Hu *Journal of Inequalities and Applications* (2018) 2018:123
https://doi.org/10.1186/s13660-018-1715-x

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

Bahadur representations of M-estimators and their applications in general linear models

Hongchang Hu[1*]

*Correspondence:
retutome@163.com
[1]School of Mathematics and Statistics, Hubei Normal University, Huangshi, China

Abstract

Consider the linear regression model

$$y_i = x_i^T \beta + e_i, \quad i = 1, 2, \ldots, n,$$

where $e_i = g(\ldots, \varepsilon_{i-1}, \varepsilon_i)$ are general dependence errors. The Bahadur representations of M-estimators of the parameter β are given, by which asymptotically the theory of M-estimation in linear regression models is unified. As applications, the normal distributions and the rates of strong convergence are investigated, while $\{\varepsilon_i, i \in Z\}$ are m-dependent, and the martingale difference and (ε, ψ)-weakly dependent.

MSC: 62J05; 62F35; 62M10

Keywords: Linear regression models; M-estimate; Bahadur representation; Normal distribution; Rate of strong convergence

1 Introduction

Consider the following linear regression model:

$$y_i = x_i^T \beta + e_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where $\beta = (\beta_1, \ldots, \beta_p)^T \in R^p$ is an unknown parametric vector, x_i^T denotes the ith row of an $n \times p$ design matrix X, and $\{e_i\}$ are stationary dependence errors with a common distribution.

An M-estimate of β is defined as any value of β minimizing

$$\sum_{i=1}^{n} \rho\bigl(y_i - x_i^T \beta\bigr) \tag{1.2}$$

for a suitable choice of the function ρ, or any solution for β of the estimating equation

$$\sum_{i=1}^{n} \psi\bigl(y_i - x_i^T \beta\bigr) x_i = 0 \tag{1.3}$$

for a suitable choice of ψ.

There is a body of statistical literature dealing with linear regression models with independent and identically distributed (i.i.d.) random errors, see e.g. Babu [1], Bai et al. [2], Chen [7], Chen and Zhao [8], He and Shao [24], Gervini and Yohai [23], Huber and Ronchetti [28], Xiong and Joseph [50], Salibian-Barrera et al. [44]. Recently, linear regression models with serially correlated errors have attracted increasing attention from statisticians; see, for example, Li [33], Wu [49], Maller [38], Pere [41], Hu [25, 26]. Over the last 40 years, M-estimators in linear regression models have been investigated by many authors. Let $\{\eta_i\}$ be i.i.d. random variables. Koul [30] discussed the asymptotic behavior of a class of M-estimators in the model (1.1) with long range dependence errors $e_i = G(\eta_i)$. Wu [49] and Zhou and Shao [52] discussed the model (1.1) with $e_i = G(\ldots, \eta_{i-1}, \eta_i)$ and derived strong Bahadur representations of M-estimators and a central limit theorem. Zhou and Wu [53] considered the model (1.1) with $e_i = \sum_{j=0}^{\infty} a_j \eta_{i-j}$, and obtained some asymptotic results including consistency of robust estimates. Fan et al. [20] investigated the model (1.1) with the errors $e_i = f(e_{i-1}) + \eta_i$ and established the moderate deviations and strong Bahadur representations for M-estimators. Wu [47] discussed strong consistency of an M-estimator in the model (1.1) for negatively associated samples. Fan [19] considered the model (1.1) with φ-mixing errors, and the moderate deviations for the M-estimators. In addition, Berlinet et al. [4], Boente and Fraiman [5], Chen et al. [6], Cheng et al. [9], Gannaz [22], Lô and Ronchetti [37], Valdora and Yohai [45] and Yang [51] have also studied some asymptotic properties of M-estimators in nonlinear models. However, no people have investigated a unified the theory of M-estimation in linear regression models with more general errors.

In this paper, we assume that

$$e_i = g(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{1.4}$$

where $g(\cdot)$ is a measurable function such that e_i is a proper random variable, and $\{\varepsilon_i, i \in Z\}$ (where Z is the set of integers) are very general random variables, including m-dependent, martingale difference, (ε, ψ)-weakly dependent, and so on.

We try to investigate the unified the theory of M-estimation in the linear regression model. In the article, we use the idea of Wu [49] to study the Bahadur representative of M-estimator, and we extend some results to general errors. The paper is organized as follows. In Sect. 2, the weak and strong linear representation of an M-estimate of the vector regression parameter β in the model (1.1) are presented. Section 3 contains some applications of our results, including the m-dependent, (ε, ψ)-weakly dependent, martingale difference. In Sect. 4, proofs of the main results are given.

2 Main results

In the section, we investigate the weak and strong linear representation of an M-estimate of the vector regression parameter β in the model (1.1). Without loss of generality, we assume that the true parameter $\beta = 0$. We start with some notation and assumptions.

For a vector $v = (v_1, \ldots, v_p)$, let $|v| = (\sum_{i=1}^{p} {v_i}^2)^{\frac{1}{2}}$. A random vector V is said to be in $L^q, q > 0$, if $E(|V|^q) < \infty$. Let $\|V\|_q = E(|V|^q)^{\frac{1}{q}}$, $\|V\| = \|V\|_2$, $\Sigma_n = \sum_{i=1}^{n} x_i x_i^T = X^T X$ and assume that Σ_n is positive definite for large enough n. Let $x_{in} = \Sigma_n^{-\frac{1}{2}} x_i$, $\beta_n = \Sigma_n^{-\frac{1}{2}} \beta$. Then the model (1.1) can be written as

$$y_i = x_{in}^T \beta_n + e_i, \quad i = 1, 2, \ldots, n, \tag{2.1}$$

with $\sum_{i=1}^{n} x_{in}x_{in}^{T} = I_p$, where I_p is an identity matrix of order p. Assume that ρ has derivative ψ. For $l \geq 0$ and a function f, write $f \in C^l$ if f has derivatives up to lth order and $f^{(l)}$ is continuous. Define the function

$$\psi_k(t;\mathcal{F}_i) = E\bigl(\psi(e_k + t)|\mathcal{F}_i\bigr), \psi_k\bigl(t;\mathcal{F}_i^*\bigr) = E\bigl(\psi\bigl(e_k^* + t\bigr)|\mathcal{F}_i^*\bigr), \quad k \geq 0, \tag{2.2}$$

where $\mathcal{F}_i^* = (\ldots,\varepsilon_{-1},\varepsilon_0',\varepsilon_1,\ldots,\varepsilon_{i-1},\varepsilon_i)$, $\mathcal{F}_i = (\ldots,\varepsilon_{-1},\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{i-1},\varepsilon_i)$, let ε_i' be an i.i.d. copy of ε_i, and $e_k^* = g(\mathcal{F}_k^*)$.

Throughout the paper, we use the following assumptions.

(A1) $\rho(\cdot)$ is a convex function, $E\psi(e_i) = 0, 0 < E\psi^2(e_i)$.

(A2) $\varphi(t) \equiv E\psi(e_i + t)$ has a strictly positive derivative at $t = 0$.

(A3) $m(t) \equiv E|\psi(e_i + t) - \psi(e_i)|^2$ is continuous at $t = 0$.

(A4) $r_n \equiv \max_{1\leq i\leq n}|x_{in}| = \max_{1\leq i\leq n}(x_i^T\Sigma_n^{-1}x_i)^{\frac{1}{2}} = o(1)$.

(A5) There exists a $\delta_0 > 0$ such that

$$L_i \equiv \sup_{|s|,|t|\leq\delta_0, s\neq t} \frac{|\psi_{i+1}(s;\mathcal{F}_i) - \psi_{i+1}(t;\mathcal{F}_i)|}{|s-t|} \in L^1. \tag{2.3}$$

(A6) Let $\psi_i(\cdot;\mathcal{F}_{i-1}) \in C^l, l \geq 0$. For some $\delta_0 > 0, \max_{1\leq i\leq n}\sup_{|\delta|\leq\delta_0}\|\psi_i^{(l)}(\delta;\mathcal{F}_{i-1})\| < \infty$ and

$$\sum_{i=0}^{\infty}\sup_{|\delta|<\delta_0}\bigl\|E\bigl(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1})|\mathcal{F}_0\bigr) - E\bigl(\psi_i^{(l)}\bigl(\delta;\mathcal{F}_{i-1}^*\bigr)|\mathcal{F}_0^*\bigr)\bigr\| < \infty. \tag{2.4}$$

(A7)

$$\sum_{i=0}^{\infty}\sup_{|\delta|<\delta_0}\bigl\|E\bigl(\psi_i^{(l)}\bigl(\delta;\mathcal{F}_{i-1}^*\bigr)|\mathcal{F}_0\bigr) - E\bigl(\psi_i^{(l)}\bigl(\delta;\mathcal{F}_{i-1}^*\bigr)|\mathcal{F}_{-1}\bigr)\bigr\| < \infty, \tag{2.5}$$

$$\sum_{i=0}^{\infty}\sup_{|\delta|<\delta_0}\bigl|E\psi^{(l)}(e_i+\delta) - E\psi^{(l)}\bigl(e_i^*+\delta\bigr)\bigr| < \infty. \tag{2.6}$$

Remark 1 Conditions (A1)–(A5) and (A6) are imposed in the M-estimation considering the theory of linear regression models with dependent errors (Wu [49]; Zhou and Shao [52]). Condition (2.6) is similar to (7) of Wu [49]. $\|E(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1})|\mathcal{F}_0) - E(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1}^*)|\mathcal{F}_0^*)\|$ measures the difference of the contribution of ε_0 and its copy ε_0' in predicting $\psi(e_i+\delta)$. However, $\|E(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1}^*)|\mathcal{F}_0) - E(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1}^*)|\mathcal{F}_{-1})\|$ measures the contribution of ε_0 in predicting $\psi(e_i+\delta)$ under the given copy of ε_0: ε_0'.

If $\{\varepsilon_i\}$ are i.i.d., then (A6) and (A7) hold. For the other settings, (A6) and (A7) are very easily satisfied. The following proposition provides some sufficient conditions for (A6) and (A7).

Proposition 2.1 *Let $F_i(u|\mathcal{F}_0) = P(e_i \leq u|\mathcal{F}_0)$ and $f_i(u|\mathcal{F}_0)$ be the conditional distribution and density function of e_i at u given $\mathcal{F}_0$, respectively. Let $f_i(u)$ and $f_i^*(u)$ be the density function of e_i and e_i^*, respectively.*

(1) *Let $f_i(\cdot|\mathcal{F}_i) \in C^l, l \geq 0$, $\omega(i) = \int_R \|f_i(u|\mathcal{F}_0) - f_i(u|\mathcal{F}_0^*)\|\psi(u;\delta_0)\,du$ and $\psi(u;\delta_0) = |\psi(u+\delta_0)| + |\psi(u-\delta_0)|$. If $\sum_{i=1}^{\infty}\omega(i) < \infty$, then* (A6) *holds.*

(2) *Let*

$$\overline{\omega}(i) = \int_R \left\| f_i^{(l)}(u|\mathcal{F}_0) - f_i^{(l)}\big(u|\mathcal{F}_0^*\big) \right\| \psi(u;\delta_0)\,du$$

and $\tilde{\omega}(i) = \int_R |f_i(u) - f_i^*(u)| \psi^{(l)}(u;\delta_0)\,du$. *If* $\sum_{i=1}^{\infty} \overline{\omega}(i) < \infty$ *and* $\sum_{i=1}^{\infty} \tilde{\omega}(i) < \infty$, *then assumption* (*A*7) *holds.*

Proof (1) By the conditions of (1), we have

$$\begin{aligned}
&\sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left\| E\big(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1})|\mathcal{F}_0\big) - E\big(\psi_i^{(l)}\big(\delta;\mathcal{F}_{i-1}^*\big)|\mathcal{F}_0^*\big) \right\| \\
&\quad= \sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left\| \int_R \psi^{(l)}(u+\delta)\big[f_i(u|\mathcal{F}_0) - f_i\big(u|\mathcal{F}_0^*\big)\big]\,du \right\| \\
&\qquad+ \sup_{|\delta|\le\delta_0} \left\| \int_R \psi^{(l)}(u+\delta)\big[f_0(u|\mathcal{F}_{-1}) - f_0(u|\mathcal{F}_{-1})\big]\,du \right\| \\
&\quad\le \sum_{i=1}^{\infty} \int_R \left\| f_i^{(l)}(u|\mathcal{F}_0) - f_i^{(l)}\big(u|\mathcal{F}_0^*\big) \right\| \psi(u+\delta_0)\,du \\
&\quad= \sum_{i=1}^{\infty} \omega(i) < \infty.
\end{aligned} \tag{2.7}$$

Namely (A6) holds.

(2) (A7) follows from

$$\begin{aligned}
&\sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left\| E\big(\psi_i^{(l)}\big(\delta;\mathcal{F}_{i-1}^*\big)|\mathcal{F}_0\big) - E\big(\psi_i^{(l)}\big(\delta;\mathcal{F}_{i-1}^*\big)|\mathcal{F}_{-1}\big) \right\| \\
&\quad= \sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left\| \int_R \psi^{(l)}(u+\delta)\big[f_i^{(l)}(u|\mathcal{F}_0) - f_i^{(l)}(u|\mathcal{F}_{-1})\big]\,du \right\| \\
&\qquad+ \sup_{|\delta|\le\delta_0} \left\| \int_R \psi^{(l)}(u+\delta)\big[f_0(u|\mathcal{F}_{-1}) - f_0\big(u|\mathcal{F}_{-1}^*\big)\big]\,du \right\| \\
&\quad\le \sum_{i=1}^{\infty} \int_R \left\| f_i^{(l)}(u|\mathcal{F}_0) - f_i^{(l)}(u|\mathcal{F}_{-1}) \right\| \psi(u+\delta_0)\,du = \sum_{i=1}^{\infty} \overline{\omega}(i) < \infty
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left| E(\psi^{(l)}(e_i+\delta) - E(\psi^{(l)}\big(e_i^*+\delta\big) \right| \\
&\quad= \sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left| \int_R \psi^{(l)}(u+\delta)\big(f_i(u) - f_i^*(u)\big) \right| \\
&\quad\le \sum_{i=1}^{\infty} \int_R \left| f_i(u) - f_i^*(u) \right| \psi^{(l)}(u;\delta_0)\,du = \sum_{i=1}^{\infty} \tilde{\omega}(i) < \infty.
\end{aligned}$$

Hence, the proposition is proved. □

Define the M-processes

$$K_n(\beta_n) = \Omega_n(\beta_n) - E\big(\Omega_n(\beta_n)\big), \qquad \tilde{K}_n(\beta) = \tilde{\Omega}_n(\beta) - E\big(\tilde{\Omega}_n(\beta)\big),$$

where

$$\Omega_n(\beta_n) = \sum_{i=1}^{n} \psi\big(e_i - x_{in}^T\beta_n\big)x_{in}, \qquad \tilde{\Omega}_n(\beta) = \sum_{i=1}^{n} \psi\big(e_i - x_i^T\beta\big)x_i.$$

Theorem 2.1 *Let* $\{\delta_n, n \in N\}$ *be a sequence of positive numbers such that* $\delta_n \to \infty$ *and* $\delta_n r_n \to 0$. *If* (A1)–(A5), *and* (A6) *and* (A7) *with* $l = 0, 1, \ldots, p$ *hold, then*

$$\sup_{|\beta|\le\delta_n} \big|K_n(\beta_n) - K_n(0)\big| = O_p\left(\sqrt{\tau_n(\delta_n)}\log n + \delta_n\sqrt{\sum_{i=1}^{n}|x_{in}|^4}\right), \tag{2.8}$$

where

$$\tau_n(\delta) = \sum_{i=1}^{n} |x_{in}|^2\big(m^2\big(|x_{in}|\delta\big) + m^2\big(-|x_{in}|\delta\big)\big), \quad \delta > 0.$$

Corollary 2.1 *Assume that* (A1)–(A5), *and* (A6) *and* (A7) *with* $l = 0, 1, \ldots, p$ *hold. If* $\varphi(t) = t\varphi'(0) + O(t^2)$ *as* $t \to 0$, $\Omega(\hat{\beta}_n) = O_p(r_n)$, *then, for* $|\hat{\beta}_n| \le \delta_n$,

$$\varphi'(0)\hat{\beta}_n - \sum_{i=1}^{n}\psi(e_i)x_{in} = O_p\big(\sqrt{\tau_n(\delta_n)}\log n + \delta_n^2 r_n\big). \tag{2.9}$$

Moreover, if, as $t \to 0$, $m(t) = O(|t|^\lambda)$ *for some* $\lambda > 0$, *then*

$$\varphi'(0)\hat{\beta}_n - \sum_{i=1}^{n}\psi(e_i)x_{in} = O_p\left(\sqrt{\sum_{i=1}^{n}|x_{in}|^{2+2\lambda}}\log n + r_n\right). \tag{2.10}$$

Remark 2 If $\{e_i\}$ i.i.d., then $|\hat{\beta}_n| \le \delta_n$ follows from (3.2) of Rao and Zhao [42]. If $\{\varepsilon_i\}$ i.i.d., then $|\hat{\beta}_n| \le \delta_n$ follows from Theorem 1 of Wu [49] and Zhou and Shao [52]. If $e_i = f(e_{i-1}) + \varepsilon_i$, where the function $f: R \times R \to R$ satisfies some condition and $\{\varepsilon_i\}$ i.i.d., then $|\hat{\beta}_n| \le \delta_n$ follows from Theorem 2.2 of Fan et al. [20]. If $\{\varepsilon_i\}$ NA, then $|\hat{\beta}_n| \le \delta_n$ follows from Theorem 1 of Wu [47]. Therefore the condition $|\hat{\beta}_n| \le \delta_n$ is not strong. In the paper, we do not discuss it.

Theorem 2.2 *Assume that* (A1)–(A3), (A5), *and* (A6) *and* (A7) *with* $l = 0, 1, \ldots, p$ *hold. Let* λ_n *be the minimum eigenvalue of* Σ_n, $b_n = n^{-\frac{1}{2}}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}$, $\upsilon > 0$, $\tilde{n} = 2^{\lceil \log n/\log 2\rceil}$ *and* $q > \frac{3}{2}$. *If* $\liminf_{n\to\infty}\lambda_n/n > 0$, $\sum_{i=1}^{n}|x_i|^2 = O(n)$ *and* $\tilde{r}_n = \max_{1\le i\le n}|x_i| = O(n^{1/2}(\log n)^{-2})$, *then*

$$\sup_{|\beta|\le b_n}\big|\tilde{K}_n(\beta) - \tilde{K}_n(0)\big| = O_{a.s.}(L_{\tilde{n}} + B_{\tilde{n}}),$$

where $L_n = \sqrt{\tilde{\tau}_n(2b_n)}(\log_n)^q$, $B_{\tilde{n}} = b_n(\sum_{i=1}^n |x_i|^4)^{1/2}(\log n)^{3/2}(\log\log n)^{(1+\upsilon)/2}$ *and*

$$\tilde{\tau}_n(\delta) = \sum_{i=1}^n |x_i|^2\big(m^2\big(|x_i|\delta\big) + m^2\big(-|x_i|\delta\big)\big), \quad \delta > 0.$$

Corollary 2.2 *Assume that* $\varphi(t) = t\varphi'(0) + O(t^2)$ *and* $m(t) = O(\sqrt{t})$ *as* $t \to 0$, *and* $\tilde{\Omega}_n = O_{a.s.}(\tilde{r}_n)$. *Under the conditions of Theorem* 2.2, *we have*:
(1) $\tilde{\beta}_n = O_{a.s.}(b_n)$;
(2) $\varphi'(0)\Sigma_n\tilde{\beta}_n - \sum_{i=1}^n \psi(e_i)x_i = O_{a.s.}(L_{\tilde{n}} + B_{\tilde{n}} + b_n^2\sum_{i=1}^n |x_i|^3 + \tilde{r}_n)$,
where $\tilde{\beta}_n$ *is the minimizer of* (1.2).

Remark 3 From the above results, we easily obtain the corresponding conclusions of Wu [49].

From the corollary below, we only derive convergence rates of $\tilde{\beta}_n$. However, it is to be regretted that we cannot give laws of the iterated logarithm $n^{1/2}(\log\log n)^{1/2}$, which is still an open problem.

Corollary 2.3 *Under the conditions of Corollary* 2.2, *we have*

$$\begin{aligned}\Sigma_n\tilde{\beta}_n = O_{a.s.}\Bigg(\max\Bigg\{&\big(n^{1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}\big),\\ &\big(n^{1/2}(\log n)^{-1/4+q}(\log\log n)^{1/4+\upsilon/2}\big), \Bigg(\sum_{i=1}^n \psi(e_i)x_i\Bigg)\Bigg\}\Bigg).\end{aligned}$$

Proof Note that $\tilde{n} = 2^{\lceil \log n/\log 2\rceil} = O(n)$ and $m(t) = O(\sqrt{t})$ as $t \to 0$; we have

$$\begin{aligned} L_{\tilde{n}} &= \sqrt{\tilde{\tau}_n(2b_n)}(\log n)^q = \sqrt{O\Bigg(\sum_{i=1}^n |x_i|^2|x_i|b_n\Bigg)(\log n)^q} \\ &= \sqrt{O\big(nn^{1/2}(\log n)^{-2}n^{-1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}\big)(\log n)^q} \\ &= O\big(n^{1/2}(\log n)^{-1/4+q}(\log\log n)^{1/4+\upsilon/2}\big), \\ B_{\tilde{n}} &= O\big(n^{-1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}\big(n\tilde{r}_n^2\big)^{1/2}(\log n)^{3/2}(\log\log n)^{(1+\upsilon)/2}\big) \\ &= O\big(n^{1/2}(\log n)(\log\log n)^{1+3\upsilon/2}\big) \end{aligned}$$

and

$$\begin{aligned} b_n^2\sum_{i=1}^n |x_i|^3 &= O\big(n^{-1}(\log n)^3(\log\log n)^{1+2\upsilon}nn^{1/2}(\log n)^{-2}\big) \\ &= O\big(n^{1/2}(\log n)(\log\log n)^{1+2\upsilon}\big). \end{aligned}$$

By Corollary 2.2, we have

$$\varphi'(0)\Sigma_n\tilde{\beta}_n = \sum_{i=1}^n \psi(e_i)x_i + O_{a.s.}\big(n^{1/2}(\log n)^{-1/4+q}(\log\log n)^{1/4+\upsilon/2}\big) \tag{2.11}$$

and

$$\Sigma_n\tilde{\beta}_n = O_{a.s.}(nb_n) = O_{a.s.}\big(n^{1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}\big). \tag{2.12}$$

Thus the conclusion follows from (2.11) and (2.12). □

3 Applications

In the following three subsections, we shall investigate some applications of our results. In Sect. 3.1, we consider that ε_i is a m-dependent random variable sequence. We shall investigate that $\{\varepsilon_i\}$ are (ε,ψ)-weakly dependent in Sect. 3.2, and martingale difference errors $\{\varepsilon_i\}$ in Sect. 3.3.

3.1 *m*-dependent process

In the subsection, we shall firstly show that the m-dependent sequence satisfies conditions (A6) and (A7) and secondly obtain the asymptotic normal distribution and strong convergence rates for M-estimators of the parameter. Koul [30] discussed the asymptotic behavior of a class of M-estimators in the model (1.1) with long range dependence errors $e_i = g(\varepsilon_i)$, where ε_i i.i.d. Here we assume that ε_i is a m-dependent sequence, of which the definition was given by Example 2.8.1 in Lehmann [32]. For m-dependent sequences or processes, there are some results (e.g., see Hu et al. [27], Romano and Wolf [43] and Valk [46]).

Proposition 3.1 *Let ε_i in* (1.4) *be a m-dependent sequence. Then* (A6) *and* (A7) *hold.*

Proof Note that ε_i is a m-dependent sequence, we have

$$\begin{aligned}
&\sum_{i=1}^{\infty}\sup_{|\delta|\le\delta_0}\big\|E\big(\psi_i^{(l)}(\delta;\mathcal{F}_{i-1})|\mathcal{F}_0\big)-E\big(\psi_i^{(l)}\big(\delta;\mathcal{F}_{i-1}^*\big)|\mathcal{F}_0^*\big)\big\|\\
&\quad=\sum_{i=1}^{\infty}\sup_{|\delta|\le\delta_0}\big\|E\big(\psi^{(l)}(e_i+\delta)|\mathcal{F}_0\big)-E\big(\psi^{(l)}\big(e_i^*+\delta\big)|\mathcal{F}_0^*\big)\big\|\\
&\qquad+\sup_{|\delta|\le\delta_0}\big\|E\big(\psi^{(l)}(e_i+\delta)|\mathcal{F}_{-1}\big)-E\big(\psi^{(l)}\big(e_i^*+\delta\big)|\mathcal{F}_{-1}\big)\big\|\\
&\quad=\sum_{i=1}^{\infty}\sup_{|\delta|\le\delta_0}\big\|E\big(\psi^{(l)}(e_i+\delta)\big)-E\big(\psi^{(l)}\big(e_i^*+\delta\big)\big)\big\|\\
&\quad=0<\infty
\end{aligned}\tag{3.1}$$

and

$$\begin{aligned}
&\sum_{i=1}^{\infty}\sup_{|\delta|\le\delta_0}\big\|E\big(\psi_i^{(l)}\big(\delta;\mathcal{F}_{i-1}^*\big)|\mathcal{F}_0\big)-E\big(\psi_i^{(l)}\big(\delta;\mathcal{F}_{i-1}^*\big)|\mathcal{F}_{-1}\big)\big\|\\
&\quad=\sum_{i=1}^{\infty}\sup_{|\delta|\le\delta_0}\big\|E\big(\psi^{(l)}\big(e_i^*+\delta\big)|\mathcal{F}_0\big)-E\big(\psi^{(l)}\big(e_i^*+\delta\big)|\mathcal{F}_{-1}\big)\big\|\\
&\qquad+\sup_{|\delta|\le\delta_0}\big\|E\big(\psi^{(l)}(e_0+\delta)|\mathcal{F}_{-1}\big)-E\big(\psi^{(l)}\big(e_0^*+\delta\big)|\mathcal{F}_{-1}\big)\big\|
\end{aligned}$$

$$= \sum_{i=1}^{\infty} \sup_{|\delta|\le\delta_0} \left\| E\psi^{(l)}(e_0+\delta) - E\psi^{(l)}\left(e_i^*+\delta\right) \right\|$$
$$= 0 < \infty. \tag{3.2}$$

Therefore, (A6) and (A7) follow from (3.1), (3.2) and $E\psi^{(l)}(e_i+\delta) = E\psi^{(l)}(e_i^*+\delta)$. □

Corollary 3.1 *Assume that* (A1)–(A5) *hold. If* $\varphi(t) = t\varphi'(0) + O(t^2)$ *and* $m(t) = O(|t|^\lambda)$ *for some* $\lambda > 0$ *as* $t \to 0$, $\Omega(\hat{\beta}_n) = 0$ *and* $0 < \sigma_\psi^2 = E[\psi(e_i)]^2 < \infty$, *then*

$$n^{-1/2}\hat{\beta}_n / \left(\left[\varphi'(0)\right]^{-1}\sigma_\psi\right) \to N(0, I_p), \quad n \to \infty.$$

In order to prove Corollary 3.1, *we give the following lemmas.*

Lemma 3.1 (Lehmann [32]) *Let* $\{\xi_i, i \ge 1\}$ *be a stationary m-dependent sequence of random variables with* $E\xi_i = 0$ *and* $0 < \sigma^2 = \mathrm{Var}(\xi_i) < \infty$, *and* $T_n = \sum_{i=1}^n \xi_i$. *Then*

$$n^{-1/2}T_n/\tau \to N(0,1),$$

where $\tau^2 = \lim_{n\to\infty} \mathrm{Var}(n^{-1/2}T_n) = \sigma^2 + 2\sum_{i=2}^{m+1} \mathrm{Cov}(\xi_1, \xi_i)$.

Using the argument of Lemma 3.1, we easily obtain the following result. Here we omit the proof.

Lemma 3.2 *Let* $\{\xi_i, i \ge 1\}$ *be a stationary m-dependent sequence of random variables with* $E\xi_i = 0$ *and* $0 < \sigma_i^2 = \mathrm{Var}(\xi_i) < \infty$, *and* $T_n = \sum_{i=1}^n \xi_i$. *Then*

$$n^{-1/2}T_n/\tau \to N(0,1),$$

where $\tau^2 = \mathrm{Var}(n^{-1/2}T_n) = n^{-1}\sum_{i=1}^n \sigma_i^2 + 2n^{-1}\sum_{i=2}^{m+1}(n-i)\,\mathrm{Cov}(\xi_1, \xi_i)$.

Proof of Corollary 3.1 By (2.10), we have

$$n^{-1/2}\hat{\beta}_n = n^{-1/2}\left[\varphi'(0)\right]^{-1}\sum_{i=1}^n \psi(e_i)x_{in} + O_p\left(n^{-1/2}r_n^\lambda \log n\right). \tag{3.3}$$

Since $\{\xi_i, i \ge 1\}$ is a stationary m-dependent sequence, so is $\{[\varphi'(0)]^{-1}\psi(e_i)x_{in}, i \ge 1\}$. Let $u \in R^p$, $|u| = 1$. Then $E(u^T[\varphi'(0)]^{-1}\psi(e_i)x_{in}) = 0$ and

$$\sigma_i^2 = E\left(u^T\left[\varphi'(0)\right]^{-1}\psi(e_i)x_{in}\right)^2 = \left[\varphi'(0)\right]^{-2}u^T x_{in}x_{in}^T u E\left[\psi(e_i)\right]^2.$$

Therefore, by $r_n = o(1)$ and $0 < \sigma_\psi^2 = E[\psi(e_i)]^2 < \infty$, we have

$$\tau^2 = n^{-1}\sum_{i=1}^n \left[\varphi'(0)\right]^{-2}u^T x_{in}x_{in}^T u E\left[\psi(e_i)\right]^2$$
$$+ 2n^{-1}\sum_{i=2}^{m+1}(n-i)\,\mathrm{Cov}\left(u^T\left[\varphi'(0)\right]^{-1}\psi(e_1)x_{1n}, u^T\left[\varphi'(0)\right]^{-1}\psi(e_i)x_{in}\right)$$

$$= \left[\varphi'(0)\right]^{-2} n^{-1} \left\{ \sum_{i=1}^{n} E\left[\psi(e_i)\right]^2 + 2\sum_{i=2}^{m+1}(n-i)u^T x_{in} x_{in}^T u \operatorname{Cov}\left(\psi(e_1), \psi(e_i)\right) \right\}$$
$$\to \left[\varphi'(0)\right]^{-2} \sigma_\psi^2. \tag{3.4}$$

Thus the corollary follows from Lemma 3.2, (3.3) and (3.4). □

Corollary 3.2 *Assume that* (A1)–(A5) *hold. If* $\varphi(t) = t\varphi'(0) + O(t^2)$ *and* $m(t) = O(\sqrt{t})$ *as* $t \to 0$, *and* $\tilde{\Omega}_n(\tilde{\beta}_n) = O_{a.s.}(\tilde{r}_n), 0 < \sigma_\psi^2 = E[\psi(e_i)]^2 < \infty$, *then*

$$\tilde{\beta}_n = O_{a.s.}\left(n^{-1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}\right).$$

Proof The corollary follows from Proposition 3.1 and Corollary 2.2. □

3.2 (ε, ψ)-weakly dependent process

In the subsection, we assume that $\{\varepsilon_i\}$ are (ε, ψ)-weakly dependent (Doukhan and Louhichi [14] and Dedecker et al. [11]) random variables. In 1999, Doukhan and Louhichi proposed a new idea of (ε, ψ)-weakly dependence which focuses on covariance rather than the total variation distance between joint distributions and the product of the corresponding marginal. It has been shown that this concept is more general than mixing and includes, under natural conditions on the process parameters, essentially all classes of processes of interest in statistics. Therefore, many researchers are interested in the (ε, ψ)-weakly dependent and related possesses, and one obtained lots of sharp results. For example, Doukhan and Louhichi [14], Dedecker and Doukhan [10], Dedecker and Prieur [12], Doukhan and Neumann [16], Doukhan and Wintenberger [17], Bardet et al. [3], Doukhan and Wintenberger [18], Doukhan et al. [13]. However, a few people (only Hwang and Shin [29], Nze et al. [40]) investigated regression models with (ε, ψ)-weakly dependent errors. Nobody has investigated a robust estimate for the regression model with (ε, ψ)-weakly dependent errors. To give the definition of the (ε, ψ)-weakly dependence, let us consider a process $\xi = \{\xi_n, n \in Z\}$ with values in a Banach space $(\mathcal{E}, \|\cdot\|)$. For $h : \mathcal{E}^u \to R$, $u \in N$, we define the Lipschitz modulus of h,

$$\operatorname{Lip} h = \sup_{y \neq x} \left|h(y) - h(x)\right| / \|y - x\|_1, \tag{3.5}$$

where we have the l_1-norm, i.e., $\|(y_1, y_2, \ldots, y_u)\|_1 = \sum_{i=1}^{u} |y_i|$.

Definition 1 (Doukhan and Louhich [14]) A process $\xi = \{\xi_n, n \in Z\}$ with values in R^d is called a (ε, ψ)-weakly dependent process if, for some classes of functions $\mathcal{E}^u, \mathcal{E}^v \to R, F_u, G_v$:

$$\varepsilon(r) = \sup_{u,v} \sup_{s_1 \geq s_2 \geq \cdots \geq s_u, t_1 \geq t_2 \geq \cdots \geq t_v, r = t_1 - s_u} \sup_{f \in F_u, g \in G_v} \frac{|\operatorname{Cov}(f(\xi_{s_1}, \xi_{s_2}, \ldots, \xi_{s_u}), g(\xi_{t_1}, \xi_{t_2}, \ldots, \xi_{t_v}))|}{\Psi(f, g)}$$
$$\to 0$$

as $r \to \infty$.

According to the definition, mixing sequences ($\alpha, \rho, \beta, \varphi$-mixing), associated sequences (positively or negatively associated), Gaussian sequences, Bernoulli shifts and Markovian models or time series bootstrap processes with discrete innovations are (ε, ψ)-weakly dependent (Doukhan et al. [15]).

From now on, assume that the classes of functions contain functions bounded by 1. Distinct functions Ψ yield η, θ, κ and a λ weak dependence of the coefficients as follows (Doukhan et al. [15]):

$$\Psi(f,g) = \begin{cases} u\mathrm{Lip}f + v\mathrm{Lip}g & \text{then denote } \varepsilon(r) = \eta(r), \\ v\mathrm{Lip}g & \text{then denote } \varepsilon(r) = \theta(r), \\ uv\mathrm{Lip}f \cdot \mathrm{Lip}g & \text{then denote } \varepsilon(r) = \kappa(r), \\ u\mathrm{Lip}f + v\mathrm{Lip}g + uv\mathrm{Lip}f \cdot \mathrm{Lip}g & \text{then denote } \varepsilon(r) = \lambda(r), \\ u\mathrm{Lip}f + v\mathrm{Lip}g + uv\mathrm{Lip}f \cdot \mathrm{Lip}g + u + v & \text{then denote } \varepsilon(r) = \omega(r). \end{cases} \tag{3.6}$$

In Corollary 3.3, we only consider λ and η-weakly dependence. Let $\{\varepsilon_i\}$ be λ or η-weakly dependent, and assume that g satisfies: for each $s \in Z$, if $x, y \in R^Z$ satisfy $x_i = y_i$ for each index $i \neq s$

$$\left|g(x) - g(y)\right| \leq b_s \Bigl(\sup_{i \neq s} |x_i|^l \vee 1\Bigr) |x_s - y_s|. \tag{3.7}$$

Lemma 3.3 (Dedecker et al. [11]) *Assume that g satisfies the condition* (3.7) *with $l \geq 0$ and some sequence $b_s \geq 0$ such that $\sum_s |s| b_s < \infty$. Assume that $E|\varepsilon_0|^{m'} < \infty$ with $lm < m'$ for some $m > 2$. Then*:

(1) *If the process $\{\varepsilon_i, i \in Z\}$ is λ-weakly dependent with coefficients $\lambda_\varepsilon(r)$, then e_n is λ-weakly dependent with coefficients*

$$\lambda_e(k) = c \inf_{r \leq [k/2]} \biggl(\sum_{i \geq r} b_i\biggr) \vee \bigl[(2r+1)^2 \lambda_\varepsilon (k-2r)^{\frac{m'-1-l}{m'-1+l}}\bigr]. \tag{3.8}$$

(2) *If the process $\{\varepsilon_i, i \in Z\}$ is η-weakly dependent with coefficients $\eta_\varepsilon(r)$, then e_n is η-weakly dependent and there exists a constant $c > 0$ such that*

$$\eta_e(k) = c \inf_{r \leq [k/2]} \biggl(\sum_{i \geq r} b_i\biggr) \vee \bigl[(2r+1)^{1+\frac{1}{m'-1}} \eta_\varepsilon (k-2r)^{\frac{m'-2}{m'-1}}\bigr].$$

Lemma 3.4 (Bardet et al. [3]) *Let $\{\xi_n, n \in Z\}$ be a sequence of R^k-valued random variables. Assume that there exists some constant $C > 0$ such that $\max_{1 \leq i \leq k} \|\xi_i\|_p \leq C, p \geq 1$. Let h be a function from R^k to R such that $h(0) = 0$ and for $x, y \in R^k$, there exist a in $[1, p]$ and $c > 0$ such that*

$$\left|h(x) - h(y)\right| \leq c|x - y| \bigl(1 + |x|^{\alpha-1} + |y|^{\alpha-1}\bigr). \tag{3.9}$$

Now we define the sequence $\{\zeta_n, n \in Z\}$ by $\zeta_n = h(\xi_n)$. Then:

(1) *If the process $\{\xi_i, i \in Z\}$ is λ-weakly dependent with coefficients $\lambda_\xi(r)$, then $\{\zeta_n, n \in Z\}$ is also with coefficients*

$$\lambda_\zeta(r) = O\bigl(\lambda_\xi^{\frac{p-a}{p+a-2}}(r)\bigr). \tag{3.10}$$

(2) *If the process* $\{\xi_i, i \in Z\}$ *is* ζ*-weakly dependent with coefficients* $\eta_\xi(r)$*, so is* $\{\zeta_n, n \in Z\}$ *with coefficients* $\eta_\zeta(r) = O(\eta_\xi^{\frac{p-a}{p-1}}(r))$.

Lemma 3.5 (Dedecker et al. [11]) *Let* $\{\xi_i, i \in Z\}$ *be a centered and stationary real-valued sequence with* $E|\xi_0|^{2+\varsigma} < \infty$, $\varsigma > 0$, $\sigma^2 = \sum_{k\in Z} \mathrm{Cov}(\xi_0, \xi_k)$ *and* $S_n = \sum_{i=1}^n \xi_i$. *If* $\lambda_\xi(r) = O(r^{-\lambda})$ *for* $\lambda > 4 + 2/\varsigma$, *then* $n^{-1/2}S_n \to N(0, \sigma^2)$ *as* $n \to \infty$.

Corollary 3.3 *Let* $\{\varepsilon_i\}$ *be* λ*-weakly dependent with coefficients* $\lambda_\varepsilon(r) = O(\exp(-r\lambda))$ *for some* $\lambda > 0$, *and* $b_i = O(\exp(-ib))$ *for some* $b > 0$. *Assume that* $\psi(0) = 0$, *and, for* $x, y \in R$, *there exists a constant* $c > 0$ *such that*

$$\left|\psi(x) - \psi(y)\right| \le c|x - y|. \tag{3.11}$$

Under the conditions of Corollary 2.1, *we have*

$$\varphi'(0)n^{-1/2}T_n \to N(0, \Sigma) \quad \text{as } n \to \infty, \tag{3.12}$$

where $\Sigma = \sum_{i=1}^n x_{1n} \mathrm{Cov}(\psi(e_1), \psi(e_i))x_{in}^T$.

Proof Note that $\{\varepsilon_i\}$ is λ-weakly dependent. By Lemma 3.3, we find that $\{e_i\}$ is λ-weakly dependent with coefficients

$$\lambda_e(r) = O\left(r^2 \exp\left(-\lambda r \frac{b(m' - 1 - l)}{b(m' - 1 + l + 2\alpha(m' - 1 - l))}\right)\right), \quad \alpha > 0 \tag{3.13}$$

from (3.8) and Proposition 3.1 in Chap. 3 (Dedecker et al. [11]).

Let $u \in R^p$, $|u| = 1$, and $\zeta_i = h(e_i) = u\psi(0)x_{in} = 0$. Then $h(0) = 0 = u\psi(0)x_{in} = 0$. Choose $p = 2, a = 1$, in (3.9), and by (3.11), we have

$$\left|h(x) - h(y)\right| = |x_{in}|\left|\psi(x) - \psi(y)\right| \le c|x - y| \tag{3.14}$$

for $x, y \in R$ and $c > 0$. Therefore, by Lemma 3.4, $\{\zeta_i, i \in N\}$ is λ-weakly dependent with coefficients

$$\lambda_\zeta(r) = O\left(r_n^{uv}\lambda_e^{\frac{p-a}{p+a-2}}(r)\right) = O\left(r_n^{uv}\lambda_e(r)\right). \tag{3.15}$$

By Corollary 2.1, we have

$$\varphi'(0)n^{-1/2}\hat{\beta}_n = n^{-1/2}\sum_{i=1}^n \psi(e_i)x_{in} + o_p(1). \tag{3.16}$$

By (3.13) and (3.15), there exist $b > 0, a > 0, l \ge 0$ and $m' > lm$ for some $m > 2$ such that

$$\lambda_\zeta(r) = O\left(r_n^{uv}r^2 \exp\left(-\lambda r \frac{b(m' - 1 - l)}{b(m' - 1 + l) + 2\alpha(m' - 1 - l)}\right)\right) = O\left(r^{-\lambda}\right) \tag{3.17}$$

for enough large r and $\lambda > 4 + 2/\varsigma$ with $\varsigma > 0$.

By Lemma 3.5 and (3.16)–(3.17), we have

$$\varphi'(0)n^{-1/2}uT_n \to N\bigl(0,\sigma^2\bigr),$$

where $\sigma^2 = \sum_{i=1}^n u^T x_{1n}\operatorname{Cov}(\psi(e_1),\psi(e_i))x_{in}^T u$. Using the Cramer device, we complete the proof of Corollary 3.3. □

Lemma 3.6 (Dedecker et al. [11]) *Suppose that* $\{\xi_i, 1 \le i \le n\}$ *are stationary real-valued random variables with* $E\xi_i = 0$ *and* $P(|\xi_i| \le M < \infty) = 1$ *for all* $i = 1,2,\ldots,n$. *Let* $\Psi : N^2 \to N$ *be one of the following functions*:

$$\begin{aligned} &\Psi(u,v) = 2v, \qquad \Psi(u,v) = u+v, \qquad \Psi(u,v) = uv, \\ &\Psi(u,v) = \alpha(u+v) + (1-\alpha)uv, \end{aligned} \tag{3.18}$$

for some $0 < \alpha < 1$. *We assume that there exist constants* $K, L_1, L_2 < \infty$, $\mu \ge 0$ *and a nonincreasing sequence of real coefficients* $\{\rho(n), n \ge 0\}$ *such that, for all u-tuples* $(s_1,\ldots,s_u)$ *and all v-tuples* $(t_1,\ldots,t_v)$ *with* $1 \le s_1 \le \cdots \le s_u \le t_1 \le \cdots \le t_v \le n$, *the following inequality is fulfilled*:

$$\bigl|\operatorname{Cov}(\xi_{s_1},\ldots,\xi_{s_u};\xi_{t_1},\ldots,\xi_{t_v})\bigr| \le K^2 M^{u+v-2}\Psi(u,v)\rho(t_1-s_u), \tag{3.19}$$

where

$$\sum_{s=0}^{\infty}(s+1)^k\rho(s) \le L_1 L_2^k (k!)^\mu, \quad \forall k \ge 0. \tag{3.20}$$

Let $S_n = \sum_{i=1}^n \xi_i$ *and* $\sigma_n^2 = \operatorname{Var}(\sum_{i=1}^n \xi_i)$. *If* $\sigma^2 = \lim_{n\to\infty}\sigma_n^2/n > 0$, *then*

$$\limsup_{n\to\infty}\frac{|S_n|}{\sigma(2n\log\log n)^{1/2}} \le 1. \tag{3.21}$$

Corollary 3.4 *Let* $\{\varepsilon_i\}$ *be* η*-weakly dependent with coefficients* $\eta_\varepsilon(r) = O(\exp(-r\eta))$ *for some* $\eta > 0$, *and* $b_i = O(\exp(-ib))$ *for some* $b > 0$. *Assume that* $\psi(0) = 0$ *and* (3.11) *hold. Under the conditions of Corollary* 2.2 *with* $\tilde{r}_n = O(n^{1/2}(\log n)^{-2})$ *replaced by* $0 < \min_{1\le i\le n}|x_{ij}| < \max_{1\le i\le n}|x_{ij}| < \infty$, *and* $0 < \sigma_\psi^2 = E\psi^2(e_i) < \infty$, *we have*:

(1) *for* $3/2 < q \le 7/4$, $\Sigma_n\tilde{\beta}_n = O_{a.s.}(nb_n) = O_{a.s.}(n^{1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon})$;

(2) *for* $q \ge 7/4$, $\Sigma_n\tilde{\beta}_n = O_{a.s.}(n^{1/2}(\log n)^{-1/4+q}(\log\log n)^{1/4+2/\upsilon})$.

Proof Let $\xi_i = \psi(e_i)x_{ij}, j = 1,\ldots,p$. Then for $\forall\mu_n \to \infty$ as $n \to \infty$

$$P\bigl\{\bigl|\psi(e_i)x_{ij}\bigr| > \mu_n\bigr\} \le \frac{E|\psi(e_i)x_{ij}|^2}{\mu_n^2} = \frac{\sigma_\psi^2\max_{1\le i\le n}|x_{ij}^2|}{\mu_n^2} \to 0. \tag{3.22}$$

Therefore, there exists some $0 < M < \infty$ such that

$$P\bigl\{\bigl|\psi(e_i)x_{ij}\bigr| \le M\bigr\} = 1. \tag{3.23}$$

Similar to the proofs of (3.13) and (3.15), we easily obtain

$$\eta_\zeta(r) = O\big(\tilde{r}_n^{uv}\eta_e^{\frac{p-a}{p+a-2}}(r)\big) = O\big(\tilde{r}_n^{uv}\eta_e(r)\big), \tag{3.24}$$

where

$$\eta_e(r) = O\left(r^{\frac{m'-1-l}{m'-1}}\exp\left(-\eta r\frac{b(m'-2)}{b(m'-1)+2\eta(m'-2)}\right)\right). \tag{3.25}$$

By (3.24) and (3.25), we have

$$\begin{aligned}
&\big|\mathrm{Cov}(\xi_{s_1},\ldots,\xi_{s_u};\xi_{t_1},\ldots,\xi_{t_v})\big| \\
&\quad\le (u+v)\eta_\zeta(r) \le (u+v)\tilde{r}_n^{uv}\eta_e(r) \\
&\quad\le (u+v)\tilde{r}_n^{uv}r^{\frac{m'-1-l}{m'-1}}\exp\left(-\eta r\frac{b(m'-2)}{b(m'-1)+2\eta(m'-2)}\right).
\end{aligned} \tag{3.26}$$

Let $\Psi(u,v) = u+v, K^2 = r_n^{uv}M^{1(u+v-2)}$ and

$$\rho(s) = r^{\frac{m'-1-l}{m'-1}}\exp\left(-\eta r\frac{b(m'-2)}{b(m'-1)+2\eta(m'-2)}\right). \tag{3.27}$$

Thus (3.19) holds. Since $\lim_{s\to\infty}\ln(s+1)/s = 0$, there exist $b>0,\eta>0,l\ge 0$ and $m'>lm$ for some $m>2$ and $m'>2$ such that

$$\exp\left(-\eta s\frac{b(m'-2)}{b(m'-1)+2\eta(m'2)}\right) \le (s+1)^{-(2+k)}, \quad \forall k\ge 0. \tag{3.28}$$

Thus

$$\begin{aligned}
\sum_{s=0}^{\infty}(s+1)^k\rho(s) &\le \sum_{s=0}^{\infty}(s+1)^{k+\frac{m'-1-l}{m'-1}}\exp\left(-\eta s\frac{b(m'-2)}{b(m'-1)+2\eta(m'-2)}\right) \\
&\le \sum_{s=0}^{\infty}(s+1)^{-2+\frac{m'-1-l}{m'-1}} < \infty,
\end{aligned} \tag{3.29}$$

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty}\left\{n^{-1}\sum_{i=1}^{n}E\psi^2(e_i)x_{ij}^2 + n^{-1}\sum_{i,k=1;i\ne k}^{n}x_{ij}x_{kj}\mathrm{Cov}\big(\psi(e_i),\psi(e_k)\big)\right\} \\
&= \lim_{n\to\infty}\left\{n^{-1}\sum_{i=1}^{n}E\psi^2(e_i)x_{ij}^2 + \frac{n-1}{2}O\big(x_{ij}^2\big)\sum_{i=1}^{n-1}(n-i)\mathrm{Cov}\big(\psi(e_1),\psi(e_{i+1})\big)\right\} \\
&= \sigma_\psi^2\bar{x}_{\cdot j} > 0.
\end{aligned} \tag{3.30}$$

By Lemma 3.6 and Corollary 2.3, we have

$$\begin{aligned}
\Sigma_n\tilde{\beta}_n &= O_{a.s.}\big((2n\log\log n)^{1/2}\big) + O_{a.s.}\big(n^{1/2}(\log n)^{-1/4+q}(\log\log n)^{1/4+\upsilon/2}\big) \\
&= O_{a.s.}\big(n^{1/2}(\log n)^{-1/4+q}(\log\log n)^{1/4+\upsilon/2}\big).
\end{aligned} \tag{3.31}$$

Therefore, by Corollary 2.3, (3.23) and (3.31), we complete the proof of Corollary 3.4. □

3.3 Linear martingale difference processes

In the subsection, we will investigate martingale difference errors $\{\varepsilon_i\}$. We shall provide some sufficient conditions for (A6) and (A7) and give the central limit theorem and strong convergence rates.

Let $\{\varepsilon_i\}$ be a martingale difference sequence, and a_j be real numbers such that $e_i = \sum_{j=0}^{\infty} a_j\varepsilon_{i-j}$ exists. It is well known that the theory of martingales provides a natural unified method for dealing with limit theorems. Under its influence, there is great interest in the martingale difference. Liang and Jing [34] were concerned with the partial linear model under the linear com of martingale differences and obtained asymptotic normality of the least squares estimator of the parameter. Nelson [39] has given conditions for the pointwise consistency of weighted least squares estimators from multivariate regression models with martingale difference errors. Lai [31] investigated stochastic regression models with martingale difference sequence errors and obtained strong consistency and asymptotic normality of the least squares estimate of the parameter.

Let F_ε be the distribution function of ε_0 and let f_ε be its density.

Proposition 3.2 *Suppose that* $E\varepsilon_0 = 0, \varepsilon_0 \in L^{4/(2-\gamma)}$, $\kappa_\gamma = \int_R \psi^2(u)\omega_{-\gamma}(du) < \infty, 1 < \gamma < 2$ *and* $\sum_{k=0}^{p}\int_R |f_\varepsilon^{(k)}(v)|^2\omega_\gamma(dv) < \infty$, *where* $\omega_\gamma(dv) = (1+|v|)^\gamma$. *If* $\sum_{j=0}^{\infty}|a_j| < \infty$, *then* $\sum_{i=0}^{\infty}\omega(i) < \infty, \sum_{i=0}^{\infty}\bar{\omega}(i) < \infty$ *and* $\sum_{i=0}^{\infty}\tilde{\omega}(i) < \infty$.

Proof Let $Z_n = \sum_{j=0}^{\infty} a_j\varepsilon_{n-j}, Z_n^* = Z_n - a_n\varepsilon_0 - a_n\varepsilon'$, and

$$R_n = \int_R \bigl[f_\varepsilon(t-U_n) - f_\varepsilon(t-U_n-a_0\varepsilon_n)\bigr]^2\omega_\gamma(dt), \tag{3.32}$$

where $U_n = Z_n - a_n\varepsilon_0$. By the Schwartz inequality, we have

$$\begin{aligned}\omega^2(n) &= \biggl(\int_R \bigl\|f_\varepsilon(t-Z_n) - f_\varepsilon\bigl(t-Z_n^*\bigr)\bigr\|\bigl(1+|t|\bigr)^\gamma \cdot \psi(t;\varepsilon_0)\bigl(1+|t|\bigr)^{-\gamma}\,dt\biggr)^2 \\ &\le \int_R \psi^2(t;\varepsilon_0)\omega_{-\gamma}(dt)\cdot\int_R \bigl\|f_\varepsilon(t-Z_n) - f_\varepsilon\bigl(t-Z_n^*\bigr)\bigr\|^2\omega_\gamma(dt) \\ &= \kappa_\gamma\int_R \bigl\|f_\varepsilon(t-Z_n) - f_\varepsilon\bigl(t-Z_n^*\bigr)\bigr\|^2\omega_\gamma(dt) \le CE(R_n).\end{aligned} \tag{3.33}$$

Note that

$$f_\varepsilon(t-U_n) - f_\varepsilon(t-U_n-a_n\varepsilon_0) = \int_0^{a_n\varepsilon_0} f_\varepsilon'(t-U_n-v)\,dv \tag{3.34}$$

and

$$\begin{aligned}\int_R \bigl[f_\varepsilon'(t-u)\bigr]^2\omega_\gamma(dt) &= \bigl(1+|u|\bigr)^\gamma\int_R \bigl[f_\varepsilon'(v)\bigr]^2\bigl(1+|u|\bigr)^{-\gamma}\bigl(1+|u+v|\bigr)^\gamma\,dv \\ &\le \bigl(1+|u|\bigr)^\gamma\int_R \bigl[f_\varepsilon'(v)\bigr]^2\omega_\gamma(dv) \le I_1\bigl(1+|u|\bigr)^\gamma.\end{aligned} \tag{3.35}$$

Let $I_k = \int_R [f_\varepsilon^{(k)}(v)]^2\omega_\gamma(dv)$. By the Schwartz inequality, we have

$$R_n \le \int_R \biggl|\int_0^{a_n\varepsilon_0} 1^2\,dv\cdot\int_0^{a_n\varepsilon_0}\bigl[f_\varepsilon'(t-U_n-v)\bigr]^2\,dv\biggr|\omega_\gamma(dt)$$

$$
\begin{aligned}
&\le |a_n\varepsilon_0| \int_0^{a_n\varepsilon_0} I_1\big(1 + |U_n + v|\big)^\gamma \, dv \\
&\le |a_n\varepsilon_0|^2\big[\big(1 + |U_n|\big)^\gamma + \big(1 + |U_n + a_n\varepsilon_0|\big)^\gamma\big] \\
&\le C|a_n\varepsilon_0|^2\big[\big(1 + |U_n|\big)^\gamma + |a_n\varepsilon_0|^\gamma\big]. \qquad (3.36)
\end{aligned}
$$

By $\sup_j E\varepsilon_j^2 < \infty$ and Chatterji's inequality (Lin and Bai [35]), we have

$$
EU_n^2 \le \sum_{j\neq n, j=1}^{\infty} a_j^2 E\varepsilon_{n-j}^2 \le \sum_{j=0}^{\infty} a_j^2. \qquad (3.37)
$$

By (3.33)–(3.37) and the Schwartz inequality, we have

$$
\begin{aligned}
E(R_n) &\le CE\big\{|a_n\varepsilon_0|^2 + |a_n\varepsilon_0|^{2+\gamma} + |a_n\varepsilon_0|^2|U_n|^\gamma\big\} \\
&\le C\big\{|1 + |a_n|^\gamma + E\big[|\varepsilon_0|^2|U_n|^\gamma\big]\big\} \\
&\le Ca_n^2\big\{|1 + |a_n|^\gamma + \big(E|U_n|^2\big)^{\gamma/2}\big\} \\
&\le Ca_n^2\left\{|1 + |a_n|^\gamma + \left(\sum_{j=0}^{\infty} a_j^2\right)^{\gamma/2}\right\}. \qquad (3.38)
\end{aligned}
$$

Note that $\sum_{j=0}^{\infty} |a_j| < \infty$ implies $\sum_{j=0}^{\infty} a_j^2 < \infty$ and $\sum_{j=0}^{\infty} |a_j|^{1+\gamma/2} < \infty$, and by (3.33) and (3.39), we have

$$
\sum_{i=0}^{\infty} \omega(i) \le \sum_{n=0}^{\infty} \max\big(|a_n|, |a_n|^{1+\gamma/2}\big) < \infty. \qquad (3.39)
$$

The general case $k \ge 1$ similarly follows. Similar to the proof of (3.39), we easily prove the other results. □

From Propositions 2.1 and 3.2, (A6) and (A7) hold. Hence, we can obtain the following two corollaries from Corollaries 2.1 and 2.2. In order to prove the following two corollaries, we first give some lemmas.

Lemma 3.7 (Liptser and Shiryayev [36]) *Let* $\xi = (\xi_k)_{-\infty<k<\infty}$ *be a strictly stationary sequence on a probability space* $(\Omega, \mathcal{F}, P)$, *and* $\mathcal{G}$ *be a* σ*-algebra of invariant sets of the sequence* ξ *and* $\mathcal{F}_k = \sigma(\dots, \xi_{k-1}, \xi_k)$. *For a certain* $p \ge 2$, *let* $E|\xi_0|^p < \infty$ *and* $\sum_{k\ge1} \gamma_k(p) < \infty$, *where* $\gamma_k(p) = \{E|E(\xi_k|F_0)|^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}$. *Then*

$$
Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k \xrightarrow{d} Z(stably),
$$

where the random variable Z *has the characteristic function* $E\exp(-\frac{1}{2}\lambda^2\sigma^2)$, *and* $\sigma^2 = E(\xi_0^2|\mathcal{G}) + 2\sum_{k\ge1} E(\xi_0\xi_k|\mathcal{G})$.

Corollary 3.5 *Assume that* (A1)–(A5) *hold,* $\varphi(t) = t\varphi'(0) + O(t^2)$ *and* $m(t) = O(|t|^\lambda)$ *for some* $\lambda > 0$ *as* $t \to 0$, $\Omega(\hat{\beta}_n) = O_p(r_n)$. *Under the conditions of Proposition* 3.2, $E|\psi(e_k)|^{\frac{p}{p-1}} <$

$\infty, p \geq 2$ and $\sum_{k=1}^{n} |x_{kn}| < \infty$, we have

$$n^{-1/2}\hat{\beta}_n \xrightarrow{d} Z(stably), \tag{3.40}$$

where the random variable Z has the characteristic function $E\exp(-\frac{1}{2}\lambda^2\sigma^2)$, and $\sigma^2 = (\varphi'(0))^{-2}x_{1n}^T x_{1n} E(\psi^2(e_1)|\mathcal{G}) + 2(\varphi'(0))^{-2}x_{1n}^T \sum_{k\geq 2} x_{kn} E(\psi(e_1)\psi(e_k)|\mathcal{G})$.

Proof By Proposition 2.1, Proposition 3.2 and Corollary 2.1, we have

$$\begin{aligned} n^{-1/2}\hat{\beta}_n &= n^{-1/2}\big(\varphi'(0)\big)^{-1}\sum_{i=1}^{n}\psi(e_i)x_{in} + O_p\big(n^{-1/2}\big(r_n^{\lambda}\log^{1/2} n + r_n\big)\big) \\ &= n^{-1/2}\big(\varphi'(0)\big)^{-1}\sum_{i=1}^{n}\psi(e_i)x_{in} + o_p(1). \end{aligned} \tag{3.41}$$

By $E|\psi(e_k)|^{\frac{p}{p-1}} < \infty$ and $\sum_{k=1}^{n}|x_{kn}| < \infty$, we have

$$\begin{aligned} \gamma_k(p) &= \big\{E\big|E\big(\psi(e_k)x_{kn}|\mathcal{F}_0\big)\big|^{\frac{p}{p-1}}\big\}^{\frac{p-1}{p}} \\ &\leq \big\{E\big[E\big(\big|\psi(e_k)x_{kn}\big|^{\frac{p}{p-1}}|\mathcal{F}_0\big)\big]\big\}^{\frac{p-1}{p}} \\ &= \big\{E\big|\psi(e_k)x_{kn}\big|^{\frac{p}{p-1}}\big\}^{\frac{p-1}{p}}\big\} \leq C|x_{kn}|, \end{aligned} \tag{3.42}$$

$$\sum_{k\geq 1}\gamma_k(p) = \sum_{k=1}^{n}|x_{kn}| < \infty \tag{3.43}$$

and

$$\begin{aligned} \sigma^2 &= \big(\varphi'(0)\big)^{-2}E\big(\big(\psi(e_1)x_{1n}\big)^2|\mathcal{G}\big) + 2\big(\varphi'(0)\big)^{-2}\sum_{k\geq 2}E\big(\psi(e_1)x_{1n}\psi(e_k)x_{kn}|\mathcal{G}\big) \\ &= \big(\varphi'(0)\big)^{-2}x_{1n}^2 E\big(\psi^2(e_1)|\mathcal{G}\big) + 2\big(\varphi'(0)\big)^{-2}x_{1n}\sum_{k\geq 2}x_{kn}E\big(\psi(e_1)\psi(e_k)|\mathcal{G}\big). \end{aligned}$$

By Proposition 2.1, Proposition 3.2 and Corollary 2.2, we easily obtain the following result. Here we omit the proof. □

Corollary 3.6 *Assume that* (A1)–(A5) *hold*, $\varphi(t) = t\varphi'(0) + O(t^2)$ *and* $m(t) = O(\sqrt{t})$ *as* $t \to 0$, $\tilde{\Omega}_n(\tilde{\beta}_n) = O_{a.s.}(\tilde{r}_n)$. *Under the conditions of Proposition* 3.2, *we have*

$$\tilde{\beta}_n = O_{a.s}\big(n^{-1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}\big), \quad \upsilon > 0.$$

4 Proofs of the main results

For the proofs of Theorem 2.1 and Theorem 2.2, we need some lemmas as follows.

Lemma 4.1 (Freedman [21]) *Let τ be a stopping time, and K a positive real number. Suppose that $P\{|\xi_i| \leq K, i \leq \tau\} = 1$, where $\{\xi_i\}$ are measurable random variables and*

$E(\xi_i|\mathcal{F}_{i-1}) = 0$. *Then, for all positive real numbers a and b,*

$$P\left\{\sum_{i=1}^{n}\xi_i \geq a \text{ and } T_n \leq b, \text{ for some } n \leq \tau\right\} \leq \left(\left(\frac{b}{Ka+b}\right)^{Ka+b} e^{Ka}\right)^{K^{-2}} \leq \exp\left(-\frac{a^2}{2(Ka+b)}\right).$$

Lemma 4.2 *Let*

$$M_n(\beta_n) = \sum_{i=1}^{n}\left\{\psi\left(e_i - x_{in}^T\beta_n\right) - E\left(\psi\left(e_i - x_{in}^T\beta_n\right)|\mathcal{F}_{i-1}\right)\right\}x_{in}. \tag{4.1}$$

Assume that (A5) *and* (A6) *hold. Then*

$$\sup_{|\beta_n|\leq\delta_n}\left|M_n(\beta_n) - M_n(0)\right| = O_p\left(\sqrt{\tau_n(\delta_n)}\log n + n^{-3}\right). \tag{4.2}$$

Proof Note that $p = \sum_{i=1}^{n} x_{in}^T x_{in} \leq (\max_{1\leq i\leq n}|x_{in}|)^2 n = nr_n^2$, and $\delta_n r_n \to 0$, we have $\delta_n = o(n^{1/2})$. For any positive sequence $\mu_n \to \infty$, let

$$\phi_n = 2\mu_n\sqrt{\tau_n(\delta_n)}\log n, \qquad t_n = \mu_n\sqrt{\tau_n(\delta_n)}/\log\mu_n, \qquad u_n = t_n^2,$$
$$\eta_i(\beta_n) = \left(\psi\left(e_i - x_{in}^T\beta_n\right) - \psi(e_i)\right)x_{in}, \qquad T_n = \max_{1\leq i\leq n}\sup_{|\beta_n|\leq\delta_n}\left|\eta_i(\beta_n)\right|$$

and

$$U_n = \sum_{i=1}^{n} E\left\{\left[\psi\left(e_i + |x_{in}|\delta_n\right) - \psi\left(e_i - |x_{in}|\delta_n\right)\right]^2|\mathcal{F}_{i-1}\right\}|x_{in}|^2.$$

By the monotonicity of ψ and $\delta \geq 0$, we have

$$\begin{aligned}\sup_{|\beta_n|\leq\delta}\left|\eta_i(\beta_n)\right| &\leq |x_{in}|\sup_{|\beta_n|\leq\delta}\left|\psi\left(e_i - x_{in}^T\beta_n\right) - \psi(e_i)\right| \\ &\leq |x_{in}|\max\left\{\psi\left(e_i - |x_{in}|\delta\right) - \psi(e_i), \psi\left(e_i + |x_{in}|\delta\right) - \psi(e_i)\right\} \\ &\leq |x_{in}|\left\{\psi\left(e_i + |x_{in}|\delta\right) - \psi\left(e_i - |x_{in}|\delta\right)\right\}. \end{aligned} \tag{4.3}$$

By (4.3), the c_r-inequality and (A3), we have

$$\begin{aligned} E\left(\sup_{|\beta_n|\leq\delta_n}\left|\eta_i(\beta_n)\right|^2\right) &\leq E\left\{|x_{in}|\left[\psi\left(e_i + |x_{in}|\delta\right) - \psi\left(e_i - |x_{in}|\delta\right)\right]\right\}^2 \\ &\leq 2|x_{in}|^2\left\{E\left[\psi\left(e_i + |x_{in}|\delta\right) - \psi(e_i)\right]^2 + E\left[\psi\left(e_i - |x_{in}|\delta\right) - \psi(e_i)\right]^2\right\} \\ &= 2|x_{in}|^2\left[m^2\left(|x_{in}|\delta\right) + m^2\left(-|x_{in}|\delta\right)\right]. \end{aligned}$$

Thus

$$E\left(T_n^2\right) = E\left(\max_{1\leq i\leq n}\sup_{|\beta_n|\leq\delta_n}\left|\eta_i(\beta_n)\right|^2\right) \leq \sum_{i=1}^{n} E\left(\sup_{|\beta_n|\leq\delta_n}\left|\eta_i(\beta_n)\right|^2\right)$$

$$\leq 2\sum_{i=1}^{n}|x_{in}|^2\big[m^2\big(|x_{in}|\delta_n\big)+m^2\big(-|x_{in}|\delta_n\big)\big]=2\tau_n(\delta_n). \tag{4.4}$$

By the Chebyshev inequality,

$$P\big(|T_n|\geq t_n\big)\leq E\big(T_n^2\big)/t_n^2\leq 2\tau_n(\delta_n)/t_n^2=2\log^2\mu_n/\mu_n^2\to 0. \tag{4.5}$$

Similarly,

$$P\big(|U_n|\geq t_n\big)\leq E(U_n)/u_n=O\big((\log\mu_n/\mu_n)^2\big)\to 0. \tag{4.6}$$

Let $x_{in}=(x_{i1n},\dots,x_{ipn})^T=(x_{i1},\dots,x_{ip})^T, D_x(i)=(2\times 1_{x_{i1}\geq 0}-1,\dots,2\times 1_{x_{ip}\geq 0}-1)\in \Pi_p, \Pi_p=\{-1,1\}^p$. For $d\in\Pi_p, j=1,2,\dots,p$, define

$$M_{n,j,d}(\beta_n)=\sum_{i=1}^{n}\big[\psi\big(e_i-x_{in}^T\beta_n\big)-E\big(\psi\big(e_i-x_{in}^T\beta_n\big)|F_{i-1}\big)\big]x_{ij}1_{D_x(i)=d}. \tag{4.7}$$

Since $M_n(\beta_n)=\sum_{d\in\Pi_p}(M_{n,1,d}(\beta_n),\dots,M_{n,p,d}(\beta_n))^T$, it suffices to prove that Lemma 4.2 holds with $M_n(\beta_n)$ replaced by $(M_{n,j,d}(\beta_n)$.

Let $|\beta_n|\leq\delta_n, \eta_{i,j,d}(\beta_n)=(\psi(e_i-x_{in}^T\beta_n)-\psi(e_i))x_{ij}1_{D_x(i)=d}$ and

$$B_n(\beta_n)=\sum_{i=1}^{n}E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|>t_n}|\mathcal{F}_{i-1}\big). \tag{4.8}$$

Note that

$$\frac{u_n}{t_n\phi_n}=\frac{t_n}{\phi_n}=\frac{\mu_n\sqrt{\tau_n(\delta_n)}/\log\mu_n}{2\mu_n\sqrt{\tau_n(\delta_n)}\log n}=\frac{1}{2\log n\log\mu_n}\to 0. \tag{4.9}$$

By (4.9), for large enough n, we have

$$\begin{aligned}P\big(\big|B_n(\beta_n)\big|\geq\phi_n,U_n\leq u_n\big)&=P\left(\left|\sum_{i=1}^{n}E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|>t_n}|\mathcal{F}_{i-1}\big)\right|\geq\phi_n,U_n\leq u_n\right)\\&\leq P\left(t_n^{-1}\sum_{i=1}^{n}E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|>t_n}|\mathcal{F}_{i-1}\big)\geq\phi_n,U_n\leq u_n\right)\\&\leq P\big(t_n^{-1}U_n\geq\phi_n,U_n\leq u_n\big)=P(t_n\phi_n\leq U_n\leq u_n)=0.\end{aligned} \tag{4.10}$$

Let the projections $\mathcal{P}_k(\cdot)=E(\cdot|\mathcal{F}_k)-E(\cdot|\mathcal{F}_{k-1})$. Since

$$\begin{aligned}&E\big\{\mathcal{P}_i\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\leq t_n}\big)|\mathcal{F}_{i-1}\big\}\\&\quad=E\big\{\big[E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\leq t_n}|\mathcal{F}_i\big)-E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\leq t_n}|\mathcal{F}_{i-1}\big)\big]|\mathcal{F}_{i-1}\big\}\\&\quad=E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\leq t_n}|\mathcal{F}_{i-1}\big)-E\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\leq t_n}|\mathcal{F}_{i-1}\big)=0.\end{aligned} \tag{4.11}$$

Note that $\{\mathcal{P}_i(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\leq t_n})\}$ are bound martingale differences. By Lemma 4.1 and (4.10), for $|\beta_n|\leq t_n$, we have

$$P\big\{\big|M_{n,j,d}(\beta_n)-M_{n,j,d}(0)\big|\geq 2\phi_n,T_n\leq t_n,U_n\leq u_n\big\}$$

$$
\begin{aligned}
&\le P\left\{\left|\sum_{i=1}^{n}\mathcal{P}_i\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|\le t_n}\big)\right|\ge\phi_n, T_n\le t_n, U_n\le u_n\right\}\\
&\quad+P\left\{\left|\sum_{i=1}^{n}\mathcal{P}_i\big(\eta_{i,j,d}(\beta_n)1_{|\eta_{i,j,d}(\beta_n)|>t_n}\big)\right|\ge\phi_n, T_n\le t_n, U_n\le u_n\right\}\\
&\le C\exp\left(-\frac{\phi_n^2}{4t_n\phi_n+2u_n}\right)+P\big(\big|B_n(\beta_n)\big|\ge\phi_n, U_n\le u_n\big)\\
&=O\left(\exp\left(-\frac{\phi_n^2}{4t_n\phi_n+2u_n}\right)\right).
\end{aligned}
\tag{4.12}
$$

Let $l=n^8$ and $K_l=\{(k_1/l,\dots,k_p/l):k_i\in Z,|k_i|\le n^9\}$. Then $\#K_l=(2n^9+1)^p$, where the symbol # denotes the number of elements of the set K_l. It is easy to show

$$
t_n\phi_n\log n=o\big(\phi_n^2\big)\quad\text{and}\quad u_n\log n=o\big(\phi_n^2\big).
\tag{4.13}
$$

By (4.12) and (4.13), for $\forall\varsigma>1$, we have

$$
\begin{aligned}
&P\Big\{\sup_{\beta_n\in K_l}\big|M_{n,j,d}(\beta_n)-M_{n,j,d}(0)\big|\ge2\phi_n, T_n\le t_n, U_n\le u_n\Big\}\\
&\quad\le\sum_{\#K_l}P\big\{\big|M_{n,j,d}(\beta_n)-M_{n,j,d}(0)\big|\ge2\phi_n, T_n\le t_n, U_n\le u_n\big\}\\
&\quad\le Cn^{9p}\exp\left(-\frac{\phi_n^2}{4t_n\phi_n+2u_n}\right)=Cn^{9p}\exp\left(-\frac{\log n}{4t_n\phi_n\log n/\phi_n^2+2u_n\log n/\phi_n^2}\right)\\
&\quad=Cn^{9p}\exp\left(-\frac{\log n}{o(1)}\right)=o\big(n^{-\varsigma p}\big).
\end{aligned}
\tag{4.14}
$$

By (4.5), (4.6) and (4.14), we have

$$
P\Big\{\sup_{\beta_n\in K_l}\big|M_{n,j,d}(\beta_n)-M_{n,j,d}(0)\big|\ge2\phi_n\Big\}\to0,\quad n\to\infty.
\tag{4.15}
$$

For a, let $\langle a\rangle_{l,-1}=\lceil a\rceil_l=\lceil al\rceil/l$ and $\langle a\rangle_{l,1}=\lfloor a\rfloor_l=\lfloor al\rfloor/l$. For a vector $\beta_n=(\beta_{1n},\dots,\beta_{pn})^T$, let $\langle\beta_n\rangle_{l,d}=(\langle\beta_{1n}\rangle_{l,d_1},\dots,\langle\beta_{pn}\rangle_{l,d_p})$.

By (A5), for $|s|,|t|\le r_n\delta_n$ and large n, we have

$$
\big|E\big\{\big[\psi(e_i-t)-\psi(e_i-s)\big]|\mathcal{F}_{i-1}\big\}\big|\le L_{i-1}|s-t|.
$$

Let $V_n=\sum_{i=1}^n L_{i-1}$. By condition (A5), the Markov inequality and $L_i\in L^1$, we have

$$
P\big(V_n\ge n^4\big)\le EV_n/n^4=\sum_{i=1}^{n}EL_{i-1}/n^4\le Cn^{-3}.
\tag{4.16}
$$

Note that $|\beta_n-\langle\beta_n\rangle_{l,d}|\le Cl^{-1}$, which implies $\max_{1\le i\le n}|x_{in}^T(\beta_n-\langle\beta_n\rangle_{l,d})|=o(l^{-1})$. Thus

$$
\sup_{|\beta_n|\le\delta_n}\left|\sum_{i=1}^{n}E\big\{\big[\eta_i\big(\langle\beta_n\rangle_{l,d}\big)-\eta_i(\beta_n)\big]|\mathcal{F}_{i-1}\big\}x_{in}\right|
$$

$$
\begin{aligned}
&\le \sup_{|\beta_n|\le\delta_n} \sum_{i=1}^{n} \bigl|E\bigl(\bigl(\psi\bigl(e_i - x_{in}^T\langle\beta_n\rangle_{l,d}\bigr) - \psi(e_i)\bigr) - \bigl(\psi\bigl(e_i - x_{in}^T\beta_n\bigr) - \psi(e_i)\bigr)|\mathcal{F}_{i-1}\bigr)x_{in}\bigr| \\
&= \sup_{|\beta_n|\le\delta_n} \sum_{i=1}^{n} \bigl|E\bigl(\psi\bigl(e_i - x_{in}^T\langle\beta_n\rangle_{l,d}\bigr) - \psi\bigl(e_i - x_{in}^T\beta_n\bigr)|\mathcal{F}_{i-1}\bigr)x_{in}\bigr| \\
&\le \sup_{|\beta_n|\le\delta_n} \sum_{i=1}^{n} |x_{in}|L_{i-1}\bigl|x_{in}^T\bigl(\langle\beta_n\rangle_{l,d} - \beta_n\bigr)\bigr| \le Cl^{-1}V_n. \qquad (4.17)
\end{aligned}
$$

Without loss of generality, assume that $j = 1$ in the following proof.

Let $d = (1,-1,1,\ldots,1)$. Then $\langle\beta_n\rangle_{l,d} = (\lfloor\beta_{1n}\rfloor_l, \lceil\beta_{2n}\rceil_l, \lfloor\beta_{3n}\rfloor_l, \ldots, \lfloor\beta_{pn}\rfloor_l)$ and $\langle\beta_n\rangle_{l,-d} = (\lceil\beta_{1n}\rceil_l, \lfloor\beta_{2n}\rfloor_l, \lceil\beta_{3n}\rceil_l, \ldots, \lceil\beta_{pn}\rceil_l)$. Since ψ is nondecreasing,

$$
\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr) \le \eta_{i,1,d}(\beta_n) \le \eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr).
$$

Note that

$$
\begin{aligned}
&\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr) - E\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr)|\mathcal{F}_{i-1}\bigr] + E\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr)|\mathcal{F}_{i-1}\bigr] \\
&\quad - E\bigl[\eta_{i,1,d}(\beta_n)|\mathcal{F}_{i-1}\bigr] \\
&\le \eta_{i,1,d}(\beta_n) - E\bigl[\eta_{i,1,d}(\beta_n)|\mathcal{F}_{i-1}\bigr] \\
&\le \eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr) - E\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr)|\mathcal{F}_{i-1}\bigr] + E\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr)|\mathcal{F}_{i-1}\bigr] \\
&\quad - E\bigl[\eta_{i,1,d}(\beta_n)|\mathcal{F}_{i-1}\bigr].
\end{aligned}
$$

Namely

$$
\begin{aligned}
&\sum_{i=1}^{n}\bigl\{\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr) - E\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr)|\mathcal{F}_{i-1}\bigr] \\
&\quad + E\bigl[\bigl(\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr) - \eta_{i,1,d}(\beta_n)\bigr)|\mathcal{F}_{i-1}\bigr]\bigr\}x_{i1}1_{D_x(i)=d} \\
&\le \sum_{i=1}^{n}\bigl\{\eta_{i,1,d}(\beta_n) - E\bigl[\eta_{i,1,d}(\beta_n)|\mathcal{F}_{i-1}\bigr]\bigr\}x_{i1}1_{D_x(i)=d} \\
&\le \sum_{i=1}^{n}\bigl\{\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr) - E\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr)|\mathcal{F}_{i-1}\bigr] \\
&\quad + E\bigl[\bigl(\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr) - \eta_{i,1,d}(\beta_n)\bigr)|\mathcal{F}_{i-1}\bigr]\bigr\}x_{i1}1_{D_x(i)=d}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&M_{n,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr) - M_{n,1,d}(0) + \sum_{i=1}^{n} E\bigl\{\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,-d}\bigr) - \eta_{i,1,d}(\beta_n)\bigr]|\mathcal{F}_{i-1}\bigr\}x_{i1}1_{D_x(i)=d} \\
&\le M_{n,1,d}(\beta_n) - M_{n,1,d}(0) \\
&\le M_{n,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr) - M_{n,1,d}(0) \\
&\quad + \sum_{i=1}^{n} E\bigl\{\bigl[\eta_{i,1,d}\bigl(\langle\beta_n\rangle_{l,d}\bigr) - \eta_{i,1,d}(\beta_n)\bigr]|\mathcal{F}_{i-1}\bigr\}x_{i1}1_{D_x(i)=d}. \qquad (4.18)
\end{aligned}
$$

By (4.17) and (4.18), we have

$$\begin{aligned} M_{n,1,d}\big(\langle\beta_n\rangle_{l,-d}\big) - M_{n,1,d}(0) - Cl^{-1}V_n &\le M_{n,1,d}(\beta_n) - M_{n,1,d}(0) \\ &\le M_{n,1,d}\big(\langle\beta_n\rangle_{l,d}\big) - M_{n,1,d}(0) + Cl^{-1}V_n. \end{aligned} \tag{4.19}$$

Note that $l^{-1}V_n = O_p(n^{-8}n^4) = O_p(n^{-4})$, (4.2) immediately follows from (4.15) and (4.19). □

Lemma 4.3 *Assume that the processes* $X_t = g(\mathcal{F}_t) \in L^2$. *Let* $g_n(\mathcal{F}_0) = E(g(\mathcal{F}_n)|\mathcal{F}_0), n \ge 0$. *Then*

$$\begin{aligned} &\big\|g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big)\big\| \le \big\|g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big\|, \\ &\|\mathcal{P}_0X_n\| \le \big\|g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big)\big\| + R, \end{aligned} \tag{4.20}$$

where $R = \|E[g_n(\mathcal{F}_0^*)|\mathcal{F}_{-1}] - E[g_n(\mathcal{F}_0^*)|\mathcal{F}_0]\|$.

Proof Since

$$\begin{aligned} E\big\{\big[g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big]|\big(\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_0\big)\big\} &= E\big[g(\mathcal{F}_n)|(\mathcal{F}_{-1}, \varepsilon_0)\big] - E\big[g\big(\mathcal{F}_n^*\big)|\big(\mathcal{F}_{-1}, \varepsilon_0'\big)\big] \\ &= g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big), \end{aligned}$$

we have

$$E\big|E\big\{\big[g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big]|\big(\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_0\big)\big\}\big|^2 = E\big|g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big)\big|^2. \tag{4.21}$$

By the Jensen inequality, we have

$$\begin{aligned} E\big|E\big\{\big[g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big]|\big(\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_0\big)\big\}\big|^2 &\le E\big\{E\big[\big|g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big|^2|\big(\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_0\big)\big]\big\} \\ &= E\big|g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big|^2. \end{aligned} \tag{4.22}$$

By (4.21) and (4.22), we have

$$E\big|g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big)\big|^2 \le E\big|g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big|^2.$$

That is,

$$\big\|g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big)\big\| \le \big\|g(\mathcal{F}_n) - g\big(\mathcal{F}_n^*\big)\big\|. \tag{4.23}$$

Note that

$$E\big[g_n(\mathcal{F}_0)|\mathcal{F}_{-1}\big] = E\big[E\big(g(\mathcal{F}_n)|\mathcal{F}_0\big)|\mathcal{F}_{-1}\big] = E\big(g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_{-1}\big) \tag{4.24}$$

and

$$E\big[g_n(\mathcal{F}_0)|\mathcal{F}_{-1}\big] = E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_0\big] + E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_{-1}\big] - E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_0\big]. \tag{4.25}$$

By (4.24), (4.25) and the Jensen inequality, we have

$$\begin{aligned}\|\mathcal{P}_0X_n\| &= \big\|E\big(g(\mathcal{F}_n)|\mathcal{F}_0\big) - E\big(g(\mathcal{F}_n)|\mathcal{F}_{-1}\big)\big\| \\ &= \big\|E\big[g_n(\mathcal{F}_0)|\mathcal{F}_0\big] - E\big[g_n(\mathcal{F}_0)|\mathcal{F}_{-1}\big]\big\| \\ &= \big\|E\big[g_n(\mathcal{F}_0)|\mathcal{F}_0\big] - E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_0\big] - E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_{-1}\big] + E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_0\big]\big\| \\ &\leq \big\|E\big[g_n(\mathcal{F}_0)|\mathcal{F}_0\big] - E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_0\big]\big\| + \big\|E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_{-1}\big] - E\big[g_n\big(\mathcal{F}_0^*\big)|\mathcal{F}_0\big]\big\| \\ &\leq \big\|g_n(\mathcal{F}_0) - g_n\big(\mathcal{F}_0^*\big)\big\| + R. \end{aligned} \tag{4.26}$$

□

Remark 4 If $\{\varepsilon_i\}$ i.i.d., then $R = 0$. In this case, the above lemma becomes Theorem 1 of Wu [48].

Lemma 4.4 *Let* $\{\delta_n, n \in N\}$ *be a sequence of positive numbers such that* $\delta_n \to \infty$ *and* $\delta_n r_n \to 0$. *If* (A6)–(A7) *hold, then*

$$\Big\|\sup_{|\beta_n|\leq\delta_n}\big|N_n(\beta_n) - N_n(0)\big|\Big\| = O\left(\sqrt{\sum_{i=1}^n |x_{in}|^4\delta_n}\right), \tag{4.27}$$

where

$$N_n(\beta_n) = \sum_{i=1}^n \big\{\psi_i\big(-x_{in}^T\beta_n;\mathcal{F}_{i-1}\big) - \varphi\big(-x_{in}^T\beta_n\big)\big\}x_{in}.$$

Proof Let $I = \{n_1,\dots,n_q\} \in \{1,2,\dots,p\}$ be a nonempty set and $1 \leq n_1 < \cdots < n_q$, and $u_I = (u_1 1_{1\in I},\dots,u_p 1_{p\in I})$, with vector $u = (u_1,\dots,u_p)$. Write

$$\begin{aligned}&\int_0^{\beta_{n,I}} \frac{\partial^q N_n(u_I)}{\partial u_I}\,du_I \\ &\quad= \int_0^{\beta_{n,m_1}}\cdots\int_0^{\beta_{n,m_q}} \frac{\partial^q N_n(u_I)}{\partial u_{m_1}\cdots\partial u_{m_q}}\,du_{m_1}\cdots du_{m_q}, w_i = x_{in}x_{im_1}\cdots x_{im_q}.\end{aligned}$$

In the following, we will prove that

$$\left|\frac{\partial^q N_n(u_I)}{\partial u_I}\right| = \left|\sum_{i=1}^n \big\{\psi_i^{(q)}\big(-x_{in}^T u_I;\mathcal{F}_{i-1}\big) - \varphi^{(q)}\big(-x_{in}^T u_I\big)\big\}w_i\right| = O\left(\sqrt{\sum_{i=1}^n |x_{in}|^{2+2q}}\right) \tag{4.28}$$

uniformly over $|u| \leq p\delta_n$.

In fact, let

$$T_n = \sum_{i=1}^n \big\{\psi_i^{(q)}\big(-x_{in}^T u_I;\mathcal{F}_{i-1}\big) - \varphi^{(q)}\big(-x_{in}^T u_I\big)\big\}w_i$$

and

$$J_k = \sum_{i=1}^n \mathcal{P}_{i-k}\big\{\psi_i^{(q)}\big(-x_{in}^T u_I;\mathcal{F}_{i-1}\big) - \varphi^{(q)}\big(-x_{in}^T u_I\big)\big\}w_i.$$

Then $T_n = \sum_{k=0}^{\infty} J_k$, and J_k are martingale differences. By the orthogonality of martingale differences and the stationarity of $\{e_i\}$, and Lemma 4.3, we have

$$\begin{aligned} \|J_k\|^2 &= \sum_{i=1}^{n} \left\| \mathcal{P}_{i-k} \left\{ \psi_i^{(q)} \left(-x_{in}^T u_I; \mathcal{F}_{i-1} \right) - \varphi^{(q)} \left(-x_{in}^T u_I \right) \right\} w_i \right\|^2 \\ &= \sum_{i=1}^{n} |w_i|^2 \left\| \mathcal{P}_0 \left\{ \psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1} \right) - \varphi^{(q)} \left(-x_{kn}^T u_I \right) \right\} \right\|^2. \end{aligned} \tag{4.29}$$

By Lemma 4.3, $\psi_i(\cdot; \mathcal{F}_{i-1}) \in C^l, l \geq 0$ and the c_r-inequality, for $k \geq 0$, we have

$$\begin{aligned} &\left\| \mathcal{P}_0 \left\{ \psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1} \right) - \varphi^{(q)} \left(-x_{kn}^T u_I \right) \right\} \right\|^2 \\ &\quad \leq \left\| E \left\{ \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1} \right) - \varphi^{(q)} \left(-x_{kn}^T u_I \right) \right] | \mathcal{F}_0 \right\} \right. \\ &\qquad \left. - E \left\{ \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1} \right) - \varphi^{(q)} \left(-x_{kn}^T u_I \right) \right] | \mathcal{F}_0^* \right\} \right\|^2 + R_k^2 \\ &\quad \leq 2 \left\| E \left\{ \psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1} \right) | \mathcal{F}_0 \right\} - E \left\{ \psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_0^* \right\} \right\|^2 \\ &\qquad + 2 \left\| E \left\{ \varphi^{(q)} \left(-x_{kn}^T u_I \right) | \mathcal{F}_0 \right\} - E \left\{ \varphi^{(q)} \left(-x_{kn}^T u_I \right) | \mathcal{F}_0^* \right\} \right\|^2 + R_k^2 \\ &\quad \leq 2 \left\| E \left\{ \psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1} \right) | \mathcal{F}_0 \right\} - E \left\{ \psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_0^* \right\} \right\|^2 \\ &\qquad + 2 \left| E\psi^{(q)} \left(e_k - x_{kn}^T u_I \right) - E\psi^{(q)} \left(e_k^* - x_{kn}^T u_I \right) \right|^2 + R_k^2, \end{aligned} \tag{4.30}$$

where

$$\begin{aligned} R_k^2 = &\left\| E \left\{ \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) - \varphi^{(q)} \left(-x_{kn}^T u_I \right) \right] | \mathcal{F}_{-1} \right\} \right. \\ &\left. - E \left\{ \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) - \varphi^{(q)} \left(-x_{kn}^T u_I \right) \right] | \mathcal{F}_0 \right\} \right\|^2. \end{aligned}$$

Note that $E\psi^{(q)}(e_i + \delta) = \frac{d^q E\psi(e_i+t)}{dt^q}|_{t=\delta}$, we have

$$\begin{aligned} R_k^2 &\leq \left\| E \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_{-1} \right] - E \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_0 \right] \right\|^2 \\ &\quad + \left\| E \left[\varphi^{(q)} \left(-x_{kn}^T u_I \right) | F_{-1} \right] - E \left[\varphi^{(q)} \left(-x_{kn}^T u_I \right) | \mathcal{F}_0 \right] \right\|^2 \\ &= \left\| E \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_{-1} \right] - E \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_0 \right] \right\|^2 \\ &\quad + \left\| E\psi_k^{(q)} \left(e_k - x_{kn}^T u_I \right) - E\psi_k^{(q)} \left(e_k - x_{kn}^T u_I \right) \right\|^2 \\ &= \left\| E \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_{-1} \right] - E \left[\psi_k^{(q)} \left(-x_{kn}^T u_I; \mathcal{F}_{k-1}^* \right) | \mathcal{F}_0 \right] \right\|^2. \end{aligned} \tag{4.31}$$

By the conditions (A6), (A7) and (4.29)–(4.31), we have

$$\begin{aligned} \|T_n\| &= \sqrt{\sum_{i=1}^{n} |w_i|^2} \sum_{k=0}^{\infty} \|J_k\| = O\left(\sqrt{\sum_{i=1}^{n} |w_i|^2} \right) \\ &= O\left(\sqrt{\sum_{i=1}^{n} |x_{in}|^{2+2q}} \right). \end{aligned}$$

Let $|u| \le p\delta_n$. By $\max_{1\le i\le n}|x_{in}u| \le p\delta_n r_n \to 0$. Note that $\delta_n \to \infty$ and $\delta_n r_n \to 0$. By (4.28), we have

$$\begin{aligned}\left\| \sup_{|\beta_n|\le\delta_n} \int_0^{\beta_{n,I}} \left| \frac{\partial^p N_n(u_I)}{\partial u_I} \right| du_I \right\| &\le \left\| \int_{-\delta_n}^{\delta_n} \cdots \int_{-\delta_n}^{\delta_n} \left| \frac{\partial^p N_n(u_I)}{\partial u_I} \right| du_I \right\| \\ &\le \int_{-\delta_n}^{\delta_n} \cdots \int_{-\delta_n}^{\delta_n} \left\| \frac{\partial^p N_n(u_I)}{\partial u_I} \right\| du_I \\ &= O\left(\delta_n^q \sqrt{\sum_{i=1}^n |x_{in}|^{2+2q}} \right) = O\left(\delta_n \sqrt{\sum_{i=1}^n |x_{in}|^4} \right). \qquad (4.32)\end{aligned}$$

Since

$$N_n(\beta_n) - N_n(0) = \sum_{I\in\{1,2,\dots,p\}} \int_0^{\beta_{n,I}} \frac{\partial^{|I|} N_n(u_I)}{\partial u_I} du_I, \qquad (4.33)$$

the result (4.27) follows from (4.32) and (4.33). □

Lemma 4.5 *Let $\pi_i, i \ge 1$ be a sequence of bounded positive numbers, and let there exist a constant $c_0 \ge 1$ such that $\max_{1\le i\le 2^d} \pi_i \le c_0 \min_{1\le i\le 2^d} \pi_i$ holds for all large n. And let $\omega_d = 2c_0\pi_{2^d}$ and $q > 3/2$. Assume that* (A5) *and $\tilde{r}_n = O(\sqrt{n})$ hold. Then as $d \to \infty$*

$$\sup_{|\beta_n|\le\omega_d} \max_{n<2^d} \left| \tilde{M}_n(\beta) - \tilde{M}_n(0) \right| = O_p\left(\sqrt{\tilde{\tau}_{2^d}(\omega_d)}\, d^q + 2^{-5d/2} \right),$$

where $\tilde{M}_n(\beta) = \sum_{i=1}^n \{\psi(e_i - x_i^T\beta) - E(\psi(e_i - x_i^T\beta)|\mathcal{F}_{i-1})\}x_i$.

Proof Let

$$\begin{aligned}&\mu_n = (\log n)^{q-1}, \qquad \tilde{\phi}_n = 2\mu_{2^d}\sqrt{\tilde{\tau}_{2^d}(\omega_d)}\log(2^d), \qquad \tilde{t}_{2^d} = \mu_{2^d}\sqrt{\tilde{\tau}_{2^d}(\omega_d)}/\log\mu_{2^d}, \\ &\tilde{\mu}_{2^d} = \tilde{t}_{2^d}^2, \qquad \tilde{\eta}_i(\beta) = \left(\psi\left(e_i - x_i^T\beta\right) - \psi(e_i)\right)x_i, \qquad \tilde{T}_{2^d} = \max_{1\le i\le 2^d} \sup_{|\beta_n|\le\omega_d} \left|\tilde{\eta}_i(\beta)\right|\end{aligned}$$

and

$$\tilde{U}_{2^d} = \sum_{i=1}^{2^d} E\left\{\left[\psi\left(e_i + |x_i|\omega_d\right) - \psi\left(e_i - |x_i|\omega_d\right)\right]^2 |\mathcal{F}_{i-1}\right\}|x_i|^2.$$

Since $q > 3/2$ and $2(q-1) > 1$, $\sum_{d=2}^\infty (\mu_{2^d}^{-1}\log\mu_{2^d})^2 < \infty$. By the argument of Lemma 4.2 and the Borel–Cantelli lemma, we have

$$P(\tilde{T}_{2^d} \ge \tilde{t}_{2^d}, i.o.) = 0 \quad \text{and} \quad P(\tilde{U}_{2^d} \ge \tilde{u}_{2^d}, i.o.) = 0. \qquad (4.34)$$

Similar to the proof of (4.12), we have

$$\begin{aligned}&P\left\{\max_{k\le 2^d}\left|\tilde{M}_{k,j,d}(\beta) - M_{k,j,d}(0)\right| \ge 2\tilde{\phi}_{2^d}, \tilde{T}_{2^d} \le \tilde{t}_{2^d}, \tilde{U}_{2^d} \le \tilde{u}_{2^d}\right\} \\ &\quad = O\left(\exp\left(-\frac{\tilde{\phi}_{2^d}^2}{4\tilde{t}_{2^d}\tilde{\phi}_{2^d} + 2\tilde{u}_{2^d}}\right)\right). \qquad (4.35)\end{aligned}$$

Let $l = n^{8d}$ and $K_l = \{(k_1/l, \ldots, k_p/l) : k_i \in Z, |k_i| \leq n^{9d}\}$. Then $\#K_l = (2n^{9d} + 1)^p$. By (4.34) and (4.35), for $\forall\varsigma > 1$, we have

$$P\Big\{\sup_{\beta\in K_l}\big|\tilde{M}_{k,j,d}(\beta) - M_{k,j,d}(0)\big| \geq 2\tilde{\phi}_{2^d}, \tilde{T}_{2^d} \leq \tilde{t}_{2^d}, \tilde{U}_{2^d} \leq \tilde{u}_{2^d}\Big\} = O\big(n^{-\varsigma dp}\big). \tag{4.36}$$

Therefore,

$$P\Big\{\sup_{\beta\in K_l}\big|\tilde{M}_{k,j,d}(\beta) - M_{k,j,d}(0)\big| \geq 2\tilde{\phi}_{2^d}, \text{i.o.}\Big\} \to 0, \quad n \to \infty. \tag{4.37}$$

Since $\tilde{r}_n = O(\sqrt{n})$ and $\max_{1\leq i\leq 2^d} |x_i^T(\beta - \langle\beta\rangle_{l,d})| = O(2^{2^d} l^{-1})$, $Cl^{-1}V$ in (4.17) can be replaced by $Cl^{-1}2^{2^d}V$, and the lemma follows from $P(V_{2^d} \geq 2^{5d}, \text{i.o.}) = 0$. □

Lemma 4.6 *Let $\pi_i, i \geq 1$ be a sequence of bounded positive numbers, and let there exist a constant $c_0 \geq 1$ such that $\max_{1\leq i\leq 2^d} \pi \leq c_0 \min_{1\leq i\leq 2^d} \pi_i$ and $\pi_n = o(n^{-1/2}(\log n)^2)$ hold for all large n. And let $\omega_d = 2c_0\pi_{2^d}$. Assume that* (A6), (A7) *and $\tilde{r}_n = O(\sqrt{n}(\log n)^{-2})$ hold. Then*

$$\Big\|\sup_{|\beta|\leq\pi_n}\big|\tilde{N}_n(\beta) - \tilde{N}_n(0)\big|\Big\| = O\left(\sqrt{\sum_{i-1}^{n}|x_i|^4\pi_n}\right), \tag{4.38}$$

and, as $d \to \infty$, for any $\upsilon > 0$,

$$\sup_{|\beta|\leq\omega_d}\max_{n<2^d}\big|\tilde{N}_n(\beta) - \tilde{N}_n(0)\big|^2 = o_{a.s.}\left(\sum_{i=0}^{2^d}|x_i|^4\omega_d^2 d^5(\log d)^{1+\upsilon}\right), \tag{4.39}$$

where $\tilde{N}_n(\beta) = \sum_{i=1}^{n}\{\psi_1(-x_i^T\beta; \mathcal{F}_{i-1}) - \varphi(-x_i^T\beta)\}x_n$.

Proof Let $Q_{n,j}(\beta) = \sum_{i=1}^{n}\psi_1(-x_i^T\beta; F_{i-1})x_{ij}, i \leq j \leq p$, and

$$S_n(\beta) = Q_{n,j}(\beta) - Q_{n,j}(0). \tag{4.40}$$

Note that

$$\pi_n\tilde{r}_n = o\big(n^{-1/2}(\log n)^2\big)O\big(\sqrt{n}(\log n)^{-2}\big) = o(1). \tag{4.41}$$

It is easy to see that the argument in the proof of Lemma 4.4 implies that there exists a positive constant $C < \infty$ such that

$$E\Big\{\Big|\sup_{|\beta|\leq\omega_d}\big|S_n(\beta) - S_{n'}(\beta)\big|\Big|^2\Big\} \leq C\sum_{q=1}^{p}\omega_d^{2q}\sum_{i=n'+1}^{n}|x_i|^2 + 2q \tag{4.42}$$

holds uniformly over $1 \leq n' < n \leq 2^d$. Therefore (4.38) holds.

Let $\Lambda = \sum_{r=0}^{d}\mu_r^{-1}$, where

$$\mu_r = \left\{\sum_{m=1}^{2^{d-r}}\Big\|\sup_{|\beta|\leq\omega_d}\big|S_{2^r m}(\beta) - S_{2^r(m-1)}(\beta)\big|\Big\|^2\right\}^{-1/2}. \tag{4.43}$$

For a positive integer $k \le 2^d$, write its dyadic expansion $k = 2^{r_1} + \cdots + 2^{r_j}$, where $0 \le r_j < \cdots < r_1 \le d$, and $k(i) = 2^{r_1} + \cdots + 2^{r_i}$. By the Schwartz inequality, we have

$$\begin{aligned}
\sup_{|\beta|\le\omega_d} \left|S_k(\beta)\right|^2 &\le \left\{\sum_{i=1}^{j} \sup_{|\beta|\le\omega_d} \left|S_{k(i)}(\beta) - S_{k(i-1)}(\beta)\right|\right\}^2 \\
&= \left\{\sum_{i-1}^{j} \mu_{r_i}^{-1/2} \cdot \mu_{r_i}^{1/2} \sup_{|\beta|\le\omega_d} \left|S_{k(i)}(\beta) - S_{k(i-1)}(\beta)\right|\right\}^2 \\
&\le \sum_{i-1}^{j} \mu_{r_i}^{-1} \sum_{i-1}^{j} \mu_{r_i} \sup_{|\beta|\le\omega_d} \left|S_{k(i)}(\beta) - S_{k(i-1)}(\beta)\right|^2 \\
&\le \Lambda \sum_{i-1}^{j} \mu_{r_i} \sum_{m=1}^{2^{d-\eta}} \sup_{|\beta|\le\omega_d} \left|S_{2^\eta m}(\beta) - S_{2^\eta (m-1)}(\beta)\right|^2 \\
&\le \Lambda \sum_{r=0}^{d} \mu_r \sum_{m=1}^{2^{d-r}} \sup_{|\beta|\le\omega_d} \left|S_{2^r m}(\beta) - S_{2^r (m-1)}(\beta)\right|^2. \qquad (4.44)
\end{aligned}$$

Thus

$$\begin{aligned}
\left\|\max_{n\le 2^d} \sup_{|\beta|\le\omega_d} \left|S_n(\beta)\right|\right\| &= \left\|\sum_{n=1}^{2^d} \sup_{|\beta|\le\omega_d} \left|S_n(\beta)\right|\right\| \\
&\le \sum_{n=1}^{2^d} \left\|\sup_{|\beta|\le\omega_d} \left|S_n(\beta)\right|\right\| \le \sum_{n=1}^{2^d} \left\{E\left|\sup_{|\beta|\le\omega_d} \left|S_n(\beta)\right|\right|^2\right\}^{1/2} \\
&\le \sum_{r=0}^{d} \left\{E\left|\Lambda \sum_{r=0}^{d} \mu_r \sum_{m=1}^{2^{d-r}} \left|\sup_{|\beta|\le\omega_d} \left|S_{2^r m}(\beta) - S_{2^r (m-1)}(\beta)\right|\right|^2\right|\right\}^{1/2} \\
&\le \sum_{r=0}^{d} \left\{\Lambda \sum_{r=0}^{d} \mu_r \sum_{m=1}^{2^{d-r}} \left\|\sup_{|\beta|\le\omega_d} \left|S_{2^r m}(\beta) - S_{2^r (m-1)}(\beta)\right|\right\|^2\right\}^{1/2} \\
&\le \sum_{r=0}^{d} \left\{\Lambda \sum_{r=0}^{d} \mu_r \mu_R^{-2}\right\}^{1/2} = \sum_{r=0}^{d} \left\{\Lambda \sum_{r=0}^{d} \mu_r^{-1}\right\}^{1/2} = d\Lambda. \qquad (4.45)
\end{aligned}$$

Since $\upsilon > 0$ and $\omega_d^{2q} \sum_{i=1}^{2^d} |x_i|^{2+2q} = O(\omega_d^2 \sum_{i=1}^{2^d} |x_i|^4)$, (4.42) implies that

$$\sum_{d=2}^{\infty} \frac{\|\max_{n\le 2^d} \sup_{|\beta|\le\omega_d} |S_n(\beta)|\|^2}{\omega_d^2 \sum_{i=1}^{2^d} |x_i|^4 d^5 (\log d)^{1+\upsilon}} = \sum_{d=2}^{\infty} \frac{O(d^2(d+1)^2)}{d^5(\log d)^{1+\upsilon}} < \infty. \qquad (4.46)$$

By the Borel–Cantelli lemma, (4.39) follows from (4.46). □

Lemma 4.7 *Under the conditions of Theorem* 2.2, *we have*:

(1) $\sup_{|\beta|\le b_n} |\tilde{K}_n(\beta) - \tilde{K}_n(0)| = O_{a.s.}(L_{\tilde{n}} + B_{\tilde{n}})$;

(2) *for and* $\upsilon > 0, \tilde{K}_n(0) = O_{a.s.}(h_n)$, *where* $h_n = n^{1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon/4}$.

Proof Observe that $\tilde{K}_n(\beta) = \tilde{M}_n(\beta) + \tilde{N}_n(\beta)$. Since $n^{-5/2} = o(B_{\tilde{n}})$, (1) follows from Lemma 4.5 and 4.6. □

As with the argument in (4.29), we have $\tilde{K}_n(0) = O(\sqrt{n})$.

Proof of Theorem 2.1 Observe that

$$\begin{aligned} K_n(\beta_n) &= \sum_{i=1}^n \psi\big(e_i - x_{in}^T\beta_n\big)x_{in} - E\left(\sum_{i=1}^n \psi\big(e_i - x_{in}^T\beta_n\big)x_{in}\right) \\ &= \sum_{i=1}^n \big\{\psi\big(e_i - x_{in}^T\beta_n\big) - E\big(\psi\big(e_i - x_{in}^T\beta_n\big)|\mathcal{F}_{i-1}\big)\big\}x_{in} \\ &\quad + \sum_{i=1}^n \big\{E\big(\psi\big(e_i - x_{in}^T\beta_n\big)|\mathcal{F}_{i-1}\big) - E\psi\big(e_i - x_{in}^T\beta_n\big)\big\}x_{in} \\ &= M_n(\beta_n) + N_n(\beta_n). \end{aligned} \tag{4.47}$$

By (4.47), Lemma 4.2 and Lemma 4.4, we have

$$\begin{aligned} \sup_{|\beta_n|\le\delta_n} \big|K_n(\beta_n) - K_n(0)\big| &\le \sup_{|\beta_n|\le\delta_n} \big|M_n(\beta_n) - M_n(0)\big| + \sup_{|\beta_n|\le\delta_n} \big|N_n(\beta_n) - N_n(0)\big| \\ &= O_p\big(\sqrt{\tau_n(\delta_n)}\log n + n^{-3}\big) + O\left(\sqrt{\sum_{i=1}^n |x_{in}|^4\delta_n}\right) \\ &= O_p\left(\sqrt{\tau_n(\delta_n)}\log n + \delta_n\sqrt{\sum_{i=1}^n |x_{in}|^4}\right). \end{aligned} \tag{4.48}$$

This completes the proof of Theorem 2.1. □

Proof of Corollary 2.1 Take an arbitrary sequence $\delta_n \to \infty$, which satisfies the assumption of Theorem 2.1. Note that

$$K_n(0) = \sum_{i=1}^n \psi(e_i)x_{in} - E\left(\sum_{i=1}^n \psi(e_i)x_{in}\right) = \sum_{i=1}^n \psi(e_i)x_{in} \tag{4.49}$$

and

$$\begin{aligned} K_n(\hat{\beta}_n) &= \sum_{i=1}^n \psi\big(e_i - x_{in}^T\hat{\beta}_n\big)x_{in} - E\left(\sum_{i=1}^n \psi\big(e_i - x_{in}^T\hat{\beta}_n\big)x_{in}\right) \\ &= \sum_{i=1}^n \psi\big(y_i - x_{in}^T\hat{\beta}_n\big)x_{in} - \sum_{i=1}^n \varphi\big(-x_{in}^T\hat{\beta}_n\big)x_{in} \\ &= -\sum_{i=1}^n \varphi\big(-x_{in}^T\hat{\beta}_n\big)x_{in} + O_P(r_n) \end{aligned} \tag{4.50}$$

for $|\hat{\beta}_n| \le \delta_n$. By Theorem 2.1 and (4.49), we have

$$K_n(\hat{\beta}_n) = \sum_{i=1}^n \psi(e_i)x_{in} + O_p\left(\sqrt{\tau_n(\delta_n)}\log n + \delta_n\sqrt{\sum_{i=1}^n |x_{in}|^4}\right). \tag{4.51}$$

By (4.50) and (4.51), we have

$$\begin{aligned}
&-\sum_{i=1}^{n}\varphi\bigl(-x_{in}^{T}\hat{\beta}_n\bigr)x_{in}+O_p(r_n)\\
&\quad=\sum_{i=1}^{n}\psi(e_i)x_{in}+O_p\left(\sqrt{\tau_n(\delta_n)}\log n+\delta_n\sqrt{\sum_{i=1}^{n}|x_{in}|^4}\right).
\end{aligned}\tag{4.52}$$

By (4.52), $\varphi(t)=t\varphi'(0)+O(t^2)$ as $t\to 0$, and $\sum_{i=1}^{n}x_{in}x_{in}^{T}=I_p$, we have

$$\begin{aligned}
&-\sum_{i=1}^{n}\bigl\{\bigl(-x_{in}^{T}\hat{\beta}_n\bigr)\varphi'(0)+O\bigl(\bigl(-x_{in}^{T}\hat{\beta}_n\bigr)^2\bigr)\bigr\}x_{in}-\sum_{i=1}^{n}\psi(e_i)x_{in}\\
&\quad=O_p\left(\sqrt{\tau_n(\delta_n)}\log n+\delta_n\sqrt{\sum_{i=1}^{n}|x_{in}|^4}\right)-O_p(r_n)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{n}x_{in}x_{in}^{T}\varphi'(0)\hat{\beta}_n-\sum_{i=1}^{n}\psi(e_i)x_{in}\\
&\quad=-\sum_{i=1}^{n}O\bigl(\bigl(-x_{in}^{T}\hat{\beta}_n\bigr)^2\bigr)x_{in}+O_p\left(\sqrt{\tau_n(\delta_n)}\log n+\delta_n\sqrt{\sum_{i=1}^{n}|x_{in}|^4}\right)-O_p(r_n).
\end{aligned}$$

Namely

$$\begin{aligned}
&\varphi'(0)\hat{\beta}_n-\sum_{i=1}^{n}\psi(e_i)x_{in}\\
&\quad=O_p\bigl(\sqrt{\tau_n(\delta_n)}\log n\bigr)+O_p\left(\delta_n\sqrt{\sum_{i=1}^{n}|x_{in}|^4}+\sum_{i=1}^{n}\bigl(-x_{in}^{T}\hat{\beta}_n\bigr)^2x_{in}+r_n\right)\\
&\quad=O_p\bigl(\sqrt{\tau_n(\delta_n)}\log n\bigr)+O_p\left(\delta_n\sqrt{\sum_{i=1}^{n}|x_{in}|^4}+|\hat{\beta}_n|^2\sum_{i=1}^{n}|x_{in}|^3+r_n\right)\\
&\quad=O_p\bigl(\sqrt{\tau_n(\delta_n)}\log n\bigr)+O_p\bigl(\delta_n r_n+\delta_n^2 r_n+r_n\bigr)\\
&\quad=O_p\bigl(\sqrt{\tau_n(\delta_n)}\log n+\delta_n^2 r_n\bigr).
\end{aligned}\tag{4.53}$$

By $m(t)=O(|t|^{\lambda})$ $(t\to 0)$ for some $\lambda>0$, we have

$$\tau_n(\delta_n)=2\sum_{i=1}^{n}|x_{in}|^2\bigl(|x_{in}|\delta_n\bigr)^{2\lambda}=2\delta_n^{2\lambda}\sum_{i=1}^{n}|x_{in}|^{2+2\lambda}.\tag{4.54}$$

Then it follows from (4.53) and (4.54) that

$$\varphi'(0)\hat{\beta}_n-\sum_{i=1}^{n}\psi(e_i)x_{in}$$

$$
\begin{aligned}
&= O_p\big(\sqrt{\tau_n(\delta_n)}\log n + \delta_n^2 r_n\big) \\
&= O_p\left(\sqrt{\sum_{i=1}^n |x_{in}|^{2+2\lambda}\delta_n^\lambda \log n + \delta_n^2 r_n}\right)
\end{aligned}
\tag{4.55}
$$

for any $\delta_n \to \infty$, which implies

$$
\varphi'(0)\hat{\beta}_n - \sum_{i=1}^n \psi(e_i)x_{in} = O_p\left(\sqrt{\sum_{i=1}^n |x_{in}|^{2+2\lambda}} \log n + r_n\right). \tag{4.56}
$$

□

Proof of Theorem 2.2 By Lemma 4.7, we have Theorem 2.2. □

Proof of Corollary 2.2 (1) By Lemma 4.7, we have

$$
\sup_{|\beta_n|\le b_n} \big|\tilde{K}_n(\beta_n)\big| \le \sup_{|\beta_n|\le b_n} \big|\tilde{K}_n(\beta_n) - \tilde{K}_n(0)\big| + \tilde{K}_n(0) = O_{a.s.}(L_{\tilde{n}} + B_{\tilde{n}} + h_n), \tag{4.57}
$$

where $b_n = n^{-1/2}(\log n)^{3/2}(\log\log n)^{1/2+\upsilon}$. Let

$$
\Theta_n(\beta) = \sum_{i=1}^n \big[\rho\big(e_i - x_i^T\beta\big) - \rho(e_i)\big] \tag{4.58}
$$

and

$$
A_n(\beta) = -\sum_{i=1}^n \int_0^1 \varphi\big(-x_i^T\beta\big)x_i^T\beta\,dt. \tag{4.59}
$$

Note that

$$
\rho(e_i) - \rho\big(e_i - x_i^T\beta\big) = \int_0^1 \psi\big(e_i - x_i^T\beta\big)x_i^T\beta\,dt. \tag{4.60}
$$

By (4.57)–(4.60), we have

$$
\begin{aligned}
\sup_{|\beta_n|\le b_n} \big|\Theta_n(\beta) - A_n(\beta)\big| &= \sup_{|\beta_n|\le b_n} \left|\sum_{i=1}^n \int_0^1 \big[\psi\big(e_i - x_i^T\beta\big) - \varphi\big(-x_i^T\beta\big)\big]x_i^T\beta\,dt\right| \\
&= \sup_{|\beta_n|\le b_n} \left|\int_0^1 \tilde{K}_n(\beta t)\beta\,dt\right| \\
&= O_{a.s.}\big((L_{\tilde{n}} + B_{\tilde{n}} + h_n)b_n\big).
\end{aligned}
\tag{4.61}
$$

It is easy to show that $b_n^3\sum_{i=1}^n |x_i|^3 = O(n\tilde{r}_n)b_n^3 = o(nb_n^2)$. By $\varphi(t) = t\varphi'(0) + O(t^2)$, we have

$$
\begin{aligned}
\inf_{|\beta_n|=b_n} A_n(\beta) &= \inf_{|\beta_n|=b_n} \left\{-\sum_{i=1}^n \int_0^1 \varphi\big(-x_i^T\beta\big)x_i^T\beta\,dt\right\} \\
&= \inf_{|\beta_n|=b_n} \left\{-\sum_{i=1}^n \int_0^1 \big[\varphi(0) + \varphi'(0)\big(-x_i^T\beta\big) + O\big(\big(-x_i^T\beta\big)^2\big)\big]x_i^T\beta\,dt\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \inf_{|\beta_n|=b_n}\left\{\sum_{i=1}^{n}\left[\frac{1}{2}\varphi'(0)\big(-x_i^T\beta\big)^2|_0^1 - \frac{1}{3}O\big(\big(-x_i^T\beta\big)^3\big)|_0^1\right]\right\} \\
&= \frac{1}{2}\varphi'(0)\sum_{i=1}^{n}x_i^Tx_ib_n^2 - \frac{1}{3}\inf_{|\beta_n|=b_n}\left\{b_n^2\sum_{i=1}^{n}O\big(x_i^Tx_i\big)x_i^T\beta\right\} \\
&\ge \frac{1}{2}\varphi'(0)S_nb_n^2 - \frac{1}{3}b_n^2\sum_{i=1}^{n}|x_i|^3b_nO(1) \\
&\ge \frac{1}{6}\varphi'(0)nb_n^2\liminf_{n\to\infty}\lambda_n/n. \qquad (4.62)
\end{aligned}
$$

By $m(t) = O(\sqrt{n})$ as $t \to 0$, we have $(L_{\tilde{n}} + B_{\tilde{n}} + h_n)b_n = o(nb_n^2)$. Thus

$$
\begin{aligned}
\inf_{|\beta_n|=b_n}\Theta_n(\beta) &\ge \inf_{|\beta_n|=b_n}A_n(\beta) - \sup_{|\beta_n|\le b_n}\big|\Theta_n(\beta) - A_n(\beta)\big| \\
&\ge \frac{1}{6}\varphi'(0)nb_n^2\liminf_{n\to\infty}\lambda_n/n + O_{a.s.}\big((L_{\tilde{n}} + B_{\tilde{n}} + h_n)b_n\big) \\
&\ge \frac{1}{4}\varphi'(0)nb_n^2\liminf_{n\to\infty}\lambda_n/n, \quad \text{a.s.} \qquad (4.63)
\end{aligned}
$$

By the convexity of the function $\Theta_n(\cdot)$, we have

$$
\begin{aligned}
&\left\{\inf_{|\beta_n|\ge b_n}\Theta_n(\beta) \ge \frac{1}{4}\varphi'(0)nb_n^2\liminf_{n\to\infty}\lambda_n/n\right\} \\
&\quad = \left\{\inf_{|\beta_n|=b_n}\Theta_n(\beta) \ge \frac{1}{4}\varphi'(0)nb_n^2\liminf_{n\to\infty}\lambda_n/n\right\}. \qquad (4.64)
\end{aligned}
$$

Therefore the minimizer $\hat{\beta}_n$ satisfies $\hat{\beta}_n = O_{a.s.}(b_n)$.

(2) Let $|\hat{\beta}_n| \le b_n$. By a Taylor expansion, we have

$$
\begin{aligned}
-\sum_{i=1}^{n}\varphi\big(-x_i^T\beta\big)x_i &= \sum_{i=1}^{n}\big[\varphi'(0)x_i^T\beta + O\big(\big|x_i^T\beta\big|^2\big)\big]x_i \\
&= \varphi'(0)\Sigma_n\beta + O\left(b_n^2\sum_{i=1}^{n}|x_i|^3\right). \qquad (4.65)
\end{aligned}
$$

Therefore (2) follows from Theorem 2.2 and (1). □

Acknowledgements
The author's work was supported by the National Natural Science Foundation of China (No. 11471105, 11471223), and the Natural Science Foundation of Hubei Province (No. 2016CFB526).

Competing interests
The author declares to have no competing interests.

Authors' contributions
The author organized and wrote this paper. Further he examined all the steps of the proofs in this paper. The author read and approved the final manuscript.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 July 2017 Accepted: 11 May 2018 Published online: 22 May 2018

References

1. Babu, G.J.: Strong representations for LAD estimators in linear models. Probab. Theory Relat. Fields **83**, 547–558 (1989)
2. Bai, Z.D., Rao, C.R., Wu, Y.: M-estimation of multivariate linear regression parameters under a convex discrepancy function. Stat. Sin. **2**, 237–254 (1992)
3. Bardet, J., Doukhan, P., Lang, G., Ragache, N.: Dependent Lindeberg central limit theorem and some applications. ESAIM Probab. Stat. **12**, 154–172 (2008)
4. Berlinet, A., Liese, F., Vaida, I.: Necessary and sufficient conditions for consistency of M-estimates in regression models with general errors. J. Stat. Plan. Inference **89**, 243–267 (2000)
5. Boente, G., Fraiman, R.: Robust nonparametric regression estimation for dependent observations. Ann. Stat. **17**(3), 1242–1256 (1989)
6. Chen, J., Li, D.G., Zhang, L.X.: Bahadur representation of nonparametric M-estimators for spatial processes. Acta Math. Sin. Engl. Ser. **24**(11), 1871–1882 (2008)
7. Chen, X.: Linear representation of parametric M-estimators in linear models. Sci. China Ser. A **23**(12), 1264–1275 (1993)
8. Chen, X., Zhao, L.: M-methods in Linear Model. Shanghai Scientific & Technical Publishers, Shanghai (1996)
9. Cheng, C.L., Van Ness, J.W.: Generalized m-estimators for errors-in-variables regression. Ann. Stat. **20**(1), 385–397 (1992)
10. Dedecker, J., Doukhan, P.: A new covariance inequality and applications. Stoch. Process. Appl. **106**, 63–80 (2003)
11. Dedecker, J., Doukhan, P., Lang, G., Leon, J.R., Louhichi, S., Prieur, C.: Weak Dependence: With Examples and Applications. Springer, New York (2007)
12. Dedecker, J., Prieur, C.: New dependence coefficients, examples and applications to statistics. Probab. Theory Relat. Fields **132**, 203–236 (2005)
13. Doukhan, P., Klesov, O., Lang, G.: Rates of convergence in some SLLN under weak dependence conditions. Acta Sci. Math. (Szeged) **76**, 683–695 (2010)
14. Doukhan, P., Louhichi, S.: A new weak dependence condition and applications to moment inequalities. Stoch. Process. Appl. **84**, 313–342 (1999)
15. Doukhan, P., Mayo, N., Truquet, L.: Weak dependence, models and some applications. Metrika **69**, 199–225 (2009)
16. Doukhan, P., Neumann, M.H.: Probability and moment inequalities for sums of weakly dependent random variables with applications. Stoch. Process. Appl. **117**, 878–903 (2007)
17. Doukhan, P., Wintenberger, O.: An invariance principle for weakly dependent stationary general models. Probab. Math. Stat. **27**(1) (2007)
18. Doukhan, P., Wintenberger, O.: Weakly dependent chains with infinite memory. Stoch. Process. Appl. **118**, 1997–2013 (2008)
19. Fan, J.: Moderate deviations for M-estimators in linear models with ϕ-mixing errors. Acta Math. Sin. Engl. Ser. **28**(6), 1275–1294 (2012)
20. Fan, J., Yan, A., Xiu, N.: Asymptotic properties for M-estimators in linear models with dependent random errors. J. Stat. Plan. Inference **148**, 49–66 (2014)
21. Freedman, D.A.: On tail probabilities for martingales. Ann. Probab. **3**(1), 100–118 (1975)
22. Gannaz, I.: Robust estimation and wavelet thresholding in partially linear models. Stat. Comput. **17**, 293–310 (2007)
23. Gervini, D., Yohai, V.J.: A class of robust and fully efficient regression estimators. Ann. Stat. **30**(2), 583–616 (2002)
24. He, X., Shao, Q.: A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. Ann. Stat. **24**(8), 2608–2630 (1996)
25. Hu, H.C.: QML estimators in linear regression models with functional coefficient autoregressive processes. Math. Probl. Eng. **2010**, Article ID 956907 (2010) https://doi.org/10.1155/2010/956907
26. Hu, H.C.: Asymptotic normality of Huber–Dutter estimators in a linear model with AR(1) processes. J. Stat. Plan. Inference **143**(3), 548–562 (2013)
27. Hu, Y., Ming, R., Yang, W.: Large deviations and moderate deviations for m-negatively associated random variables. Acta Math. Sci. **27B**(4), 886–896 (2007)
28. Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn. John Wiley & Sons, New Jersey (2009)
29. Hwang, E., Shin, D.: Semiparametric estimation for partially linear regression models with ψ-weak dependent errors. J. Korean Stat. Soc. **40**, 411–424 (2011)
30. Koul, H.L.: M-estimators in linear regression models with long range dependent errors. Stat. Probab. Lett. **14**, 153–164 (1992)
31. Lai, T.L.: Asymptotic properties of nonlinear least squares estimates in stochastic regression models. Ann. Stat. **22**(4), 1917–1930 (1994)
32. Lehmann, E.L.: Elements of Large-Sample Theory. Springer, New York (1998)
33. Li, I.: On Koul's minimum distance estimators in the regression models with long memory moving averages. Stoch. Process. Appl. **105**, 257–269 (2003)
34. Liang, H., Jing, B.: Asymptotic normality in partial linear models based on dependent errors. J. Stat. Plan. Inference **139**, 1357–1371 (2009)
35. Lin, Z., Bai, Z.: Probability Inequalities. Science Press, Beijing (2010)
36. Liptser, R.S., Shiryayev, A.N.: Theory of Martingale. Kluwer Academic Publishers, London (1989)
37. Lô, S.N., Ronchetti, E.: Robust and accurate inference for generalized linear models. J. Multivar. Anal. **100**, 2126–2136 (2009)
38. Maller, R.A.: Asymptotics of regressions with stationary and nonstationary residuals. Stoch. Process. Appl. **105**, 33–67 (2003)
39. Nelson, P.I.: A note on strong consistency of least squares estimators in regression models with martingale difference errors. Ann. Stat. **8**(5), 1057–1064 (1980)
40. Nze, P.A., Bühlmann, P., Doukhan, P.: Weak dependence beyond mixing and asymptotics for nonparametric regression. Ann. Stat. **30**(2), 397–430 (2002)
41. Pere, P.: Adjusted estimates and Wald statistics for the AR(1) model with constant. J. Econom. **98**, 335–363 (2000)
42. Rao, C.R., Zhao, L.C.: Linear representation of M-estimates in linear models. Can. J. Stat. **20**(4), 359–368 (1992)

43. Romano, J.P., Wolf, M.: A more general central limit theorem for m-dependent random variables with unbounded m. Stat. Probab. Lett. **47**, 115–124 (2000)
44. Salibian-Barrera, M., Aelst, S.V., Yohai, V.J.: Robust tests for linear regression models based on τ-estimates. Comput. Stat. Data Anal. **93**, 436–455 (2016) https://doi.org/10.1016/j.csda
45. Valdora, M., Yohai, V.J.: Robust estimators for generalized linear models. J. Stat. Plan. Inference **146**, 31–48 (2014)
46. Valk, V.D.: Hilbert space representations of m-dependent processes. Ann. Probab. **21**(3), 1550–1570 (1993)
47. Wu, Q.: Strong consistency of M estimator in linear model for negatively associated samples. J. Syst. Sci. Complex. **19**, 592–600 (2006)
48. Wu, W.B.: Nonlinear system theory: another look at dependence. Proc. Natl. Acad. Sci. USA **102**(40), 14150–14154 (2005)
49. Wu, W.B.: M-estimation of linear models with dependent errors. Ann. Stat. **35**(2), 495–521 (2007)
50. Xiong, S., Joseph, V.R.: Regression with outlier shrinkage. J. Stat. Plan. Inference **143**, 1988–2001 (2013)
51. Yang, Y.: Asymptotics of M-estimation in non-linear regression. Acta Math. Sin. Engl. Ser. **20**(4), 749–760 (2004)
52. Zhou, Z., Shao, X.: Inference for linear models with dependent errors. J. R. Stat. Soc. Ser. B **75**(2), 323–343 (2013)
53. Zhou, Z., Wu, W.B.: On linear models with long memory and heavy-tailed errors. J. Multivar. Anal. **102**, 349–362 (2011)