rT}\left(S(0)\exp\left((r-\phi)T+X_T^{VG}\right)-K\right)^+\right)}{\partial S(0)}\\ &= e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\ge K\right\}}\frac{S^{VG}(T)}{S(0)}.\end{aligned}$$

2. $\partial V/\partial T$ (Theta). The IPA estimator is given by

$$\begin{aligned}\frac{\partial\left(e^{-rT}\left(S^{VG}(T)-K\right)^+\right)}{\partial T} &= \frac{\partial\left(e^{-rT}\left(S(0)\exp\left((r-\phi)T+X_T^{VG}\right)-K\right)^+\right)}{\partial T}\\ &= -re^{-rT}\left(S^{VG}(T)-K\right)^+ + e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\ge K\right\}}S^{VG}(T)\left((r-\phi)+\frac{\partial X_T^{VG}}{\partial T}\right),\end{aligned}$$

where $\partial X_T^{VG}/\partial T$ is given in Appendix B.1 in Cao (2011).

3. $\partial V/\partial r$ (Rho). The IPA estimator is given by

$$\begin{aligned}\frac{\partial\left(e^{-rT}\left(S^{VG}(T)-K\right)^+\right)}{\partial r} &= \frac{\partial\left(e^{-rT}\left(S(0)\exp\left((r-\phi)T+X_T^{VG}\right)-K\right)^+\right)}{\partial r}\\ &= -Te^{-rT}\left(S^{VG}(T)-K\right)^+ + e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\ge K\right\}}TS^{VG}(T).\end{aligned}$$

4. $\partial V/\partial\theta$. The IPA estimator is given by

$$\begin{aligned}\frac{\partial\left(e^{-rT}\left(S^{VG}(T)-K\right)^+\right)}{\partial \theta} &= \frac{\partial\left(e^{-rT}\left(S(0)\exp\left((r-\phi)T+X_T^{VG}\right)-K\right)^+\right)}{\partial \theta}\\ &= e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\ge K\right\}}\frac{\partial\left(S(0)\exp\left((r-\phi)T+\theta G_T+\sigma W_{G_T}\right)\right)}{\partial\theta}\\ &= e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\ge K\right\}}S^{VG}(T)\left(-\frac{T}{1-\theta\nu-\sigma^2\nu/2}+G_T\right).\end{aligned}$$

5. $\partial V/\partial\nu$. The IPA estimator is given by

$$\begin{aligned}\frac{\partial\left(e^{-rT}\left(S^{VG}(T)-K\right)^+\right)}{\partial \nu} &= \frac{\partial\left(e^{-rT}\left(S(0)\exp\left((r-\phi)T+X_T^{VG}\right)-K\right)^+\right)}{\partial \nu}\\ &= e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\ge K\right\}}S^{VG}(T)\left(-\frac{T}{\nu}\left(\frac{\theta+\sigma^2/2}{1-\theta\nu-\sigma^2\nu/2}-\phi\right)+\frac{\partial X_T^{VG}}{\partial\nu}\right),\end{aligned}$$

where $\partial X_T^{VG}/\partial\nu$ is given in Appendix B.2 in Cao (2011).

6. $\partial V/\partial\sigma$. The IPA estimator is given by

$$\begin{aligned}
\frac{\partial\left(e^{-rT}\left(S^{VG}(T)-K\right)^{+}\right)}{\partial\sigma} &= \frac{\partial\left(e^{-rT}\left(S(0)\exp\left((r-\phi)T+X_T^{VG}\right)-K\right)^{+}\right)}{\partial\sigma} \\
&= e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\geq K\right\}}\frac{\partial\left(S(0)\exp\left((r-\phi)T+\theta G_T+\sigma W_{G_T}\right)\right)}{\partial\sigma} \\
&= e^{-rT}\mathbf{1}_{\left\{S^{VG}(T)\geq K\right\}}S^{VG}(T)\left(-\frac{T\sigma}{1-\theta\nu-\sigma^2\nu/2}+W_{G_t}\right).
\end{aligned}$$

EC.13. Parameter Setting in Section 6.3

Table vi Parameter settings

Parameters	Values
Initial Values	$S_{i,1}(0)=9+i, i=1,2,\dots,40$; $S_{i,2}(0)=19+i, i=1,2,\dots,40$.
Volatilities	$\sigma_{i,1}=0.1, i=1,\dots,10$; $\sigma_{i,1}=0.15, i=11,\dots,20$; $\sigma_{i,1}=0.2, i=21,\dots,30$; $\sigma_{i,1}=0.15, i=31,\dots,40$ $\sigma_{i,2}=0.2, i=1,\dots,20$; $\sigma_{i,2}=0.25, i=21,\dots,30$; $\sigma_{i,2}=0.2, i=31,\dots,40$.
Correlations	$\rho_i=0.1, i=1,\dots,10$; $\rho_i=0.5, i=11,\dots,20$; $\rho_i=-0.5, i=21,\dots,30$; $\rho_i=0, i=31,\dots,40$.
Strike Prices	$K_{i,1}=9.3+0.7i, i=1,\dots,40$; $K_{i,2}=7.4+0.6i, i=1,\dots,40$; $K_{i,3}=11.2+0.8i, i=1,\dots,40$; $K_{i,4}=16.8+1.2i, i=1,\dots,40$; $K_{i,5}=15.9+1.1i, i=1,\dots,40$; $K_{i,6}=18.7+1.3i, i=1,\dots,40$;
Dividends	$q_i=0, i=1,2,\dots,40$;

References

Billingsley P (1995) *Probability and Measure* (John Wiley & Sons), 3rd edition.

Cao L (2011) *Optimal Variance Swaps Portfolios and Estimating Greeks for Variance-Gamma*. Ph.D. thesis, University of Maryland, College Park, MD, USA.

Carr P, Madan DB (1999) Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2:61–73.

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

Isaacson D (1969) Stochastic integrals and derivatives. *The Annals of Mathematical Statistics* 40:1610–1616.

Morris MD, Mitchell TJ (1995) Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference* 43(3):381–402.

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

Stein ML (1999) *Interpolation of Spatial Data: Some Theory for Kriging* (New York: Springer).