Journal of Multivariate Analysis 102 (2011) 349–362

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# On linear models with long memory and heavy-tailed errors

Zhou Zhou *, Wei Biao Wu

*University of Toronto, Canada*
*University of Chicago, United States*

ARTICLE INFO

*Article history:*
Received 18 October 2007
Available online 22 September 2010

*AMS subject classifications:*
62J05
60G18
60G51

*Keywords:*
Bahadur representation
Heavy tails
Long memory
*M*-estimation

ABSTRACT

We consider the robust estimation of regression parameters in linear models with long memory and heavy-tailed errors. Asymptotic Bahadur-type representations of robust estimates are developed and their limiting distributions are obtained. It is shown that the limiting distributions are very different from those obtained under short memory. A simulation study is carried out to compare the performance of various asymptotic representations.



## 1. Introduction

The estimation of unknown regression parameters in linear models has been extensively studied. The least squares estimate (LSE) is widely used in practice and its finite and asymptotic distribution theory has been well developed; see for example the texts by Davidson and MacKinnon [11] and Rao and Toutenburg [26]. For linear models with heavy-tailed errors, the LSE may perform poorly and robust estimates are attractive alternatives. The last three decades have witnessed a rapid growth in quantile estimation and other robust procedures. See [17,14,19] for excellent treatments.

Consider the $p$-variate linear model:

$$y_i = \mathbf{x}_i^T \beta + e_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\mathbf{A}^T$ denotes the matrix transpose, and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$, $1 \le i \le n$, are $p \times 1$ known deterministic design vectors. As a typical robust estimation procedure, let $\rho$ be a convex function and we estimate the unknown parameter vector $\beta$ by the minimizer

$$\hat{\beta}_n = \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^T \beta). \tag{2}$$

Note that, $\rho(u) = u^2$ leads to LSE. Other popular choices of $\rho$ include quantile regression with $\rho(x) = \alpha x_+ + (1-\alpha)(-x)_+$, $0 < \alpha < 1$, where $x_+ = \max(0, x)$, Huber's procedure [17] with $\rho(x) = (x^2 \mathbf{1}\{|x| > c\})/2 + (cx - c^2/2)\mathbf{1}\{|x| \le c\}$, $c > 0$, and $\mathcal{L}^q$ regression with $\rho(x) = |x|^q$, $1 \le q \le 2$. In the literature, asymptotic properties of $\hat{\beta}_n - \beta$ have been studied mainly under the assumption that the errors are independent (Bassett and Koenker [5], Babu [3], Bai et al. [4] and He and Shao [15] among

* Corresponding address: Department of Statistics, 100 St. George St. Toronto, M5S 3G3, Canada.
*E-mail addresses:* zhou@utstat.toronto.edu, zhouzhou@uchicago.edu (Z. Zhou), wbwu@galton.uchicago.edu (W.B. Wu).

0047-259X/$ – see front matter 
doi:10.1016/j.jmva.2010.09.009

others) or strong mixing [13,23,10] or short-range dependent [33]. See the latter paper for additional references for robust estimation under independence and weak dependence. Hampel et al. [14] argued that many science and engineering data exhibit significant temporal dependence and the assumption of independence is violated; see Chapter 8 therein. However, there seem to be few results on robust estimators of linear models with long memory (or long-range dependent) errors.

Recently processes with both heavy tails and long memory have received considerable attention. Willinger et al. [32] showed that self-similarity and heavy tails exist in network traffic data, while Rachev and Mittnik [25] did an extensive empirical study and showed that high frequency asset return data exhibited both long memory and heavy tails. To the best of our knowledge, most of the existing results focused on estimation and inference of long memory and heavy-tail parameters, while little attention was paid to regression analysis. The latter problem is clearly of great interest if one wants to include covariates or predictors into the model for explanatory purpose.

The paper aims to study properties of $\hat{\beta}_n$ under the assumption that the errors $e_i$ in model (1) are long memory as well as heavy tailed; see Section 2.1 for assumptions on the error structure. It is shown that asymptotic behavior of $\hat{\beta}_n$ is very different from that obtained under independence and weak dependence. We will also provide Bahadur representations of the robust estimates of model (1). Those representations are useful for further analysis of the asymptotics of robust estimates.

The rest of the paper is structured as follows. Regularity conditions are given in Section 2. Section 3 presents main results including consistency and asymptotic distributions of robust estimates. Section 5 provides proofs and Section 4 presents a simulation study.

## 2. Preliminaries

We now introduce some notation. For a vector $\mathbf{v} = (v_1, \ldots, v_p) \in \mathbb{R}^p$, let $|\mathbf{v}| = (\sum_{i=1}^p v_i^2)^{1/2}$. For a $p \times p$ matrix $A$, define $|A| = \sup\{|A\mathbf{v}| : |\mathbf{v}| = 1\}$. For a random vector $\mathbf{V}$, write $\mathbf{V} \in \mathcal{L}^q$ $(q > 0)$ if $\|\mathbf{V}\|_q := [\mathbb{E}(|\mathbf{V}|^q)]^{1/q} < \infty$. Let $(\eta_n)$ be a sequence of random variables and $(d_n)$ a positive sequence. We write $\eta_n = o_p(d_n)$ if $\eta_n/d_n \to 0$ in probability and $\eta_n = O_p(d_n)$ if $\eta_n/d_n$ is bounded in probability. Denote by $\Rightarrow$ the weak convergence. Let $\mathcal{C}^i$, $i \in \mathbb{N}$, be the collection of functions that have $i$th order continuous derivatives. Let $C$ denote a generic constant independent of $n$ and its value may vary from place to place.

### 2.1. The error structure

We assume that $(e_i)$ is a moving average process

$$e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}, \tag{3}$$

where $\varepsilon_j, j \in \mathbb{Z}$, are independent and identically distributed (iid) random variables with mean 0 and $\varepsilon_j \in \mathcal{D}(\alpha)$, $\alpha \in (1, 2)$. Here $\mathcal{D}(\alpha)$ denotes the $\alpha$-stable domain of attraction (see [9]); namely, there exists real sequences $(A_n)$ and $(B_n)$ such that $A_n^{-1}(\varepsilon_1 + \cdots + \varepsilon_n) - B_n$ converges to an $\alpha$-stable law whose characteristic function is

$$\varphi(t) = \exp(-\sigma^\alpha |t|^\alpha (1 - \sqrt{-1}\varrho w_\alpha(t)) + \sqrt{-1}\mu t), \quad \text{where } w_\alpha(t) = \tan \frac{\pi\alpha \operatorname{sgn}(t)}{2}. \tag{4}$$

Here $\sigma$, $\mu$ and $\varrho$ $(-1 \le \varrho \le 1)$ are the scale, shift and skewness parameters, respectively, and $\sqrt{-1}$ is the imaginary unit. Let $F_\varepsilon$ be the distribution function of $\varepsilon_j$ and $f_\varepsilon = F_\varepsilon'$ be its density. Then $\varepsilon_j \in \mathcal{D}(\alpha)$ can be characterized by

$$1 - F_\varepsilon(u) = \frac{c_1 + o(1)}{u^\alpha} L(u) \quad \text{and} \quad F_\varepsilon(-u) = \frac{c_2 + o(1)}{u^\alpha} L(u) \quad \text{as } u \to \infty, \tag{5}$$

where $c_1, c_2 \ge 0$, $c_1 + c_2 > 0$ and $L$ is a slowly varying function, i.e., $\lim_{x\to\infty} L(tx)/L(x) = 1$, for all $t > 0$ (cf. [7]). It is easy to see that $\inf\{x : \mathbb{P}(|\varepsilon_i| > x) \le 1/n\} = n^{1/\alpha} L_1(n)$, where $L_1$ is also a slowly varying function. Observe that $\varepsilon_i \in \mathcal{L}^{\alpha'}$, for all $\alpha' \in (0, \alpha)$, and $\alpha$ is called the heavy-tail index, and $\mathbb{E}(\varepsilon_i^2) = \infty$. If $\varrho = \mu = 0$, then (4) becomes the symmetric-$\alpha$-stable ($S\alpha S$) law. In this case (5) holds with $L(t) = 1$, $c_1 = c_2 = \sigma^\alpha/(2C_\alpha)$, where $C_\alpha = \cos(\alpha\pi/2)\Gamma(2-\alpha)/(1-\alpha)$.

Let $\varepsilon_\alpha(u)$, $u \in \mathbb{R}$, be a two-sided Levy $\alpha$-stable process [28] with independent increments, $\varepsilon_\alpha(0) = 0$, and $\varepsilon_\alpha(u+t) - \varepsilon_\alpha(u)$ having characteristic function $\varphi(t)$ (cf. (4)) with $\mu = 0$. By Theorem 2.7 in [29], in the space $\mathcal{D}[0, 1]$ of functions that are right continuous and have left limit, we have the weak convergence

$$\lim_{n\to\infty} \left\{ \frac{1}{n^{1/\alpha} L_1(n)} \sum_{i=1}^{\lfloor nu \rfloor} \varepsilon_i, 0 \le u \le 1 \right\} = \{\varepsilon_\alpha(u), 0 \le u \le 1\}, \tag{6}$$

where $\lfloor v \rfloor = \max\{j \in \mathbb{Z} : j \le v\}$. See also [2].

For the coefficients $(a_j)_0^\infty$, we assume $a_0 = 1$, and, for $j \ge 1$,

$$a_j = j^{-\gamma} l(j), \quad 1/\alpha < \gamma < 1, \text{ where } l(\cdot) \text{ is a slowly varying function.} \tag{7}$$

By Kolmogorov's three series theorem [9], under (7), $e_i$ is well defined. The partial sum process $e_1 + \cdots + e_k$, after proper normalization, converges to linear fractional stable motion in an appropriate sense; see [2].

Under (7), $\sum_{i=0}^{\infty} |a_i| = \infty$, which implies strong dependence. Note that, under our model, all autocovariances of the process $(e_i)$ equal infinity. Hence our definition of long memory is different from the usual one which says that the autocovariances decay slowly. McElroy and Politis [22] give an example of a heavy-tailed long memory process that has infinite variance and finite autocovariances. The parameter $\gamma$ controls the magnitude of the memory, with smaller $\gamma$ indicating stronger dependence. An important special case is the fractionally integrated ARIMA (FARIMA) processes [16]. For such processes $a_j j^{1-d} \to c_0$, where $c_0$ is a constant. So (7) holds with $\gamma = 1 - d$, where $d \in (0, 1 - \alpha^{-1})$.

### 2.2. Regularity conditions

Without loss of generality, we assume throughout the paper that the true parameter $\beta_0 = 0$. Define $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$. Assume that $\Sigma_n := \mathbf{X}_n^T \mathbf{X}_n = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T$ is non-singular for sufficiently large $n$. We shall consider the transformed model

$$y_i = \mathbf{z}_{i,n}^T \theta_n + e_i, \quad \text{where } \mathbf{z}_{i,n} = \Sigma_n^{-1/2} \mathbf{x}_i \text{ and } \theta_n = \Sigma_n^{1/2} \beta. \tag{8}$$

Observe that $\hat{\theta}_n = \Sigma_n^{1/2} \hat{\beta}_n$ is a minimizer of $\sum_{i=1}^n \rho(e_i - \mathbf{z}_{i,n}' \theta)$ and that, by definition, $\mathbf{z}_{i,n} = (z_{i,1,n}, z_{i,2,n}, \ldots, z_{i,p,n})^T$ satisfies $\sum_{i=1}^n \mathbf{z}_{i,n} \mathbf{z}_{i,n}^T = \mathrm{Id}_p$, the $p \times p$ identity matrix. For $q > 0$, let $s_n(q) = \sum_{i=1}^n |\mathbf{z}_{i,n}|^q$. Let $m_n = \max_{1 \le i \le n} |\mathbf{z}_{i,n}|$. Then $s_n(2) = p \le n m_n^2$. It may occur that $n m_n^2 \to \infty$. For example, let $p = 1$, $\mathbf{x}_i = i^{-1/3}$. Then $\Sigma_n/(3n^{1/3}) \to 1$ and $m_n(3n^{1/3})^{1/2} \to 1$ and $n^{1/2} m_n \to \infty$. Since $\Sigma_n$ is non-singular for all large $n$, we can pick $p$ linearly independent $\mathbf{x}_i$s and denote this $p \times p$ sub-matrix of $\mathbf{X}$ by $\mathbf{X}_*$. Then $\mathbf{X}_*$ is non-singular and $\mathbf{X}_* \Sigma_n^{-1} \mathbf{X}_*^T = O(m_n^2)$. Hence

$$|\Sigma_n^{-1}| = O(m_n). \tag{9}$$

Assume that $\rho$ is absolutely continuous with derivative $\psi = \rho'$. We now impose some regularity conditions on $\psi$, $\mathbf{z}_{i,n}$ and $\varepsilon_i$.

(A1) $|\psi(x)| \le C(1 + |x|)$ for all $x \in \mathbb{R}$, and $\mathbb{E}[\psi(e_i)] = 0$.
(A2) $f_\varepsilon \in \mathcal{C}^r$, and $|f_\varepsilon^{(i)}(x)| \le C(1 + |x|)^{-\alpha-1} L(|x|)$ for all $x \in \mathbb{R}$, $i = 0, \ldots, r$.
(A3) $\phi(x) := \mathbb{E}[\psi(e_i + x)]$ has a strictly positive derivative at 0.
(A4) Let $d_n = n^{1-\gamma+1/\alpha} L_1(n) l(n)$ and $k_n = |d_n| m_n$. Assume $m_n k_n = o(1)$.

A few remarks are in order. Condition (A1) controls the tail of $\psi$. For $M$-estimation with long memory and heavy-tailed errors, previous results require that $\psi$ is bounded [21]. The latter restriction excludes the important $\mathcal{L}^q$ regression with $1 < q \le 2$ [35]. Our (A1) allows a wider range of $\rho$.

Condition (A2) is for technical purpose and it is not the weakest possible. Its purpose is to guarantee sufficient smoothness of conditional expectations of $\psi(e_i + x)$; see Lemma 1 below. It is satisfied if $\varepsilon_i$ is $S\alpha S$. In this case $f_\varepsilon$ is $r$ times differentiable and $f_\varepsilon^{(r)}(x) \sim C|x|^{-1-r-\alpha} L(|x|)$ as $x \to \pm\infty$ [18].

Condition (A3) is a natural condition for $\theta_n$ to be estimable. Under (A1) and (A2) with $r = 1$, $\phi$ is differentiable (see Lemma 1).

Under Condition (A4), since $n m_n^2 \ge p$, $k_n \ge p^{1/2} n^{-1/2} |d_n| \ge n^{1/2-\gamma+1/\alpha} |L_1(n) l(n)|$. Observe that $1/2 - \gamma + 1/\alpha > 0$ in view of $1/\alpha < \gamma < 1$ and $1 < \alpha < 2$. Then for any $c \in (0, 1/2 - \gamma + 1/\alpha)$, we have $n^c = o(k_n)$ and $m_n = o(n^{-c})$. Condition (A4) is needed for the weak consistency of the $M$-estimate $\hat{\beta}_n$. For linear models with independent errors, the condition $m_n = o(1)$ is needed for consistency [12]. In this case $\hat{\theta}_n = O_p(1)$. So $\hat{\beta}_n = o_p(1)$ under $m_n = o(1)$. On the other hand, under long memory, $|\hat{\theta}_n| = O_p(k_n)$ (Theorem 1), so (A4) is natural. Note that, if $m_n = O(n^{-1/2})$, then (A4) always holds.

## 3. Main results

### 3.1. Consistency

**Theorem 1.** *Let $\hat{\beta}_{n,ls}$ be the LSE of (1); let $U_n = \sum_{i=1}^n \mathbf{z}_{i,n} e_i$. Suppose that* (A1), (A3), (A4) *holds and that* (A2) *holds with $r = 2$. Then $|\hat{\theta}_n| = O_p(k_n)$ and $|\hat{\theta}_n - U_n| = o_p(k_n)$. Consequently $|\hat{\beta}_n| = o_p(1)$ and $|\hat{\beta}_n - \hat{\beta}_{n,ls}| = o_p(1)$.*

Theorem 1 asserts that $\hat{\theta}_n$ can be approximated by $U_n$, which is often easier to deal with due to its linearity structure. It also asserts that the $M$-estimate $\hat{\beta}_n$ and the LSE $\hat{\beta}_{n,ls}$ are asymptotically first order equivalent. Several authors have already noticed this phenomenon, but under more restrictive conditions; see [6] for subordinated Gaussian processes and [21] for bounded $\psi$ with a special design matrix under which $m_n = O(n^{-1/2})$. We will obtain a more precise order of $\hat{\beta}_n - \hat{\beta}_{n,ls}$ via Bahadur representations of $\hat{\theta}_n$ (cf. Theorem 3).

### 3.2. Bahadur representations

To establish an asymptotic expansion of $\hat{\theta}_n$, we need

(A5) Let the $M$-process $\Xi_n(\theta) = \sum_{i=1}^n \psi(e_i - \mathbf{z}_{i,n}^T \theta) \mathbf{z}_{i,n}$. Assume $|\Xi_n(\hat{\theta}_n)| = O_p(m_n)$.

Condition (A5) is natural. If $\psi$ is continuous, $\Xi_n(\hat\theta_n) = 0$. For quantile regression, $\psi$ is discontinuous, by Babu [3], $|\Xi_n(\hat\theta_n)| \le (p+1)m_n$ almost surely if (A2) holds with $r = 0$.

For Bahadur representations of $M$-estimates, we approximate $\hat\theta_n$ by the linear form

$$V_n := \sum_{i=1}^n \psi(e_i)\mathbf{z}_{i,n}. \tag{10}$$

(See [15,3,33].) Under (A5) and slightly stronger conditions than those in Theorem 1, we have Theorems 2 and 3, which concern approximations of $\hat\theta_n$ by $V_n$ and $U_n$, respectively.

**Theorem 2.** *Let* (A2) *hold with* $r = p + 1$. *Assume* (A3)–(A5) *and*

$$|\psi(x)| \le C(1+|x|)^{\alpha_0} \quad \text{for some } \alpha_0 < \alpha/2. \tag{11}$$

*Let* $\pi(x) = \|\psi(e_i + x) - \psi(e_i)\|_2$*; let* $(c_n)$ *be a positive sequence with* $c_n \to \infty$. *Then*

$$|\phi'(0)\hat\theta_n - V_n| = O_p((\tau_n^{1/2}(r_n k_n)\log n + r_n m_n k_n^2)) \tag{12}$$

*where* $r_n = \min\{c_n, (k_n m_n)^{-1/2}\}$ *and* $\tau_n(x) = \sum_{i=1}^n |\mathbf{z}_{i,n}|^2[\pi^2(|\mathbf{z}_{i,n}|x) + \pi^2(-|\mathbf{z}_{i,n}|x)]$.

Clearly (11) implies that $\pi(x)$ exists. Under our setting, $\pi(x) \to 0$ as $x \to 0$. To demonstrate this, observe that since $\psi$ is nondecreasing and $e_i$ has a continuous distribution function, $\psi(e_i + x) - \psi(e_i) \to 0$ almost surely as $x \to 0$. Since $\|\psi(e_i+1) - \psi(e_i-1)\|_2^2 < \infty$, $\pi(x) \to 0$ as $x \to 0$ by the Lebesgue Dominant Convergence Theorem (LDCT). In particular, if $\pi(x) \to 0$ at some polynomial rate, we have the following corollary.

**Corollary 1.** *Recall* $s_n(q) = \sum_{i=1}^n |\mathbf{z}_{i,n}|^q$. *Let* $\pi(x) = O(|x|^\lambda)$ *for some* $\lambda > 0$. *Then under the conditions of* Theorem 2,

$$|\phi'(0)\hat\theta_n - V_n| = O_p(k_n^\lambda s_n^{1/2}(2+2\lambda)\log n + m_n k_n^2). \tag{13}$$

By Theorem 2, for any positive sequence $(c_n)$ with $c_n \to \infty$, we have $|\phi'(0)\hat\theta_n - V_n| = O_p(c_n^\lambda k_n^\lambda s_n^{1/2}(2+2\lambda)\log n + c_n m_n k_n^2)$. Since $c_n \to \infty$ can be arbitrarily slow, (13) follows.

**Example 1.** Assume (A2) with $r = 0$. For quantile regression with $\rho(x) = \alpha x_+ + (1-\alpha)(-x)_+$, $\psi(x) = \alpha - I\{x \le 0\}$ and $\|\psi(e_i + x) - \psi(e_i)\|_2^2 = |\int_{-x}^0 f_e(t)\mathrm{d}t|$. So $\pi(x)^2 = O(|x|)$. For Huber's function $\rho(x) = (x^2 I\{|x| \le c\})/2 + (c|x| - c^2/2) I\{|x| > c\}$, $c > 0$, we have $\psi'(x) = I\{-c \le x \le c\}$. Thus $\pi(x)^2 = \mathbb{E}[\int_0^x \psi'(e_i + t)\mathrm{d}t]^2 \le x^2$. Then $\lambda = 1$. Since $s_n(4) \le m_n^2 \sum_{i=1}^n |\mathbf{z}_{i,n}|^2 = p m_n^2$ and $n^c = o(k_n)$ for $c \in (0, 1/2 - \gamma + 1/\alpha)$, we have $k_n s_n^{1/2}(4)\log n = o(m_n k_n^2)$. So the bound in (13) becomes $O(k_n^2 m_n)$. For $\mathcal{L}^q$ regression with $1 < q < 2$, Arcones [1] showed that if $q \ne 3/2$, then $\pi(x) = O(|x|^{\min(2,2q-1)/2})$. If $q = 3/2$, then $\pi(x) = O(|x|\log(1/|x|))$. □

**Theorem 3.** *Let the conditions of* Theorem 2 *be satisfied.* (i) *We have*

$$|\hat\theta_n - U_n| = O_p(n^\eta + \tau_n^{1/2}(r_n k_n)\log n + r_n m_n k_n^2) \tag{14}$$

*for* $\eta$ *satisfying* $1/2 - \gamma + 1/\alpha > \eta > \eta_0 := 1/2 - \gamma + 1/\alpha - (\gamma - 1/\alpha)^2/\gamma$. (ii) *If* $\pi(x) = O(|x|^\lambda)$ *for some* $\lambda > 0$ *as* $x \to 0$. *Then*

$$|\hat\beta_n - \hat\beta_{n,ls}| = O_p(m_n n^\eta + m_n k_n^\lambda s_n^{1/2}(2+2\lambda)\log n + m_n^2 k_n^2). \tag{15}$$

The bound (14) in Theorem 3 is sharper than the one in Theorem 1. To see this, let $r_n = (m_n k_n)^{-1/2}$. Under (A4), $r_n \to \infty$, $r_n m_n k_n^2 = o(k_n)$, and $\tau_n(r_n k_n) = o(1)$ since $\pi(x) \to 0$ as $x \to 0$. Note that, $n^\eta = o(k_n)$. Hence (14) is sharper.

**Remark 1.** Suppose $m_n = O(n^{-1/2})$. Then $m_n^2 k_n^2 = O(m_n n^\eta)$. By simple calculations,

$$n|d_n|^{-1}|\hat\beta_n - \hat\beta_{n,ls}| = O_p(n^{-\eta_1}), \quad \text{for any } 0 < \eta_1 < \eta^*, \tag{16}$$

where $\eta^* = \min\{(\gamma - 1/\alpha)^2/\gamma, 1/2 + (\lambda - 1)(\gamma - 1/\alpha)\}$. This is sharper than the $o_p(1)$ bound in [21]. More importantly, we see from (16) that, even though $\hat\beta_n$ and $\hat\beta_{n,ls}$ are asymptotically equivalent, magnitude of the dependence and the tail of the errors as well as the quantity $\lambda$ determine the speed at which $\hat\beta_n - \hat\beta_{n,ls}$ converges to 0. We assume for now $\lambda \ge 1/2$, as it will be satisfied by all quantile, Huber and $\mathcal{L}^q$ regressions (cf. Example 1). Let $K = \gamma - 1/\alpha$. Since $1/\alpha < \gamma < 1$ and $1 < \alpha < 2$, we have by elementary calculations that $K^2/\gamma < 1/2 + (\lambda - 1)(\gamma - 1/\alpha)$. Then $\eta^* = K^2/\gamma$. Note that, $H = 1 - K$ is the Hurst index of the error process $(e_i)$. So lighter tails and weaker dependence lead to faster convergence when $\lambda \ge 1/2$. □

### 3.3. Limiting distributions of *M*-estimates

We now present a limit theory for our *M*-estimators under conditions (B1) and (B2) below. Let $\mathbf{x}^j = (x_{1j}, \ldots, x_{nj})$, $j = 1, \ldots, p$. For $\mathbf{v} = (v_1, \ldots, v_n)$, define the function $h_{\mathbf{v}}(\cdot) : [0, 1] \to \mathbb{R}$ by $h_{\mathbf{v}}(t) = v_{\lfloor nt \rfloor+1}$, $0 \le t < 1$, and $h_{\mathbf{v}}(1) = v_n$.

(B1) $h_{\mathbf{x}^j}(\cdot)$ converge uniformly on [0, 1] to continuous functions $g_j(\cdot)$, $j = 1, \ldots, p$.

(B2) Let $\mathbf{g} = (g_1, \ldots, g_p)^T$. Assume that the matrix $\mathcal{G} := \int_0^1 \mathbf{g}(t)\mathbf{g}^T(t)\mathrm{d}t$ is non-singular.

**Theorem 4.** *Assume* (A1), (A3), (B1), (B2), *and that* (A2) *holds with* $r = 2$. *Then*

$$\frac{n^{\gamma-1/\alpha}}{L_1(n)l(n)}\hat{\beta}_n \Rightarrow \mathcal{G}^{-1}\int_{-\infty}^{\infty}\left[\int_0^1 \mathbf{g}(x)(x-u)_+^{-\gamma}\mathrm{d}x\right]\mathrm{d}\varepsilon_\alpha(u) := \mathcal{G}^{-1}\mathcal{L}_{\mathbf{g}}(\varepsilon). \tag{17}$$

**Remark 2.** Under (B1) and (B2), we have $n^{-1}\Sigma_n \to \mathcal{G}$ and $\max_{1\le i\le n}|\mathbf{x}_i| = O(1)$. Thus $m_n = \max_{1\le i\le n}|\Sigma_n^{-1/2}\mathbf{x}_i| = O(n^{-1/2})$. In particular, condition (A4) is always satisfied. As a special case, if $x_{ij} = g_j(i/n)$, where $g_j(\cdot)$ is continuous on [0, 1], then condition (B1) is always satisfied. This type of design was discussed in [21]. □

**Remark 3.** We see from the above theorem that the *M*-estimator and the LSE are not only first order equivalent, but are also equally efficient. Since the *M*-estimator is robust to additive outliers, it is preferred when long memory and heavy tails are present. □

**Remark 4.** To apply Theorem 4 for statistical inference of $\beta$, we need to know or estimate slowly varying functions $L_1(n)$ and $l(n)$ and parameters $\alpha, \sigma, \varrho$ and $\gamma$. The distribution of $\mathcal{L}_{\mathbf{g}}(\varepsilon)$ can be approximated by plugging in the estimated values of $\alpha, \sigma, \varrho$ and $\gamma$. Kokoszka and Taqqu [20] and Taqqu and Teverovsky [30] discussed parameter estimation of heavy-tailed FARIMA processes. Resnick and Stărică [27] considered heavy tail index estimation for long memory and heavy-tailed linear processes. Chapter 8 of [24] contains a discussion on estimating slowly varying functions using subsampling. We expect that, using their techniques, we can estimate $L_1(n), l(n), \alpha, \sigma, \varrho$ and $\gamma$ from the estimated residuals $\hat{e}_i = y_i - \mathbf{x}_i^T\hat{\beta}_n$ of model (1), and establish a related asymptotic theory. However, the latter problem seems nontrivial and we leave it as an open problem. □

**Example 2** (*Polynomial Design*)**.** Let $x_{ij} = (i/n)^{j-1}$, $i = 1, \ldots, n, j = 1, \ldots, p$. Then condition (B1) holds with $g_j(x) = x^j$. So $\mathcal{G}_{ij} = 1/(i+j+1)$, $1 \le i, j \le p$. $\mathcal{G}$ is the Hilbert matrix and it is non-singular. Hence Theorem 4 and previous theorems are applicable.

## 4. A simulation study

The Bahadur representation plays an important role in understanding the asymptotic behavior of the estimates. In Section 3.2, we proposed two such representations under long memory and heavy tails. The accuracy of the approximations depends on the strength of the dependence and the thickness of the tails of the error process. Here we shall carry out a simulation study and compare the sensitivity of the $V_n$ and $U_n$ representations to parameter values $\alpha$ and $\gamma$. Accuracy of the two approximations is compared in Section 4.2.

### 4.1. Performance of representations

Consider the simple regression model:

$$y_i = \beta_1 + \beta_2\frac{i}{n} + e_i, \quad i = 1, 2, \ldots, n, \tag{18}$$

where $e_i$ has the form (3), $a_i = (i+1)^{-\gamma}$ and $\varepsilon_i$ are iid standard $S\alpha S$ random variables. Let the true parameter $(\beta_1, \beta_2) = (0, 0)$. In our simulation we use the LAD regression, which corresponds to Example 1 with $\alpha = 0.5$. Then (16) holds with $\eta^* = K^2/\gamma$, $K = \gamma - 1/\alpha$; see Remark 1. From (13) and the discussion in Example 2, we have

$$|\hat{\beta}_n - U_n^*| = O_p(n^{-\eta_1}), \quad \text{for any } 0 < \eta_1 < \eta_u, \tag{19}$$

where $U_n^* = \hat{\beta}_{n,ls} = \Sigma_n^{-1/2}U_n$ and $\eta_u = (\gamma - 1/\alpha)(2 - 1/(\alpha\gamma)) = K^2/\gamma + K$, and

$$|\hat{\beta}_n - V_n^*| = O_p(n^{-\eta_2}), \quad \text{for any } 0 < \eta_2 < \eta_v, \tag{20}$$

where $V_n^* = \Sigma_n^{-1/2}V_nf_e^{-1}(0)$, $f_e$ is the density of $e_i$, and $\eta_v = \min\{K/2 + 1/2, 2K\}$.

Let $n = 1000$. We generate standard $S\alpha S$ variables by the algorithm in [8]. Using the convolution structure in process (3), we can apply the circular embedding and the fast Fourier transform algorithm; see [34]. Applying a version of the algorithm in their paper, we generate 6000 series of length 1000. For each generated series, quantile estimation of $\beta$ was carried

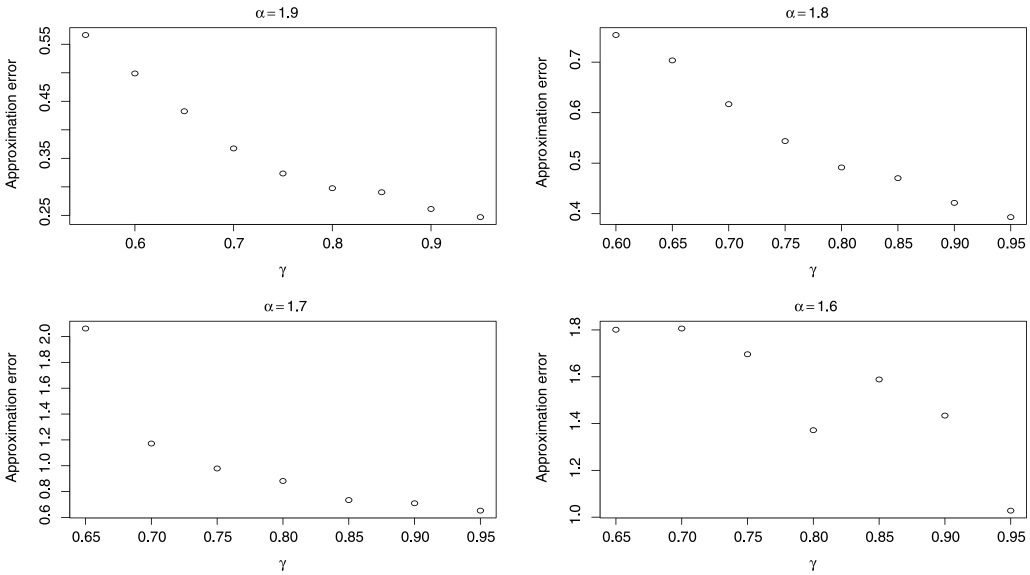


**Fig. 1.** Estimated absolute deviation errors of $U_n^*$ representation for fixed $\alpha$.

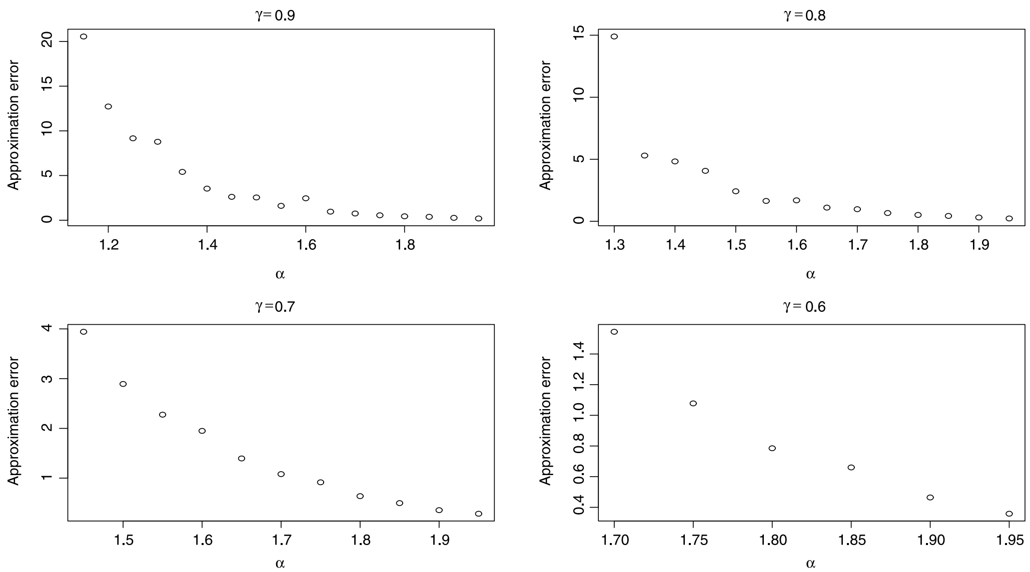


**Fig. 2.** Estimated absolute deviation errors of $U_n^*$ representation for fixed $\gamma$.

out and $U_n^*$ and $V_n^*$ were calculated. To assess the accuracy of the approximations, we compute the sample mean absolute deviation errors (MADE) of $\hat{\beta}_n - U_n^*$ and $\hat{\beta}_n - V_n^*$.

Figs. 1–4 show the $U_n^*$ and $V_n^*$ approximations for the slope parameter $\beta_2$. The results for the intercept parameter $\beta_1$ are similar. From those four graphs we see that for both the $U_n^*$ and $V_n^*$ representations, the mean absolute deviation error is small when the heavy-tail index $\alpha$ or the memory index $\gamma$ is sufficiently large. Furthermore, for fixed $\alpha$, approximation error decreases as the memory index $\gamma$ increases, which indicates that shorter memory leads to more accurate approximations. This finding is consistent with our theoretical assertions in (19) and (20). Similar conclusions can be drawn in the case of fixed memory index $\gamma$. On the other hand, we see relatively large approximation errors on the left edges of some of the graphs. This is because parameters $\eta_u$ and $\eta_v$ are very small on the left edge of the graphs.

### *4.2. Comparison of the $U_n$ and $V_n$ approximations*

We see from Figs. 1–4 that, for the same combination of $\alpha$'s and $\gamma$'s, the $U_n^*$ approximation usually starts with some smaller value than the $V_n^*$ approximation on the left region, yet ends with relatively larger errors. This suggests that the $U_n^*$ representation is better when the Hurst index $H$ is large, but worse when $H$ is small. For a more detailed study, we choose

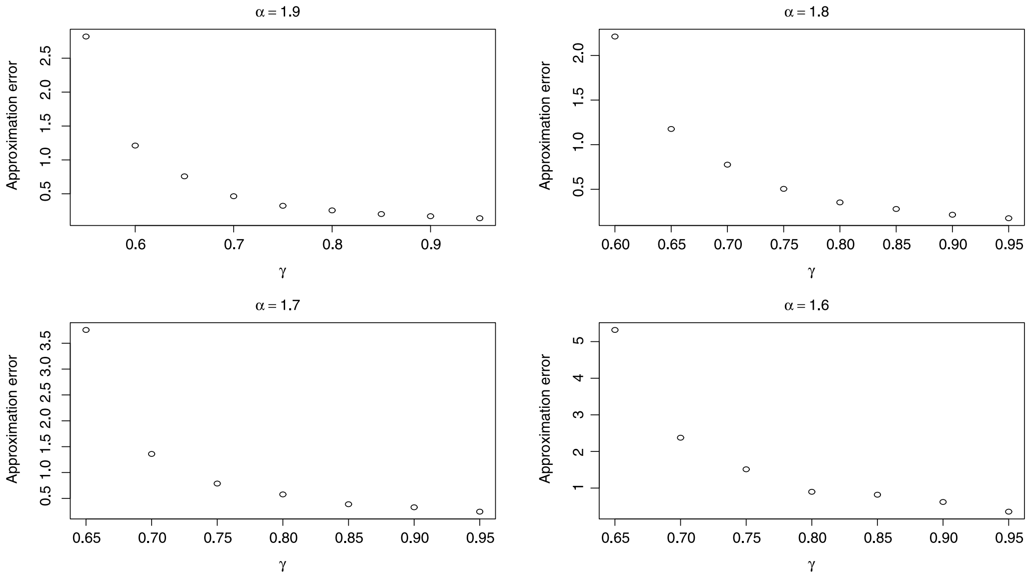


**Fig. 3.** Estimated absolute deviation errors of $V_n^*$ representation for fixed $\alpha$.

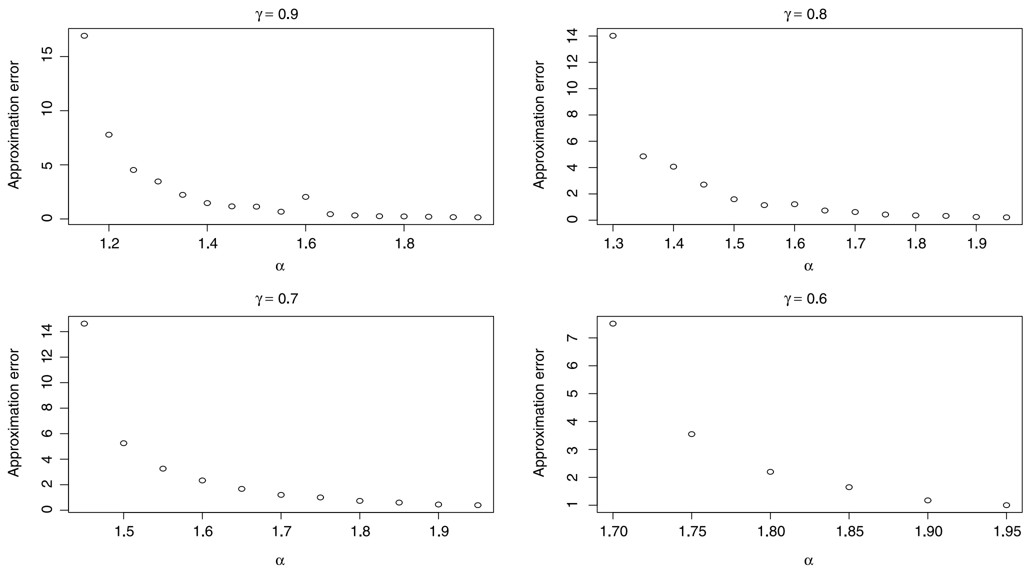


**Fig. 4.** Estimated absolute deviation errors of $V_n^*$ representation for fixed $\gamma$.

the same model and the simulation method as those in the last subsection; the only difference is that we now fix the Hurst index $H$, and according to this $H$, we choose different combinations of $\alpha$ and $\gamma$. We use two levels of $H$, $H = .95$ as simulation for large index and $H = .7$ for small index. The results are showed in Figs. 5 and 6, respectively.

Our simulation results support our claim. Note from (19) and (20), order of the $U_n^*$ approximation is always higher than that of the $V_n^*$ approximation, as $\eta_u \le \eta_v$. However, as we found in our results, when the Hurst index is large, and thus the large sample behavior is violated, the $U_n^*$ representation may perform better. A new theory is needed in order to explain this interesting phenomena.

Note also that, the $U_n^*$ representation decreases fast as the heavy tail index $\alpha$ increases, while the $V_n^*$ representation is relatively stable. This is consistent with (19) and (20) in the sense that $\eta_v$ increases as $\alpha$ increases, while $\eta_u$ does not change if we fix $H$.

## 5. Proofs of results in Section 3

We first introduce some notation. For $k \in \mathbb{Z}$ define the projection operator

$$\mathcal{P}_k\mathbf{V} = \mathbb{E}[\mathbf{V}|\mathcal{F}_k] - \mathbb{E}[\mathbf{V}|\mathcal{F}_{k-1}], \quad \mathbf{V} \in \mathcal{L}^1, \text{ where } \mathcal{F}_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k).$$

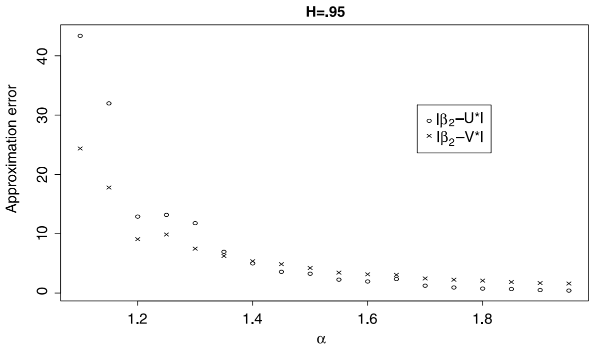


**Fig. 5.** Estimated absolute deviation errors of $U_n^*$ and $V_n^*$ representations for fixed Hurst index $H = .95$.

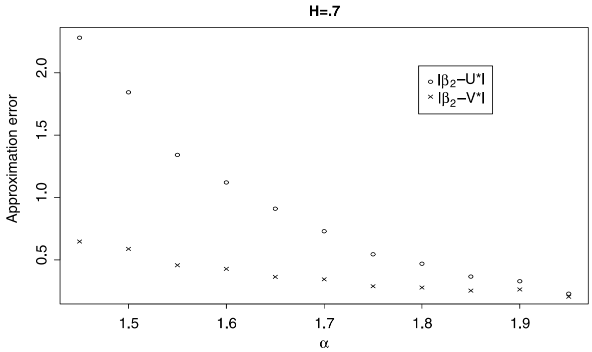


**Fig. 6.** Estimated absolute deviation errors of $U_n^*$ and $V_n^*$ representations for fixed Hurst index $H = .7$.

Introduce the truncated processes

$$\underline{e}_{n,k} = \sum_{i=-\infty}^{k} a_{n-i}\varepsilon_i \quad \text{and} \quad \bar{e}_{n,k} = \sum_{i=k}^{n} a_{n-i}\varepsilon_i. \tag{21}$$

Then $\underline{e}_{n,k}$ is $\mathcal{F}_k$-measurable. Recall $a_0 = 1$. Let

$$\psi_{n,\delta}(x) = \mathbb{E}[\psi(x+\delta+\bar{e}_{n,1})] \quad \text{and} \quad \psi_{\infty,\delta}(x) = \mathbb{E}[\psi(x+\delta+e_n)]. \tag{22}$$

Note $\phi(x) = \psi_{\infty,0}(x) = \psi_{\infty,x}(0)$. For a function $f$, let $f(x;\delta) = \sup_{|t|\le\delta} |f(x+t)|$, $\delta > 0$.

**Lemma 1.** *Let $n \in \mathbb{N}$ and $\delta \in [-1, 1]$. Under Conditions* (A1) *and* (A2), *$\psi_{n,\delta}(\cdot)$ and $\psi_{\infty,\delta}(\cdot)$ are $r$ times differentiable. Furthermore, we have $|\psi_{n,\delta}^{(i)}(x)| \le C(1+|x|)$ uniformly in $n \in \mathbb{N}$ and $\delta \in [-1, 1]$, and $|\psi_{\infty,\delta}^{(i)}(x)| \le C(1+|x|)$, $i = 0, 1, \ldots, r$.*

**Proof.** Let $F_{n,1}(x) := \mathbb{P}(\bar{e}_{n,1} \le x)$ be the distribution function of $\bar{e}_{n,1}$; let $f_{n,1} = F'_{n,1}$ be its density. Let $\Gamma_n = \bar{e}_{n,1} - \varepsilon_n$. Then $F_{n,1}(x) = \mathbb{E}[F_\varepsilon(x - \Gamma_n)]$. By the LDCT,

$$f_{n,1}^{(i)}(x) = \mathbb{E}[f_\varepsilon^{(i)}(x - \Gamma_n)] \quad \text{for } i = 0, 1, \ldots, r. \tag{23}$$

By Condition (A2), for $i = 0, 1, \ldots, r$, $f_\varepsilon^{(i)}(u; 1) \le C(1+|u|)^{-2-\iota}$ for some $\iota > 0$. So

$$\begin{aligned}
\int_{\mathbb{R}} |\psi(y)| f_{n,1}^{(i)}(y-\delta-x; 1)\mathrm{d}y &= \mathbb{E}\left[\int_{\mathbb{R}} |\psi(y)| f_\varepsilon^{(i)}(y-\delta-x-\Gamma_n; 1)\mathrm{d}y\right] \\
&= \mathbb{E}\left[\int_{\mathbb{R}} |\psi(u+\delta+x+\Gamma_n)| f_\varepsilon^{(i)}(u; 1)\mathrm{d}u\right] \\
&\le C\mathbb{E}\left[\int_{\mathbb{R}} (1+|u|)(1+|\delta|)(1+|x|)(1+|\Gamma_n|) f_\varepsilon^{(i)}(u; 1)\mathrm{d}u\right] \\
&\le C(1+|x|)\int_{\mathbb{R}} (1+|u|) f_\varepsilon^{(i)}(u; 1)\mathrm{d}u \le C(1+|x|).
\end{aligned}$$

Since $\psi_{n,\delta}(x) = -\int_{\mathbb{R}} \psi(y) f_{n,1}(y-\delta-x)\mathrm{d}y$, by the LDCT, $\psi_{n,\delta}(x)$ is $r$ times differentiable, and $|\psi_{n,\delta}^{(i)}(x)| \le C(1+|x|)$. The conclusions about $\psi_{\infty,\delta}^{(i)}(x)$, $i \ge 1$, similarly follow. □

**Lemma 2.** *Let $m \in \mathbb{N}$. For a triangular array of $m$-vectors $\{\gamma_{i,n}, i = 1, 2, \ldots, n\}$ with $\max_{1\le i\le n} |\gamma_{i,n}| = O(1)$, we have for sufficiently small $\eta > 0$ that*

$$\left\| \sum_{i=1}^{n} \gamma_{i,n} e_i \right\|_{\alpha-\eta} = O(|d_n|). \tag{24}$$

**Proof.** We only need to consider $m = 1$, since the general case follows by considering each coordinate of $\sum_{i=1}^{n} \gamma_{i,n} e_i$. Write $a_i = 0$ if $i < 0$. Recall $d_n = n^{1-\gamma+1/\alpha} L_1(n) l(n)$.

Let $\chi_{j,n} = n^{\gamma-1}|l^{-1}(n)| \sum_{i=1}^{n} \gamma_{i,n} a_{i-j}$. By Lemma 3(a) in [2],

$$\mathbb{E}\left| \sum_{i=1}^{n} \frac{\gamma_{i,n} e_i}{d_n} \right|^{\alpha-\eta} = \mathbb{E}\left| \sum_{j=-\infty}^{\infty} \frac{\chi_{j,n}\varepsilon_j}{n^{1/\alpha} L_1(n)} \right|^{\alpha-\eta} \le \frac{C}{n} \max\left\{ \sum_j |\chi_{j,n}|^{\alpha-\eta}, \sum_j |\chi_{j,n}|^{\alpha+\eta} \right\}. \tag{25}$$

Write $\mathrm{I}_n = \sum_{j=-\infty}^{-n-1} |\chi_{j,n}|^{\alpha-\eta}$, $\mathrm{II}_n = \sum_{j=-n}^{0} |\chi_{j,n}|^{\alpha-\eta}$ and $\mathrm{III}_n = \sum_{j=1}^{n} |\chi_{j,n}|^{\alpha-\eta}$. By Karamata's theorem, $\mathrm{III}_n \le Cn[n^{\gamma-1} |l^{-1}(n)| \sum_{i=1}^{n} |i^{-\gamma} l(i)|]^{\alpha-\eta} \le Cn$. Let $S_a(n) = \sum_{i=1}^{n} |a_i|$. Then $S_a(n) = O(n^{1-\gamma}|l(n)|)$, and, for sufficiently small $\eta > 0$,

$$\begin{aligned} \mathrm{I}_n &\le C \sum_{j=-\infty}^{-n-1} [n^{\gamma-1}|l^{-1}(n)|(S_a(n-j) - S_a(-j))]^{\alpha-\eta} \\ &\le C \sum_{j=n+1}^{\infty} [n^{\gamma-1}|l^{-1}(n)| n |a_{1+j}|]^{\alpha-\eta} \le Cn. \end{aligned}$$

Similarly, $\mathrm{II}_n \le Cn$. So $\sum_j |\chi_{j,n}|^{\alpha-\eta} = O(n)$. Similar arguments also imply $\sum_j |\chi_{j,n}|^{\alpha+\eta} = O(n)$. Hence (24) follows from (25). □

**Lemma 3.** *Assume* (A1) *and* (A2) *with $r = 2$. Let $\alpha_1 \in (1/\gamma, \alpha)$ and $v \in [1, \alpha_1]$. Then*

$$\sup_{|\delta|\le 1} \|\psi_{n-1,\delta}(\underline{e}_{n,1}) - \psi_{n,\delta}(\underline{e}_{n,0}) - \psi'_{\infty,\delta}(0) a_{n-1}\varepsilon_1\|_v = O(|a_{n-1}|^{\alpha_1/v} + |a_{n-1}| A_n(\alpha_1)^{1/\alpha_1}), \tag{26}$$

*where $A_n(q) = \sum_{i=n}^{\infty} |a_i|^q$, $q > 0$.*

**Proof.** Let $\varpi = a_{n-1}\varepsilon_1$, $U = \psi_{n,\delta}(\underline{e}_{n,1}) - \psi_{n,\delta}(\underline{e}_{n,0}) - \psi'_{n,\delta}(\underline{e}_{n,0})\varpi$ and $V = \psi_{n,\delta}(\underline{e}_{n,1}) - \psi_{n-1,\delta}(\underline{e}_{n,1})$. Below we shall show that $\|U\|_v = O(|\alpha_{n-1}|^{\alpha_1/v})$, $\|V\|_v = O(|a_{n-1}|^{\alpha_1/v})$ and

$$\|\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{\infty,\delta}(0)\|_v = O(A_n^{1/\alpha_1}(\alpha_1)). \tag{27}$$

Then (26) follows from

$$\psi_{n-1,\delta}(\underline{e}_{n,1}) - \psi_{n,\delta}(\underline{e}_{n,0}) - \psi'_{\infty,\delta}(0)\varpi = U - V - (\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{\infty,\delta}(0))\varpi.$$

For $U$, note that,

$$\begin{aligned} \mathbb{E}[|U|^v]/2^{v-1} &\le \mathbb{E}[|U\mathbf{1}_{|\varpi|\le 1}|^v] + \mathbb{E}[|U\mathbf{1}_{|\varpi|>1}|^v] \\ &\le \mathbb{E}[|\psi''_{n,\delta}(\underline{e}_{n,0};1)|\varpi|^2 \mathbf{1}_{|\varpi|\le 1}|^v] + 3^{v-1}\mathbb{E}[|\psi_{n,\delta}(\underline{e}_{n,1})|^v \mathbf{1}_{|\varpi|>1}] \\ &\quad + 3^{v-1}\mathbb{E}[|\psi_{n,\delta}(\underline{e}_{n,0})|^v \mathbf{1}_{|\varpi|>1}] + 3^{v-1}\mathbb{E}[|\psi'_{n,0}(\underline{e}_{n,0})|^v |\varpi|^v \mathbf{1}_{|\varpi|>1}] \\ &=: \mathrm{I}_n^* + \mathrm{II}_n^* + \mathrm{III}_n^* + \mathrm{IV}_n^*. \end{aligned}$$

By Lemma 1, since $\underline{e}_{n,0}$ and $\varpi$ are independent and $2v > \alpha_1$, we have $\mathrm{I}_n^* = O(|a_{n-1}|^{\alpha_1})$. Similarly, $\mathrm{III}_n^* + \mathrm{IV}_n^* = O(|a_{n-1}|^{\alpha_1})$. Then $\|U\|_v = O(|\alpha_{n-1}|^{\alpha_1/v})$, since, again by Lemma 1,

$$\begin{aligned} \mathrm{II}_n^*/3^{v-1} &= \mathbb{E}[|\psi_{n,\delta}(\underline{e}_{n,0} + \varpi)|^v \mathbf{1}_{|\varpi>1|}] \\ &\le C\mathbb{E}[1 + |\underline{e}_{n,0}|]^v \mathbb{E}[(1+|\varpi|)^v \mathbf{1}_{|\varpi|>1}] \\ &\le C\mathbb{E}[(1+|\varpi|)^v \mathbf{1}_{|\varpi|>1}] \\ &\le C\mathbb{E}[|\varpi|^v \mathbf{1}_{|\varpi|>1}] \le C\mathbb{E}|\varpi|^{\alpha_1} = O(|a_{n-1}|^{\alpha_1}). \end{aligned}$$

Let $V^* = \psi_{n-1,\delta}(\underline{e}_{n,1} + a_{n-1}\varepsilon'_1) - \psi_{n-1,\delta}(\underline{e}_{n,1}) - \psi'_{n-1,\delta}(\underline{e}_{n,1}) a_{n-1}\varepsilon'_1$, where $\{\varepsilon'_i, i \in \mathbb{Z}\}$ is an iid copy of $\{\varepsilon_i, i \in \mathbb{Z}\}$. Similarly as $U$, $\|V^*\|_v = O(|a_{n-1}|^{\alpha_1/v})$. Hence, we have $\|V\|_v = O(|a_{n-1}|^{\alpha_1/v})$ in view of $\psi_{n,\delta}(x) - \psi_{n-1,\delta}(x) = \mathbb{E}[\psi_{n-1,\delta}(x + a_{n-1}\varepsilon'_1) - \psi_{n-1,\delta}(x) - \psi'_{n-1,\delta}(x) a_{n-1}\varepsilon'_1]$, and Jensen's inequality.

Now we show (27). By the LDCT, $\psi'_{\infty,\delta}(0) = \mathbb{E}[\psi'_{n,\delta}(\underline{e}_{n,0})]$. Let $\underline{e}^*_{n,0} = \sum_{i=-\infty}^{0} a_{n-i}\varepsilon'_i$. By the Bahr–Esseen inequality [31], $\|\underline{e}_{n,0}\|_{\alpha_1} = \|\underline{e}^*_{n,0}\|_{\alpha_1} \le CA_n(\alpha_1)^{1/\alpha_1}$. Note that,

$$\begin{aligned}
\|\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{\infty,\delta}(0)\|_v &= \|\mathbb{E}[\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{n,\delta}(\underline{e}^*_{n,0})|\mathcal{F}_0]\|_v \\
&\le \|\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{n,\delta}(\underline{e}^*_{n,0})\|_v \\
&\le \|\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{n,\delta}(0)\|_v + \|\psi'_{n,\delta}(\underline{e}^*_{n,0}) - \psi'_{n,\delta}(0)\|_v \\
&= 2\|\psi'_{n,\delta}(\underline{e}_{n,0}) - \psi'_{n,\delta}(0)\|_v \\
&\le 2\mathrm{I}'_n + 2\mathrm{II}'_n + 2\mathrm{III}'_n,
\end{aligned}$$

where, by Taylor's expansion, $\mathrm{I}'_n = \|\psi''_{n,\delta}(0;1)\underline{e}_{n,0}\mathbf{1}_{|\underline{e}_{n,0}|\le 1}\|_v$, $\mathrm{II}'_n = \|\psi'_{n,\delta}(0)\mathbf{1}_{|\underline{e}_{n,0}|>1}\|_v$ and $\mathrm{III}'_n = \|\psi'_{n,\delta}(\underline{e}_{n,0})\mathbf{1}_{|\underline{e}_{n,0}|>1}\|_v$. Note that, $\mathrm{I}'_n \le C\|\underline{e}_{n,0}\|_v \le C\|\underline{e}_{n,0}\|_{\alpha_1} = O(A_n^{1/\alpha_1}(\alpha_1))$, $\mathrm{II}'_n \le C\mathbb{P}(|\underline{e}_{n,0}| \ge 1)^{1/v} \le C\|\underline{e}_{n,0}\|_{\alpha_1}^{1/v} = O(A_n^{1/v}(\alpha_1))$ and $\mathrm{III}'_n$ satisfies

$$\begin{aligned}
\mathrm{III}'_n &\le C\mathbb{E}[(1+|\underline{e}_{n,0}|)^v\mathbf{1}_{|\underline{e}_{n,0}|>1}]^{1/v} \le C\mathbb{E}[(1+|\underline{e}_{n,0}|)^{\alpha_1}\mathbf{1}_{|\underline{e}_{n,0}|>1}]^{1/v} \\
&\le C\mathbb{E}[|\underline{e}_{n,0}|^{\alpha_1}]^{1/v} = O(A_n^{1/v}(\alpha_1)) = O(A_n^{1/\alpha_1}(\alpha_1)).
\end{aligned}$$

So (27) holds. □

**Proposition 1.** *Let $v > \alpha_1/(2\alpha_1\gamma - 1)$. Under the assumptions of Lemma 3, for any triangular array of $m$-vectors $\{\mathbf{c}_{in}, i = 1, 2, \dots, n\}$, $m \in \mathbb{N}$, we have*

$$\sup_{|\delta|\le 1}\left\|\sum_{i=1}^{n}[\psi(e_i+\delta) - \mathbb{E}[\psi(e_i+\delta)] - \phi'(\delta)e_i]\mathbf{c}_{in}\right\|_v = O(n^{1/2-\gamma'+1/v}\varsigma^{1/2}(n)), \tag{28}$$

*where $1/v < \gamma' < \gamma_0 := \min\{(\alpha_1\gamma)/v, 2\gamma - 1/\alpha_1\}$ and $\varsigma(n) = \sum_{i=1}^n |\mathbf{c}_{in}|^2$.*

**Proof.** Let $T_i = \psi(e_i+\delta) - \mathbb{E}[\psi(e_i+\delta)] - \phi'(\delta)e_i$, $\omega_i = \|\mathcal{P}_1 T_i\|_v$ and $\Omega_i = \sum_{j=i}^{\infty}\omega_j^v$. Then $\sum_{i=1}^n T_i\mathbf{c}_{in} = \sum_{j=-\infty}^{n}\mathcal{P}_j[\sum_{i=1}^n T_i\mathbf{c}_{in}]$ and $\{\mathcal{P}_j[\sum_{i=1}^n T_i\mathbf{c}_{in}], j \in \mathbb{Z}\}$ is a sequence of martingale differences. By the Bahr–Esseen inequality [31],

$$\begin{aligned}
\left\|\sum_{i=1}^n T_i\mathbf{c}_{in}\right\|_v^v &\le 2\sum_{j=-\infty}^{n}\left\|\mathcal{P}_j\left[\sum_{i=1}^n T_i\mathbf{c}_{in}\right]\right\|_v^v \\
&\le 2\sum_{j=-\infty}^{n}\left[\sum_{i=1}^n \omega_{i-j+1}|\mathbf{c}_{in}|\right]^v \le 2\sum_{j=-\infty}^{n}\varsigma^{v/2}(n)\left[\sum_{i=1}^n \omega_{i-j+1}^2\right]^{v/2} \\
&= 2\varsigma^{v/2}(n)\left\{\left[\sum_{j=-\infty}^{-n} + \sum_{j=-n+1}^{0} + \sum_{j=1}^{n}\right]\left(\sum_{i=1}^n \omega_{i-j+1}^2\right)^{v/2}\right\} \\
&=: 2\varsigma^{v/2}(n)(I_n^{**} + II_n^{**} + III_n^{**}).
\end{aligned}$$

By Lemma 3 and since $\mathcal{P}_1 T_n = \psi_{n-1,\delta}(\underline{e}_{n,1}) - \psi_{n,\delta}(\underline{e}_{n,0}) - \psi'_{\infty,\delta}(0)a_{n-1}\varepsilon_1$, we have $\omega_n = O(n^{-\gamma'})$ and $\Omega_n = O(n^{1-v\gamma'})$. Using similar arguments as in the proof of Lemma 2, we have $I_n^{**} + II_n^{**} + III_n^{**} = O(n^{v/2+1-v\gamma'})$. So (28) follows. □

**Remark 5.** Clearly $1/v < \gamma_0$ if and only if $v > \alpha_1/(2\alpha_1\gamma - 1)$. If so, $\gamma' \in (1/v, \gamma_0)$ exists.

**Corollary 2.** *Assume* (A1) *and* (A2) *with $r = 2$. Then for any triangular array of $m$-vectors $\{\mathbf{c}_{in},\ i = 1, 2, \dots, n\}$, we have for some $v_0 \in (1, \alpha)$ that*

$$\sup_{|\delta|\le 1}\left\|\sum_{i=1}^{n}[\psi(e_i+\delta) - \mathbb{E}[\psi(e_i+\delta)] - \phi'(\delta)e_i]\mathbf{c}_{in}\right\|_{v_0} = O(n^{\eta}\varsigma^{1/2}(n)) \tag{29}$$

*for all $\eta$ satisfying $1/2 - \gamma + 1/\alpha > \eta > 1/2 - \gamma + 1/\alpha - (\gamma - 1/\alpha)^2/\gamma$.*

**Proof.** In (28), let $v = (\gamma\alpha_1^2)/(2\gamma\alpha_1 - 1)$ and $\alpha_1 \uparrow \alpha$. Then the order is arbitrarily close to $n^{\eta_0}\varsigma^{1/2}(n)$ with $\eta_0 = 1/2 - \gamma + 1/\alpha - (\gamma - 1/\alpha)^2/\gamma$. So Corollary 2 holds. □

**Proof of Theorem 1.** Let $z_{\theta,i,n} = \mathbf{z}_{i,n}^T\theta$, where $|\theta| \le Ck_n$, $\theta \in \mathbb{R}^p$. Recall $k_n = |d_n|m_n$. By (A4), $\max_{1\le i\le n}|z_{\theta,i,n}| \to 0$. By Corollary 2, for $n$ large enough,

$$\sup_{0\le t\le 1}\left\|\sum_{i=1}^{n}[\psi(e_i - z_{\theta,i,n}t) - \mathbb{E}[\psi(e_i - z_{\theta,i,n}t)] - \phi'(-z_{\theta,i,n}t)e_i]z_{\theta,i,n}\right\|_{v_0} = O(n^{\eta}k_n), \tag{30}$$

where $\eta$ is defined in Corollary 2. Note that, $n^\eta k_n = o(k_n^2)$. By Lemma 1, $\sup_{0\le t\le 1}|\phi'(0)-\phi'(-z_{\theta,i,n}t)| = O(|z_{\theta,i,n}|)$. Since $m_n k_n = m_n^2|d_n| \to 0$, by Lemma 2,

$$\sup_{0\le t\le 1}\left\|\sum_{i=1}^n z_{\theta,i,n}[\phi'(0)-\phi'(-z_{\theta,i,n}t)]e_i\right\|_{\alpha-\eta} \le C(m_n k_n)^2|d_n| = o(k_n^2). \tag{31}$$

Let $\pi_i(\theta) = \rho(e_i - z_{\theta,i,n}) - \rho(e_i) + z_{\theta,i,n}\phi'(0)e_i$. Note that,

$$\sum_{i=1}^n[\pi_i(\theta) - \mathbb{E}\pi_i(\theta)] = -\int_0^1\sum_{i=1}^n[\psi(e_i - z_{\theta,i,n}t) - \mathbb{E}[\psi(e_i - z_{\theta,i,n}t)] - \phi'(-z_{\theta,i,n}t)e_i]z_{\theta,i,n}\mathrm{d}t$$
$$+\int_0^1\sum_{i=1}^n z_{\theta,i,n}[\phi'(0)-\phi'(-z_{\theta,i,n}t)]e_i\mathrm{d}t.$$

Hence by (30) and (31), we have for any fixed $\theta$ with $|\theta|\le Ck_n$ that

$$\left|\sum_{i=1}^n[\pi_i(\theta) - \mathbb{E}\pi_i(\theta)]\right| = o_p(k_n^2). \tag{32}$$

By Lemma 1 in [4],

$$\begin{aligned}\mathbb{E}[\pi_i(\theta)] &= \frac{1}{2}\phi'(0)\sum_{i=1}^n|\mathbf{z}_{i,n}^T\theta|^2 + o\left(\sum_{i=1}^n|\mathbf{z}_{i,n}^T\theta|^2\right)\\ &= \frac{1}{2}\phi'(0)\sum_{i=1}^n\theta^T\mathbf{z}_{i,n}\mathbf{z}_{i,n}^T\theta + o\left(\sum_{i=1}^n\theta^T\mathbf{z}_{i,n}\mathbf{z}_{i,n}^T\theta\right)\\ &= \frac{1}{2}\phi'(0)|\theta|^2 + o(k_n^2).\end{aligned} \tag{33}$$

Hence, we have by (32) and (33) that

$$\left|\sum_{i=1}^n[\rho(e_i - \mathbf{z}_{i,n}^T\theta) - \rho(e_i) + \phi'(0)\mathbf{z}_{i,n}^T\theta e_i] - \phi'(0)|\theta|^2/2\right| = o_p(k_n^2). \tag{34}$$

Now a standard argument using properties of convex functions entails $|\hat\theta_n - U_n| = o_p(k_n)$; see the proofs of Theorems 2.2 and 2.4 in [4]. Details are omitted. By (9) and (A4), since $\hat\beta_n = \Sigma_n^{-1/2}\hat\theta_n$ and $\hat\beta_{n,ls} = \Sigma_n^{-1/2}U_n$, the rest of the theorem easily follows. □

**Proof of Theorem 2.** Let $\psi_1(x) = \mathbb{E}[\psi(x+\varepsilon_i)]$ and $\tilde e_i = e_i - \varepsilon_i$. From Lemma 1 we see that $\psi_1(\cdot)$ and $\phi(\cdot)$ are $r = p+1$ times differentiable, and $|\psi_1^{(i)}(x)| + |\phi^{(i)}(x)| \le C(1+|x|)$ for $i = 0,\ldots,r$. Note that, $\Xi_n(\theta) = \sum_{i=1}^n\psi(e_i - z_{\theta,i,n})\mathbf{z}_{i,n}$, $z_{\theta,i,n} = \mathbf{z}_{i,n}^T\theta$. Then

$$\Delta_n(\theta) := \Xi_n(\theta) - \mathbb{E}[\Xi_n(\theta)] = M_n(\theta) + N_n(\theta) + G_n(\theta), \tag{35}$$

where

$$M_n(\theta) = \sum_{i=1}^n[\psi(e_i - z_{\theta,i,n})] - [\psi_1(\tilde e_i - z_{\theta,i,n})]\mathbf{z}_{i,n},$$
$$N_n(\theta) = \sum_{i=1}^n[\psi_1(\tilde e_i - z_{\theta,i,n}) - \phi(-z_{\theta,i,n}) - \phi'(-z_{\theta,i,n})\tilde e_i]\mathbf{z}_{i,n},$$
$$G_n(\theta) = \sum_{i=1}^n\phi'(-z_{\theta,i,n})\tilde e_i\mathbf{z}_{i,n}.$$

The summands of $M_n(\theta)$ form an $\mathcal{L}^2$ martingale difference with respect to the filtration $\sigma(\mathcal{F}_i)$. We will use Lemma 4 in [33] to bound the oscillation rate of $M_n(\theta)$. Our Lemma 1 implies that condition (A5) in [33] holds. On the other hand, since $r_nk_n\to\infty$ and $r_nk_nm_n\to 0$, his condition (16) will be satisfied if we choose $r_nk_n$ as $\delta_n$ there. Thus,

$$\sup_{|\theta|\le r_nk_n}|M_n(\theta) - M_n(0)| = O_p(\tau_n^{1/2}(r_nk_n)\log n + n^{-3}). \tag{36}$$

Let $J = \{j_1, \ldots, j_q\} \subseteq \{1, \ldots, p\}$ be a nonempty index set, $1 \le j_1 < \cdots < j_q$. For $\mathbf{u} = (u_1, \cdots, u_p) \in \mathbb{R}^p$, let $\mathbf{u}_J = (u_1\mathbf{1}_{1\in J}, \ldots, u_p\mathbf{1}_{p\in J})$. Write $\mathbf{w}_{i,J} = \mathbf{z}_{i,n} \times z_{i,j_1,n} \times \cdots \times z_{i,j_q,n}$. Recall that $\mathbf{z}_{i,n} = (z_{i,1,n}, \ldots, z_{i,p,n})^T$. Write

$$\int_0^{\theta_J} \frac{\partial^q N_n(\mathbf{u}_J)}{\partial \mathbf{u}_J} \mathrm{d}\mathbf{u}_J = \int_0^{\theta_{j_1}} \cdots \int_0^{\theta_{j_q}} \frac{\partial^q N_n(\mathbf{u}_J)}{\partial u_{j_1} \cdots \partial u_{j_q}} \mathrm{d}u_{j_1} \cdots \mathrm{d}u_{j_q}.$$

By Lemma 1 and the LDCT, $\phi^{(q)}(-z_{\mathbf{u}_J,i,n}) = \mathbb{E}[\psi_1^{(q)}(\tilde{e}_i - z_{\mathbf{u}_J,i,n})]$. If $|\mathbf{u}| \le p^{1/2} r_n k_n$, then $\max_{1\le i\le n} |z_{\mathbf{u}_J,i,n}| = O(r_n k_n m_n) \to 0$. By similar arguments as those of Corollary 2,

$$\begin{aligned}
\left\| \frac{\partial^q N_n(\mathbf{u}_J)}{\partial \mathbf{u}_J} \right\|_{v_0} &= \left\| \sum_{i=1}^n [\psi_1^{(q)}(\tilde{e}_i - z_{\mathbf{u}_J,i,n}) - \phi^{(q)}(-z_{\mathbf{u}_J,i,n}) - \phi^{(q+1)}(-z_{\mathbf{u}_J,i,n})\tilde{e}_i]\mathbf{w}_{i,J} \right\|_{v_0} \\
&= O\left( n^\eta \left( \sum_{i=1}^n |\mathbf{w}_{i,J}|^2 \right)^{1/2} \right) = O(n^\eta s_n^{1/2}(2+2q)),
\end{aligned} \tag{37}$$

uniformly over $|\mathbf{u}| \le p^{1/2} r_n k_n$. Hence

$$\begin{aligned}
\left\| \sup_{|\theta|\le r_n k_n} \int_0^{\theta_J} \left| \frac{\partial^q N_n(\mathbf{u}_J)}{\partial \mathbf{u}_J} \right| \mathrm{d}\mathbf{u}_J \right\|_{v_0} &\le \left\| \int_{-r_n k_n}^{r_n k_n} \cdots \int_{-r_n k_n}^{r_n k_n} \left| \frac{\partial^q N_n(\mathbf{u}_J)}{\partial \mathbf{u}_J} \right| \mathrm{d}\mathbf{u}_J \right\|_{v_0} \\
&\le \int_{-r_n k_n}^{r_n k_n} \cdots \int_{-r_n k_n}^{r_n k_n} \left\| \frac{\partial^q N_n(\mathbf{u}_J)}{\partial \mathbf{u}_J} \right\|_{v_0} \mathrm{d}\mathbf{u}_J \\
&= O(r_n^q k_n^q n^\eta s_n^{1/2}(2+2q)).
\end{aligned} \tag{38}$$

Since $s_n^{1/2}(2+2q) \le m_n^{q-1} s_n^{1/2}(4)$ and $r_n k_n m_n = o(1)$, (38) implies

$$\begin{aligned}
\left\| \sup_{|\theta|\le r_n k_n} |N_n(\theta) - N_n(0)| \right\|_{v_0} &= \left\| \sup_{|\theta|\le r_n k_n} \left| \sum_{J\subseteq\{1,\ldots,p\}} \int_0^{\theta_J} [\partial^{|J|} N_n(\mathbf{u}_J)/\mathbf{u}_J] \mathrm{d}\mathbf{u}_J \right| \right\|_{v_0} \\
&= O(r_n k_n n^\eta s_n^{1/2}(4)).
\end{aligned} \tag{39}$$

Similarly, $\phi'(\cdot)$ is $p$ times differentiable, and $|\phi^{r+1}(x)| \le C(1+|x|)$ for $r = 0, \ldots, p$. Furthermore, by Lemma 2, for any $\mathbf{u}$ with $|\mathbf{u}| \le p^{1/2} r_n k_n$,

$$\left\| \frac{\partial^q G_n(\mathbf{u}_J)}{\partial \mathbf{u}_J} \right\|_{\alpha-\eta} = \left\| \sum_{i=1}^n \phi^{q+1}(-z_{\mathbf{u}_J,i,n}) \tilde{e}_i \mathbf{w}_{i,J} \right\|_{\alpha-\eta} = O(m_n^{q+1}|d_n|).$$

Therefore, we obtain by $r_n k_n m_n = o(1)$ that

$$\left\| \sup_{|\theta|\le r_n k_n} |G_n(\theta) - G_n(0)| \right\|_{\alpha-\eta} \le C r_n k_n m_n^2 |d_n| = O(r_n m_n k_n^2). \tag{40}$$

Since $s_n(4) = O(m_n^2)$ and $n^\eta = o(k_n)$, by (36), (39) and (40),

$$\sup_{|\theta|\le r_n k_n} |\Delta_n(\theta) - \Delta_n(0)| = O_p(\tau_n^{1/2}(r_n k_n)\log n + r_n m_n k_n^2). \tag{41}$$

Since $\hat{\theta}_n = O_p(k_n)$, we have by (41) that

$$|\Delta_n(\hat{\theta}_n) - \Delta_n(0)| = O_p(\tau_n^{1/2}(r_n k_n)\log n + r_n m_n k_n^2). \tag{42}$$

Furthermore, by Lemma 1 and Taylor's expansion, we get

$$\begin{aligned}
\sum_{i=1}^n \phi(-\mathbf{z}_{i,n}^T \hat{\theta}_n)\mathbf{z}_{i,n} &= -\phi'(0) \sum_{i=1}^n \mathbf{z}_{i,n}\mathbf{z}_{i,n}^T \hat{\theta}_n + \sum_{i=1}^n O(|\mathbf{z}_{i,n}^T \hat{\theta}_n|^2 |\mathbf{z}_{i,n}|) \\
&= -\phi'(0)\hat{\theta}_n + O\left( m_n \sum_{i=1}^n \hat{\theta}_n^T \mathbf{z}_{i,n}\mathbf{z}_{i,n}^T \hat{\theta}_n \right) \\
&= -\phi'(0)\hat{\theta}_n + O_p(m_n k_n^2).
\end{aligned} \tag{43}$$

Plugging (43) into (42), since $m_n = o(r_n m_n k_n^2)$, we have (12) in view of (A5). $\square$

**Proof of Theorem 3.** Applying Corollary 2 with $\delta = 0$ and $\mathbf{c}_{in} = \mathbf{z}_{i,n}$, we have

$$|V_n - \phi'(0)U_n| = \left|\sum_{i=1}^{n}[\psi(e_i) - \phi'(0)e_i]\mathbf{z}_{i,n}\right| = O_p(n^{\eta}).$$

By (12), (i) follows. Since $\hat{\beta}_n - \hat{\beta}_{n,ls} = \Sigma_n^{-1/2}(\hat{\theta}_n - U_n)$, by (9), we have (ii). □

To prove Theorem 4, we need the following lemma.

**Lemma 4.** *Under the conditions of Theorem 4, we have*

$$d_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i e_i \Rightarrow \int_{-\infty}^{\infty}\left[\int_0^1 \mathbf{g}(x)(x-u)_+^{-\gamma}\mathrm{d}x\right]\mathrm{d}\varepsilon_\alpha(u). \tag{44}$$

**Proof.** Let $\mathbf{h}_n(u) = (h_{\mathbf{x}^1}(u), \dots, h_{\mathbf{x}^p}(u))^T$. For $c \in \mathbb{R}^p$ with $|c| = 1$, let $h_{n,c}(u) = c^T\mathbf{h}_n(u)$ and $g_c(u) = c^T\mathbf{g}(u)$. By the Cramer–Wold device, to prove (44), it suffices to show that

$$d_n^{-1}\sum_{i=1}^{n} h_{n,c}\left(\frac{i-1}{n}\right)e_i \Rightarrow \int_{-\infty}^{\infty}\left[\int_0^1 g_c(x)(x-u)_+^{-\gamma}\mathrm{d}x\right]\mathrm{d}\varepsilon_\alpha(u). \tag{45}$$

To prove (45), we shall apply Theorem 4 in [2]. By (B1), $h_{n,c}(\cdot)$ converges to $g_c(\cdot)$ uniformly on [0,1] and $g_c(\cdot)$ is continuous. Let $a_i = 0$ if $i < 0$, $l^*(x) = l(x)$ if $x \geq 0$ and $l^*(x) = 0$ if $x < 0$. Define

$$\zeta_n(u) = \sum_{i=1}^{n} h_{n,c}\left(\frac{i-1}{n}\right)(i - \lfloor nu \rfloor)_+^{-\gamma} l^*(i - \lfloor nu \rfloor).$$

Interpret $0^{-\gamma} = 0$ in the above definition. Let $\varepsilon_i^* = \varepsilon_i/(n^{1/\alpha}L_1(n))$. Then

$$\begin{aligned}\sum_{i=1}^{n} h_{n,c}\left(\frac{i-1}{n}\right)e_i/(n^{1/\alpha}L_1(n)) &= \sum_{j=-\infty}^{\infty}\left[\sum_{i=1}^{n} h_{n,c}\left(\frac{i-1}{n}\right)a_{i-j}\right]\varepsilon_j^* \\ &= \sum_{j=-\infty}^{\infty}\zeta_n(j/n)\varepsilon_j^* + \vartheta(n),\end{aligned} \tag{46}$$

where $\vartheta(n) = \sum_{i=1}^{n} h_{n,c}((i-1)/n)\varepsilon_i^*$. Since $\max_{1\leq i\leq n}|h_{n,c}((i-1)/n)| = O(1)$, it is easy to see that $|\vartheta(n)| = O_p(1)$.

Let $\zeta_n^*(u) = \zeta_n(u)/(n^{1-\gamma}l(n))$. By (46),

$$d_n^{-1}\sum_{i=1}^{n} h_{n,c}\left(\frac{i-1}{n}\right)e_i = \sum_{j=-\infty}^{\infty}\zeta_n^*(j/n)\varepsilon_j^* + o_p(1). \tag{47}$$

By the uniform convergence theorem for slowly varying functions (see Theorem 1.2.1 of [7]),

$$\zeta_n^*(u) \to \int_0^1 g_c(x)(x-u)_+^{-\gamma}\mathrm{d}x \tag{48}$$

point-wise on $\mathbb{R}$. On the other hand, we see that for $n \in \mathbb{N}$,

$$|\zeta_n^*(u)|^{\alpha\pm\eta} + \left|\int_0^1 g_c(x)(x-u)_+^{-\gamma}\mathrm{d}x\right|^{\alpha\pm\eta} \leq C\left(\int_0^1 (x-u)_+^{-\gamma}\mathrm{d}x\right)^{\alpha\pm\eta}.$$

Note that, the right-hand side of the above inequality is integrable over $\mathbb{R}$ for sufficiently small positive $\eta$. Thus (48) also holds in the sense of convergence in $L_{\alpha,\eta} = \{f : \|f\|_{\alpha,\eta} < \infty\}$ [2], where $\|f\|_{\alpha,\eta} = \max(\|f\|_{\alpha-\eta}, \|f\|_{\alpha+\eta})^{\alpha-\eta}$, $\|f\|_q = [\int_{\mathbb{R}} |f(x)|^q \mathrm{d}x]^{1/q}$. Hence by their Theorem 4, we have

$$\sum_{j=-\infty}^{\infty}\zeta_n^*(j/n)\varepsilon_j^* \Rightarrow \int_{-\infty}^{\infty}\left[\int_0^1 g_c(x)(x-u)_+^{-\gamma}\mathrm{d}x\right]\mathrm{d}\varepsilon_\alpha(u). \tag{49}$$

Together with (47), we conclude that Lemma 4 holds. □

**Proof of Theorem 4.** Let $\kappa_n = n^{-1/2}d_n$. Under assumptions of Theorem 4, there exists $0 < C_1 \le C_2 < \infty$, such that $C_1 n^{-1/2} \le m_n \le C_2 n^{-1/2}$ for sufficiently large $n$. Hence, we have $k_n/C_2 \le |\kappa_n| \le k_n/C_1$. Recall that $k_n = m_n|d_n|$. Therefore, by Theorem 1,

$$|\kappa_n^{-1}\hat{\theta}_n - \kappa_n^{-1}U_n| = o_p(1), \quad \text{where } \kappa_n^{-1}U_n = n^{1/2}\Sigma_n^{-1/2}d_n^{-1}\sum_{i=1}^n \mathbf{x}_i e_i. \tag{50}$$

Recall $\mathcal{L}_{\mathbf{g}}(\varepsilon) = \int_{-\infty}^{\infty}[\int_0^1 \mathbf{g}(x)(x-u)_+^{-\gamma}\,\mathrm{d}x]d\varepsilon_\alpha(u)$. By Lemma 4 and (50),

$$\kappa_n^{-1}\hat{\theta}_n \Rightarrow \mathcal{G}^{-1/2}\mathcal{L}_{\mathbf{g}}(\varepsilon) \tag{51}$$

since $n^{1/2}\Sigma_n^{-1/2} \to \mathcal{G}^{-1/2}$. Noting that $\hat{\theta}_n = \Sigma_n^{1/2}\hat{\beta}_n$, Theorem 4 follows. □

## Acknowledgment

We are grateful to the referee for his/her many helpful comments.

## References

[1] M.A. Arcones, The Bahadur–Kiefer representations of the least absolute deviation regression estimator, Annals of the Institute of Statistical Mathematics 50 (1996) 87–117.
[2] F. Avram, M.S. Taqqu, Weak convergence of moving averages with infinite variance, in: Eberlein, Taqqu (Eds.), Dependence in Probability and Statistics: A Survey of Recent Results, Birkhauser, Boston, 1986, pp. 399–416.
[3] G.J. Babu, Strong representation for LAD estimators in liner models, Probability Theory and Related Fields 83 (1989) 547–558.
[4] Z.D. Bai, C.R. Rao, Y. Wu, $M$-estimation of multivariate linear regression parameters under a convex discrepancy function, Statistica Sinica 2 (1992) 237–254.
[5] G. Bassett, R. Koenker, Asymptotic theory of least absolute error regression, Journal of the American Statistical Association 73 (1978) 618–622.
[6] J. Beran, $M$-estimators of location for Gaussian and related processes with slowly decaying serial correlations, Journal of the American Statistical Association 86 (1991) 704–708.
[7] N.H. Bingham, C.M. Goldie, J.L. Teugels, Regular Variation, Cambridge University Press, Cambridge, 1987.
[8] J.M. Chambers, C.L. Mallows, B.W. Stuck, A method for simulating stable random variables, Journal of the American Statistical Association 71 (1976) 340–344. Correction: (1987) Journal of the American Statistical Association, 82, 704.
[9] Y.S. Chow, H. Teicher, Probability Theory, Springer Verlag, New York, 1988.
[10] H. Cui, X. He, K.W. Ng, $M$-estimation for linear models with spatially-correlated errors, Statistics & Probability Letters 66 (2004) 383–393.
[11] R. Davidson, J.G. MacKinnon, Econometric Theory and Methods, Oxford University Press, New York, 2003.
[12] F. Eicker, Asymptotic normality and consistency of the least squares estimators for families of linear regressions, The Annals of Mathematical Statistics 34 (1963) 447–456.
[13] J.L. Gastwirth, H. Rubin, The behavior of robust estimators on dependent data, The Annals of Statistics 3 (1975) 1070–1100.
[14] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, Robust Statistics, Wiley, New York, 1986.
[15] X. He, Q.M. Shao, A general Bahadur representation of $M$-estimators and its application to linear regression with nonstochastic designs, The Annals of Statistics 24 (1996) 2608–2630.
[16] J.R.M. Hosking, Fractional differencing, Biometrika 68 (1981) 165–176.
[17] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[18] I.A. Ibragimov, Yu.V. Linnik, Independent and Stationary Sequences of Random Variables, Wolters-Noordhoff, Groningen, 1971.
[19] R. Koenker, Quantile Regression, Cambridge University Press, Cambridge, 2005.
[20] P.S. Kokoszka, M.S. Taqqu, Parameter estimation for infinite variance fractional ARIMA, The Annals of Statistics 24 (1996) 1880–1913.
[21] H. Koul, D. Surgailis, Asymptotics of empirical processes of long memory moving averages with infinite variance, Stochastic Processes and their Applications 100 (2001) 255–274.
[22] T. McElroy, D. Politis, Self-normalization for heavy-tailed time series with long memory, Statistica Sinica 17 (2007) 199–220.
[23] P.C.B. Phillips, A shortcut to LAD estimator asymptotics, Econometric Theory 7 (1991) 450–463.
[24] D.N. Politis, J.P. Romano, M. Wolf, Subsampling, Springer, New York, 1999.
[25] S. Rachev, S. Mittnik, Stable Paretian Models in Finance, Wiley, Chichester, 2000.
[26] C.R. Rao, H. Toutenburg, Linear Models, 2nd edition, Springer, New York, 2005.
[27] S. Resnick, C. Stărică, Consistency of Hill's estimator for dependent data, Journal of Applied Probability 32 (1995) 139–167.
[28] G. Samorodnitsky, M.S. Taqqu, Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance. Stochastic Modeling, Chapman & Hall, New York, 1994.
[29] A.V. Skorokhod, Limit theorems for stochastic processes with independent increments, Theory of Probability and its Applications 2 (1957) 138–171.
[30] M.S. Taqqu, V. Teverovsky, On estimating the intensity of long-range dependence in finite and infinite variance time series, in: R.J. Adler, R.E. Feldman, M.S. Taqqu (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhäuser, 1998, pp. 177–217.
[31] B. von Bahr, C.G. Esseen, Inequalities for the $r$th absolute moment of a sum of random variables, $1 \le r \le 2$, The Annals of Mathematical Statistics 36 (1965) 299–303.
[32] W. Willinger, V. Paxson, M.S. Taqqu, Self-similarity and heavy tails: structural modeling of network traffic, in: R.J. Adler, R.E. Feldman, M.S. Taqqu (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhäuser, 1998, pp. 27–53.
[33] W.B. Wu, $M$-estimation of linear models with dependent errors, The Annals of Statistics 35 (2007) 495–521.
[34] W.B. Wu, G. Michailidis, D. Zhang, Simulating sample paths of linear fractional stable motion, IEEE Transactions on Information Theory 50 (6) (2004) 1086–1096.
[35] R. Zeckhauser, M. Thompson, Linear regression with non-normal error terms, Review Economics and Statistics 52 (1970) 280–286.