Journal of Mathematical Research & Exposition
May, 2010, Vol. 30, No. 3, pp. 507–518
DOI:10.3770/j.issn:1000-341X.2010.03.015
Http://jmre.dlut.edu.cn

The Integration of Dual-Domain Method for Estimating the Volatility of Financial Assets

Xue Qiao DU, Xu Guo YE*
Department of Mathematics, Hefei University of Technology, Anhui 230009*, P. R. China*

Abstract Time- and state-domain methods are two common approaches for nonparametrically estimating the volatility of financial assets. Economic conditions vary over time in real financial market. It is reasonable to expect that volatility depends on both time and price level for a given state variable. Recently, Fan, et al (2007) proposed the idea of dynamically integrated method in both time-and state domain. This idea has become an interesting topic in the estimation of volatility. In this paper, our purpose is to discuss the integrated method in the estimation of volatility. Simulations are conducted to demonstrate that the newly integrated method outperforms some old ones, and the results of simulations demonstrate this fact. Furthermore, we establish its asymptotic properties.

Keywords Volatility; estimation; integrated estimator.

Document code A
MR(2000) Subject Classification 60J60; 62G05; 62G20; 91B28
Chinese Library Classification O212.7

1. Introduction

We consider the problem of estimating the diffusion coefficient, $\sigma(\cdot)$, for a continuous-time diffusion process X_t following the stochastic differential equation

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad 0 \le t \le T, \tag{1}$$

where $\mu(\cdot)$ is drift coefficient, W_t is a standard one-dimensional Brownian motion. This model has been widely used for describing the price process of the financial assets.

There is a large literature on the estimation of the drift and volatility coefficient. Yao and Tong [1] proposed the direct estimator $\widehat{\sigma_t^2}(x) = \widehat{v}(x) - \{\widehat{m}(x)\}^2$. There mainly exist two problems. First, it cannot completely hold for $\widehat{\sigma_t^2}(x) \ge 0$; Secondly, there is a big bias. Härdle and Tsybakov [2] applied the local polynomial regression estimator to reduce the large bias, but they did not solve the first one, and this method is not fully effective to the unknown drift

Received April 15, 2008; Accepted January 1, 2009
Supported by the Hefei University of Technology's Excellent Courses - Probability Theory and Mathematical Statistics and "Eleventh Five -Year" national scientific and technological support projects-GIS -based cluster of manufacturing industries to support R&D and modeling of technical service system - issue 5 - auto parts industry cluster system modeling (Grant No. 2008BAF35B05).
* Corresponding author
E-mail address: dxq1006@126.com (X. Q. DU); yexuguo522@126.com (X. G. YE)

coefficient, $m(\cdot)$. Others proposed the difference-based estimator. Rice [3, 4], Gasser, et al. [5], and Müller and StadtMüller [6] used a high-pass filter to remove the regression function from the data sequence $\{Y_i\}$, but Hall, Kay and Titterigton [7] pointed out the resulting estimator was inefficient even in some homoscedastic models with optimal filters. Then Hall, et al. [7] employed the optimal difference sequence for a Gaussian model and got a certain degree of effectiveness. For the purpose of acquiring the higher efficiency of the methods of estimation, the residual-based estimators had been proposed for the coming years. Stone [8]. Hall and Carroll [9], Müller and StadtMüller [10], Neumann [11], Fan and Gijbel [12],Gallant and Tauchen [13], Stanton [14], Jiang and Knight [15], Fan and Yao [16], Arfi [17], Fan, et al.[18], Fan and Zhang [19] had proposed various nonparametric methods and respectively proved their advantages. But now the local linear estimator is attractive to estimate the variance function. Because we know economic conditions vary over time, it is reasonable to expect that the volatility depends on both time and price level for a given state variable. But we find there is no sufficient information to estimate the bivariate functions in (1) without further restrictions. Because of consistently estimating the volatility function, $\sigma(x,t)$, we need to have data that eventually fill up a neighborhood of the point (x,t). Fan, et al.[20] firstly proposed the dynamic integration of time-and state-domain methods for volatility estimation, and further established the aggregation of nonparametric estimators for volatility matrix. It is indicated that the new concept for estimating volatility is further developed in modern financial analysis.

For most practical situations, we use the Euler scheme to approximate the diffusion process on the basis of the fact that while higher order can possibly reduce approximation errors, it increases variances of data substantially. Furthermore, it is reported that the difference between the Euler approximation scheme and the strong order-one approximation is negligible by simulations. For an overview, see the recent literature by Fan [21]. Here we use the Euler approximation scheme. Suppose that we have historic data $\{X_{t_i}\}_{i=1}^{n+1}$ from the process (1) with a sampling interval Δ at time t. Set $Y_i = (X_{t_{i+1}} - X_{t_i})/\Delta^{1/2}$. Then for the model (1), we have

$$Y_i \approx \mu(X_{t_i})\Delta^{\frac{1}{2}} + \sigma(X_{t_i})\varepsilon_i, \tag{2}$$

where $\varepsilon_i = (W_{t_{i+1}} - W_{t_i})/\Delta^{1/2}$ and $\varepsilon_i \sim_{\text{i.i.d}} N(0,1)$ for $i = 1, 2, \ldots, n$.

In this paper, firstly we will focus on comparing the integrated estimators and establishing asymptotic properties of the proposed estimator, and then illustrate our ideas by numerical simulations and analysis. Finally, we collect outlines of the conditions and the proofs in the Appendix.

2. Estimation of volatility

The volatility estimation is an important part of modern financial econometrics and time series analysis, which almost refers to every aspect of the financial field. There is a large literature on the estimation of volatility based on time-domain and state-domain smoothing. For an overview, see the recent book and literature in references.

2.1. Time-domain estimator

We firstly introduce the following smoothing estimator

$$\widehat{\sigma^2}_{WS,t} = \sum_{i=1}^{n} \alpha_i Y_{t-i}^2, \tag{3}$$

where α_i $(i = 1, 2, \ldots, n)$ are the weights of local data $\{Y_i\}$ in the estimator and $\sum_{i=1}^{n} \alpha_i = 1$.

Some popular versions of time-domain smoothing estimator are proposed by adapting the above estimator. With the development of methodology, time-domain smoothing method had been extensively discussed in modern financial analysis, we refer to [21] for an overview. Recently we discuss the exponential weighted moving estimtor ($\widehat{\sigma^2}_{EWM} = \frac{1-b}{1-b^n} \sum_{i=1}^{n} b^{i-1} Y_{t-i}^2$, where b is smoothing parameter). Since the weight decays exponentially, it essentially uses the recent data. But one notes that the weight drops rapidly, which means that the relation of the returns between Y_{t-i} and Y_{t-j} $(i \neq j)$ is rapidly decreasing. Here we choose a nonparametric kernel estimator (one side kernel estimator) for estimating the volatility. We choose the Epanechnikov kernel in the nonparametric kernel estimator. Indeed it is the neighbor estimator and illustrates its advantages over the adaptive moving estimator. So we use the kernel estimator to estimate the volatility.

$$\widehat{\sigma^2}_{KS,t} = \sum_{i=1}^{n} K(\frac{i}{n}) Y_{t-i}^2 / \sum_{i=1}^{n} K(\frac{i}{n}). \tag{4}$$

All of the time domain smoothing is based on the assumption that the asset returns Y_{t-1}, Y_{t-2}, Y_{t-3}, ... , Y_{t-n} have approximately the same volatility.

Theorem 1 *If $\Delta \to 0$ and $n\Delta \to 0$ $(n \to +\infty)$, conditions (1)–(2) and lemmas in Appendix hold. We have*

$$\sqrt{nh}(\widehat{\sigma^2}_{KS,t} - \sigma_t^2) \xrightarrow{D} N\Big(0, 2\sigma_t^4 \frac{\int K^2(u) \mathrm{d}u}{p(t)}\Big).$$

2.2. State-domain estimator

For practical analysis of financial data, it is hard to determine whether the sampling interval tends to zero. It is reasonable that a method is applicable whether or not "Δ" is small. So here we propose the following nonparametric estimators in state-domain.

To estimate the volatility, we need estimate $f = \mu(x)\Delta^{1/2}$ in (2). Here we use the local linear estimator to estimate it (Indeed for simplifying the process, the estimator $\widehat{\sigma^2}_{S,t}$ behaves as if the drift function f is known). We exclude the n data points used in the time-domain estimator. The historical data at time t are $\{(X_{t_i}, Y_i), i = 1, \ldots, N-n\}$. Denote the squared residuals by $\hat{R}_i = \{Y_i - \hat{f}(X_{t_i})\}^2$. The local constant estimator($\hat{a}$) and the local linear estimator($\hat{\alpha}$) respectively are given by

$$\hat{a} = \arg\min_{a} \sum_{i=1}^{N-n} \{\hat{R}_i - a\}^2 U_{h_1}(X_{t_i} - x) \tag{5}$$

and

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{N-n} \{\hat{R}_i - \alpha - \beta(X_{t_i} - x)\}^2 W_{h_2}(X_{t_i} - x), \tag{6}$$

where $U_{h_1}(\cdot) = U(\cdot/h_1)/h_1$ and $W_{h_1}(\cdot) = W(\cdot/h_1)/h_1$ are kernel functions, h_1 and h_2 are the smoothing parameters.

Note 1 For practical application, Fan and Gijbels [22] recommended the use of the local linear fit ($q = 1$). For an overview, see the recent book by Fan and Gijbels [22]. However, to avoid zero in the denominator, we add N^{-2} in the denominator. This point is revised for practical application in the pointwise estimation. But it has no impact on understanding the following theorems and theoretically analyzing this method.

Theorem 2 *Set $v_j = \int u^j K^2(u)du$ for $j = 0, 1, 2$. Suppose that the second derivatives $\mu(\cdot)$ and $\sigma^2(\cdot)$ exist in a neighborhood of x, $h_n \to 0$ and $nh_n \to \infty (n \to \infty)$. Conditions (3)–(5) in the Appendix hold. Then we have*

(A) Asymptotic normality of $\widehat{\sigma^2}_{KE,S}$:

$$\sqrt{(N-n)h_1}(\widehat{\sigma^2}_{KE,S} - \sigma_t^2) \xrightarrow{D} N(0, 2\sigma_t^4 v_0 p(x)^{-1})$$

(B) Asymptotic normality of $\widehat{\sigma^2}_{LE,S}$:

$$\sqrt{(N-n)h_2}(\widehat{\sigma^2}_{LE,S} - \sigma_t^2 - \theta_n) \xrightarrow{D} N(0, 2\sigma_t^4 p(x)^{-1} e_1^T (H^{-1}S^{-1})^T S^* H^{-1} S^{-1} e_1)$$

where $\theta_n = \frac{1}{2} h_2^2 \ddot{\sigma^2}(x) \sigma_w^2 + o(h_2^2)$, $\sigma_w^2 = \int u^2 w(u)$du.

3. Integrated estimator

In this section, we firstly prove that the time-domain and state-domain estimator are nearly independent (namely, asymptotic independent property), then we choose the dynamic weights by the corresponding criterion.

Theorem 3 *Suppose that the conditions of Theorems 1 and 2 are satisfied. Then we have*

(a) Asymptotic independence:

$$\Sigma_1^{-\frac{1}{2}} \begin{pmatrix} \sqrt{nh}[\widehat{\sigma^2}_{KS,t} - \sigma_t^2)] \\ \sqrt{(N-n)h_1}[\widehat{\sigma^2}_{KE,s} - \sigma_t^2] \end{pmatrix} \longrightarrow N(0, I_2),$$

where $\Sigma_1^{-\frac{1}{2}} = \mathrm{diag}\{V_2, 2\sigma_t^4 v_0 p(x)^{-1}\}$ and $V_2 = 2\sigma_t^4 \frac{\int K^2(u)\mathrm{d}u}{p(t)}$.

(a') Asymptotic independence:

$$\Sigma_2^{-\frac{1}{2}} \begin{pmatrix} \sqrt{nh}[\widehat{\sigma^2}_{KS,t} - \sigma_t^2)] \\ \sqrt{(N-n)h_2}[\widehat{\sigma^2}_{LE,S} - \sigma_t^2 - \theta_n] \end{pmatrix} \longrightarrow N(0, I_2),$$

where $\Sigma_2^{-\frac{1}{2}} = \mathrm{diag}\{V_2, 2\sigma_t^4 p(x)^{-1} e_1^T H^{-1} S^{-1} S^* (H^{-1}S^{-1})^T e_1\}$ and $V_2 = 2\sigma_t^4 \frac{\int K^2(u)\mathrm{d}u}{p(t)}$.

On the basis of time- and state-domain smoothing, we propose the integrated form with the dynamic optimal weights

$$\widehat{\sigma^2}_{I,(s,t)} = W_t\widehat{\sigma^2}_{t,\mathrm{time}} + (1-W_t)\widehat{\sigma^2}_{t,\mathrm{state}}, \tag{7}$$

where W_t $(0 \leq W_t \leq 1)$. Furthermore, we can get the dynamic optimal weights by minimizing the variance of the integrated estimator, $W_t = \frac{\mathrm{Var}(\widehat{\sigma^2}_{t,\mathrm{state}})}{\mathrm{Var}(\widehat{\sigma^2}_{t,\mathrm{state}})+\mathrm{Var}(\widehat{\sigma^2}_{t,\mathrm{time}})}$. Therefore, we get the integrated estimator:

$$\widehat{\sigma^2}_{I,(s,t)} = \widehat{W_t}\widehat{\sigma^2}_{t,time} + (1-\widehat{W_t})\widehat{\sigma^2}_{t,\mathrm{state}}. \tag{8}$$

4. Numerical simulations and analysis

As an illustration, we use the simple abbreviation in Table 1 to denote several estimators.

KET: the kernel estimator in time-domain
KES: the kernel estimator in state-domain
LLE: the local linear estimator in state-domain
IE: the integrated estimator in Fan, et al.([20],$\lambda = 0.94$)
NIE 1: the New integrated estimator of time and state domain in (5)
NIE 2: the New integrated estimator of time and state domain in (6)

Table 1 Abbreviations of time-domain, state-domain and integrated estimators

Example 1 As a first illustration, we consider the well-known Cox-Ingersoll-Ross (CIR) model:

$$\mathrm{d}X_t = k(\theta - X_t)\mathrm{d}t + \sigma X_t^{1/2}\mathrm{d}W_t, \quad t \geq t_0, \tag{9}$$

where the spot rate X_t, moves around a central location or long-run equilibrium level $\theta = 0.08571$ at speed $k = 0.21459$, here σ is set to be 0.07830 ([19]).

Here we use the approximation scheme to generate 10000 samples. The scheme takes the form

$$X_{t_{i+1}} - X_{t_i} \approx k(\theta - X_{t_i})\Delta + \sigma X_{t_i}^{1/2}\Delta^{1/2}\varepsilon_i \text{ for } \leq i \leq n, \tag{10}$$

where $\varepsilon_i = \frac{W_{t_{i+1}} - W_{t_i}}{\Delta^{1/2}}$, $\varepsilon_i \sim_{\mathrm{i.i.d}} N(0,1)$. For each simulation experiment, we generate from a sample path of length 10000. We replicate the experiments 1000 times. Although choices of kernel function play an import role to estimate the volatility, choices of kernel function depend purely on individual references. Here we use the Epanechnikov kernel ($K(u) = 0.75(1-u^2)I(|u| \leq 1)$), where $I(\cdot)$ stands for the indicator function. For a given kernel function, choices of an effective bandwidth parameter is very important to the performance of a non-parametric estimator [11, 22, 23, 24]. For the purpose of illustration, here we set $h = 0.04$. In our implementation, we usually set $\Delta = 1/12$ (yearly); $1/52$ (monthly); or $1/252$ (daily) in the state domain. Here we set $\Delta = 1/252$. The initial value can be got from the distribution of X_t ([19]). We focus on an interior state point $x = 0.1$ in the simulated calculation.

Estimator	Measure	NP	IMADE
KET	Ave	0.006133086435664	7.493756945247402e-005
	Std	4.575213916349301e-009	1.769163422049167e-009
KES	Ave	0.006133610087099	6.757124298279898e-005
	Std	3.624624686197756e-009	1.343031278355880e-009
LLE	Ave	0.00613190566218	8.500395125373502e-005
	Std	5.711765001730663e-009	2.101030167524508e-009
IE	Ave	0.006122234037852	7.716647442301703e-005
	Std	1.800156175593209e-008	1.503926863713015e-008
NIE 1	Ave	0.006134398849847	6.963277222752401e-005
	Std	3.818917458841625e-009	1.395951926723230e-009
NIE 2	Ave	0.006129193172399	7.551176036654199e-005
	Std	4.405727101088140e-009	1.556270394655766e-009

Table 2 Comparison of time-domain, state-domain and integrated estimators and $\sigma^2 = 0.00613089$

To assess the performance of the six estimators in Table 1, we compute the average and the root mean squared error of each of the two measures over 1000 simulations. "NP" stands for the number properties of estimating data : average and root mean squared error of the sampleing data. "IMADE" stands for the ideal mean absolute deviation error (IMADE $= m^{-1}\sum_{i=T+1}^{T+m}|\widehat{\sigma_i^2} - \sigma_t^2|$).

The performance of each volatility estimation is described in Table 2, which shows that the performance of the integrated estimators outperform the Fan' estimator. Table 2 shows that two estimators (KET, KES) outperform the latter (LLE) for the same kernel function and sample number. It indicates that the integrated estimator from combining KET and KES should be better than the integrated estimator from KET and KES, this guess is illustrated in Table 2. It shows that the performance of the two new integrated estimators uniformly dominate the other estimator. But it depends on the performance of the time-and state-domain estimators. These results are derived by simulations under the same conditions, which only indicate some facts. If we want to get the better results, we should consider every aspect in financial markets. But further study on this topic is beyond our purpose in this paper.

Example 2 We now consider another familiar example of Geometric Brownian Motion (GBM) model:

$$\mathrm{d}X_t = (\mu + 2^{-1}\sigma^2)X_t\mathrm{d}t + \sigma X_t\mathrm{d}W_t, \quad 0 < t < T, \tag{11}$$

where W_t is a standard one-dimensional Brownian motion. For Example 2, apparently both the drift and diffusion are linear, and thus $\{X_t\}$ is Markovian [19].

We simulate in time $[0, T]$ with $T = 20$, the corresponding approximate process with parameters $\mu = 0.087$ and $\sigma = 0.178$. We choose the order-0.5 approximation scheme.

$$\frac{X_{t_{i+1}} - X_{t_{i+1}}}{\Delta^{1/2}} = (\mu + 2^{-1}\sigma^2)X_{t_i}\Delta^{1/2} + \sigma X_{t_i}\varepsilon_i \quad \text{for } 1 \le i \le n, \tag{12}$$

where $\varepsilon_i = \frac{W_{t_{i+1}} - W_{t_i}}{\Delta^{1/2}}$ and $\varepsilon_i \sim_{\text{i.i.d}} N(0,1)$. Alternatively, we could directly use the explicit solution $X_t = X_0 \exp\{\mu t + \sigma W_t\}$ for equation (11). We focus on an interior state point $X_0 = 1.0$; 1,000 sample paths of length 50, 000 are generated. The bandwidth parameter, $h = 0.004$ and $\Delta = 1/250$, is used respectively for local smoothing. Some demands are the same as in Example 1.

Estimator	Measure	NP	IMADE
KET	Ave	0.031684527781019	1.787526876911020e-004
	Std	2.451424005418594e-008	8.546524900690302e-009
KES	Ave	0.031785657829251	8.876827267321372e-004
	Std	1.087330178264748e-006	6.940339004959433e-007
LLE	Ave	0.031816822310177	0.001005895072185
	Std	1.892250407349083e-006	1.387725684910996e-006
IE	Ave	0.031689301327384	9.323532566736667e-004
	Std	1.233307439126375e-006	7.994656071184975e-007
NIE 1	Ave	0.031679627805908	1.794410737083469e-004
	Std	2.507962808329603e-008	8.989067684178025e-009
NIE 2	Ave	0.031680848561931	1.860336227279028e-004
	Std	2.835474432049124e-008	1.106155147933674e-008

Table 3 Comparison of time-domain, state-domain and integrated estimators and $\sigma^2 = 0.031684$

To assess the performance of the six estimators in Table 1, we compute the average and the root mean squared error of each of the two measures over 1000 simulations. "NP" stands for the number properties of estimating data: average and root mean squared error of the sampleing data. "IMADE" stands for the ideal mean absolute deviation error (IMADE $= m^{-1} \sum_{i=T+1}^{T+m} |\widehat{\sigma_i^2} - \sigma_t^2|$).

The results from the simulated data are reported in Table 3. In order to study the influence of the sample number on the effect of estimators, we respectively use 10000 smaples in Example 1 and 50000 samples in Example 2 to illustrate our method. Comparing Table 2 with Table 3 shows that the time-domain estimator with Epanechnikov kernel is attractive to estimate the volatility and depends tightly on sample number. Furthermore, we use the difference bandwidth between Examples 1 and 2 to illustrate the influence of the difference bandwidth on the performance of the estimators. It is also shown that the integrated estimator by integrating the good time-and state-domain estimator can get more efficient estimator in Tables 2 and 3. By simulations, Tables 2 and 3 report that the new integrated estimator outperforms the estimator of the integration method of time and state domains Fan, et al.([20],$\lambda = 0.94$).

5. Appendix

We always use C to denote a generic constant which may be different at different places. We introduce the following conditions.

Condition 1 $\sigma^2(x)$ is Lipschitz conditions.

Condition 2 There exists a constant $C > 0$, such that $E|\mu(X_s)|^{2(p+\delta)} \leq C$ and $E|\sigma(X_s)|^{2(p+\delta)} \leq C$ for any $s \in [t-\eta, t]$, where η is some positive constant, p is an integer not less than 2 and $\delta > 0$.

Condition 3 The strictly stationary process $\{(X_{t_i}, Y_i), i = 1, \ldots, n-1\}$ is absolutely regular, i.e.,

$$\beta(j) = \sup_{i \geq 0} E\{\sup_{A \in \omega_{i+j}^{\infty}} |Pr(A|\omega_1^i)| - Pr(A)\} \to 0,$$

as $j \to \infty$, where ω_i^j is the $\sigma-$field generated by $\{(X_{t_i}, Y_i), i = 1, \ldots, n-1\}$ $(j \geq i)$. Furthermore, for the same δ as in Condition 2, $\sum_{j=0}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}}(j) < \infty$, we use the convention $0^0 = 0$.

Condition 4 The discrete observations $\{X_{t_i}\}_{i=1}^{n+1}$ satisfy the stationarity condition of Banon [25]. Furthermore, a stationary process X_t is said to satisfy the condition $G_2(s, \alpha)$ of Rosenblatt [26].

Condition 5 The conditional density $p_l(y|x)$ of $X_{t_{i+1}}$ for given X_{t_i} is continuous in the arguments (x, y) and is bounded by a constant (independent of l).

Lemma 1 *Under Conditions (1)–(2) in the Appendix, we have*

$$|\sigma^2(X_s) - \sigma^2(X_u)| \leq \lambda |s-u|^{\frac{p-1}{2p}}$$

for any $s, u \in [t-\eta, t]$, where the coefficient C satisfies $E(\lambda^{2(p+\delta)}) < \infty$ and η is a positive constant.

Proof of Theorem 1 Let $Z_{i,s} = (r_s - r_{t_i})^2$. Applying Itô formula to $Z_{i,s}$, we obtain

$$\begin{aligned}
\mathrm{d}Z_{i,s} &= 2\Big(\int_{t_i}^{s} \mu_u \mathrm{d}u + \int_{t_i}^{s} \sigma_u \mathrm{d}W_u\Big)(\mu_s \mathrm{d}s + \sigma_s \mathrm{d}W_s) + \sigma_s^2 \mathrm{d}s \\
&= 2\Big[(\int_{t_i}^{s} \mu_u \mathrm{d}u + \int_{t_i}^{s} \sigma_u \mathrm{d}W_u)\mu_s \mathrm{d}s + \sigma_s(\mu_s \mathrm{d}s + \sigma_s \mathrm{d}W_s)\Big] + 2\Big(\int_{t_i}^{s} \sigma_u \mathrm{d}W_u\Big)\sigma_s \mathrm{d}W_s + \sigma_s^2 \mathrm{d}s.
\end{aligned}$$

Then Y_i^2 can be decomposed as $Y_i^2 = 2a_i + 2b_i + \tilde{\sigma_i^2}$, where

$$a_i = \Delta^{-1}\Big[\int_{t_i}^{t_{i+1}} \mu_u \mathrm{d}s \int_{t_i}^{s} \mu_u \mathrm{d}u + \int_{t_i}^{t_{i+1}} \mu_s \mathrm{d}s \int_{t_i}^{s} \sigma_u \mathrm{d}W_u + \int_{t_i}^{t_{i+1}} \sigma_s \mathrm{d}W_s \int_{t_i}^{s} \mu_u \mathrm{d}u\Big]$$

$$b_i = \Delta^{-1} \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \sigma_u \mathrm{d}W_u \mathrm{d}\sigma_s W_s, \ \tilde{\sigma_i^2} = \Delta^{-1} \int_{t_i}^{t_{i+1}} \sigma_s^2 \mathrm{d}s.$$

Therefore,

$$\begin{aligned}
\widehat{\sigma^2}_{KS,t} - \sigma_t^2 &= \sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}[2a_i + 2b_i + \tilde{\sigma_i^2} - \sigma_t^2] \\
&= 2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})} a_i + 2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})} b_i + \sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}(\tilde{\sigma_i^2} - \sigma_t^2).
\end{aligned}$$

Lemma 2 *Under Conditions (1)–2, if $n \to \infty$, $\Delta \to 0$ and $n\Delta \to 0$, then*

$$\Big|\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}(\tilde{\sigma_i^2} - \sigma_t^2)\Big| = \Big|\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}(\tilde{\sigma_i^2} - \sigma_t^2)\Big| \to 0 \quad (n \to +\infty),$$

$$\Big|2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}a_i\Big| = 2\Big|\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}a_i\Big| \to 0 \quad (n \to +\infty).$$

Lemma 3 *If Conditions (1)–(2) are satisfied, then*

$$b_j = \Delta^{-1}\int_{t_i}^{t_{i+1}}\int_{t_i}^{s}\sigma_u \mathrm{d}W_u \sigma_s \mathrm{d}W_s = \sigma_t^2\Delta^{-1}\int_{t_i}^{t_{i+1}}(W_s - W_{t_i})\mathrm{d}W_s + \gamma_j,$$

where

$$\gamma_j = \Delta^{-1}\int_{t_i}^{t_{i+1}}(\sigma_s - \sigma_t)\Big[\int_{t_i}^{s}\sigma_u \mathrm{d}W_u\Big]\mathrm{d}W_s + \Delta^{-1}\sigma_t\int_{t_i}^{t_{i+1}}\Big[\int_{t_i}^{s}(\sigma_s - \sigma_t)\mathrm{d}W_u\Big]\mathrm{d}W_s.$$

Because

$$2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}b_i = 2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\sigma_t^2\Delta^{-1}\int_{t_i}^{t_{i+1}}(W_s - W_{t_i})\mathrm{d}W_s + 2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\gamma_i,$$

we have $|2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\gamma_i| = 2|\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\gamma_i| \to 0$ $(n \to +\infty)$. Therefore, we only consider

$$\begin{aligned}
&2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\sigma_t^2\Delta^{-1}\int_{t_i}^{t_{i+1}}(W_s - W_{t_i})\mathrm{d}W_s \\
&= 2\sigma_t^2\Delta^{-1}\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\Big[\int_{t_i}^{t_{i+1}} W_s \mathrm{d}W_s - W_{t_i}\mathrm{d}W_s\Big] \\
&= 2\sigma_t^2\Delta^{-1}\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\Big[\frac{1}{2}(W_{t_{i+1}}^2 - W_{t_i}^2) - \frac{\Delta}{2} - W_{t_i}(W_{t_{i+1}} - W_{t_i})\Big] \\
&= 2\sigma_t^2\Delta^{-1}\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\frac{1}{2}(W_{t_{i+1}} - W_{t_i})^2 - \sigma_t^2 \\
&= \sigma_t^2\Delta^{-1}\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\Big(\frac{W_{t_{i+1}} - W_{t_i}}{\Delta^{1/2}}\Big)^2 - \sigma_t^2 \\
&= \sigma_t^2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}\Big[(\frac{W_{t_{i+1}} - W_{t_i}}{\Delta^{1/2}})^2 - 1\Big] \\
&= \sigma_t^2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}[\varepsilon_i^2 - 1].
\end{aligned}$$

Conditions 2 and 3 imply that $E\{W_{ni}(t)\sigma^2(X_{t_i})(\varepsilon_i^2 - 1)\}^{2+\frac{\delta}{2}} < \infty$, by Theorem 1 of Fan and Yao [16]. $\sigma_t^2\sum_{i=1}^{n} \frac{K(\frac{t_i-t}{h})}{\sum_{i=1}^{n} K(\frac{t_i-t}{h})}[(\frac{W_{t_{i+1}} - W_{t_i}}{\Delta^{1/2}})^2 - 1]$ is asymptotically normal mean 0 and variance $2\sigma_t^4\frac{\int K^2(u)\mathrm{d}u}{nhp(t)}$. Therefore, we easily get

$$\sqrt{nh}(\widehat{\sigma^2}_{KS,t} - \sigma_t^2) \to N\Big(0, 2\sigma_t^4\frac{\int K^2(u)\mathrm{d}u}{p(t)}\Big).$$

Parts of proofs for Lemma 1-3 and Theorem 1 can be found in Fan's literatrue [21]. □

Proof of Theorem 2 Without loss of generality, we assume that $f(x) = 0$, hence $\widehat{R_i} = Y_i^2$. (A) is easily proved [16]. By [27], let $Y = (Y_1^2, Y_2^2, \ldots, Y_{N-n}^2)$, $W = \mathrm{diag}\{W(\frac{X_{t_1}-x}{h_2}), W(\frac{X_{t_2}-x}{h_2}), \ldots, W(\frac{X_{t_{N-n}}-x}{h_2})\}$ and

$$X = \begin{pmatrix} 1 & X_{t_1} - x \\ \vdots & \vdots \\ 1 & X_{t_{N-n}} - x \end{pmatrix}.$$

Denote by $m_i = E[Y_i^2|X_{t_i}]$, $\mathbf{m} = (m_1, m_2, \ldots, m_{N-n})^{\mathrm{T}}$ and $e_1 = (1,0)^{\mathrm{T}}$. Then it can be written that

$$\begin{aligned} \widehat{\sigma^2}_{LE,S} &= e_1^{\mathrm{T}}(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}WY, \\ \widehat{\sigma^2}_{LE,S} - \sigma_t^2 &= e_1^{\mathrm{T}}(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W\{m - X\beta_N\} + e_1^{\mathrm{T}}(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W\{Y - m\} \\ &= e_1^{\mathrm{T}}B + e_1^{\mathrm{T}}b, \end{aligned}$$

where $\beta_N = (m(x), m'(x))^{\mathrm{T}}$ with $m(x) = E[Y_1^2|X_{t_i} = x]$. By [16], the bias vector B converges in probability to a vector B with $B = O(h_2^2) = o(1/\sqrt{(N-n)h_2})$. In the following, we will show that the centralized vector b is asymptotically normal.

Put $u = (N-n)^{-1}H^{-1}X^{\mathrm{T}}W(Y-m)$ where $H = \mathrm{diag}\{1, h_2\}$, then by [27], the vector b can be written as

$$b = p^{-1}(x)H^{-1}S^{-1}u(1 + o_p(1)), \tag{13}$$

where $S = (\mu_{i+j-2})$, $i,j = 1,2$ with $\mu_j = \int u^j k(u)\mathrm{d}u$. For any constant vector c, define

$$Q_N = c^T u = \frac{1}{2}\sum_{i=1}^{N-n}\{Y_i^2 - m_i\}C_{h_2}(X_{t_i - x}),$$

where $C(u) = c_1W(u) + c_2uW(u)$ with $C_u(h_2) = C(u/h_2)/h$. Applying the "big-block" and "small-block" arguments in [27, Theorem 6.3], we obtain

$$\theta^{-1}(x)\sqrt{(N-n)h_2}Q_N \to N(0,1), \tag{14}$$

where $\theta^2(x) = 2p(x)\sigma_t^4\int C^2(u)\mathrm{d}u$. Namely,

$$\sqrt{(N-n)h_2}c^Tu \to N\Big(0, 2p(x)\sigma_t^4\int C^2(u)\mathrm{d}u\Big).$$

Because Q_N is a linear transform of $\mathbf{u}$, we have

$$\sqrt{(N-n)h_2}u \to N(0, 2\sigma_t^4p(x)S^*/(N-n)h_2),$$

where $S^* = (v_{i+j-2})$, $i,j = 1,2$, with $v_j = \int u^jK^2(u)\mathrm{d}u$. So we can reduce to $u \to N(0, \frac{2\sigma_t^4p(x)S^*}{(N-n)h_2})$.

Because that $b = p^{-1}(x)H^{-1}S^{-1}u(1 + o_p(1))$, we have

$$b \to N(0, 2\sigma_t^4p(x)^{-1}(H^{-1}S^{-1})^{\mathrm{T}}S^*H^{-1}S^{-1}/(N-n)h_2).$$

Furthermore, we easily get $\widehat{\sigma^2}_{LE,S} - \sigma_t^2 \to N(0, 2\sigma_t^4p(x)^{-1}e_1^{\mathrm{T}}(H^{-1}S^{-1})^{\mathrm{T}}S^*e_1H^{-1}S^{-1}/(N-n)h_2)$. □

Proof of Theorem 3 If Theorem 3 (a') is proved, then we can easily get the proof of Theorem 3 (a). Because of the above results, we only prove nearly independent. In the following, we will decompose Q_N into two parts $Q_N^{/}$ and $Q_N^{//}$, which satisfy that

(i) $(N-n)h_2E[\theta^{-1}(x)Q_N^{/}]^2 \leq \frac{h_2}{N-n}(h_2^{-1}a_{N-n}(1+o(1))+(N-n)o(h_2^{-1})) \to 0$;

(ii) $Q_N^{//}$ is identically distributed as Q_N and is asymptotically independent of $\widehat{\sigma^2}_{KS,t}$.

Define

$$Q_N^{/} = \frac{1}{N-n}\sum_{i=1}^{a_{N-n}}\{Y_i^2 - E[Y_i^2|X_{t_i}]\}C_{h_2}(X_{t_i}-x) \tag{15}$$

and $Q_N^{//} = Q_N - Q_N^{/}$, where a_{N-n} is a positive integer with $a_{N-n} = o(N-n)$ and $a_{N-n}\Delta \to \infty$. Let $v_{N,l} = v_{l+1}\sqrt{h_2}$ and $v_i = \{Y_i - m_i\}C_{h_2}(X_{t_i}-x)$ $(i=1,2,\ldots,n)$. Then by [27]

$$\mathbf{VAR}[\theta^{-1}(x)v_{N,0}] = (1+o(1)) \quad \text{and} \quad \sum_{l=1}^{n-1}|\mathbf{Cov}(v_{N,0}, v_{N,l+1})| = o(1),$$

which yields the result in (i). Combining this with equation (7), (i) and equation (8) leads to

$$\theta^{-1}(x)\sqrt{(N-n)h_N}Q_N^{//} \to N(0,1).$$

According to the stationarity conditions of Banon [25], a stationary process X_t is said to satisfy the condition $G_2(s,\alpha)$ of Rosenblatt [26] and the Proposition 2.6 of Fan and Yao [27] imply that the $\rho(l)$ of $\{X_{t_i}\}$ decays exponentially and the strong-mixing coefficient $\alpha(l) \leq \rho(l)$. It follows that

$$|E\exp\{i\zeta(Q_N^{//} + \widehat{\sigma^2}_{KS,t})\} - E\exp\{i\zeta(Q_N^{//}\}E\exp\{i\zeta\widehat{\sigma^2}_{KS,t})\}| \leq 16\alpha(s_N) \to 0$$

for any $\zeta \in R$. By theorem of Volkonskii and Rozanov (1959) or Revuz and Yor [28], we get the asymptotic independence of $\widehat{\sigma^2}_{KS,t}$ and $Q_N^{//}$.

By (i), $\sqrt{(N-n)h_2}Q_N'$ is asymptotically negligible. This together with Theorem1 leads to

$$d_1\theta^{-1}(x)\sqrt{(N-n)h_2}Q_N + d_2V_2^{-1/2}\sqrt{nh}[\widehat{\sigma^2}_{KS,t} - \sigma_t^2] \to N(0, d_1^2 + d_2^2)$$

for any $d_1, d_2 \in R$, where $V_2 = 2\sigma_t^4\frac{\int K^2(u)\mathrm{d}u}{p(t)}$. Since Q_N is a linear transform of $\mathbf{u}$,

$$V^{-\frac{1}{2}}\begin{pmatrix}\sqrt{(N-n)h_N}u \\ \sqrt{nh}[\widehat{\sigma^2}_{KS,t} - \sigma_t^2]\end{pmatrix} \longrightarrow N(0, I_2),$$

where $V = \mathrm{diag}\{V_1, V_2\}$ with $V_1 = 2\sigma_t^4p(x)S^*$, where $S^* = (v_{i+j-2})$, $i,j = 1,2$ with $v_j = \int u^jK^2(u)\mathrm{d}u$. This combined with equation (13) gives the joint asymptotic normality of b and $\widehat{\sigma^2}_{KS,t}$. Note that $B = o_p(1/\sqrt{(N-n)h_2})$, it follows that

$$\Sigma_2^{-\frac{1}{2}}\begin{pmatrix}\sqrt{(N-n)h_N}[\widehat{\sigma^2}_{LE,S} - \sigma_t^2 - \theta_n] \\ \sqrt{nh}[\widehat{\sigma^2}_{KS,t} - \sigma_t^2]\end{pmatrix} \longrightarrow N(0, I_2),$$

where $\Sigma_2^{-\frac{1}{2}} = \mathrm{diag}\{2\sigma_t^4p(x)^{-1}e_1^{\mathrm{T}}(H^{-1}S^{-1})^{\mathrm{T}}S^*H^{-1}S^{-1}e_1, V_2\}$ and $V_2 = 2\sigma_t^4\frac{\int K^2(u)\mathrm{d}u}{p(t)}$. Note that $\widehat{\sigma^2}_{t,\mathrm{time}}$ and $\widehat{\sigma^2}_{t,\mathrm{state}}$ are asymptotically independent, it follows that the asymptotical normality of $\widehat{\sigma^2}_{I,(s,t)}$ holds. □

References

[1] YAO Qiwei, TONG H. *Quantifying the influence of initial values on non-linear prediction* [J]. J. Roy. Statist. Soc. Ser. B, 1994, **56**(4): 701–725.

[2] HÄRDLE W, TSYBAKOV A. *Local polynomial estimators of the volatility function in nonparametric autoregression* [J]. J. Econometrics, 1997, **81**(1): 223–242.

[3] RICE J A. *Bandwidth choice for nonparametric kernel regression* [J]. Ann. Statist, Ser.A, 1984, **12**: 1215–1230.

[4] RICE J A. *Boundary modification for nonparametric regression* [J]. Commun. Statistics Theory and Mothods, Ser. B, 1984, **13**: 893–900.

[5] GASSER T, SROKA L, JENNEN-STEINMETZ C. *Residual variance and residual pattern in nonlinear regression* [J]. Biometrika, 1986, **73**(3): 625–633.

[6] MÜLLER H, STADTMÜLLER U. *Estimation of heteroscedasticity in regression analysis* [J]. Ann. Statist., 1987, **15**(2): 610–625.

[7] HALL P, KAY J W, TITTERINGTON D M. *Asymptotically optimal difference-based estimation of variance in nonparametric regression* [J]. Biometrika, 1990, **77**: 521–528.

[8] STONE C J. *Consistent nonparametric regression* [J]. Ann. Statist, 1974, **5**: 595–645.

[9] HALL P, CARROLL R J. *Variance function estimation in regression: the effect of estimation of the mean* [J]. J. R. Statist. Soc, B, 1989, **51**: 3–14.

[10] MÜLLER H, STADTMÜLLER U. *On variance function estimation with quadratic forms* [J]. J.Statist.Plan.Inf., 1993, **35**:213–231.

[11] NEUMANN M H. *Fully data-driven nonparametric variance estimators* [J]. Statistics, 1994, **25**(3): 189–212.

[12] FAN Jianqing, GIJBELS I. *Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation* [J]. J. Roy. Statist. Soc. Ser. B, 1995, **57**(2): 371–394.

[13] GALLANT A R, TAUCHEN G. *Estimation of continuous time models for stock returns and interest rates* [J]. Macroecon. Dyn, 1997, **1**: 135–168.

[14] STANTON R. *A Nonparametric model of term structure dynamics and the Market price of interest rate risk* [J]. J. Finance, 1997, **52**: 1973–2002.

[15] 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.

[16] FAN Jianqing, YAO Qiwei. *Efficient estimation of conditional variance functions in stochastic regression* [J]. Biometrika, 1998, **85**(3): 645–660.

[17] ARFI M. *Non-parametric variance estimation from ergodic samples* [J]. Scand. J. Statist., 1998, **25**(1): 225–234.

[18] FAN Jianqing, JIANG Jiancheng, ZHANG Chunming. *Time-dependent diffusion models for term structure dynamics. Statistical applications in financial econometrics* [J]. Statist. Sinica, 2003, **13**(4): 965–992.

[19] FAN Jianqing, ZHANG Chunming. *A reexamination of diffusion estimators with applications to financial model validation* [J]. J. Amer. Statist. Assoc., 2003, **98**(461): 118–134.

[20] FAN Jianqing, FAN Yingying, JIANG Jiancheng. *Dynamic integration of time- and state-domain methods for volatility estimation* [J]. J. Amer. Statist. Assoc., 2007, **102**(478): 618–631.

[21] FAN Jianqing. *A selective overview of nonparametric methods in financial econometrics* [J]. Statist. Sci., 2005, **20**(4): 317–357.

[22] FAN Jianqing, GIJBELS I. *Local Polynomial Modelling and Its Applications* [M]. Chapman & Hall, London, 1996.

[23] ALLEN D M. *The relationship between variable and data augmentation and a method of prediction* [J]. Technometrics, 1974, **16**: 125–127.

[24] WAHBA G. *A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them* [M]. North-Holland, Amsterdam, 1977.

[25] BANON G. *Nonparametric identification for diffusion processes* [J]. SIAM J. Control Optim., 1978, **16**(3): 380–395.

[26] ROSENBLATT M. *Density Estimates and Markov Sequences* [M]. Cambridge Univ. Press, London, 1970.

[27] FAN Jianqing, YAO Qiwei. *Nonlinear Time Series. Nonparametric and Parametric Methods* [M]. Springer-Verlag, New York, 2003.

[28] REVUZ D, YOR M. *Continuous Martingales and Brownian Motion* [M]. Springer-Verlag, Berlin, 1999.