M-ESTIMATION OF LINEAR MODELS WITH DEPENDENT ERRORS

Wei Biao Wu

TECHNICAL REPORT NO. 551

October 25, 2004

5734 S. University Avenue
Chicago, IL 60637
\textbf{M-ESTIMATION OF LINEAR MODELS WITH DEPENDENT ERRORS}

\textbf{BY WEI BIAO WU}

October 25, 2004

\textit{University of Chicago}

\textit{Abstract:} We study the asymptotic behavior of \(M\)-estimates of regression parameters in multiple linear models where errors are dependent random variables. A Bahadur representation of the \(M\)-estimates is derived and a central limit theorem is established. The results are applied to linear models with errors being short-range dependent linear processes, heavy-tailed linear processes and some widely used nonlinear time series.

\section{Introduction}

Consider the linear model
\[ y_i = x_i' \beta + e_i, \quad 1 \leq i \leq n, \] (1)

where \( \beta \) is a \( p \times 1 \) unknown regression coefficient vector, \( x_i = (x_{i1}, \ldots, x_{ip})' \) are \( p \times 1 \) known (non-stochastic) design vectors and \( e_i \) are errors. We estimate the unknown parameter vector \( \beta \) by minimizing
\[ \sum_{i=1}^{n} \Phi(y_i - x_i' \hat{\beta}), \] (2)

where \( \Phi \) is a convex function. Important examples include Huber’s estimators with \( \Phi(x) = \min(c|x| - c^2/2, x^2/2) \), \( c > 0 \), \( L^q \) regression estimators with \( \Phi(x) = |x|^q \), \( 1 \leq q \leq 2 \), and regression quantiles with \( \Phi(x) = \max(x, 0) + (1 - \rho)\max(-x, 0) \), \( 0 < \rho < 1 \). In particular, if \( q = 1 \) or \( \rho = 1/2 \), then the minimizer of (2) is called the least absolute deviation estimator. See Zeckhauser and Thompson (1970) and Arcones (1996) for \( L^q \) regression estimators and Koenker and Bassett (1978) for regression quantiles.

Let \( \hat{\beta} \) be the minimizer of (2) and \( \beta_0 \) the true parameter. There is a substantial amount of work concerning asymptotic properties of \( \hat{\beta} - \beta_0 \) for various forms of \( \Phi \) (not necessarily convex); see for example Yohai (1974), Bassett and Koenker (1978), Huber
(1981), Bloomfield and Steiger (1983), Ronner (1984), Welsh (1986), Babu (1989), Chen, Bai, Zhao and Y. Wu (1990), Bai, Rao and Y. Wu (1992), Jurečková and Sen (1996), He and Shao (1996), Arcones (1996) and Zhao (2000) among others. Deep results such as Bahadur representations have also been obtained. However, in majority of the previous work it is assumed that the errors $e_i$ are independent. The asymptotic problem of $M$-estimation of linear models with dependent errors is practically important, however, theoretically challenging. Huber (1973, 1981) commented that the assumption of independence is a serious restriction.

In this paper we shall relax the independence assumption in the classical $M$-estimation theory so that a very general class of dependent errors is allowed. Specifically, we shall establish a Bahadur representation and a central limit theorem for $\hat{\theta}_n - \theta_0$ for the multiple linear model (1) with the errors $(e_i)$ being short-range dependent (SRD) stationary causal processes [cf. (3) and Condition (A6)]. In the early literature very restrictive assumptions have been imposed on the error process $(e_i)$. Typical examples are strongly mixing processes of various types. See Koul (1977), Deniau, Oppenheim and Viano (1977), Phillips (1991) and Cui, He and Ng (2004) among others for strong ($\varphi$) mixing processes and Prakasa Rao (1981) for $\varphi$-mixing processes. Berlinet, Liese and Vajda (2000) obtained consistency of $M$-estimators for regression models with errors being strong mixing processes. Gastwirth and Rubin (1975) considered the behavior of $L$-estimators of strong mixing Gaussian processes and the first order autoregressive process with double exponential marginal. It is generally not easy to verify strong mixing conditions. For example, for linear processes to be strong mixing, very restrictive conditions are needed on the decay rate of the coefficients [Doukhan (1994), Gorodetskii (1977)]. Portnoy (1977, 1979) and Lee and Martin (1986) investigated the effect of dependence on robust location estimators by assuming that the errors are autoregressive moving processes with finite orders.

To the best of our knowledge, it seems that the problem of Bahadur representations has been rarely studied for $M$-estimators of multiple linear models with errors being dependent. The Bahadur-type representations provide significant insight into the asymptotic behavior of an estimator by approximating it by a linear form. The recent work by Wu and Mielniczuk (2002) and Wu (2003a, 2004a, 2004b) shed new light on the asymptotic theory for dependent random variables. In the current paper we shall generalize the method in these
papers and perform a systematic study of the asymptotic behavior of the $M$-estimator $\hat{\theta}_n$ of (1).

For the errors $(e_i)$ we confine ourselves to stationary causal processes. Namely, let

$$e_i = G(\ldots, e_{i-1}, e_i),$$

where $e_k$, $k \in \mathbb{Z}$, are independent and identically distributed (iid) random variables and $G$ is a measurable function such that $e_i$ is a proper random variable. The framework (3) represents a huge class of stationary processes which appear frequently in practice. Let the shift process $\mathcal{F}_k = (\ldots, e_{k-1}, e_k)$; let $F_\varepsilon(u|\mathcal{F}_i) = \mathbb{P}(e_{i+1} \leq u|\mathcal{F}_i)$ [resp. $f_\varepsilon(u|\mathcal{F}_i)$] be the conditional distribution (resp. density) function of $e_{i+1}$ given $\mathcal{F}_i$ and $f$ the marginal density of $e_{i+1}$. For $\tau \in \mathbb{N}$ let $f^{(\tau)}_\varepsilon(u|\mathcal{F}_i) = \partial^\tau f_\varepsilon(u|\mathcal{F}_i)/\partial u^\tau$ be the $\tau$-th order derivative if exists. Our short-range dependence condition is expressed in terms of $f_\varepsilon(u|\mathcal{F}_i)$; see (6). Since the SRD condition is directly related to the data-generating mechanism of $(e_i)$, it is often easily verifiable; see applications in Section 3.

The paper is structured as follows. Section 2 presents the main results on Bahadur representations and central limit theorems for $\hat{\theta}_n$. Proofs are given in Section 4. Section 3 contains applications to linear models with errors being short-range dependent linear processes, heavy-tailed linear processes where $M$-estimation is particularly relevant, and some widely used nonlinear time series.

### 2 Main Results

Without loss of generality, assume throughout the paper that true parameter $\theta_0 = 0$ and the conditional density $f_{\varepsilon}(u|\mathcal{F}_i)$ exists. We first introduce some notation. For a $p$-dimensional vector $v = (v_1, \ldots, v_p)$ let $|v| = (\sum_{i=1}^p v_i^2)^{1/2}$. A random vector $\xi$ is said to be in $L^q$, $q > 0$, if $\mathbb{E}(|\xi|^q) < \infty$. In this case write $\|\xi\|_q = [\mathbb{E}(|\xi|^q)]^{1/q}$ and $\|\xi\| = \|\xi\|_2$. Let the covariance matrix of a $p$-dimensional vector $\xi$ be $\text{var}(\xi) = \mathbb{E}(\xi \xi') - \mathbb{E}(\xi)\mathbb{E}(\xi')$. For $\xi \in L^1$ define projection operators $P_k \xi = \mathbb{E}(\xi|\mathcal{F}_k) - \mathbb{E}(\xi|\mathcal{F}_{k-1})$, $k \in \mathbb{Z}$. Let $w$ $(du) = (1 + |u|) \, du$ be a weighted measure. The notation $C$ denotes a generic constant which may vary from place to place. Let the model matrix $X_n = (x_1, \ldots, x_n)'$ and $S_n = \sum_{i=1}^n x_ix_i' = X_n'X_n$. Assume that $S_n$ is non-singular for sufficiently large $n$. It is convenient to consider the
transferred model

\[ y_i = z_i'\theta + e_i, \quad (4) \]

where \( z_i = z_{i,n} = S_n^{-1/2}x_i \) and \( \theta = \theta_n = S_n^{1/2} \). To study the asymptotic behavior of \( \hat{\theta}_n \), it is equivalent to studying that of \( \hat{\theta}_n = S_n^{1/2} \hat{\theta}_n \), which is a minimizer of \( \sum_{i=1}^n \Phi(e_i - z_i'\theta) \). Observe that \( \sum_{i=1}^n z_i z_i' = Id_p \), a \( p \times p \) identity matrix. For \( q \geq 0 \) define

\[ \Gamma_n(q) = \sum_{i=1}^n |z_i|^q. \quad (5) \]

We make the following assumptions that will be used in the main results Theorems 1 and 2. Let \( \Phi \) be absolutely continuous with derivative \( \phi = \Phi' \).

\( (A1) \) \( \Phi \) is a convex function, \( E[\phi(e_1)] = 0 \) and \( \|\phi(e_1)\|^2 > 0. \)

\( (A2) \) \( \varphi(t) := E[\phi(e_1 + t)] \) has a strictly positive derivative at \( t = 0. \)

\( (A3) \) \( m(t) := \|\phi(e_1 + t) - \phi(e_1)\| \) is continuous at \( t = 0. \)

\( (A4) \) \( \kappa := \int_{\mathbb{R}} \phi^2(t)w_\tau \, (dt) < \infty \) for some \( \tau \geq 0. \)

\( (A5) \) \( r_n := \max_{i \leq n} |z_i| = \max_{i \leq n} [x_i' S_n^{-1} x_i]^{1/2} = o(1). \)

\( (A6) \) (short-range dependence) For all \( \tau = 0, \ldots, p, \)

\[ \sum_{i=0}^{\infty} \sqrt{\int_{\mathbb{R}} \|P_0 f^{(\tau)}_\varepsilon(t|x_i)\|^2 w \, (dt)} < \infty. \quad (6) \]

Conditions \( (A1)-(A3) \) and \( (A5) \) are standard and they are often imposed in the \( M \)-estimation theory of linear models with independent errors; see for example Bai, Rao and Y. Wu (1992). In \( (A1) \), the error process \( (e_i) \) itself is allowed to have infinite variance, which is actually one of the primary reasons for using the \( M \)-estimation technique. Section 3 contains an application to linear processes with stable distributions, which are heavy-tailed and have infinite variances. Condition \( (A2) \) guarantees that \( \theta \) is estimable or separable. Condition \( (A3) \) is very mild. Note that \( \phi \) is nondecreasing and it has countably many discontinuous points. If \( e_i \) has a continuous distribution function and \( \|\phi(e_1 + t_0)\| + \|\phi(e_1 - t_0)\| < \infty \) for some \( t_0 > 0 \), then \( \lim_{t \to 0} \phi(e_1 + t) = \phi(e_1) \) almost surely and \( (A3) \) follows from the Lebesgue dominated convergence theorem. The uniform asymptotic negligibility condition \( (A5) \) is basically the Lindeberg-Feller condition and it states that the diagonal
elements of the hat matrix $X_n S_n^{-1} X_n'$ are uniformly negligible. For the regression model (1) with iid errors $e_i$ having mean 0 and finite variance, (A5) is a necessary and sufficient condition for the least squares estimator $S_n^{-1} X_n'(y_1, \ldots, y_n)'$ to be asymptotic normal; see Huber (1981, Section 7.2) and Gleser (1965). The short-range dependence condition of type (A6) is also adopted in Wu (2003a, 2004a). In a variety of situations it is easily verifiable.

Theorem 1 asserts that the $M$-estimator $\hat{\theta}_n$ is consistent. Theorem 2 presents a local oscillation rate for the $M$-process

$$K_n(\theta) = \sum_{i=1}^n [\phi(e_i - z'_i \theta) - \varphi(-z'_i \theta)]z_i, \quad \theta \in \mathbb{R}^p,$$

which plays an important role in the study of $M$-estimation theory. Welsh (1989) considered the behavior of $M$-processes for linear models with iid errors. Based on the oscillation rate given in Theorem 2, we present in Corollary 1 a Bahadur representation and a central limit theorem for $\hat{\theta}_n$. Corollary 2 concerns the special case of $M$-estimates of location parameters.

If $\phi$ is continuous, then $\hat{\theta}_n$ solves the equation

$$\sum_{i=1}^n \phi(y_i - z'_i \theta)z_i = 0.\quad (8)$$

In the case that $\phi$ is discontinuous, for example, $\phi(x) = d|x|/dx = \text{sgn}(x)$, where $\text{sgn}(0) = 0$, $\text{sgn}(t) = 1$ if $t > 0$ and $-1$ if $t < 0$, (8) may not have a solution. To overcome this difficulty, we propose approximate equations (12) and (14) and the solutions $\hat{\theta}_n$ are said to be approximate $M$-estimators.

**Theorem 1.** Assume (A1)-(A5) and that (6) holds with $\tau = 1$. Let $\hat{\theta}_n$ minimize (2). Then $|\hat{\theta}_n| = O_P(1)$ and $\hat{\theta}_n = o_P(1)$.

It is generally not trivial to establish the consistency of $M$-estimators. The convexity condition is quite useful in proving consistency; see Haberman (1989), Niemiro (1992) and Bai, Rao and Y. Wu (1992) among others for regression models with independent errors. Recently, Berlinet, Liese and Vajda (2000) considered consistency of $M$-estimates in regression models with strong mixing errors. The latter paper requires that the regressors
\( \mathbf{x} \) satisfy the condition that \( n^{-1} \sum_{i=1}^{n} \delta_{\mathbf{x}_i} \) converges to some probability measure, where \( \delta \) is the Dirac measure. This condition is really restrictive and it excludes many interesting cases (cf. Remark 2).

**Theorem 2.** Assume (A1)-(A6). Let \((\delta_n)_{n \in \mathbb{N}}\) be a sequence of positive numbers such that

\[
\delta_n \to \infty \text{ and } \delta_n r_n = \delta_n \max_{i \leq n} |z_i| \to 0. \tag{9}
\]

Then

\[
\sup_{|\theta| \leq \delta_n} |K_n(\theta) - K_n(0)| = O_{\mathbb{P}}[\sqrt{\tau_n(\delta_n)(\log n)^q} + \delta_n \sqrt{\Gamma_n(4)}], \tag{10}
\]

where

\[
\tau_n(\delta) = \sum_{i=1}^{n} |z_i|^2[m^2(|z_i|)\delta + m^2(-|z_i|)\delta], \quad \delta > 0. \tag{11}
\]

**Corollary 1.** Let (A1)-(A6) be satisfied and \( q > 1 \). (i) (Bahadur representation) Assume that \( \varphi(t) = t\varphi'(0) + O(t^2) \) as \( t \to 0 \) and that \( \hat{\theta}_n \) satisfies

\[
\sum_{i=1}^{n} \phi(e_i - z_i'\hat{\theta}_n)z_i = O_{\mathbb{P}}[\sqrt{\tau_n(1)}\log n + r_n]. \tag{12}
\]

Then for any sequence \( c_n \to \infty \),

\[
\varphi'(0)\hat{\theta}_n - \sum_{i=1}^{n} \phi(e_i)z_i = O_{\mathbb{P}}[\sqrt{\tau_n(\delta_n)}(\log n)^q + \delta_n r_n], \tag{13}
\]

where \( \delta_n = \min(c_n, r_n^{-1/2}) \). (ii) (Asymptotic normality). Assume that \( \hat{\theta}_n \) satisfies

\[
\sum_{i=1}^{n} \phi(e_i - z_i'\hat{\theta}_n)z_i = o_{\mathbb{P}}(1), \tag{14}
\]

that \( \sqrt{\tau_n(c)(\log n)^q} \to 0 \) for each \( c > 0 \) and that

\[
\sum_{i=1}^{n-k} z_i z_{i+k} \to \Delta_k \tag{15}
\]

for each \( k \in \mathbb{Z} \). Then

\[
\hat{\theta}_n \Rightarrow N(0, [\varphi'(0)]^{-2}\Delta), \text{ where } \Delta = \sum_{k \in \mathbb{Z}} \mathbb{E}[\phi(e_0)\phi(e_k)]\Delta_k. \tag{16}
\]
Remark 1. If \( \phi \) has finitely many discontinuous points, then the approximate equations (12) and (14) always have solutions since \( r_n = \max_{i \leq n} \left| z_i \right| \rightarrow 0. \)

The representation (13) asserts that \( \hat{\theta}_n \) can be approximated by the linear form \( T_n = \sum_{i=1}^{n} \phi(e_i)z_i \) with a higher-order remainder under suitable conditions on \( z_i \). It is usually easier to deal with \( T_n \) due to its linearity structure (cf. Lemma 1). Bahadur representations are very useful in the study of asymptotic behavior of statistical estimators. In the context of \( M \)-estimation under the assumption of independent errors, various Bahadur representations have been derived; see for example Arcones (1996), Babu (1989), Carroll (1978), Rao and Zhao (1992), He and Shao (1996) and Pollard (1991) among others. In particular, He and Shao (1996) obtained a sharp almost sure bound under very general conditions on \( \Phi \). In the case of multiple linear models with dependent errors, it is not clear how to obtain an almost sure bound of (13). If \( \mathbf{x}_i \equiv 1 \) and \( \Phi(x) = \max(x, 0) + (1 - ) \max(-x, 0) \), then \( \hat{\theta}_n \) corresponds to sample quantiles. Wu (2004b) established Bahadur representations of sample quantiles for dependent random variables with optimal bounds in the sense that they are as sharp as those obtained under the independence assumption.

Remark 2. Many of the earlier results require that \( \mathbf{x}_i, 1 \leq i \leq n, \) satisfy the condition that \( S_n/n \) converges to a positive definite matrix [Bassett and Koenker (1978), He and Shao (1996), Jurečková and Sen (1996) among others]. The latter condition is not required in our setting. Consider the polynomial regression problem with design vectors \( \mathbf{x}_i = (1, i, \ldots, i^{p-1})', 1 \leq i \leq n. \) Then \( S_n/n \) does not have a limit. Elementary but tedious calculations show that (A5) is satisfied and (15) holds with \( \Delta_k = \text{Id}_p. \)

Remark 3. In the expression of \( \Delta \) in (16), the presence of the terms \( \mathbb{E}[\phi(e_0)\phi(e_k)]\Delta_k, k \neq 0, \) is due to the dependence of \( e_i. \)

To apply Theorem 2 and Corollary 1, we need to know the order of magnitude of \( m(\cdot) \); see the definition of \( \tau_n(\delta) \) by (11). Proposition 1 below gives the orders of \( m(\cdot) \) for three commonly used types of \( \Phi \): absolutely continuous functions, power functions \( \Phi(x) = |x|^q, 1 \leq q \leq 2, \) and \( \Phi(x) = \max(x, 0) + (1 - ) \max(-x, 0), 0 < < 1. \) Recall that \( f \) is the density of \( e_i. \)
**Proposition 1.** Assume (9). (i) If \( \phi \) is absolutely continuous with derivative \( \phi' \) such that 
\[
\sup_{|u| \leq \delta} \| \phi'(e_1 + u) \| < \infty \quad \text{for some } \delta > 0, \text{ then } m(t) = O(|t|) \text{ as } t \to 0 \text{ and } \tau_n(\delta_n) = O[\Gamma_n(4)\delta_n^2].
\]
(ii) Let \( \Phi(t) = |t|^q, 1 \leq q \leq 2 \) and assume \( \sup_v f(v) < \infty \). If \( q \neq 3/2 \), then 
\[
m(t) = O(|t|^{q/2}) \text{ and } \tau_n(\delta_n) = O[\Gamma_n(2 + q')\delta_n^q], \quad \text{where } q' = \min(2,2q-1).
\]
If \( q = 3/2 \), then 
\[
m(t) = O(|t|\log(1/|t|)) \quad \text{and} \quad \tau_n(\delta_n) = \sum_{i=1}^n |z_i|^t(\log |z_i|)^2O(\delta_n^2).
\]
(iii) Let \( \Phi(x) = \max(x,0) + (1 - )\max(-x,0), 0 < < 1, \) and assume \( \sup_v f(v) < \infty \). Then 
\[
m(t) = O(|t|^{1/2}) \text{ and } \tau_n(\delta_n) = O[\Gamma_n(3)\delta_n].
\]

**Proof of Proposition 1.** Case (i) is trivial in view of 
\[
\phi(e_1 + t) - \phi(e_1) = \int_0^t \phi'(e_1 + u)du \quad \text{and hence } \quad m^2(t) \leq t \int_0^t \| \phi'(e_1 + u) \|^2du = O(t^2).
\]
For case (ii), the bound of \( m(t) \) follows from Arcones (1996). The bound of \( \tau_n(\delta_n) \) when \( q \neq 3/2 \) can be easily obtained. If 
\[ q = 3/2, \] since \( \tau_n\delta_n \to 0, \] then 
\[
| \log |z_i|\delta_n| \leq 2| \log |z_i|| \quad \text{for sufficiently large } n \text{ and the stated bound for } \tau_n(\delta_n) \text{ follows.}
\]
For (iii), noting that 
\[
\phi(x) = 1_{x>0} - (1 - )1_{x<0} \quad \text{and} \quad \mathbb{E}[1_{e_1+t>0} - 1_{e_1>0}] = f(0)t + o(t) \text{ as } t \to 0, \text{ we have } m^2(t) = f(0)t + o(t).
\]

**Corollary 2.** Let \( p = 1, \ x_i \equiv 1 \). Assume (A1), (A4) and that (6) holds with \( \tau = 0 \) and 1. Further assume \( \varphi(t) = t\varphi'(0) + O(t^2) \) and 
\[
m(t) = O(|t|^\lambda|\log |t||) \quad \text{as } t \to 0, \text{ where } 0 < \lambda \leq 1 \quad \text{and} \quad \in \mathbb{R}, \text{ and that } \hat{\theta}_n \text{ satisfies}
\]
\[
\sum_{i=1}^n \phi(e_i - \hat{\theta}_n/\sqrt{n}) = O_p[n^{(1-\lambda)/2}(\log n)^{1+\zeta}]. \quad (17)
\]

Then for every \( q > 1, \)
\[
\varphi'(0)\hat{\theta}_n - \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(e_i) = O_p[n^{-\lambda/2}(\log n)^{q+\zeta}]. \quad (18)
\]

Corollary 2 easily follows from Corollary 1 since \( r_n = 1/\sqrt{n} \). The following example concerns some common \( M \)-estimates of location parameters.

**Example 1.** (i) (Huber estimates) For \( \Phi(x) = \min(c|x| - c^2/2, x^2/2), c > 0, \) we have 
\[
\phi(x) = \max[\min(x,c), -c] \text{ and } \varphi'(t) = \mathbb{P}(|e_1 + t| \leq c) = F(c - t) - F(-c - t).
\]
If \( \sup_x f(x) < \infty, \) then 
\[
\varphi(t) = t\varphi'(0) + O(t^2) \quad \text{and} \quad m(t) = O(|t|) \text{ as } t \to 0. \quad \text{Suppose (6) holds with } \tau = 0 \text{ and 1 and } \eta > 1. \text{ Let } \hat{\theta}_n \text{ satisfy (17) with } \lambda = 1 \text{ and } \eta = 0. \text{ Then we have (18) with the error bound } O_p[n^{-1/2}(\log n)^{1+\eta}] \text{ for each } \eta > 0.
(ii) \(L^q\)-regression estimates) Let \(\Phi(x) = |x|^q, 1 < q < 2\). Assume that \(\sup_x [f(x) + |f'(x)|] < \infty\) and that (A4) and (6) hold with \(\tau > 2q - 1\) and \(\tau = 0, 1\). Then the bound in (18) is \(O_p[n^{-q'/4}(\log n)^{1+\eta}]\) (resp. \(O_p[n^{-1/2}(\log n)^{2+\eta}]\)) if \(q \neq 3/2\) (resp. \(q = 3/2\)) for each \(\eta > 0\), where \(q' = \min(2, 2q - 1)\). To this end, by Corollary 2 and Proposition 1, it suffices to verify that \(\varphi(t) = t\varphi'(0) + O(t^2)\). Note that \(\varphi(x) = q|x|^{q-1}\text{sgn}(x), \varphi'(x) = q(q-1)|x|^{q-2}\text{sgn}(x)\) and \(\varphi'(x) = E[\varphi'(x + e_1)]\). Let \(|x| \leq 1\) and \(|\delta| \leq 1\). If \(|e_1| > 3\), then \(|\varphi'(x + e_1 + \delta) - \varphi'(x + e_1)| \leq |\delta|\). On the other hand,

\[
E\{[\varphi'(x + e_1 + \delta) - \varphi'(x + e_1)]1_{|e_1| \leq 3}\} = \int_{-3}^{3} \varphi'(x + u)[f(u) - f(u - \delta)]du \\
+ \left[\int_{-3}^{-3+\delta} - \int_{3}^{3+\delta}\right] \varphi'(x + u)f(u - \delta)du,
\]
which is also of the order \(O(\delta)\) since \(\sup_x |f'(x)| < \infty\) and \(\int_{-3}^{3} |\varphi'(u)|du < \infty\). Therefore \(\varphi'(x + \delta) - \varphi'(x) = O(\delta)\) and consequently \(\varphi(t) - \varphi(0) = t\varphi'(0) + O(t^2)\).

(iii) (Regression quantiles) Let \(\Phi(x) = \max(x, 0) + (1 - \lambda)\max(-x, 0), 0 < \lambda < 1\). Assume that \(\sup_x [f(x) + |f'(x)|] < \infty\). Then \(\varphi(t) = -1_{x \leq 0}, \varphi(t) = -F(-t)\) and \(\varphi(t) - \varphi(0) = t\varphi'(0) + O(t^2)\). On the other hand, \(m^2(t) = E(|1_{e_{1+1} \leq 0} - 1_{e_{1+1} \leq 0}|) = O(t)\). So \(\lambda = 1/2\) and the bound in (18) becomes \(O_p[n^{-1/4}(\log n)^{1+\eta}]\) for each \(\eta > 0\).

**Remark 4.** If (6) holds with \(\tau = 0\) and 1, then \(\sup_t f(t) < \infty\). Actually, since \(f(t) = E[f_\epsilon(t)|\mathcal{F}_0]\), it suffices to verify that \(E[\sup_t f_\epsilon^2(t)|\mathcal{F}_0] < \infty\). Since \(\mathcal{P}_i, i \in \mathbb{Z},\) are orthogonal operators and \(\|\mathcal{P}_{-i}f_\epsilon^{(\tau)}(t)|\mathcal{F}_0\| = \|\mathcal{P}_0f_\epsilon^{(\tau)}(t)|\mathcal{F}_0\|\), (6) implies

\[
\int_R \|f_\epsilon^{(\tau)}(t)|\mathcal{F}_0\|w dt = \sum_{i=0}^{\infty} \int_R \|\mathcal{P}_{-i}f_\epsilon^{(\tau)}(t)|\mathcal{F}_0\|w dt < \infty.
\]

Let \(H(t) = f_\epsilon(t)|\mathcal{F}_0\). By the maximal inequality \(\sup_t H^2(t) \leq 2 \int_{\mathbb{R}} H^2(u) + [H'(u)]^2du\) [cf. Lemma 4 in Wu (2003a)], \(E[\sup_t H^2(t)] \leq \infty\) follows.

\[
\Box
\]

### 3 Applications

This section contains applications of Corollary 1 in Section 2 to linear models with errors being (i) heavy-tailed linear processes, (ii) linear processes with finite variances and (iii) some widely used nonlinear time series, which are given in Sections 3.1-3.3 respectively. For such processes the condition (6) can be easily verified.

9
3.1 Linear processes with heavy-tailed innovations.

For the linear model (1) with errors being dependent and heavy-tailed, it is more desirable to apply the $M$-estimation technique to estimate the unknown parameter since the linear squares procedure may result in estimators with erratic behavior. A popular model for such heavy-tailed processes is infinite order moving average linear processes, namely $e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$, where $\varepsilon_i$ are iid random variables with stable distributions and $a_i$ are coefficients such that $e_i$ is a proper random variable. Recently there has been a substantial interest in linear processes with heavy-tailed innovations; see Hsing (1999), Surgailis (2002), Wu (2003b) among others. Davis, Knight and Liu (1992) studied the behavior of $M$-estimator in casual autoregressive models, while Davis and Wu (1997) considered $M$-estimation in linear models. In the latter two papers the errors are assumed to be heavy-tailed, however, independent.

For simplicity in presentation we assume that $\varepsilon_i$ are iid standard symmetric-stable (S S) random variables with index $\alpha \in (1, 2)$. Then the characteristic function of $\varepsilon_i$ is $\mathbb{E}[\exp(\sqrt{-1}u\varepsilon_1)] = \exp(-|u|^{\alpha})$, $u \in \mathbb{R}$. Note that the case $\alpha = 2$ corresponds to Gaussian distributions. Let $f_\varepsilon$ be the density function of $\varepsilon_i$ and recall that $f$ is the density function of $e_i$.

**Proposition 2.** Assume that $\varepsilon_i$ are iid standard S S random variables with index $1 < \alpha < 2$. Let $1 < \gamma < 0$ and $(a_i)_{i=0}^{\infty}$ be a sequence of real numbers such that $e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$ is a proper random variable. Then for every integer $\tau \geq 0$,

$$
\int_{\mathbb{R}} \|P_0 f_\varepsilon^{(\tau)}(t|\mathcal{F}_{n-1})\|^2 w (dt) = O(|a_n|^\gamma).
$$

(19)

Proposition 2 gives an easy way to verify the short-range condition (6). Its proof is presented in Section 5. The following Corollary 3 deals with Huber functions and regression quantiles. Applications to $L^q$ regression estimators can be similarly made.

**Corollary 3.** Assume (A5) and $e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$, where $\varepsilon_i$ be iid standard S S random variables with index $\alpha \in (1, 2)$. Further assume that for some $\gamma_0 \in (1, \gamma)$,

$$
\sum_{j=0}^{\infty} |a_j|^{\alpha/2} < \infty.
$$

(20)
(i) Let $\phi(x) = \min[\max(x,-c),c]$, $c > 0$, $q > 1$ and $\hat{\theta}_n$ satisfy
\[
\sum_{i=1}^{n} \phi(e_i - z_i^0 \hat{\theta}_n)z_i = o_P[r_n(\log n)^q].
\] (21)

Then $\varphi'(0) = P(|e_1| \leq c) > 0$ and
\[
P(|e_1| \leq c)\hat{\theta}_n - \sum_{i=1}^{n} \phi(e_i)z_i = O_P[r_n(\log n)^q].
\] (22)

If in addition $r_n(\log n)^q = o(1)$ and (15) is satisfied, then the central limit theorem (16) holds. (ii) (Regression quantiles) Let $\lambda \in (0,1)$ and $\xi_\lambda$ be the $\lambda$-th quantile of $e_i$; let $\phi(x) = \lambda - 1_{x<\xi_\lambda}$. Then for any $1 < q' < 2$,
\[
f(\xi_\lambda)\hat{\theta}_n - \sum_{i=1}^{n} \phi(e_i)z_i = O_P[\sqrt{r_n}(\log n)^{q'}].
\] (23)

The central limit theorem (16) holds if (15) is satisfied and $\sqrt{r_n}(\log n)^{q'} = o(1)$.

Remark 5. Wu (2004) considers Bahadur representations for sample quantiles of linear processes, which corresponds to the special case of (ii) of Corollary 3 with $p = 1$ and $x_1 = \ldots = x_n = 1$, or $z_1 = \ldots = z_n = 1/\sqrt{n}$. In this case it is easily seen that $\hat{\theta}_n + \xi_\lambda$ is the $\lambda$-th sample quantile of $e_1, \ldots, e_n$. The Bahadur representation (23) gives the bound $O_P[n^{-1/4}(\log n)^{q'}]$, which is slightly less accurate than the bound $O_{a.s.}[n^{-1/4}(\log \log n)^{-3/4}]$ given in Wu (2004b). The latter almost sure bound is optimal and it is obtained under stronger conditions on the coefficients $(a_i)_{i=0}^\infty$. \hfill \diamond

Remark 6. The condition (20) seems almost necessary for the asymptotic normality of $\hat{\theta}_n$. Let $a_n \sim n^{-\gamma}$, $n \in \mathbb{N}$. Then (20) is reduced to $\gamma > 2$. Surgailis (2002) show that, if $\gamma < 2$, then the empirical process of $e_i$ have a non-central limit theorem and the normalizing sequence is no longer $\sqrt{n}$. It is unclear how to obtain the asymptotic distribution of $\hat{\theta}_n$ for the multiple linear model (1) when $\gamma < 2$. \hfill \diamond

Remark 7. Davis et al (1992) pointed out that, for a casual autoregressive model with iid heavy-tailed innovations, the $M$-estimator of the autoregressive coefficients may not be asymptotically normal. \hfill \diamond

11
3.2 Linear processes with finite variances.

Consider the linear model (1) with errors \( e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j} \), where \( \varepsilon_i \) are iid random variables with mean 0 and finite variance and \( a_i \) are square summable. The process \((e_i)\) is said to be short-range dependent if its covariances are absolutely summable and long-range dependent (LRD) if otherwise. It seems that the asymptotic robust estimation problem of (1) with SRD errors has been rarely investigated in the literature. In the LRD case the problem has been studied a lot; see Koul and Surgailis (2000) and references therein.

Corollary 4 imposes very mild conditions on \( z_i \) and on the coefficients \((a_j)\) to ensure a Bahadur representation of \( \hat{\theta}_n \).

**Corollary 4.** Assume \((A5)\), \( \mathbb{E}(\varepsilon_0) = 0 \), \( \mathbb{E}(\varepsilon_0^2) < \infty \) and

\[
\sum_{j=0}^{\infty} |a_j| < \infty. \tag{24}
\]

Further assume that there is a \( \sigma \in (1, 2) \) such that

\[
\sum_{\tau=0}^{p+1} \int_{\mathbb{R}} |f^{(\tau)}(v)|^2 w(\, dv) < \infty. \tag{25}
\]

Then (i) [resp. (ii)] of Corollary 3 holds if \( \mathbb{P}(|e_1| \leq c) > 0 \) [resp. \( f(\xi_\lambda) > 0 \)].

**Remark 8.** The summability condition (24) is a natural condition to ensure the short-range dependence of \((e_i)\) in the case that \( e_i \) has finite variance.

**Remark 9.** It is interesting to note that, in the derivation of asymptotic expansions of empirical processes of long-range dependent sequences, Wu (2003a) also adopted the same condition (25).

3.3 Nonlinear time series.

Many nonlinear time series models assume the form of iterated random functions

\[
e_i = R(e_{i-1}, \varepsilon_i), \tag{26}
\]

where \( R \) is a measurable function and \( \varepsilon_i \) are iid random variables. See Diaconis and Freedman (1999) for a review. Under suitable conditions on \( R(\cdot, \cdot) \), (26) has a stationary and
unique solution. In this case iterations of (26) give (3). Let $F_\varepsilon(u|v) = \mathbb{P}[R(v, \varepsilon_i) \leq u]$ and $f_\varepsilon(u|v) = \partial F_\varepsilon(u|v)/\partial u$ be the transition distribution and density functions respectively. Then $f_\varepsilon(u|\mathcal{F}_i) = f_\varepsilon(u|\varepsilon_i)$. Let $(\varepsilon_i^*)_{i \in \mathbb{Z}}$ be an iid copy of $(\varepsilon_i)_{i \in \mathbb{Z}}$ and, for $i \geq 0$,
\begin{equation}
\varepsilon_i' = G(\ldots, \varepsilon_{i-1}, \varepsilon_i^*, \varepsilon_1, \ldots, \varepsilon_i).
\end{equation}
Wu (2004a) obtained the inequality
\begin{equation}
\int_R \|\mathcal{P_0}f_\varepsilon^{(\tau)}(t|\mathcal{F}_i)\|^2 w (dt) \leq \|\rho(\tau)(\varepsilon_i, \varepsilon_i')\|^2,
\end{equation}
where $\rho(\tau)$ is the weighted distance
\begin{equation}
\rho(\tau)(a, b) = \int_a^b [H(\tau)(v)]^{1/2} dv, \quad H(\tau)(v) = \int_{\mathbb{R}} \frac{\partial}{\partial v} f_\varepsilon^{(\tau)}(u|v)^2 w (du).
\end{equation}
So (6) holds if
\begin{equation}
\sum_{i=0}^{\infty} \|\rho(\tau)(\varepsilon_i, \varepsilon_i')\| < \infty.
\end{equation}
Since the preceding condition only involves the conditional density $f_\varepsilon(u|v)$ rather than the marginal density $f$, it is manageable in many situations. Consider the special model
\begin{equation}
\varepsilon_i = m(e_{i-1}) + \varepsilon_i
\end{equation}
where $m$ is a Lipschitz continuous function such that the Lipschitz constant
\begin{equation}
L_m := \sup_{a \neq b} \frac{|m(a) - m(b)|}{|a - b|} < 1
\end{equation}
and $\mathbb{E}(|\varepsilon_i|) < \infty$ for some $\gamma > 1$. Note that the condition $L_m < 1$ ensures that the nonlinear time series model (29) has a unique stationary distribution. A prominent example of (29) is the threshold autoregressive model $e_{i+1} = 1 \max(e_i, 0) + 2 \max(-e_i, 0) + \varepsilon_{i+1}$, where $1, 2$ are real coefficients [cf. Tong (1990)]. In this example (30) is satisfied if $\max(|1|, |2|) < 1$. If the process (26) is of the form (29), then the condition (6) can be simplified.

**Corollary 5.** Assume that (29) satisfies (30) and $\varepsilon_i \in L$ for some $\gamma > 1$. Further assume (25) and $1 < \nu < \tau < +2$. Then there exists $\chi \in (0, 1)$ such that
\begin{equation}
\int_R \|\mathcal{P_0}f_\varepsilon^{(\tau)}(t|\mathcal{F}_i)\|^2 w (dt) = O(\chi^\gamma).
\end{equation}
Consequently, (i) [resp. (ii)] of Corollary 3 holds if $\mathbb{P}(|e_1| \leq c) > 0$ [resp. $f(\xi_\lambda) > 0$].
Remark 10. In Corollary 5, $\varepsilon_i$ are allowed to have infinite variances. \hfill $\Diamond$

4 Proofs of results in Section 2.

In proving Theorem 1 we apply the method in Bai, Rao and Y. Wu (2002), where the assumption of the convexity of $\Phi$ plays a critical role. Techniques in Wu and Mielniczuk (2002) and Wu (2003a, 2004a, 2004b) are used in the proof of Theorem 2.

Lemma 1. Let $T_n = \sum_{i=1}^{n} \phi(e_i)z_i$. Assume (A4) and (6) with $\tau = 0$. Then $\|T_n\| = O(1)$. If in addition (15) holds, then $T_n \Rightarrow N(0, \Delta)$.

Proof of Lemma 1. For $k \in \mathbb{N}$ let $J_k = \sum_{i=1}^{n} P_{i-k}\phi(e_i)z_i$. Then the summands of $J_k$ form martingale differences and

$$\|J_k\|^2 = \sum_{i=1}^{n} \|P_{i-k}\phi(e_i)z_i\|^2 = \sum_{i=1}^{n} |z_i|^2 \|P_0\phi(e_k)\|^2.$$ 

Note that $\mathbb{E}[\phi(e_j)|\mathcal{F}_{j-1}] = \int_{\mathbb{R}} \phi(u)f_z(u|\mathcal{F}_{j-1})du$ and $P_0\phi(e_k) = P_0\mathbb{E}[\phi(e_k)|\mathcal{F}_{k-1}], k \geq 1$. By Schwarz’s inequality,

$$\|P_0\phi(e_k)\|^2 \leq \int_{\mathbb{R}} \phi(u)P_0f_z(u|\mathcal{F}_{k-1})du \leq \kappa \int_{\mathbb{R}} \|P_0f_z(u|\mathcal{F}_{k-1})\|^2 w(du),$$

which by (6) implies that $\sum_{k=0}^{\infty} \|P_0\phi(e_k)\| < \infty$. So $\|T_n\| = O(1)$ since $T_n = \sum_{k=1}^{\infty} J_k$. That $T_n \Rightarrow N(0, \Delta)$ follows from Theorem 1 in Hannan (1973) in view of the Cramer-Wold device. \hfill $\Diamond$

Lemma 2. Let $(n_i)_{i=1}^{n}$, $n \in \mathbb{N}$, be a triangular array of real numbers such that $\sum_{i=1}^{n} \frac{2}{n_i} \leq 1$ and $\rho_n := \max_{i \leq n} |n_i| \to 0$. Assume (A1), (A3), (A4) and (6) with $\tau = 1$. Let $\eta_{n,i} = \Phi(e_i - n_i) - \Phi(e_i) + n_i \phi(e_i)$. Then $\text{var}(\sum_{i=1}^{n} \eta_{n,i}) \to 0$.

Proof of Lemma 2. By (A1), $|\eta_{n,i}| \leq |n_i||\phi(e_i - n_i) - \phi(e_i)|$. Since $\rho_n \to 0$, by (A3),

$$\sum_{i=1}^{n} [\eta_{n,i} - \mathbb{E}(\eta_{n,i}|\mathcal{F}_{i-1})]^2 \leq \sum_{i=1}^{n} \|\eta_{n,i} - \mathbb{E}(\eta_{n,i}|\mathcal{F}_{i-1})\|^2 \leq \sum_{i=1}^{n} \|\eta_{n,i}\|^2 \leq \sum_{i=1}^{n} \|n_i[\phi(e_i - n_i) - \phi(e_i)]\|^2.$$
\[
\leq \sum_{i=1}^{n} \frac{2}{ni} \sup_{|t| \leq \rho_n} m_i(t) \leq \sup_{|t| \leq \rho_n} m_i(t) \to 0.
\]

It remains to show that \( \text{var} \left[ \sum_{i=1}^{n} \mathbb{E}(\eta_{n,i} | \mathcal{F}_{i-1}) \right] \to 0 \). Since \( \eta_{n,i} = \int_{0}^{t} n \phi(e_i) - \phi(e_i - u) du \),

\[
\mathbb{E}(\eta_{n,i} | \mathcal{F}_{i-1}) = \int_{0}^{n} \mathbb{E}(\phi(e_i) - \phi(e_i - u) | \mathcal{F}_{i-1}) du
\]

\[
= \int_{\mathbb{R}} \phi(x) \left[ \int_{0}^{n} f_\varepsilon(x + u | \mathcal{F}_{i-1}) - f_\varepsilon(x | \mathcal{F}_{i-1}) du \right] dx
\]

\[
= \int_{\mathbb{R}} \phi(x) \left[ \int_{0}^{1} (1 - t) \frac{2}{ni} f'_\varepsilon(x + nt | \mathcal{F}_{i-1}) dt \right] dx.
\]

For \( k \in \mathbb{N} \) let \( \mu_k = \int_{\mathbb{R}} \| P_0 f'_\varepsilon(v | \mathcal{F}_{k-1}) \|^2 w (dv) \), \( \Lambda = \sum_{k=1}^{\infty} \sqrt{\mu_k} \) and

\[
I_k(x,t) = I_{n,k}(x,t) = \sum_{i=1}^{n} \frac{2}{ni} P_{i-k} f'_\varepsilon(x + nt | \mathcal{F}_{i-1}).
\]

Observe that the summands of \( I_k \) form (triangular array) martingale differences. Hence for \( 0 \leq t \leq 1 \), by the inequality \( 1 + |v - u| \leq (1 + |v|)(1 + |u|) \),

\[
\int_{\mathbb{R}} \| I_k(x,t) \|^2 w (dx) = \sum_{i=1}^{n} \frac{4}{ni} \int_{\mathbb{R}} \| P_{i-k} f'_\varepsilon(x + nt | \mathcal{F}_{i-1}) \|^2 w (dx)
\]

\[
= \sum_{i=1}^{n} \frac{4}{ni} \int_{\mathbb{R}} \| P_{i-k} f'_\varepsilon(v | \mathcal{F}_{i-1}) \|^2 (1 + |v - nt|) dv
\]

\[
\leq \sum_{i=1}^{n} \frac{4}{ni} (1 + |nt|) \mu_k = O(\rho_n^2 \mu_k) \sum_{i=1}^{n} \frac{2}{ni} = O(\rho_n^2 \mu_k)
\]

holds since \( \rho_n \to 0 \). Let

\[
H(x,t) = H_n(x,t) = \sum_{i=1}^{n} \frac{2}{ni} [ f'_\varepsilon(x + nt | \mathcal{F}_{i-1}) - f'(x + nt)].
\]

Then \( H(x,t) = \sum_{k=1}^{\infty} I_k(x,t) \) and by Schwarz’s inequality, for \( 0 \leq t \leq 1 \),

\[
|H(x,t)|^2 \leq \Lambda \sum_{k=1}^{\infty} \frac{1}{\sqrt{\mu_k}} |I_k(x,t)|^2.
\]

Consequently

\[
\int_{\mathbb{R}} \| H(x,t) \|^2 w (dx) \leq \Lambda \sum_{k=1}^{\infty} \frac{1}{\sqrt{\mu_k}} \int_{\mathbb{R}} \| I_k(x,t) \|^2 w (dx) = O(\rho_n^2).
\]
Therefore, we have
\[
\sum_{i=1}^{n} \left[ \mathbb{E}(\eta_{ni}|\mathcal{F}_{i-1}) - \mathbb{E}(\eta_{ni}) \right]^2 = \int_{\mathbb{R}} \phi(x) \left[ \int_{0}^{1} H(x,t)(1-t)dt \right] dx \\
\leq \kappa \int_{\mathbb{R}} \left[ \int_{0}^{1} H(x,t)(1-t) dt \right]^2 w(dx) \\
\leq \kappa \int_{\mathbb{R}} \left[ \int_{0}^{1} \|H(x,t)\| dt \right]^2 w(dx) \\
\leq \kappa \int_{\mathbb{R}} \int_{0}^{1} \|H(x,t)\|^2 dt w(dx) = 2\kappa \ O(\rho_n^2),
\]
completing the proof since \(\rho_n \to 0\). \(\diamondsuit\)

**Proposition 3.** Assume (A1)-(A5) and (6) with \(\tau = 1\). Then for any \(c > 0\),
\[
D_n(c) := \sup_{|\theta| \leq c} \sum_{i=1}^{n} \left[ \Phi(e_i - z_i'\theta) - \Phi(e_i) + z_i'\theta\phi(e_i) \right] - \frac{\varphi'(0)}{2} |\theta|^2 \to 0 \tag{32}
\]
in probability.

Proposition 3 is a generalization of Theorem 1 of Bai, Rao and Y. Wu (1992) which deals with iid errors.

**Proof of Proposition 3.** We should use the argument in Bai, Rao and Y. Wu (1992). Let \(\eta_i(\theta) = \Phi(e_i - z_i'\theta) - \Phi(e_i) + z_i'\theta\phi(e_i)\). For a fixed vector \(\theta\) with \(|\theta| \leq c\), let \(n_i = z_i'\theta\). Then \(\sum_{i=1}^{n} \frac{2}{n_i} \leq c^2\) and \(\max_{i \leq n} |n_i| \leq cr_n \to 0\). Applying Lemma 2 with \(n_i = z_i'\theta\), we have \(\text{var} [\sum_{i=1}^{n} \eta_i(\theta)] \to 0\). By Lemma 1 in Bai et al (1992), under (A1) and (A2), the bias
\[
\sum_{i=1}^{n} \mathbb{E}[\eta_i(\theta)] - \frac{\varphi'(0)}{2} |\theta|^2 = \sum_{i=1}^{n} \left[ \frac{\varphi'(0)}{2} |z_i'\theta|^2 + o(|z_i'\theta|^2) \right] - \frac{\varphi'(0)}{2} |\theta|^2 \\
= \frac{\varphi'(0)}{2} \sum_{i=1}^{n} [\theta'z_i\theta + o(|z_i'\theta|^2)] - \frac{\varphi'(0)}{2} |\theta|^2 = o(1)
\]
since \(\sum_{i=1}^{n} z_i z_i' = \text{Id}_p\). So (32) holds pointwisely. Since \(\eta_i(\theta)\), \(1 \leq i \leq n\), are convex functions of \(\theta\), the uniform convergence follows from Theorem 10.8 of Rockafellar (1970, p.90). See Appendix II in Andersen and Gill (1982) and Bai et al (1992) for more details. \(\diamondsuit\)

**Proof of Theorem 1.** By Proposition 3, that \(|\hat{\theta}_n| = O_{\mathbb{P}}(1)\) easily follows from properties of convex functions; see for example the proof of Theorem 2.2 in Bai et al (1992). For
the sake of completeness, we present that standard argument here. It suffices to show that \( \mathbb{P}(|\hat{\theta}_n| \leq c_n) \to 1 \) holds for any sequence \( c_n \to \infty \). By Proposition 3, it is easily seen that there exists a sequence \( c'_n \to \infty \) such that (32) holds along this sequence. Let

\[
\inf_{|\theta| = n} V_n(\theta) \geq \frac{\varphi'(0)}{2} \frac{2}{n} - n \sum_{i=1}^{n} z'_i \phi(e_i) - D_n(\theta) \]

which by Lemma 1 and the convexity of \( \Phi \) and \( V_n \) implies that

\[
\mathbb{P} \left\{ \inf_{|\theta| = n} V_n(\theta) \geq \frac{\varphi'(0)}{3} \frac{2}{n} \right\} = \mathbb{P} \left\{ \inf_{|\theta| \geq n} V_n(\theta) \geq \frac{\varphi'(0)}{3} \frac{2}{n} \right\} \to 1
\]

as \( n \to \infty \). So \( \mathbb{P}(|\hat{\theta}_n| \leq c_n) \geq \mathbb{P}(|\hat{\theta}_n| \leq n) \to 1 \).

Now we show \( n = o_P(1) \). Let \( S_n \) have the eigen-decomposition \( S_n = Q_n \Lambda_n Q_n' \), where \( Q_n \) is an orthonormal matrix and \( \Lambda_n = \text{Diag}[\lambda_{1,n}, \ldots, \lambda_{p,n}] \) is a diagonal matrix with \( \lambda_{i,n}, 1 \leq i \leq p \) being eigenvalues. Assume without loss of generality that the matrix \( P = (x_1, \ldots, x_p) \) has full rank. By (A5), \( P'S_n^{-1}P \to 0 \), hence \( PP'S_n^{-1}PP' \to 0 \). Since \( PP' \) is positive definite, we have \( S_n^{-1} \to 0 \), which implies that \( \min_{i \leq p} \lambda_{i,n} \to \infty \). If errors are independent, the last condition is necessary and sufficient for least squares estimators to be consistent [Drygas (1976), Lai, Robbins and Wei (1978)]. So we have \( \hat{\theta}_n = S_n^{1/2} \hat{\theta}_n = O_P(1) \) in view of \( \hat{\theta}_n = S_n^{1/2} \hat{\theta}_n = O_P(1) \).

\[ \hat{\theta}_n = S_n^{1/2} \hat{\theta}_n = O_P(1) \]

\section{Proof of Theorem 2.} The idea of the proof is to write \( K_n = M_n + N_n \), where

\[
M_n(\theta) = \sum_{i=1}^{n} \left\{ \phi(e_i - z'_i \theta) - \mathbb{E}[\phi(e_i - z'_i \theta) | \mathcal{F}_{i-1}] \right\} z_i
\]

and

\[
N_n(\theta) = \sum_{i=1}^{n} \left\{ \mathbb{E}[\phi(e_i - z'_i \theta) | \mathcal{F}_{i-1}] - \varphi(-z'_i \theta) \right\} z_i.
\]

Note that the summands of \( M_n \) form (triangular array) martingale differences with respect to the filter \( \sigma(\mathcal{F}_i) \). Martingale theory is then applicable. Since \( \mathbb{E}[\phi(e_i - t) | \mathcal{F}_{i-1}] = \int_\mathbb{R} \phi(v)f_z(t + v | \mathcal{F}_{i-1})dv \), \( N_n \) has the useful representation

\[
N_n(\theta) = \int_\mathbb{R} \phi(v) \sum_{i=1}^{n} \left[ f_z(v + z'_i \theta | \mathcal{F}_{i-1}) - f(v + z'_i \theta) \right] z_i dv,
\]

17
which considerably facilitates its analysis. Lemmas 3 and 4 deal with $M_n$ and $N_n$ respectively. Since $\sum_{i=1}^n z_i z_i' = \text{Id}_p$, we have $n^{-1/2} = O(r_n)$ and by (9), $n^{-2} = O(r_n^4) = O[\delta_n \Gamma_n(4)]$. So (10) follows from Lemmas 3 and 4 since the term $n^{-2}$ is absorbed into $O[\delta_n \Gamma_n(4)]$. \(\diamondsuit\)

**Lemma 3.** Assume (A4), (A5), (9) and

\[
\int_{\mathbb{R}} \|f'_e(v|\mathcal{F}_0)\|^2 w(\,dv\,) < \infty. \tag{36}
\]

Then for any $q > 1$,

\[
\sup_{|\theta| \leq \delta_n} |M_n(\theta) - M_n(0)| = o_P[\sqrt{\tau_n(\delta_n)(\log n)^q + n^{-2}}]. \tag{37}
\]

**Remark 11.** Since $\|f'_e(v|\mathcal{F}_0)\|^2 = \sum_{k=0}^\infty \|P_{-k} f'_e(v|\mathcal{F}_0)\|^2$, it is easily seen that (36) follows from (6) with $\tau = 1$. \(\diamondsuit\)

**Proof of Lemma 3.** Since $p = \sum_{i=1}^n z_i z_i' \leq nr_n^2$, (9) implies that $\delta_n = o(\sqrt{n})$. Let $b_n = [(\log n)^{-q-1}\tau_n(\delta_n)]^{1/2}$, $r_n = \sqrt{\tau_n(\delta_n)(\log n)^q}$, $\eta_i(\theta) = [\phi(e_i - z_i'\theta) - \phi(e_i)]z_i$ and $T_n = \max_{i \leq n} \sup_{|\theta| \leq \delta_n} |\eta_i(\theta)|$. Since $\phi$ is monotone, for $\delta \geq 0$,

\[
\sup_{|\theta| \leq \delta} |\eta_i(\theta)| \leq |z_i| \max[|\phi(e_i) - |z_i|\delta) - \phi(e_i)|, |\phi(e_i + |z_i|\delta) - \phi(e_i)|].
\]

So $\mathbb{E}(T_n^2) \leq \tau_n(\delta_n)$ and

\[
\mathbb{P}(T_n \geq b_n) \leq b_n^{-2}\mathbb{E}(T_n^2) \leq b_n^{-2}\tau_n(\delta_n) = (\log n)^{-q} \to 0. \tag{38}
\]

Let $|\theta| \leq \delta$. Since $M_n$ is a martingale,

\[
\|M_n(\theta) - M_n(0)\|^2 = \sum_{i=1}^n \|\eta_i(\theta) - \mathbb{E}[\eta_i(\theta)|\mathcal{F}_{i-1}]\|^2 \leq \sum_{i=1}^n \|\eta_i(\theta)\|^2
\]

\[
= \sum_{i=1}^n m^2(-z_i'\theta)z_i^2 \leq \tau_n(\delta).
\]

By Freedman’s (1975) inequality,

\[
\mathbb{P}(|M_n(\theta) - M_n(0)| \geq r_n, T_n \leq b_n) \leq 2 \exp\{-r_n^2/[4b_n r_n + 2\tau_n(\delta_n)]\}
\]

18
Let $\ell = n^5$ and $\Theta_\ell = \{(k_1/\ell, \ldots, k_p/\ell) : k_i \in \mathbb{Z}, |k_i| \leq n^6\}$. Note that $\Theta_\ell$ has $(2n^6 + 1)^p$ points. By (39),

$$
\mathbb{P}\left[ \sup_{\theta \in \Theta_\ell} |M_n(\theta) - M_n(0)| \geq r_n, T_n \leq b_n \right] = O(n^{6p}) \exp\left[ -(\log n)^{(q+1)/2}/5 \right].
$$

(40)

For $a \in \mathbb{R}$ let $[a]_\ell = [a \ell]/\ell$ and $[a]_\ell = [a \ell]/\ell$, where $[\cdot]$ and $\lfloor \cdot \rfloor$ are the usual ceiling and floor functions. For a vector $\theta = (\theta_1, \ldots, \theta_p)'$ let $[\theta]_\ell = ([\theta_1]_\ell, \ldots, [\theta_p]_\ell)'$ and similarly define $[\theta]_\ell$. Since $\phi$ is nondecreasing, it is easily seen that $\eta_i([\theta]_\ell) \geq \eta_i(\theta) \geq \eta_i([\theta]_\ell)$, where the relation $u \geq v$ means that $u_i \geq v_i$ for all $1 \leq i \leq p$. By exchanging the order of integration and Schwarz’s inequality,

$$
|\mathbb{E}[\phi(e_i - t) - \phi(e_i - s)|F_{i-1}]| = \int_{\mathbb{R}} \phi(v)[f_\ell(t + v|F_{i-1}) - f_\ell(s + v|F_{i-1})]dv
\leq \int_{\mathbb{R}} \left| \phi(v) \right| \int_s^t \left| f_\ell'(x + v|F_{i-1}) \right| dx dv
\leq \int_s^t \left[ \int_{\mathbb{R}} \left| f_\ell'(x + v|F_{i-1}) \right|^2 w(du) \left| f_\ell'(u|F_{i-1}) \right|^2 w(du) \right]^{1/2} dx
\leq \kappa^{1/2} \int_s^t (1 + |x|)^{1/2}dx \left[ \int_{\mathbb{R}} \left| f_\ell'(u|F_{i-1}) \right|^2 w(du) \right]^{1/2},
$$

where the last inequality is due to the fact that $1 + |u - x| \leq (1 + |u|)(1 + |x|)$. Let

$$
V_n = \sum_{i=1}^n \left[ \int_{\mathbb{R}} \left| f_\ell'(u|F_{i-1}) \right|^2 w(du) \right]^{1/2}.
$$

By (36), $\mathbb{P}(V_n \geq n^2) \leq n^{-2} \mathbb{E}(V_n) = O(n^{-1})$. Since $|\theta - [\theta]_\ell| = O(\ell^{-1})$, we have $\max_{i \leq n} |z_i'(\theta - [\theta]_\ell)| = o(\ell^{-1})$, and by (9),

$$
\sup_{|\theta| \leq \delta_n} \sum_{i=1}^n |\mathbb{E}[\eta_i([\theta]_\ell) - \eta_i(\theta)|F_{i-1}]| \leq C\ell^{-1}V_n = Cn^{-5}V_n.
$$

Therefore, for all $|\theta| \leq \delta_n$,

$$
M_n([\theta]_\ell) - M_n(0) + Cn^{-5}V_n \geq M_n(\theta) - M_n(0) \geq M_n([\theta]_\ell) - M_n(0) - Cn^{-5}V_n,
$$

which implies (37) in view of (38), (40) and $\mathbb{P}(V_n \geq n^2) = O(n^{-1})$. \hfill \Box
Lemma 4. Under the conditions of Theorem 2, we have

\[
\sup_{|n| \leq \delta_n} |N_n(\cdot) - N_n(0)| = O(\sqrt{\Gamma_n(4)\delta_n}).
\]

Proof of Lemma 4. Let \( I = \{1, \ldots, r\} \subseteq\{1, \ldots, p\} \) be a non-empty set. Assume that \( 1 \leq 1 < \ldots < r \). For a \( p \)-dimensional vector \( u = (u_1, \ldots, u_p) \) let \( u_I = (u_11_{1 \in I}, \ldots, u_p1_{p \in I}) \).

Note that the \( j \)-th component of \( u_I \) is 0 if \( j \not\in I \), \( 1 \leq j \leq p \). Let

\[
H(v, u) = H'(v, u) = \sum_{i=1}^{n} |f_\varepsilon(x) (v + z_i u|F_{i-1}) - f_\varepsilon(v + z_i u)| z_i z_{i, \alpha_1} \ldots z_{i, \alpha_r},
\]

\[
J_k(v, u) = J_{n,k}(v, u) = \sum_{i=1}^{n} |z_i|^2 (z_{i, \alpha_1} \ldots z_{i, \alpha_r})^2 \|P_{i-k}f_\varepsilon(v + z_i u|F_{i-1})\|^2
\]

and \( \mu_k = \int_{\mathbb{R}} \|P_0f_\varepsilon(v|F_{-1})\|^2 w (dv) \). By (6), \( \Lambda := \sum_{k=1}^{\infty} \sqrt{\mu_k} < \infty \). Using the inequality \( 1 + |x + y| \leq (1 + |x|)(1 + |y|) \), we have by (9) that

\[
\int_{\mathbb{R}} J_k(v, u) w (dv) \leq \sum_{i=1}^{n} |z_i|^2 (z_{i, \alpha_1} \ldots z_{i, \alpha_r})^2 \int_{\mathbb{R}} \|P_{i-k}f_\varepsilon(v|F_{i-1})\|^2 w (dv)(1 + |z_i u|)
\]

\[
= O(\Gamma_n(2 + 2\tau) \mu_k)
\]

holds uniformly over \( |u| \leq \delta_n \sqrt{p} \) since \( \max_{i \leq n} |z_i'| u| \leq \delta_n \sqrt{p} r_n \rightarrow 0 \). For each \( k \in \mathbb{N} \), observing that \( P_{i-k}f_\varepsilon(v + z_i'u|F_{i-1}) \), \( i = 1, \ldots, n \), form martingale differences,

\[
\|H(v, u)\| = \sum_{k=1}^{\infty} \sum_{i=1}^{n} P_{i-k}f_\varepsilon(v + z_i'u|F_{i-1}) z_i z_{i, \alpha_1} \ldots z_{i, \alpha_r}
\]

\[
\leq \sum_{k=1}^{\infty} \sum_{i=1}^{n} \|P_{i-k}f_\varepsilon(v + z_i'u|F_{i-1})\|^2 z_i z_{i, \alpha_1} \ldots z_{i, \alpha_r}
\]

\[
= \sum_{k=1}^{\infty} [J_k(v, u)]^{1/2},
\]

which by Schwarz’s inequality implies that

\[
\int_{\mathbb{R}} \|H(v, u)\|^2 w (dv) \leq \int_{\mathbb{R}} \Lambda \left[ \sum_{k=1}^{\infty} \frac{J_k(v, u)}{\sqrt{\mu_k}} \right] w (du) = O(\Gamma_n(2 + 2\tau)).
\]

Write

\[
\int_0^t \frac{\partial |I|}{\partial u_I} N_n(u_I) du_I = \int_0^1 \cdots \int_0^\tau \frac{\partial^r N_n(u_I)}{\partial u_1 \ldots \partial u_r} du_1 \ldots du_r.
\]

20
By the representation (35),
\[
\frac{\partial^{|l|}N_n(u_l)}{\partial u_l} = \int_{\mathbb{R}} \phi(v)H(v,u)w\left(du\right).
\]
Again by Schwarz’s inequality,
\[
\left|\frac{\partial^{|l|}N_n(u_l)}{\partial u_l}\right|^2 \leq \int_{\mathbb{R}} \phi^2(v)w\left(du\right) \int_{\mathbb{R}} \|H^{(\tau)}(v,u)\|^2 \left(du\right) = O(\Gamma_n(2 + 2\tau)]
\]
holds uniformly over \(|u| \leq \delta_n\sqrt{p}\). Consequently,
\[
\sup_{|t| \leq \delta_n} \int_{0}^{t} \frac{\partial^{|l|}N_n(u_l)}{\partial u_l} \, du_l \leq \left(\int_{-\delta_n}^{\delta_n} \, du_l\right)^{\frac{1}{2}} \leq \int_{-\delta_n}^{\delta_n} \, du_l = O(\delta_n\sqrt{\Gamma_n(2 + 2\tau)}).
\]
By (9), \(\delta_n\sqrt{\Gamma_n(2 + 2|I|)} = O(\delta_n\sqrt{\Gamma_n(4)})\). So (41) follows from the identity
\[
N_n(\cdot) - N_n(0) = \sum_{I \subseteq \{1, \ldots, p\}} \int_{0}^{t} \frac{\partial^{|l|}N_n(u_l)}{\partial u_l} \, du_l,
\]
where the summation is over all the non-empty \(2^p - 1\) subsets of \(\{1, \ldots, p\}\).

**Proof of Corollary 1.** (i) The sequence \(\delta_n\) clearly satisfies (9). Theorem 1 implies that \(|\hat{\theta}_n| = O_P(1) = o_P(\delta_n)\). Note that \(K_n(0) = \sum_{i=1}^{n} \phi(e_i)z_i\) and by (12),
\[
K_n(\hat{\theta}_n) = -\sum_{i=1}^{n} \varphi(-z_i'(\hat{\theta}_n))z_i + O_P[\sqrt{\tau_n(1) \log n + r_n}].
\]
By Theorem 2, (13) follows from
\[
\sum_{i=1}^{n} \varphi(-z_i'(\hat{\theta}_n))z_i = \sum_{i=1}^{n} O(|z_i'(\hat{\theta}_n)|^2)z_i = O_P[\Gamma_n(3)]
\]
in view of \(\sum_{i=1}^{n} z_i z_i' = \text{Id}_p\), \(\Gamma_n(2) = p\), \(\Gamma_n(4) \leq r_n^2 \Gamma_n(2)\) and \(\Gamma_n(3) \leq r_n^p\). (ii) Since \(\tau_n(c)\) is nondecreasing in \(c > 0\), the condition that \(\sqrt{\tau_n(c)(\log n)^q} \to 0\) holds for all \(c\) implies that there exists a sequence \(c_n \to \infty\) such that \(\sqrt{\tau_n(c_n)(\log n)^q} \to 0\). As in (i), let \(\delta_n = \min(c_n, r_n^{-1/2})\). Then \(\delta_n\sqrt{\Gamma_n(4)} = O_P(\delta_n r_n) \to 0\) and
\[
\sum_{i=1}^{n} |[\varphi(-z_i'(\hat{\theta}_n)) + z_i'(\hat{\theta}_n) \varphi'(0)]z_i| = \sum_{i=1}^{n} o(|z_i'(\hat{\theta}_n)||z_i|) = o_P[\Gamma_n(2)] = o_P(1).
\]
Hence by (14), \( \varphi'(0)\hat{\theta}_n - \sum_{i=1}^{n} \phi(e_i)z_i = o_{\mathbb{P}}(1) \) and the central limit theorem follows from Lemma 1.

\[ \triangle \]

5 Proofs of results in Section 3.

A key issue in applying Corollary 1 is to verify the short-range dependent condition (A6). The following lemma is needed in the proof of Proposition 2.

**Lemma 5.** Let \( \tau < 1 + 2 \). Then for every integer \( \tau \geq 0 \), \( I_{\tau} := \int_{\mathbb{R}}[f_\varepsilon(t)]^2w \, (dt) < \infty. \)

**Proof of Lemma 5.** Let \( c = \pi^{-1}\Gamma(1+\tau)\sin(\pi/2) \). By Theorem 2.4.2 in Ibragimov and Linnik (1971), we have \( f(t) \sim c \, |t|^{-1-\tau} \) as \( |t| \to \infty \). A similar argument shows that \( f^{(\tau)}(t) \sim c_{\alpha,\tau} |t|^{-1-\tau} \), where \( c_{\alpha,\tau} = c \sum_{i=1}^{\tau} (-i - ) \). Then the lemma easily follows since \( -2 - 2 + < -1. \)

\[ \triangle \]

**Proof of Proposition 2.** We first consider the case \( \tau = 0 \). Let \( V_n = \sum_{j=1}^{\infty} a_j\varepsilon_{n-j}, U_n = V_n - a_n\varepsilon_0 \); let \( (\varepsilon_i)_{i \in \mathbb{Z}} \) be an iid copy of \( (\varepsilon_i)_{i \in \mathbb{Z}} \). Without loss of generality assume \( a_0 = 1 \). Note that \( f_\varepsilon(t|\mathcal{F}_{n-1}) = f_\varepsilon(t - V_n) \). Then

\[
\| \mathcal{P}_0 f_\varepsilon(t|\mathcal{F}_{n-1}) \| = \| \mathbb{E}[f_\varepsilon(t - V_n)|\mathcal{F}_0] - \mathbb{E}[f_\varepsilon(t - V_n)|\mathcal{F}_{n-1}] \| \\
= \| \mathbb{E}[f_\varepsilon(t - U_n - a_n\varepsilon_0)|\mathcal{F}_0] - \mathbb{E}[f_\varepsilon(t - U_n - a_n\varepsilon_0^*)|\mathcal{F}_0] \| \\
\leq \| f_\varepsilon(t - V_n) - f_\varepsilon(t - U_n - a_n\varepsilon_0^*) \| \\
\leq \| f_\varepsilon(t - V_n) - f_\varepsilon(t - U_n) \| + \| f_\varepsilon(t - U_n) - f_\varepsilon(t - U_n - a_n\varepsilon_0^*) \| \\
= 2\| f_\varepsilon(t - U_n - a_n\varepsilon_0) - f_\varepsilon(t - U_n) \|. \tag{43}
\]

Let

\[
R_n = \int_{\mathbb{R}}[f_\varepsilon(t - U_n) - f_\varepsilon(t - U_n - a_n\varepsilon_0)]^2w \, (dt).
\]

We shall show that \( \mathbb{E}(R_n) = O(|a_n|^\gamma) \), from which (19) follows. Recall Lemma 5 for the definition of \( I_{\tau} \). Observe that

\[
\int_{\mathbb{R}} f_\varepsilon^2(t - u)w \, (dt) \leq (1 + |u|) \int_{\mathbb{R}} f_\varepsilon^2(t)w \, (dt) = I_0(1 + |u|) .
\]

22
Then
\[
R_n \leq C[(1 + |U_n|) + (1 + |U_n + a_n\varepsilon_0|)] \\
\leq C[(1 + |U_n|) + |a_n\varepsilon_0|].
\] (44)

Here the constant \(C\) may change from line to line. On the other hand, since
\[
|f_\varepsilon(t - U_n) - f_\varepsilon(t - U_n - a_n\varepsilon_0)|^2 = \int_0^{a_n\varepsilon_0} f_\varepsilon'(t - U_n - v) dv^2 \\
\leq |a_n\varepsilon_0| \int_0^{a_n\varepsilon_0} [f_\varepsilon'(t - U_n - v)]^2 dv,
\]
we have
\[
\int_\mathbb{R} \int_0^{a_n\varepsilon_0} [f_\varepsilon'(t - U_n - v)]^2 dv \ w (dt) \leq \int_0^{a_n\varepsilon_0} \int_\mathbb{R} [f_\varepsilon'(t - U_n - v)]^2 w (dt) dv \\
\leq I_1 \int_0^{a_n\varepsilon_0} (1 + |U_n + v|) \ dv \\
\leq C|a_n\varepsilon_0|[\{1 + |U_n|\} + (1 + |U_n + a_n\varepsilon_0|)] \\
\leq C|a_n\varepsilon_0|[(1 + |U_n|) + |a_n\varepsilon_0|]
\]
and consequently by Schwarz’s inequality,
\[
R_n \leq |a_n\varepsilon_0| \int_\mathbb{R} \int_0^{a_n\varepsilon_0} [f_\varepsilon'(t - U_n - v)]^2 dv \ w (dt) \\
\leq C|a_n\varepsilon_0|^2[(1 + |U_n|) + |a_n\varepsilon_0|].
\] (45)

Recall \(1 < \varepsilon_0 < 2\). Since \(a_n\varepsilon_0\) and \(U_n\) are independent and \(\mathbb{E}(|U_n|) < \infty\), we have by (44) and (44) that
\[
\mathbb{E}(R_n) \leq C\mathbb{E}[\min(|a_n\varepsilon_0|^2, 1)(1 + |U_n|)] + C\mathbb{E}[\min(|a_n\varepsilon_0|^2, 1)|a_n\varepsilon_0|] \\
\leq C\mathbb{E}[|a_n\varepsilon_0| \circ (1 + |U_n|)] + C\mathbb{E}[|a_n\varepsilon_0| \circ |a_n\varepsilon_0|] = O(|a_n|)
\]

The general case \(\tau \geq 1\) similarly follows. \(\diamondsuit\)

_Proof of Corollary 3._ We shall apply (i) of Corollary 1 and Proposition 2. Let \(\tau = (1 + \varepsilon_0)/2\). Then \(1 < \varepsilon_0 < 2\). (i) By Proposition 2, (A6) holds. It is easily verified that \(\varphi'(0) = \mathbb{P}(|e_1| \leq c) > 0, \varphi(0) = 0\) and (A1)-(A3) holds. (A4) is satisfied since \(\phi\) is bounded and \(\tau > 1\). Note that \(m(t) = O(|t|)\). So \(\tau_n(\delta) = \Gamma_n(4O(\delta^2))\), and (22)
follows from Corollary 1 by letting \( \delta_n = (\log n)^{q-1} \). The central limit theorem is an easy consequence of Lemma 1. (ii) Since \( \mathbb{E}[1_{e_1 + t \leq \xi} - 1_{e_1 \leq \xi}] \sim f(\xi)|t| \) as \( t \to 0 \), \( m(t) = O(|t|^{1/2}) \) and \( \tau_n(\delta) = \sqrt{\Gamma_n(3)}O(\delta) \). Let \( q = (1 + q')/2 \) and \( \delta_n = (\log n)^{q'-1} \). By Corollary 1, (23) follows from

\[
\sqrt{\tau_n(\delta_n)}(\log n)^q + r_n\delta_n = O[\sqrt{r_n}(\log n)^q]
\]
since \( \Gamma_n(3) \leq r_n\Gamma_n(2) \to 0 \).

**Proof of Corollary 4.** As in Proposition 2, let \( V_n = \sum_{j=1}^\infty a_j \varepsilon_{n-j} \) and \( U_n = V_n - a_n \varepsilon_0 \). It is easily seen from the proof of Proposition 2 that, under the condition (25), we have

\[
R_n^\tau = \int_\mathbb{R} \left[ f^{(\tau)}(t - U_n) - f^{(\tau)}(t - U_n - a_n \varepsilon_0) \right]^2 w(\,dt) \leq C \min(1, |a_n \varepsilon_0|^2) [(1 + |U_n|) + |a_n \varepsilon_0| ]
\]

for \( \tau = 0, \ldots, p \). Since \( \mathbb{E}(\varepsilon_j^2) < \infty \), \( \mathbb{E}[1 + |U_n|] = O(1) \). Therefore, it follows that \( \mathbb{E}(R_n^\tau) = O(a_n^2) \) in view of the independence of \( a_n \varepsilon_0 \) and \( U_n \) and \( \min(1, |a_n \varepsilon_0|^2)|a_n \varepsilon_0| \leq |a_n \varepsilon_0|^2 \). By (43),

\[
\sum_{i=0}^\infty \sqrt{\int_R \| P_0 f^{(\tau)}(t|F_i) \|^2 w(\,dt)} = \sum_{i=0}^\infty O(|a_i|) < \infty.
\]

Hence (A6) holds. By Lemma 4 in Wu (2003b), (25) implies that \( \sup_\tau |f^{(\tau)}(v)| < \infty \) for \( \tau = 0, \ldots, p \), which entails the desired results in the same manner as the proof of Corollary 3.

**Proof of Corollary 5.** By Theorem 2 in Wu and Shao (2003), the conditions (30) and \( \mathbb{E}(|\varepsilon_i|) < \infty \) imply that \( e_i \in L \) and \( \mathbb{E}(|e_i - e'_i|) \leq CL\overline{\alpha} \). Note that \( f^{(\tau)}(t|F_i) = f^{(\tau)}(t - m(e_i)) \)

and

\[
\mathbb{E}\{f^{(\tau)}(t - m(e_i))|F_{-1}\} = \mathbb{E}\{f^{(\tau)}(t - m(e'_i))|F_0\}.
\]

Then by Schwarz’s inequality,

\[
\| P_0 f^{(\tau)}(t - m(e_i)) \| = \| \mathbb{E}\{f^{(\tau)}(t - m(e_i))|F_0\} - \mathbb{E}\{f^{(\tau)}(t - m(e_i))|F_{-1}\} \|
\leq \| f^{(\tau)}(t - m(e_i)) - f^{(\tau)}(t - m(e'_i)) \|.
\]

24
We first consider the case \( \tau = 0 \). By (44) and (45), since \( |m(a)| \leq |m(0)| + |a| \),

\[
R_i = \int_{\mathbb{R}} \left\{ f_\varepsilon(t - m(e_i)) - f_\varepsilon(t - m(e'_i)) \right\}^2 w \, dt
\leq C \min(1, |e_i - e'_i|^2) \{ [1 + |m(e_i)|] + [1 + |m(e'_i)|] \}
\leq C \min(1, |e_i - e'_i|^2) [1 + |e_i| + |e'_i|]
\leq |e_i - e'_i| - [1 + |e_i| + |e'_i|].
\]

By Hölder’s inequality,

\[
\mathbb{E}(R_n) \leq C \mathbb{E}(\{e_i - e'_i\})^{(1/2) - 1} \times \mathbb{E}(1 + |e_i| + |e'_i|)^{\alpha / \gamma} = O[L_m^{\alpha / \gamma}].
\]

So (31) follows by letting \( \chi = L_m^{-\gamma} \). The general case \( \tau \geq 1 \) similarly follows.

Acknowledgments. The author is grateful to Jan Mielniczuk for many useful suggestions.

REFERENCES


Department of Statistics,
The University of Chicago,
5734 S. University Avenue, Chicago, IL 60637
wbwu@galton.uchicago.edu