Available online at www.tjnsa.com
J. Nonlinear Sci. Appl. 9 (2016), 3048–3065
Research Article

Journal of Nonlinear Science and Applications
Print: ISSN 2008-1898 Online: ISSN 2008-1901

Nonparametric robust function estimation for integrated diffusion processes

Yunyan Wang, Mingtian Tang*

School of Science, Jiangxi University of Science and Technology, No. 86, Hongqi Ave., Ganzhou, 341000, Jiangxi, P. R. China.

Communicated by R. Saadati

Abstract

This paper considers local M-estimation of the unknown drift and diffusion functions of integrated diffusion processes. We show that under appropriate conditions, the proposed estimators for drift and diffusion functions in the integrated process are consistent, and the conditions that ensure the asymptotic normality of these local M-estimators are also stated. The simulation studies show that the proposed estimators perform better than the kernel estimator in robustness.

Keywords: Integrated diffusion process, local linear estimator, M-estimation, robust estimation.
2010 MSC: 62G35, 62G20.

1. Introduction

In this paper, we consider the stochastic process $Y_t = \int_0^t X_s ds$ where X is a one-dimensional diffusion process given by

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{1.1}$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion (or Wiener process), and $\mu(\cdot)$ and $\sigma(\cdot)$, the drift and diffusion of the process $\{X_t\}$, are functions only of the contemporaneous value of X_t. The integrated diffusion process (Y_t, X_t) solves the following second-order stochastic differential equation:

$$\begin{cases} dY_t = X_t dt, \\ dX_t = \mu(X_t)dt + \sigma(X_t)dB_t. \end{cases} \tag{1.2}$$

*Corresponding author
Email addresses: `yywang@zju.edu.cn` (Yunyan Wang), `mttang_csu@yahoo.com` (Mingtian Tang)

Received 2015-10-09

This integrated diffusion processes can model integrated and differentiated diffusion processes, and at the same time overcome the difficulties associated with the nondifferentiability of the Brownian motion, so these models play an important role in the financial and economic field.

As we all know, the diffusion models (1.1) represent a widely accepted class of processes in finance and economics. Estimation and inference about infinitesimal mean function $\mu(\cdot)$ and infinitesimal variance function $\sigma(\cdot)$ in (1.1) has been an important area in modern econometrics, especially the estimators proposed and studied based on a discrete time observation of the sample path. For example, one can view [3, 4] for parametric estimation and [18] for nonparametric estimation based on low-frequency data, as for the high-frequency data, one can view [2, 6, 14] and so on.

However, all sample functions of a diffusion process (1.1) driven by a Brownian motion are of unbounded variation and nowhere differentiable, so it can not model integrated and differentiated diffusion processes. On the other hand, integrated diffusion processes play an important role in the field of finance, engineering and physics. For example, X_t may represent the velocity of a particle and Y_t its coordinate, see e.g. [11, 23]. P. D. Ditlevsen and et al. in [10] used integrated diffusion process to model and analyze the ice-core data.

Statistical inference for discretely observed integrated diffusion processes has been considered recently. For example, parametric inference for integrated diffusion processes has been addressed by [11, 15, 16, 17]. As for the nonparametric inferences, [22] obtained the Nadaraya-Waston estimators for drift and diffusion functions, [24] studied the local linear estimators for the two functions, and [25] developed the re-weighted estimator of the diffusion function.

All the above nonparametric estimators are based on local polynomial regression smoothers. However, local polynomial regression smoothers are not robust, so there is a growing literature on the robust methods. One popular robust technique is the so-called M-estimator, which is the easiest to cope with as far as asymptotic theory is concerned as pointed out by [20]. Therefore, M-estimation has been studied by many authors such as [8, 9, 19] and references therein. Furthermore, some modified M-estimators were proposed, such as local M-estimator, which is a combination of the local linear smoothing technique and the M-estimation technique. The local M-estimators inherit the nice properties from not only M-estimators but also local linear estimators. For example, [13] proposed local linear M-estimator with variable bandwidth for regression function, [21] developed a robust estimator of the regression function by using local polynomial regression techniques.

So in the present paper, we study robust nonparametric statistical inference for drift and diffusion functions of integrated diffusions. We wish to construct local M-estimators of the drift and diffusion coefficients in integrated diffusion processes (1.2) based on local linear smoothing technique and the M-estimation technique.

The remainder of the paper is organized as follows. Section 2 introduces the local M-estimators of drift and diffusion coefficients in integrated diffusion model and develops the asymptotic results of the estimators under some mild conditions. Simulation studies are developed in Section 3. Some useful lemmas and all mathematical proofs are presented in Section 4.

2. Local M-estimator and Asymptotic Results

2.1. Local M-estimators

Just as [22, 25] pointed out that there is a difficulty to estimate the integrated diffusion processes, i.e., the value of X in model (1.2) at time t_i is impossible to obtain, and at the same time the estimation of model (1.2) can not be based on the observations $\{Y_{t_i}, i = 1, 2, \cdots\}$.

To simplify we use the notation $t_i = i\Delta$,where $\Delta = t_i - t_{i-1}$, by $Y_t = \int_0^t X_u du$, we have

$$\frac{Y_{i\Delta} - Y_{(i-1)\Delta}}{\Delta} = \frac{1}{\Delta}\left(\int_0^{i\Delta} X_u du - \int_0^{(i-1)\Delta} X_u du\right) = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} X_u du,$$

when Δ tends to zero, the values of $X_{i\Delta}$, $X_{(i-1)\Delta}$ and $\frac{Y_{i\Delta}-Y_{(i-1)\Delta}}{\Delta}$ become nearly to each other, so our estimation procedure will be based on the following equivalent set of data,

$$\tilde{X}_{i\Delta} = \frac{Y_{i\Delta} - Y_{(i-1)\Delta}}{\Delta}.$$

The estimation of drift and diffusion functions in the integrated diffusion model (1.2) depends on the following equations:

$$E\left(\frac{\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta}}{\Delta}\middle| \mathcal{F}_{(i-1)\Delta}\right) = \mu(X_{(i-1)\Delta}) + o(1), \quad \Delta \to 0, \tag{2.1}$$

$$E\left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta})^2}{\Delta}\middle| \mathcal{F}_{(i-1)\Delta}\right) = \sigma^2(X_{(i-1)\Delta}) + o(1), \quad \Delta \to 0, \tag{2.2}$$

where $\mathcal{F}_t = \sigma\{X_s, s \le t\}$. Eqs. (2.1) and (2.2) can be obtained from Lemma 4.1 in Section 4, readers can also refer to [7] or [22] for more details about them.

By (2.1), the local linear estimator for $\mu(x)$ is defined as the solution to the following problem: Choose a_1 and b_1 to minimize the following weighted sum

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

and by (2.2) the local linear estimator for $\sigma^2(x)$ is defined as the solution to the following problem: Choose a_2 and b_2 to minimize the following weighted sum

$$\sum_{i=1}^{n}\left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta})^2}{\Delta} - a_2 - b_2(\tilde{X}_{i\Delta} - x)\right)^2 K\left(\frac{\tilde{X}_{(i-1)\Delta} - x}{h}\right),$$

where $K(\cdot)$ is the kernel function and $h = h_n$ is bandwidth.

However, the above local linear estimators are not robust. Therefore, let us choose a_1 and b_1 to minimize

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

and a_2 and b_2 to minimize

$$\sum_{i=1}^{n}\rho_2\left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta})^2}{\Delta} - a_2 - b_2(\tilde{X}_{i\Delta} - x)\right) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x}{h}\right),$$

or to satisfy the following equations:

$$\sum_{i=1}^{n}\psi_1\left(\frac{\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta}}{\Delta} - a_1 - b_1(\tilde{X}_{i\Delta} - x)\right) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x}{h}\end{pmatrix} = \begin{pmatrix} 0 \\ 0\end{pmatrix}, \tag{2.3}$$

and

$$\sum_{i=1}^{n}\psi_2\left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta})^2}{\Delta} - a_2 - b_2(\tilde{X}_{i\Delta} - x)\right) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x}{h}\end{pmatrix} = \begin{pmatrix} 0 \\ 0\end{pmatrix}, \tag{2.4}$$

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

The local M-estimators of $\mu(x)$ and $\mu'(x)$ are denoted as $\hat{\mu}(x) = \hat{a}_1$ and $\hat{\mu}'(x) = \hat{b}_1$, which are the solutions of (2.3), the local M-estimators of $\sigma^2(x)$ and $(\sigma^2(x))'$ are denoted as $\hat{\sigma}^2(x) = \hat{a}_2$ and $(\hat{\sigma}^2(x))' = \hat{b}_2$, which are the solutions of (2.4).

2.2. Assumptions and Asymptotic Results

Suppose x_0 is a given point, the asymptotic results for our local M-estimators of drift and diffusion functions in integrated diffusion processes need the following conditions.

(**A1**). ([22])

(i) Let interval $\mathcal{D} = (l, r)$ be the state space of X, $s(z) = \exp\{-\int_{z_0}^{z} \frac{2\mu(x)}{\sigma^2(x)} dx\}$ is the scale density function (z_0 is an arbitrary point inside $\mathcal{D}$). For $x \in \mathcal{D}$, $l < x_1 < x < x_2 < r$, let

$$S(l, x] = \lim_{x_1 \to l} \int_{x_1}^{x} s(u)du = \infty, \quad S[x, r) = \lim_{x_2 \to r} \int_{x}^{x_2} s(u)du = \infty;$$

(ii) $\int_l^r m(x)dx < \infty$, where $m(x) = (\sigma^2(x)s(x))^{-1}$ is a speed density function;

(iii) $X_0 = x$ has distribution P^0, where P^0 is the invariant distribution of the ergodic process $(X_t)_{t\in[0,\infty)}$.

(**A2**). Let interval $\mathcal{D} = (l, r)$ be the state space of X, we assume that

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

Remark 2.1. (A1) assures that X is stationary ([1]), the stationary density of X denotes as $p(x)$ in this paper. (A2) assures that the process X is $\alpha-$mixing. Under assumptions (A1) and (A2), we know that $\{\tilde{X}_{i\Delta}, i = 0, 1, \cdots\}$ is stationary and $\alpha-$mixing.

Let $\alpha(k)$ be the mixing coefficient of process X, we assume that $\alpha(k)$ satisfy the following condition.

(**A3**). The mixing coefficient $\alpha(k)$ of process X satisfies $\sum\limits_{k\geq 1} k^a (\alpha(k))^{\gamma/(2+\gamma)} < \infty$ for some $a > \gamma/(2+\gamma)$, where γ is given in assumption (A10).

The kernel function $K(\cdot)$ and bandwidth h satisfy the following assumptions (A4) and (A5).

(**A4**). The kernel $K(\cdot)$ is a continuously differentiable, symmetric density function compactly supported on $[-1, 1]$.

(**A5**). $\Delta \to 0$, $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$.

Remark 2.2. To simplify our purpose, we consider at this stage only positive and symmetrical kernels which are the most classical ones. As for how to choose bandwidth the book of [12] is recommended.

(**A6**). (i) $\mu(x)$ and $\sigma(x)$ have continuous derivative of order 4 and satisfy $|\mu(x)| \leq C(1 + |x|)^\lambda$ and $|\sigma(x)| \leq C(1 + |x|)^\lambda$ for some $\lambda > 0$;

(ii) $E[X_0^r] < \infty$, where $r = \max\{4\lambda, 1 + 3\lambda, -1 + 5\lambda, -2 + 6\lambda\}$.

Remark 2.3. Assumption (A6) guarantees that Lemma 4.1 can be used properly throughout the paper.

(**A7**). The density function $p(x)$ of the process X is continuous at x_0, and $p(x_0) > 0$. Furthermore, the joint density of $X_{i\Delta}$ and $X_{j\Delta}$ are bounded for all i, j.

(**A8**). (i) $\lim\limits_{h\to 0} \frac{1}{h} E(|K'(\xi_{ni})|^2) < \infty$;

(ii) $\lim\limits_{h\to 0} \frac{1}{h} E(|K'(\xi_{ni})|^4) < \infty$, where

$$\xi_{ni} = \theta((X_{(i-1)\Delta} - x)/h) + (1 - \theta)((\tilde{X}_{(i-1)\Delta} - x)/h), \ 0 \leq \theta \leq 1.$$

(**A9**). (i) $E[\psi_1(u_{i\Delta})|\, X_{(i-1)\Delta}=x]=o(1)$ with $u_{i\Delta}=\frac{\tilde{X}_{(i+1)\Delta}-\tilde{X}_{i\Delta}}{\Delta}-\mu(X_{(i-1)\Delta})$;

(ii) $E[\psi_2(v_{i\Delta})|\, X_{(i-1)\Delta}=x]=o(1)$ with $v_{i\Delta}=\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta}-\tilde{X}_{i\Delta})^2}{\Delta}-\sigma^2(X_{(i-1)\Delta})$.

(**A10**). (i) The function $\psi_1(\cdot)$ is continuous and has a derivative $\psi_1'(\cdot)$ almost everywhere. Furthermore, it is assumed that functions

$$E[\psi_1'(u_{i\Delta})\big|\, X_{(i-1)\Delta}=x]>0,\ E[\psi_1^2(u_{i\Delta})\big|\, X_{(i-1)\Delta}=x]>0,\ E[{\psi_1'}^2(u_{i\Delta})\Big|\, X_{(i-1)\Delta}=x]>0$$

and continuous at the point x_0, and there exists a constant $\gamma>0$ such that

$$E[|\psi_1(u_{i\Delta})|^{2+\gamma}\Big|\, X_{(i-1)\Delta}=x],\quad E[\left|\psi_1'(u_{i\Delta})\right|^{2+\gamma}\Big|\, X_{(i-1)\Delta}=x]$$

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. Furthermore, it is assumed that functions $E[\psi_2'(v_{i\Delta})|\, X_{(i-1)\Delta}=x]>0$, $E[\psi_2^2(v_{i\Delta})\big|\, X_{(i-1)\Delta}=x]>0$, $E[{\psi_2'}^2(v_{i\Delta})\Big|\, X_{(i-1)\Delta}=x]>0$, and continuous at the point x_0, and there exists a constant $\gamma>0$ such that $E[|\psi_2(v_{i\Delta})|^{2+\gamma}\Big|\, X_{(i-1)\Delta}=x]$, $E[|\psi_2'(v_{i\Delta})|^{2+\gamma}\Big|\, X_{(i-1)\Delta}=x]$ are bounded in a neighborhood of x_0.

(**A11**). (i) For any i,j,

$$E[\psi_1^2(u_{i\Delta})+\psi_1^2(u_{j\Delta})\big|\, X_{(i-1)\Delta}=x, X_{(j-1)\Delta}=y],$$
$$E[{\psi_1'}^2(u_{i\Delta})+{\psi_1'}^2(u_{j\Delta})\Big|\, X_{(i-1)\Delta}=x, X_{(j-1)\Delta}=y]$$

are bounded in the neighborhood of x_0;

(ii) For any i,j,

$$E[\psi_2^2(v_{i\Delta})+\psi_2^2(v_{j\Delta})\big|\, X_{(i-1)\Delta}=x, X_{(j-1)\Delta}=y],$$
$$E[{\psi_2'}^2(v_{i\Delta})+{\psi_2'}^2(v_{j\Delta})\Big|\, X_{(i-1)\Delta}=x, X_{(j-1)\Delta}=y]$$

are bounded in the neighborhood of x_0.

(**A12**). (i) The function $\psi_1'(\cdot)$ satisfies

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

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

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

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

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

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

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

Remark 2.4. The conditions in assumptions (A9)-(A12) imposed on $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are mild and satisfied for many applications. Particularly, they are fulfilled for Huber's $\psi(\cdot)$ function. In this paper, we do not need the monotonicity and boundedness of $\psi_i(\cdot), i=1,2$ and the convexity of the function $\rho_i(\cdot), i=1,2$. For more details about these conditions we refer to [5] or [13].

(**A13**). Assume that there exists 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 $n \to \infty$.

(**A14**). There exists $\tau > 2+\gamma$, where γ is given in Assumption 10, such that $E\{|\psi_1(u_{i\Delta})|^\tau | X_{(i-1)\Delta} = x\}$, $E\{|\psi_2(v_{i\Delta})|^\tau | X_{(i-1)\Delta} = x\}$ are bounded for all x in a neighborhood of x_0, and $\alpha(n) = O(n^{-\theta})$, where $\theta \geq (2+\gamma)\tau/\{2(\tau-2-\gamma)\}$.

(**A15**). $n^{-\gamma/4}h^{(2+\gamma)/\tau-1-\gamma/4} = O(1)$, where γ and τ are given in assumptions (A10) and (A14), respectively.

Remark 2.5. Obviously, Assumption 15 is automatically satisfied for $\gamma \geq 1$ and it is also fulfilled for $0<\gamma<1$ if τ satisfies $\gamma < \tau - 2 \leq \gamma/(1-\gamma)$.

Throughout the whole paper, let

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

$$U = \begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix}, \quad V = \begin{pmatrix} J_0 & J_1 \\ J_1 & J_2 \end{pmatrix}, \quad A = \begin{pmatrix} K_2 \\ K_3 \end{pmatrix},$$

$$G_1(x) = E[\psi_1'(u_{i\Delta})\big| X_{(i-1)\Delta} = x], \quad G_2(x) = E[\psi_1^2(u_{i\Delta})\big| X_{(i-1)\Delta} = x],$$

$$H_1(x) = E[\psi_2'(v_{i\Delta})\big| X_{(i-1)\Delta} = x], \quad H_2(x) = E[\psi_2^2(v_{i\Delta})\big| X_{(i-1)\Delta} = x].$$

Our main results are as follows:

Theorem 2.6. *Under assumptions (A1)-(A7) and the conditions (i) of the assumptions (A8)-(A12), there exist solutions $\hat{\mu}(x_0)$ and $\hat{\mu}'(x_0)$ to equation* (2.3) *such that*

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

(ii) Furthermore, if assumptions (A13)-(A15) hold, then

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

where "$\xrightarrow{P}$" means convergence in probability, "$\xrightarrow{D}$" means convergence in distribution, and

$$\Sigma_1 = \frac{G_2(x_0)}{G_1^2(x_0)p(x_0)}U^{-1}VU^{-1}.$$

Theorem 2.7. *Under assumptions (A1)-(A7) and the conditions (ii) of the assumptions (A8)-(A12), there exist solutions $\hat{\sigma}^2(x_0)$ and $(\hat{\sigma}^2(x_0))'$ to equation* (2.4) *such that*

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

(ii) Furthermore, if assumptions (A13)-(A15) hold, then

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

where "$\xrightarrow{P}$" means convergence in probability, "$\xrightarrow{D}$" means convergence in distribution, and

$$\Sigma_2 = \frac{H_2(x_0)}{H_1^2(x_0)p(x_0)}U^{-1}VU^{-1}.$$

3. Simulations

In this section, we perform a Monte Carlo experiment to show the performance of the local M-estimators discussed in the paper by comparing the mean square error (MSE) between the new estimators and the following kernel-type estimators for $\mu(x)$ and $\sigma^2(x)$(see [22]):

$$\bar{\mu}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta}-x}{h}\right)\frac{(\tilde{X}_{(i+1)\Delta}-\tilde{X}_{i\Delta})}{\Delta}}{\sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta}-x}{h}\right)},$$

$$\bar{\sigma}^2(x) = \frac{\sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta}-x}{h}\right)\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta}-\tilde{X}_{i\Delta})^2}{\Delta}}{\sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta}-x}{h}\right)}.$$

Our experiment is based on the simulation of $Y_t = Y_0 + \int_0^t X_u du$, where X is an ergodic process governed by the following stochastic differential equation:

$$dX_t = -10X_t dt + \sqrt{0.1 + 0.1X_t^2} dB_t, \tag{3.1}$$

and we use the Euler-Maruyama method to approximate the numerical solution of our stochastic differential equation.

Throughout we take Huber's function $\psi_1(z) = \max\{-c, \min(c, z)\}$ with $c = 0.135$ and we smooth using a Gauss kernel, and the bandwidth used in our simulation is $h = h_{\text{opt}}$, where h_{opt} is the optimal bandwidth, which minimize the mean square error (MSE):

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

where $\{x_i, i = 1, 2, \cdots, n\}$ are chosen uniformly to cover the range of sample path of X_t, and we obtain $\hat{\mu}(\cdot)$ by iteration since it has no explicit expression. For any initial value $\hat{\mu}_0(x)$, we have

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

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

$$W_n(a_1, b_1) = \begin{pmatrix} \frac{\partial \Psi_n(a_1,b_1)}{\partial a_1}, & \frac{\partial \Psi_n(a_1,b_1)}{\partial b_1} \end{pmatrix},$$

$$\Psi_n(a_1, b_1) = \sum_{i=1}^{n} \psi_1\left(\frac{\tilde{X}_{(i+1)\Delta} - \tilde{X}_{i\Delta}}{\Delta} - a_1 - b_1(\tilde{X}_{i\Delta} - x)\right) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x}{h} \end{pmatrix}.$$

This procedure terminates when

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

In order to illustrate continuous time integrated processes, Figure 1 presents a simulated path of X which is defined by (3.1) and Figure 2 presents a simulated path of $Y_t = Y_0 + \int_0^t X_u du$ where $t \in [0, T] = [0, 10]$ and X is governed by (3.1).

In order to compare the robustness of the local M-estimators with that of the Nadaraya-Watson estimators, Table 1 and Table 2 show the performance of the local M-estimator and the Nadaraya-Watson estimator for drift function $\mu(\cdot)$ and diffusion function $\sigma^2(\cdot)$ in terms of the MSE, respectively. We can see that the local M-estimator performs better than the Nadaraya-Watson estimator, and the performances of the both estimators are improved as the sample size increases.

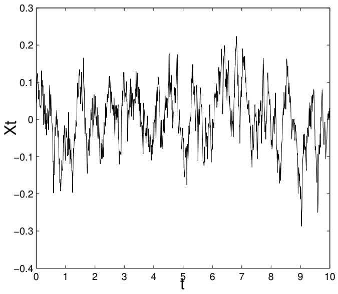

Figure 1: THE SAMPLE PATH OF THE PROCESS X_t.

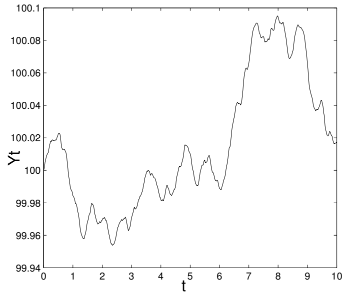

Figure 2: THE SAMPLE PATH OF THE INTEGRATED PROCESS Y_t.

Table 1: The mean square error of Nadaraya-Watson estimator (MSE_{11}) and local M-estimator (MSE_{12}) for drift function $\mu(\cdot)$.

Sample size n	MSE_{11}	MSE_{12}
$n = 100$	0.9084	0.1652
$n = 200$	0.7499	0.1214
$n = 400$	0.5754	0.1272
$n = 800$	0.3742	0.1171

Table 2: The mean square error of Nadaraya-Watson estimator (MSE_{21}) and local M-estimator (MSE_{22}) for diffusion function $\sigma^2(\cdot)$.

Sample size n	MSE_{21}	MSE_{22}
$n = 100$	0.0988	0.0459
$n = 200$	0.0891	0.0458
$n = 400$	0.0881	0.0390
$n = 800$	0.0862	0.0392

4. Lemmas and Proofs

In order to prove Theorem 2.6 and Theorem 2.7 , we need the following lemmas.

Lemma 4.1 ([22]). *Let Z be a d-dimensional diffusion process governed by the stochastic integral equation*

$$Z_t = Z_0 + \int_0^t \mu(Z_s)ds + \int_0^t \sigma(Z_s)dB_s,$$

where $\mu(z) = [\mu_i(z)]_{d\times 1}$ is a $d \times 1$ vector, $\sigma(z) = [\sigma_{ij}(z)]_{d\times d}$ is a $d \times d$ diagonal matrix, and B_t is a $d \times 1$ vector of independent Brownian motions. Assume that μ and σ have continuous partial derivatives of order $2s$. Let $f(z)$ be a continuous function defined on $\mathbb{R}^d$ with values in $\mathbb{R}^d$ and with continuous partial derivative of order $2s+2$. Then

$$E[f(Z_{i\Delta})|Z_{(i-1)\Delta}] = \sum_{k=0}^{s} L^k f(Z_{(i-1)\Delta})\frac{\Delta^k}{k!} + R,$$

where L is a second-order differential operator defined as

$$L = \sum_{i=1}^{d} \mu_i(z)\frac{\partial}{\partial z_i} + \frac{1}{2}(\sigma_{11}^2(z)\frac{\partial^2}{\partial z_1^2} + \sigma_{22}^2(z)\frac{\partial^2}{\partial z_2^2} + ... + \sigma_{dd}^2(z)\frac{\partial^2}{\partial z_d^2}),$$

and

$$R = \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{u_1}\int_{(i-1)\Delta}^{u_2}\cdots\int_{(i-1)\Delta}^{u_s} E[L^{s+1}f(Z_{u_{s+1}})\big|\, Z_{(i-1)\Delta}]du_{s+1}du_s\cdots du_1$$

is a stochastic function of order Δ^{s+1}.

Lemma 4.2 ([22]). *Let*

$$\xi_{ni} = \theta((X_{(i-1)\Delta} - x)/h) + (1-\theta)((\tilde{X}_{(i-1)\Delta} - x)/h),\ 0 \le \theta \le 1,$$

and

$$\varepsilon_{1,n} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta} - x}{h}\right) g(\tilde{X}_{(i-1)\Delta}, \tilde{X}_{i\Delta}),$$

$$\varepsilon_{2,n} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) g(\tilde{X}_{(i-1)\Delta}, \tilde{X}_{i\Delta}),$$

where g is a measurable function on $\mathbb{R}\times\mathbb{R}$. Assume that Assumption 1, Assumption 4, and $\sqrt{\Delta}/h \to 0$. If one of the following two conditions holds,

(i) $E[(g(\tilde{X}_{(i-1)\Delta}, \tilde{X}_{i\Delta}))^2] < \infty$ *and Assumption 8 (ii).*

(ii) $h^{-1}E[|(\tilde{X}_{(i-1)\Delta} - X_{(i-1)\Delta})g(\tilde{X}_{(i-1)\Delta}, \tilde{X}_{i\Delta})|^2] < \infty$ *and Assumption 8 (i).*

Then

$$|\varepsilon_{1n} - \varepsilon_{2n}| \xrightarrow{P} 0.$$

Lemma 4.3. *Under Assumptions 1-7 and the conditions (i) of the Assumptions 8-12, for any random sequence $\{\eta_j\}_{j=1}^n$, if $\max\limits_{1\le j\le n} |\eta_j| = o_p(1)$, we have*

(i) $\sum\limits_{i=1}^{n} \psi_1'(u_{i\Delta} + \eta_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l = nh^{l+1}G_1(x_0)p(x_0)K_l(1 + o_p(1)),$

(ii) $\sum\limits_{i=1}^{n} \psi_1'(u_{i\Delta} + \eta_{i\Delta})R_1(X_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l = nh^{l+3}\frac{G_1(x_0)}{2}\mu''(x_0)p(x_0)K_{l+2}(1 + o_p(1)),$

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

Proof. (i) Obviously, we have

$$\begin{aligned}
&\sum_{i=1}^{n}\psi_1'(u_{i\Delta} + \eta_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l \\
&\quad= \sum_{i=1}^{n}\psi_1'(u_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l \\
&\qquad+ \sum_{i=1}^{n}[\psi_1'(u_{i\Delta} + \eta_{i\Delta}) - \psi_1'(u_{i\Delta})]K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l \\
&\quad:= T_{n1} + T_{n2}.
\end{aligned}$$

For T_{n1}, by Lemma 4.2, we need to consider

$$\begin{aligned}
&E\left[\sum_{i=1}^{n}\psi_1'(u_{i\Delta})K\left(\frac{X_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l\right] \\
&\quad= E\left[\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l E[\psi_1'(u_{i\Delta})\,|X_{(i-1)\Delta} = x_0]\right] \\
&\quad= G_1(x_0)E\left[\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta} - x_0}{h}\right)(\tilde{X}_{i\Delta} - x_0)^l\right].
\end{aligned}$$

Next, we will show that

$$E\left[\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)(\tilde{X}_{i\Delta}-x_0)^l\right]=nh^{l+1}p(x_0)K_l(1+o(1)).$$

In the same lines of arguments as in Lemma 1 of [5], we have

$$E\left[\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)(X_{(i-1)\Delta}-x_0)^l\right]=nh^{l+1}p(x_0)K_l(1+o(1)),$$

so it suffices to show that

$$\frac{1}{nh^{l+1}}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)(\tilde{X}_{i\Delta}-x_0)^l-\frac{1}{nh^{l+1}}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)(X_{(i-1)\Delta}-x_0)^l\xrightarrow{P}0. \tag{4.1}$$

For (4.1), let

$$\delta_{1n}=\frac{1}{nh^{l+1}}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)[(\tilde{X}_{i\Delta}-x_0)^l-(X_{(i-1)\Delta}-x_0)^l],$$

it suffices to show that

$$\lim_{n\to\infty}E(\delta_{1n})=0,\qquad \lim_{n\to\infty}Var(\delta_{1n})=0.$$

By the stationarity and Lemma 4.1, we have

$$\begin{aligned}
E(\delta_{1n})=&E\left[\frac{1}{nh^{l+1}}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)((\tilde{X}_{i\Delta}-x_0)^l-(X_{(i-1)\Delta}-x_0)^l)\right]\\
=&E\left[\frac{1}{h^{l+1}}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)((\tilde{X}_{i\Delta}-x_0)^l-(X_{(i-1)\Delta}-x_0)^l)\right]\\
=&E\left[\frac{1}{h^{l+1}}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l(X_{(i-1)\Delta}-x_0)^{l-1}(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})\right]+o(1)\\
=&E\left[\frac{1}{h^{l+1}}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l(X_{(i-1)\Delta}-x_0)^{l-1}E[(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})\,|X_{(i-1)\Delta}]\right]+o(1)\\
=&\frac{\Delta}{2h}E\left[\frac{1}{h}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)^{l-1}\mu(X_{(i-1)\Delta})\right]+O(\Delta^2),
\end{aligned}$$

which implies that $\lim_{n\to\infty}E(\delta_{1n})=0$.

On the other hand,

$$\begin{aligned}
Var(\delta_{1n})&=Var\left[\frac{1}{nh^{l+1}}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)((\tilde{X}_{i\Delta}-x_0)^l-(X_{(i-1)\Delta}-x_0)^l)\right]\\
&=\frac{1}{nh^2}Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)^{l-1}(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})\right]+o(1)\\
&=\frac{1}{nh^2}Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_{il}\right]+o(1),
\end{aligned}$$

where $f_{il}=\frac{1}{h}K\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)^{l-1}(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})$, and we have

$$Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_{il}\right]=\frac{1}{n}\sum_{i=1}^{n}Var(f_{il})+\frac{2}{n}\sum_{i=j+1}^{n}\sum_{j=1}^{n-1}Cov(f_{il},f_{jl}).$$

Next we will show that $Var\left[\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} f_{il}\right] < \infty$, with the same arguments as [23], we only need to show $E(f_{il}^2) < \infty$. In fact,

$$\begin{aligned} E(f_{il}^2) &= E\left(\frac{1}{h^2}K^2\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l^2\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)^{2l-2}\left(\tilde{X}_{i\Delta}-X_{(i-1)\Delta}\right)^2\right) \\ &= E\left(\frac{1}{h^2}K^2\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)l^2\left(\frac{X_{(i-1)\Delta}-x_0}{h}\right)^{2l-2}E[(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})^2\left|X_{(i-1)\Delta}\right.]\right), \end{aligned}$$

and

$$E[(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})^2\left|X_{(i-1)\Delta}\right.] = E[\tilde{X}_{i\Delta}^2\left|X_{(i-1)\Delta}\right.] - 2E[\tilde{X}_{i\Delta}X_{(i-1)\Delta}\left|X_{(i-1)\Delta}\right.] + E[X_{(i-1)\Delta}^2\left|X_{(i-1)\Delta}\right.].$$

By Lemma 4.1, for the first term, we have

$$E[\tilde{X}_{i\Delta}^2\left|X_{(i-1)\Delta}\right.] = E\left(\frac{(Y_{i\Delta}-Y_{(i-1)\Delta})^2}{\Delta^2}\left|X_{(i-1)\Delta}\right.\right) = X_{(i-1)\Delta}^2+\frac{1}{3}\sigma^2(X_{(i-1)\Delta})\Delta+X_{(i-1)\Delta}\mu(X_{(i-1)\Delta})\Delta+R_1,$$

where R_1 is a stochastic function of order Δ^2. For the second term, we have

$$E[\tilde{X}_{i\Delta}X_{(i-1)\Delta}\left|X_{(i-1)\Delta}\right.] = E\left(\frac{Y_{i\Delta}-Y_{(i-1)\Delta}}{\Delta}X_{(i-1)\Delta}\left|X_{(i-1)\Delta}\right.\right) = X_{(i-1)\Delta}^2+\frac{1}{2}X_{(i-1)\Delta}\mu(X_{(i-1)\Delta})\Delta+R_2,$$

where R_2 is a stochastic function of order Δ^2. For the third term, we have

$$E[X_{(i-1)\Delta}^2\left|X_{(i-1)\Delta}\right.] = X_{(i-1)\Delta}^2.$$

Then we have

$$E[(\tilde{X}_{i\Delta}-X_{(i-1)\Delta})^2\left|X_{(i-1)\Delta}\right.] = \frac{1}{3}\sigma^2(X_{(i-1)\Delta})\Delta+R,$$

where $R = R_1 - 2R_2$. So we have $E(f_{il}^2) < \infty$, and $\lim\limits_{n\to\infty} Var(\delta_{1n}) = 0$ is immediate from $nh^2 \to \infty$.

For T_{n2}, by Assumption 5 and Assumption 12(i), with the similar arguments as in Lemma 5.1 of [13], we can get $T_{n2} = o_p(nh^{l+1})$. This completes the proof of lemma.

(ii) This part can be proved by the same arguments with the first part of this lemma, so we omit the details. □

Lemma 4.4. *Under Assumptions 1-7 and the conditions (ii) of the Assumptions 8-12, for any random sequence* $\{\eta_j\}_{j=1}^n$, *if* $\max\limits_{1\le j\le n}|\eta_j| = o_p(1)$, *we have*

(i) $\sum\limits_{i=1}^{n}\psi_2'(v_{i\Delta}+\eta_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\tilde{X}_{i\Delta}-x_0)^l = nh^{l+1}H_1(x_0)p(x_0)K_l(1+o_p(1))$,

(ii) $\sum\limits_{i=1}^{n}\psi_2'(v_{i\Delta}+\eta_{i\Delta})R_2(X_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)(\tilde{X}_{i\Delta}-x_0)^l = nh^{l+3}\frac{H_1(x_0)}{2}p(x_0)(\sigma^2(x_0))''K_{l+2}(1+o_p(1))$, *where* $R_2(\tilde{X}_{i\Delta}) = \sigma^2(\tilde{X}_{i\Delta}) - \sigma^2(x_0) - (\sigma^2(x_0))'(\tilde{X}_{i\Delta}-x_0)$.

Proof. The proof of this lemma is similar to Lemma 4.3, so we omit the details. □

The proofs of the following two lemmas are similar to Theorem 1 in [5], so we omit the details.

Lemma 4.5. *Under Assumptions 1-7, the conditions (i) of the Assumptions 8-11 and Assumptions 13-15, we have*

$$\frac{1}{\sqrt{nh}}\begin{pmatrix} \sum_{i=1}^{n}\psi_1(u_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right) \\ \sum_{i=1}^{n}\psi_1(u_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\frac{\tilde{X}_{i\Delta}-x_0}{h} \end{pmatrix} \xrightarrow{D} N(0,\Sigma_3),$$

where $\Sigma_3 = G_2(x_0)p(x_0)V$.

Lemma 4.6. *Under Assumptions 1-7, the conditions (ii) of the Assumptions 8-11 and Assumptions 13-15, we have*

$$\frac{1}{\sqrt{nh}}\begin{pmatrix} \sum_{i=1}^{n}\psi_2(v_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right) \\ \sum_{i=1}^{n}\psi_2(v_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\frac{\tilde{X}_{i\Delta}-x_0}{h} \end{pmatrix} \xrightarrow{D} N(0,\Sigma_4),$$

where $\Sigma_4 = H_2(x_0)p(x_0)V$.

The proof of Theorem 2.6. (i) We prove the consistency of the local M-estimators of $\mu(x)$ and $\mu'(x)$. Let

$$r = (a_1, hb_1)^T, \quad r_0 = (\mu(x_0), h\mu'(x_0))^T, \quad r_{i\Delta} = (r-r_0)^T\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x_0}{h}\end{pmatrix},$$

and

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

Then we have

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

Let S_δ be the circle centered at r_0 with radius δ. We will show that for any sufficiently small δ,

$$\lim_{n\to\infty} P\{\inf_{r\in S_\delta} L_n(r) > L_n(r_0)\} = 1. \tag{4.2}$$

In fact, for $r \in S_\delta$, we have

$$L_n(r) - L_n(r_0) = \sum_{i=1}^{n}\rho_1\left(\frac{\tilde{X}_{(i+1)\Delta}-\tilde{X}_{i\Delta}}{\Delta} - a_1 - b_1(\tilde{X}_{i\Delta}-x_0)\right)K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)$$

$$
\begin{aligned}
&-\sum_{i=1}^{n}\rho_1\left(\frac{\tilde{X}_{(i+1)\Delta}-\tilde{X}_{i\Delta}}{\Delta}-\mu(x_0)-\mu'(x_0)(\tilde{X}_{i\Delta}-x_0)\right)K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\\
=&\sum_{i=1}^{n}K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)[\rho_1(u_{i\Delta}+R_1(\tilde{X}_{i\Delta})-r_{i\Delta})-\rho_1(u_{i\Delta}+R_1(\tilde{X}_{i\Delta}))]\\
=&\sum_{i=1}^{n}K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\int_{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})}^{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})-r_{i\Delta}}\psi_1(t)dt\\
=&\sum_{i=1}^{n}K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\int_{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})}^{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})-r_{i\Delta}}\psi_1(u_{i\Delta})dt\\
&+\sum_{i=1}^{n}K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\int_{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})}^{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})-r_{i\Delta}}\psi_1'(u_{i\Delta})(t-u_{i\Delta})dt\\
&+\sum_{i=1}^{n}K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\int_{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})}^{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})-r_{i\Delta}}[\psi_1(t)-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})(t-u_{i\Delta})]dt\\
:=&L_{n1}+L_{n2}+L_{n3}.
\end{aligned}
$$

Next, we will show that

$$L_{n1}=o_p(nh\delta), \tag{4.3}$$

$$L_{n2}=\frac{nh}{2}(r-r_0)^TG_1(x_0)p(x_0)U(1+o_p(1))(r-r_0)+O_p(nh^3\delta), \tag{4.4}$$

$$L_{n3}=o_p(nh\delta^2). \tag{4.5}$$

For (4.3), we have

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

where

$$W_n=\begin{pmatrix}\sum\limits_{i=1}^{n}\psi_1(u_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\\ \sum\limits_{i=1}^{n}\psi_1(u_{i\Delta})K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\frac{\tilde{X}_{i\Delta}-x_0}{h}\end{pmatrix}.$$

By Lemma 4.5, we have $E(W_n)=o(1)$, and

$$Var(W_n)=nhG_2(x)p(x_0)V(1+o(1)).$$

Note that

$$W_n=E(W_n)+O_p\left(\sqrt{Var(W_n)}\right),$$

so we have $W_n = O_p(\sqrt{nh})$, which implies that (4.3) holds.

For (4.4), we have

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

From Lemma 4.3 with $l = 0, l = 1$ and $l = 2$, respectively, we have

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

and

$$\begin{aligned}
L_{n22} &= -\sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\psi_1'(u_{i\Delta})R_1(\tilde{X}_{i\Delta})r_{i\Delta} \\
&= -(r-r_0)^T \sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\psi_1'(u_{i\Delta})R_1(\tilde{X}_{i\Delta}) \begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x_0}{h} \end{pmatrix} \\
&= -\frac{nh^3}{2}(r-r_0)^T G_1(x_0)\mu''(x_0)p(x_0) \begin{pmatrix} K_2 \\ K_3 \end{pmatrix} (1+o_p(1)) \\
&= O_p(nh^3\delta).
\end{aligned}$$

Therefore

$$L_{n2} = L_{n21} + L_{n22} = \frac{nh}{2}(r-r_0)^T G_1(x_0)p(x_0)U(1+o_p(1))(r-r_0) + O_p(nh^3\delta).$$

For (4.5), we have

$$\begin{aligned}
L_{n3} &= \sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right) \int_{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})}^{u_{i\Delta}+R_1(\tilde{X}_{i\Delta})-r_{i\Delta}} [\psi_1(t) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})(t-u_{i\Delta})]dt \\
&= \sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right) \int_{R_1(\tilde{X}_{i\Delta})}^{R_1(\tilde{X}_{i\Delta})-r_{i\Delta}} [\psi_1(t+u_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})t]dt
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)[\psi_1(z_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})z_{i\Delta}](-r_{i\Delta}) \\
&= -(r-r_0)^T \sum_{i=1}^{n} K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)[\psi_1(z_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})z_{i\Delta}]\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x_0}{h}\end{pmatrix},
\end{aligned}
$$

where the second-to-last equality follows from the integral mean value theorem and $z_{i\Delta}$ lies between $R_1(\tilde{X}_{i\Delta})$ and $R_1(\tilde{X}_{i\Delta})-r_{i\Delta}$, for $i=1,2,\cdots,n$. Since $\left|\tilde{X}_{i\Delta}-x_0\right|\le h$, we have

$$
\max_i |z_{i\Delta}| \le \max_i \left|R_1(\tilde{X}_{i\Delta})\right| + \left|(r-r_0)^T\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta}-x_0}{h}\end{pmatrix}\right| \le \max_i \left|R_1(\tilde{X}_{i\Delta})\right| + 2\delta, \tag{4.6}
$$

and by Taylor's expansion,

$$
\max_i \left|R_1(\tilde{X}_{i\Delta})\right| = \max_i \left|\mu(\tilde{X}_{i\Delta})-\mu(x_0)-\mu'(x_0)(\tilde{X}_{i\Delta}-x_0)\right| = \max_i \left|\frac{1}{2}\mu''(\xi_i)(\tilde{X}_{i\Delta}-x_0)^2\right| \le O_p(h^2), \tag{4.7}
$$

where ξ_i lies between $\tilde{X}_{i\Delta}$ and x_0, for $i=1,2,\cdots,n$.

For any given $\eta>0$, let $D_\eta=\{(\delta_{1\Delta},\delta_{2\Delta},\cdots,\delta_{n\Delta})^T : |\delta_{i\Delta}|\le\eta, \forall i\le n\}$, by Assumption 12(i) and $\left|\tilde{X}_{i\Delta}-x_0\right|\le h$, we have

$$
\begin{aligned}
&E\left[\sup_{D_\eta}\left|\sum_{i=1}^{n}[\psi_1(\delta_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})\delta_{i\Delta}]K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)(\tilde{X}_{i\Delta}-x_0)^l\right|\right] \\
&\le E\left[\sum_{i=1}^{n}\sup_{D_\eta}\left|\psi_1(\delta_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})\delta_{i\Delta}\right|K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\left|\tilde{X}_{i\Delta}-x_0\right|^l\right] \\
&\le a_\eta\delta E\left[\sum_{i=1}^{n}K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)\left|\tilde{X}_{i\Delta}-x_0\right|^l\right] \\
&\le b_\eta\delta,
\end{aligned}
$$

where a_η and b_η are two sequences of positive numbers, tending to zero as $\eta\to 0$. Therefore by (4.6) and (4.7), we have

$$
\sum_{i=1}^{n}[\psi_1(z_{i\Delta}+u_{i\Delta})-\psi_1(u_{i\Delta})-\psi_1'(u_{i\Delta})z_{i\Delta}]K\left(\frac{\tilde{X}_{(i-1)\Delta}-x_0}{h}\right)(\tilde{X}_{i\Delta}-x_0)^l = o_p(nh^{l+1}\delta),
$$

which implies that (4.5) holds.

Let λ be the smallest eigenvalue of the positive definite matrix U. Then, for any $r\in S_\delta$, we have

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

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

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

which implies that (4.2) holds. From (4.2), we know that $L_n(r)$ has a local minimum in the interior of S_δ, so there exists solutions to equation (2.3). Let $(\hat{\mu}(x_0), h\hat{\mu}'(x_0))^T$ be the closest solutions to $r_0 = (\mu(x_0), h\mu'(x_0))^T$, then

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

which implies the consistency of the local M-estimators of $\mu(x)$ and $\mu'(x)$.

(ii) We prove the asymptotic normality of the local M-estimators of $\mu(x)$ and $\mu'(x)$. Let

$$\hat{\eta}_{i\Delta} = R_1(\tilde{X}_{i\Delta}) - (\hat{\mu}(x_0) - \mu(x_0)) - (\hat{\mu}'(x_0) - \mu'(x_0))(\tilde{X}_{i\Delta} - x_0). \tag{4.8}$$

Then we have

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

Therefore by (2.3), we have

$$\sum_{i=1}^n \psi_1(u_{i\Delta} + \hat{\eta}_{i\Delta}) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta} - x}{h}\end{pmatrix} = 0. \tag{4.9}$$

Let

$$\begin{aligned} T_{n1} &= \sum_{i=1}^n \psi_1(u_{i\Delta}) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta} - x}{h}\end{pmatrix} = W_n,\\ T_{n2} &= \sum_{i=1}^n \psi_1'(u_{i\Delta})\hat{\eta}_{i\Delta} K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta} - x}{h}\end{pmatrix},\\ T_{n3} &= \sum_{i=1}^n [\psi_1(u_{i\Delta} + \hat{\eta}_{i\Delta}) - \psi_1(u_{i\Delta}) - \psi_1'(u_{i\Delta})\hat{\eta}_{i\Delta}] K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta} - x}{h}\end{pmatrix}.\end{aligned}$$

Then by (4.9), we have $T_{n1} + T_{n2} + T_{n3} = 0$. And by (4.8), we have

$$\begin{aligned} T_{n2} =& \sum_{i=1}^n \psi_1'(u_{i\Delta}) R_1(\tilde{X}_{i\Delta}) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta} - x}{h}\end{pmatrix}\\ &- \sum_{i=1}^n \psi_1'(u_{i\Delta}) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} (\hat{\mu}(x_0) - \mu(x_0)) + (\hat{\mu}'(x_0) - \mu'(x_0))(\tilde{X}_{i\Delta} - x_0) \\ \frac{\tilde{X}_{i\Delta} - x}{h}[(\hat{\mu}(x_0) - \mu(x_0)) + (\hat{\mu}'(x_0) - \mu'(x_0))(\tilde{X}_{i\Delta} - x_0)]\end{pmatrix}\\ =& \sum_{i=1}^n \psi_1'(u_{i\Delta}) R_1(\tilde{X}_{i\Delta}) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 \\ \frac{\tilde{X}_{i\Delta} - x}{h}\end{pmatrix}\\ &- \sum_{i=1}^n \psi_1'(u_{i\Delta}) K\left(\frac{\tilde{X}_{(i-1)\Delta} - x_0}{h}\right)\begin{pmatrix} 1 & \frac{\tilde{X}_{i\Delta} - x}{h} \\ \frac{\tilde{X}_{i\Delta} - x}{h} & \frac{(\tilde{X}_{i\Delta} - x)^2}{h^2}\end{pmatrix}\begin{pmatrix} \hat{\mu}(x_0) - \mu(x_0) \\ h(\hat{\mu}'(x_0) - \mu'(x_0))\end{pmatrix}\\ =& \frac{nh^3}{2} G_1(x_0)\mu''(x_0)p(x_0)\begin{pmatrix} K_2 \\ K_3\end{pmatrix}(1 + o_p(1))\end{aligned}$$

$$
\begin{aligned}
& - nhG_1(x_0)p(x_0)\begin{pmatrix} K_0 & K_1 \\ K_1 & K_2 \end{pmatrix}(1+o_p(1))\begin{pmatrix} \hat{\mu}(x_0)-\mu(x_0) \\ h(\hat{\mu}'(x_0)-\mu'(x_0)) \end{pmatrix} \\
=& \frac{nh^3G_1(x_0)\mu''(x_0)p(x_0)}{2}A(1+o_p(1)) - nhG_1(x_0)p(x_0)U(1+o_p(1))\begin{pmatrix} \hat{\mu}(x_0)-\mu(x_0) \\ h(\hat{\mu}'(x_0)-\mu'(x_0)) \end{pmatrix} \\
:=& T_{n21}+T_{n22},
\end{aligned}
$$

where the third equality follows from Lemma 4.3.

Noting that for $\left|\tilde{X}_{i\Delta}-x_0\right| \leq h$, we have

$$
\begin{aligned}
\sup_i |\hat{\eta}_{i\Delta}| &= \sup_i \left| R_1(\tilde{X}_{i\Delta}) - (\hat{\mu}(x_0)-\mu(x_0)) - (\hat{\mu}'(x_0)-\mu'(x_0))(\tilde{X}_{i\Delta}-x_0)\right| \\
&\leq \sup_i \left|R_1(\tilde{X}_{i\Delta})\right| + |\hat{\mu}(x_0)-\mu(x_0)| + h\left|\hat{\mu}'(x_0)-\mu'(x_0)\right| \\
&= O_p(h^2 + (\hat{\mu}(x_0)-\mu(x_0)) + h(\hat{\mu}'(x_0)-\mu'(x_0))) \\
&= o_p(1),
\end{aligned}
$$

where the last equality follows from the consistency of $(\hat{\mu}(x_0), h\hat{\mu}'(x_0))$. Then, by the Assumption 12(i) and the same argument as that in the first part of Theorem 2.6, we have

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

Therefore, by $T_{n1}+T_{n2}+T_{n3}=0$, we have

$$
\begin{pmatrix} \hat{\mu}(x_0)-\mu(x_0) \\ h(\hat{\mu}'(x_0)-\mu'(x_0)) \end{pmatrix} = \frac{1}{nh}G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}(1+o_p(1))W_n + \frac{h^2}{2}\mu''(x_0)U^{-1}A(1+o_p(1)).
$$

It follows that

$$
\sqrt{nh}\left[\begin{pmatrix} \hat{\mu}(x_0)-\mu(x_0) \\ h(\hat{\mu}'(x_0)-\mu'(x_0)) \end{pmatrix} - \frac{h^2\mu''(x_0)}{2}U^{-1}A(1+o_p(1))\right] = G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}(1+o_p(1))\frac{1}{\sqrt{nh}}W_n.
$$

So by Lemma 4.5, Assumption 5 and the Slutsky's theorem, we have

$$
\begin{aligned}
\sqrt{nh}\left[\begin{pmatrix} \hat{\mu}(x_0)-\mu(x_0) \\ h(\hat{\mu}'(x_0)-\mu'(x_0)) \end{pmatrix} - \tfrac{h^2\mu''(x_0)}{2}U^{-1}A\right] &\xrightarrow{D} G_1^{-1}(x_0)p^{-1}(x_0)U^{-1}N(0,\Sigma_3) \\
&= N\left(0, \tfrac{G_2(x_0)}{G_1^2(x_0)p(x_0)}U^{-1}VU^{-1}\right) = N(0,\Sigma_1).
\end{aligned}
$$

This completes the proof. □

The proof of Theorem 2.7. The proof follows from Theorem 2.6 using Lemmas 4.4 and 4.6, and is omitted here. □

Acknowledgements

This research is supported by the NSFC (No.11461032 and No.11401267) and the SF of Jiangxi Provincial Education Department (No. GJJ150687).

References

[1] L. Arnold, *Stochastic differential equations: Theory and Application*, Wiley, New York, (1974). 2.1

[2] F. Bandi, P. Phillips, *Fully nonparametric estimation of scalar diffusion models*, Econometrica, **71** (2003), 241–283. 1

[3] B. M. Bibby, M. Jacobsen, M. Sørensen, *Estimating functions for discretely sampled diffusion-type models*, In: Handbook of Financial Econometrics, Amsterdam: North-Holland, (2002). 1

[4] B. M. Bibby, M. Sørensen, *Martingale estimation functions for discretely observed diffusion processes*, Bernoulli., **1** (1995), 17–39. 1

[5] Z. Cai, E. Ould-Said, *Local M-estimator for nonparametric time series*, Statist. Probab. Lett., **65** (2003), 433–449. 2.4, 4, 4

[6] F. Comte, V. Genon-Catalot, Y. Rozenholc, *Penalized nonparametric mean square estimation of coefficients of diffusion processes*, Bernoulli., **13** (2007), 514–543. 1

[7] F. Comte, V. Genon-Catalot, Y. Rozenholc, *Nonparametric adaptive estimation for integrated diffusions*, Stoch. Proc. Appl., **119** (2009), 811–834. 2.1

[8] R. A. Davis, K. Knight, J. Liuc, *M-estimation for autoregressions with infinite variance*, Stoch. Proc. Appl., **40** (1992), 145–180. 1

[9] M. Delecroix, M. Hristache, V. Patilea, *On semiparametric M-estimation in single-index regression*, J. Statist. Plann. Inference, **136** (2006), 730–769. 1

[10] P. D. Ditlevsen, S. Ditlevsen, K. K. Andersen, *The fast climate fluctuations during the stadial and interstadial climate states*, Ann. Glaciol., **35** (2002), 457–462. 1

[11] S. Ditlevsen, M. Sørensen, *Inference for observations of integrated diffusion processes*, Scand. J. Statist., **31** (2004), 417–429. 1

[12] J. Fan, I. Gijbels, *Local polynomial modeling and its applications*, Chapman & Hall, London, (1996). 2.2

[13] J. Fan, J. Jiang, *Variable bandwidth and one-step local M-estimator*, Sci. China Ser. A, **43** (2000), 65–81. 1, 2.4, 4

[14] J. Fan, C. Zhang, *A re-examination of diffusion estimations with applications to financial model validation*, J. Amer. Statist. Assoc., **98** (2003), 118–134. 1

[15] A. Gloter, *Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient*, ESAIM: Probability Statist., **4** (2000), 205–227. 1

[16] A. Gloter, *Parameter estimation for a discretely observed integrated diffusion process*, Scand. J. Statist., **33** (2006), 83–104. 1

[17] A. Gloter, E. Gobet, *LAMN property for hidden processes: The case of integrated diffusions*, Ann. de l'Institut Henri Poincaré, **44** (2008), 104–128. 1

[18] E. Gobet, M. Hoffmann, M. Reiss, *Nonparametric estimation of scalar diffusions based on low frequency data*, Ann. Statist., **32** (2004), 2223–2253. 1

[19] P. Hall, M. C. Jones, *Adaptive M-estimation in nonparametric regression*, Ann. Statist., **18** (1990), 1712–1728. 1

[20] P. J. Huber, *Robust regression*, Ann. Statist., **1** (1973), 799–821. 1

[21] J. C. Jiang, Y. P. Mack, *Robust local polynomial regression for dependent data*, Statist. Sinica, **11** (2001), 705–722. 1

[22] J. Nicolau, *Nonparametric estimation of scend-order stochastic differential equations*, Economet. Theory, **23** (2007), 880–898. 1, 2.1, 2.1, 2.2, 3, 4.1, 4.2

[23] L. C. G. Rogers, D. Williams, *Diffusions, Markov processes, and Martingales*, Vol 2, Cambridge University Press, Cambridge, (2000). 1, 4

[24] H. C. Wang, Z. Y. Lin, *Local linear estimation of second-order diffusion models*, Commun. Statist., **40** (2011), 394–407. 1

[25] Y. Y. Wang, L. X. Zhang, M. T. Tang, *Re-weighted functional estimation of second-order diffusion processes*, Metrika., **75** (2012), 1129–1151. 1, 2.1