Ann Inst Stat Math (2009) 61:391–411
DOI 10.1007/s10463-007-0142-4

M-estimation in nonparametric regression under strong dependence and infinite variance

Ngai Hang Chan · **Rongmao Zhang**

Received: 8 September 2006 / Revised: 11 April 2007 / Published online: 10 August 2007

Abstract A robust local linear regression smoothing estimator for a nonparametric regression model with heavy-tailed dependent errors is considered in this paper. Under certain regularity conditions, the weak consistency and asymptotic distribution of the proposed estimators are obtained. If the errors are short-range dependent, then the limiting distribution of the estimator is normal. If the data are long-range dependent, then the limiting distribution of the estimator is a stable distribution.

Keywords Heavy-tailed · Long-range dependence · M-estimation · Nonparametric regression · Stable distribution

1 Introduction

Let $\{\xi_j, j \in \mathbf{Z}\}$ be i.i.d. random variables that belong to the domain of attraction of a symmetric stable law with index $\alpha \in (0, 2]$, denoted by $\xi_i \in D(\alpha)$. That is, as $n \to \infty$,

$$\frac{1}{n^{1/\alpha} L_\varepsilon(n)} \sum_{i=1}^{[nx]} \xi_i \to Z_\alpha(x) \quad \text{in distribution,}$$

where $Z_\alpha(x)$ is a Lévy-stable process of index α with Lévy measure

$$\nu(dx) = \alpha\{C_+ I(x > 0) + C_- I(x < 0)\}|x|^{-\alpha-1} dx$$

N. H. Chan (✉)
Department of Statistics, Chinese University of Hong Kong, Shatin, NT, Hong Kong
e-mail: nhchan@sta.cuhk.edu.hk

R. Zhang
Department of Mathematics, Zhejiang University, Hangzhou 310027, China
e-mail: rmzhang@zju.edu.cn

and $L_\varepsilon(x)$ is a slowly varying function that satisfies

$$n^{1/\alpha} L_\varepsilon(n) = \inf\{x : P\{|\xi_1| \geq x\} \leq 1/n\}.$$

Let

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

be a nonparametric regression model with fixed design $x_i = i/n$, where $m(\cdot)$ is an unknown measurable function on [0, 1] and $\{\varepsilon_i\}$ is a discrete stationary process given by

$$\varepsilon_i = \sum_{j=-\infty}^{i} a_{i-j}\xi_j$$

with $0 < \sum_{j=0}^{\infty} |a_j|^\alpha < \infty$. Various results are available for the estimation of $m(\cdot)$ when $\alpha = 2$. For example, when $\{\varepsilon_i, i \in \mathbf{Z}\}$ is a short-range dependent process in the sense that its autocorrelation functions are summable, Hall and Hart (1990) showed that the asymptotic distribution of the estimators of $m(\cdot)$ is asymptotic normal with the same convergence rate as in the uncorrelated case. For long-range dependent data, Csörgö and Mielniczuk (1995) and Hall and Hart (1990) established the asymptotic normality of the kernel estimator of $m(\cdot)$. Beran et al. (2003) considered the nonparametric kernel M-estimators of $m(\cdot)$ when the errors are Gaussian and long-range dependent. Guo and Koul (2007) studied the nonparametric kernel estimation when the errors are heteroscedastic and long-memory. Robinson (1994) and Ray and Tsay (1997) studied the optimal bandwidth problems and Robinson (1997) studied large sample inference for nonparametric regression. Masry (2001) considered strong consistency and convergence rates for the local linear estimator. Koul and Surgailis (2002) established asymptotic expansions for the empirical processes of long-memory moving averages models. For more information for long-range dependence, we refer to the books of Beran (1994) and Doukhan et al. (2003). For $\alpha < 2$, however, there are few results on the asymptotic properties of nonparametric estimators of $m(\cdot)$ and no result seems to be available for when the errors are heavy-tailed.

Let $a_i = i^{-\beta} l(i)$, where $l(i)$ is a slowly varying function. Similar to $\alpha = 2$, we say that $\{\varepsilon_i\}$ is short-range dependent (short-memory) or long-range dependent (long-memory) according to $\sum_{i=0}^{\infty} |a_i|^{\alpha/2} < \infty$ or $= \infty$. From this definition, it follows that for $\beta > 2/\alpha$, $\{\varepsilon_i\}$ is a short-memory process and for $1/\alpha < \beta < 2/\alpha$, $\{\varepsilon_i\}$ is a long-memory process. Recently, Peng and Yao (2004) showed that if $\{\xi_i, i \in \mathbf{Z}\}$ are symmetric stable random variables, that is,

$$E \exp\{it\xi_i\} = \exp\{-|t|^\alpha\},$$

and $\{\varepsilon_i\}$ is a short-range dependent process, then the local least absolute deviations (LAD) estimator of $m(\cdot)$ for model (1) is asymptotic normal with the same convergence rate as in the uncorrelated case. For long-range dependent $\{\varepsilon_i\}$, the asymptotic distribution of the local LAD estimator is a stable distribution. Knight (1993) provides related results of the LAD estimation of parametric linear dynamic models with infinite variance. In contrast to the least squares estimator (LSE), the LAD estimator gives less weight to the outliers and is more efficient. However, for $\alpha < 1$, it is known that $\rho(x) = |x|^\alpha$ is the optimal choice among loss functions of the form $\rho(x) = x^\gamma$: see Davis et al. (1992). This observation suggests that LAD may not be an efficient estimator for $\alpha < 1$. On the other hand, in the presence of heavy-tailed errors, M-estimators are among the most efficient robust estimators for regression functions: see Huber (1973) and Takeuchi et al. (2002). It is therefore interesting to study the asymptotic properties of the M-estimator of $m(x)$ in (1) under long-range dependent heavy-tailed errors. In this paper, it is shown that when $\xi_i \in D(\alpha)$ with $\beta > 2/\alpha$, i.e., $\{\varepsilon_i\}$ is short-range dependent, the M-estimator of $m(\cdot)$ is asymptotic normal and the optimal bandwidth is of the order $n^{-1/5}$. However, when $1/\alpha < \beta < 2/\alpha$, the limiting distribution of the M-estimator of $m(\cdot)$ becomes stable. Similar results were given by Koul and Surgailis (2001) for multiple linear regression models with long-memory and heavy-tailed errors when $1/\alpha < \beta < 1$. In the sequel, C denotes an unspecified positive and finite constant, which may vary in each appearance.

2 Local linear M-estimator

For any fixed x_0, define the local linear M-estimator of $m(x)$ and its derivative $m'(x)$ at x_0 as $\widehat{m}(x_0) = \widehat{a}$ and $\widehat{m}'(x_0) = \widehat{b}$, where $\widehat{a}$ and $\widehat{b}$ are defined by

$$(\widehat{a}, \widehat{b}) = \operatorname{argmin}_{a,b} \sum_{i=1}^{n} \rho(Y_i - a - b(x_i - x_0)) K\left(\frac{x_i - x_0}{h}\right),$$

$K(\cdot) \geq 0$ is a kernel function, h is a bandwidth. In the sequel, we assume the following conditions.

1. $m(\cdot)$ has continuous second derivatives in a neighborhood of x_0.
2. The kernel $K(\cdot)$ is a symmetric, bounded function with $\int K(x)\,dx = 1$ and has a compact support, say $[-1, 1]$.
3. $\rho(\cdot)$ is a convex function and has a right continuous, bounded derivative function $\varphi(\cdot)$, almost everywhere.
4. $E\varphi(\varepsilon_0) = 0$ and $E\varphi(\varepsilon_0 + x) = Ax + o(x)$ as $x \to 0$.
5. The density $p(x)$ of ξ_i satisfies the inequalities that for some $\varepsilon > \max\{0, \alpha - 1\}$, $|p'(x)| \leq C(1 + |x|)^{-(1+\varepsilon)}$ for all $x \in \mathbf{R}$ and $|p'(x) - p'(y)| \leq C|x - y|(1 + |x| + |y|)^{-(1+\varepsilon)}$ for all $x, y \in \mathbf{R}$ with $|x - y| < 1$.

Under these conditions, we have the following theorems.

Theorem 1 *For $\beta > \alpha/2$,*

$$\sqrt{nh}\begin{pmatrix} \widehat{m}(x_0) - m(x_0) \\ h(\widehat{m}'(x_0) - m'(x_0)) \end{pmatrix} - \frac{1}{2}m''(x_0)h^2\sqrt{nh}\begin{pmatrix} \int_{-1}^{1} u^2 K(u)\,du \\ 0 \end{pmatrix} + o(h^{5/2}) \longrightarrow^{d} N(0, \Sigma),$$

where Σ is a finite positive covariance matrix of

$$\frac{\sum_{i=1}^{n}\varphi(\varepsilon_i)K(\frac{x_i-x_0}{h})}{A\sqrt{nh}} \quad \text{and} \quad \sum_{i=1}^{n}\frac{(x_i-x_0)\varphi(\varepsilon_i)K(\frac{x_i-x_0}{h})}{A\sqrt{nh}h\int_{-1}^{1}u^2K(u)\,\mathrm{d}u}.$$

In particular,

$$\sqrt{nh}\left(\widehat{m}(x_0)-m(x_0)-\frac{1}{2}\ddot{m}(x_0)h^2\int_{-1}^{1}u^2K(u)\,\mathrm{d}u\right)\longrightarrow^{d} N(0,\sigma^2), \tag{2}$$

where

$$\sigma^2 = E\left(\frac{\sum_{i=1}^{n}\varphi(\varepsilon_i)K(\frac{x_i-x_0}{h})}{A\sqrt{nh}}\right)^2 < \infty.$$

Remark 1 If $\rho(x)=|x|$, then elementary computations show that $A=2p(0)$ and the conclusion of Peng and Yao (2004) for short-range dependent data follows from (2).

Theorem 2 *For $\alpha > 1$ and $1/\alpha < \beta < 1$, then*

$$\frac{(nh)^{\beta-1/\alpha}}{l(nh)L_\varepsilon(nh)}\begin{pmatrix}\widehat{m}(x_0)-m(x_0)\\ h(\widehat{m}'(x_0)-m'(x_0))\end{pmatrix} - \frac{m''(x_0)h^2(nh)^{\beta-1/\alpha}}{2l(nh)L_\varepsilon(nh)}\begin{pmatrix}\int_{-1}^{1}u^2K(u)\,du\\ 0\end{pmatrix} + o(h^{2+\beta-1/\alpha})$$
$$\longrightarrow^{d}\begin{pmatrix}\int_R\int_{-1}^{1}(y-x)_+^{-\beta}k(y)\,\mathrm{d}y\,\mathrm{d}Z_\alpha(x)\\ h\int_R\int_{-1}^{1}y(y-x)_+^{-\beta}k(y)\,\mathrm{d}y\,\mathrm{d}Z_\alpha(x)\end{pmatrix},$$

where $Z_\alpha(x)$ and $L_\varepsilon(x)$ are defined in Sect. 1.

Let $Z_{\alpha\beta}(x)$ and $Z^*_{\alpha\beta}(x)$ be two independent Levy-stable processes of index $\alpha\beta$ with Lévy measures $\nu(\mathrm{d}x)=\alpha C_+I(x>0)|x|^{-\alpha-1}\,\mathrm{d}x$ and $\nu(\mathrm{d}x)=\alpha C_-I(x<0)|x|^{-\alpha-1}\,\mathrm{d}x$ respectively. Let $\varphi_\infty(x)=E\varphi(x+\varepsilon_0)$ and denote $c_1=\frac{1}{\beta}\int_0^\infty\varphi_\infty(a)C_0^{1/\beta}a^{-1-1/\beta}\,\mathrm{d}a$, $c_2=\frac{1}{\beta}\int_0^\infty\varphi_\infty(-a)C_0^{1/\beta}a^{-1-1/\beta}\,\mathrm{d}a$. We have the following theorems.

Theorem 3 *Suppose that $a_i i^{-\beta}\to C_0$, $\lim_{x\to\infty}P(\xi>x)/P(|\xi|>x)=1/2$, then for $\alpha>1$ and $1<\beta<2/\alpha$,*

$$\frac{(nh)^{1-\frac{1}{\alpha\beta}}}{L'_\varepsilon(nh)}\begin{pmatrix}\widehat{m}(x_0)-m(x_0)\\ h(\widehat{m}'(x_0)-m'(x_0))\end{pmatrix}$$

$$-\frac{m''(x_0)h^2(nh)^{1-\frac{1}{\alpha\beta}}}{2L'_\varepsilon(nh)}\begin{pmatrix}\int_{-1}^1 u^2K(u)\,du\\ 0\end{pmatrix}+o(h^{3-\frac{1}{\alpha\beta}})$$

$$\longrightarrow^d \begin{pmatrix} c_1\int_{-1}^1 K(x)\,\mathrm{d}Z_{\alpha\beta}(x)+c_2\int_{-1}^1 K(x)\,\mathrm{d}Z^*_{\alpha\beta}(x)\\ c_1h\int_{-1}^1 xK(x)\,\mathrm{d}Z_{\alpha\beta}(x)+c_2h\int_{-1}^1 xK(x)\,\mathrm{d}Z^*_{\alpha\beta}(x)\end{pmatrix}$$

where $L'_\varepsilon(nh)$ *is a slowly varying function that satisfies*

$$n^{1/(\alpha\beta)}L'_\varepsilon(n)=\inf\{x: P(\xi_1^{1/\beta}\geq x)\leq 1/n\}.$$

Remark 2 By taking $\rho(x)=|x|, \alpha_j/j^{-\beta}\to C_0$, from Theorems 2 and 3, we arrive at the conclusion of Peng and Yao (2004) for long-range dependent cases. Furthermore, the symmetric condition of $K(\cdot)$ in the theorems is not necessary.

3 Auxiliary lemmas

Lemma 1 *Under the condition of Theorem 1, we have*

$$E[\rho(\varepsilon_1+a)-\rho(\varepsilon_1)]=a^2A/2+o(a^2)\quad \text{as } a\to 0.$$

Proof By the convexity of ρ,

$$\sum_{i=1}^n \frac{a}{n}\varphi\left(\varepsilon_1+\frac{i-1}{n}a\right)\leq\sum_{i=1}^n E\left[\rho\left(\varepsilon_1+\frac{ia}{n}\right)-\rho\left(\varepsilon_1+\frac{(i-1)a}{n}\right)\right]$$
$$\leq\sum_{i=1}^n\frac{a}{n}\varphi\left(\varepsilon_1+\frac{ia}{n}\right).$$

From the condition $E\varphi(\varepsilon_1+x)=Ax+o(x)$ as $x\to 0$, it follows that

$$\sum_{i=1}^n\frac{(i-1)a^2A}{n^2}\leq E[\rho(\varepsilon_1+a)-\rho(\varepsilon_1)]\leq\sum_{i=1}^n\frac{ia^2A}{n^2},$$

i.e.,

$$\frac{Aa^2n(n-1)}{2n^2}\leq E[\rho(\varepsilon_1+a)-\rho(\varepsilon_1)]\leq\frac{Aa^2n(n+1)}{2n^2}.$$

Letting $n\to\infty$, we obtain the desired conclusion. □

Lemma 2 *If* $E|X|^\alpha<\infty$, *then for any* $q_1\geq\alpha\geq q_2\geq 0$, *as* $M\to\infty$,

$$E[|X/M|^{q_1}I(|X|\leq M)]+E[|X/M|^{q_2}I(|X|>M)]=O(1/M^\alpha).$$

Proof See Lemma 1 of Wu (2003). □

Similar to the argument in Lemma 4.2 of Koul and Surgailis (2001), we have the following results.

Lemma 3 *Under condition 5, the density functions $f(x)$, $f_{ij}(x)$ of ε_0 and $\overline{X}_{ij} = \sum_{k=j+1}^{i} a_{i-k}\xi_k$ respectively, have the following properties:*

(i) $|f'(x)| + |f'_{ij}(x)| \le C(1+|x|)^{-(1+\varepsilon)}$,

(ii) $|f'(x) - f'(y)| + |f'_{ij}(x) - f'_{ij}(y)| \le C|x-y|(1+|x|)^{-(1+\varepsilon)}$

for all $x, y \in \mathbf{R}$, $|x-y| \le 1$.

Let $b_i = m(x_i) - m(x_0) - m'(x_0)(x_i - x_0)$, $X_i = (1, \frac{x_i - x_0}{h})'$, $K_i = K((x_i - x_0)/h)$ and $\widehat{\theta} = \{m(x_0) - \widehat{m}(x_0), h(m'(x_0) - \widehat{m}'(x_0))\}(nh)^d$, where

$$d = \begin{cases} 1/2 & \text{for } \beta > 2/\alpha, \\ \beta - 1/\alpha & \text{for } 1/\alpha < \beta < 1, \\ 1 - 1/(\alpha\beta) & \text{for } 1 < \beta < 2/\alpha. \end{cases}$$

The next lemma generalizes Lemma 4.3 of Koul and Surgailis (2001) to all the three cases of d as prescribed in the preceding equation. Note that Lemma 4.3 of Koul and Surgailis (2001) only deals with the case $1/\alpha < \beta < 1$. Although it is conceivable that the method given in Koul and Surgailis (2001) can be modified to deal with this general case, we adopt a different approach given in Wu (2003) to establish the next lemma. In this way, we can use the moment's inequality in Lemma 2 to directly obtain $E|M_{nj}|^v$ for any $1 \le v \le 2$. Specifically, we have the following lemma.

Lemma 4 *Suppose that $h = O(n^{-d/(2+d)})$, under the conditions of Theorem 1, for any $C < \infty$, we have*

$$\sup_{|\theta|\le C} \Bigg| \sum_{i=1}^{n} (nh)^{2d-1} \Big[\rho(\varepsilon_i + b_i + \theta X_i/(nh)^d) - \rho(\varepsilon_i + b_i) - \theta X_i/(nh)^d \varphi(\varepsilon_i + b_i) - \frac{1}{2} A \Big(\theta X_i/(nh)^d\Big)^2 \Big] K((x_i - x_0)/h) \Bigg| \longrightarrow^p 0,$$

where $\theta = (\theta_1, \theta_2) = (m(x_0) - a, h(m'(x) - b))(nh)^d \in \mathbf{R}^2$.

Proof The main ideas of this proof is to adopt an approach of Wu (2003) together with that of Koul and Surgailis (2001). Let $g(\varepsilon_i, s) = \varphi(\varepsilon_i + b_i + s) - \varphi(\varepsilon_i + b_i)$, $\mathcal{F}_j = \sigma(\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots, \xi_{j-1}, \xi_j)$ be the σ-algebra generated by $\{\xi_k : -\infty \le k \le j\}$. Then,

$$h(\varepsilon_i) := \rho\left(\varepsilon_i + b_i + \frac{\theta X_i}{(nh)^d}\right) - \rho(\varepsilon_i + b_i) - \frac{\theta X_i}{(nh)^d}\varphi(\varepsilon_i + b_i) = \int_0^{\theta X_i/(nh)^d} \varphi(\varepsilon_i + b_i + s) - \varphi(\varepsilon_i + b_i)\,\mathrm{d}s = \int_0^{\theta X_i/(nh)^d} g(\varepsilon_i, s)\,\mathrm{d}s.$$

Thus,

$$
\begin{aligned}
h(\varepsilon_i) - Eh(\varepsilon_i) &= \sum_{j=-\infty}^{i} \Big(E[h(\varepsilon_i|\mathcal{F}_j) - h(\varepsilon_i|\mathcal{F}_{j-1})] \Big) \\
&= \sum_{j=-\infty}^{i} \int_0^{\theta X_i/(nh)^d} E(g(\varepsilon_i, s)|\mathcal{F}_j) - E(g(\varepsilon_i, s)|\mathcal{F}_{j-1})\, \mathrm{d}s
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{i=1}^{n}\{h(\varepsilon_i) - Eh(\varepsilon_i)\}K_i &= \sum_{i=1}^{n}\sum_{j=-\infty}^{i} K_i \int_0^{\theta X_i/(nh)^d} E(g(\varepsilon_i, s)|\mathcal{F}_j) - E(g(\varepsilon_i, s)|\mathcal{F}_{j-1})\, \mathrm{d}s \\
&= \sum_{j=-\infty}^{n}\sum_{i=1\vee j}^{n} \left[\int_0^{\theta X_i/(nh)^d} E(g(\varepsilon_i, s)|\mathcal{F}_j) - E(g(\varepsilon_i, s)|\mathcal{F}_{j-1})\mathrm{d}s\right] K_i \\
&=: \sum_{j=-\infty}^{n} M_{nj}.
\end{aligned}
$$

Note that $\{M_{nj}, j \le n\}$ is a martingale difference sequence. From Bahr–Esseen's well-known inequality for martingales, it follows that for any $1 \le v \le 2$,

$$
E\left|\sum_{j=-\infty}^{n} M_{nj}\right|^{v} \le 2 \sum_{j=-\infty}^{n} E|M_{nj}|^{v}. \tag{3}
$$

By Minkowski's inequality, the right-hand side of (3) is less than

$$
2\sum_{j=-\infty}^{n}\left\{\sum_{i=1\vee j}^{n}\left[E\left|\left(\int_0^{\theta X_i/(nh)^d} E(g(\varepsilon_i, s)|\mathcal{F}_j) - E(g(\varepsilon_i, s)|\mathcal{F}_{j-1})\, \mathrm{d}s\right)K_i\right|^{v}\right]^{1/v}\right\}^{v}. \tag{4}
$$

Because $K(\cdot)$ is supported on $[-1, 1]$ and bounded, it follows that there is a constant M_0 such that $K_i \le M_0$ and K_i is not zero when $-1 \le (i/n - x_0)/h \le 1$, i.e., $n(x_0 - h) \le i \le n(x_0 + h)$. By (4),

$$
\begin{aligned}
E\left|\sum_{j=-\infty}^{n} M_{nj}\right|^{v} \le 2M_0 \sum_{j=-\infty}^{n}\Bigg\{\sum_{i=[n(x_0-h)]}^{[n(x_0+h)]}\Bigg[E\Bigg|\int_0^{\theta X_i/(nh)^d} & E(g(\varepsilon_i, s)|\mathcal{F}_j) \\
& - E(g(\varepsilon_i, s)|\mathcal{F}_{j-1})\, \mathrm{d}s)\Bigg|^{v}\Bigg]^{1/v}\Bigg\}^{v}.
\end{aligned} \tag{5}
$$

Let $\underline{X}_{ij} = \sum_{k=-\infty}^{j-1} a_{i-k}\xi_k$ and $\overline{X}_{ij}$ be defined as in Lemma 3. Let $Eg(x + \overline{X}_{ij}, s) = g_{ij}(x, s)$ and $a_{i-j}(\xi_j - \xi'_j) = \Delta$, where $\{\xi'_j, j \in \mathbf{Z}\}$ is an independent copy of

$\{\xi_j, j \in \mathbf{Z}\}$. Then,

$$
\begin{aligned}
&E\left|\int_0^{\theta X_i/(nh)^d} E(g(\varepsilon_i, s)|\mathcal{F}_j) - E(g(\varepsilon_i, s)|\mathcal{F}_{j-1})\,\mathrm{d}s\right|^v \\
&= E\left|\int_0^{\theta X_i/(nh)^d} E\left[\left(g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j, s) - g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j', s)\right)|\mathcal{F}_j\right]\mathrm{d}s\right|^v \\
&\le \int_0^{\theta X_i/(nh)^d} E|g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j, s) - g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j', s)|^v\,\mathrm{d}s \\
&= \int_0^{\theta X_i/(nh)^d} E\Big[\Big(|g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j, s) - g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j', s)|^v I(|\Delta| \le 1)\Big) \\
&\quad + \Big(E|g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j, s) - g_{ij}(\underline{X}_{ij} + a_{i-j}\xi_j', s)|^v I(|\Delta| > 1)\Big)\Big]\,\mathrm{d}s \\
&=: \int_0^{\theta X_i/(nh)^d} (J_1 + J_2)\,\mathrm{d}s. \qquad (6)
\end{aligned}
$$

For simplicity, write

$$h_{ij}(x, y, t, z, u) := \varphi(x + a_{i-j}y + z + b_i + us) - \varphi(x + a_{i-j}t + z + b_i + us)$$

and let $h(x, 1) = \sup_{|y|\le 1}|h(x + y)|$ be the maximal function of $h(\cdot)$. Let $p_{ij}(x)$, $p_j(y)$, $f_j(t)$ and f_{ij} be the density functions of $\underline{X}_{ij}$, ξ_j, ξ_j' and $\overline{X}_{ij}$ respectively. Then,

$$
\begin{aligned}
J_1 &= \int_R\int_R\int_R\left|\int_R \frac{1}{a_{i-j}(y-t)}(h_{ij}(x, y, t, z, 1) - h_{ij}(x, y, t, z, 0))a_{i-j}(y-t)\right. \\
&\quad \left.\times I(|a_{i-j}(y-t)| \le 1)f_{ij}(z)\,\mathrm{d}z\right|^v p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&= \int_R\int_R\int_R\left|\int_R [\varphi(x + z + b_i + s) - \varphi(x + z + b_i)]\left[\frac{1}{a_{i-j}(y-t)}\right.\right. \\
&\quad \left.\left.\times (f_{ij}(z - a_{i-j}y) - f_{ij}(z - a_{i-j}t))\right]\mathrm{d}z\right|^v |a_{i-j}(y-t)|^v \\
&\quad \times I(|a_{i-j}(y-t)| \le 1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t. \qquad (7)
\end{aligned}
$$

Note that $\varphi(\cdot)$ is a bounded function and there is an M such that $|\varphi(x)| \leq M$ for all $x \in R$. By (7) and Lemma 3,

$$\begin{aligned}
J_1 &= \int_R\int_R\int_R\bigg|\int_R \varphi(x+z+b_i)\bigg[\frac{1}{a_{i-j}(y-t)}(f_{ij}(z-a_{i-j}y-s)-f_{ij}(z-a_{i-j}y) \\
&\quad -(f_{ij}(z-a_{i-j}t-s)-f_{ij}(z-a_{i-j}t)))\bigg]\mathrm{d}z\bigg|^v |a_{i-j}(y-t)|^v \\
&\quad \times I(|a_{i-j}(y-t)|\leq 1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&= \int_R\int_R\int_R\bigg|\int_R \varphi(x+z+b_i)\int_0^s\frac{1}{a_{i-j}(y-t)}(f'_{ij}(z-a_{i-j}y-u)-f'_{ij}(z-a_{i-j}t-u)) \\
&\quad \times \mathrm{d}u\,\mathrm{d}z\bigg|^v |a_{i-j}(y-t)|^v I(|a_{i-j}(y-t)|\leq 1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&\leq \int_R\int_R\int_R\bigg|\int_R \varphi(x+z+b_i)\int_0^s(1+|z-u|)^{-(1+\varepsilon)}\,\mathrm{d}u\,\mathrm{d}z\bigg|^v |a_{i-j}(y-t)|^v \\
&\quad \times I(|a_{i-j}(y-t)|\leq 1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&\leq \int_R\int_R\int_R\left(Ms\int_R(1+|z|)^{-(1+\varepsilon)}\,\mathrm{d}z\right)^v |a_{i-j}(y-t)|^v I(|a_{i-j}(y-t)|\leq 1) \\
&\quad \times p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&\leq C(Ms)^v E|a_{i-j}(\xi_j-\xi'_j|^v I(|a_{i-j}\left(\xi_j-\xi'_j\right)|\leq 1)\leq C(Ms)^v|a_{i-j}|^{\alpha'}, \qquad (8)
\end{aligned}$$

for all $\alpha' < \alpha$ and any $v \geq \max\{1, \alpha'\}$ by Lemma 2. Furthermore,

$$\begin{aligned}
J_2 &= \int_R\int_R\int_R\bigg|\int_R \varphi(x+z+b_i)\bigg[f_{ij}(z-a_{i-j}y-s) \\
&\quad -f_{ij}(z-a_{i-j}y)-(f_{ij}(z-a_{i-j}t-s) \\
&\quad -f_{ij}(z-a_{i-j}t))]\,\mathrm{d}z\bigg|^v I(|a_{i-j}(y-t)|>1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&= \int_R\int_R\int_R\bigg|\int_R \varphi(x+z+b_i)\int_0^s f'_{ij}(z-a_{i-j}y-u)-f'_{ij}(z-a_{i-j}t-u)\,\mathrm{d}u\,\mathrm{d}z\bigg|^v \\
&\quad \times I(|a_{i-j}(y-t)|>1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&\leq \int_R\int_R\int_R\left[\int_R 2Msf'_{ij}(z,1)\,\mathrm{d}z\right]^v I(|a_{i-j}(y-t)|>1)p_{ij}(x)p_j(y)f_j(t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&=: 2|Ms|^v\left[\int_R f'_{ij}(z,1)\mathrm{d}z\right]^v EI(|a_{i-j}(\xi_j-\xi_j)'|>1)\leq C(Ms)^v|a_{i-j}|^{\alpha'}, \qquad (9)
\end{aligned}$$

by virtue of Lemma 2 again. In view of (8) and (9),

$$J_1 + J_2 \leq C|\theta X_i/(nh)^d|^v|a_{i-j}|^{\alpha'}.$$

By (5), (6) and $|s| \le |\theta X_i/(nh)^d|$, we have for $\alpha' < \alpha$ and $v \ge \max\{1, \alpha\}$,

$$E\left|\sum_{j=-\infty}^{n} M_{nj}\right|^v \le \frac{C}{(nh)^{d(v+1)}} \sum_{j=-\infty}^{[n(x_0+h)]} \left[\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} |a_{i-j}|^{\alpha'/v}\right]^v. \tag{10}$$

For $\beta > 2/\alpha$, by (10) and taking $v = 2$ and α' near α such that $\beta\alpha' > 2$, we have

$$\begin{aligned}
E\left|\sum_{j=-\infty}^{n} M_{nj}\right|^v &\le \frac{C}{(nh)^{3/2}} \sum_{j=-\infty}^{[n(x_0+h)]} \left[\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} |a_{i-j}|^{\alpha'/2}\right]^2 \\
&\le \frac{C}{(nh)^{3/2}} \left\{ \sum_{j=-\infty}^{n(x_0-2h)} \left[\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} |a_{i-j}|^{\alpha'/2}\right]^2 \right. \\
&\quad + \sum_{j=n(x_0-2h)}^{[n(x_0-h)]} \left[\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} |a_{i-j}|^{\alpha'/2}\right]^2 \\
&\quad \left. + \sum_{j=[n(x_0-h)]}^{[n(x_0+h)]} \left[\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} |a_{i-j}|^{\alpha'/2}\right]^2 \right\} \\
&\le \frac{C}{(nh)^{3/2}} \left[\int_{-\infty}^{n(x_0-2h)} \left[\int_{n(x_0-h)}^{n(x_0+h)} \left((x-y)^{-\beta} L(x-y)^{-\beta} L(x-y)\right)^{\alpha'/v} \mathrm{d}x\right]^v \mathrm{d}y + nh\right] \\
&\le C(nh)^{-3/2}[(nh)^{-\alpha'\beta+v+1} L(nh)^{\alpha'} + nh] \\
&\le 2C(nh)^{-1/2}.
\end{aligned} \tag{11}$$

Note that $b_i = O(h^2) \to 0$ and $h = O(n^{-d/(2d+1)})$. By condition 4 and Lemma 1, we have

$$\sum_{i=1}^{n} Eh(\varepsilon_i)K_i = \sum_{i=1}^{n} (A/2)(\theta X_i/(nh)^d)^2 K_i. \tag{12}$$

Combining this with (11) leads to the conclusion of Lemma 4 for the case of $\beta > 2/\alpha$.

For $1 < \beta < 2/\alpha$ and $\alpha > 1$, by (10) and taking v with $\alpha \le v \le \alpha\beta$, α' near to α, a similar argument to the case of $\beta > 2/\alpha$ leads to

$$\sum_{j=-\infty}^{[n(x_0+h)]} \left[\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} |a_{i-j}|^{\alpha'/v}\right]^v = O(nh).$$

Therefore, by (10), we have

$$
\begin{aligned}
E\left|\sum_{i=1}^{n}(nh)^{2d-1}\{h(\varepsilon_i)-Eh(\varepsilon_i)\}K_i\right|^{v} &\le C(nh)^{(2d-1)v-d(v+1)}O(nh)\\
&\le C(nh)^{(d-1)v-d+1}=C(nh)^{-(v-1)/(\alpha\beta)}\\
&\to 0.
\end{aligned}
$$

The conclusion follows from (12).

For $1/\alpha<\beta<1$, elementary computations show

$$
\begin{aligned}
&\sum_{j=-\infty}^{[n(x_0+h)]}\left\{\sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]}|a_{i-j}|^{\alpha'/v}\right\}^{v}\\
&\le C\left\{\int_{-\infty}^{n(x_0-h)}\left(\int_{n(x_0-h)}^{n(x_0+h)}[(x-y)^{-\beta}L(x-y)]^{\alpha'}\mathrm{d}x\right)^{v}\mathrm{d}y\right.\\
&\quad\left.+\int_{n(x_0-h)}^{n(x_0+h)}\left(\int_{y}^{[n(x_0+h)]}[(x-y)^{-\beta}L(x-y)]^{\alpha'}\mathrm{d}x\right)^{v}dy\right\}\\
&\le C(nh)^{-\alpha'\beta+v+1}L(nh)^{\alpha'}.
\end{aligned}
$$

Therefore, by (10) and taking $v=\alpha>\alpha'$ and α' near α such that $\beta(\alpha-\alpha')<\beta-1/\alpha$, we have

$$
\begin{aligned}
E\left|\sum_{i=1}^{n}(nh)^{2d-1}\{h(\varepsilon_i)-Eh(\varepsilon_i)\}K_i\right|^{v} &\le C(nh)^{(2d-1)v-d(v+1)}C(nh)^{-\alpha'\beta+v+1}L(nh)^{\alpha'}\\
&\le C(nh)^{(d-1)v-d+1}=C(nh)^{-(v-1)(1-\beta+1/\alpha)}\\
&\to 0\,.
\end{aligned}
$$

The conclusion follows from (12) and the proof of Lemma 4 is completed. □

Similar to the argument of Lemma 4, we have

Lemma 5 *Under the conditions of Lemma 4,*

$$
(nh)^{d-1}\left[\sum_{i=1}^{n}[(\varphi(\varepsilon_i+b_i)-\varphi(\varepsilon_i))-E\varphi(\varepsilon_i+b_i)]K_i\right]\longrightarrow^{p}0.
$$

Remark 3 Intuitively, from these two lemmas we see that if $\{\varepsilon_i\}$ is short-range dependent then $d=1/2$, and so the optimal bandwidth is of the order $n^{-d/(2+d)}=n^{-1/5}$. However, as $\{\varepsilon_i\}$ is a long-memory process, the optimal bandwidth is completely different: it depends on the values of both α and β.

The following two lemmas address the uniform reduction principle for the weighted empirical distribution of $\{\varepsilon_i\}$.

Lemma 6 *If* $1/\alpha < \beta < 1$ *and* $1 < \alpha < 2$, *then for any* $\varepsilon > 0$,

$$P\left\{\sup_{x\in\mathbf{R}} \frac{1}{n^{1+1/\alpha-\beta}}\left|\sum_{i=1}^{n} v_i(I(\varepsilon_i \le x) - F(x) + f(x)\varepsilon_i)\right| \ge \varepsilon\right\} \le Cn^{-\tau}$$

for some $\tau > 0$, *where* $\{v_i, 1 \le i \le n\}$ *are non-random real-valued sequences with* $\max_{1\le i\le n} |v_i| = O(1)$, $F(x)$ *and* $f(x)$ *are the distribution and density of* ε_0 *respectively.*

Proof The proof is similar to that of Koul and Surgailis (2001). Although $a_i = i^{-\beta} l(i)$, because $l(x)$ is a slowly varying function, their conclusion still follows and the details are omitted. □

Lemma 7 *If* $1 < \beta < 2/\alpha$, $1 < \alpha < 2$ *and for any* $\varepsilon > 0$,

$$P\left\{\sup_{x\in\mathbf{R}} \frac{1}{n^{1/(\alpha\beta)}}\left(\left|\sum_{i=1}^{n} v_i(I(\varepsilon_i \le x) - F(x) - \eta(\xi_i, x))\right|\right) \ge \varepsilon\right\} \le Cn^{-\tau}$$

for some $\tau > 0$, *where* $\{v_i, 1 \le i \le n\}$ *is defined in Lemma 6 and*

$$\eta(\xi_i, x) = \sum_{j=1}^{\infty}(F(x - a_j\xi_j) - EF(x - a_j\xi_j)),$$

$F(x)$ *is the distribution of* ε_0.

Proof By Lemma 2.3 of Surgailis (2002), there is a $\tau > 0$ such that

$$P\left\{\sup_{x\in\mathbf{R}} \frac{1}{n^{1/(\alpha\beta)}}\left(\left|\sum_{i=1}^{n} I(\varepsilon_i \le x) - F(x) - \eta(\xi_i, x)\right|\right) \ge \varepsilon\right\} \le Cn^{-\tau}. \quad (13)$$

Note that $\max_{1\le i\le n} |v_i| = O(1)$, by (13), the conclusion of Lemma 7 follows. □

4 Proofs of main results

Proof of Theorem 1. By Lemma 4 and using the argument of Pollard (1991) (see also Peng and Yao 2004), we have that the difference between the minimum of

$$\sum_{i=1}^{n}[\rho(\varepsilon_i + b_i + \theta X_i/(nh)^{1/2}) - \rho(\varepsilon_i + b_i)]K_i$$

and the minimum of

$$\sum_{i=1}^{n}\left[\frac{\theta X_i}{\sqrt{nh}}\varphi(\varepsilon_i+b_i)+\frac{A(\theta X_i)^2}{2nh}\right]K_i$$
$$=\sum_{i=1}^{n}\left[\frac{\theta X_i}{\sqrt{nh}}\varphi(\varepsilon_i+b_i)\right]K_i+\frac{A}{2}\int_{-1}^{1}(\theta_1+\theta_2 u)^2K(u)\,\mathrm{d}u$$
$$=\sum_{i=1}^{n}\left[\frac{\theta X_i}{\sqrt{nh}}\varphi(\varepsilon_i+b_i)\right]K_i+\frac{A}{2}\int_{-1}^{1}(\theta_1^2+\theta_2^2u^2)K(u)\,\mathrm{d}u$$

converges to 0 in probability. By the definition of $\widehat{\theta}$, we see $\widehat{\theta}$ is just the minimum of

$$\sum_{i=1}^{n}[\rho(\varepsilon_i+b_i+\theta X_i/(nh)^{1/2})-\rho(\varepsilon_i+b_i)]K_i.$$

This implies that

$$\sqrt{nh}(\widehat{m}(x_0)-m(x_0))=\frac{\sum_{i=1}^n\varphi(\varepsilon_i+b_i)K_i}{A\sqrt{nh}\int_{-1}^1K(u)\,\mathrm{d}u}+o_p(1),$$
$$h\sqrt{nh}(\widehat{m}'(x_0)-m'(x_0))=\frac{\sum_{i=1}^n\left(\frac{x_i-x_0}{h}\right)\varphi(\varepsilon_i+b_i)K_i}{A\sqrt{nh}\int_{-1}^1u^2K(u)\,\mathrm{d}u}+o_p(1).$$

Note that $\int_{-1}^1K(u)\,\mathrm{d}u=1$. By Lemma 5 and condition 4, it follows that

$$\sqrt{nh}(\widehat{m}(x_0)-m(x_0))=\frac{\sum_{i=1}^n\varphi(\varepsilon_i)K_i}{A\sqrt{nh}}+\frac{1}{2}\sqrt{nh}m''(x_0)h^2\int_{-1}^1u^2K(u)\,\mathrm{d}u+o_p(1),$$
$$h\sqrt{nh}(\widehat{m}'(x_0)-m'(x_0))=\frac{\sum_{i=1}^n\left(\frac{x_i-x_0}{h}\right)\varphi(\varepsilon_i)K_i}{A\sqrt{nh}\int_{-1}^1u^2K(u)\,\mathrm{d}u}+\frac{1}{2}\sqrt{nh}m''(x_0)h^2\frac{\int_{-1}^1u^3K(u)\,\mathrm{d}u}{\int_{-1}^1u^2K(u)\,\mathrm{d}u}+o_p(1).$$

From these two equalities, proving Theorem 1 is equivalent to showing that

$$\frac{\sum_{i=1}^n\varphi(\varepsilon_i)K_i}{A\sqrt{nh}}\longrightarrow^d N(0,\sigma^2). \tag{14}$$

We proceed with a method based on the limit theory for additive functions of Markov chains: see Wu (2003). Let $g(\varepsilon_i)=\varphi(\varepsilon_i)K_i$, $g_n(x)=E[g(\varepsilon_i)|\underline{X}_{n,1}+a_{n-1}\varepsilon_1=x]=: Eg(x+\overline{X}_{n,1})$ and $f_{n1}(z)$ be the density of $\overline{X}_{n,1}$, where $\underline{X}_{n,1}$ and $\overline{X}_{n,1}$ are defined as those in Lemma 4. From the proof of Theorem 1 of Wu (2003), to show (14) it is sufficient to show that

(i) $Eg^2(\varepsilon_i) = E(\varphi(\varepsilon_i)K_i)^2 < \infty$,
(ii) $\sup_{x\in\mathbf{R}} |g_n(x)| < \infty$,
(iii) $\sup_{x\in\mathbf{R}} \sup_{y\neq x:|x-y|\leq 1} |g_n(x) - g_n(y)|/|y-x| < \infty$.

Note that both $\varphi(\cdot)$ and $K(\cdot)$ are bounded functions. It follows that

$$Eg^2(\varepsilon_i) = E(\varphi(\varepsilon_i)K_i)^2 < \infty$$

and

$$\sup_{x\in\mathbf{R}} |g_n(x)| = \sup_{x\in\mathbf{R}} Eg(x+\overline{X}_{n,1}) = \sup_{x\in\mathbf{R}} \int_{y\in\mathbf{R}} \varphi(x+z)K_i f_{n1}(z)\,\mathrm{d}z \leq MM_0.$$

Furthermore, by condition 5,

$$\begin{aligned}
&\sup_x \sup_{y\neq x:|x-y|\leq 1} \frac{|g_n(x)-g_n(y)|}{|y-x|} \\
&\quad = \sup_x \sup_{y\neq x:|x-y|\leq 1} \frac{1}{|y-x|} \int_{\mathbf{R}} |(\varphi(x+z)-\varphi(y+z))K_i| f_{n1}(z)\,\mathrm{d}z \\
&\quad = \sup_x \sup_{y\neq x:|x-y|\leq 1} \frac{1}{|y-x|} \int_{\mathbf{R}} |\varphi(z)K_i| \cdot |f_{n1}(z-x) - f_{n1}(z-y)|\,\mathrm{d}z \\
&\quad \leq \sup_x \sup_{y\neq x:|x-y|\leq 1} \int_{\mathbf{R}} |\varphi(z+x)K_i|(1+|z|)^{-(1+\varepsilon)}\,\mathrm{d}z < \infty.
\end{aligned}$$

The proof of Theorem 1 is completed. □

Proof of Theorem 2. Similarly, by Lemma 4 and the argument of Pollard (1991), we have for $1/\alpha < \beta < 1$, the difference between the minimum of

$$\sum_{i=1}^n n^{2d-1}[\rho(\varepsilon_i + b_i + \theta X_i/(nh)^d) - \rho(\varepsilon_i + b_i)]K_i$$

and the minimum of

$$\begin{aligned}
&\sum_{i=1}^n n^{2d-1}\left[\frac{\theta X_i}{(nh)^d}\varphi(\varepsilon_i+b_i) + \frac{A(\theta X_i)^2}{2(nh)^{2d}}\right]K_i \\
&\quad = \sum_{i=1}^n \left[\frac{\theta X_i}{(nh)^{1-d}}\varphi(\varepsilon_i+b_i) + \frac{A(\theta X_i)^2}{2nh}\right]K_i \\
&\quad = \sum_{i=1}^n \left[\frac{\theta X_i}{(nh)^{1-d}}\varphi(\varepsilon_i+b_i)\right]K_i + \frac{A}{2}\int_{-1}^1 (\theta_1+\theta_2 u)^2 K(u)\,\mathrm{d}u \\
&\quad = \sum_{i=1}^n \left[\frac{\theta X_i}{(nh)^{1-d}}\varphi(\varepsilon_i+b_i)\right]K_i + \frac{A}{2}\int_{-1}^1 (\theta_1^2+\theta_2^2 u^2) K(u)\,\mathrm{d}u
\end{aligned}$$

converging to 0 in probability. By the definition of $\widehat{\theta}$, we see $\widehat{\theta}$ is the minimum of

$$\sum_{i=1}^{n} n^{2d-1}[\rho(\varepsilon_i + b_i + \theta X_i/(nh)^d) - \rho(\varepsilon_i + b_i)]K_i.$$

Therefore, by Lemma 5 and condition 4, we have

$$\begin{aligned}(nh)^d(\widehat{m}(x_0) - m(x_0)) &= \frac{\sum_{i=1}^n \varphi(\varepsilon_i + b_i)K_i}{A(nh)^{1-d}\int_{-1}^1 K(u)\,du} + o_p(1)\\ &= \frac{\sum_{i=1}^n \varphi(\varepsilon_i)K_i}{A(nh)^{1-d}} + \frac{(nh)^d m''(x_0)h^2}{2}\int_{-1}^1 u^2K(u)\,\mathrm{d}u + o_p(1).\end{aligned} \tag{15}$$

Observe that

$$\begin{aligned}&h(nh)^d(\widehat{m}'(x_0) - m'(x_0))\\ &= \left[\sum_{i=1}^n \left(\frac{x_i - x_0}{h}\right)\varphi(\varepsilon_i + b_i)K_i\right] \Big/ \left[A(nh)^{1-d}\int_{-1}^1 u^2K(u)\,\mathrm{d}u\right] + o_p(1)\\ &= \frac{\sum_{i=1}^n \left(\frac{x_i - x_0}{h}\right)\varphi(\varepsilon_i)K_i}{A(nh)^{1-d}\int_{-1}^1 u^2K(u)\,\mathrm{d}u} + \frac{(nh)^d m''(x_0)h^2}{2}\frac{\int_{-1}^1 u^3K(u)\,\mathrm{d}u}{\int_{-1}^1 u^2K(u)\,\mathrm{d}u} + o_p(1).\end{aligned} \tag{16}$$

Next, we show for $1/\alpha < \beta < 1$,

$$\sum_{i=1}^n \frac{(\varphi(\varepsilon_i) - A\varepsilon_i)K_i}{(nh)^{1+1/\alpha-\beta}} \longrightarrow^p 0. \tag{17}$$

Because $\varphi(\cdot)$ is bounded, similar to the argument of Koul and Surgailis (2001), we can represent

$$\sum_{i=1}^n \frac{(\varphi(\varepsilon_i) - A\varepsilon_i)K_i}{(nh)^{1+1/\alpha-\beta}} \quad \text{as} \quad \sum_{i=1}^\infty \frac{\left[\int_{\mathbf{R}} I(\varepsilon_i \le x) - F(x) - f(x)\varepsilon_i\,\mathrm{d}\varphi(x)\right]K_i}{(nh)^{1+1/\alpha-\beta}}. \tag{18}$$

By noting that $K(\cdot)$ is supported on $[-1, 1]$ and $K(x) \le M_0$ for all $x \in \mathbf{R}$, it follows that the right-hand side of (18) is smaller than

$$\begin{aligned}&\frac{1}{(nh)^{1+1/\alpha-\beta}}\left(M_0\left|\int_{\mathbf{R}}\sum_{i=[n(x_0-h)]}^{[n(x_0+h)]} I(\varepsilon_i \le x) - F(x) - f(x)\varepsilon_i\,d\varphi(x)\right|\right)\\ &= \frac{1}{(nh)^{1+1/\alpha-\beta}}\left(M_0\left|\int_{\mathbf{R}}\sum_{i=1}^{2nh} I(\varepsilon_i \le x) - F(x) - f(x)\varepsilon_i\,\mathrm{d}\varphi(x)\right|\right) =: S_{nh}.\end{aligned}$$

By Lemma 6, we have for any $\delta > 0$,

$$
\begin{aligned}
S_{nh} &\le M_0|\varphi(\infty) - \varphi(-\infty)| \sup_{x\in\mathbf{R}} \frac{1}{(nh)^{1+1/\alpha-\beta}}\left(\left|\sum_{i=1}^{n}[I(\varepsilon_i \le x) - F(x) - f(x)\varepsilon_i]\right|\right) \\
&\le C \sup_{x\in\mathbf{R}} \frac{1}{(nh)^{1+1/\alpha-\beta}}\left(\left|\sum_{i=1}^{n}[I(\varepsilon_i \le x) - F(x) - f(x)\varepsilon_i]\right|\right) \\
&\le \delta \quad \text{in probability,}
\end{aligned}
$$

which yields (17). In the following, we show that

$$
\begin{aligned}
&\left(\sum_{i=1}^{n} \frac{\varepsilon_i K_i}{(nh)^{1+1/\alpha-\beta} l(nh) L_\varepsilon(nh)}, \sum_{i=1}^{n} \frac{\left(\frac{x_i - x_0}{h}\right)\varepsilon_i K_i}{(nh)^{1+1/\alpha-\beta} l(nh) L_\varepsilon(nh)}\right) \\
&\longrightarrow^d \left(\int_{-\infty}^{\infty}\int_{-1}^{1} K(y)(y-x)_+^{-\beta}\,\mathrm{d}y\,\mathrm{d}Z_\alpha(x), \int_{-\infty}^{\infty}\int_{-1}^{1} yK(y)(y-x)_+^{-\beta}\,\mathrm{d}y\,\mathrm{d}Z_\alpha(x)\right).
\end{aligned} \tag{19}
$$

Let $c_n = n^{1/\alpha} L_\varepsilon(n)$. Then,

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{\varepsilon_i K_i}{(nh)^{1+1/\alpha-\beta} l(nh) L_\varepsilon(nh)} \\
&= \sum_{i=1}^{n} \frac{\sum_{j=-\infty}^{i} a_{i-j}\xi_j K_i}{(nh)^{1+1/\alpha-\beta} l(nh) L_\varepsilon(nh)} \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=1\vee j}^{n} \frac{K((x_i - x_0)/h)(i-j)^{-\beta} l(i-j)}{(nh)^{1+1/\alpha-\beta} l(nh) L_\varepsilon(nh)} \xi_j \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=[n(x_0-h)]\vee j}^{[n(x_0+h)]} \frac{K((x_i - x_0)/h)(i-j)^{-\beta} l(i-j)}{(nh)^{1-\beta} l(nh)} \frac{\xi_j}{c_{nh}} \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=[-nh]\vee j}^{[nh)]} \frac{K(\frac{i}{nh})(i-j)^{-\beta} l(i-j)}{(nh)^{1-\beta} l(nh)} \frac{\xi_j}{c_{nh}},
\end{aligned} \tag{20}
$$

in distribution. Similarly,

$$
\sum_{i=1}^{n} \frac{\left(\frac{x_i - x_0}{h}\right)\varepsilon_i K_i}{(nh)^{1+1/\alpha-\beta} l(nh) L_\varepsilon(nh)}
$$

has the same distribution as

$$\sum_{j=-\infty}^{\infty} \sum_{i=[-nh]\vee j}^{[nh)]} \frac{\frac{i}{nh} K(\frac{i}{nh})(i-j)^{-\beta} l(i-j)}{(nh)^{1-\beta} l(nh)} \frac{\xi_j}{c_n}. \tag{21}$$

By (20), (21) and Theorem 4 of Avram and Taqqu (1986) (see also Kasahara and Maejima 1988), we have (19). Note that in this case $d = \beta - 1/\alpha$, by (15), (16), (17) and (19), we have the conclusion of Theorem 2. □

Proof of Theorem 3. Similar to the argument for the case of $1/\alpha < \beta < 1$, we have for $1 < \beta < 2/\alpha$, the conclusions of (15) and (16) are also true, but in this case $d = 1 - 1/(\alpha\beta)$. Next, we show that

$$\frac{1}{(nh)^{1/(\alpha\beta)}} \sum_{i=1}^{n} \left(\varphi(\varepsilon_i) - \int_{\mathbf{R}} \eta(\xi_i, x)\, \mathrm{d}\varphi(x) \right) K_i \longrightarrow^p 0. \tag{22}$$

Because $E\varphi(\varepsilon_i) = 0$,

$$\begin{aligned} &\sum_{i=1}^{n} \left(\varphi(\varepsilon_i) - \int_{\mathbf{R}} \eta(\xi_i, x) \mathrm{d}\varphi(x) \right) \\ &= \sum_{i=1}^{n} \left(\int_{\mathbf{R}} [I(\varepsilon_i \le x) - F(x)]\, \mathrm{d}\varphi(x) - \int_{\mathbf{R}} \eta(\xi_i, x)\, \mathrm{d}\varphi(x) \right) \\ &= \sum_{i=1}^{n} \int_{\mathbf{R}} [I(\varepsilon_i \le x) - F(x) - \eta(\xi_i, x)]\, \mathrm{d}\varphi(x). \end{aligned}$$

Combining this fact with Lemma 7, it follows that for any $\delta > 0$,

$$\begin{aligned} &\frac{1}{(nh)^{1/(\alpha\beta)}} \sum_{i=1}^{n} \left(\varphi(\varepsilon_i) - \int_{\mathbf{R}} \eta(\xi_i, x)\, \mathrm{d}\varphi(x) \right) K_i \\ &\le \frac{1}{(nh)^{1/(\alpha\beta)}} \left(|\varphi(\infty) - \varphi(-\infty)| \sup_{x \in \mathbf{R}} \left| \sum_{i=1}^{n} I(\varepsilon_i \le x) - F(x) - \eta(\xi_i, x) \right| K_i \right) \\ &\le 2M \sup_{x \in \mathbf{R}} \left| \sum_{i=1}^{n} I(\varepsilon_i \le x) - F(x) - \eta(\xi_i, x) \right| K_i / (nh)^{1/(\alpha\beta)} \le \delta. \end{aligned}$$

By the arbitrariness of δ, we have (22). We now show that

$$S_n := \frac{1}{(nh)^{1/(\alpha\beta)} L_\varepsilon(nh)'} \sum_{i=1}^{n} K_i \int_{\mathbf{R}} \eta(\xi_i, x)\, \mathrm{d}\varphi(x)$$
$$= \frac{1}{(nh)^{1/(\alpha\beta)} L_\varepsilon(nh)'} \sum_{i=1}^{n} K_i \int_{\mathbf{R}} \varphi(x) \left[\sum_{j=1}^{\infty} \big(f(x - a_j\xi_i) - Ef(x - a_j\xi_i) \big) \right] \mathrm{d}x, \tag{23}$$

where $f(x)$ is the density of ε_0. Let $H(x) = \sum_{j=1}^{\infty} [\varphi_\infty(a_j x) - E\varphi_\infty(a_j\xi_i)]$ and $A_{nh} = (nh)^{1/(\alpha\beta)} L_\varepsilon(nh)'$. Then from (23), we have

$$\begin{aligned} S_n &= \frac{1}{A_{nh}} \sum_{i=1}^{n} \sum_{i=1}^{n} K_i \int_{\mathbf{R}} \varphi(x) \left[\sum_{j=1}^{\infty} \big(f(x - a_j\xi_i) - Ef(x - a_j\xi_i) \big) \right] \mathrm{d}x \\ &= \sum_{i=1}^{n} \sum_{j=1}^{\infty} [\varphi_\infty(a_j\xi_i) - E\varphi_\infty(a_j\xi_i)] K_i / A_{nh} \\ &= \sum_{i=1}^{n} H(\xi_i) K_i / A_{nh}. \end{aligned}$$

Set $\zeta_i^1 = c_1 \xi_i^{1/\beta} I(\xi_i > 0)$, $\zeta_i^2 = c_2(-\xi_i)^{1/\beta} I(\xi_i < 0)$. We next show that

$$\sum_{i=1}^{n} H(\xi_i) K_i / A_{nh} = \sum_{i=1}^{n} (\zeta_i^1 + \zeta_i^2) K_i / A_{nh} \quad \text{in probability.} \tag{24}$$

Because $\varphi_\infty(0) = E\varphi(\varepsilon_0) = 0$,

$$\begin{aligned} \varphi_\infty(x) &= \int_{\mathbf{R}} (\varphi(x + y) - \varphi(y)) f(y)\, \mathrm{d}y \\ &= \int_{\mathbf{R}} \varphi(y) \int_0^x f'(y - a)\, \mathrm{d}a\, \mathrm{d}y \\ &= \int_0^x \int_{\mathbf{R}} f'(y) \varphi(y + a)\, \mathrm{d}y\, \mathrm{d}a \le Cx. \end{aligned} \tag{25}$$

On the other hand, because $|\varphi(x)| \le M$ for all $x \in \mathbf{R}$, we have

$$\varphi_\infty(x) = \int_{\mathbf{R}} (\varphi(x + y) - \varphi(y)) f(y)\, \mathrm{d}y \le 2M. \tag{26}$$

Therefore, by (25) and (26),

$$|\varphi_\infty(x)| \le \max\{2M, C\}\min\{1, x\} =: C\min\{1, x\}.$$

This implies that for $z > 1$,

$$\sum_{j=1}^{\infty} |\varphi_\infty(a_j z)| \le \sum_{j=1}^{\infty} \min\{1, a_j z\} \le C|z|^{1/\beta}. \tag{27}$$

From (27), we have

$$H(x) = \sum_{j=1}^{\infty} \varphi_\infty(a_j x) - O(CE|\xi_i|^{1/\beta}). \tag{28}$$

Let $H_1(x) = \int_0^\infty \varphi_\infty(C_0 t^{-\beta} x)\,\mathrm{d}t$ and $H_2(x) = \sum_{j=1}^\infty \varphi_\infty(a_j x) - H_1(x)$, where C_0 is given in the condition of a_i in Theorem 3. As $x > 0$,

$$H_1(x) = \frac{1}{\beta}\int_0^\infty \varphi_\infty(a)(C_0 x)^{1/\beta} a^{-1-1/\beta}\,\mathrm{d}a,$$

which yields

$$x^{-1/\beta} H_1(x) = \frac{1}{\beta}\int_0^\infty \varphi_\infty(a) C_0^{1/\beta} a^{-1-1/\beta}\,\mathrm{d}a = c_1. \tag{29}$$

By argument (8) of Surgailis (2002), we have

$$\lim_{x\to\infty} x^{-1/\beta} H_2(x) = 0. \tag{30}$$

By (28), (29) and (30), we have

$$\lim_{x\to\infty} x^{-1/\beta} H(x) = c_1.$$

Similarly, we also have

$$\lim_{x\to-\infty} (-x)^{-1/\beta} H(x) = c_2 = \frac{1}{\beta}\int_0^\infty \varphi_\infty(-a) C_0^{1/\beta} a^{-1-1/\beta}\,\mathrm{d}a.$$

Therefore, for any $\varepsilon > 0$ there is a constant $B > 0$ such that for any $x > B$,

$$(c_1 - \varepsilon)x^{1/\beta} \le g(x) \le (c_1 + \varepsilon)x^{1/\beta} \tag{31}$$

and for any $x < -B$,

$$(c_2 - \varepsilon)(-x)^{1/\beta} \le g(x) \le (c_2 + \varepsilon)(-x)^{1/\beta}. \tag{32}$$

Furthermore, because $H(\xi_i), i = 1, 2, \ldots, n$ are i.i.d. random variables with mean zeros, it follows that for any $\varepsilon > 0$,

$$\begin{aligned}
E\left(\frac{1}{A_{nh}}\sum_{i=1}^{n} H(\xi_i)I(|\xi_i| \le B)K_i\right)^2 &= \frac{1}{A_{nh}^2}\sum_{i=1}^{n} E(H(\xi_i)I(|\xi_i| \le B)K_i)^2 \\
&\le \frac{M_0^2}{A_{nh}^2}\sum_{i=[n(x_0-h)]}^{[n(x_0+h)]} E[H(\xi_i)I(|\xi_i| \le B)]^2 \\
&\le (nh)^{1+\varepsilon-2/(\alpha\beta)} E[H(\xi_i)I(|\xi_i| \le B)]^2 \\
&= (nh)^{1+\varepsilon-2/(\alpha\beta)} \int_{-B}^{B} H^2(x)f(x)\,\mathrm{d}x \\
&\le C(nh)^{1+\varepsilon-2/(\alpha\beta)} \int_{-B}^{B} x^{2/\beta} f(x)\,\mathrm{d}x \to 0
\end{aligned}$$

by (26) and the fact that $\alpha\beta < 2$. Similarly, we have

$$E\left(\sum_{i=1}^{n}\left[c_1\xi_i^{1/\beta}I(0 < \xi_i < B) + c_2(-\xi_i)^{1/\beta}I(-A < \xi_i < 0)\right]/A_{nh}\right)^2 \to 0.$$

Consequently, by virtue of (31) and (32), we have (24). Note also that

$$\sum_{i=1}^{n}\xi_i^{1/\beta}I(\xi_i > 0)/A_{nh} \longrightarrow^d Z_{\alpha\beta}(1),$$
$$\sum_{i=1}^{n}(-\xi_i)^{1/\beta}I(\xi_i < 0)/A_{nh} \longrightarrow^d Z_{\alpha\beta}^*(1).$$

Combining these with (23), (24) and Theorem 4 of Avram and Taqqu (1986) again, we have

$$\frac{\sum_{i=1}^{n} K_i \int_{\mathbf{R}} \eta(\xi_i, x)\,\mathrm{d}\varphi(x)}{(nh)^{1/(\alpha\beta)}L_\varepsilon(nh)^{1/\beta}} \longrightarrow^d c_1\int_{-1}^{1} K(x)\,\mathrm{d}Z_{\alpha\beta}(x) + c_2\int_{-1}^{1} K(x)\,\mathrm{d}Z_{\alpha\beta}^*(x).$$

By (22),

$$\frac{\sum_{i=1}^{n} K_i\varphi(\varepsilon_i)}{(nh)^{1/(\alpha\beta)}L_\varepsilon(nh)^{1/\beta}} \longrightarrow^d c_1\int_{-1}^{1} K(x)\,\mathrm{d}Z_{\alpha\beta}(x) + c_2\int_{-1}^{1} K(x)\,\mathrm{d}Z_{\alpha\beta}^*(x).$$

Therefore, by (15) and (16) with $d = 1 - 1/(\alpha\beta)$, we establish the conclusion of Theorem 3. □

Acknowledgements The authors thank the two referees for helpful comments and references. This research was supported in part by HKSAR-RGC Grants CUHK-4043/02P, 400305 and 400306.

References

Avram, F., Taqqu, M. S. (1986). Weak convergence of moving averages with infinite variance. In E. Eberlein, & M.S. Taqqu (Eds.), *Dependence in probability and statistics*, pp. 399–415. Boston: Birkhäuser.

Beran, J., Ghosh, S., Sibbertsen, P. (2003). Nonparametric M-estimation with long-memory errors. *Journal of Statistical Planning and Inference, 117*, 199–205.

Beran, J. (1994). *Statistics for long-memory processes*. New York: Champman and Hall.

Chiu, C. K., Marron, J. S. (1991). Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics, 19*, 1906–1918.

Csörgő, S., Mielniczuk, J. (1995). Nonparametric regression under long-dependent normal errors. *The Annals of Statistics, 23*, 1000–1014.

Davis, R. A., Knight, K., Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications*, *40*, 145–180.

Doukhan, P., Oppenheim, G., Taqqu, M. S. (Eds.) (2003). *Theory and application of long-range dependence*. Boston: Birkhäuser.

Guo, H. W., Koul, H. L. (2007). Nonparametric regression with heteroscedastic long-memory errors. *Journal of Statistical Planning and Inference, 137*, 379–404.

Hall, P., Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and their Applications*, *36*, 339–351.

Huber, P. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, *1*, 799–821.

Kasahara, Y., Maejima, M. (1988). Weighted sums of i.i.d. random variables attracted to integrals of stable processes. *Probability Theory and Related Fields, 78*, 75–96.

Knight, K. (1993). Estimation in dynamic linear regression models with infinite variance errors. *Econometric Theory*, *9*, 570–588.

Koul, H. L., Surgailis, D. (2001). Asymptotics of empirical processes of long-memory moving averages with infinite variance. *Stochastic Processes and their Applications, 91*, 309–336.

Koul, H. L., Surgailis, D. (2002). Asymptotics expansion of the empirical process of long-memory moving averages. In H. Dehling, T. Mikosch, & M. Soreson (Eds.), *Empirical process technique for dependent data*, pp. 213–239. Boston: Birkhäuser.

Masry, E. (2001). Local linear regression estimation under long-range dependence: strong consistency and rates. *IEEE Transaction on Information Theory, 47*, 2863–2875.

Peng, L., Yao, Q. W. (2004). Nonparametric regression under dependent errors with infinite variance. *Annals of the Institute of Statistical Mathematics, 56*, 73–86.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, *7*, 186–198.

Ray, B. K., Tsay, R. S. (1997). Bandwidth selection for kernel regression for long-range dependence. *Biometrika, 84*, 791–802.

Robinson, P. M. (1994). Rates of convergence and optimal bandwidth choice for long-range dependence. *Probability Theory and Related Fields, 99*, 443–473.

Robinson, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *The Annals of Statistics, 25*, 2054–2083.

Surgailis, D. (2002). Stable limits of empirical processes of moving averages with infinite variance. *Stochastic Processes and their Applications, 100*, 255–274.

Takeuchi, I., Bengio, Y., Kanamori, T. (2002). Robust regression with asymmetric heavy-tail noise distribution. *Neural Computation, 14*, 2469–2496.

Wu, W. B. (2003). Additive functionals of infinite-variance moving averages. *Statistica Sinica, 13*, 1259–1267.