

文章编号:2095-6134(2018)01-0010-08

A restricted B-spline estimation of diffusion processes with applications to financial data^{*}

HOU Yangyang[†], XU Tiange

(School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

(Received 27 December 2016; Revised 1 March 2017)

Hou Y Y, Xu T G. A restricted B-spline estimation of diffusion processes with applications to financial data[J]. Journal of University of Chinese Academy of Sciences, 2018,35(1):10-17.

Abstract Time-homogeneous diffusion process plays an important role in the financial market, and it is widely used for describing the stochastic dynamics of the underlying economic variables. In this work, we study nonparametric estimation of the drift and diffusion functions for the time-homogeneous diffusion process, and propose a new nonparametric regression technique based on higher-order approximations, which is called B-spline approach. The nonnegativity of the diffusion function is guaranteed by the restricted B-spline method. Our simulation results show that our method indeed outperforms the local polynomial method.

Keywords B-spline; time-homogeneous diffusion process; volatility

CLC number: TB114 **Document code:** A **doi:**10. 7523/j. issn. 2095-6134. 2018. 01. 002

扩散过程系数的约束 B 样条估计及其在金融数据中的应用

侯扬扬,徐天戈

(中国科学院大学数学科学学院,北京 100049)

摘 要 时齐扩散过程在金融领域具有重要作用,它被广泛应用于描述基础资产变量的随机波动。研究时齐扩散过程的漂移系数和扩散系数的非参数估计问题,在高阶近似的基础上给出新的非参数估计方法——B 样条估计。该方法通过对 B 样条拟合系数加以非负约束来保证扩散系数的非负性。模拟结果验证了该方法在对扩散系数的估计上优于局部多项式估计方法。

关键词 B 样条;时齐扩散过程;波动率

In the last decades, various well-known financial models based on the time-homogeneous diffusion processes have arisen. In this work, we are concerned about how to estimate the drift and diffusion coefficients of the processes. We generally consider the time-homogeneous Markov process $\{X_t\}$ given by the stochastic differential equations

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad (1)$$

^{*} Supported by the National Nature Science Foundation of China (11101419)
[†] Corresponding author, E-mail: houyangyang14@mails.ucas.ac.cn

where $\{B_t, t > 0\}$ is a standard one-dimensional Brownian motion, and $\mu(\cdot)$ and $\sigma(\cdot)$ are the drift and diffusion coefficients, respectively, dependent on $\{X_t\}$. There have been a great amount of studies on parametric methods to estimate $\mu(\cdot)$ and $\sigma(\cdot)$ [1,2,3]. One of the major limitations of these methods is that the specific parametric forms of $\mu(\cdot)$ and $\sigma(\cdot)$ have to be assumed first. In order to model more complex drift and diffusion functions and reduce the bias of the estimators, nonparametric regression techniques have been introduced in this area. Florens-Zmirou [4] first proposed nonparametric estimation with discrete sampling observations for the diffusion term, while the nonparametric kernel estimators of the drift and diffusion functions of a stationary process based on a set of discrete sampling observations were developed by Jiang and Knight [5].

Stanton [6] constructed the first-, second-, and third-order approximation formulas for $\mu(\cdot)$ and $\sigma(\cdot)$, in which the unknown conditional expectations estimated by Nadaraya-Watson kernel regression were given, and claimed the superiority of higher-order approximations. By deriving the higher-order approximations and computing the asymptotic variances, Fan and Zhang [7] concluded that higher-order approximations could reduce the numerical approximation errors of asymptotic biases, but the asymptotic variances would escalate nearly exponentially with the order of approximation. To overcome this difficulty, they adopted the local polynomial techniques to estimate the drift and diffusion functions and applied such techniques to some well-known short-term interest rate models. It turned out that this approach might produce the negative results for the diffusion functions when the higher-order approximations were used. To ensure the nonnegativity of the diffusion functions, Ziegelmann [8] and Yu and Jones [9] proposed positive nonparametric conditional variance estimators and Xu [10] proposed reweighted functional estimation of the diffusion coefficient based on reweighting the conventional Nadaraya-Watson estimator. In addition, Yu et al. [11] developed the log P-splines and local log-linear method to semiparametrically

estimate a class of time-inhomogeneous diffusion processes. Due to the heavy reliance on the logarithmic transformation of the diffusion function, which brings additional bias term, the method is not asymptotically equivalent to the local linear estimator.

In this work, we make use of B-spline basis functions to estimate the drift and diffusion functions based on higher-order approximations, and compare it with the local polynomial techniques. Our simulation results show that the B-spline method indeed outperforms the local polynomial method.

1 The estimators of the drift and diffusion functions

We introduce the estimators of the drift and diffusion functions based on higher-order approximations [7]. Given a time-homogeneous diffusion process $\{X_t\}$ satisfying (0.1) and an arbitrary $n + 1$ order continuous differentiable function $f(\cdot)$, the conditional expected increment can be expressed as

$$E[f(X_{t+\Delta}) - f(X_t) | X_t] = \mathcal{L}f(X_t)\Delta + \frac{1}{2}\mathcal{L}^2f(X_t)\Delta^2 + \cdots + \frac{1}{n!}\mathcal{L}^nf(X_t)\Delta^n + O(\Delta^{n+1}), \quad (2)$$

where the infinitesimal generator \mathcal{L} of the process $\{X_t\}$ is defined as

$$\mathcal{L}f(x) = f'(x)\mu(x) + \frac{1}{2}f''(x)\sigma^2(x). \quad (3)$$

Setting $f(x) = x - X_t$ and $(x - X_t)^2$ in (3), we have $\mathcal{L}f(x) = \mu(X_t)$ and $\mathcal{L}f(x) = \sigma^2(X_t)$.

Furthermore, we rewrite (2) at the different time steps $j\Delta, j = 1, \dots, k$, and get a k th-order approximation scheme

$$\frac{1}{\Delta} \sum_{j=1}^k a_{k,j} E[f(X_{t+j\Delta}) - f(X_t) | X_t] \approx \mathcal{L}f(X_t) \quad (4)$$

with the approximation error $(-1)^{k+1} \frac{\mathcal{L}^{k+1}f(X_t)}{(k+1)!} \Delta^k + O(\Delta^{k+1})$, where $a_{k,j} = (-1)^{j+1} \binom{k}{j} / j, j = 1, \dots, k$ (see Ref. [7] for details). Setting $f(x) = x - X_t$ and $(x - X_t)^2$ in (4), we reach the k th-order

approximation schemes of $\mu(\cdot)$ and $\sigma^2(\cdot)$

$$\begin{aligned}\mu(X_t) &= \frac{1}{\Delta} \sum_{j=1}^k a_{k,j} E(X_{t+j\Delta} - X_t | X_t) + O(\Delta^k), \\ \sigma^2(X_t) &= \frac{1}{\Delta} \sum_{j=1}^k a_{k,j} E((X_{t+j\Delta} - X_t)^2 | X_t) + \\ &\quad O(\Delta^k).\end{aligned}\quad (5)$$

2 B-spline techniques

Fan and Zhang^[7] proposed the q th-degree ($q \geq 0$) local polynomial techniques for estimating the conditional expectations in (5), and computed the asymptotic variances of nonparametric estimators. In this section, we will adopt B-spline techniques to estimate $\mu(\cdot)$ and $\sigma^2(\cdot)$.

2.1 Estimation for $\mu(\cdot)$ using B-Spline

Our goal is to obtain the smooth function $\mu(x) = E(Y_{i\Delta} | X_{i\Delta})$. To this end, we fit the discrete $(n-k)$ pairs of synthetic data $(X_{i\Delta}, Y_{i\Delta})$, $X_{i\Delta} \in [\alpha, \beta]$, $i = 1, \dots, n-k$ by the model $Y_{i\Delta} = \mu(X_{i\Delta}) + \varepsilon_i$, where $Y_{i\Delta} = \frac{1}{\Delta} \sum_{j=1}^k a_{k,j} (X_{(i+j)\Delta} - X_{i\Delta})$. Let p_1 be the order of the B-spline and define a sequence of knots $\alpha = x_1 = x_2 = \dots = x_{p_1} < x_{p_1+1} < \dots < x_{p_1+J_1+1} = x_{p_1+J_1+2} = \dots = x_{2p_1+J_1} = \beta$ on an interval $[\alpha, \beta]$. Then $\mu(x)$ can be expanded in terms of B-spline basis functions as $\mu(x) = \mathbf{B}_1(x)\mathbf{a}$, where $\mathbf{a} = (a_1, \dots, a_{m_1})'$, $\mathbf{B}_1(x) = (B_1(x), \dots, B_{m_1}(x))$, $m_1 = J_1 + p_1$, and $B_j(x)$ is defined as

$$\begin{aligned}B_{j,1}(x) &= \begin{cases} 1, & x_j < \dots < x_{j+1} \\ 0, & \text{otherwise.} \end{cases} \\ B_{j,p_1}(x) &= \frac{x - x_j}{x_{j+p_1-1} - x_j} B_{j,p_1-1}(x) + \\ &\quad \frac{x_{j+p_1} - x}{x_{j+p_1} - x_{j+1}} B_{j+1,p_1-1}(x),\end{aligned}$$

for $j = 1, \dots, m_1$ (the convention $x/0 = 0$ is used). When p_1 is chosen as the order of B-spline basis functions, $B_j(x) = B_{j,p_1}(x)$, $j = 1, \dots, m_1$ for the simplicity.

To avoid the excessively “wiggly” estimated curve $\hat{\mu}(x)$, we take the roughness penalty $\mathbf{a}'\mathbf{R}_1\mathbf{a}$ into account^[12]. Then the coefficients a_j can be obtained by minimizing

$$\text{PENSSE}(\mathbf{a}) = (\mathbf{Y} - \mathbf{H}_1\mathbf{a})'(\mathbf{Y} - \mathbf{H}_1\mathbf{a}) + \lambda_1\mathbf{a}'\mathbf{R}_1\mathbf{a}, \quad (6)$$

where $\mathbf{Y} = (Y_{1\Delta}, \dots, Y_{(n-k)\Delta})'$, $\mathbf{R}_1 = \int D^2\mathbf{B}_1(x)D^2\mathbf{B}_1'(x)dx$, $\mathbf{H}_1 = (\mathbf{B}_1(X_{1\Delta}), \dots, \mathbf{B}_1(X_{(n-k)\Delta}))'$, D^2 is the second-order differential operator, λ_1 is the smoothing parameter which controls the smoothness of a fitted curve.

In this work, we choose the best value of λ_1 by minimizing $\text{GCV}(\lambda_1)$ (generalized cross-validation)

$$\text{GCV}(\lambda_1) = \frac{\text{RSS}(\lambda_1)}{\{1 - (n-k)^{-1}\text{tr}(\mathbf{S}_{\lambda_1})\}^2},$$

where $\text{RSS}(\lambda_1) = (\mathbf{Y} - \mathbf{H}_1\hat{\mathbf{a}})'(\mathbf{Y} - \mathbf{H}_1\hat{\mathbf{a}})$ is the residual sum of squares and the smoother matrix of $\mathbf{S}_{\lambda_1} = \mathbf{H}_1(\mathbf{H}_1'\mathbf{H}_1 + \lambda_1\mathbf{R}_1)^{-1}\mathbf{H}_1'$, $\text{tr}(\mathbf{S}_{\lambda_1})$ is the trace of \mathbf{S}_{λ_1} . By minimizing (2.8), we get the penalized B-spline regression estimator of the vector of coefficients $\hat{\mathbf{a}} = (\mathbf{H}_1'\mathbf{H}_1 + \lambda_1\mathbf{R}_1)^{-1}\mathbf{H}_1'\mathbf{Y}$, and furthermore reach the estimator for the drift function

$$\hat{\mu}(x) = \mathbf{B}_1(x)\hat{\mathbf{a}} = \mathbf{B}_1(x)(\mathbf{H}_1'\mathbf{H}_1 + \lambda_1\mathbf{R}_1)^{-1}\mathbf{H}_1'\mathbf{Y}. \quad (7)$$

2.2 Estimation for $\sigma^2(\cdot)$ using B-spline with nonnegative restrictions

Similarly, the square of diffusion function $\sigma^2(x)$ can be also expressed as a linear combination of the B-splines $B_j(x)$ in the form $\sigma^2(x) = \mathbf{B}_2(x)\mathbf{c}$, where $\mathbf{c} = (c_1, \dots, c_{m_2})$, $\mathbf{B}_2(x) = (B_1(x), \dots, B_{m_2}(x))$, $m_2 = J_2 + p_2$, J_2 and p_2 are the numbers of interval knots and the order of B-spline, respectively. In order to ensure the positivity of $\sigma^2(x)$, we impose the reasonable constraint $c_j \geq 0$, ($j = 1, \dots, m_2$) due to the property of the coefficients c_j of B-spline $B_j(x)$ that there are no more sign changes in $\sigma^2(x)$ than those in the sequence c_j .

Under the constraint $\mathbf{c} \geq 0$, the estimation of \mathbf{c} can be achieved by the following optimization problem with a linear inequality constraint

$$\begin{aligned}\text{minimise } & \longrightarrow (\mathbf{Z} - \mathbf{H}_2\mathbf{c})'(\mathbf{Z} - \mathbf{H}_2\mathbf{c}) + \lambda_2\mathbf{c}'\mathbf{R}_2\mathbf{c} \\ \text{with } & \mathbf{c} \geq 0,\end{aligned}\quad (8)$$

where $\mathbf{Z} = (Z_{1\Delta}, \dots, Z_{(n-k)\Delta})'$, $Z_{i\Delta} = \frac{1}{\Delta} \sum_{j=1}^k a_{k,j} (X_{(i+j)\Delta} - X_{i\Delta})^2$, $i = 1, \dots, n-k$, $\mathbf{R}_2 = \int D^2\mathbf{B}_2(x)D^2\mathbf{B}_2'(x)dx$, and $\mathbf{H}_2 = (\mathbf{B}_2(Z_{1\Delta}), \dots, \mathbf{B}_2(Z_{(n-k)\Delta}))'$,

The smoothing parameter λ_2 can be chosen by minimizing $GCV(\lambda_2)$,

$$GCV(\lambda_2) = \frac{RSS(\lambda_2)}{\{1 - (n - k)^{-1} \text{tr}(\mathbf{S}_{\lambda_2})\}^2},$$

where $RSS(\lambda_2) = (\mathbf{Z} - \mathbf{H}_2 \hat{\mathbf{c}})'(\mathbf{Z} - \mathbf{H}_2 \hat{\mathbf{c}})$ is the residual sum of squares, $\hat{\mathbf{c}}$ is the estimator of \mathbf{c} , and the smoother matrix $\mathbf{S}_{\lambda_2} = \mathbf{H}_2(\mathbf{H}_2' \mathbf{H}_2 + \lambda_2 \mathbf{R}_2)^{-1} \mathbf{H}_2'$, and $\text{tr}(\mathbf{S}_{\lambda_2})$ is the trace of \mathbf{S}_{λ_2} .

2.3 Choosing the order k of approximation

According to the view of Stanton^[6], the approximations to the drift and diffusion functions converge pointwise to $\mu(\cdot)$ and $\sigma(\cdot)$ at a rate Δ^k , and the first-order approximation performs well when the data are sampled weekly or more frequently. However, as the frequency of data decreases, the approximation error of the first-order approximation increases. In order to reduce asymptotic bias of the approximation error, Fan and Zhang^[7] proposed higher-order approximation method. However, the asymptotic variances of such approximations escalate nearly exponentially with the approximations order.

In this work, we balance the asymptotic bias and variance by choosing $k = 1$ for the data sampled weekly or more frequently and opting for the second-order approximation when the sampling frequency decreases or the data over a small range of values.

2.4 Selection of the number of knots and order p

Ruppert^[13] pointed out that the number of the interval knots was not so crucial as the penalty parameter λ . When the chosen number of knot is enough to fit the feature of data, any increase in J will have little effect on the fitting result. Moreover, the computation time will increase as J rises. As a trade-off, Yoshimoto et al.^[14] proposed the statistical criterion BIC to select a suitable J . Apart from BIC statistical criterion, one might use a simple default value. In Ref. [13], Matt Wand suggested that $\min(\lfloor \frac{n-k}{4} \rfloor, 35)$ interval knots were chosen ($\lfloor x \rfloor$ is the integral part of x), where n is the number of observations, and k is the order of approximation.

To simplify the computation, we set $J_1 = J_2 = \min(\lfloor \frac{n-k}{4} \rfloor, 35)$ and choose equally spaced knots. This turns out not to affect the results. Other choices of the number of knots have also been tried, and similar results display.

As for the selection of p , we choose $p_1 = p_2 = 4$. The reason for such selection is that the simulation results show the order 4 B-splines fit the drift and diffusion functions well enough and the higher order spline functions will increase the computation, which overwhelms the approximation precision.

3 Simulation

In this section, we apply our proposed method to three examples. Then we compare our method with local polynomial method. To assess the performance of the two estimation methods, we use the following two measures^[15].

Measure 1 Mean absolute deviation error

$$MADE(\hat{\sigma}_i^2) = m^{-1} \sum_{i=1}^m |Z_{i\Delta} - \hat{\sigma}_i^2|,$$

Measure 2 Ideal mean absolute deviation error

$$IMADE(\hat{\sigma}_i^2) = m^{-1} \sum_{i=1}^m |\hat{\sigma}_i^2 - \sigma_i^2|,$$

where m is the length of the testing sample, $\hat{\sigma}_i$ is the estimation of the volatility, and σ_i is the true. These measures are used to assess the performance of different procedures for estimating the volatility. However, for the real data analysis, Measure 2 is not applicable. For the drift function, these measures can also be employed.

3.1 Cox-ingersoll-ross squared-root (CIR SR) model

In this example, we consider the well-known CIR SR model

$$dX_t = (\alpha + \beta X_t) dt + \sigma X_t^{1/2} dB_t, t \geq t_0, \quad (9)$$

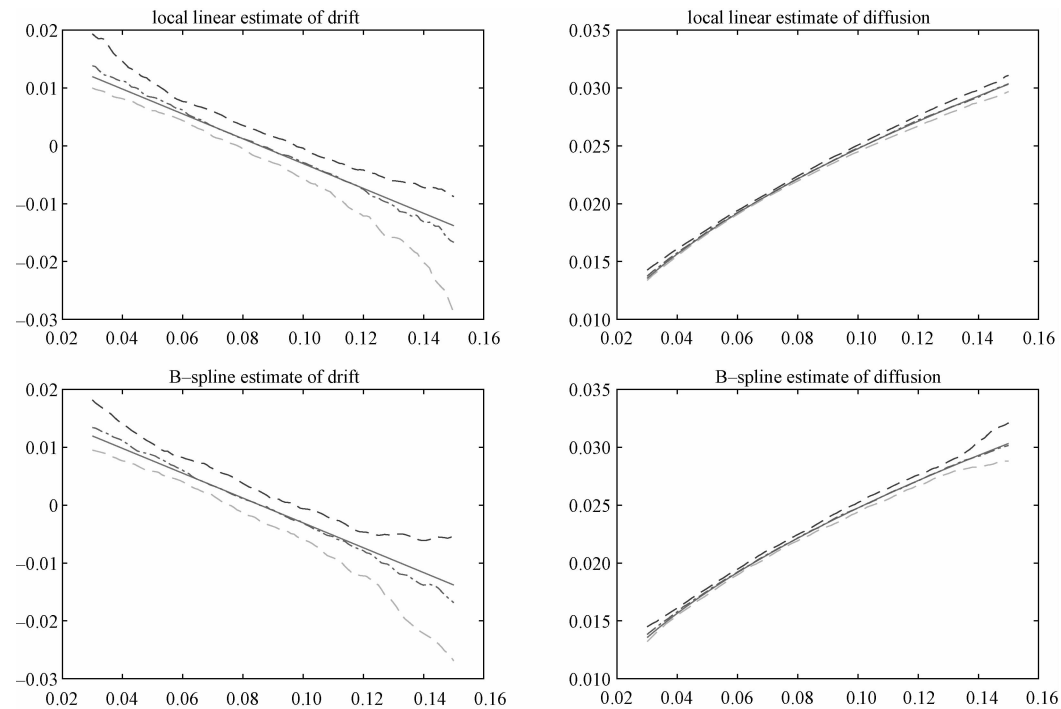
where $\{X_t\}$ is a measurable Markov process. B_t is a standard Brownian motion. Our simulation data are generated by using the discrete time order 1.0 strong approximation scheme (see Ref. [16] for details) as follows:

$$\begin{aligned} X_{t_{i+1}} \approx & X_{t_i} + [\alpha + \beta X_{t_i} - 4^{-1} \sigma^2] \Delta + \\ & 2^{-1} \sigma \{ [X_{t_i} + (\alpha + \beta X_{t_i} - 4^{-1} \sigma^2) \Delta + \\ & \sigma (X_{t_i})_+^{1/2} \varepsilon_i \sqrt{\Delta}]_+^{1/2} + (X_{t_i})_+^{1/2} \} \varepsilon_i \sqrt{\Delta}, \\ & \text{for } i = 1, \cdots, n-1, \end{aligned} \tag{10}$$

where $\{\varepsilon_i\}_{i=1}^{n-1}$ are independent and standard normal variables, $x_+ = \max(x, 0)$, the time points $t_1 < \cdots < t_n$, are equally spaced, and Δ is the step size.

In order to compare with local polynomial techniques, we adopt the same values of model parameters for simulation: $\alpha = 0.018\,392\,5$, $\beta = -0.214\,59$, $\sigma = 0.078\,30$, and $\Delta = 1/52$. To begin with, a sample path of length 5 000 is

generated by the discretization scheme (10) and setting the initial value $X_0 = 0.1$, where the first 4 500 observations are in-sample data and the remaining 500 observations out-sample data. Then the estimators of the drift and diffusion functions are calculated respectively through k th-order approximation. Here we choose $k = 1$, as the simulation data are weekly ones. For the purpose of comparison, we simulate 100 sample paths with range interval $\vartriangle = [0.03, 0.15]$, and set the number of interval knots $J_1 = J_2 = 35$ on the interval \vartriangle . The simulation results are presented in Fig. 1 and Table 1.



Solid; true function, dashed-dotted; the median of the estimates, and dashed; the 25th and 75th percentiles of the estimates.

Fig. 1 Estimated drift and diffusions among 100 simulations for CIR SR model

Table 1 Comparisons between B-spline and local polynomial methods				
	measure	empirical formula	local polynomial	B-spline
Drift	MADE	ave	0.178 7	0.125 0
		std	0.529 4	0.017 9
	IMADE	ave	0.058 2	0.003 9
		std	0.527 6	0.003 0
Volitivity	MADE	ave($\times 10^{-4}$)	9.153 0	4.819 3
		std($\times 10^{-3}$)	4.300 0	0.140 7
	IMADE	ave($\times 10^{-4}$)	4.532 5	0.603 3
		std($\times 10^{-3}$)	4.300 0	0.145 1

In this example, smoothing parameters λ_1 and λ_2 are chosen as 1×10^{-3} and 1×10^{-4} by

minimizing $GCV(\lambda_1)$ and $GCV(\lambda_2)$, respectively. The bandwidth parameter for the local linear estimation is generated from the rule

$$h = C \times \text{std}(X_\Delta, \cdots, X_{(n-k)\Delta}) n^{-1/5}, \tag{11}$$

where $C = 6$. In Table 1 it is shown that the B-spline method possesses certain superior performance because of its lower MADE and IMADE.

3.2 Geometric Brownian motion (GBM) model

We consider another familiar example of geometric Brownian motion determined by

$$dX_t = (\mu + \frac{1}{2}\sigma^2)X_t dt + \sigma X_t dB_t, t \geq t_0, \quad (12)$$

where B_t is a standard one-dimensional Brownian motion. The used parameter values are the same as these in Ref. [7]. $\mu = 0.087$, $\sigma = 0.178$, and the state value $X_0 = 1, \Delta = 1/250, h = 2 \times \text{std}(X_\Delta, \dots, X_{(n-k)\Delta})n^{-1/5}$. Using these values and following the order 1.0 scheme, we generate 1 000 sample paths of length 1 000 observations,

$$X_{t_{i+1}} \approx X_{t_i} + (\mu + 2^{-1}\sigma^2)X_{t_i}\Delta + \sigma X_{t_i}\varepsilon_i \sqrt{\Delta} + 2^{-1}\sigma^2 X_{t_i}(\varepsilon_i^2 - 1)\Delta. \quad (13)$$

For each simulation, we set the first 900 observations as in-sample data and the remaining observations as out-sample data. In this example, smoothing parameters $\lambda_1 = 0.01$ and $\lambda_2 = 0.005$. Other parameters are the same as these in subsection 3. 1. The results are summarized in Table 2.

3.3 Verifying the nonnegativity of diffusion function

In this part, we generate 600 observations from model (9), and use the first 500 observations as in-sample observations and the remaining as out-sample observations. The simulation data set is presented

Table 2 Comparison between the two methods				
	measure	empirical formula	local polynomial	B-spline
Drift	MADE	ave	7. 338 0	3. 399 7
		std	18. 526 0	1. 273 8
	IMADE	ave	4. 909 8	0. 357 1
		std	18. 662 0	0. 338 6
Volitivity	MADE	ave	0. 210 9	0. 080 6
		std	0. 887 2	0. 072 6
	IMADE	ave	0. 154 9	0. 034 3
		std	0. 884 7	0. 056 9

in Fig. 2(a). Stanton^[6] proposed that the first-order approximation had a bad performance over a small range of values or when the sampling frequency decreased. In order to illustrate that the local polynomial techniques are not suitable for the higher-order approximations, we use the second-order approximation to estimate the squared volatility and set $h = 0.016$, $\lambda_2 = 1 \times 10^{-5}$. In Fig. 2(b) we display the discrete points of second-order approximation of squared volatility, many of which are negative. The results are shown in Fig. 3 and Table 3. It is seen in Fig. 3 that the fitting results of the squared volatility have negative values at the right boundary when we use the local polynomial method.

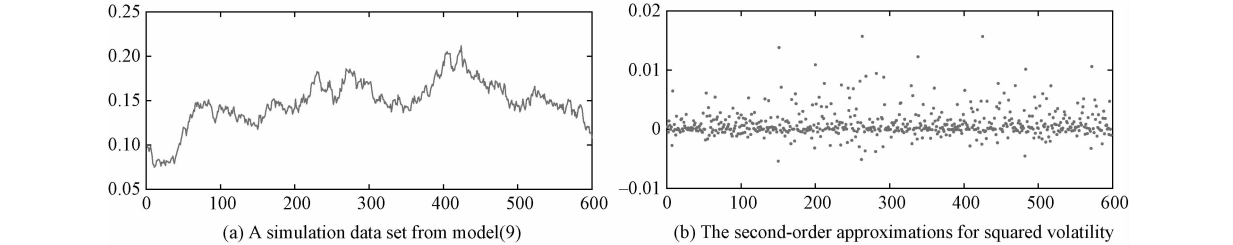


Fig. 2 Simulation data

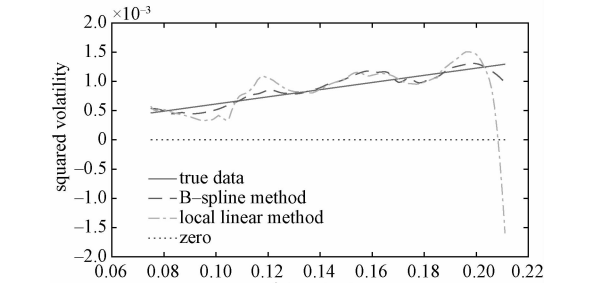


Fig. 3 Estimated squared volatility

This example shows that the restricted B-spline method not only has a lower MADE and IMADE, but also guarantees the nonnegativity of the volatility which is shown in Fig. 3.

Table 3 Comparison between the two volatility estimation methods			
	measure	local polynomial	B-spline
drift	MADE	0. 289 3	0. 289 0
	IMADE	0. 030 7	0. 030 1
volitivity	MADE($\times 10^{-3}$)	1. 500 0	1. 500 0
	IMADE($\times 10^{-4}$)	1. 133 3	0. 998 4

4 Application

4.1 Application of the B-spline techniques to HS300 index

To understand the dynamics of the stock market of China, we apply our proposed techniques to the HS300 index data. The data consist of 2 610 daily

observations from 8 April 2005 to 31 December 2015, and are presented in Fig. 4.

Because the data are the daily ones, the first-order approximation is a good choice in the light of the estimation effect and variance inflation. Following the conventional practice in finance research, we first take the logarithmic transformation of the price index and then obtain the range interval $\varpi = [6.7069, 8.6788]$. We set the number of equally spaced interval knots $J_1 = J_2 = 35$. By using the rule proposed in section 2, smoothing parameters $\lambda_1 = 0.1$ and $\lambda_2 = 0.02$ are chosen. The estimation results

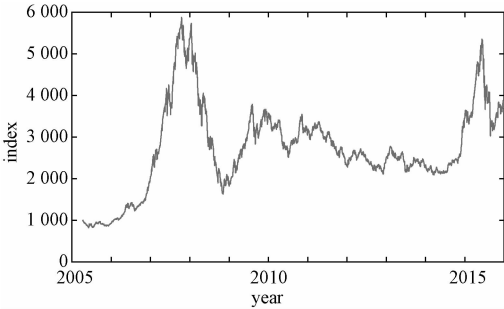
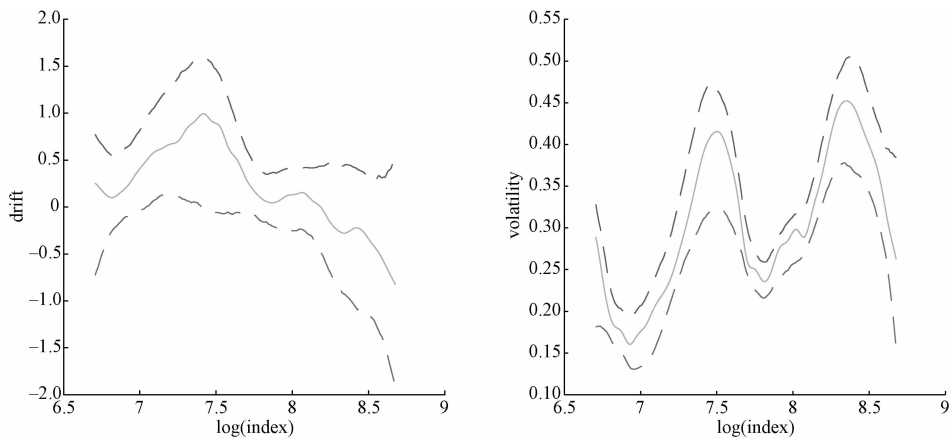


Fig. 4 Index of HS300 from 2005-04-08 to 2015-12-31

of drift and volatility are presented in Fig. 5. The solid curves are the estimation, and the dashed curves are the 95% bootstrap confidence intervals.



Solid—estimator, dashed—95% bootstrap confidence bond.

Fig. 5 Estimation of the coefficients of diffusion processes using B-spline

Regression bootstrap technique is used to calculate the bootstrap confidence intervals (see Ref [17] for details). We denote $S_{t_i} = X_{t_{i+1}} - X_{t_i}$ for the original HS300 data $\{X_{t_i}, i = 1, \cdots, n\}$, and then generate the bootstrap responses $\{S_{t_i}^*\}$ using $S_{t_i}^* = \hat{\mu}(X_{t_i})\Delta + \hat{\sigma}(X_{t_i})\sqrt{\Delta}\varepsilon_i^*$, $i = 1, \cdots, n - 1$, where the bootstrap residuals $\{\varepsilon_i^*, i = 1, \cdots, n - 1\}$ are sampled randomly with replacement from $\{\hat{\varepsilon}_i, i = 1, \cdots, n - 1\}$ and $\hat{\varepsilon}_i = \frac{S_{t_i} - \hat{\mu}(X_{t_i})\Delta}{\hat{\sigma}(X_{t_i})\sqrt{\Delta}}$. It is easy to

obtain the bootstrap samples $\{(X_{t_i}, Y_{i\Delta}^*)_{i=1}^{n-1}\}$ and $\{(X_{t_i}, Z_{i\Delta}^*)_{i=1}^{n-1}\}$ when the first-order approximation is used. If a higher-order approximation is involved, we just need to set $X_{t_1}^* = X_{t_1}$, and then generate a regression bootstrap sample $\{X_{t_i}^*, i = 1, \cdots, n\}$. Similarly, the bootstrap sample $\{(X_{t_i}, Y_{i\Delta}^*)_{i=1}^{n-1}\}$ and $\{(X_{t_i}, Z_{i\Delta}^*)_{i=1}^{n-1}\}$ can be obtained. In our studies, the

bootstrap confidence intervals are calculated based on 1 000 bootstrap samples.

4.2 Treasury bond case study

We apply the two techniques proposed above to the three-year treasury bond from the secondary market rates on the last trading day of each month and compare the estimation effects of squared volatility between the two methods. The data set contains 753 monthly observations from 30 April 1953 to 31 December 2015, and is presented in Fig. 6.

Here we use the second-order approximation to estimate the squared volatility of model (1) for the monthly three-year treasury bond data. The estimation of squared volatility with the 95% bootstrap confidence bond based on 1 000 bootstrap samples by the penalized B-spline with nonnegative restriction techniques is displayed in Fig. 7 (the solid curve). In Fig. 7 the dashed-dotted curve

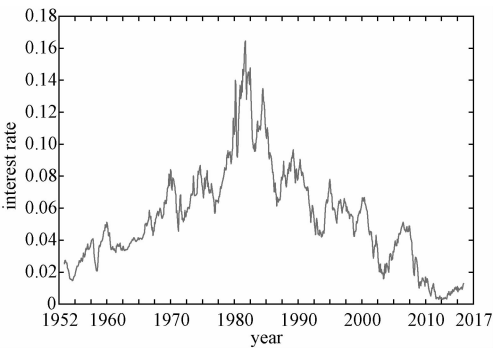


Fig. 6 Three-year treasury bond data set

denotes the estimation obtained by local polynomial techniques. It is obvious that the squared volatility estimated by local polynomial techniques with $h = 0.0289$ is negative when the treasury bond rate is equal to 0.16. This result shows that the fitting effect of penalized B-spline with nonnegative restriction techniques is better than that of the local polynomial. Such effect will be very obvious when the higher-order approximation is used. In this study, we choose $\varGamma = [0.003\ 0, 0.164\ 5]$, $J_2 = 35$, and $\lambda_2 = 1 \times 10^{-5}$.

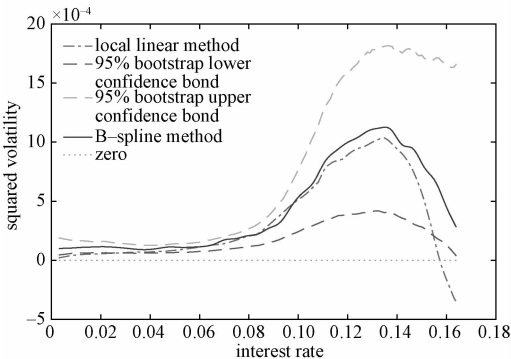


Fig. 7 Estimated squared volatility with 95% bootstrap confidence bond

5 Conclusion

The B-spline method gives a good performance for the drift and volatility estimation. When a higher-order approximation is adopted, the B-spline method is a better choice because it not only has lower MADE and IMADE but also guarantees the nonnegativity of the volatility, while the local polynomial method does not. Our techniques will be extended to time-inhomogeneous diffusion models in the future.

References

[1] Hansen L P, Scheinkman J A. Back to the future: generating

moment implications for continuous-time Markov processes [J]. *Econometrica*, 1995, 63 (4) : 767-804.

[2] Pedersen A R. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes [J]. *Bernoulli*, 1995, 1 (3) : 257-279.

[3] Ait-Sahalia, Yacine. Maximum likelihood estimation of discretely sampled diffusions; a closed-form approximation approach [J]. *Econometrica*, 2002, 70 (1) : 223-262.

[4] Florens-Zmirou D. On estimating the diffusion coefficient from discrete observations [J]. *Journal of Applied Probability*, 1993, 30 (4) : 790-804.

[5] Jiang G J, Knight J L. A nonparametric approach to the estimation of diffusion processes; with an application to a short-term interest rate model [J]. *Econometric Theory*, 1997, 13 (5) : 615-645.

[6] Stanton R. A nonparametric model of term structure dynamics and the market price of interest rate risk [J]. *The Journal of Finance*, 1997, 52 (5) : 1 973-2 002.

[7] Fan J, Zhang C. A reexamination of diffusion estimators with applications to financial model validation [J]. *Journal of the American Statistical Association*, 2003, 98 (461) : 118-134.

[8] Ziegelmann F A. Nonparametric estimation of volatility functions: the local exponential estimator [J]. *Econometric Theory*, 2002, 18 (4) : 985-991.

[9] Yu K, Jones M C. Likelihood-based local linear estimation of the conditional variance function [J]. *Journal of the American Statistical Association*, 2004, 99 (465) : 139-144.

[10] Xu K L. Reweighted functional estimation of diffusion models [J]. *Econometric Theory*, 2010, 26 (2) : 541-563.

[11] Yu Y, Yu K, Wang H, et al. Semiparametric estimation for a class of time-inhomogeneous diffusion processes [J]. *Statistica Sinica*, 2009, 19 (2) : 843-867.

[12] Ramsay J O, Silverman B W. *Functional data analysis* [M]. New York : Springer, 2005.

[13] Ruppert D. Selecting the number of knots for penalized splines [J]. *Journal of computational and graphical statistics*, 2002, 11 (4) : 735-757.

[14] Yoshimoto F, Harada T, Yoshimoto Y. Data fitting with a spline using a real-coded genetic algorithm [J]. *Computer-Aided Design*, 2003, 35 (8) : 751-760.

[15] Fan J, Fan Y, Jiang J. Dynamic integration of time-and state-domain methods for volatility estimation [J]. *Journal of the American Statistical Association*, 2007, 102 (478) : 618-631.

[16] Kloeden P E. On effects of discretization on estimators of drift parameters for diffusion processes [J]. *Journal of Applied Probability*, 1996, 33 (4) : 606-610.

[17] Franke J, Kreiss J P, Mammen E. Bootstrap of kernel smoothing in nonlinear time series [C] // *Bernoulli*. Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 1997: 1-37.