Tang and Wang *Advances in Difference Equations* (2018) 2018:51
https://doi.org/10.1186/s13662-018-1506-1

Advances in Difference Equations
a SpringerOpen Journal

RESEARCH Open Access

Variable bandwidth local maximum likelihood type estimation for diffusion processes

Ming T. Tang and Yun Y. Wang*

*Correspondence:
yywang@zju.edu.cn
School of Science, Jiangxi University of Science and Technology, Ganzhou, People's Republic of China

Abstract

The method of robust approach is applied to estimate drift function and diffusion function of diffusion processes with discrete-time observations. The proposed method combines the ideas of local linear regression technique and maximum likelihood type estimation technique, so the advantages of local linear estimators persist and overcome the disadvantages of least-squares estimator. Moreover, a variable bandwidth instead of a constant bandwidth is considered in the local maximum likelihood type estimators. The consistency and asymptotic normality of the local maximum likelihood type estimators for drift and diffusion functions are developed under some given conditions. We perform a simulation study to evaluate the robust performances of the proposed estimators.

MSC: 62G35; 62G20

Keywords: Asymptotic normality; Consistency; Robust estimation; Stochastic differential equation; Variable bandwidth

1 Introduction

Diffusion processes X defined by the following stochastic differential equation are considered in this article:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t, \tag{1}$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion, $\mu(\cdot)$ is an unknown measurable function (drift function) and $\sigma(\cdot)$ is an unknown positive function (diffusion function). It is well known that diffusion processes driven by Brown motion have been widely used in the financial and economic fields, and it is often used to model and analyze dynamic changes in asset prices, interest rates, and exchange rates. For example, [1] investigated an application of Ornstein-Uhlenbeck process to commodity pricing in Thailand. Reference [2] improved estimation of drift parameters of diffusion processes for interest rates by incorporating information in bond prices. So, recently in the literature, the statistical inference for diffusion processes based on discrete observations has often been of concern; for example, see [3–6] and its references for parametric estimation, see [7–9] and the references therein for a semi-parametric estimation and see [10–16] and the references therein for a nonparametric estimation.

As is well known, the first to consider nonparametric estimation for the diffusion coefficient in model (1) with discrete-time observation was [17], where a kernel type estimator was considered. Thereafter, [18] proposed a nonparametric identification and an estimation procedure for the diffusion function after [17], and derived a consistent nonparametric estimator for the drift function by combining their estimator of the diffusion function. Reference [19] constructed the first-, second-, and third-order approximation formulas for drift and diffusion functions by using an infinitesimal generator and Taylor expansion. Reference [12] generalized Stanton's idea and introduced the local polynomial estimators for drift and diffusion functions. Since a local linear method may produce negative values for the diffusion function, [14] proposed a new nonparametric estimation procedure of the diffusion function based on re-weighting the Nadaraya–Watson estimator.

However, local linear regression methods are very sensitive to outliers, and individual outliers can lead to large changes in the results of statistical inference, therefore leading to irrational and even erroneous conclusions. Such statistical methods like the local linear regression approach are not strong enough to adapt the complex changing reality, in other words, the local linear regression method is not robust when it comes to outliers or heavy-tailed distributions. In recent years, various robust methods have been proposed for abnormal observation, which has become increasingly crucial and frequent in many research fields. Reference [20] defined maximum likelihood type robust estimates of regression and investigated the asymptotic properties. From then on, the maximum likelihood type robust estimation (M-estimation) has been discussed by many authors, for example, [21, 22] and the references therein. Meanwhile, some modified maximum likelihood type estimators were developed, such as the local maximum likelihood type estimator (local M-estimators), which is a combination of the local linear regression and the M-estimation regression, so the nice properties of local linear estimator and M-estimator persist. For instance, [23] constructed variable bandwidth local linear M-estimator for a regression function. Reference [24] proposed a nonparametric estimator of the regression function by combining local polynomial regression and M-estimation regression. Reference [25] developed robust version of local linear regression smoothers for stationary time series sequence. Reference [26] considered local M-estimation of the unknown drift and diffusion functions of integrated diffusion processes.

The purpose of this paper is to investigate the local linear and variable bandwidth M-estimators of the drift and diffusion functions in model (1) based on high-frequency data, that is, the sample observations are only selected at discrete-time points, say at n equally spaced $\{i\Delta, i = 0, 1, \dots, n\}$, where Δ is the sampling interval, and $\Delta \to 0$ as $n \to \infty$. Furthermore, the consistency and asymptotic normality of the local M-estimators will be proved under general assumptions. Since the new the local linear and variable bandwidth M-estimators have a good control of outliers, which are common in financial, economic, physical, engineering and other fields, the proposed estimators in this paper greatly promote the application of diffusion model (1) in these fields, and they provide the theory and application foundation for dynamic modeling in these fields.

The article is organized as follows. The second section constructs the variable bandwidth local M-estimators of drift and diffusion functions, and the consistency and asymptotic normality of the new estimators are developed in the same section. Section 3 presents the results of a simulation study. Proofs and auxiliary results are given in Section 4.

2 Local M-estimators and asymptotic theory

Local M-estimation of drift function $\mu(x)$ and diffusion function $\sigma^2(x)$ depend on the equations

$$E\left(\frac{X_{(i+1)\Delta}-X_{i\Delta}}{\Delta}\middle|X_{i\Delta}=x\right)=\mu(x)+o(1), \tag{2}$$

$$E\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta}\middle|X_{i\Delta}=x\right)=\sigma^2(x)+o(1), \tag{3}$$

as $\Delta\to 0$. The reader can refer to [19] or [27] for more details as regards (2) and (3).

Neglecting the smaller-order terms, the local linear estimator with the variable bandwidth for $\mu(x)$ is defined as the solution to the problem: Choose a_1 and b_1 to minimize the weighted sum as

$$\sum_{i=1}^{n}\left(\frac{X_{(i+1)\Delta}-X_{i\Delta}}{\Delta}-a_1-b_1(X_{i\Delta}-x)\right)^2\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h}\beta_1(X_{i\Delta})\right),$$

and the local linear estimator with the variable bandwidth for $\sigma^2(x)$ is defined as the solution to the problem: Choose a_2 and b_2 to minimize the weighted sum as follows:

$$\sum_{i=1}^{n}\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta}-a_2-b_2(X_{i\Delta}-x)\right)^2\beta_2(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h}\beta_2(X_{i\Delta})\right),$$

where $K(\cdot)$ is kernel function and $h=h_n$ is the bandwidth. $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are nonnegative functions reflecting the variable amount of smoothing at each data point. $h/\beta_1(X_{i\Delta})$ and $h/\beta_2(X_{i\Delta})$ are called variable bandwidth. For more detailed information on variable bandwidths, see [28–30], among others.

In fact, the aforementioned method used to establish estimators are based on least-squares approach and are not robust. As a result, we choose a_1 and b_1 to minimize

$$\sum_{i=1}^{n}\rho_1\left(\frac{X_{(i+1)\Delta}-X_{i\Delta}}{\Delta}-a_1-b_1(X_{i\Delta}-x)\right)\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h}\beta_1(X_{i\Delta})\right)$$

and a_2 and b_2 to minimize

$$\sum_{i=1}^{n}\rho_2\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta}-a_2-b_2(X_{i\Delta}-x)\right)\beta_2(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h}\beta_2(X_{i\Delta})\right),$$

or to satisfy the following estimation equations:

$$\sum_{i=1}^{n}\psi_1\left(\frac{X_{(i+1)\Delta}-X_{i\Delta}}{\Delta}-a_1-b_1(X_{i\Delta}-x)\right)\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix}1\\ \frac{X_{i\Delta}-x}{h}\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix} \tag{4}$$

and

$$\sum_{i=1}^{n} \psi_2\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta} - a_2 - b_2(X_{i\Delta}-x)\right)\beta_2(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h/\beta_2(X_{i\Delta})}\right)\begin{pmatrix}1\\ \frac{X_{i\Delta}-x}{h}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}, \tag{5}$$

where $\rho_1(\cdot)$ and $\rho_2(\cdot)$ are given outlier-resistant functions and $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are the derivatives of $\rho_1(\cdot)$ and $\rho_2(\cdot)$, respectively.

The maximum likelihood type estimators of $\mu(x)$ and $\mu'(x)$ are denoted $\hat{\mu}(x) = \hat{a}_1$ and $\hat{\mu}'(x) = \hat{b}_1$, which are the solutions of (4), the maximum likelihood type estimators of $\sigma^2(x)$ and $(\sigma^2(x))'$ are denoted $\hat{\sigma}^2(x) = \hat{a}_2$ and $(\hat{\sigma}^2(x))' = \hat{b}_2$, which are the solutions of (5).

For a given point x_0, our lemmas and asymptotic theory results are based on the following conditions.

C 1 ([10])

(i) The initial condition $X_0 \in L^2$ and is considered to be independent of $\{B_t, t \geq 0\}$.

(ii) The unknown functions $\mu(\cdot)$ and $\sigma(\cdot)$ are time-homogeneous and measurable functions on $\mathcal{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$. We also assume the two functions are at least twice continuously differentiable, and satisfy local Lipschitz and growth conditions, that is, for any compact subset $J \subseteq \mathcal{D}$, for all $x, y \in J$, there exist constants L_1 and L_2 such that

$$\left|\mu(x)-\mu(y)\right| + \left|\sigma(x)-\sigma(y)\right| \leq L_1|x-y|,$$

and

$$\left|\mu(x)\right| + \left|\sigma(x)\right| \leq L_2\left[1+|x|\right].$$

(iii) $\sigma^2(\cdot) > 0$ on $\mathcal{D}$;

(iv) Let $S(z) = \int_{z_0}^{z} \exp(\int_{z_0}^{y} \frac{-2\mu(x)}{\sigma^2(x)}\,dx)\,dy$, $z_0 \in \mathcal{D}$, suppose $S(z)$ satisfies

$$\lim_{z\to l} S(z) = -\infty,$$

$$\lim_{z\to u} S(z) = \infty.$$

Remark 1 The condition C1 ensures the existence and uniqueness of a strong solution to model (1), see [31] for details.

C 2

(i) $\int_l^u s(x)\,dx < \infty$, where $s(x) = 2/S'(x)\sigma^2(x)$.

(ii) The initial point X_0 has a stationary distribution P^0, where P^0 is the invariant distribution of the process $(X_t)_{t\in[0,\infty)}$.

Remark 2 The conditions C1 and C2 ensure that X is stationary, and from Kolmogorov forward equation we can get the stationary density $p(x)$ of X:

$$p(x) = \frac{s(x)}{\int_l^u s(x)\,dx} = \frac{\xi}{\sigma^2(x)} \exp\left\{\int_{z_0}^x \frac{2\mu(x)}{\sigma^2(x)}\,dx\right\},$$

where z_0 is an arbitrary point inside $\mathcal{D}$ and ξ is a normalizing constant.

C 3 Let $\mathcal{D} = (l, u)$ be the state space of X, suppose that

$$\limsup_{x\to u}\left(\frac{\mu(x)}{\sigma(x)} - \frac{\sigma'(x)}{2}\right) < 0,$$
$$\limsup_{x\to l}\left(\frac{\mu(x)}{\sigma(x)} - \frac{\sigma'(x)}{2}\right) > 0.$$

Moreover, the mixing coefficient $\alpha(k)$ satisfies $\sum_{k\geq 1} k^a(\alpha(k))^{\gamma/(2+\gamma)} < \infty$ for some $a > \gamma/(2+\gamma)$, where γ is given in the condition C8.

Remark 3 The condition C3 guarantees that the process X is α-mixing; see [32] for details.

C 4

(i) The kernel function $K(\cdot)$ is a continuous probability density function compactly supported on $[-1, 1]$.
(ii) The bandwidth h satisfies $h \to 0$, $nh \to \infty$ and $nh\Delta \to \infty$ as $n \to \infty$.

C 5 The density function $p(x)$ of the process X is continuous at the point x_0, and $p(x_0) > 0$. Moreover, the joint density function of $X_{i\Delta}$ and $X_{j\Delta}$ is bounded for all i, j.

C 6

(i) $\min_x \beta_1(x) > 0$, and $\beta_1(\cdot)$ is continuous at the point x_0;
(ii) $\min_x \beta_2(x) > 0$, and $\beta_2(\cdot)$ is continuous at the point x_0.

C 7

(i) $E[\psi_1(u_{i\Delta})|X_{i\Delta} = x] = o(1)$ with $u_{i\Delta} = \frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} - \mu(X_{i\Delta})$;
(ii) $E[\psi_2(v_{i\Delta})|X_{i\Delta} = x] = o(1)$ with $v_{i\Delta} = \frac{(X_{(i+1)\Delta} - X_{i\Delta})^2}{\Delta} - \sigma^2(X_{i\Delta})$.

C 8

(i) The function $\psi_1(\cdot)$ is continuous and has a derivative $\psi_1'(\cdot)$ almost everywhere. Additionally, we assume that the following three functions: $E[\psi_1'(u_{i\Delta})|X_{i\Delta} = x]$, $E[\psi_1^2(u_{i\Delta})|X_{i\Delta} = x]$, $E[{\psi_1'}^2(u_{i\Delta})|X_{i\Delta} = x]$ are all positive for any x and continuous at the point x_0, and there exists a constant $\gamma > 0$ such that $E[|\psi_1(u_{i\Delta})|^{2+\gamma}|X_{i\Delta} = x]$, $E[|\psi_1'(u_{i\Delta})|^{2+\gamma}|X_{i\Delta} = x]$ are bounded in a neighborhood of x_0.
(ii) The function $\psi_2(\cdot)$ is continuous and has a derivative $\psi_2'(\cdot)$ almost everywhere. Additionally, we assume that the following three functions: $E[\psi_2'(v_{i\Delta})|X_{i\Delta} = x]$, $E[\psi_2^2(v_{i\Delta})|X_{i\Delta} = x]$, $E[{\psi_2'}^2(v_{i\Delta})|X_{i\Delta} = x]$ are all positive for any x and continuous at the point x_0, and there exists a constant $\gamma > 0$ such that $E[|\psi_2(v_{i\Delta})|^{2+\gamma}|X_{i\Delta} = x]$, $E[|\psi_2'(v_{i\Delta})|^{2+\gamma}|X_{i\Delta} = x]$ are bounded in a neighborhood of x_0.

C 9

(i) The function $\psi_1'(\cdot)$ satisfies

$$E\Big[\sup_{|z|\le\delta}\big|\psi_1'(u_{i\Delta}+z)-\psi_1'(u_{i\Delta})\big||X_{i\Delta}=x\Big]=o(1)$$

and

$$E\Big[\sup_{|z|\le\delta}\big|\psi_1(u_{i\Delta}+z)-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})z\big||X_{i\Delta}=x\Big]=o(\delta),$$

as $\delta\to 0$ uniformly in x in a neighborhood of x_0;

(ii) The function $\psi_2'(\cdot)$ satisfies

$$E\Big[\sup_{|z|\le\delta}\big|\psi_2'(v_{i\Delta}+z)-\psi_2'(v_{i\Delta})\big||X_{i\Delta}=x\Big]=o(1)$$

and

$$E\Big[\sup_{|z|\le\delta}\big|\psi_2(v_{i\Delta}+z)-\psi_2(v_{i\Delta})-\psi_2'(v_{i\Delta})z\big||X_{i\Delta}=x\Big]=o(\delta),$$

as $\delta\to 0$ uniformly in x in a neighborhood of x_0.

C 10

(i) For any i, j, suppose that

$$E\big[\psi_1^2(u_{i\Delta})+\psi_1^2(u_{j\Delta})|X_{i\Delta}=x,X_{j\Delta}=y\big],$$
$$E\big[{\psi_1'}^2(u_{i\Delta})+{\psi_1'}^2(u_{j\Delta})|X_{i\Delta}=x,X_{j\Delta}=y\big]$$

are bounded in the neighborhood of x_0;

(ii) For any i, j, suppose that

$$E\big[\psi_2^2(v_{i\Delta})+\psi_2^2(v_{j\Delta})|X_{i\Delta}=x,X_{j\Delta}=y\big],$$
$$E\big[{\psi_2'}^2(v_{i\Delta})+{\psi_2'}^2(v_{j\Delta})|X_{i\Delta}=x,X_{j\Delta}=y\big]$$

are bounded in the neighborhood of x_0.

Remark 4 In fact, the conditions C7–C10 imposed on $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are mild and satisfied for many applications, such as Huber's $\psi(\cdot)$ function. For more detailed information on these conditions please refer to [23] or [25].

C 11 Suppose that there exist a sequence of positive integers q_n such that $q_n\to\infty$, $q_n=o((nh)^{1/2})$ and $(n/h)^{1/2}\alpha(q_n)\to 0$ as $n\to\infty$.

C 12 For γ in the condition C8 and all x in a neighborhood of x_0, there exists $\tau>2+\gamma$ such that the two functions $E\{|\psi_1(u_{i\Delta})|^\tau|X_{i\Delta}=x\}$, $E\{|\psi_2(v_{i\Delta})|^\tau|X_{i\Delta}=x\}$ are bounded. Furthermore, we assume $\alpha(n)=O(n^{-\theta})$, where $\theta\ge(2+\gamma)\tau/\{2(\tau-2-\gamma)\}$.

C 13 Assume $n^{-\gamma/4}h^{(2+\gamma)/\tau-1-\gamma/4} = O(1)$, where γ is given in the condition C8 and τ is given in the condition C12.

Remark 5 The assumptions in condition C3 and C11 on mixing coefficient $\alpha(k)$ is sufficient conditions for mixing coefficient, [25] pointed out this assumptions on mixing coefficient are satisfied given some general conditions. Condition C13 is also satisfied under some weak constraints on γ or τ.

Throughout the whole paper, let

$$K_l = \int K(u)u^l\,du, \qquad J_l = \int u^l K^2(u)\,du, \quad \text{for } l \geq 0.$$

$$U_1 = \begin{pmatrix} K_0 & \frac{K_1}{\beta_1(x_0)} \\ \frac{K_1}{\beta_1(x_0)} & \frac{K_2}{\beta_1^2(x_0)} \end{pmatrix}, \qquad V_1 = \begin{pmatrix} J_0 & \frac{J_1}{\beta_1(x_0)} \\ \frac{J_1}{\beta_1(x_0)} & \frac{J_2}{\beta_1^2(x_0)} \end{pmatrix}, \qquad A_1 = \begin{pmatrix} K_2 \\ \frac{K_3}{\beta_1(x_0)} \end{pmatrix},$$

$$U_2 = \begin{pmatrix} K_0 & \frac{K_1}{\beta_2(x_0)} \\ \frac{K_1}{\beta_2(x_0)} & \frac{K_2}{\beta_2^2(x_0)} \end{pmatrix}, \qquad V_2 = \begin{pmatrix} J_0 & \frac{J_1}{\beta_2(x_0)} \\ \frac{J_1}{\beta_2(x_0)} & \frac{J_2}{\beta_2^2(x_0)} \end{pmatrix}, \qquad A_2 = \begin{pmatrix} K_2 \\ \frac{K_3}{\beta_2(x_0)} \end{pmatrix},$$

$$G_1(x) = E\bigl[\psi_1'(u_{i\Delta})|X_{i\Delta} = x\bigr], \qquad G_2(x) = E\bigl[\psi_1^2(u_{i\Delta})|X_{i\Delta} = x\bigr],$$

$$G_3(x) = E\bigl[\psi_1'^2(u_{i\Delta})|X_{i\Delta} = x\bigr], \qquad H_1(x) = E\bigl[\psi_2'(v_{i\Delta})|X_{i\Delta} = x\bigr],$$

$$H_2(x) = E\bigl[\psi_2^2(v_{i\Delta})|X_{i\Delta} = x\bigr], \qquad H_3(x) = E\bigl[\psi_2'^2(v_{i\Delta})|X_{i\Delta} = x\bigr].$$

We now develop the asymptotic theory for the proposed local M-estimators:

Theorem 1 *Under the conditions* C1–C5 *and the conditions* (i) *of* C6–C10, *there exist solutions* $\hat{\mu}(x_0)$ *and* $\hat{\mu}'(x_0)$ *to equations* (4) *such that*

(i)

$$\begin{pmatrix} \hat{\mu}(x_0) - \mu(x_0) \\ h(\hat{\mu}'(x_0) - \mu'(x_0)) \end{pmatrix} \xrightarrow{P} 0, \quad n \to \infty.$$

(ii) *Furthermore, if the conditions* C11–C13 *hold, then*

$$\sqrt{nh}\left[\begin{pmatrix} \hat{\mu}(x_0) - \mu(x_0) \\ h(\hat{\mu}'(x_0) - \mu'(x_0)) \end{pmatrix} - \frac{h^2\mu''(x_0)}{2\beta_1^2(x_0)}U_1^{-1}A_1\right] \xrightarrow{D} N(0, \Sigma_1),$$

where

$$\Sigma_1 = \frac{G_2(x_0)\beta_1(x_0)}{G_1^2(x_0)p(x_0)}U_1^{-1}V_1U_1^{-1}.$$

Theorem 2 *Under the conditions* C1–C5 *and the conditions* (ii) *of* C6–C10, *there exist solutions* $\hat{\sigma}^2(x_0)$ *and* $(\hat{\sigma}^2(x_0))'$ *to equations* (5) *such that*

(i)

$$\begin{pmatrix} \hat{\sigma}^2(x_0) - \sigma^2(x_0) \\ h((\hat{\sigma}^2(x_0))' - (\sigma^2(x_0))') \end{pmatrix} \xrightarrow{P} 0, \quad n \to \infty.$$

(ii) *Furthermore, if the conditions* C11–C13 *hold, then*

$$\sqrt{nh}\left[\begin{pmatrix}\hat{\sigma}^2(x_0)-\sigma^2(x_0)\\ h[(\hat{\sigma}^2(x_0))'-(\sigma^2(x_0))']\end{pmatrix}-\frac{h^2(\sigma^2(x_0))''}{2\beta_2^2(x_0)}U_2^{-1}A_2\right]\xrightarrow{D}N(0,\Sigma_2),$$

where

$$\Sigma_2=\frac{H_2(x_0)\beta_2(x_0)}{H_1^2(x_0)p(x_0)}U_2^{-1}V_2U_2^{-1}.$$

3 Simulation study

We now perform a Monte Carlo simulation study to evaluate the finite sample performance of the variable bandwidth local M-estimators by comparing the mean square error (MSE) between them and the Nadaraya–Watson estimators.

We consider the following diffusion process X:

$$dX_t=\left(-X_t+0.5\sqrt{1+X_t^2}\right)dt+0.1\,dB_t$$

for $t\in[0,T]=[0,100]$. Throughout the simulation, we take Huber's function $\psi_1(z)=\max\{-c,\min(c,z)\}$ with $c=0.135$. The uniform kernel $K(u)=\frac{1}{2}I(|u|\leq 1)$ is selected as the kernel function, and the bandwidth h is chosen by minimizing the MSE as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mu}(x_i)-\mu(x_i)\right)^2,$$

where $\{x_i, i=1,2,\dots,n\}$ are chosen uniformly to cover the range of sample path of X_t.

Throughout the study, we use iterative method to obtain $\hat{\mu}(\cdot)$, for any initial value $\hat{\mu}_0(x)$, we have

$$\begin{pmatrix}\hat{\mu}_t(x)\\ \hat{\mu}_t'(x)\end{pmatrix}=\begin{pmatrix}\hat{\mu}_{t-1}(x)\\ \hat{\mu}_{t-1}'(x)\end{pmatrix}-\left[W_n\left(\hat{\mu}_{t-1}(x),\hat{\mu}_{t-1}'(x)\right)\right]^{-1}\Psi_n\left(\hat{\mu}_{t-1}(x),\hat{\mu}_{t-1}'(x)\right),$$

where $\hat{\mu}_{t-1}(x)$ and $\hat{\mu}_{t-1}'(x)$ are the tth iteration value of $\hat{\mu}'(x)$ and $\hat{\mu}(x)$, and

$$\begin{aligned}W_n(a_1,b_1)&=\left(\frac{\partial\Psi_n(a_1,b_1)}{\partial a_1},\frac{\partial\Psi_n(a_1,b_1)}{\partial b_1}\right),\\ \Psi_n(a_1,b_1)&=\sum_{i=1}^{n}\psi_1\left(\frac{X_{(i+1)\Delta}-X_{i\Delta}}{\Delta}-a_1-b_1(X_{i\Delta}-x)\right)\beta_1(X_{i\Delta})\\ &\quad\times K\left(\frac{X_{i\Delta}-x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix}1\\ \frac{X_{i\Delta}-x}{h}\end{pmatrix}.\end{aligned}$$

When the following conditions are satisfied, the procedure terminates:

$$\left\|\begin{pmatrix}\hat{\mu}_t(x)\\ \hat{\mu}_t'(x)\end{pmatrix}-\begin{pmatrix}\hat{\mu}_{t-1}(x)\\ \hat{\mu}_{t-1}'(x)\end{pmatrix}\right\|\leq 1\times 10^{-4}.$$

Figure 1 shows the five sample paths of the process X_t. Table 1 lists the MSEs of the Nadaraya–Watson estimator and the variable bandwidth local M-estimator for the drift function $\mu(\cdot)$ when $n=100$, $n=500$, $n=1000$, $n=5000$, $n=10{,}000$. The figures in Table 1

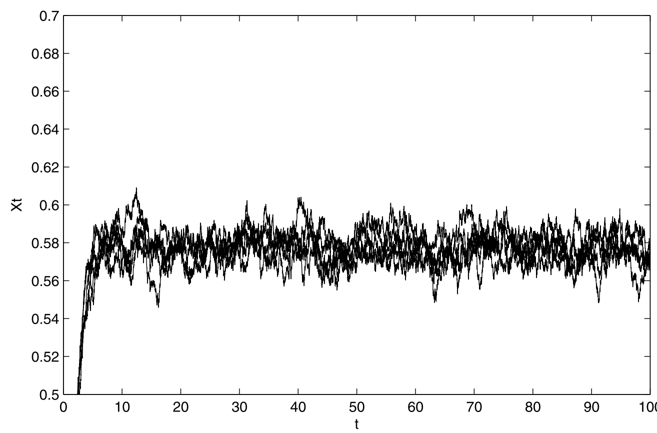

Figure 1 The five sample paths of the process X_t

Table 1 The MSEs of Nadaraya–Watson estimator (MSE_1) and variable bandwidth local M-estimator (MSE_2) for drift function $\mu(\cdot)$

Sample size n	MSE_1	MSE_2
$n = 100$	0.0628	0.0667
$n = 500$	0.0042	0.0036
$n = 1000$	0.0023	0.0019
$n = 5000$	0.0014	0.0011
$n = 10{,}000$	0.0012	0.0007

indicate that

(i) The MSEs of the two types of estimators decrease toward zero as the sample size n increases.

(ii) The variable bandwidth local M-estimator performs better than the Nadaraya–Watson estimator.

4 Lemmas and proofs

The following lemmas are needed to prove the main results of this paper.

Lemma 1 *Under the conditions* C1–C5 *and the conditions* (i) *of the* C6–C10, *we have*

$$\begin{aligned} &\sum_{i=1}^{n} \psi_1'(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l \\ &\quad = nh^{l+1}\frac{G_1(x_0)}{\beta_1^l(x_0)}p(x_0)K_l\bigl(1+o_p(1)\bigr) \end{aligned}$$

and

$$\begin{aligned} &\sum_{i=1}^{n} \psi_1'(u_{i\Delta})R_1(X_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l \\ &\quad = nh^{l+3}\frac{G_1(x_0)}{2\beta_1^{l+2}(x_0)}\mu''(x_0)p(x_0)K_{l+2}\bigl(1+o_p(1)\bigr), \end{aligned}$$

where $R_1(X_{i\Delta}) = \mu(X_{i\Delta}) - \mu(x_0) - \mu'(x_0)(X_{i\Delta}-x_0)$.

Proof of Lemma 1 Since the second part of Lemma 1 can be proved by the same arguments as the first one, we only prove the first part. Let

$$Z_{n,i} = \psi_1'(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l.$$

By a change of variable and the continuity at the point x_0 of $\beta_1(\cdot)$, $K(\cdot)$, $G_1(\cdot)$ and $p(\cdot)$, we obtain

$$\begin{aligned}
E(Z_{n,1}) &= \int G_1(x)\beta_1(x)K\left(\frac{x-x_0}{h/\beta_1(x)}\right)(x-x_0)^l p(x)\,dx \\
&= \int G_1(x_0+yh)\beta_1(x_0+yh)K\bigl(y\beta_1(x_0+yh)\bigr)(yh)^l p(x_0+yh)h\,dy \\
&= h^{l+1}G_1(x_0)\beta_1(x_0)p(x_0)\int K\bigl(y\beta_1(x_0)\bigr)y^l\,dy\bigl(1+o(1)\bigr) \\
&= h^{l+1}G_1(x_0)\frac{p(x_0)}{\beta_1^l(x_0)}\int K(u)u^l\,du\bigl(1+o(1)\bigr) \\
&= h^{l+1}G_1(x_0)\frac{p(x_0)}{\beta_1^l(x_0)}K^l\bigl(1+o(1)\bigr).
\end{aligned}$$

Therefore, we have

$$E\left(\sum_{i=1}^n \psi_1'(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l\right) = nh^{l+1}G_1(x_0)\frac{p(x_0)}{\beta_1^l(x_0)}K^l\bigl(1+o(1)\bigr).$$

Note that

$$\sum_{i=1}^n Z_{n,i} = E\left(\sum_{i=1}^n Z_{n,i}\right) + O_p\left(\sqrt{\operatorname{Var}\left(\sum_{i=1}^n Z_{n,i}\right)}\right)$$

and

$$\operatorname{Var}\left(\sum_{i=1}^n Z_{n,i}\right) = nEZ_{n,1}^2 + 2\sum_{j=2}^n (n-j+1)\operatorname{Cov}(Z_{n,1},Z_{n,j}).$$

By a change of variable and the continuity at the point x_0 of $\beta_1(\cdot)$, $K(\cdot)$, $G_3(\cdot)$ and $p(\cdot)$, we obtain

$$\begin{aligned}
EZ_{n,1}^2 &= \int G_3(x)\beta_1^2(x)K^2\left(\frac{x-x_0}{h/\beta_1(x)}\right)(x-x_0)^{2l}p(x)\,dx \\
&= \int G_3(x_0+yh)\beta_1^2(x_0+yh)K^2\bigl(y\beta_1(x_0+yh)\bigr)(yh)^{2l}p(x_0+yh)h\,dy \\
&= h^{2l+1}G_3(x_0)\beta_1^2(x_0)p(x_0)\int K^2\bigl(y\beta_1(x_0)\bigr)y^{2l}\,dy\bigl(1+o(1)\bigr) \\
&= h^{2l+1}G_3(x_0)\beta_1^{1-2l}(x_0)p(x_0)\int K^2(u)u^{2l}\,du\bigl(1+o(1)\bigr) \\
&= O\bigl(h^{2l+1}\bigr).
\end{aligned}$$

Let d_n be a sequence of positive integers satisfying $d_n \to \infty$ and $hd_n \to 0$. Then we have

$$\sum_{j=2}^{n} \left|\operatorname{Cov}(Z_{n,1}, Z_{n,j})\right| = \sum_{j=2}^{d_n} \left|\operatorname{Cov}(Z_{n,1}, Z_{n,j})\right| + \sum_{j=d_n+1}^{n} \left|\operatorname{Cov}(Z_{n,1}, Z_{n,j})\right|.$$

By the conditions C6(i), C10(i) and the bounded support of $K(\cdot)$, we have

$$\begin{aligned}
&|EZ_{n,i}Z_{n,j}| \\
&\quad \le E|Z_{n,i}Z_{n,j}| \\
&\quad = E\left|E\left[\psi_1'(u_{i\Delta})\psi_1'(u_{j\Delta})|X_{i\Delta}, X_{j\Delta}\right]\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l \right.\\
&\qquad \left. \times \beta_1(X_{j\Delta})K\left(\frac{X_{j\Delta}-x_0}{h/\beta_1(X_{j\Delta})}\right)(X_{j\Delta}-x_0)^l\right| \\
&\quad \le C_1 E\left|\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l \beta_1(X_{j\Delta})K\left(\frac{X_{j\Delta}-x_0}{h/\beta_1(X_{j\Delta})}\right)(X_{j\Delta}-x_0)^l\right| \\
&\quad \le C_2 h^{2l+2},
\end{aligned}$$

where C_1 and C_2 are constants. Therefore, we have

$$\sum_{j=2}^{d_n} \left|\operatorname{Cov}(Z_{n,1}, Z_{n,j})\right| \le C_2 h^{2l+2} \sum_{j=2}^{d_n} 1 = o\left(nh^{2l+1}\right).$$

By using the Davydov inequality, we have

$$\left|\operatorname{Cov}(Z_{n,1}, Z_{n,j})\right| \le C_3\left[\alpha(j-1)\right]^{\gamma/(2+\gamma)}\left(E|Z_{n,1}|^{2+\gamma}\right)^{2/(2+\gamma)},$$

and by the condition C8(i), we have

$$\begin{aligned}
E|Z_{n,i}|^{2+\gamma} &= E\left|E\left[\psi_1'(u_{i\Delta})|X_{i\Delta}\right]\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l\right|^{2+\gamma} \\
&\le C_4 E\left|\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l\right|^{2+\gamma} \\
&\le C_5 h^{(2+\gamma)l+1},
\end{aligned}$$

where C_3, C_4 and C_5 are constants. Therefore, by using the condition C3 and choosing d_n such that $d_n^a h^{\gamma/(2+\gamma)} = O(1)$, we have

$$\begin{aligned}
\sum_{j=d_n+1}^{n} \left|\operatorname{Cov}(Z_{n,1}, Z_{n,j})\right| &\le C_6 \sum_{j=d_n+1}^{n} \left[\alpha(j-1)\right]^{\gamma/2+\gamma}\left(h^{(2+\gamma)l+1}\right)^{2/(2+\gamma)} \\
&= C_6 h^{2l+2/(2+\gamma)} \sum_{k=d_n}^{n} \left[\alpha(k)\right]^{\gamma/2+\gamma} \\
&\le C_6 d_n^{-a} h^{2l+2/(2+\gamma)} \sum_{k=d_n}^{n} k^a\left[\alpha(k)\right]^{\gamma/2+\gamma} = o\left(nh^{2l+1}\right),
\end{aligned}$$

where C_6 is a constant. In summary, we have

$$\operatorname{Var}\left(\sum_{i=1}^{n} Z_{n,i}\right) = O\bigl(nh^{2l+1}\bigr).$$

Therefore,

$$\sum_{i=1}^{n} \psi_1'(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l = nh^{l+1}\frac{G_1(x_0)}{\beta_1^l(x_0)}p(x_0)K_l\bigl(1+o_p(1)\bigr).$$

This completes the lemma. □

Lemma 2 *Under the conditions* C1–C5 *and the conditions* (ii) *of* C6–C10, *we have*

$$\begin{aligned}&\sum_{i=1}^{n} \psi_2'(v_{i\Delta})\beta_2(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_2(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l \\ &\quad = nh^{l+1}\frac{H_1(x_0)}{\beta_2^l(x_0)}p(x_0)K_l\bigl(1+o_p(1)\bigr)\end{aligned}$$

and

$$\begin{aligned}&\sum_{i=1}^{n} \psi_2'(v_{i\Delta})R_2(X_{i\Delta})\beta_2(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_2(X_{i\Delta})}\right)(X_{i\Delta}-x_0)^l \\ &\quad = nh^{l+3}\frac{H_1(x_0)}{2\beta_2^{l+2}(x_0)}p(x_0)\bigl(\sigma^2(x_0)\bigr)''K_{l+2}\bigl(1+o_p(1)\bigr),\end{aligned}$$

where $R_2(X_{i\Delta}) = \sigma^2(X_{i\Delta}) - \sigma^2(x_0) - (\sigma^2(x_0))'(X_{i\Delta}-x_0)$.

Proof of Lemma 2 Using the same techniques in proving Lemma 1, we omit the proof process. □

Lemma 3 *Under the conditions* C1–C5, C6–C8(i) *and* C10(i)–13, *we have*

$$\frac{1}{\sqrt{nh}}\begin{pmatrix}\sum_{i=1}^{n}\psi_1(u_{i\Delta})\beta_1(X_{i\Delta})K(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}) \\ \sum_{i=1}^{n}\psi_1(u_{i\Delta})\beta_1(X_{i\Delta})K(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})})\frac{X_{i\Delta}-x_0}{h}\end{pmatrix} \xrightarrow{D} N(0,\Sigma_3),$$

where $\Sigma_3 = G_2(x_0)p(x_0)\beta_1(x_0)V_1$.

Proof of Lemma 3 Let

$$W_n = \sum_{i=1}^{n} W_{n,i} = \sum_{i=1}^{n} \psi_1(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix}1 \\ \frac{X_{i\Delta}-x_0}{h}\end{pmatrix},$$

then by the condition C7(i), we have $EW_n = 0$, and

$$\operatorname{Var} W_n = \operatorname{Var}\left(\sum_{i=1}^{n} W_{n,i}\right) = nEW_{n,1}^2 + 2\sum_{j=2}^{n}(n-j+1)\operatorname{Cov}(W_{n,1},W_{n,j}).$$

Similar to the proof methods in Lemma 1, we have

$$\operatorname{Var} W_n = nhG_2(x)p(x_0)\beta_1(x_0)V_1\big(1 + o(1)\big).$$

Next, we will show the asymptotic normality of $\frac{1}{\sqrt{nh}}W_n$, and this can be shown by using similar methods to Theorem 2 of [25]. This completes the lemma. □

Lemma 4 *Under the conditions* C1–C5, C6–C8(ii) *and* C10(ii)–13, *we have*

$$\frac{1}{\sqrt{nh}}\begin{pmatrix} \sum_{i=1}^n \psi_2(v_{i\Delta})\beta_2(X_{i\Delta})K(\frac{X_{i\Delta}-x_0}{h/\beta_2(X_{i\Delta})}) \\ \sum_{i=1}^n \psi_2(v_{i\Delta})\beta_2(X_{i\Delta})K(\frac{X_{i\Delta}-x_0}{h/\beta_2(X_{i\Delta})})\frac{X_{i\Delta}-x_0}{h} \end{pmatrix} \xrightarrow{D} N(0,\Sigma_4),$$

where $\Sigma_4 = H_2(x_0)p(x_0)\beta_2(x_0)V_2$.

Proof of Lemma 4 The proof methods are similar to those used in Lemma 3. □

Proof of Theorem 1 (i) We now show that the new robust estimators of $\mu(x)$ and $\mu'(x)$ are consistent. Let

$$\begin{aligned} &r = (a_1, hb_1)^T, \qquad r_0 = \big(\mu(x_0), h\mu'(x_0)\big)^T, \\ &r_{i\Delta} = (r - r_0)^T \begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix}, \end{aligned}$$

and

$$L_n(r) = \sum_{i=1}^n \rho_1\left(\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} - a_1 - b_1(X_{i\Delta} - x_0)\right)\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right).$$

Then we have

$$\begin{aligned} r_{i\Delta} &= (r - r_0)^T \begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix} \\ &= \big(a_1 - \mu(x_0), hb_1 - h\mu'(x_0)\big)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix} \\ &= a_1 - \mu(x_0) + \big(hb_1 - h\mu'(x_0)\big)\frac{X_{i\Delta} - x_0}{h} \\ &= a_1 - \mu(x_0) + \big(b_1 - \mu'(x_0)\big)(X_{i\Delta} - x_0) \\ &= a_1 + b_1(X_{i\Delta} - x_0) - \mu(x_0) - \mu'(x_0)(X_{i\Delta} - x_0) \\ &= a_1 + b_1(X_{i\Delta} - x_0) + R_1(X_{i\Delta}) - \mu(X_{i\Delta}) \\ &= a_1 + b_1(X_{i\Delta} - x_0) + R_1(X_{i\Delta}) - \left(\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} - u_{i\Delta}\right). \end{aligned}$$

W denote the circle centered at r_0 and with radius δ by S_δ. $\forall\delta > 0$, we prove

$$\lim_{n\to\infty} P\Big\{\inf_{r\in S_\delta} L_n(r) > L_n(r_0)\Big\} = 1. \tag{6}$$

For $r \in S_\delta$,

$$
\begin{aligned}
&L_n(r) - L_n(r_0) \\
&\quad = \sum_{i=1}^{n} \rho_1\left(\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} - a_1 - b_1(X_{i\Delta} - x_0)\right)\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \\
&\qquad - \sum_{i=1}^{n} \rho_1\left(\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} - \mu(x_0) - \mu'(x_0)(X_{i\Delta} - x_0)\right)\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \\
&\quad = \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\left[\rho_1\left(u_{i\Delta} + R_1(X_{i\Delta}) - r_{i\Delta}\right) - \rho_1\left(u_{i\Delta} + R_1(X_{i\Delta})\right)\right] \\
&\quad = \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \psi_1(t)\,dt \\
&\quad = \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \psi_1(u_{i\Delta})\,dt \\
&\qquad + \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \psi_1'(u_{i\Delta})(t - u_{i\Delta})\,dt \\
&\qquad + \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \\
&\qquad \times \int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \left[\psi_1(t) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})(t - u_{i\Delta})\right]dt \\
&\quad =: L_{n1} + L_{n2} + L_{n3}.
\end{aligned}
$$

Next, we will show that

$$L_{n1} = o_p(nh\delta), \tag{7}$$

$$L_{n2} = \frac{nh}{2}(r - r_0)^T G_1(x_0)p(x_0)U_1\left(1 + o_p(1)\right)(r - r_0) + O_p\left(nh^3\delta\right), \tag{8}$$

$$L_{n3} = o_p\left(nh\delta^2\right). \tag{9}$$

For (7), we have

$$
\begin{aligned}
L_{n1} &= \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right) \int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \psi_1(u_{i\Delta})\,dt \\
&= \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1(u_{i\Delta})(-r_{i\Delta}) \\
&= -(r - r_0)^T \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1(u_{i\Delta}) \begin{pmatrix} 1 \\ \frac{X_{i\Delta} - x_0}{h} \end{pmatrix} \\
&= -(r - r_0)^T W_n,
\end{aligned}
$$

where

$$W_n = \sum_{i=1}^{n} W_{n,i} = \sum_{i=1}^{n} \psi_1(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix}.$$

By the proof of Lemma 3, we have $EW_n = 0$, and

$$\operatorname{Var} W_n = nhG_2(x)p(x_0)\beta_1(x_0)V_1\bigl(1 + o(1)\bigr).$$

Note that

$$W_n = \sum_{i=1}^{n} W_{n,i} = E\left(\sum_{i=1}^{n} W_{n,i}\right) + O_p\left(\sqrt{\operatorname{Var}\left(\sum_{i=1}^{n} W_{n,i}\right)}\right),$$

so we have $W_n = O_p(\sqrt{nh})$, which means that (7) holds.

For (8), we have

$$\begin{aligned}
L_{n2} &= \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \bigl[\psi_1'(u_{i\Delta})(t - u_{i\Delta})\bigr]\,dt \\
&= \frac{1}{2}\sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1'(u_{i\Delta})\bigl(r_{i\Delta}^2 - 2R_1(X_{i\Delta})r_{i\Delta}\bigr) \\
&= \frac{1}{2}\sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1'(u_{i\Delta})(r - r_0)^T\begin{pmatrix} 1 & \frac{X_{i\Delta}-x_0}{h} \\ \frac{X_{i\Delta}-x_0}{h} & \frac{(X_{i\Delta}-x_0)^2}{h^2} \end{pmatrix}(r - r_0) \\
&\quad - \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1'(u_{i\Delta})R_1(X_{i\Delta})r_{i\Delta} \\
&=: L_{n21} + L_{n22}.
\end{aligned}$$

From Lemma 1 with $l = 0$, $l = 1$ and $l = 2$, we get

$$\begin{aligned}
L_{n21} &= \frac{1}{2}\sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1'(u_{i\Delta})(r - r_0)^T\begin{pmatrix} 1 & \frac{X_{i\Delta}-x_0}{h} \\ \frac{X_{i\Delta}-x_0}{h} & \frac{(X_{i\Delta}-x_0)^2}{h^2} \end{pmatrix}(r - r_0) \\
&= \frac{nh}{2}(r - r_0)^T G_1(x_0)p(x_0)\begin{pmatrix} K_0 & \frac{K_1}{\beta_1(x_0)} \\ \frac{K_1}{\beta_1(x_0)} & \frac{K_2}{\beta_1^2(x_0)} \end{pmatrix}\bigl(1 + o_p(1)\bigr)(r - r_0) \\
&= \frac{nh}{2}(r - r_0)^T G_1(x_0)p(x_0)U_1\bigl(1 + o_p(1)\bigr)(r - r_0)
\end{aligned}$$

and

$$\begin{aligned}
L_{n22} &= -\sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1'(u_{i\Delta})R_1(X_{i\Delta})r_{i\Delta} \\
&= -(r - r_0)^T\sum_{i=1}^{n} \beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)\psi_1'(u_{i\Delta})R_1(X_{i\Delta})\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{nh^3}{2}(r-r_0)^T G_1(x_0)\mu''(x_0)p(x_0)\begin{pmatrix} \frac{K_2}{\beta_1^2(x_0)} \\ \frac{K_3}{\beta_1^3(x_0)} \end{pmatrix}\bigl(1+o_p(1)\bigr) \\
&= O_p\bigl(nh^3\delta\bigr).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
L_{n2} &= L_{n21} + L_{n22} \\
&= \frac{nh}{2}(r-r_0)^T G_1(x_0)p(x_0)U_1\bigl(1+o_p(1)\bigr)(r-r_0) + O_p\bigl(nh^3\delta\bigr).
\end{aligned}
$$

For (9), by the integral mean value theorem, we have

$$
\begin{aligned}
L_{n3} &= \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\biggl(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\biggr) \\
&\quad \times \int_{u_{i\Delta}+R_1(X_{i\Delta})}^{u_{i\Delta}+R_1(X_{i\Delta})-r_{i\Delta}} \bigl[\psi_1(t)-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})(t-u_{i\Delta})\bigr]\,dt \\
&= \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\biggl(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\biggr)\int_{R_1(X_{i\Delta})}^{R_1(X_{i\Delta})-r_{i\Delta}} \bigl[\psi_1(t+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})t\bigr]\,dt \\
&= \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\biggl(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\biggr)\bigl[\psi_1(z_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})z_{i\Delta}\bigr](-r_{i\Delta}) \\
&= -(r-r_0)^T \sum_{i=1}^{n} \beta_1(X_{i\Delta})K\biggl(\frac{X_{i\Delta}-x_0}{h/\beta_1(X_{i\Delta})}\biggr)\bigl[\psi_1(z_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})z_{i\Delta}\bigr] \\
&\quad \times \begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix},
\end{aligned}
$$

where $z_{i\Delta}$ $(i = 1, 2, \dots, n)$ lies between $R_1(X_{i\Delta})$ and $R_1(X_{i\Delta}) - r_{i\Delta}$.

By $|X_{i\Delta} - x_0| \le \frac{h}{\min_x \beta_1(x)}$, we have

$$
\begin{aligned}
\max_i |z_{i\Delta}| &\le \max_i \bigl|R_1(X_{i\Delta})\bigr| + \left|(r-r_0)^T \begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x_0}{h} \end{pmatrix}\right| \\
&\le \max_i \bigl|R_1(X_{i\Delta})\bigr| + \biggl(1+\frac{1}{\min_x \beta_1(x)}\biggr)\delta,
\end{aligned}
\tag{10}
$$

and according to Taylor's expansion,

$$
\begin{aligned}
\max_i \bigl|R_1(X_{i\Delta})\bigr| &= \max_i \bigl|\mu(X_{i\Delta})-\mu(x_0)-\mu'(x_0)(X_{i\Delta}-x_0)\bigr| \\
&= \max_i \left|\frac{1}{2}\mu''(\xi_i)(X_{i\Delta}-x_0)^2\right| \\
&\le O_p\bigl(h^2\bigr),
\end{aligned}
\tag{11}
$$

where ξ_i lies between x_0 and $X_{i\Delta}$, $i = 1, 2, \dots, n$.

$\forall \eta > 0$, let $D_\eta = \{(\delta_{1\Delta}, \delta_{2\Delta}, \ldots, \delta_{n\Delta})^T : |\delta_{i\Delta}| \leq \eta, \forall i \leq n\}$, by the condition C9(i) and $|X_{i\Delta} - x_0| \leq \frac{h}{\min_x \beta_1(x)}$, we get

$$\begin{aligned}
&E\left[\sup_{D_\eta}\left|\sum_{i=1}^{n}\left[\psi_1(\delta_{i\Delta} + u_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})\delta_{i\Delta}\right]\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta} - x_0)^l\right|\right] \\
&\quad\leq E\left[\sum_{i=1}^{n}\sup_{D_\eta}\left|\psi_1(\delta_{i\Delta} + u_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})\delta_{i\Delta}\right|\beta_1(X_{i\Delta})\right. \\
&\qquad\left.\times K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)|X_{i\Delta} - x_0|^l\right] \\
&\quad\leq a_\eta \delta E\left[\sum_{i=1}^{n}\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)|X_{i\Delta} - x_0|^l\right] \\
&\quad\leq b_\eta \delta,
\end{aligned}$$

where $a_\eta > 0$, $b_\eta > 0$ are two sequences, and satisfy $a_\eta \to 0$ and $b_\eta \to 0$ as $\eta \to 0$. From (10) and (11), we can see that

$$\begin{aligned}
&\sum_{i=1}^{n}\left[\psi_1(z_{i\Delta} + u_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})z_{i\Delta}\right]\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x_0}{h/\beta_1(X_{i\Delta})}\right)(X_{i\Delta} - x_0)^l \\
&\quad= o_p\left(nh^{l+1}\delta\right),
\end{aligned}$$

we get (9) immediately.

Let U_1 be a positive definite matrix, λ be the smallest eigenvalue of the U_1. Accordingly, for any $r \in S_\delta$,

$$\begin{aligned}
&L_n(r) - L_n(r_0) \\
&\quad= L_{n1} + L_{n2} + L_{n3} \\
&\quad= \frac{nh}{2}G_1(x_0)p(x_0)(r - r_0)^T U_1(r - r_0)\left(1 + o_p(1)\right) + O_p\left(nh^3\delta\right) \\
&\quad\geq \frac{nh}{2}G_1(x_0)p(x_0)\lambda\delta^2\left(1 + o_p(1)\right) + O_p\left(nh^3\delta\right).
\end{aligned}$$

So as $n \to \infty$, we have

$$P\left\{\inf_{r\in S_\delta} L_n(r) - L_n(r_0) > \frac{nh}{2}G_1(x_0)p(x_0)\lambda\delta^2 > 0\right\} \to 1,$$

it follows that (6) holds. In view of (6), one can easily see that $L_n(r)$ has a local minimum in the interior of S_δ, thus there exist solutions to equation (4). Let $(h\hat{\mu}(x_0), h\hat{\mu}'(x_0))^T$ denote the closest solution to $r_0 = (\mu(x_0), \mu'(x_0))^T$, then

$$\lim_{n\to\infty} P\left\{\left(\hat{\mu}(x_0) - \mu(x_0)\right)^2 + h^2\left(\hat{\mu}'(x_0) - \mu'(x_0)\right)^2 \leq \delta^2\right\} = 1,$$

we finish the proof of the consistency of the proposed estimators of $\mu(x)$ and $\mu'(x)$.

(ii) We derive the asymptotic normality of the new robust estimators of $\mu(x)$ and $\mu'(x)$. Let

$$\hat{\eta}_{i\Delta} = R_1(X_{i\Delta}) - \big(\hat{\mu}(x_0) - \mu(x_0)\big) - \big(\hat{\mu}'(x_0) - \mu'(x_0)\big)(X_{i\Delta} - x_0). \tag{12}$$

Then we have

$$\begin{aligned}\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} &= \mu(X_{i\Delta}) + u_{i\Delta}\\ &= u_{i\Delta} + \mu(X_{i\Delta}) - \mu(x_0) - \mu'(x_0)(X_{i\Delta} - x_0) + \mu(x_0) + \mu'(x_0)(X_{i\Delta} - x_0)\\ &= u_{i\Delta} + R_1(X_{i\Delta}) + \hat{\mu}(x_0) + \hat{\mu}'(x_0)(X_{i\Delta} - x_0) + \hat{\eta}_{i\Delta} - R_1(X_{i\Delta})\\ &= \hat{\mu}(x_0) + \hat{\mu}'(x_0)(X_{i\Delta} - x_0) + u_{i\Delta} + \hat{\eta}_{i\Delta}.\end{aligned}$$

Therefore by (4), we have

$$\sum_{i=1}^{n} \psi_1(u_{i\Delta} + \hat{\eta}_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x}{h}\end{pmatrix} = 0. \tag{13}$$

Let

$$\begin{aligned}T_{n1} &= \sum_{i=1}^{n} \psi_1(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x}{h}\end{pmatrix} = W_n,\\ T_{n2} &= \sum_{i=1}^{n} \psi_1'(u_{i\Delta})\hat{\eta}_{i\Delta}\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x}{h}\end{pmatrix},\\ T_{n3} &= \sum_{i=1}^{n} \big[\psi_1(u_{i\Delta} + \hat{\eta}_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})\hat{\eta}_{i\Delta}\big]\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x}{h}\end{pmatrix}.\end{aligned}$$

In view of (13), one can get $T_{n1} + T_{n2} + T_{n3} = 0$. Lemma 1 and (12) imply that

$$\begin{aligned}T_{n2} &= \sum_{i=1}^{n} \psi_1'(u_{i\Delta})R_1(X_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x}{h}\end{pmatrix} - \sum_{i=1}^{n} \psi_1'(u_{i\Delta})\beta_1(X_{i\Delta})\\ &\quad\cdot K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} (\hat{\mu}(x_0) - \mu(x_0)) + (\hat{\mu}'(x_0) - \mu'(x_0))(X_{i\Delta} - x_0) \\ \frac{X_{i\Delta}-x}{h}[(\hat{\mu}(x_0) - \mu(x_0)) + (\hat{\mu}'(x_0) - \mu'(x_0))(X_{i\Delta} - x_0)]\end{pmatrix}\\ &= \sum_{i=1}^{n} \psi_1'(u_{i\Delta})R_1(X_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 \\ \frac{X_{i\Delta}-x}{h}\end{pmatrix}\\ &\quad - \sum_{i=1}^{n} \psi_1'(u_{i\Delta})\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta} - x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix} 1 & \frac{X_{i\Delta}-x}{h} \\ \frac{X_{i\Delta}-x}{h} & \frac{(X_{i\Delta}-x)^2}{h^2}\end{pmatrix}\begin{pmatrix} \hat{\mu}(x_0) - \mu(x_0) \\ h(\hat{\mu}'(x_0) - \mu'(x_0))\end{pmatrix}\\ &= \frac{nh^3}{2}G_1(x_0)\mu''(x_0)p(x_0)\begin{pmatrix} \frac{K_2}{\beta_1^2(x_0)} \\ \frac{K_3}{\beta_1^3(x_0)}\end{pmatrix}\big(1 + o_p(1)\big)\\ &\quad - nhG_1(x_0)p(x_0)\begin{pmatrix} K_0 & \frac{K_1}{\beta_1(x_0)} \\ \frac{K_1}{\beta_1(x_0)} & \frac{K_2}{\beta_1^2(x_0)}\end{pmatrix}\big(1 + o_p(1)\big)\begin{pmatrix} \hat{\mu}(x_0) - \mu(x_0) \\ h(\hat{\mu}'(x_0) - \mu'(x_0))\end{pmatrix}\end{aligned}$$

$$= \frac{nh^3 G_1(x_0)\mu''(x_0)p(x_0)}{2\beta_1^2(x_0)} A_1\big(1+o_p(1)\big)$$
$$- nhG_1(x_0)p(x_0)U_1\big(1+o_p(1)\big)\begin{pmatrix}\hat{\mu}(x_0)-\mu(x_0)\\ h(\hat{\mu}'(x_0)-\mu'(x_0))\end{pmatrix} =: T_{n21}+T_{n22}.$$

Since we have already got the consistency of $(\hat{\mu}(x_0), h\hat{\mu}'(x_0))$ and using $|X_{i\Delta}-x_0| \le \frac{h}{\min_x \beta_1(x)}$, we have

$$\begin{aligned}
\sup_i |\hat{\eta}_{i\Delta}| &= \sup_i \big|R_1(X_{i\Delta}) - \big(\hat{\mu}(x_0)-\mu(x_0)\big) - \big(\hat{\mu}'(x_0)-\mu'(x_0)\big)(X_{i\Delta}-x_0)\big| \\
&\le \sup_i \big|R_1(X_{i\Delta})\big| + \big|\hat{\mu}(x_0)-\mu(x_0)\big| + \frac{h}{\min_x \beta_1(x)}\big|\hat{\mu}'(x_0)-\mu'(x_0)\big| \\
&= O_p\big(h^2 + \big(\hat{\mu}(x_0)-\mu(x_0)\big) + h\big(\hat{\mu}'(x_0)-\mu'(x_0)\big)\big) = o_p(1),
\end{aligned}$$

then, by the condition C9(i) and the same argument as that in the first part of Theorem 1, we have

$$\begin{aligned}
T_{n3} &= \sum_{i=1}^{n} \big[\psi_1(u_{i\Delta}+\hat{\eta}_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})\hat{\eta}_{i\Delta}\big]\beta_1(X_{i\Delta})K\left(\frac{X_{i\Delta}-x}{h/\beta_1(X_{i\Delta})}\right)\begin{pmatrix}1\\ \frac{X_{i\Delta}-x}{h}\end{pmatrix} \\
&= o_p(nh)\big[h^2 + \big(\hat{\mu}(x_0)-\mu(x_0)\big) + h\big(\hat{\mu}'(x_0)-\mu'(x_0)\big)\big] = o_p(T_{n22}).
\end{aligned}$$

Therefore, using the fact that $T_{n1}+T_{n2}+T_{n3}=0$, we have

$$\begin{aligned}
&\begin{pmatrix}\hat{\mu}(x_0)-\mu(x_0)\\ h(\hat{\mu}'(x_0)-\mu'(x_0))\end{pmatrix} \\
&\quad = \frac{1}{nh}G_1^{-1}(x_0)p^{-1}(x_0)U_1^{-1}\big(1+o_p(1)\big)W_n + \frac{h^2}{2\beta_1^2(x_0)}\mu''(x_0)U_1^{-1}A_1\big(1+o_p(1)\big),
\end{aligned}$$

it follows that

$$\begin{aligned}
&\sqrt{nh}\left[\begin{pmatrix}\hat{\mu}(x_0)-\mu(x_0)\\ h(\hat{\mu}'(x_0)-\mu'(x_0))\end{pmatrix} - \frac{h^2\mu''(x_0)}{2\beta_1^2(x_0)}U_1^{-1}A_1\big(1+o_p(1)\big)\right] \\
&\quad = G_1^{-1}(x_0)p^{-1}(x_0)U_1^{-1}\big(1+o_p(1)\big)\frac{1}{\sqrt{nh}}W_n.
\end{aligned}$$

According to Lemma 3 and the Slutsky theorem, we have

$$\begin{aligned}
&\sqrt{nh}\left[\begin{pmatrix}\hat{\mu}(x_0)-\mu(x_0)\\ h(\hat{\mu}'(x_0)-\mu'(x_0))\end{pmatrix} - \frac{h^2\mu''(x_0)}{2\beta_1^2(x_0)}U_1^{-1}A_1\right] \\
&\quad \xrightarrow{D} G_1^{-1}(x_0)p^{-1}(x_0)U_1^{-1}N(0,\Sigma_3) \\
&\quad = N\left(0, \frac{G_2(x_0)\beta_1(x_0)}{G_1^2(x_0)p(x_0)}U_1^{-1}V_1U_1^{-1}\right) \\
&\quad = N(0,\Sigma_1).
\end{aligned}$$

This completes the proof. □

Proof of Theorem 2 By using Lemma 2 and Lemma 4 instead of Lemma 1 and 3, the proof of this theorem is similar to Theorem 1, so it is omitted here. □

5 Conclusions

In this paper, new variable bandwidth nonparametric robust estimators for the drift function and diffusion function of diffusion processes based on discrete-time observations are devised. The new estimators are based on the local linear regression technique and the maximum likelihood type estimation technique, and they have a good control of outliers. The proposed estimators are proved to be consistent and asymptotically normal.

The authors of [23–25] developed robust version estimators of regression function for stationary time series sequence under independent data and dependent data, respectively. Based on their research, in this paper, we studied a continuous-time diffusion process determined by a stochastic differential equation, and our robust estimators based on local linear regression techniques; the reader can consider a robust estimator for a diffusion process by using local polynomial regression techniques. Moreover, the basic ideas of this paper have good generality and can be extended to other continuous-time stochastic models.

In addition, in this paper, we only considered robust estimators for the one dimensional diffusion process; in fact, the basic ideas of our methodology hold for the case of multidimensional diffusion processes and situations.

Acknowledgements
The authors would like to thank the editor and referees for their valuable suggestions, which improved the structure and the presentation of the paper.

Funding
This research is supported by the NSFC (No.11461032, No.11401267, No.11401169, No.11561028), the Program of Qingjiang Excellent Young Talents, Jiangxi University of Science and Technology, and the NSF of Jiangxi Province (No. 20161BAB211014 and No. 20171BAB201008).

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors jointly worked on the results and they read and approved the final manuscript.

Publisher's Note

Received: 10 August 2017 Accepted: 29 January 2018 Published online: 06 February 2018

References

1. Chaiyapo, N., Phewchean, N.: An application of Ornstein–Uhlenbeck process to commodity pricing in Thailand. Adv. Differ. Equ. **2017**, 179 (2017)
2. Zou, T., Chen, S.X.: Enhancing estimation for interest rate diffusion models with bond prices. J. Bus. Econ. Stat. **35**, 486–498 (2017)
3. Tang, C.Y., Chen, S.X.: Parameter estimation and bias correction for diffusion processes. J. Econom. **149**, 65–81 (2009)
4. Leonenko, N.N., Šuvak, N.: Statistical inference for student diffusion process. Stoch. Anal. Appl. **28**, 972–1002 (2010)
5. Leonenko, N.N., Šuvak, N.: Statistical inference for reciprocal gamma diffusion process. J. Stat. Plan. Inference **140**, 30–51 (2010)
6. Li, J., Wu, J.L.: On drift parameter estimation for mean-reversion type stochastic differential equations with discrete observations. Adv. Differ. Equ. **2016**, 90 (2016)
7. Nishiyama, Y.: Asymptotic theory of semiparametric Z-estimators for stochastic processes with applications to ergodic diffusions and time series. Ann. Stat. **37**, 3555–3579 (2009)
8. Kristensen, D.: Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. J. Econom. **156**, 239–259 (2010)
9. Koo, B., Linton, O.: Estimation of semiparametric locally stationary diffusion models. J. Econom. **170**, 210–233 (2012)
10. Bandi, F., Phillips, P.: Fully nonparametric estimation of scalar diffusion models. Econometrica **71**, 241–283 (2003)

11. Comte, F., Genon-Catalot, V., Rozenholc, Y.: Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli **13**, 514–543 (2007)
12. Fan, J., Zhang, C.: A re-examination of Stanton's diffusion estimations with applications to financial model validation. J. Am. Stat. Assoc. **98**, 118–134 (2003)
13. Xu, K.L.: Empirical likelihood-based inference for nonparametric recurrent diffusions. J. Econom. **153**, 65–82 (2009)
14. Xu, K.L.: Re-weighted functional estimation of diffusion models. Econom. Theory **26**, 541–563 (2010)
15. Kristensen, D.: Nonparametric filtering of the realized spot volatility: a kernel-based approach. Econom. Theory **26**, 60–93 (2010)
16. van der Meulen, F., Schauer, M., van Zanten, H.: Reversible jump MCMC for nonparametric drift estimation for diffusion processes. Comput. Stat. Data Anal. **71**, 615–632 (2014)
17. Florens-Zmirou, D.: On estimating the diffusion coefficient from discrete observations. J. Appl. Probab. **30**, 790–804 (1993)
18. 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. Econom. Theory **13**, 615–645 (1997)
19. Stanton, R.: A nonparametric model of term structure dynamics and the market price of interest rate risk. J. Finance **52**, 1973–2002 (1997)
20. Huber, P.J.: Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Stat. **1**, 799–821 (1973)
21. Delecroix, M., Hristache, M., Patilea, V.: On semiparametric M-estimation in single-index regression. J. Stat. Plan. Inference **136**, 730–769 (2006)
22. Davis, R.A., Knight, K., Liuc, J.: M-estimation for autoregressions with infinite variance. Stoch. Process. Appl. **40**, 145–180 (1992)
23. Fan, J., Jiang, J.: Variable bandwidth and one-step local M-estimator. Sci. China Math. **43**, 65–81 (2000)
24. Jiang, J.C., Mack, Y.P.: Robust local polynomial regression for dependent data. Stat. Sin. **11**, 705–722 (2001)
25. Cai, Z., Ould-Said, E.: Local M-estimator for nonparametric time series. Stat. Probab. Lett. **65**, 433–449 (2003)
26. Wang, Y.Y., Tang, M.T.: Nonparametric robust function estimation for integrated diffusion processes. J. Nonlinear Sci. Appl. **9**, 3048–3065 (2016)
27. Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Springer, New York (2005)
28. Fan, J., Gijbels, I.: Variable bandwidth and local linear regression smoothers. Ann. Stat. **29**, 2008–2036 (1992)
29. Fan, J., Gijbels, I.: Data-driven bandwidth selection in hold polynomial fitting: variable bandwidth and spatial adaption. J. R. Stat. Soc. B **57**, 371–394 (1995)
30. Hall, P., Hu, T.C., Marron, J.S.: Improved variable window estimators of probability densities. Ann. Stat. **23**, 1–10 (1995)
31. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus. Springer, New York (2012)
32. Hansen, L., Scheinkman, J.: Back to the future: generating moment implications for continuous-time Markov processes. Econometrica **63**, 767–804 (1995)