Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Nonparametric Regression for Dependent Data in the Errors-in-Variables Problem

Toshio Honda

Hi-Stat Discussion Paper

Nonparametric regression for dependent data in the errors-in-variables problem

Toshio Honda*

Graduate School of Economics, Hitotsubashi University, Tokyo 186-8601, JAPAN

Abstract

We consider the nonparametric estimation of the regression functions for dependent data. Suppose that the covariates are observed with additive errors in the data and we employ nonparametric deconvolution kernel techniques to estimate the regression functions in this paper. We investigate how the strength of time dependence affects the asymptotic properties of the local constant and linear estimators. We treat both short-range dependent and long-range dependent linear processes in a unified way and demonstrate that the long-range dependence (LRD) of the covariates affects the asymptotic properties of the nonparametric estimators as well as the LRD of regression errors does.

Key words: local polynomial regression, errors-in-variables, deconvolution, ordinary smooth case, supersmooth case, linear processes, long-range dependence

1. Introduction

When we have dependent data, (Y_i, X_i), $i = 1, \ldots, n$, and need to know the regression function of the response Y_i on the covariate X_i with no parametric assumption, we usually appeal to nonparametric regression techniques such as local polynomial regression. See Fan and Gijbels (1996) for more on nonparametric regression. In this paper, we consider the case where the covariate X_i is observed with an additive measurement error and we estimate

---

*Corresponding author
Email address: honda@econ.hit-u.ac.jp (Toshio Honda)

the regression function of Y_i on X_i by using deconvolution kernel techniques. This is one of errors-in-variables problems.

The errors-in-variables problems have been receiving a lot of attention for decades. For example, see Carroll et al. (2006) for general results and applications and Meister (2009) for results on nonparametric estimation including deconvolution kernel density estimation. When we carry out nonparametric estimation of regression functions, one of the most familiar estimators has been the local constant type estimator of Fan and Truong (1993). The estimator is based on the idea of the deconvolution kernel density estimator and is very recently extended to the higher order versions in Delaigle et al. (2009). There are some other kinds of nonparametric regression estimators such as those of Schennach (2004), Comte and Taupin (2007), and Hu and Schennach (2008). Hereafter we concentrate on deconvolution kernel estimators of Fan and Truong (1993) and Delaigle et al. (2009).

Most of the existing results on nonparametric estimation in the errors-in-variables problems are focused on the cases of i.i.d. or short-range dependent observations. There are too many papers to mention in the cases of i.i.d. observations and we just refer to Carroll et al. (2006), Meister (2009) and the references therein. Masry (1991,1993a,1993b,2003) and Fan and Masry (1992) deal with the cases of short-range dependent observations by assuming some mixing conditions or positive association. Those papers contain the results on deconvolution kernel density estimation and deconvolution kernel nonparametric regression estimation. van Zanten and Zareba (2008) is a recent paper on wavelet deconvolution density estimation for short-range dependent observations. The settings of those papers are different from that of the present paper since those papers assume mixing conditions or similar ones and the assumptions of them exclude the cases of slowly decaying autocovariances.

It is well known that we have the same asymptotics for short-range dependent observations as for i.i.d observations when we employ nonparametric kernel estimation without measurement error. See Fan and Yao (2003) for more details on this phenomenon. This is also true of the errors-in-variables problems. The asymptotic properties given in Masry (1991,1993a,1993b,2003) and Fan and Masry (1992) are the same as in the i.i.d. situations. However, things change when dependent data exhibit LRD stronger than a level in density estimation and nonparametric regression. When the LRD exceeds the level determined by the bandwidth, the effect of the LRD becomes dominant and we have the asymptotics completely different from those in the

i.i.d. situations. For example, see Masry and Mielniczuk (1999), Wu and Mielniczuk (2002), and Mielniczuk and Wu (2004) for those results in the cases of no measurement error. In nonparametric regression, only the LRD of the regression error terms can affect the asymptotic properties when we observe the covariates and have the same data generating process as in this paper. It is easy to see this from the arguments in Mielniczuk and Wu (2004).

This transition of the asymptotic properties also happens when we employ deconvolution kernel density estimators in the errors-in-variables problems. See Kulik (2008) for this phenomenon in density estimation. Kulik (2008) has examined how the LRD affects the asymptotic properties of deconvolution kernel density estimators by following Wu and Mielniczuk (2002). It is shown that the effect of the LRD appears when the strength of LRD exceeds a level as in Wu and Mielniczuk (2002). However, no result has been obtained for nonparametric regression under LRD in the errors-in-variables problems.

In this paper, we examine how the LRD of (Y_i, X_i) affects asymptotic properties of the local constant and linear estimators of Fan and Truong (1993) and Delaigle et al. (2009). We also follow Wu and Mielniczuk (2002) and use the results of Giraitis et al. (1996) as in Kulik (2008). However, there are two sources of LRD in nonparametric regression, covariates and regression error terms, and we have to deal with unbounded functions. Thus the asymptotics and mathematical treatment are more complicated than in density estimation.

In the setting of this paper, we have found a similar transition of the asymptotic properties. However, the LRD of covariates can affect the asymptotic properties of the nonparametric estimators as the LRD of regression error terms can. Note again that the LRD of covariates does not affect the asymptotic properties in the cases of no measurement error under the data generating process (DGP) of this paper.

We assume that (Y_i, X_i), $i = 1, \ldots, n$, are generated by linear processes and we treat short-range dependent and long-range dependent linear processes in a unified way. We call $\{X_i\}$ a short-range dependent process when $\sum_{i=1}^{\infty} \mathrm{Cov}(X_1, X_i) < \infty$. When $\sum_{i=1}^{\infty} \mathrm{Cov}(X_1, X_i) = \infty$, $\{X_i\}$ is a long-range dependent process. See Beran (1994) for more details on short-range dependence (SRD) and LRD. Note that SRD and weak dependence are often used for the same meaning in the literature.

We could deal with general local polynomial estimators. However, we deal with only the local constant and linear estimators for simplicity of presentation. We follow a series of papers on deconvolution kernel techniques

and treat the ordinary smooth case and the supersmooth case separately. The asymptotic normality is established in the ordinary smooth case. In the supersmooth case, we investigate the asymptotic variance of the local constant estimator and show how the effect of LRD appears in the variance. The investigation of the asymptotic distribution is a subject of future research.

The paper is organized as follows. We define the estimators and the DGP in section 2. A brief exposition on long-range dependent linear processes is also given. We present the results on the ordinary smooth case and the supersmooth case in sections 3 and 4, respectively. The proofs of the propositions in sections 3 and 4 are postponed to section 5. The technical lemmas and the proofs are in section 6.

In this paper C, C_l, δ, and δ_l are generic positive constants and their values change from place to place. We write a^T for the transpose of a vector a.

2. Estimators and data generating process

In this section we describe the DGP and define the local constant and linear estimators.

Let the dependent data (Y_i, X_i), $i = 1, \ldots, n$, be generated by

$$Y_i = m(X_i) + \sigma(X_i)\eta_i, \quad i = 1, \ldots, n, \tag{1}$$

where $m(x)$, $x \in R$, is an unknown function and $\sigma(X_i)\eta_i$ is the regression error term. We do not observe the covariate X_i and we have n observations of (Y_i, W_i), where $W_i = X_i + U_i$ and U_i, $i = 1, \ldots, n$, are i.i.d. and we assume that $\{U_i\}$, $\{X_i\}$ and $\{\eta_i\}$ are mutually independent. We denote the density function and the characteristic function of U_i by $f_U(u)$ and $\phi_U(t)$, respectively and we assume that $\phi_U(t)$ is known.

We estimate $m(x_0)$ for a fixed x_0 from (Y_i, W_i), $i = 1, \ldots, n$ with no parametric assumption on $m(x)$. The asymptotic properties of the estimators crucially depend on the tail behavior of $\phi_U(t)$ and we consider the ordinary smooth case and the supersmooth case as in the other papers in this field.

It would be almost impossible to investigate the asymptotic properties of the estimators without any specific assumptions on the DGP of $\{X_i\}$, and $\{\eta_i\}$ when the time dependence is strong. Hence we assume that $\{X_i\}$ and $\{\eta_i\}$ are linear processes defined in (2) below.

$$\eta_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j} \quad \text{and} \quad X_i = \sum_{j=0}^{\infty} b_j \xi_{i-j} \tag{2}$$

where $\{\epsilon_j\}$ and $\{\xi_j\}$ are sequences of i.i.d. random variables with zero mean, $a_0 = 1$, $\mathrm{E}(\eta_j^2) = \mathrm{E}(\epsilon_1^2)\sum_{j=0}^{\infty} a_j^2 = 1$, $\mathrm{E}(\xi_j^2) = \sigma_\xi^2$, $b_0 = 1$, and $\sum_{j=0}^{\infty} b_j^2 < \infty$. Note that that $\{U_j\}$, $\{\epsilon_j\}$, and $\{\xi_j\}$ are mutually independent. Notice that existing results mentioned in section 1 do not cover the cases where the coefficients $\{a_j\}$ and $\{b_j\}$ decay not very fast.

We define the local constant and linear estimators of $m(x_0)$, $\hat{m}_c(x_0)$ and $\hat{m}_l(x_0)$, by following Delaigle et al. (2009). First we choose a symmetric kernel function $K(v)$ satisfying $\int K(v)dv = 1$ and put

$$\phi_K(t) = \int \exp(itv)K(v)dv \quad \text{and} \quad K_h(x) = K\left(\frac{x}{h}\right),$$

where i in $\exp(itv)$ is the imaginary unit and h is the bandwidth converging to 0. The range of integration is the whole real line when it is omitted.

Some definitions are necessary to define $\hat{m}_c(x_0)$ and $\hat{m}_l(x_0)$. We define $\hat{T}_{n,k}(x_0)$, $k = 0, 1$, and $\hat{S}_{n,k}(x_0)$, $k = 0, 1, 2$, by

$$\hat{T}_{n,k}(x_0) = n^{-1}\sum_{j=1}^{n} Y_j \left(\frac{W_j - x_0}{h}\right)^k L_{k,h}(W_j - x_0), \tag{3}$$

$$\hat{S}_{n,k}(x_0) = n^{-1}\sum_{j=1}^{n} \left(\frac{W_j - x_0}{h}\right)^k L_{k,h}(W_j - x_0), \tag{4}$$

where

$$\left(\frac{w - x_0}{h}\right)^k L_{k,h}(w - x_0) = K_{U,k}((w - x_0)/h),$$

$$K_{U,k}(v) = \frac{1}{2i^k\pi}\int \exp(-itv)\frac{\phi_K^{(k)}(v)}{\phi_U(-t/h)}dt.$$

The above definition of $K_{U,k}$ is motivated by the equation

$$\mathrm{E}\left\{\left(\frac{W_j - x_0}{h}\right)^k L_{k,h}(W_j - x_0)\Big|X_j\right\} = \left(\frac{X_j - x_0}{h}\right)^k K_h(X_j - x_0). \tag{5}$$

Then put
$$\hat{T}_n = (\hat{T}_{n,0}(x_0), \hat{T}_{n,1}(x_0))^T$$

and let $\hat{S}_n$ be a 2×2 matrix whose (i, j) element is $\hat{S}_{n,i+j-2}(x_0)$.

The local linear estimator $\hat{m}_l(x_0)$ is defined by

$$\hat{m}_l(x_0) = (1, 0)\hat{S}_n^{-1}\hat{T}_n. \tag{6}$$

5

The local constant estimator $\hat{m}_c(x_0)$ is given by

$$\hat{m}_c(x_0) = \hat{S}_n^{-1}\hat{T}_n \tag{7}$$

with $\hat{T}_n = \hat{T}_{n,0}(x_0)$ and $\hat{S}_n = \hat{S}_{n,0}(x_0)$.

Next we are more specific about $\{\eta_j\}$ and $\{X_j\}$ and give some important properties of the autocovariances. In this paper, we consider the following two cases for $\{\eta_j\}$.

$$\sum_{j=0}^{\infty} |a_j| < \infty. \tag{8}$$

$$\sum_{j=0}^{\infty} |a_j| = \infty \quad \text{and} \quad a_j = (j+1)^{-(1+\gamma_a)/2}l_a(j), \tag{9}$$

where $0 < \gamma_a \le 1$ and $l_a(j)$ is a slowly varying function. It is easy to see $\{\eta_j\}$ is a short-range dependent linear process in the case of (8) and that $\{\eta_j\}$ is a long-range dependent linear process in the case of (9).

Put $\sigma_\eta(i-j) = \mathrm{E}(\eta_i\eta_j)$ and then we have in the case of (8),

$$\mathrm{Var}(\sum_{j=1}^{n} \eta_j) = \sum_{1 \le i,j \le n} \sigma_\eta(i-j) = O(n). \tag{10}$$

In the case of (9), we have

$$\mathrm{Var}(\sum_{j=1}^{n} \eta_j) = \sum_{1 \le i,j \le n} \sigma_\eta(i-j) = \begin{cases} O(n|\sum_{j=1}^{n} l_a(j)/j|^2), & \gamma_a = 1, \\ O(n^{2-\gamma_a}l_a^2(n)), & 0 < \gamma_a < 1. \end{cases} \tag{11}$$

Especially when $0 < \gamma_a < 1$,

$$\lim_{n\to\infty} (n^{2-\gamma_a}l_a^2(n))^{-1} \sum_{1 \le i,j \le n} \sigma_\eta(i-j) = D(\gamma_a)\mathrm{E}(\epsilon_1^2), \tag{12}$$

where

$$D(\gamma) = \{(1-\gamma)(1-\gamma/2)\}^{-1} \int_0^\infty (x+x^2)^{-(1+\gamma)/2}dx.$$

We also consider the following two cases for $\{X_j\}$.

$$\sum_{j=0}^{\infty} |b_j| < \infty. \tag{13}$$

6

$$\sum_{j=0}^{\infty} |b_j| = \infty \quad \text{and} \quad b_j = (j+1)^{-(1+\gamma_b)/2} l_b(j), \tag{14}$$

where $0 < \gamma_b \le 1$ and $l_b(j)$ is a slowly varying function. By setting $\sigma_X(i-j) = \mathrm{E}(X_i X_j)$ for $\{X_j\}$, we have the results similar to (10)-(12). When we do not specify (8), (9), (13), and (14) in the propositions and lemmas, we mean that the propositions and lemmas are true in any cases.

We examine the asymptotic properties of the estimators by employing the martingale decomposition approach as in Ho and Hsing (1996, 1997) and the technique of Wu and Mielniczuk (2002). Finally in this section, we define several σ-fields which we use in applying the approach and technique to our setting.

$$
\begin{aligned}
\mathcal{F}_j &= \sigma\{\xi_j, \xi_{j-1}, \ldots\}, \\
\mathcal{H}_j &= \sigma\{U_j, \xi_j, U_{j-1}, \xi_{j-1}, \ldots\}, \\
\bar{\mathcal{H}}_j &= \sigma\{\{U_j, \xi_j, U_{j-1}, \xi_{j-1}, \ldots\} \cup \{\epsilon_i\}_{i=-\infty}^{\infty}\},
\end{aligned}
$$

where $\sigma\{\cdots\}$ means σ-field generated by the random variables inside the braces.

3. Ordinary smooth case

In this section, we consider the ordinary smooth case where $\phi_U(t)$ is known and satisfies Assumption CU below. The Gamma and Laplace distributions and the convolutions of those distributions satisfy Assumption CU. First we state assumptions and the main theorems. Then we describe the propositions and the proofs of the main theorems. As for the bandwidth, we take $h = c_1 n^{-\alpha}$ $(0 < \alpha < 1)$ in this section.

Assumption CU
(i) $|\phi_U(t)| > 0$ for all t.
(ii) $\phi_U(t)$ is twice continuously differentiable.
(iii) There is a β larger than 1 such that

$$\lim_{t \to \pm\infty} |t|^{\beta} \phi_U(t), \quad \lim_{t \to \pm\infty} |t|^{\beta+1} \phi_U'(t), \quad \text{and} \quad \lim_{t \to \pm\infty} |t|^{\beta+2} \phi_U''(t)$$

exist and are finite. Especially $\lim_{t \to \pm\infty} |t|^{\beta} \phi_U(t) \ne 0$.

Assumption CK below is about the kernel function $K(v)$ and just a standard one in the literature on deconvolution kernel techniques.

Assumption CK
(i) $\phi_K(t)$ is four times continuously differentiable and $\phi_K(t)$ and all the derivative functions are bounded.
(ii) For $k = 0, 1, 2$,

$$\int (|t|^\beta |\phi_K^{(k)}(t)| + |t|^{\beta-1} |\phi_K^{(k+1)}(t)| + |t|^\beta |\phi_K^{(k+2)}(t)|)dt < \infty.$$

When Assumptions CU and CK hold, Lemma 3 of Masry (1991) and Lemma A of Kulik (2008) imply that

$$h^\beta |K_{U,k}(v)| \leq C\Big(1 \wedge \frac{1}{|v|^2}\Big). \tag{15}$$

Hence we can define Δ_k for $k = 0, 1, 2$ by

$$\Delta_k = \lim_{n \to \infty} h^\beta \int K_{U,k}(v)dv = \int \tilde{L}_k(v)dv, \tag{16}$$

where $\tilde{L}_k(v) = \lim_{n \to \infty} h^\beta K_{U,k}(v)$. The existence of $\tilde{L}_k(v)$ is guaranteed by Assumption CU. Next we define $\Omega_{k+1,l+1}$ for $k = 0, 1$ and $l = 0, 1$ by

$$\Omega_{k+1,l+1} = \lim_{n \to \infty} h^{2\beta} \int K_{U,k}(v)K_{U,l}(v)dv = \int \tilde{L}_k(v)\tilde{L}_l(v)dv. \tag{17}$$

The definitions of Δ_k and $\Omega_{k+1,l+1}$ in (16) and (17) are justified by inequality (15) and the dominated convergence theorem.

Put $\Omega = \Omega_{11}$ for $\hat{m}_c(x_0)$ and

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

for $\hat{m}_l(x_0)$. We need another assumption on the kernel function $K(v)$ and some notation. The assumption is necessary to evaluate the bias term of the estimators.

Assumption MK
$\int |v|^3 |K(v)|dv < \infty$. Write $\mu_j = \int v^j K(v)dv$, for $j = 0, 1, 2, 3$. Note that $\mu_0 = 1$, $\mu_1 = 0$, and $\mu_3 = 0$.

The former of Assumption MM below is about the moment of X_i. Let γ_m be positive for technical reasons even if $m(x)$ and $\sigma(x)$ is bounded functions. The latter is used to evaluate the bias term of the estimators.

8

Assumption MM

(i) $|m(x)| \leq C(1 + |x|)^{\gamma_m}$ and $|\sigma(x)| \leq C(1 + |x|)^{\gamma_m}$ for some positive γ_m such that $0 < \gamma_m \leq 3$.

(ii) $m(x)$ is twice continuously differentiable and

$$\sup_{v} \sup_{|\theta| \leq 1} |m''(x_0 + \theta v) f_X(x_0 + v)| < C,$$

where $f_X(x)$ is the density of X_i and it exists due to Assumption X below.

Assumption X

(i) $\mathrm{E}(|\xi_1|^{2\gamma_m + 2}) < \infty$

(ii) Write $\phi_\xi(t)$ for the characteristic function of ξ_1. Then there is a positive δ such that $|\phi_\xi(t)| \leq C(1 + |t|)^{-\delta}$.

We can derive the two facts D1 and D2 below from Assumption X by following the arguments of Lemma 2 of Giraitis et al. (1996) and Lemma 5.1 of Koul and Surgailis (2002).

Fact D1: There is an integer l_0 such that the density function of $\sum_{j=0}^{l-1} b_j \xi_{i-j}$ exist for any $l \geq l_0$. We write $f_l(x)$ and $f_X(x)$ for the density functions of $\sum_{j=0}^{l-1} b_j \xi_{i-j}$ and X_1, respectively.

Fact D2: When $s = 0, 1, 2$ and $l \geq l_0$, we have

$$|f_X^{(s)}(x)| + |f_l^{(s)}(x)| \leq \frac{C}{(1 + |x|)^{\gamma_1}} \text{ and } |f_l^{(s)}(x) - f_{l+1}^{(s)}(x)| \leq \frac{Cb_l^2}{(1 + |x|)^{\gamma_1}},$$

where γ_1 is the integer part of $2\gamma_m + 2$.

Hereafter in this paper we take $l_0 = 1$ to simplify our presentation and write $f_\xi(x)$ for $f_1(x)$. We can derive the same results as Theorems 1 and 2 below by appealing to a complicated and lengthy argument in as Honda (2009) when l_0 is larger than 1.

Put

$$X_{j,l} = \sum_{j=l}^{\infty} b_j \xi_{i-j}.$$

The random variable $X_{j,1}$ plays an important role in the proofs of Theorems 1 and 2 since $l_0 = 1$. In the case where $\{X_i\}$ is a long-range dependent linear process, it is easy to see that

$$\lim_{n \to \infty} \frac{\sum_{1 \leq i,j \leq n} \mathrm{E}(X_i X_j)}{\sum_{1 \leq i,j \leq n} \mathrm{E}(X_{i,1} X_{j,1})} = 1.$$

9

When $\{X_i\}$ is short-range dependent, $\{X_{i,1}\}$ is also short-range dependent. Therefore we do not have to discriminate between $\mathrm{E}(X_i X_j)$ and $\mathrm{E}(X_{i,1} X_{j,1})$ and we use $\sigma_X(i-j)$ to denote $\mathrm{E}(X_{i,1} X_{j,1})$ in the proofs of Theorems 1 and 2, Propositions 1-6, and the proofs.

Here we state Theorems 1 and 2 and related remarks. Theorem 1 deals with the case where both $\{\eta_i\}$ and $\{X_i\}$ are short-range dependent linear processes. Let $\xrightarrow{d}$ and $\xrightarrow{p}$ denote convergence in distribution and probability, respectively and $(g_1 * g_2)(x)$ stands for the convolution of $g_1(x)$ and $g_2(x)$.

Theorem 1. *Suppose that Assumptions CU, CK, MK, MM, and X, (8), and (13) hold. Besides we assume that $\alpha(\beta+1/2) < 1/2$ for the bandwidth. Then for $j = c$ and l, we have*

$$h^\beta \sqrt{nh}(\hat{m}_j(x_0) - m(x_0) - Bias) \xrightarrow{d} N(0, \Omega_{11}(f_U * (\tau^2 f_X))(x_0)/f_X^2(x_0)),$$

where $\tau^2(x) = \mathrm{E}\{(Y_i - m(x_0))^2 | X_i = x\}$ and Ω_{11} is defined in (17). As for the bias term Bias, we have

$$Bias = \begin{cases} h^2(m'(x_0)f_X'(x_0) + m''(x_0)f_X(x_0)/2)\mu_2/f_X(x_0) + o(h^2), & j = c, \\ h^2 m''(x_0)\mu_2/2 + o(h^2), & j = l. \end{cases}$$

We have the same asymptotic properties for i.i.d. observations in Theorem 1 and the optimal rate of convergence is achieved when $\alpha = 1/(2\beta + 5)$.

Theorem 2 below deals with the cases where either of $\{\eta_i\}$ or $\{X_i\}$ has LRD. We only deal with typical three cases (a)-(c) to simplify the presentation. When we observe X_i without measurement error, the effect of the LRD of the covariate does not appear and we have only to consider two cases. The asymptotic distribution is independent of h when we observe X_i without measurement error and the effect of the LRD appears. There are also only two cases in density estimation.

Theorem 2. *Suppose that Assumptions CU, CK, MK, MM, and X hold and that either of (9) or (14) holds. Then the nonparametric regression estimators $\hat{m}_c(x_0)$ and $\hat{m}_l(x_0)$ have the following asymptotic properties. The bias term is the same as in Theorem 1. Recall that Δ_0 is defined in (16)*
(a) When $\sum_{1 \le i,j \le n} \sigma_X(i-j) / \sum_{1 \le i,j \le n} \sigma_\eta(i-j) \to 0$, $(nh)^{1/2} n^{-\gamma_a/2}|l_a(n)| \to \infty$, and $2\alpha\beta < \gamma_a$, we have for $j = c$ and l,

$$h^\beta n^{\gamma_a/2}(l_a^2(n)D(\gamma_a)\mathrm{E}(\epsilon_1^2))^{-1/2}(\hat{m}_j(x_0) - m(x_0) - Bias)$$
$$\xrightarrow{d} N(0, \Delta_0^2(f_U * (\sigma f_X))^2(x_0)/f_X^2(x_0)).$$

10

(b) When $\sum_{1\le i,j\le n}\sigma_X(i-j)/\sum_{1\le i,j\le n}\sigma_\eta(i-j)\to\infty$, $(nh)^{1/2}n^{-\gamma_b/2}|l_b(n)|\to$ ∞, and $2\alpha\beta<\gamma_b$, we have for $j=c$ and l,

$$h^\beta n^{\gamma_b/2}(l_b^2(n)D(\gamma_b)\mathrm{E}(\xi_1^2))^{-1/2}(\hat{m}_j(x_0)-m(x_0)-Bias)$$
$$\xrightarrow{d}\ N(0,\Delta_0^2(f_U*((m-m(x_0))f_X')^2(x_0)/f_X^2(x_0)).$$

(c) When $nh\max\{n^{-2}\sum_{1\le i,j\le n}\sigma_X(i-j),n^{-2}\sum_{1\le i,j\le n}\sigma_\eta(i-j)\}\to 0$ and $\alpha(\beta+1/2)<1/2$, we have the same result as in Theorem 1.

Letting $l_a(i)$ and $l_b(i)$ be constant functions, we give some remarks on the results of Theorem 2. It is easy to see that

(a) $\Leftrightarrow \gamma_a<\gamma_b$, $1>\alpha+\gamma_a$, and $2\alpha\beta<\gamma_a$,
(b) $\Leftrightarrow \gamma_b<\gamma_a$, $1>\alpha+\gamma_b$, and $2\alpha\beta<\gamma_b$,
(c) $\Leftrightarrow 1<\alpha+\gamma_a$, $1<\alpha+\gamma_b$, and $\alpha(\beta+1/2)<1/2$.

In the case of (a), the effect of the LRD of $\{\eta_i\}$ is dominant and the optimal rate of convergence is achieved with $\alpha=\gamma_a/\{2(2+\beta)\}$. In the case of (b), the effect of the LRD of $\{X_i\}$ is dominant and the optimal rate of convergence is achieved with $\alpha=\gamma_b/\{2(2+\beta)\}$.

Bandwidth selection is one of the important problems for nonparametric estimation. It seems very difficult to use plug-in rules because of the complicated forms of the asymptotic variances. Some kind of cross validation may be possible.

Before we state the propositions, we give some useful expressions and introduce some notation. We concentrate on $\hat{m}_l(x_0)$ since $\hat{m}_c(x_0)$ can be treated in the same way. We define $\hat{T}_{n,k}^*(x_0)$ for $k=0,1$ and $\hat{T}_n^*$ as in Delaigle et al. (2009) to represent $\hat{m}_l(x_0)$ in a more tractable form.

$$\hat{T}_{n,k}^*(x_0)\ =\ \hat{T}_{n,k}(x_0)-\sum_{j=0}^1 h^j\frac{m^{(j)}(x_0)}{j!}\hat{S}_{n.k+j}(x_0),\qquad(18)$$

$$\hat{T}_n^*\ =\ (\hat{T}_{n,0}^*(x_0),\hat{T}_{n,1}^*(x_0))^T.$$

Then we have

$$\hat{m}_l(x_0)-m(x_0)\qquad\qquad\qquad\qquad\qquad\qquad(19)$$
$$=\ (1,0)\hat{S}_n^{-1}\{\hat{T}_n-\hat{S}_n(m(x_0),hm'(x_0))^T\}\ =\ (1,0)\hat{S}_n^{-1}\hat{T}_n^*.$$

To examine $\hat{T}_n^*$, we rewrite $\hat{T}_{n,k}^*(x_0)-\mathrm{E}(\hat{T}_{n,k}^*(x_0))$ as

$$\hat{T}_{n,k}^*(x_0)-\mathrm{E}(\hat{T}_{n,k}^*(x_0))=V_1^{(k)}-V_2^{(k)}+V_3^{(k)},\qquad(20)$$

11

where

$$
\begin{aligned}
V_1^{(k)} &= n^{-1}\sum_{j=1}^{n}\frac{1}{h}[K_{U,k}((W_j - x_0)/h)(m(X_j) - m(x_0)) \\
&\qquad - \mathrm{E}\{K_{U,k}((W_j - x_0)/h)(m(X_j) - m(x_0))|\mathcal{H}_{j-1}\}] \\
&\quad + n^{-1}\sum_{j=1}^{n}\frac{1}{h}[\mathrm{E}\{K_{U,k}((W_j - x_0)/h)(m(X_j) - m(x_0))|\mathcal{H}_{j-1}\} \\
&\qquad - \mathrm{E}\{K_{U,k}((W_j - x_0)/h)(m(X_j) - m(x_0))\}] \\
&= V_{11}^{(k)} + V_{12}^{(k)}, \tag{21} \\
V_2^{(k)} &= hn^{-1}\sum_{j=1}^{n}\frac{m'(x_0)}{h}[K_{U,k+1}((W_j - x_0)/h) - \mathrm{E}\{K_{U,k+1}((W_j - x_0)/h)\}], \\
V_3^{(k)} &= n^{-1}\sum_{j=1}^{n}\frac{\eta_j}{h}[\sigma(X_j)K_{U,k}((W_j - x_0)/h) \\
&\qquad - \mathrm{E}\{\sigma(X_j)K_{U,k}((W_j - x_0)/h)|\mathcal{H}_{j-1}\}] \\
&\quad + n^{-1}\sum_{j=1}^{n}\frac{\eta_j}{h}\mathrm{E}\{\sigma(X_j)K_{U,k}((W_j - x_0)/h)|\mathcal{H}_{j-1}\} \\
&= V_{31}^{(k)} + V_{32}^{(k)}. \tag{22}
\end{aligned}
$$

Note that $V_{11}^{(k)}$, $V_{12}^{(k)}$, $V_{31}^{(k)}$, and $V_{32}^{(k)}$ are defined in (21) and (22). We find the same kind of decomposition in Wu and Mielniczuk (2002) and Kulik (2008). Recall that we have taken $l_0 = 1$ for simplicity of presentation. The martingale CLT is applied to $V_{11}^{(k)}$ and $V_{31}^{(k)}$ in Proposition 2. In Propositions 3 and 4, we treat $V_{12}^{(k)}$ and $V_{32}^{(k)}$ by appealing to the theory of linear processes as in Ho and Hsing (1996,1997) and Koul and Surgailis (2002). We show in Proposition 5 that $V_2^{(k)}$ is negligible. Proposition 6 deals with $\hat{S}_n$. We deal with SRD and LRD cases in a unified way in the propositions below. The proofs of the propositions are postponed to section 5.

Proposition 1. *Suppose that Assumptions KM, MM, and X hold. Then we have for $k = 0, 1$,*

$$
\mathrm{E}(\hat{T}_{n,k}^*(x_0)) = \frac{h^2}{2}\mu_{k+2}m''(x_0)f_X(x_0) + o(h^2).
$$

Define V_m by $V_m = (V_{11}^{(0)} + V_{31}^{(0)}, V_{11}^{(1)} + V_{31}^{(1)})^T$. Proposition 2 is about the asymptotic distribution of V_m.

Proposition 2. *Suppose that Assumptions CU, CK, MM, and X hold. In addition $\alpha(\beta + 1/2) < 1/2$. Then we have*

$$h^\beta (nh)^{1/2} V_m \xrightarrow{d} N(0, \Omega(f_U * (\tau^2 f_X))(x_0)).$$

Before stating Proposition 3, we define Θ_n^2 by

$$\Theta_n^2 = \sum_{l=1}^{n}(\sum_{j=1}^{l}\theta_j)^2 + \sum_{l=0}^{\infty}(\sum_{j=1+l}^{n+l}\theta_j)^2, \tag{23}$$

where $\theta_j^2 = b_j^{4 \wedge (2\gamma_m + 2)} + B_j^2 b_j^2$ and $B_j^2 = \sum_{i=j}^{\infty} b_i^2$. The results similar to Proposition 3 are common in the literature of long-range dependent linear processes.

Proposition 3. *Suppose that Assumptions CU, CK, MM, and X hold. Then we have for $k = 0, 1$,*

$$
\begin{aligned}
h^\beta V_{12}^{(k)} &= -n^{-1}\Delta_k(f_U * ((m - m(x_0))f_X'))(x_0)\sum_{j=1}^{n} X_{j,1} \\
&\quad + O_p\Big(\frac{\Theta_n}{n}\Big) + o_p\Big(\frac{(\sum_{1 \le i,j \le n} \sigma_X(i-j))^{1/2}}{n}\Big).
\end{aligned}
$$

Proposition 4. *Suppose that Assumptions CU, CK, MM, and X hold. Then we have for $k = 0, 1$,*

$$
\begin{aligned}
h^\beta V_{32}^{(k)} &= -n^{-1}\Delta_k(f_U * (\sigma f_X))(x_0)\sum_{j=1}^{n} \eta_j \\
&\quad + O_p\Big(\frac{\Theta_n}{n}\Big) + o_p\Big(\frac{(\sum_{1 \le i,j \le n} \sigma_\eta(i-j))^{1/2}}{n}\Big).
\end{aligned}
$$

Proposition 5. *Suppose that Assumptions CU, CK, and X hold. In addition $\alpha(\beta + 1/2) < 1/2$. Then we have for $k = 0, 1$,*

$$h^\beta V_2^{(k)} = hO_p\Big((nh)^{-1/2}\Big) + hO_p\Big(\frac{\Theta_n}{n}\Big) + hO_p\Big(\frac{(\sum_{1 \le i,j \le n} \sigma_X(i-j))^{1/2}}{n}\Big).$$

13

Proposition 6. *Suppose that Assumptions CU, CK, KM, and X hold. In addition $\alpha(\beta + 1/2) < 1/2$. Then we have for $k = 0, 1, 2$,*

$$
\begin{aligned}
\hat{S}_{n,k}(x_0) \\
= \quad & f_X(x_0)\mu_k + O(h) + \frac{1}{h^\beta}O_p\Big((nh)^{-1/2}\Big) \\
& + \frac{1}{h^\beta}O_p\Big(\frac{\Theta_n}{n}\Big) + \frac{1}{h^\beta}O_p\Big(\frac{(\sum_{1 \le i,j \le n}\sigma_X(i-j))^{1/2}}{n}\Big).
\end{aligned}
$$

We prove Theorem 1 here. Recall that we have concentrated on $\hat{m}_l(x_0)$ in the proofs and that the expressions (19) and (20) are crucial in the proof. **Proof of Theorem 1.** When (8) and (13) hold, we have

$$
\Theta_n^2 = O(n), \quad \sum_{1 \le i,j \le n}\sigma_X(i-j) = O(n), \quad \text{and} \quad \sum_{1 \le i,j \le n}\sigma_\eta(i-j) = O(n).
$$

(24)

It follows from Proposition 6, (24), and the assumption on h that

$$
\hat{S}_n \xrightarrow{p} f_X(x_0)\text{diag}\{1, \mu_2\}.
$$

(25)

Propositions 3-5 and (24) imply that

$$
h^\beta(nh)^{1/2}(V_{12}^{(k)} + V_{32}^{(k)} + V_2^{(k)}) = o_p(1).
$$

Then from (20)-(22), we obtain

$$
h^\beta(nh)^{1/2}(\hat{T}_{n,k}^*(x_0) - \text{E}(\hat{T}_{n,k}^*(x_0))) = h^\beta(nh)^{1/2}(V_{11}^{(k)} + V_{31}^{(k)}) + o_p(1). \quad (26)
$$

The equations (19) and (26), Proposition 2, and (25) yield the convergence in distribution in Theorem 1.

The bias term comes from $\text{E}(\hat{T}_{n,k}^*(x_0))$ and the assertion on the bias term *Bias* follows from Proposition 1 and (25). Hence the proof of Theorem 1 is complete.

Finally we prove Theorem 2.
Proof of Theorem 2. Note that we can treat the bias term *Bias* as in the proof of Theorem 1 and we only give the details on the stochastic term.

First notice that

$$
\frac{\Theta_n}{n} = O(n^{-\gamma_b/2-\delta}) \quad \text{and} \quad \frac{\Theta_n}{n} = o\Big(\frac{(\sum_{1 \le i,j \le n}\sigma_X(i-j))^{1/2}}{n}\Big)
$$

14

for some positive δ in the case of (14).

(a) Propositions 2-5 and the assumptions on $\sigma_X(i)$, $\sigma_\eta(i)$, and the bandwidth imply that

$$h^\beta(V_{11}^{(k)} + V_{31}^{(k)}) = o_p((\sum_{1 \le i,j \le n} \sigma_\eta)^{1/2}/n),$$

$$h^\beta V_{12}^{(k)} = o_p((\sum_{1 \le i,j \le n} \sigma_\eta)^{1/2}/n),$$

$$h^\beta V_2^{(k)} = o_p((\sum_{1 \le i,j \le n} \sigma_\eta)^{1/2}/n).$$

Then we have from (20) and Proposition 4 that

$$h^\beta(\sum_{1 \le i,j \le n} \sigma_\eta(i-j))^{1/2}/n)^{-1}(\hat{T}_{n,k}^*(x_0) - \mathrm{E}(\hat{T}_{n,k}^*(x_0)))$$

$$= h^\beta(\sum_{1 \le i,j \le n} \sigma_\eta(i-j))^{1/2}/n)^{-1}V_{32}^{(k)} + o_p(1)$$

$$\xrightarrow{d} N(0, \Delta_k^2(f_U * (\sigma f_X))^2(x_0)).$$

See also Theorem 5.2.3 of Taniguchi and Kakizawa (2000) about the last line.

Proposition 6 yield that $\hat{S}_n \xrightarrow{p} f_X(x_0)\mathrm{diag}\{1, \mu_2\}$ again. Thus the assertion in the case of (a) follows from (19).

(b) We can proceed as in the case of (a) and we give just the difference. Propositions 2-5 and the assumptions on $\sigma_X(i)$, $\sigma_\eta(i)$, and the bandwidth imply that

$$h^\beta(V_{11}^{(k)} + V_{31}^{(k)}) = o_p((\sum_{1 \le i,j \le n} \sigma_X)^{1/2}/n),$$

$$h^\beta V_{32}^{(k)} = o_p((\sum_{1 \le i,j \le n} \sigma_X)^{1/2}/n),$$

$$h^\beta V_2^{(k)} = o_p((\sum_{1 \le i,j \le n} \sigma_X)^{1/2}/n).$$

Then we have from (20) that

$$h^\beta(\sum_{1 \le i,j \le n} \sigma_X(i-j))^{1/2}/n)^{-1}(\hat{T}_{n,k}^*(x_0) - \mathrm{E}(\hat{T}_{n,k}^*(x_0)))$$

$$= h^\beta \big(\sum_{1 \le i,j \le n} \sigma_X(i-j) \big)^{1/2}/n \big)^{-1} V_{12}^{(k)} + o_p(1)$$

$$\xrightarrow{d} N(0, \Delta_k^2 (f_U * ((m - m(x_0)) f_X')^2(x_0)).$$

(c) Propositions 2-5 and the assumptions on $\sigma_X(i)$, $\sigma_\eta(i)$, and the bandwidth imply that

$$h^\beta V_{12}^{(k)} = o_p((nh)^{1/2}), \quad h^\beta V_{32}^{(k)} = o_p((nh)^{1/2}), \quad h^\beta V_2^{(k)} = o_p((nh)^{1/2}).$$

Therefore we can prove the assertion as in the proof of Theorem 1. The details are omitted.

Hence the proof is complete.

4. Supersmooth case

In this section, we consider the supersmooth case by following subsection 3.2 of Kulik (2008), which crucially depends on the results in Giraitis et al. (1996). We examine the variance of the estimator more closely and show the effect of LRD in the variance.

In the supersmooth case, $\phi_U(t)$ is known and satisfies Assumption CU' below. The normal and Cauchy distributions and the convolutions of them satisfy Assumption CU'.

For simplicity of presentation, we only deal with the case where both (9) and (14) hold. However, the approximation to the joint densities in (33) below will hold in the other cases and we can derive similar results. Note that we examine only the bias and the variance of $\hat{m}_c(x_0)$ since we need to use the results on the bounds of $K_{U,0}$. in (27) and (28).

Assumption CU'
(i) $|\phi_U(t)| > 0$ for all t.
(ii) $a_0|t|^{\beta_0} \exp(-a_S|t|^{\beta_1}) \le |\phi_U(t)| \le a_1|t|^{\beta_0} \exp(-a_S|t|^{\beta_1})$ as $|t| \to \infty$, where a_0, a_1, a_S, and β_1 are positive constants and β_0 is a constant.
(iii) Let $\phi_U(t) = \tilde{R}_U(t) + i\tilde{I}_U(t)$, where $\tilde{R}_U(t)$ and $\tilde{I}_U(t)$ are the real and imaginary part of $\phi_U(t)$, respectively. Then assume that we have

$$\tilde{I}_U(t) = o(|\tilde{R}_U(t)|) \quad \text{or} \quad \tilde{R}_U(t) = o(|\tilde{I}_U(t)|).$$

Assumption CK'
(i) The function $\phi_K(t)$ has a finite support $(-d_S, d_S)$.

16

(ii) There are positive constants a_3, l, and δ such that $|\phi_K(t)| \leq a_3(d_S - t)^l$ for $t \in (d_S - \delta, d_S)$.

(iii) There are positive constants a_4, l, and δ such that $\phi_K(t) \geq a_4(d_S - t)^l$ for $t \in (d_S - \delta, d_S)$.

When Assumptions CU' and CK' hold, there are positive constants C_1 and C_2 such that

$$\sup_v |K_{U,0}(v)| \leq C_1 h^{(l+1)\beta_1 + \beta_0} (\log(1/h))^l \exp\{a_S(d_S/h)^{\beta_1}\} \qquad (27)$$

and uniformly in v on a bounded interval,

$$|K_{U,0}(v)| \geq C_2 h^{(l+1)\beta_1 + \beta_0} \exp\{a_S(d_S/h)^{\beta_1}\} \tilde{H}(v), \qquad (28)$$

where
$$\tilde{H}(v) = \begin{cases} |\cos(d_S v)|, & \text{when } \tilde{I}_U(t) = o(|\tilde{R}_U(t)|), \\ |\sin(d_S v)|, & \text{when } \tilde{R}_U(t) = o(|\tilde{I}_U(t)|). \end{cases}$$

Assumption MK'
$\int |v|^2 |K(v)| dv < \infty$.

Assumption MM'
(i) Both $|m(x)|$ and $|\sigma(x)|$ are bounded functions.
(ii) $m(x)$ is twice continuously differentiable and

$$\sup_v \sup_{|\theta| \leq 1} |m''(x_0 + \theta v) f_X(x_0 + v)| < C.$$

Assumption X'
(i) Let the random variable ξ_1 have the moment of any order.
(ii) Write $\phi_\xi(t)$ for the characteristic function of ξ_1. Then there is a positive δ such that $|\phi_\xi(t)| \leq C(1 + |t|)^{-\delta}$.

We have to strengthen Assumptions MM and X to those as above to use the argument in the proof of Lemma 3 of Giraitis et al. (1996).

Here we specify the bandwidth by taking $h = c_2(1/\log n)^{1/\beta_1}$, where $c_2 > d_S(2a_S)^{1/\beta_1}$. Further assume that

$$m'(x_0) f_X'(x_0) + \frac{1}{2} m''(x_0) f_X(x_0) \neq 0, \qquad (29)$$

$$m'(x_0) f_X''(x_0) + \frac{1}{2} m''(x_0) f_X'(x_0) \neq 0. \qquad (30)$$

We give three typical cases (a)-(c), which are considered in Theorem 3 below.

(a) $-1 + 2a_S d_S^{\beta_1} / c_2^{\beta_1} > -\min\{\gamma_a, \gamma_b\}$
(b) $-\gamma_a > -\min\{1 - 2a_S d_S^{\beta_1} / c_2^{\beta_1}, \gamma_b\}$
(c) $-\gamma_b > -\min\{1 - 2a_S d_S^{\beta_1} / c_2^{\beta_1}, \gamma_a\}$

In the case of (a), we have the same result as for i.i.d. observations. The effect of the LRD of $\{\eta_j\}$ and $\{X_j\}$ is dominant in the stochastic term in the case of (b) and (c), respectively. In all the cases in Theorem 3, the bias term *Bias* dominates the stochastic term.

We state Theorem 3 and the propositions. The proof of Theorem 3 follows the propositions. The proof of Proposition 7 is given in section 5. The proofs of the other propositions are omitted since they can be established in almost the same way as Propositions 1 and 7.

Theorem 3. *Suppose that Assumptions CU', CK', MK', MM', and X' hold and that both of (9) and (14) hold. Then we have*

$$\hat{m}_c(x_0) - m(x_0) = (f_X(x_0))^{-1}\hat{Z}_n + Bias + o_p((\mathrm{Var}(\hat{Z}_n))^{1/2}),$$

where

$$Bias = h^2 \mu_2 (m'(x_0) f_X'(x_0) + \frac{1}{2} m''(x_0) f_X(x_0)) / f_X(x_0) + o(h^2).$$

(a) When $-1 + 2a_S d_S^{\beta_1} / c_2^{\beta_1} > -\min\{\gamma_a, \gamma_b\}$, *there exist positive constants* C_1 *and* C_2 *such that*

$$C_1 \frac{h^{2\{(l+1)\beta_1 + \beta_0 - 1/2\}} \exp\{2a_S(d_S/h)^{\beta_1}\}}{n}$$

$$\leq \mathrm{Var}(\hat{Z}_n) \leq C_2 \frac{h^{2\{(l+1)\beta_1 + \beta_0 - 1\}} (\log(1/h))^{2l} \exp\{2a_S(d_S/h)^{\beta_1}\}}{n}.$$

(b) When $-\gamma_a > -\min\{1 - 2a_S d_S^{\beta_1} / c_2^{\beta_1}, \gamma_b\}$, *we have*

$$\mathrm{Var}(\hat{Z}_n) = \frac{(\sigma(x_0) f_X(x_0))^2}{n^2} \sum_{1 \leq i,j \leq n} \sigma_\eta(i-j)(1 + o(1)).$$

(c) When $-\gamma_b > -\min\{1 - 2a_S d_S^{\beta_1} / c_2^{\beta_1}, \gamma_a\}$, *we have*

$$\mathrm{Var}(\hat{Z}_n) = \frac{(m'(x_0) f_X''(x_0) + m''(x_0) f_X'(x_0)/2)^2 \mu_2^2 h^4}{n^2} \sum_{1 \leq i,j \leq n} \sigma_X(i-j)(1+o(1)).$$

18

Since

$$\hat{m}_c(x_0) - m(x_0) = \hat{S}_n^{-1}(\hat{T}_n^* - \mathrm{E}\{\hat{T}_n^*\}) + \hat{S}_n^{-1}\mathrm{E}\{\hat{T}_n^*\}, \tag{31}$$

we should examine $\hat{T}_n^* - \mathrm{E}\{\hat{T}_n^*\}$ and $\hat{S}_n$ to prove Theorem 3. We can treat $\mathrm{E}\{\hat{T}_n^*\}$ in the same way as in section 3 and the details are omitted. To examine $\hat{T}_n^* - \mathrm{E}\{\hat{T}_n^*\}$, we decompose the exprssion into V_1, V_2, and V_3 as

$$
\begin{aligned}
\hat{T}_n^* &- \mathrm{E}\{\hat{T}_n^*\} \tag{32}\\
=\ & n^{-1}\sum_{j=1}^{n}\Big[\frac{1}{h}(m(X_j) - m(x_0))K_{U,0}((W_j - x_0)/h)\\
&\quad -\mathrm{E}\Big\{\frac{1}{h}(m(X_j) - m(x_0))K_{U,0}((W_j - x_0)/h)\Big\}\Big]\\
&+n^{-1}\sum_{j=1}^{n}\eta_j\mathrm{E}\Big\{\frac{\sigma(X_j)}{h}K_{U,0}((W_j - x_0)/h)\Big\}\\
&+n^{-1}\sum_{j=1}^{n}\eta_j\Big[\frac{\sigma(X_j)}{h}K_{U,0}((W_j - x_0)/h)\\
&\quad -\mathrm{E}\Big\{\frac{\sigma(X_j)}{h}K_{U,0}((W_j - x_0)/h)\Big\}\Big]\\
=\ & V_1 + V_2 + V_3.
\end{aligned}
$$

We evaluate V_1, V_2, V_3, and $\hat{S}_n$ in Propositions 7, 8, 9, and 10, respectively.

The argument in Lemma 3 of Giraitis et al. (1996) is based on the approximation of the joint density of (X_i, X_j), which is denoted by $f_{X,i-j}(x_i, x_j)$. The approximation below is given in Lemma 2 of Giraitis et al. (1996).

$$
\begin{aligned}
f_{X,i-j}&(x_i, x_j) \tag{33}\\
=\ & f_X(x_i)f_X(x_j) + \sigma_X(i - j)f'_X(x_i)f'_X(x_j) + |\sigma_X(i - j)|^{1+\delta}\rho_{X,i-j}(x_i, x_j)
\end{aligned}
$$

for some positive δ, where $\rho_{X,i-j}(x_i, x_j)$ is is uniformly bounded in x_i, x_j, and $i - j$.

Proposition 7. *Suppose all the assumptions of Theorem 3 hold. Then we have for some positive δ_1,*

$$
\begin{aligned}
\mathrm{Var}&(V_1) \tag{34}\\
=\ & \frac{1}{n}\mathrm{Var}\Big\{\frac{1}{h}(m(X_1) - m(x_0))K_{U,0}((W_1 - x_0)/h)\Big\}
\end{aligned}
$$

19

$$+2(m'(x_0)f_X''(x_0) + \frac{1}{2}m''(x_0)f_X'(x_0))^2\mu_2 h^4$$
$$\times \frac{1}{n^2} \sum_{1 \le i < j \le n} \sigma_X(i-j)(1+o(1))$$
$$+O\Big(\frac{1}{n^2h^2} \sum_{1 \le i < j \le n} |\sigma_X(i-j)|^{1+\delta_1}\Big).$$

Besides, for some positive constants C_1 and C_2,

$$C_1 h^{2\{(l+1)\beta_1+\beta_0-1/2\}} \exp\{2a_S(d_S/h)^{\beta_1}\}$$
$$\le \mathrm{Var}\Big\{\frac{1}{h}(m(X_1) - m(x_0))K_{U,0}((W_1 - x_0)/h)\Big\}$$
$$\le C_2 h^{2\{(l+1)\beta_1+\beta_0-1\}}(\log(1/h))^{2l} \exp\{2a_S(d_S/h)^{\beta_1}\}.$$

Proposition 8. *Suppose all the assumptions of Theorem 3 hold. Then we have*
$$\mathrm{Var}(V_2) = \frac{(\sigma(x_0)f_X(x_0))^2}{n^2} \sum_{1 \le i,j \le n} \sigma_\eta(i-j)(1+o(1)).$$

Proposition 9. *Suppose all the assumptions of Theorem 3 hold. Then we have for some positive δ_1,*

$$\mathrm{Var}(V_3) \tag{35}$$
$$= \frac{1}{n}\mathrm{E}(\eta_1^2)\mathrm{Var}\Big\{\frac{\sigma(X_1)}{h}K_{U,0}((W_1 - x_0)/h)\Big\}$$
$$+O\Big(\frac{1}{n^2} \sum_{1 \le i,j \le n} \sigma_X(i-j)\sigma_\eta(i-j)\Big) + O\Big(\frac{1}{n^2h^2} \sum_{1 \le i,j \le n} |\sigma_X(i-j)|^{1+\delta_1}\Big).$$

Besides, for some positive constants C_1 and C_2,

$$C_1 h^{2\{(l+1)\beta_1+\beta_0-1/2\}} \exp\{2a_S(d_S/h)^{\beta_1}\}$$
$$\le \mathrm{Var}\Big\{\frac{\sigma(X_1)}{h}K_{U,0}((W_1 - x_0)/h)\Big\}$$
$$\le C_2 h^{2\{(l+1)\beta_1+\beta_0-1\}}(\log(1/h))^{2l} \exp\{2a_S(d_S/h)^{\beta_1}\}.$$

Proposition 10. *Suppose all the assumptions of Theorem 3 hold. Then we have for some positive δ_1,*

$$\mathrm{Var}(\hat{S}_{n,0}(x_0))$$

$$= O\left(\frac{1}{n}h^{2\{(l+1)\beta_1+\beta_0-1\}}(\log(1/h))^{2l}\exp\{2a_S(d_S/h)^{\beta_1}\}\right)$$
$$+ O\left(\frac{1}{n^2}\sum_{1\le i,j\le n}\sigma_X(i-j)\right) + \left(\frac{1}{n^2h^2}\sum_{1\le i,j\le n}|\sigma_X(i-j)|^{1+\delta_1}\right).$$

Besides, we have
$$\mathrm{E}(\hat{S}_{n,0}(x_0)) = f_X(x_0) + O(h^2).$$

We prove Theorem 3 by exploiting Propositions 7-10 and (31).

Proof of Theorem 3. The definition of the bandwidth h in this section and Proposition 10 imply that

$$\hat{m}_c(x_0) - m(x_0) = \frac{1}{f_X(x_0)}(\hat{T}^* - \mathrm{E}(\hat{T}^*)) + \hat{S}_n^{-1}\mathrm{E}(\hat{T}^*) + o_p((\mathrm{Var}(\hat{T}^*))^{1/2}).$$

Take $\hat{Z}_n = \hat{T}^* - \mathrm{E}(\hat{T}^*) = V_1 + V_2 + V_3$ and note that $\exp\{2a_S(d_S/h)^{\beta_1}\}/n = n^{-1+2a_S d_S^{\beta_1}/c_2^{\beta_1}}$. Then the assertion on $\mathrm{Var}(\hat{T}^*)$ follows from Propositions 7-9. Actually, in the case of (a), the first terms of the right-hand side of (34) and (35) are dominant. It is easy to see that $\mathrm{Var}(V_2)$ and the second term of the right-hand side of (34) is dominant in the csaes of (b) and (c), respectively.

We can treat the bias term *Bias* in the same way as in Theorems 1 and 2. Hence the proof of Theorem 3 is complete.

5. Proofs of propositions

Proof of Proposition 1. We have from (5) and Assumption MM that

$$\mathrm{E}\{\hat{T}_{n,k}(x_0) - m(x_0)\hat{S}_{n,k}(x_0)\}$$
$$= \int v^k K(v)(m(x_0 + vh) - m(x_0))f_X(x_0 + vh)dv$$
$$= \int v^k K(v)\left(m'(x_0)vh + m''(x_0)\frac{(vh)^2}{2}\right)f_X(x_0 + vh)dv$$
$$+ \int \frac{(vh)^2}{2}K(v)(m''(x_0 + \theta vh) - m''(x_0))f_X(x_0 + vh)dv$$
$$= m'(x_0)(h\mu_{k+1}f_X(x_0) + h^2\mu_{k+2}f'_X(x_0))\frac{m''(x_0)h^2\mu_{k+2}f_X(x_0)}{2} + o(h^2)$$

Similarly,
$$m'(x_0)h\mathrm{E}(\hat{S}_{n,k+1}(x_0)) = m'(x_0)(h\mu_{k+1}f_X(x_0) + h^2\mu_{k+2}f'_X(x_0)) + o(h^2).$$

Hence the assertion of the proposition follows from (18).

Proof of Proposition 2. We only give the details on $V_{31}^{(k)}$ because the rest of the proof follows from the similar argument, the Cramér-Wold device, and the independence among $\{\epsilon_i\}$, $\{\xi_i\}$, and $\{U_i\}$. We apply the martingale CLT since $V_{31}^{(k)}$ is wirtten as

$$
V_{31}^{(k)} = \frac{1}{nh} \sum_{j=1}^{n} \eta_j [\sigma(X_j) K_{U,k}((W_j - x_0)/h)
$$
$$
- \mathrm{E}\{\sigma(X_j) K_{U,k}((W_j - x_0)/h) | \bar{\mathcal{H}}_{j-1}\}].
$$

and it is the sum of the martingale differences w.r.t. $\{\bar{\mathcal{H}}_j\}$.

All we have to do is to establish the Lindeberg condition and the convergence in probability of the conditional variance w.r.t. $\{\bar{\mathcal{H}}_j\}$. The conditional variance is written as

$$
\frac{1}{n} \sum_{j=1}^{n} \eta_j^2 \int h^{2\beta} K_{U,k}^2(v) \left\{ \int f_U(x_0 - t + vh) f_\xi(t - X_{j,1}) \sigma^2(t) dt \right\} dv + o_p(1)
$$
$$
= \frac{1}{n} \sum_{j=1}^{n} \eta_j^2 \int \tilde{L}_k^2(v) \left\{ \int f_U(x_0 - t + vh) f_\xi(t - X_{j,1}) \sigma^2(t) dt \right\} dv + o_p(1)
$$
$$
= \frac{1}{n} \sum_{j=1}^{n} \eta_j^2 \int \tilde{L}_k^2(v) dv \int f_U(x_0 - t) f_\xi(t - X_{j,1}) \sigma^2(t) dt + o_p(1)
$$
$$
\xrightarrow{p} \Omega_{kk} \mathrm{E}(\eta_1^2)(f_U * (\sigma^2 f_X))(x_0).
$$

Here we used the properties of $\tilde{L}_k(v)$, the boundedness of $(\sigma^2 f_X)(t)$, and the ergodic theorem. We can show in the same way that the conditional variance of $V_{11}^{(k)}$ converges in probability to $\Omega_{kk}(f_U * ((m - m(x_0))^2 f_X))(x_0)$.

Finally we establish the Lindeberg condition. We write Z_j for the j-th term of $h^\beta(nh)^{1/2} V_{31}^{(k)}$. Assumptions MM and X imply that there is a positive constant δ s.t. $(\sigma^{2+\delta} f_X)(t)$ is bounded. Then we have

$$
n\mathrm{E}(|Z_1|^{2+\delta})
$$
$$
\leq C(nh)^{-\delta/2} \int h^{\beta(2+\delta)} |K_{U,k}(v)|^{2+\delta}
$$
$$
\times \left\{ \int f_U(x_0 - t + vh)(\sigma^{2+\delta} f_X)(t) dt \right\} dv
$$
$$
\leq C(nh)^{-\delta/2} \to 0.
$$

22

Therefore for any positive ϵ,

$$\sum_{j=1}^{n} \mathrm{E}\{|Z_j|^2 I(|Z_j| > \epsilon)|\bar{\mathcal{H}}_{j-1}\} \xrightarrow{p} 0.$$

The convergence of the conditional variance and the Lindeberg condition yield that

$$h^\beta (nh)^{1/2} V_{31}^{(k)} \to N(0, \Omega_{kk}(f_U * (\sigma^2 f_X))(x_0)).$$

Hence the proof of the proposition is complete.

Proof of Proposition 3. By applying Lemma 1 and Lemma 2 with $M(x) = m(x) - m(x_0)$ and $d_j = 1$, respectively, we get

$$
\begin{aligned}
h^\beta & V_{12}^{(k)} \\
= & \ n^{-1} \sum_{j=1}^{n} h^\beta \int K_{U,k}(v) \Big\{ \int f_U(x_0 - t + vh)(f_\xi(t - X_{j,1}) - f_X(t)) \\
& \times (m(t) - m(x_0)) dt \Big\} dv \\
= & \ n^{-1} \sum_{j=1}^{n} h^\beta \int K_{U,k}(v) \Big\{ \int f_U(x_0 - t + vh) f_X'(t) \\
& \times (m(t) - m(x_0)) dt \Big\} dv X_{j,1} + O_p\Big(\frac{\Theta_n}{n}\Big) \\
= & \ -n^{-1} \Delta_k(f_U * ((m - m(x_0)) f_X'))(x_0) \sum_{j=1}^{n} X_{j,1} \\
& + o_p\Big(\frac{(\sum_{1 \le i,j \le n} \sigma_X(i-j))^{1/2}}{n}\Big) + O_p\Big(\frac{\Theta_n}{n}\Big).
\end{aligned}
$$

Hence the proof of the proposition is complete.

Proof of Proposition 4. By applying Lemma 1 and Lemma 2 with $M(x) = \sigma(x)$ and $d_j = \eta_j$, respectively, we get

$$
\begin{aligned}
h^\beta & \sum_{j=1}^{n} \mathrm{E}\Big\{ \frac{\sigma(X_j)}{h} K_{U,k}((W_j - x_0)/h)|\mathcal{H}_{j-1}\Big\} \\
= & \sum_{j=1}^{n} \eta_j h^\beta \int K_{U,k}(v) \Big\{ \int f_U(x_0 - t + vh)(f_\xi(t - X_{j,1}) - f_X(t)
\end{aligned}
$$

$$+ f_X'(t)X_{j,1})\sigma(t)dt\Big\}dv$$

$$+ \sum_{j=1}^{n} \eta_j h^\beta \int K_{U,k}(v)\Big\{\int f_U(x_0 - t + vh)f_X(t)\sigma(t)dt\Big\}dv$$

$$- \sum_{j=1}^{n} \eta_j X_{j,1} h^\beta \int K_{U,k}(v)\Big\{\int f_U(x_0 - t + vh)f_X'(t)\sigma(t)dt\Big\}dv$$

$$= \Delta_k(f_U * (\sigma f_X))(x_0) \sum_{i=1}^{n} \eta_j + O_p\Big(\frac{\Theta_n}{n}\Big) + o_p\Big(\frac{(\sum_{1\le i,j\le n}\sigma_\eta(i-j))^{1/2}}{n}\Big).$$

Hence the proof of the proposition is complete.
Proof of Proposition 5. Since

$$h^\beta V_2^{(k)}$$
$$= \frac{h^{\beta+1}m'(x_0)}{nh} \sum_{j=1}^{n} [K_{U,k+1}((W_j - x_0)/h) - \mathrm{E}\{K_{U,k+1}((W_j - x_0)/h)|\mathcal{H}_{j-1}\}$$
$$+ \{\mathrm{E}\{K_{U,k+1}((W_j - x_0)/h)|\mathcal{H}_{j-1}\} - \mathrm{E}\{K_{U,k+1}((W_j - x_0)/h)\}\}],$$

the assertion of the proposition follows from the arguments in the proofs of Propositions 2 and 3.
Proof of Proposition 6. Write

$$\hat{S}_{n,k}(x_0) - \mathrm{E}(\hat{S}_{n,k}(x_0))$$
$$= h^{-\beta}[h^\beta n^{-1} \sum_{j=1}^{n} \{K_{U,k}((W_j - x_0)/h) - \mathrm{E}\{K_{U,k}((W_j - x_0)/h)|\mathcal{H}_{j-1}\}\}$$
$$+ h^\beta n^{-1} \sum_{j=1}^{n} \{\mathrm{E}\{K_{U,k}((W_j - x_0)/h)|\mathcal{H}_{j-1}\} - \mathrm{E}\{K_{U,k}((W_j - x_0)/h)\}\}]$$
$$= \frac{1}{h^\beta}O_p\Big((nh)^{-1/2}\Big) + \frac{1}{h^\beta}O_p\Big(\frac{(\sum_{1\le i,j\le n}\sigma_X(i-j))^{1/2}}{n}\Big) + \frac{1}{h^\beta}O_p\Big(\frac{\Theta_n}{n}\Big)$$

The last line follows from the arguments in the proofs of Propositions 2 and 3. We can evaluate $\mathrm{E}(\hat{S}_{n,k}(x_0))$ in the same way as in the proof of Proposition 1. The details are omitted. Hence the proof of the proposition is complete.
Proof of Proposition 7. Write

$$Z_j = \frac{1}{h}(m(X_j) - m(x_0))K_{U,k}((W_j - x_0)/h).$$

24

Then

$$\operatorname{Var}(V_1) = \frac{1}{n^2} \sum_{j=1}^{n} \operatorname{Var}(Z_j) + \frac{2}{n^2} \sum_{1 \le i < j \le n} \operatorname{Cov}(Z_i, Z_j). \tag{36}$$

First we consider the second term of (36) by using (33) and establish the second assertion of the proposition at the end of the proof.

Notice that the integration w.r.t. $f_X(x_i) f_X(x_j) f_U(u_i) f_U(u_j)$ vanishes in $\operatorname{Cov}(Z_i, Z_j)$ when we replace $f_{X,i-j}(x_i, x_j)$ with the right-hand side of (33). Thus we have only to evaluate in $\operatorname{Cov}(Z_i, Z_j)$ the integration w.r.t. $\sigma_X(i-j) f'_X(x_i) f'_X(x_j) f_U(u_i) f_U(u_j)$ and $|\sigma_X(i-j)|^{1+\delta} \rho_{X,i-j}(x_i, x_j) f_U(u_i) f_U(u_j)$.

The sum of the integration w.r.t. $\sigma_X(i-j) f'_X(x_i) f'_X(x_j) f_U(u_i) f_U(u_j)$ is written as

$$\frac{2}{n^2} \sum_{1 \le i < j \le n} \sigma_X(i-j)$$

$$\times \int \int \int \int \left\{ \frac{1}{h}(m(x_i) - m(x_0)) K_{U,0}((x_i + u_i - x_0)/h) - \mathrm{E}(Z_i) \right\}$$

$$\times \left\{ \frac{1}{h}(m(x_j) - m(x_0)) K_{U,0}((x_j + u_j - x_0)/h) - \mathrm{E}(Z_j) \right\}$$

$$\times f'_X(x_i) f'_X(x_j) f_U(u_i) f_U(u_j) dx_i dx_j du_i du_j$$

$$= \frac{2}{n^2} \sum_{1 \le i < j \le n} \sigma_X(i-j) \int \int \left\{ \frac{1}{h}(m(x_i) - m(x_0)) K((x_i - x_0)/h) - \mathrm{E}(Z_i) \right\}$$

$$\times \left\{ \frac{1}{h}(m(x_j) - m(x_0)) K((x_j - x_0)/h) - \mathrm{E}(Z_j) \right\} f'_X(x_i) f'_X(x_j) dx_i dx_j$$

$$= \left(m'(x_0) f''_X(x_0) + \frac{m''(x_0) f'_X(x_0)}{2} \right)^2 \mu_2^2 h^4 \tag{37}$$

$$\times \frac{2}{n^2} \sum_{1 \le i < j \le n} \sigma_X(i-j)(1 + o(1)).$$

The sum of the integration w.r.t. $|\sigma_X(i-j)|^{1+\delta} \rho_{X,i-j}(x_i, x_j) f_U(u_i) f_U(u_j)$ is evaluated as

$$\frac{2}{n^2} \sum_{1 \le i < j \le n} |\sigma_X(i-j)|^{1+\delta}$$

$$\times \int \int \int \int \left\{ \frac{1}{h}(m(x_i) - m(x_0)) K_{U,0}((x_i + u_i - x_0)/h) - \mathrm{E}(Z_i) \right\}$$

$$\times \left\{ \frac{1}{h}(m(x_j) - m(x_0)) K_{U,0}((x_j + u_j - x_0)/h) - \mathrm{E}(Z_j) \right\}$$

25

$$\times \rho_{X,i-j}(x_i, x_j) f_U(u_i) f_U(u_j) dx_i dx_j du_i du_j$$

$$= \frac{2}{n^2} \sum_{1 \le i < j \le n} |\sigma_X(i-j)|^{1+\delta}$$

$$\times \int \int \left\{ \frac{1}{h}(m(x_i) - m(x_0)) K((x_i - x_0)/h) - \mathrm{E}(Z_i) \right\}$$

$$\times \left\{ \frac{1}{h}(m(x_j) - m(x_0)) K((x_j - x_0)/h) - \mathrm{E}(Z_j) \right\} \rho_{X,i-j}(x_i, x_j) dx_i dx_j$$

$$= O\left(\frac{1}{n^2 h^2} \sum_{1 \le i < j \le n} |\sigma_X(i-j)|^{1+\delta_1} \right) \tag{38}$$

for some positive δ_1. The argument in the proof of Lemma 3 of Giraitis et al. (1996) is applied to derive the last equality. See Giraitis et al. (1996) for the details. Thus the first assertion of the proposition follows from (36)-(38).

Finally we consider $\mathrm{Var}(Z_1)$, which is represented as

$$\mathrm{Var}(Z_1)$$
$$= \int \int \frac{1}{h^2}(m(x) - m(x_0))^2 K_{U,0}^2((x + u - x_0)/h) \tag{39}$$
$$\times f_X(x) f_U(u) dx du$$
$$= \frac{1}{h} \int \{((m - m(x_0))^2 f_X) * f_U\}(x_0 + vh) K_{U,0}^2(v) dv. \tag{40}$$

We apply the upper bound (27) and the lower bound (28) to (39) and (40), respectively and the second assertion of the proposition is established.

Hence the proof of the proposition is complete.

6. Technical lemmas

In this section, Lemmas 1-3 and the proofs are given. Lemma 3 is similar to Lemma 5.2 of Koul and Surgailis (2002) and the proof is omitted. We find lemmas similar to Lemmas 1 and 2 in the literature of long-range dependent linear processes, for example, in Koul and Surgailis (2002). However, I have not found any lemmas which can cover the result of Lemma 1. Lemma 2 is just given for reference.

Lemma 1. *Suppose that Assumptions CU, CK, and X hold. Let $M(t)$ be a function satisfying $|M(t)| \le C(1 + |t|)^{\gamma_M}$ $(0 < \gamma_M \le \gamma_m)$. Set*

$$H_j = \int f_U(x_0 - t + vh)(f_\xi(t - X_{j,1}) - f_X(t) + f_X'(t) X_{j,1}) M(t) dt$$

26

for $j \geq 2$. Then we have for $k = 0, 1, 2$,

$$h^{2\beta} E\left[\left\{\int K_{U,k}(v)(E\{H_j(v)|\mathcal{F}_1\} - E\{H_j(v)|\mathcal{F}_0\})dv\right\}^2\right]$$

$$\leq C\theta_{j-1}^2\left(h^\beta \int |K_{U,k}(v)|dv\right)^2,$$

where C is independent of j.

PROOF. Since (15) implies that $h^\beta \int |K_{U,k}(v)|dv < \infty$, we have only to verify that

$$\sup_v E\{|E\{H_j(v)|\mathcal{F}_1\} - E\{H_j(v)|\mathcal{F}_0\}|^2\} \leq C\theta_{j-1}^2. \qquad (41)$$

Rewrite $E\{H_j(v)|\mathcal{F}_1\} - E\{H_j(v)|\mathcal{F}_0\}$ and define $U_{j,1}$, $U_{j,2}$, and $U_{j,3}$ by

$$E\{H_j(v)|\mathcal{F}_1\} - E\{H_j(v)|\mathcal{F}_0\}$$
$$= \int f_U(x_0 - t + vh)\{f_j(t - X_{j,j-1}) - f_j(t - X_{j,j})$$
$$+ b_{j-1}\xi_1 f_j'(t - X_{j,j})\}M(t)dt$$
$$+ \int f_U(x_0 - t + vh)(f_{j-1}(t - X_{j,j-1}) - f_j(t - X_{j,j-1})M(t)dt$$
$$- b_{j-1}\xi_1 \int f_U(x_0 - t + vh)(f_j'(t - X_{j,j}) - f_X'(t))M(t)dt$$
$$= U_{j,1} + U_{j,2} + U_{j,3}.$$

By representing $U_{j,1}$ as

$$U_{j,1} = \int_0^{-b_{j-1}\xi_1}\left\{\int f_U(x_0 - t + vh)\{f_j'(t - X_{j,j} + z)\right.$$
$$\left. - f_j'(t - X_{j,j})\}M(t)dt\right\}dz$$

and applying Lemma 3 below, we get

$$|U_{j,1}|^2 \leq C\{|b_{j-1}\xi_1|^4 I(|b_{j-1}\xi_1| \leq 1)$$
$$+ |b_{j-1}\xi_1|^{2\gamma_M+2}I(|b_{j-1}\xi_1| > 1)\}(1 + |X_{j,j}|)^{2\gamma_M}.$$

Taking the expectation of the above expression, we obtain

$$E\{|U_{j,1}|^2\} \leq C|b_{j-1}|^{4\wedge(2\gamma_M+2)}.$$

27

Next we consider $U_{j,2}$. By noting that that $|M(t+v)| \le C(1+|t|)^{\gamma_M}(1+|v|)^{\gamma_M}$ and exploiting Fact D2, we have

$$|U_{j,2}|^2 \le Cb_{j-1}^4(1 + |X_{j,j-1}|)^{2\gamma_M}.$$

Taking the expectation of the above expression, we obtain

$$\mathrm{E}\{|U_{j,2}|^2\} \le Cb_{j-1}^4.$$

Finally we deal with $U_{j,3}$. We represent $U_{j,3}$ as

$$
\begin{aligned}
U_{j,3} \\
&= -b_{j-1}\xi_1 \int f_U(x_0 - t + vh)\tilde{\mathrm{E}}\{f_j'(t - X_{j,j}) - f_j'(t - \tilde{X}_{j,j})\}M(t)dt \\
&= -b_{j-1}\xi_1 \int f_U(x_0 - t + vh)(f_j'(t - X_{j,j}) - f_j'(t))M(t)dt \\
&\quad + \tilde{\mathrm{E}}\left[b_{j-1}\xi_1 \int f_U(x_0 - t + vh)(f_j'(t - X_{j,j}) - f_j'(t))M(t)dt\right] \quad (42)
\end{aligned}
$$

where $\tilde{X}_{j,j}$ is an independent copy of $X_{j,j}$ and $\tilde{\mathrm{E}}\{\cdot\}$ is the expectation w.r.t. $\tilde{X}_{j,j}$.

Let T_1 be the first term of (42). Then we apply Lemma 3 and get

$$|T_1|^2 \le C|b_{j-1}\xi_1|^2\{|X_{j,j}|^2 I(|X_{j,j}| \le 1) + |X_{j,j}|^{2\gamma_M} I(|X_{j,j}| > 1)\}.$$

The expectation is bounded from above by

$$\mathrm{E}(|T_1|^2) \le Cb_{j-1}^2 B_j^2.$$

We can treat the second term of (42) in the same way and the details are omitted.

Hence the proof of the lemma is complete.

Lemma 2. *Let $\{d_j\}$ be a sequence of random variables independent of $\{\xi_j\}$. In addition $\mathrm{E}(|d_j|^2) < \infty$ and $g(x)$ is a function such that*

$$\mathrm{E}\{|\mathrm{E}\{g(X_{j,1})|\mathcal{F}_1\} - \mathrm{E}\{g(X_{j,1})|\mathcal{F}_0\}|^2\} \le C_g\theta_{j-1}^2.$$

for a positive constant C_g. Then we have

$$\mathrm{E}\left\{\left(\sum_{j=1}^n d_j g(X_{j,1})\right)^2\right\} \le C_g \sup_{1 \le j \le n} \mathrm{E}(|d_j|^2)\Theta_n^2$$

PROOF. Represent $\sum_{j=1}^{n} d_j g(X_{j,1})$ as

$$
\begin{aligned}
\sum_{j=1}^{n} d_j g(X_{j,1}) &= \sum_{l=1}^{n-1} \sum_{j=l+1}^{n} d_j [\mathrm{E}\{g(X_{j,1})|\mathcal{F}_l\} - \mathrm{E}\{g(X_{j,1})|\mathcal{F}_{l-1}\}]) \qquad (43) \\
&\quad + \sum_{l=0}^{\infty} \sum_{j=1}^{n} d_j [\mathrm{E}\{g(X_{j,1})|\mathcal{F}_{-l}\} - \mathrm{E}\{g(X_{j,1})|\mathcal{F}_{-l-1}\}].
\end{aligned}
$$

By exploiting the properties of sums of martingale differences of (43) and applying the triangle inequality, we can derive the assertion of the lemma.

Note that $f_j(x)$, $f_j'(x)$, $j = 1, 2, \ldots$, satisfy the assumption on $\phi(x)$ in Lemma 3 with $\gamma = \gamma_1$ and C_γ independent of j.

Lemma 3. *Let $M(t)$ be a function satisfying $|M(t)| \le C(1 + |t|)^{\gamma_M}$ $(0 < \gamma_M \le \gamma_m)$. Besides, There is another function $\phi(x)$ which satisfies the following conditions.*

$$
\begin{aligned}
|\phi(x) - \phi(y)| &\le C_\gamma h_\gamma(x)|x - y|, \quad |x - y| \le 1 \\
|\phi(x)| &\le C_\gamma h_\gamma(x),
\end{aligned}
$$

where $h_\gamma(x) = (1 + |x|)^{-\gamma}(\gamma > \gamma_M + 1)$. Then we have

$$
\begin{aligned}
&\int |\phi(t + v) - \phi(t)||M(t)|dt \\
&\le C\{|v|I(|v| \le 1) + |v|^{\gamma_M} I(|v| > 1)\}, \\
&\int_0^v \left\{ \int |\phi(t + w - z) - \phi(t - z)||M(t)|dt \right\} dw \\
&\le C\{|v|^2 I(|v| \le 1) + |v|^{\gamma_M + 1} I(|v| > 1)\}(1 + |z|)^{\gamma_M}.
\end{aligned}
$$

References

[1] Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall, New York.

[2] Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models, 2nd Ed. Chapman & Hall/CRC, Boca Raton.

[3] Comte, F., Taupin, M.-L., 2007. Nonparametric estimation of the regression function in an errors-in-variables model. Statist Sinica 17, 1065–1090.

[4] Delaigle, A., Fan, J., Carroll, R.J., 2009. A design-adaptive local polynomial estimator for the errors-in-variables problem. J. Amer. Statist.Assoc. 104, 348–359.

[5] Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman & Hall, London.

[6] Fan, J., Masry, E., 1992. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. J. Multivariate Anal. 43, 237–271.

[7] Fan, J., Truong, Y.K., 1993. Nonparametric regression with errors in variables. Ann. Statist. 21, 1900–1925.

[8] Fan, J., Yao, Q., 2003. Nonlinear Time Series. Springer, Nwe York.

[9] Giraitis, L., Koul, H.L., Surgailis, D., 1996. Asymptotic normality of regression estimators with long memory errors. Statist. & Probab. Letters 29, 317–335.

[10] Ho, H.-C., Hsing, T., 1996. On the asymptotic expansion of the empirical process of long-memory moving averages. Ann. Statist. 24, 992–1024.

[11] Ho, H.-C., Hsing, T., 1997. Limit theorems for functionals of moving averages. Ann. Probab. 25, 1636–1669.

[12] Honda, T., 2009. Nonparametric density estimation for linear processes with infinite variance. Ann. Inst. Statist. Math. 61, 413–439.

[13] Hu, Y., Schennach, S.M., 2008. Instrumental variable treatment of non-classical measurement error models. Econometrica 76, 195–216.

[14] Koul, H.L., Surgailis, D., 2002. Asymptotic expansion of the empirical process of long memory moving averages. In: Dehling, H., Mikosch, M., Sørensen, M.(Eds.), Empirical Process Techniques for Dependent Data, Birkhäser, Boston, 213–239.

[15] Kulik, R., 2008. Nonparametric deconvolution problem for dependent sequences. Electronic J. of Statist. 2, 722-740.

[16] Masry, E., 1991. Multivariate probability density deconvolution for stationary random processes. IEEE Trans. Information Theory 37, 1105–1115.

[17] Masry, E., 1993a. Strong consistency and rates for deconvolution of multivariate densities of stationary processes. Stoch. Proc. Appl. 47, 53–74.

[18] Masry, E., 1993b. Asymptotic normality for deconvolving estimators of multivariate densities for stationary processes. J. Multivariate Anal. 44, 47–68.

[19] Masry, E., 2003. Deconvolving multivariate kernel density estimates from contaminated associated observations. IEEE Trans. Information Theory 49, 2941–2952.

[20] Masry, E., Mielniczuk, J., 1999. Local linear regression estimation for time series with long-range dependence. Stoch. Proc. Appl. 82, 173–193.

[21] Meister, A., 2009. Deconvolution Problems in Nonparametric Statistics. Lecture Notes in Statistics 193, Springer, Berlin-Heidelberg.

[22] Mielniczuk, J., Wu, W.B., 2004. On random-design model with dependent errors. Statist. Sinica 14, 1105–1126.

[23] Schennach, S.M., 2004. Nonparametric regression in the presence of measurement error. Econometric Theory 20, 1046–1093.

[24] Taniguchi, M., Kakizawa, Y., 2000. Asymptotic Theory of Statistical Inference for Time Series. Springer, Nwe York.

[25] van Zanten, H., Zareba, P., 2008. A note on wavelet density deconvolution for weakly dependent data. Statist. Inference Stoch. Proc. 11, 207–219.

[26] Wu, W.B., Mielniczuk, J., 2002. Kernel density estimation for linear processes. Ann. Statist. 30, 1441–1459.