Nonparametric Regression Under Dependent Errors With Infinite Variance

Liang Peng
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160
Email: peng@math.gatech.edu

Qiwei Yao
Department of Statistics
London School of Economics
London, WC2A 2AE, UK
Email: q.yao@lse.ac.uk

Abstract

We consider local least absolute deviation (LLAD) estimation for trend functions of time series with heavy tails which are characterised via a symmetric stable law distribution. The setting includes both causal stable ARMA model and fractional stable ARIMA model as special cases. The asymptotic limit of the estimator is established under the assumption that the process has either short or long memory autocorrelation. For a short memory process, the estimator admits the same convergence rate as if the process has the finite variance. The optimal rate of convergence $n^{-2/5}$ is obtainable by using appropriate bandwidths. This is distinctly different from local least squares estimation, of which the convergence is slowed down due to the existence of heavy tails. On the other hand, the rate of convergence of the LLAD estimator for a long memory process is always slower than $n^{-2/5}$ and the limit is no longer normal.

Keywords: ARMA, fractional ARIMA, heavy tail, least absolute deviation estimation, long memory, median, stable distribution, time series.

Short title. Nonparametric regression with infinite variances

1 Introduction and Models

A substantial literature now exists on using kernel-type smoothers to estimate a smooth trend in time series data. These methods are important, in part, because they allow estimation of a smooth trend without prior specification of the form of the trend. A popular setting which has attracted much attention in the last decade is a fixed-design regression with dependent 'errors', under which the observations $Y_1, \ldots, Y_n$ follow the model

$$Y_t = m(t/n) + \varepsilon_t, \quad t = 1, \ldots n, \tag{1.1}$$

where $m(.)$ is a smooth function defined on [0,1], and $\{\varepsilon_t\}$ is a stationary process such as ARMA time series. If $\{\varepsilon_t\}$ are correlated but with only short-range dependence in the sense that its autocorrelation functions are absolutely summable, it has been proved that nonparametric regression estimators for $m(.)$ are asymptotically normal at the same convergence rate as in the case of uncorrelated $\{\varepsilon_t\}$, although the asymptotic variances have one more factor due to the dependence in the data; see Hall and Hart (1990). When there exists a long-range dependence, Hall and Hart (1990) shows that the estimators have a *slower* convergence rate for the long-range dependent normal errors; see also Csörgő and Mielniczuk (1995). Since the mean squared errors of the estimators are different from those with independent data, research has been carried out to modify the standard kernel regression techniques, including the bandwidth selection procedures, to incorporate various dependence structures. This includes Altman (1990), Chiu and Marron (1991), Truong (1991), Hart (1991, 1994), Herrmann, Gasser and Kneip (1992), Roussas, Tran and Ioannides (1992), Tran, Roussas, Yakowits and Truong Van (1996) for short-range dependence data, and Ray and Tsay (1997) and Robinson (1994, 1997) for long-range dependence data. A common practice in all the aforementioned literature is to assume that ε_t has the zero-mean and a finite variance.

In this paper, we also deal with the kernel regression estimation for function $m(.)$ but with *heavy tailed* error terms such that $E(\varepsilon_t^2) = \infty$ or $E|\varepsilon_t| = \infty$. More specifically, we assume that in model (1.1) ε_t is a linear process defined as

$$\varepsilon_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \tag{1.2}$$

where $\{Z_t\}$ are independent random variables sharing the same standard symmetric stable law distribution with index $\alpha \in (0, 2)$, i.e. the characteristic function Z_t has the form

$$E(e^{itZ_t}) = \exp\{-|t|^\alpha\}.$$

Under the condition

$$0 < \sum_{j=0}^{\infty} |c_j|^\alpha < \infty, \tag{1.3}$$

the infinite sum in (1.2) is well-defined. Moreover we may say that $\{\varepsilon_t\}$ has short memory or long memory according as $\sum_{j=0}^{\infty} |c_j|^{\alpha/2} < \infty$ or $= \infty$ respectively. Note that $E|\varepsilon_t| = \infty$ when $\alpha \le 1$ and $E(\varepsilon_t^2) = \infty$ when $\alpha < 2$, and $m(t/n)$ is the median (as well as mean when $\alpha > 1$) of Y_t. The setting (1.2) – (1.3) includes the causal stable ARMA model (Mikosch et al. 1995, and Klüppelberg and Mikosch 1996) and the causal fractional stable ARIMA model (Kokoszka and Taqqu, 1995, 1996) as special cases. A causal infinite variance fractional ARIMA(p, d, q) time series may be defined as

$$\Phi(B)\varepsilon_t = \Theta(B)(1 - B)^{-d} Z_t, \tag{1.4}$$

where $d \in (0, 1)$ is a self-similarity parameter, B denotes the backshift operator, $\Phi(\cdot)$ and $\Theta(\cdot)$ are polynomials with degrees p and q respectively, and all the roots of the equation

$\Phi(z) = 0$ are outside the unit circle. Because of the presence of d, the process $\{\varepsilon_t\}$ has infinite variance as well as long-range dependence. It in fact admits the MA(∞)-representation (1.2) with $\sum_{j=0}^{\infty} |c_j|^{\alpha/2} = \infty$. For further information on ARMA models with heavy tails and their applications, we refer to Adler, Feldman and Gallagher (1997) and Resnick (1997).

It is known that for regression models with heavy tailed noises, the conventional least squares estimators typically have slow convergence rates, are only consistent when the tail index $\alpha \in (1, 2)$, and even then the limiting distribution is non-normal; see Davis, Knight and Liu (1992) and references within for parametric regression, and Hall, Peng and Yao (2002) for nonparametric regression. We consider in this paper the local linear least absolute deviations estimator for $m(\cdot)$. The asymptotic limit of the estimator is established for both short and long memory cases (Theorem 2.1 in Section 2 below). The limit is normal in case of short memory, however it is a stable law in case of long memory. The proof is based on a combined use of the convex lemma (Pollard 1991) and the asymptotic results for stable moving average processes of Hsing (1999), Koul and Surgailis (2001) and Surgailis (2002). When $\{\varepsilon_t\}$ has short memory, the convergence rate as well as the first order asymptotic mean and variance are the same as if ε_t had a finite variance. We can reproduce the optimal rate of convergence $n^{-2/5}$ by choosing the bandwidth of the order $n^{-1/5}$, which is a folklore in conventional (one-dimensional) kernel regression. In this sense, the least absolute deviations estimation is adaptive to heavy tails. However when the process has a long memory, the convergence rate is always slower than $n^{-2/5}$ and the limit is no longer normal.

There is a substantial literature on nonparametric regression in the least absolute deviation setting. Mallows (1980), Velleman (1980), Truong (1989) and Fan and Hall (1994) addressed local median smoothing for independent data, Tsybakov (1986) and Fan, Hu and Truong (1994) developed robust methods for fitting local polynomials. In the time series context, Truong and Stone (1992) and Truong (1991, 1992a,b) discussed robust nonparametric regression for random-design models, Yao and Tong (1994) suggested robust conditional quantile estimation. All of them considered regression models with random designs and none of them addressed estimation with infinite-variance data. Hall, Peng and Yao (2002) considered nonparametric least squares as well as least absolute deviations estimation for heavy tailed regressive models with random design under the assumption that the processes fulfill certain mixing conditions which rule out the possibility of long memory properties.

2 Estimators and Main Results

Let $x_t = t/n$ for $t = 1, \ldots, n$ and $x \in (0,1)$ fixed. The local linear least absolute deviations estimator is defined as $\hat{m}(x) = \hat{a}$, where

$$(\hat{a}, \hat{b}) = \underset{(a,b)}{\operatorname{argmin}} \sum_{t=1}^{n} |Y_t - a - b(x_t - x)| K\left(\frac{x_t - x}{h}\right).$$

In the above expression, $K(\cdot) \geq 0$ is a density function on R^1 and $h > 0$ is a bandwidth. We write $\hat{m}_1(x) = \hat{b}$ which is an estimator for $\dot{m}(x) \equiv \frac{d}{dx} m(x)$.

In the sequel, we always assume that $x \in (0,1)$ is fixed. Let $p(\cdot)$ denote the marginal density function of ε_t, $\sigma_0^2 = \int u^2 K(u) du$, and $D(\xi) = I(\xi > 0) - I(\xi \leq 0)$. We use C to denote some generic constant which may be different at different places.

(C1) For fixed x, $m(\cdot)$ has second continuous derivative in a neighbourhood of x.

(C2) The kernel K is a symmetric, bounded and non-negative function with support $[-1, 1]$. Further $|K(z_1) - K(z_2)| \leq C|z_1 - z_2|$ for any $z_1, z_2 \in R^1$.

(C3) $h = h(n) \to 0$ and $nh^3 \to \infty$ as $n \to \infty$.

Theorem 2.1. Let Conditions (C1) – (C3) hold for the process defined in (1.1) – (1.3).

(i) If $\sum_{l=0}^{\infty} |c_l|^{\alpha/2} < \infty$, then

$$\sqrt{nh}\{\hat{m}(x) - m(x) - \frac{1}{2}h^2\sigma_0^2\ddot{m}(x)\} \xrightarrow{d} N(0, \sigma^2),$$

where $\ddot{m}(x) = \frac{d^2}{dx^2}m(x)$, $\varepsilon_{i,j} = \sum_{l=0}^{j} c_l Z_{i-l}$, and

$$\sigma^2 = \lim_{j\to\infty} E\{\frac{1}{2\sqrt{nh}\ p(0)} \sum_{i=1}^{n} D(\varepsilon_{i,j})K(\frac{x_i - x}{h})\}^2.$$

The limit on the RHS of the above expression exists and is finite.

(ii) Suppose $c_j/j^{-\beta} \to b_0$ as $j \to \infty$ for some $\beta \in (\alpha^{-1}, 1)$. Then

$$(nh)^{\beta-1/\alpha}\{\hat{m}(x) - m(x) - \frac{1}{2}h^2\sigma_0^2\ddot{m}(x)\} \xrightarrow{d} L_\alpha,$$

where L_α is a stable law with characteristic function

$$Ee^{itL_\alpha} = \exp\{-|t|^\alpha b_0^\alpha \int_{-\infty}^{-1} (\int_{-1}^{1} K(u)(u-v)^{-\beta}\, du)^\alpha\, dv\}.$$

(iii) Suppose $c_j/j^{-\beta} \to b_0$ as $j \to \infty$ for some $\beta \in (1, 2/\alpha)$. Let G denote the distribution function of ε_i. Then

$$(nh)^{1-1/(\alpha\beta)}\{\hat{m}(x) - m(x) - \frac{1}{2}h^2\sigma_0^2\ddot{m}(x)\} \xrightarrow{d} \frac{c^+}{p(0)}L_{\alpha\beta}^+ + \frac{c^-}{p(0)}L_{\alpha\beta}^-,$$

where

$$c^\pm = \sigma^*(\int_{-1}^{1} K(s)\, ds)(\int_0^\infty (G_\infty(\pm t) - G_\infty(0))t^{-1-1/\beta}\, dt),$$

$$G_\infty(x) = EG(x + Z_i),$$

$$\sigma^* = \{\frac{b_0^\alpha(\alpha\beta - 1)}{\Gamma(2-\alpha\beta)\cos\frac{\pi\alpha\beta}{2}\beta^{\alpha\beta}}\}^{1/(\alpha\beta)},$$

and $L_{\alpha\beta}^+$ and $L_{\alpha\beta}^-$ are independent copies of a stable law $L_{\alpha\beta}$ with characteristic function

$$Ee^{itL_{\alpha\beta}} = \exp\{-|t|^{\alpha\beta}(1 - isgn(t)tan\frac{\pi\alpha\beta}{2})\}.$$

Remark. (i) If $\{\varepsilon_t\}$ has the short range dependence, i.e. $\sum_{j=0}^{\infty} |c_j|^{\alpha/2} < \infty$, Theorem 2.1(i) indicates that the asymptotic distribution of the least absolute deviations estimator

$\hat{m}(\cdot)$ is of the same form as if ε_t had a finite variance. Note that the first order asymptotic approximation for the mean squared error of $\hat{m}(x)$ is

$$\frac{1}{4}h^4\sigma_0^4\{\ddot{m}(x)\}^2 + \frac{1}{nh}\sigma^2.$$

Minimising this approximation over h, we obtain an optimum bandwidth of the order $n^{-1/5}$, which is the same as for (one-dimensional) nonparametric regression estimation with finite variances. By using the optimum bandwidth, the estimator $\hat{m}(x)$ converges at the rate $1/\sqrt{nh} = O(n^{-2/5})$.

(ii) In Theorem 2.1(ii) and (iii) the condition $c_j \sim b_0 j^{-\beta}$ implies $\sum_{j=0}^{\infty} |c_j|^{\alpha/2} = \infty$. The asymptotic stable law has been established for the subclass of long memory processes fulfilling this condition.

3 Proofs

In this section, we always assume the regularity conditions (C1) – (C3) hold. We introduce some notation first.

Let $Y_t^* = Y_t - m(x) - \dot{m}(x)(x_t - x)$, $z_t = (1, \frac{x_t - x}{h})^T$, $K_t = K(\frac{x_t-x}{h})$, and $\widehat{\theta} = \sqrt{nh}\{\hat{m}(x) - m(x), h(\hat{m}_1(x) - \dot{m}(x))\}^T$. For $\theta = (\theta_1, \theta_2)^T$, we define

$$G(\theta) = \sum_{t=1}^{n}\{|Y_t^* - \theta^T z_t/\sqrt{nh}| - |Y_t^*|\}K_t,$$

$$R(\theta) = G(\theta) - p(0)(\theta_1^2 + \theta_2^2\sigma_0^2) + \frac{\theta^T}{\sqrt{nh}}\sum_{t=1}^{n} z_t D(Y_t^*)K_t.$$

Obviously, $\widehat{\theta}$ is the minimiser of $G(\theta)$. We split the proof into several lemmas. Lemma 3.1 below follows easily from condition (1.3) and the proof of Lemma 3 of Hsing (1999).

Lemma 3.1. The marginal density of ε_t is positive and continuous at zero.

Lemma 3.2. Let u and v be real numbers. For $0 < q \leq 1$, $|u+v|^q \leq |u|^q + |v|^q$. Further for $1 < q < \infty$, $|u+v|^q \le 2^{q-1}(|u|^q + |v|^q)$ and

$$||u+v|^q - |u|^q - |v|^q| \le |u||v|^{q-1} + |u|^{q-1}|v|.$$

Proof. The first two inequalities follow from Lemma 2.7.13 of Samorodnitsky and Taqqu (1994). The last one follows from the inequalities

$$|u+v|^q \le (|u| + |v|)|u+v|^{q-1} \le (|u| + |v|)(|u|^{q-1} + |v|^{q-1})$$

and

$$|u+v|^q \ge (|u| - |v|)(|u|^{q-1} - |v|^{q-1}) \ge ||u| - |v|| \times |u+v|^{q-1}.$$

Lemma 3.3. Let $p_t(x, y)$ be the joint density of $(\varepsilon_1, \varepsilon_t)$ for $t > 1$. It holds that $\sup_{t\ge 2} p_t(0, 0) < \infty$.

Proof. Note that the characteristic function of $(\varepsilon_1, \varepsilon_{j+1})$ is

$$\begin{aligned} f(t_1, t_2) &\equiv Ee^{i(t_1\varepsilon_1 + t_2\varepsilon_{j+1})} \\ &= E[\exp\{i(t_1 \sum_{k=0}^{\infty} c_k Z_{1-k} + t_2 \sum_{k=0}^{\infty} c_k Z_{j+1-k})\}] \\ &= E[\exp\{i \sum_{k=0}^{\infty} (t_1 c_k + t_2 c_{k+j}) Z_{1-k} + i \sum_{l=-j}^{-1} t_2 c_{l+j} Z_{1-l}\}] \\ &= \exp\{- \sum_{k=0}^{\infty} |t_1 c_k + t_2 c_{k+j}|^{\alpha} - \sum_{l=-j}^{-1} |t_2 c_{l+j}|^{\alpha}\}. \end{aligned}$$

We consider the case $\alpha \in (0, 1]$ first. By the inverse formula and Lemma 3.2, the density function $p_{j+1}(0, 0)$ is equal to

$$\begin{aligned} &\lim_{x\to 0} \frac{1}{x^2}\{P(\varepsilon_1 < x, \varepsilon_{j+1} < x) - P(\varepsilon_1 < x, \varepsilon_{j+1} < 0) - P(\varepsilon_1 < 0, \varepsilon_{j+1} < x) + P(\varepsilon_1 < 0, \varepsilon_{j+1} < 0)\} \\ &= \lim_{x\to 0} \frac{1}{4\pi^2 x^2} \int\int \{e^{-i(t_1 x + t_2 x)} - e^{-it_1 x} - e^{-it_2 x} + 1\} f(t_1, t_2)\, dt_1 dt_2 \\ &\le \int\int O(|t_1|) O(|t_2|) \exp\{- \sum_{k=0}^{\infty} |t_1 c_k + t_2 c_{k+j}|^{\alpha} - \sum_{l=-j}^{-1} |t_2 c_{l+j}|^{\alpha}\}\, dt_1 dt_2 \\ &\le \int\int O(|t_1|) O(|t_2|) \exp\{- \sum_{k=0}^{\infty} |t_1|^{\alpha} |c_k|^{\alpha} + \sum_{k=0}^{\infty} |t_2|^{\alpha} |c_{k+j}|^{\alpha} - \sum_{l=-j}^{-1} |t_2|^{\alpha} |c_{l+j}|^{\alpha}\}\, dt_1 dt_2 \\ &\le \int O(|t_1|) \exp\{-|t_1|^{\alpha} \sum_{k=0}^{\infty} |c_k|^{\alpha}\}\, dt_1 \int O(|t_2|) \exp\{-|t_2|^{\alpha} (\sum_{k=0}^{j-1} |c_k|^{\alpha} - \sum_{l=j}^{\infty} |c_l|^{\alpha})\}\, dt_2. \end{aligned}$$

Note that (1.3) implies

$$0 < \sum_{k=0}^{j-1} |c_k|^\alpha - \sum_{l=j}^{\infty} |c_l|^\alpha \le \sum_{k=0}^{\infty} |c_k|^\alpha$$

for all large j's. Hence there exists $j_0 > 0$ such that $\sup_{j \ge j_0} p_{j+1}(0,0) < \infty$.

When $\alpha \in (1,2)$, the required inequality can be derived from

$$|t_1 c_k + t_2 c_{k+j}|^\alpha \ge \frac{|t_1 c_k|^\alpha}{2} - |t_2 c_{k+j}|^\alpha$$

in a similar manner and the above relation is implied by Lemma 3.2. This completes the proof of Lemma 3.3.

Lemma 3.4. As $n \to \infty$, $E\{R(\theta)\} \to 0$.

Proof. Let $d_t = \theta^T z_t/\sqrt{nh}$. Without loss of generality, we may assume that $d_t \ge 0$. Then

(3.1)
$$\begin{aligned}
& \{|Y_t^* - \theta^T z_t/\sqrt{nh}| - |Y_t^*|\}K_t \\
= \;& K_t[Y_t^*\{I(Y_t^* > d_t) - I(Y_t^* \le d_t) - I(Y_t^* > 0) + I(Y_t^* \le 0)\} \\
+ \;& d_t\{I(Y_t^* \le d_t) - I(Y_t^* > d_t)\}] \\
= \;& -2K_t Y_t^* I(0 < Y_t^* \le d_t) \\
+ \;& d_t K_t\{I(Y_t^* \le 0) + I(0 < Y_t^* \le d_t) - I(Y_t^* > 0) + I(0 < Y_t^* \le d_t)\} \\
= \;& -d_t K_t D(Y_t^*) + 2K_t(d_t - Y_t^*) I(0 \le d_t - Y_t^* < d_t).
\end{aligned}$$

Note that $(C2)$ implies $d_t \to 0$ as $n \to \infty$. We have

$$E\{(d_t - Y_t^*)I(0 \le d_t - Y_t^* < d_t)\} = p(0) d_t^2\{1 + o(1)\}.$$

Combining the above equation with (3.1) we have

$$E\{(|Y_t^* - \theta^T z_t/\sqrt{nh}| - |Y_t^*| + d_t D(Y_t^*))K_t\} = p(0)K_t d_t^2\{1 + o(1)\}.$$

Thus

$$\begin{aligned}
E\{G(\theta) + \sum_{t=1}^{n} d_t D(Y_t^*) K_t\} &= p(0)\sum_{t=1}^{n} d_t^2 K_t\{1 + o(1)\} \\
\to \; p(0)\int_{-\infty}^{\infty} (\theta_1 + \theta_2 u)^2 K(u) du &= p(0)(\theta_1^2 + \theta_2^2\sigma_0^2),
\end{aligned}$$

since $nh \to \infty$. This completes the proof of the lemma.

Lemma 3.5. As $n \to \infty$, $R(\theta)$ converges to 0 in probability.

Proof. Note that $R(\theta) = \sum_{i=1}^n T_t - p(0)(\theta_1^2 + \theta_2^2\sigma_0^2)$ with

$$T_t = K_t[|Y_t^* - \theta^T z_t/\sqrt{nh}| - |Y_t^*| + \theta^T z_t D(Y_t^*)/\sqrt{nh}].$$

It follows from Lemma 3.4 that we only need to prove that $\sum_{t=1}^n (T_t - ET_t) \xrightarrow{P} 0$. Note that

$$\begin{aligned} & P\{|\sum_{i=1}^n (T_i - ET_i)| > \varepsilon) \\ \le\ & \frac{1}{\varepsilon^2}\sum_{i=1}^n E(T_i - ET_i)^2 + \frac{2}{\varepsilon^2}\sum_{i=1}^{n-1}(n-i)\{E(T_1T_{i+1}) - ET_1ET_{i+1}\}. \end{aligned}$$

From (3.1) we have

$$\begin{aligned} \sum_{i=1}^n ET_i^2 &= \sum_{i=1}^n 4p(0)K_i\frac{d_i^3}{3}(1+o(1)) \\ &= \frac{4}{3}p(0)\frac{1}{\sqrt{nh}}\int(\theta_1+\theta_2 u)^3K^2(u)\,du(1+o(1)) \ \to\ 0, \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^n (ET_i)^2 &= \sum_{i=1}^n p^2(0)K_i^2d_i^4(1+0(1)) \\ &= p^2(0)\frac{1}{nh}\int(\theta_1+\theta_2 u)^4K^2(u)\,du(1+o(1)) \ \to\ 0, \end{aligned}$$

and

$$\begin{aligned} & \sum_{i=1}^{n-1}(n-i)|ET_1T_{i+1}| \\ =\ & \sum_{i=1}^{n-1}(n-i)p_{i+1}(0,0)K_1d_1^2K_{i+1}d_{i+1}^2(1+o(1)) \\ \le\ & \sup_{j\ge1} p_{j+1}(0,0)nK_1d_1^2\sum_{i=1}^{n-1}K_{i+1}d_{i+1}^2(1+o(1)) \\ \le\ & \sup_{j\ge1} p_{j+1}(0,0)Ch^{-1}K(\frac{1/n-x}{h})\int(\theta_1+\theta_2 u)^2K(u)\,du(1+o(1)). \end{aligned}$$

Note that by (C2)

$$K((1/n - x - 1/(nh))/h) = 0$$

as n large enough. Hence by (C2) and (C3)

$$h^{-1}K(\frac{1/n-x}{h}) = \frac{|K((1/n-x)/h) - K((1/n-x-1/(nh))/h)|}{h} \leq C/(nh^3) \to 0.$$

Thus by Lemma 3.3

$$\sum_{i=1}^{n-1}(n-i)ET_1T_{i+1} \to 0.$$

By the same arguments as above we have

$$\begin{aligned} |\sum_{i=1}^{n-1}(n-i)ET_1ET_{i+1}| &\leq n\sum_{i=1}^{n-1}|ET_1ET_{i+1}| \\ = \quad & n\sum_{i=1}^{n-1}p^2(0)K_1d_1^2K_{i+1}d_{i+1}^2(1+o(1)) \to 0. \end{aligned}$$

Hence the lemma is proven.

Lemma 3.6. If $\sum_{l=0}^{\infty}|c_l|^{\alpha/2} < \infty$, then

$$\lim_{l\to\infty} E(D(\varepsilon_1) - D(\varepsilon_{1,l}))^2 = 0.$$

Proof. Let $W_1 = \varepsilon_{1,l}$ and $W_2 = \varepsilon_1 - W_1$. Then W_1 and W_2 are independent. It follows from the symmetric distributions of ε_1 and $\varepsilon_{1,l}$ that

$$E(D(\varepsilon_1) - D(\varepsilon_{1,l}))^2 = 2P(\varepsilon_1 > 0, \varepsilon_{1,l} < 0).$$

Let g_{W_1} and G_{W_2} denote the density of W_1 and the distribution of W_2, respectively. Then

$$P(\varepsilon_1 > 0, \varepsilon_{1,l} < 0) = \int_{-\infty}^{0} g_{W_1}(y)[1 - G_{W_2}(-y)]\,dy.$$

Note that

$$G_{W_2}(y) = P(Z_1 \leq (\sum_{j>l}|c_j|^{\alpha})^{-1/\alpha}y)$$

and g_{W_1} is uniformly bounded (see Lemma 3 of Hsing (1999)). Hence by Potter bounds (see Geluk and de Haan (1987))

$$\begin{aligned} P(\varepsilon_1 > 0, \varepsilon_{1,l} < 0) &= \int_{-\infty}^{-\delta} g_{W_1}(y)[1 - G_{W_2}(-y)]\, dy + \int_{-\delta}^{0} g_{W_1}(y)[1 - G_{W_2}(y)]\, dy \\ &= O(1 - P\{Z_1 > (\sum_{j>l} |c_j|^\alpha)^{-1/\alpha}\}) + O(\delta). \end{aligned}$$

Therefore the lemma follows by letting $l \to \infty$ first, and then $\delta \to 0$.

Lemma 3.7. If $\sum_{l=0}^{\infty} |c_l|^{\alpha/2} < \infty$, then

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} D(\varepsilon_i) K_i \to N(0, \sigma_1^2)$$

where

$$\sigma_1^2 = \lim_{l \to \infty} E\{\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} D(\varepsilon_{i,l}) K_i\}^2$$

exists and is finite.

Proof. Let $\mathcal{F}_{-\infty,l}$ be the σ-field generated by $\{Z_i, i \le l\}$. Note that

$$\begin{aligned} \sum_{i=1}^{n} D(\varepsilon_i) K_i - \sum_{i=1}^{n} D(\varepsilon_{i,l}) K_i &= \sum_{i=1}^{n} \{D(\varepsilon_i) - D(\varepsilon_{i,l})\} K_i I(|\frac{i/n - x}{h}| \le 1) \\ = \sum_{i=[n(x-h)]}^{[n(x+h)]} \{D(\varepsilon_i) - D(\varepsilon_{i,l})\} K_i &= \sum_{i=[n(x-h)]}^{[n(x+h)]} \sum_{j=1}^{\infty} K_i U_{i,j,l}, \end{aligned}$$

where

$$\begin{aligned} U_{i,j,l} &= \{E(D(\varepsilon_i)|\mathcal{F}_{-\infty,i-j}) - E(D(\varepsilon_i)|\mathcal{F}_{-\infty,i-(j+1)})\} \\ &- \{E(D(\varepsilon_{i,l})|\mathcal{F}_{-\infty,i-j}) - E(D(\varepsilon_{i,l})|\mathcal{F}_{-\infty,i-(j+1)})\} I(j \le l). \end{aligned}$$

Then the lemma follows from Lemma 3.7, the boundedness of K_i and the proof of Theorem 1 of Hsing (1999).

Lemma 3.8. Suppose $c_j/j^{-\beta} \to b_0 > 0$ as $j \to \infty$, where $\beta \in (\alpha^{-1}, 1)$. Then there exists $\delta_0 > 0$ such that for any $\delta_1 > 0$

$$P\{\sup_x (nh)^{-1+\beta-1/\alpha} |\sum_{i=1}^{n} K_i(I(\varepsilon_i \le x) - P(\varepsilon_i \le x) + p(x)\varepsilon_i)| \ge \delta_1\} = O((nh)^{-\delta_0}).$$

Proof. It is similar to the proof of Theorem 2.1 in Koul and Surgailis (2001) by replacing $I(-s_1 \le j \le n)$ and $\sum_{t=1\vee j}^{n}(1\wedge(t-j)^{-\beta(1+\gamma)})$ in Lemma 4.3 of Koul and Surgailis (2001) by $I(-s_1 \le j \le n)nh$ and $\sum_{t=1\vee j}^{n}(1\wedge(t-j)^{-\beta(1+\gamma)})I([n(x-h)] \le t \le [n(x+h)])$, respectively.

Lemma 3.9. Suppose $c_j/j^{-\beta} \to b_0$ as $j \to \infty$, where $\beta \in (\alpha^{-1}, 1)$. Then $(nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} K_i D(\varepsilon_i)$ converges in distribution to a stable law with characteristic function

$$exp\{-|t|^{\alpha}(2p(0)b_0)^{\alpha}\int_{-\infty}^{-1}(\int_{-1}^{1} K(u)(u-v)^{-\beta}\,du)^{\alpha}\,dv\}.$$

Proof. Define $c_j = 0$ if $j < 0$. Note that

$$(nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} K_i D(\varepsilon_i) = (nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} 2K_i(I(\varepsilon_i \le 0) - \frac{1}{2}),$$

$$(nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} 2p(0)K_i\varepsilon_i = (nh)^{-1-1/\alpha+\beta}\sum_{j=-\infty}^{\infty}\sum_{i=1}^{n} 2p(0)K_i c_{i-j} Z_j,$$

$$\begin{aligned}
&\textstyle\sum_{l=-\infty}^{\infty}((nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} 2p(0)K_i c_{i-j})^{\alpha}\\
=\;&\textstyle\sum_{l\le nx-\delta nh}((nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} 2p(0)K_i c_{i-j})^{\alpha}\\
&\textstyle+\sum_{l> nx-\delta nh}((nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} 2p(0)K_i c_{i-j})^{\alpha}\\
=\;&\Delta_1 + \Delta_2.
\end{aligned}$$

It is easy to check that for any $\delta > 1$

$$\Delta_1 \to (2p(0)b_0)^{\alpha}\int_{-\infty}^{-\delta}(\int_{-1}^{1} K(u)(u-v)^{-\beta}\,du)^{\alpha}\,dv$$

and

$$\Delta_2 \le \sum_{l=nx-\delta nh}^{nx+nh}(2p(0)(nh)^{-1-1/\alpha+\beta}\sup K(x)\sum_{j=0}^{\infty} c_j)^{\alpha} \to 0.$$

Hence the limiting characteristic function of $(nh)^{-1-1/\alpha+\beta}\sum_{i=1}^{n} 2p(0)K_i\varepsilon_i$ is

$$\exp\{-|t|^{\alpha}(2p(0)b_0)^{\alpha}\int_{-\infty}^{-1}(\int_{-1}^{1} K(u)(u-v)^{-\beta}\,du)^{\alpha}\,dv\}.$$

Thus Lemma 3.9 follows from Lemma 3.8.

Lemma 3.10. Suppose $c_j/j^{-\beta} \to b_0$ as $j \to \infty$, where $\beta \in (1, 2/\alpha)$. Let G denote the distribution function of ε_i. Then there exists $\delta_0 > 0$ such that for any $\delta_1 > 0$

$$P\{\sup_x |(nh)^{-1/(\alpha\beta)} \sum_{i=1}^n K_i \Big(I(\varepsilon_i \le x) - G(x) - \sum_{j=1}^\infty (G(x - b_j Z_i) - EG(x - b_j Z_i))\Big)| > \delta_1\} = O((nh)^{-\delta_0}).$$

Proof. It is similar to the proof of Lemma 2.3 of Surgailis (2002).

Lemma 3.11. Suppose $c_j/j^{-\beta} \to b_0$ as $j \to \infty$, where $\beta \in (1, 2/\alpha)$. Let G denote the distribution function of ε_i. Then

$$(nh)^{-1/(\alpha\beta)} \sum_{i=1}^n K_i D(\varepsilon_i) \xrightarrow{d} c^+ L^+_{\alpha\beta} + c^- L^-_{\alpha\beta},$$

where $c^\pm$ and $L^\pm_{\alpha\beta}$ are defined in Theorem 2.1 (iii).

Proof. It follows from Lemma 2.4 of Surgailis (2002) and Theorem 3.1 of Kasahara and Maejima (1988) that

$$(nh)^{-1/(\alpha\beta)} \sum_{i=1}^n K_i \sum_{j=1}^\infty (G(-c_j Z_i) - EG(-c_j Z_i)) \xrightarrow{d} c^+ L^+_{\alpha\beta} + c^- L^-_{\alpha\beta}.$$

Hence Lemma 3.11 follows from Lemma 3.10 and the fact that

$$\sum_{i=1}^n K_i D(\varepsilon_i) = 2 \sum_{i=1}^n K_i (I(\varepsilon_i \le 0) - \frac{1}{2}).$$

Proof of Theorem 2.1. Since $R(\theta) \xrightarrow{P} 0$, the convex function

$$G(\theta) - (nh)^{-1/2}\, \theta^T \sum_{1 \le i \le n} z_i\, D(Y_i^*)\, K_i$$

converges to $p(0)\,(\theta_1^2 + \theta_2^2 \sigma_0^2)$. By the convexity lemma (Pollard, 1991), the convergence is uniform on compact sets in R^2. Using the arguments of Pollard (1991, p. 193) we can show that the difference between the minimiser of $\widehat{\theta}$ of $G(\theta)$ and the minimiser of

$$-\frac{1}{\sqrt{nh}}\, \theta^T \sum_{i=1}^n z_i\, D(Y_i^*)\, K_i + p(0)\,(\theta_1^2 + \theta_2^2 \sigma_0^2)$$

converges to 0 in probability. This implies that

$$\sqrt{nh}\,\{\hat{m}(x) - m(x)\} = \frac{1}{2\sqrt{nh}\;p(0)} \sum_{i=1}^{n} D(Y_i^*)\,K_i + o_p(1)\,. \tag{3.2}$$

We may assume that $\ddot{m}(x) > 0$. Then

$$D(Y_t^*) = D(\varepsilon_t) + 2I\{0 < -\varepsilon_t \le \ddot{m}(x)(x_t - x)^2/2 + o(h^2)\},$$

and

$$E(I\{0 < -\varepsilon_t \le \ddot{m}(x)(x_t - x)^2/2 + o(h^2)\}) = p(0)\ddot{m}(x)(x_t - x)^2/2\{1 + o(1)\}.$$

Hence

$$\frac{1}{nh^3}\sum_{t=1}^{n} K_t E(I\{0 < -\varepsilon_t \le \ddot{m}(x)(x_t - x)^2/2\}) \to \frac{1}{2}\ddot{m}(x)p(0).$$

Let $W_i = K_i I(0 < -\varepsilon_i \le \ddot{m}(x)(x_i - x)^2/2)$. Then

$$\begin{aligned} & P\Big(\Big|\frac{1}{nh^3}\sum_{i=1}^{n}(W_i - EW_i)\Big| > \varepsilon\Big) \\ \le\ & \frac{1}{\varepsilon^2 n^2 h^6}\sum_{i=1}^{n} E(W_i - EW_i)^2 + \frac{2}{\varepsilon^2 n^2 h^6}\sum_{i=1}^{n-1}(n-i)(E(W_1 W_{i+1}) - EW_1 EW_{i+1}). \end{aligned}$$

Note that $EW_i = K_i p(0)\ddot{m}(x)(x_i - x)^2/2$. We have

$$\begin{aligned} \frac{1}{n^2h^6}\sum_{i=1}^{n}(EW_i)^2 &= \frac{1}{n^2h^6}\sum_{i=1}^{n} p^2(0)K_i^2\ddot{m}^2(x)\frac{(x_i - x)^4}{4h^4}h^4(1 + o(1)) \\ &= \frac{p(0)^2\ddot{m}^2(x)}{4n^2}\sum_{i=1}^{n}\frac{1}{n^2h^2}K_h^2(x_i - 1)^4(1 + o(1)) \\ &= \frac{p(0)^2\ddot{m}^2(x)}{4nh}\Big(\int K^2(u)u^4\,du\Big)(1 + o(1)) \;\to\; 0. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{1}{n^2h^6}\sum_{i=1}^{n} EW_i^2 &= \frac{1}{n^2h^6}\sum_{i=1}^{n} K_i^2 p(0)\ddot{m}(x)(x_i - x)^2/2(1 + o(1)) \\ &= \frac{p(0)\ddot{m}(x)}{2nh^3}\Big(\int K^2(u)u^2\,du\Big)(1 + o(1)) \;\to\; 0, \end{aligned}$$

and

$$
\begin{aligned}
& \frac{1}{n^2h^6}\sum_{i=1}^{n-1}(n-i)EW_1EW_{i+1} \\
= \quad & \frac{1}{n^2h^6}\sum_{i=1}^{n-1}K_1K_{i+1}p_{i+1}(0,0)\ddot{m}^2(x)\frac{(x_1-1)^2}{2}\frac{(x_{i+1}-x)^2}{2}(1+o(1)) \\
\leq \quad & \sup_{i\geq 1}p_{i+1}(0,0)\,\frac{\ddot{m}^2(x)}{4}h^{-1}K(\frac{x_1-x}{h})\frac{1}{n}(\int K(u)u^2\,du)(1+o(1)) \\
\to \quad & 0.
\end{aligned}
$$

The last limit was ensured by Lemma 3.3. Combining all the above arguments together, we have that

$$\sqrt{nh}\,\{\hat{m}(x)-m(x)-\frac{1}{2}h^2\sigma_0^2\ddot{m}(x)\} = \frac{1}{2\sqrt{nh}\;p(0)}\sum_{i=1}^{n}D(\varepsilon_i)K_i+o_p(1).$$

Now the theorem follows from Lemma 3.7, Lemma 3.9 and Lemma 3.11 immediately. The proof is completed.

Acknowledgements. We thank Dr. Shi-Qing Ling for drawing our attention to the paper of Hsing (1999) and Professor Howell Tong for helpful comments. We thank a reviewer and an associate editor for their comments. QY's research was partially supported by a BBSRC/EPSRC grant.

References

[1] Adler, R.J., Feldman, R.E. and Gallagher, C. (1997). Analysing stable time series. In: *A User's Guide to Heavy Tails: Statistical Techniques For Analysing Heavy Tailed Distributions and Processes*, Ed. R.J. Adler, R.E. Feldman and M. Taqqu. Birkhäuser, Boston.

[2] Altman, N.S. (1990). Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.*, **85**, 749-759.

[3] Chiu, C.K. and Marron, J.S. (1991). Comparison of two bandwidth selectors with dependent errors. *Ann. Statist.* **19**, 1906–1918.

[4] Csörgő, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23**, 1000–1014.

[5] Davis, R.A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variances. *Stoch. Processes Appl.* **40**, 145-80.

[6] Fan, J. and Hall, P. (1994). On curve estimation by minimizing mean absolute deviation and its implications. *Ann. Statist.* **22**, 867–885.

[7] Fan, J., Hu, T.-C. and Truong, Y.K. (1994). Robust nonparametric function estimation. *Scand. J. Statist.* **21**, 433–446.

[8] Geluk, J. and de Haan, L. (1987). Regular Variation, Extensions and Tauberian Theorems. *CWI Tract 40, Amsterdam.*

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

[10] Hall, P., Peng, L. and Yao, Q. (2002). Prediction and nonparametric estimation for time series with heavy tails. *Journal of Time Series Analysis.* **23(3)**, 313 - 331.

[11] Hart, J.D. (1991). Kernel regression estimation with time series errors. *J. Roy. Statist. Soc.*, **B**, **53**, 173-187.

[12] Hart, J.D. (1994). Automated kernel smoothing of dependent data by using time series crossing-validation. *J. Roy. Statist. Soc.*, **B**, **56**, 529-542.

[13] Herrmann, E., Gasser, T. and Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika.* **79(4)**, 783 - 795.

[14] Hsing, T. (1999). On the asymptotic distributions of partial sums of functionals of infinite-variance moving averages. *Ann. Probab.* **27**, 1579 - 1599.

[15] Kasahara, Y. and Maejima, M. (1988). Weighted sums of i.i.d. random variables attracted to integrals of stable processes. *Probab. Th. Rel. Fields*, **78**, 75 - 96.

[16] Kokoszka, P. and Taqqu, M.S. (1995). Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications.* **60**, 19 - 47.

[17] Kokoszka, P. and Taqqu, M.S. (1996). Parameter estimation for infinite variance fractional ARIMA. *Ann. Statist.* **24**, 1880 - 1913.

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

[19] Mallows, C.L. (1980). Some theory of nonlinear smoothers. *Ann. Statist.* **8**, 695–715.

[20] Mikosch, T., Gadrich, T., Klüppelberg, C. and Alder, R.J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.* **23**, 305 - 326.

[21] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186-98.

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

[23] Resnick, S.I. (1997). Heavy tail modeling and teletraffic data (with discussion). *Ann. Statist.* **25**, 1805-1869.

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

[25] Robinson, P.M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25**, 2054–2083.

[26] Roussas, G.G., Tran, L.T. and Ioannides, O.A. (1992). Fixed design regression for time series: asymptotic normality. *J. Multivariate Anal.* **40**, 262–291.

[27] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York.

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

[29] Tran, L.T., Roussas, G.G., Yakowitz, S. and Truong Van, B. (1996). Fixed-design regression for linear time series. *Ann. Statist.* **24**, 975–991.

[30] Truong, Y.K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17**, 606–617.

[31] Truong, Y.K. (1991). Nonparametric curve estimation with time series errors. *J. Statist. Plann. Inf.* **28**, 167–183.

[32] Truong, Y.K. (1992a). Robust nonparametric regression in time series. *J. Multivar. Anal.* **41**, 163–177.

[33] Truong, Y.K. (1992b). A nonparametric approach for time series analysis. *IMA Volumes in Mathematics and its Applications* **45**, 371–386.

[34] Truong, Y.K. and Stone, C.J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20**, 77–97.

[35] Tsybakov, A.B. (1986). Robust reconstruction of functions by local-approximation method. *Problems Inform. Transmission* **22**, 133–146.

[36] Velleman, P.F. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. *J. Amer. Statist. Assoc.* **75**, 609–615.

[37] Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *J. Nonparametric Statist*, **6**, 273-292.