The Annals of Statistics
1998, Vol. 26, No. 2, 741–754

WEAK CONVERGENCE OF THE SEQUENTIAL EMPIRICAL PROCESSES OF RESIDUALS IN NONSTATIONARY AUTOREGRESSIVE MODELS

By Shiqing Ling

University of Hong Kong and University of Western Australia

This paper establishes the weak convergence of the sequential empirical process $\hat{K}_n$ of the estimated residuals in nonstationary autoregressive models. Under some regular conditions, it is shown that $\hat{K}_n$ converges weakly to a Kiefer process when the characteristic polynomial does not include the unit root 1; otherwise $\hat{K}_n$ converges weakly to a Kiefer process plus a functional of stochastic integrals in terms of the standard Brownian motion. The latter differs not only from that given by Koul and Levental for an explosive AR(1) model but also from that given by Bai for a stationary ARMA model.

1. Introduction and main results. Empirical processes based on the estimated residuals in a variety of models have been studied for a long time. In the field of time series, Boldin (1982) and Kreiss (1991) examined their weak convergence for some stationary ARMA(p, q) models and Koul and Levental (1989) investigated their weak convergence for an explosive AR(1) model. Bai (1994) extended Boldin's results to stationary ARMA models by considering the sequential empirical process based on estimated residuals. Under some conditions, these authors proved that the estimated residual empirical processes have identical weak convergence properties to those of the residual empirical processes. Many important applications can be found in the cited literature and Koul (1991). In this paper, my interest is to investigate the weak convergence of the sequential empirical processes when the estimated residuals come from nonstationary autoregressive models.

Consider the autoregressive model

$$y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t, \tag{1.1}$$

where $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) random disturbances, y_t is the observation with starting value $(y_0, y_{-1}, \dots, y_{1-p})$ independent of $\{\varepsilon_t\}$ and the characteristic polynomial $\phi(z) = 1 - \beta_1 z - \cdots - \beta_p z^p$

Received August 1996; revised September 1997.

AMS 1991 *subject classifications*. Primary 62G30, 60F17; secondary 62M10, 62F05.

Key words and phrases. Brownian motions, Kiefer process, sequential empirical processes, nonstationary autoregressive model, weak convergence.

has the decomposition,

$$\phi(z) = \psi(z)(1-z)^a(1+z)^b \times \prod_{k=1}^{l} \left[(1 - z\exp(i\theta_k))(1 - z\exp(-i\theta_k))\right]^{d_k}, \tag{1.2}$$

where $a, b, l, d_k, k = 1, \dots, l$ are nonnegative intergers, $0 < \theta_k < \pi$ and $\psi(z)$ is a polynomial of degree $q = p - [a + b + 2(d_1 + \cdots + d_l)]$ with all roots outside the unit circle. The model (1.1) is a general nonstationary autoregressive time series. In the last ten years, a huge amount of statistical literature has been devoted to the study of nonstationary time series. Some general results on the estimation theory can be found in Chan and Wei (1988) and Jeganathan (1991).

Given $n + p$ observations, $y_{1-p}, \dots, y_0, y_1, \dots, y_n$. Let $\hat{\alpha}$ be any estimator of the parameter $\alpha = (\beta_1, \dots, \beta_p)^T$. The estimated residual $\hat{\varepsilon}_t$ is defined by

$$\hat{\varepsilon}_t = y_t - \hat{\alpha}^T X_{t-1}, \tag{1.3}$$

where $t = 1, \dots, n$, the superscript T of A^T denotes the transpose of a vector or matrix A and $X_t = (y_t, \dots, y_{t-p+1})^T$. Define the sequential empirical processes based on estimated residuals as

$$\hat{K}_n(s, x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left[I(\hat{\varepsilon}_t \le x) - F(x)\right], \qquad 0 \le s \le 1, \tag{1.4}$$

where $I(\cdot)$ is the indicator function. Similarly, define the sequential empirical processes $K_n(s, x)$ with $\hat{\varepsilon}_t$ replaced by ε_t. When $s = 1$, $K_n(s, x)$ reduces to the empirical process $G_n(x) = (1/\sqrt{n})\sum_{t=1}^{n}[I(\varepsilon_t \le x) - F(x)]$. My result can be stated by the following theorem.

THEOREM. *Suppose that the following conditions are satisfied*:

(i) *The* ε_t *are i.i.d. with* $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, *and a common distribution* $F(x)$;

(ii) $F(x)$ *admits a uniformly continuous density function* $f(x)$, $f(x) > 0$ *a.e.*;

(iii) $\delta_n^{-1}(\hat{\alpha} - \alpha) = O_p(1)$.

Then

$$\sup_{s\in[0,1],\, x\in R} \left|\hat{K}_n(s, x) - K_n(s, x) - R_n(s, x)\right| = o_p(1), \tag{1.5}$$

where $O_p(1)$ [*or* $o_p(1)$] *stands for a series of random variables that is bounded* (*or converges to zero*) *in probability*, δ_n *is defined in Lemma* 2.1 *and* $R_n(s, x) = (\hat{\alpha} - \alpha)^T \sum_{t=1}^{[ns]} X_{t-1} f(x)/\sqrt{n}$.

REMARK. Assumptions (i) and (ii) are identical as those given in Koul (1991) and Bai (1994). Assumption (iii) is satisfied by the usual least squares estimator as in Chan and Wei (1988). The asymptotic behavior of $R_n(s, x)$ depends on the locations of the unit roots of $\phi(z)$ and hence it also affects the weak convergence of $\hat{K}_n(s, x)$. Further discussion is divided into the following two cases.

CASE 1. When $\phi(z)$ does not include the unit root 1, by assumption (iii) and Lemma 2.1(b) in the next section, $R_n(s, x) = o_p(1)$ uniformly for all $s \in [0, 1]$ and all $x \in R$. From Bickel and Wichura (1971), $K_n(s, F^{-1}(\tau))$ converges weakly in D_2 to a Kiefer process $K(s, \tau)$, a two-parameter Gaussian process with zero mean and covariance function

$$\operatorname{cov}(K(s_1, \tau_1), K(s_2, \tau_2)) = (s_1 \wedge s_2)(\tau_1 \wedge \tau_2 - \tau_1 \tau_2),$$

where D_2 denotes the space of functions $f(s, \tau)$ on $[0, 1]^2$, which is defined and equipped with the Skorokhod topology in Straf (1970) and Bickel and Wichura (1971). Thus the theorem actually implies that $\hat{K}_n(s, F^{-1}(\tau))$ converges weakly to a Kiefer process $K(s, \tau)$ in D_2. These results are the same as those already known in stationary cases and hence some statistics based on $K_n(s, x)$ can be reconstructed by employing $\hat{K}_n(s, x)$ to replace $K_n(s, x)$. All applications as in Boldin (1982), Koul and Levental (1989) and Bai (1994), and other references can be carried over to these nonstationary cases.

CASE 2. When $\phi(z)$ includes the unit root 1 with multiplicities a, if we further assume that $\delta_n^{-1}(\hat{\alpha} - \alpha)$ converges in distribution to a random variable $\tilde{\xi}$ and $([\delta_n^{-1}(\hat{\alpha} - \alpha)]^T, \sum_{t=1}^{[ns]} X_{t-1}^T \delta_n / \sqrt{n})$ converges weakly in $R^p \times D^p$, then by the continuous mapping theorem [Billingsley (1968), Theorem 5.1] and Lemma 2.1(a), $R_n(s, F^{-1}(\tau))$ converges weakly to $(\xi^T(s), O)\tilde{\xi} f(F^{-1}(\tau))$ in D_2, where $\xi(s)$ is defined in Lemma 2.1 and $D^n = D \times D \times \cdots \times D$ denotes the product space of n-D spaces. In particular, if $\hat{\alpha}$ is the least squares estimator of α, that is, $\hat{\alpha} = (\sum_{t=1}^n X_{t-1} X_{t-1}^T)^{-1} \sum_{t=1}^n X_{t-1} y_t$, by Theorem 2.2 and Theorem 3.5.1 of Chan and Wei (1988) and Lemma 2.1(a), $([\delta_n^{-1}(\hat{\alpha} - \alpha)]^T, \sum_{t=1}^{[ns]} X_{t-1}^T \delta_n / \sqrt{n})$ converges weakly in $R^p \times D^p$ and $(\xi^T(s), O)\tilde{\xi} = \xi^T(s) \Omega \zeta$, where $\Omega = (\sigma_{ij})$, $\sigma_{ij} = \int_0^1 g_{i-1}(\tau) g_{j-1}(\tau)\, d\tau$, for $i, j = 1, \dots, a$, $\zeta = (\int_0^1 g_0(\tau)\, dW(\tau), \dots, \int_0^1 g_{a-1}(\tau)\, dW(\tau))^T$, and $g_i(\tau)$ and $W(\tau)$ are defined in Lemma 2.1. In this case the theorem implies that $\hat{K}_n(\cdot, \cdot)$ converges weakly to a Kiefer process plus a functional of stochastic integrals in terms of the standard Brownian motion. This is different from those given by Koul and Levental (1989) and Bai (1994). Since the limiting distribution of $\hat{K}_n(\cdot, \cdot)$ is no longer distribution free, the prototypical Kolmogorov–Smirnov tests based on the estimated residuals cannot be used. Statistical inferences related to innovations of these nonstationary time series will become more difficult.

The proof of the theorem will be shown in the next section and the following notation will be used: $\Rightarrow$ denotes convergence in distribution and $\|\cdot\|$ denotes the Euclidean norm.

2. Proof of the theorem. Before giving the proof of the theorem, we first present several lemmas.

LEMMA 2.1. *Suppose that* $\{y_t\}$ *is generated by the nonstationary AR*(p) *model* (1.1) *and assumption* (i) *in the theorem is satisfied. Then we have the following.*

(a) *If* $a \neq 0$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \delta_n^T X_{t-1} = \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(N_1^{-1} U_{t-1} \right)^T, o_p(1) \right)^T \Rightarrow \left(\xi^T(s), O \right)^T \quad \text{in } D^p;$$

(b) *If* $a = 0$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \delta_n^T X_{t-1} = o_p(1);$$

(c) $$\sup_{1 \le t \le n} \left\| \delta_n^T X_{t-1} \right\| = o_p(1);$$

(d) $$\sup_{1 \le t \le n} E \left\| \delta_n^T X_{t-1} \right\|^2 = O(n^{-1});$$

(e) $$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\| \delta_n^T X_{t-1} \right\| = O_p(1);$$

(f) $$\sum_{t=1}^{n} \left\| \delta_n^T X_{t-1} \right\|^2 = O_p(1);$$

(g) $$\sum_{t=1}^{n} E \left\| \delta_n^T X_{t-1} \right\|^2 = O(1),$$

where $\delta_n = G^T J_n^{-1}$, N_1, U_t, J_n *and* G *are defined below*; *the* $o_p(1)$ *in* (a) *and* (b) *holds uniformly in* $s \in [0,1]$; $\xi(s) = (\int_0^s g_i(\tau)\, d\tau,\ i = 0, 1, \ldots, a-1)^T$, $g_0(\tau) = W(\tau)$, $g_j(\tau) = \int_0^\tau g_{j-1}(\tau)\, d\tau$, $j = 1, \ldots, a$; *and* $W(\tau)$ *is the standard Brownian motion.*

REMARK. The proof of this lemma mainly uses the idea and some results of Chan and Wei (1988), abbreviated henceforth as CW. The results of CW are obtained under the assumption that $\{\varepsilon_t\}$ is a series of martingale differences and $\sup_t E|\varepsilon_t|^{2+\kappa} < \infty$, where κ is a positive constant. Since $\{\varepsilon_t\}$ here is a sequence of i.i.d. random variables, assumption (i) is sufficient for their results [cf. Jeganathan (1991)].

PROOF. For simplicity, in the following we will assume that the starting values $y_0 = y_{-1} = \cdots = y_{1-p} = 0$. Denote $N_1 = \operatorname{diag}(n, n^2, \ldots, n^a)$, $N_2 = \operatorname{diag}(n, n^2, \ldots, n^b)$, $N_{k+2} = \operatorname{diag}(nI_2, \ldots, n^{d_k} I_2)$, $k = 1, \ldots, l$ and $J_n = \operatorname{diag}(N_1, N_2, \ldots, N_{l+2}, \sqrt{n}\, I_q)$, where I_k is the $k \times k$ identity matrix.

Let $u_t = \phi(B)(1-B)^{-a} y_t$, $\tilde{u}_t = (u_t, \ldots, u_{t-a+1})^T$, $v_t = \phi(B)(1+B)^{-b} y_t$, $\tilde{v}_t = (v_t, \ldots, v_{t-b+1})^T$, $z_t = \phi(B)\psi^{-1}(B) y_t$, $\tilde{z}_t = (z_t, \ldots, z_{t-q+1})^T$, $x_t(k) = \phi(B)(1 - 2B\cos\theta_k + B^2)^{-d_k} y_t$ and $\tilde{x}_t(k) = (x_t(k), \ldots, x_{t-d_k+1}(k))^T$, where B is a backshift operator and $k = 1, \ldots, l$. As shown in (3.2) of CW, there exists a nonsingular matrix Q such that

$$QX_t = \left(\tilde{u}_t^T, \tilde{v}_t^T, \tilde{x}_t^T(1), \dots, \tilde{x}_t^T(l), \tilde{z}_t^T\right)^T. \tag{2.1}$$

Further let $U_t(j) = (1-B)^{a-j}u_t$ for $j = 0, 1, \dots, a$, $U_t = (U_t(1), \dots, U_t(a))^T$, $V_t(j) = (1+B)^{b-j}v_t$ for $j = 0, 1, \dots, b$, $V_t = (V_t(1), \dots, V_t(b))^T$, $Y_t(k, j) = (1 - 2B\cos\theta_k + B^2)^{d_k - j}x_t(k)$ for $j = 0, 1, \dots, d_k$, $k = 1, \dots, l$, and $Y_t(k) = (Y_t(k,1), Y_{t-1}(k,1), \dots, Y_t(k, d_k), Y_{t-1}(k, d_k))^T$, where $k = 1, \dots, l$. Then there exist nonsingular matrices $M, \tilde{M}, C_k, k = 1, \dots, l$ such that

$$M\tilde{u}_t = U_t, \qquad \tilde{M}\tilde{v}_t = V_t, \qquad C_k\tilde{x}_t(k) = Y_t(k), \qquad k = 1, \dots, l.$$

Denote $G = \operatorname{diag}(M, \tilde{M}, C_1, \dots, C_l, I_q)Q$. We have

$$GX_t = \left(U_t^T, V_t^T, Y_t^T(1), \dots, Y_t^T(l), \tilde{z}_t^T\right)^T. \tag{2.2}$$

For (a), note that

$$U_t(1) = \sum_{i=1}^{t} U_i(0) = \sum_{i=1}^{t} \varepsilon_i, \qquad U_t(j+1) = \sum_{k=1}^{t} U_k(j),$$

where $j = 0, \dots, a-1$. By Therorem 2.2 and Theorem 2.3 of CW,

$$n^{(1/2)-j}U_{[n\tau]}(j) \Rightarrow g_{j-1}(\tau) \quad \text{in } D \text{ for } j = 1, \dots, a.$$

Again by Theorem 2.3 of CW, we obtain

$$\sqrt{n}\,N_1^{-1}U_{[n\tau]} \Rightarrow \left(g_0(\tau), \dots, g_{a-1}(\tau)\right)^T \quad \text{in } D^a. \tag{2.3}$$

By (2.3) and the continuous mapping theorem [Billingsley (1968), Theorem 5.1],

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} N_1^{-1}U_{t-1} = \frac{1}{n} \sum_{t=1}^{[ns]} \left(\sqrt{n}\,N_1^{-1}U_{t-1}\right) \Rightarrow \xi(s) \quad \text{in } D^a. \tag{2.4}$$

Similarly to (2.3) (see Theorem 3.2.1 of CW), we can obtain

$$\sqrt{n}\,N_2^{-1}(-1)^{[n\tau]}V_{[n\tau]} \Rightarrow -\left(\tilde{g}_0(\tau), \dots, \tilde{g}_{b-1}(\tau)\right)^T \quad \text{in } D^b, \tag{2.5}$$

where $\tilde{g}_j(\tau)$, $j = 0, \dots, b-1$, are defined as in Theorem 3.5.1 of CW. By Proposition 8 of Jeganathan (1991),

$$\begin{aligned} &\max_{1 \le j \le n} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{j} N_2^{-1}V_{t-1} \right\| \\ &\quad = \max_{1 \le j \le n} \left\| \frac{1}{n} \sum_{t=1}^{j} \exp((t-1)i\pi)\sqrt{n}\,N_2^{-1}(-1)^{t-1}V_{t-1} \right\| = o_p(1). \end{aligned} \tag{2.6}$$

Let

$$S_t(k, j) = \sum_{i=1}^{t} Y_i(k, j)\sin\theta_k \quad \text{and} \quad T_t(k, j) = \sum_{i=1}^{t} Y_i(k, j)\cos\theta_k,$$

where $k = 1, \dots, l$, $j = 0, \dots, d_k$. By a direct verification or Lemma 3.3.1 of CW, we have

$$(2.7) \quad Y_t(k,j)\sin\theta_k = S_t(k,j-1)\sin(t+1)\theta_k - T_t(k,j-1)\cos(t+1)\theta_k,$$

where $j = 1, \dots, d_k$. By Lemma 3.3.7 of CW,

$$(2.8) \quad \sqrt{2}\, n^{-j-1/2}\big(S_{[n\tau]}(k,j), T_{[ns]}(k,j)\big) \Rightarrow \big(f_{kj}(\tau), g_{kj}(s)\big) \quad \text{in } D^2,$$

where $k = 1, \dots, l$, $j = 0, \dots, d_k - 1$, $f_{kj}(\tau)$ and $g_{kj}(s)$ are defined in Theorem 3.5.1 of CW. Again by Proposition 8 of Jeganathan (1991), we obtain

$$(2.9) \quad \max_{1 \le i \le n} \left| \frac{1}{n} \sum_{t=1}^{i} n^{-(j-1)-1/2} S_{t-1}(k,j-1)\sin t\theta_k \right| = o_p(1),$$

$$(2.10) \quad \max_{1 \le i \le n} \left| \frac{1}{n} \sum_{t=1}^{i} n^{-(j-1)-1/2} T_{t-1}(k,j-1)\cos t\theta_k \right| = o_p(1),$$

where $j = 1, \dots, d_k$. By (2.7), (2.9) and (2.10), we have

$$(2.11) \quad \max_{1 \le i \le n} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{i} N_{k+2}^{-1} Y_{t-1}(k) \right\| = o_p(1) \qquad k = 1, \dots, l.$$

Since z_t is generated by model $\psi(B) z_t = \varepsilon_t$, $\{\tilde{z}\}$ is a stationary and ergodic process. Similarly to the proof of Theorem 1 in Bai (1993), we can show

$$(2.12) \quad \max_{1 \le j \le n} \left\| \frac{1}{n} \sum_{t=1}^{j} \tilde{z}_{t-1} \right\| = o_p(1).$$

When $a \neq 0$, by (2.2), (2.4), (2.6), (2.11) and (2.12), we obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \delta_n^T X_{t-1} = \left(\frac{1}{n} \sum_{t=1}^{[ns]} \left[\sqrt{n}\, N_1^{-1} U_{t-1}\right]^T, o_p(1) \right)^T \Rightarrow \big(\xi^T(s), O\big)^T \quad \text{in } D^p,$$

where $o_p(1)$ holds uniformly in $s \in [0,1]$. That is, (a) holds. By (2.2), (2.6), (2.11) and (2.12), we know that (b) holds.

For (c), by (2.3) and the continuous mapping theorem,

$$\max_{1 \le t \le n} \|\sqrt{n}\, N_1^{-1} U_t\| \quad \text{converges to} \quad \max_{0 \le \tau \le 1} \left[\sum_{i=0}^{a-1} g_i^2(\tau) \right]^{1/2}$$

in distribution and thus

$$(2.13) \quad \max_{1 \le t \le n} \left\| N_1^{-1} U_t \right\| = o_p(1).$$

Similarly we have

$$\max_{1 \le t \le n} \left\| N_2^{-1} V_t \right\| = o_p(1), \tag{2.14}$$

$$\max_{1 \le t \le n} \left\| n^{-j} S_t(k, j-1) \right\| = o_p(1) \quad \text{and} \qquad \max_{1 \le t \le n} \left\| n^{-j} T_t(k, j-1) \right\| = o_p(1) \tag{2.15}$$

for $k = 1, \dots, l$ and $j = 1, \dots, d_k$. By (2.7) and (2.15), we obtain

$$\max_{1 \le t \le n} \left\| N_{k+2}^{-1} Y_t(k) \right\| = o_p(1). \tag{2.16}$$

Since $\|\tilde{z}_t\|$ has identical distribution with finite variance,

$$\max_{1 \le t \le n} \|n^{-1/2} \tilde{z}_t\| = o_p(1)$$

[See Chung (1968), page 93 or the proof of Lemma 1(b) in Bai (1994)]. Further by (2.13), (2.14) and (2.16), (c) holds.

For (d), we first show that, by induction on j,

$$E\big(U_t^2(j)\big) = O(t^{2(j-1)+1}), \qquad j = 1, \dots, a. \tag{2.17}$$

As $j = 1$, (2.17) holds. Assume that (2.17) holds as $j = k$. Then

$$\begin{aligned} EU_t^2(j+1) &= E\left(\sum_{i=1}^{t} U_i(j) \right)^2 \le t \sum_{i=1}^{t} EU_i^2(j) \\ &= t \sum_{i=1}^{t} O(t^{2(j-1)+1}) = O(t^{2j+1}). \end{aligned} \tag{2.18}$$

So (2.17) holds for $j = 1, \dots, a$. Thus

$$\sup_{1 \le t \le n} E \left\| N_1^{-1} U_t \right\|^2 = O(n^{-1}). \tag{2.19}$$

Similarly we can show

$$\sup_{1 \le t \le n} E \left\| N_2^{-1} V_t \right\|^2 = O(n^{-1}). \tag{2.20}$$

By Lemma 3.3.5 of CW, for $k = 1, \dots, l$ and $j = 0, \dots, d_k - 1$,

$$ES_t^2(k, j) = O(t^{2j+1}) \quad \text{and} \quad ET_t^2(k, j) = O(t^{2j+1})$$

and further by (2.7), we can obtain

$$\sup_{1 \le t \le n} E \left\| N_{k+2}^{-1} Y_t(k) \right\|^2 = O(n^{-1}). \tag{2.21}$$

Since z_t is strictly stationary and has a finite variance,

$$\sup_{1 \le t \le n} E \left\| n^{-1/2} \tilde{z}_t \right\|^2 = O(n^{-1}). \tag{2.22}$$

By (2.19)–(2.22), it is easy to know that (d) holds. Then (e)–(g) come directly from (d). This completes the proof. □

Denote

$$g_t(u, \lambda) = u^T \delta_n^T X_{t-1} + \lambda \left\| \delta_n^T X_{t-1} \right\|, \tag{2.23}$$

where $u \in R^p$ and $\lambda \in R$.

LEMMA 2.2. *Suppose that* $\{y_t\}$ *is generated by the nonstationary* $AR(p)$ *model* (1.1) *and assumptions in the theorem hold. Then for any* $d \in (0, 1/2)$, *every* $u \in D_\Delta$ *and* $\lambda \in R$,

$$\sup_{(x, y) \in B_{n,d}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| F(y + g_t(u, \lambda)) - F(x + g_t(u, \lambda)) \right| = o_p(1), \tag{2.24}$$

where $B_{n,d} = \{(x, y) \in R \times R, |F(x) - F(y)| \le n^{-(1/2)-d}\}$ *and* $D_\Delta = [-\Delta, \Delta]^p \subset R^p$.

PROOF. By Lemma 2.1(c) and (e), $\max_{1 \le t \le n} |g_t(u, \lambda)| = o_p(1)$ and $\sum_{t=1}^n |g_t(u, \lambda)| / \sqrt{n} = O_p(1)$. The remaining proof is similar to the arguments of Lemma 2.1 in Koul (1991) and hence is omitted. This completes the proof. □

Define

$$\begin{aligned} \tilde{Z}_n(x, s, u, \lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \big[& I(\varepsilon_t \le x + g_t(u, \lambda)) - F(x + g_t(u, \lambda)) \\ & - I(\varepsilon_t \le x) + F(x) \big] \end{aligned} \tag{2.25}$$

and

$$H_n(x, s, u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left[F\left(x + u^T \delta_n^T X_{t-1}\right) - F(x) - u^T \delta_n^T X_{t-1} f(x) \right], \tag{2.26}$$

where $g_t(u, \lambda)$ is defined by (2.23), $u \in R^p$ and $\lambda \in R$.

LEMMA 2.3. *Under the assumptions of the theorem, for any* $u \in D_\Delta$ *and* $\lambda \in R$,

(a) $$\sup_{s \in [0,1],\, x \in R} |\tilde{Z}_n(x, s, u, \lambda)| = o_p(1);$$

(b) $$\sup_{s \in [0,1],\, x \in R} |H_n(x, s, u)| = o_p(1),$$

where Δ *is any fixed positive number and* D_Δ *is defined as in Lemma* 2.2.

PROOF. (a) Following the ideas of Boldin (1982) and Bai (1994), let $N(n) = [n^{1/2+d}] + 1$, where $d \in (0, 1/2)$ and partition the real line into $N(n)$

parts by the points

$$-\infty = x_0 \le x_1 \le \cdots \le x_N(n) = \infty \quad \text{where } F(x_i) = i/N(n).$$

Since $I(\varepsilon_t \le x)$ and $F(x)$ are nondecreasing, for any $x \in (x_r, x_{r+1}]$, we have

$$\begin{aligned}\tilde{Z}_n(x, s, u, \lambda) \le \tilde{Z}_n\left(x_{r+1}, \frac{j}{n}, u, \lambda\right) &+ \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} [F(x_{r+1} + g_t) - F(x + g_t)] \\ &+ \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} [I(\varepsilon_t \le x_{r+1}) - F(x_{r+1}) - I(\varepsilon_t \le x) + F(x)]\end{aligned}$$

and a reverse inequality with x_{r+1} replaced by x_r, where g_t denotes $g_t(u, \lambda)$ and $j = [ns]$. Therefore

$$\sup_{s \in [0,1],\, x \in R} \left|\tilde{Z}_n(x, s, u, \lambda)\right| \tag{2.27}$$

$$\le \max_r \max_j \left|\tilde{Z}_n\left(x_r, \frac{j}{n}, u, \lambda\right)\right| \tag{2.28}$$

$$+ \max_r \sup_{x \in (x_r, x_{r+1}]} \sup_s \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left|F(x_{r+1} + g_t) - F(x + g_t)\right| \tag{2.29}$$

$$\begin{aligned}+ \sup_{s,\, |t_1 - t_2| \le N^{-1}(n)} \frac{1}{\sqrt{n}} \Bigg| \sum_{t=1}^{[ns]} \Big[& I\big(\varepsilon_t \le F^{-1}(t_1)\big) \\ & - t_1 - I\big(\varepsilon_t \le F^{-1}(t_2)\big) + t_2 \Big] \Bigg|.\end{aligned} \tag{2.30}$$

By the tightness of the sequential empirical processes based on i.i.d. random variables [see Bickel and Wichura (1971)] and $N^{-1}(n) = o(1)$, we know that (2.30) converges to zero in probability. By Lemma 2.2, (2.29) also converges to zero in probability. In the following, we will show that (2.28) converges to zero.

First note that

$$\begin{aligned} &P\left(\max_r \max_j \left|\tilde{Z}_n\left(x_r, \frac{j}{n}, u, \lambda\right)\right| > \varepsilon\right) \\ &\qquad \le N(n) \max_r P\left(\max_j \left|\tilde{Z}_n\left(x_r, \frac{j}{n}, u, \lambda\right)\right| > \varepsilon\right). \end{aligned} \tag{2.31}$$

Define

$$a_{nt} = I(\varepsilon_t \le x + g_t) - F(x + g_t) - I(\varepsilon_t \le x) + F(x), \qquad 1 \le t \le n.$$

Then $S_{n,m} = \sum_{t=1}^{m} a_{nt}$ is a martingale array with respect to $\mathscr{F}_m = \sigma\{\varepsilon_t, t \le m\}$ and

$$\tilde{Z}_n\left(x, \frac{j}{n}, u, \lambda\right) = \frac{1}{\sqrt{n}} S_{n,j}. \tag{2.32}$$

By the Doob inequality, for any small $\varepsilon > 0$,

$$P\left(\max_j \left|\tilde{Z}_n\left(x, \frac{j}{n}, u, \lambda\right)\right| > \varepsilon\right) \le \varepsilon^{-4} n^{-2} E\left(S_{nn}^4\right). \tag{2.33}$$

By the Rosenthall inequality [Hall and Heyde (1980), page 23],

$$E\left(S_{nn}^4\right) \le cE\left[\sum_{t=1}^{n} E\left(a_{nt}^2 \mid \mathscr{F}_{t-1}\right)\right]^2 + c\sum_{t=1}^{n} E\left(a_{nt}^4\right), \tag{2.34}$$

for some constant c. By the assumptions of model (1.1), X_{t-1} is measureable with respect to $\mathscr{F}_{t-1}$ and hence

$$E\left(a_{nt}^2 \mid \mathscr{F}_{t-1}\right) \le |F(x+g_t) - F(x)| \le |g_t| \sup_x |f(x)|. \tag{2.35}$$

By (2.35), we have

$$\begin{aligned} E\left[\sum_{t=1}^{n} E\left(a_{nt}^2 \mid \mathscr{F}_{t-1}\right)\right]^2 &\le \left(\sup_x |f(x)|\right)^2 E\left[\sum_{t=1}^{n} |g_t|\right]^2 \\ &\le n\left(\sup_x |f(x)|\right)^2 \left[\sum_{t=1}^{n} E|g_t|^2\right] \\ &\le n\left(\sup_x |f(x)|\right)^2 (\|u\| + |\lambda|)^2 \sum_{t=1}^{n} E\left\|\delta_n^{-1} X_{t-1}\right\|^2 \\ &= O(n), \end{aligned} \tag{2.36}$$

where the last equation holds by lemma 2.1(g). Next, since $|a_{nt}| \le 2$, we have $\sum_{t=1}^{n} E(a_{nt}^4) \le 16n$. Further by (2.33), (2.34) and (2.36), we obtain

$$\begin{aligned} N(n) P\left(\max_j \left|\tilde{Z}_n\left(x, \frac{j}{n}, u, \lambda\right)\right| > \varepsilon\right) &\le N(n)\varepsilon^{-4} n^{-2} O(n) \\ &\le n^{1/2+d} \varepsilon^{-4} n^{-2} O(n) = o(1) \end{aligned}$$

for $d \in (0, 1/2)$, where $o(1)$ does not depend on x. By (2.31), (2.28) converges to zero in probability. Summarizing the discussion for (2.27)–(2.30), we complete the proof of Lemma 2.3(a).

(b) By Taylor's expansion,

$$|H_n(x, s, u)| = \frac{1}{\sqrt{n}} \left|\sum_{t=1}^{[ns]} [f(\xi_t) - f(x)]\left(u^T \delta_n^T X_{t-1}\right)\right|,$$

where ξ_t is between x and $x + u^T \delta_n^T X_{t-1}$. By Lemma 2.1(c), $\sup_{1 \le t \le n} |\xi_t - x| \le \|u\| \sup_{1 \le t \le n} \|\delta_n^T X_{t-1}\| = o_p(1)$ uniformly in x. By assumption (ii), $\sup_{1 \le t \le n} |f(\xi_t) - f(x)| = o_p(1)$ uniformly in x. Further by Lemma 2.1(e), we have

$$\sup_{s\in[0,1],\,x\in R} |H_n(x,s,u)| \le \sup_{x\in R}\sup_{1\le t\le n} |f(\xi_t)-f(x)|\|u\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\|\delta_n^T X_{t-1}\| = o_p(1).$$

This completes the proof of Lemma 2.3. □

PROOF OF THE THEOREM. Note that

$$\hat{\varepsilon}_t = \varepsilon_t - (\hat{\alpha}-\alpha)^T X_{t-1} = \varepsilon_t - \left[\delta_n^{-1}(\hat{\alpha}-\alpha)\right]^T(\delta_n^T X_{t-1}).$$

Denote $\hat{u} = \delta_n^{-1}(\hat{\alpha}-\alpha)$. Then

$$\begin{aligned}(2.37)\quad &\hat{K}_n(s,x) - K_n(s,x) - \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]} f(x)\hat{u}^T\delta_n^T X_{t-1}\\ &\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\left[I(\varepsilon_t \le x + \hat{u}^T\delta_n^T X_{t-1}) - I(\varepsilon_t\le x) - f(x)\hat{u}^T\delta_n^T X_{t-1}\right].\end{aligned}$$

To study the process $\hat{K}_n(s,x) - K(s,x) - (1/\sqrt{n})\sum_{t=1}^{[ns]} f(x)\hat{u}^T\delta_n^T X_{t-1}$, we only need to study the process

$$\begin{aligned}(2.38)\quad A_n(x,s,u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\big[&I(\varepsilon_t\le x + u^T\delta_n^T X_{t-1})\\ &-I(\varepsilon_t\le x) - f(x)u^T\delta_n^T X_{t-1}\big]\end{aligned}$$

for all $u\in R^p$ and all $x\in R$. By assumption (iii), $\hat{u} = O_p(1)$ and thus the theorem is proved if

$$(2.39)\quad \sup_{u\in D_\Delta}\sup_{s\in[0,1],\,x\in R}|A_n(x,s,u)| = o_p(1)\quad\text{for every }\Delta>0,$$

where D_Δ is defined as in Lemma 2.2. Denote

$$\begin{aligned}(2.40)\quad Z_n(x,s,u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\big[&I(\varepsilon_t\le x+u^T\delta_n^T X_{t-1})\\ &-F(x+u^T\delta_n^T X_{t-1}) - I(\varepsilon_t\le x) + F(x)\big].\end{aligned}$$

By the triangle inequality, $|A_n(x,s,u)| \le |Z_n(x,s,u)| + |H_n(x,s,u)|$, where $H_n(x,s,u)$ is defined by (2.26). Therefore, to prove (2.39), it is sufficient to show that, for every $\Delta > 0$.

$$(2.41)\quad \sup_{u\in D_\Delta}\sup_{s\in[0,1],\,x\in R}|Z_n(x,s,u)| = o_p(1)$$

and

$$(2.42)\quad \sup_{u\in D_\Delta}\sup_{s\in[0,1],\,x\in R}|H_n(x,s,u)| = o_p(1).$$

Since D_Δ is a bounded and closed region of R^p, for every $\kappa > 0$, there is a finite number of open subsets $\Delta_i(\kappa)$, $i = 1, \dots, m$, each with diameter κ, such that $\bigcup_{i=1}^m \Delta_i(\kappa) \supset D_\Delta$ and $\Delta_i(\kappa) \cap D_\Delta$ is not empty. Let u_r be any fixed point in $\Delta_r(\kappa) \cap D_\Delta$. Then for any $u \in \tilde{\Delta}_r = \Delta_r(\kappa) \cap D_\Delta$, we have

$$|g_t(u, \lambda) - g_t(u_r, \lambda)| \le \|u - u_r\| \|\delta_n^T X_{t-1}\| \le \kappa \|\delta_n^T X_{t-1}\|, \tag{2.43}$$

that is,

$$g_t(u_r, \lambda - \kappa) \le g_t(u, \lambda) \le g_t(u_r, \lambda + \kappa), \tag{2.44}$$

where $g_t(u, \lambda)$ is defined by (2.33).

Note that $Z_n(x, s, u) = \tilde{Z}_n(x, s, u, 0)$, where $\tilde{Z}_n(x, s, u, \lambda)$ is defined by (2.25). By the monotonicity of the indicator function, we obtain

$$\begin{aligned} Z_n(x, s, u) &\le \tilde{Z}_n(x, s, u_r, \kappa) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left[F(x + g_t(u_r, k)) - F(x + g_t(u, 0)) \right] \end{aligned} \tag{2.45}$$

and a reverse inequality with κ replaced by $-\kappa$, for all $u \in \tilde{\Delta}_r$. However since assumption (ii) implies that $\sup_x |f(x)| < \infty$, by the mean value theorem,

$$\begin{aligned} &\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left[F(x + g_t(u_r, \pm\kappa)) - F(x + g_t(u, 0)) \right] \right| \\ &\quad \le \sup_x |f(x)| \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} |g_t(u_r, \pm\kappa) - g_t(u, 0)| \\ &\quad \le \frac{2\kappa \sup_x |f(x)|}{\sqrt{n}} \sum_{t=1}^{n} \|\delta_n^T X_{t-1}\| = \kappa O_p(1), \end{aligned} \tag{2.46}$$

where the last equation holds by Lemma 2.1(e) and $O_p(1)$ uniformly holds for all $s \in [0, 1]$, all $x \in R$, all $u \in \tilde{\Delta}_r$ and all $r \in \{1, \dots, m\}$.

Given any small $\varepsilon > 0$ and $\eta > 0$, by (2.46), there exists a $\kappa = \kappa(\varepsilon, \eta) > 0$ such that

$$\begin{aligned} P\Bigg\{ \frac{1}{\sqrt{n}} \max_r \sup_{u \in \tilde{\Delta}_r} \sup_s \sup_x \Bigg| \sum_{t=1}^{[ns]} &\left[F(x + g_t(u_r, \pm\kappa)) \right. \\ &\left. - F(x + g_t(u, 0)) \right] \Bigg| \ge \frac{\varepsilon}{3} \Bigg\} < \eta, \end{aligned} \tag{2.47}$$

for all n. Next for the $\pm\kappa$, by Lemma 2.3(a), we can find $n_0 = n_0(\varepsilon, \eta)$ such that, for $n > n_0$,

$$\begin{aligned} &P\left\{ \max_r \sup_{s \in [0,1],\, x \in R} |\tilde{Z}_n(x, s, u_r, \pm\kappa)| \ge \frac{\varepsilon}{3} \right\} \\ &\quad \le m \max_r P\left\{ \sup_{s \in [0,1],\, x \in R} |\tilde{Z}_n(x, s, u_r, \pm\kappa)| \ge \frac{\varepsilon}{3} \right\} < \eta \end{aligned} \tag{2.48}$$

because κ is fixed and the number m of open subsets is also fixed. So when $n > n_0$, by (2.45), (2.47) and (2.48), we have

$$
\begin{aligned}
&P\Big\{\sup_{u\in D_\Delta}\ \sup_{s\in[0,1],\,x\in R} |Z_n(x,s,u)| \ge \varepsilon\Big\}\\
&\quad\le P\Big\{\max_r\ \sup_{s\in[0,1],\,x\in R} \big|\tilde{Z}_n(x,s,u_r,\kappa)\big| \ge \frac{\varepsilon}{3}\Big\}\\
&\qquad + P\Big\{\max_r\ \sup_{s\in[0,1],\,x\in R} \big|\tilde{Z}_n(x,s,u_r,-\kappa)\big| \ge \frac{\varepsilon}{3}\Big\}\\
&\qquad + P\Big\{\frac{1}{\sqrt{n}}\max_r\ \sup_{u\in\tilde{\Delta}_r}\ \sup_s\ \sup_x\Big|\sum_{t=1}^{[ns]}\big[F(x+g_t(u_r,\pm\kappa))\\
&\qquad\qquad\qquad - F(x+g_t(u,0))\big]\Big| \ge \frac{\varepsilon}{3}\Big\}\\
&\quad\le 3\eta.
\end{aligned}
\tag{2.49}
$$

So (2.41) holds.

Since $F(x)$ is a nondecreasing function, we obtain that, as $u \in \tilde{\Delta}_r$,

$$
\begin{aligned}
&H_n(x,s,u)\\
&\quad= \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\big[F\big(x+u^T\delta_n^TX_{t-1}\big)-F(x)-f(x)u^T\delta_n^TX_{t-1}\big]\\
&\quad\le \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\big[F\big(x+u_r^T\delta_n^TX_{t-1}+\kappa\big\|\delta_n^TX_{t-1}\big\|\big)\\
&\qquad\qquad -F(x)-f(x)u^T\delta_n^TX_{t-1}\big]\\
&\quad= H_n(x,s,u_r)+\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\big[F\big(x+u_r^T\delta_n^TX_{t-1}+\kappa\big\|\delta_n^TX_{t-1}\big\|\big)\\
&\qquad\qquad -F\big(x+u_r^T\delta_n^TX_{t-1}\big)+f(x)u_r^T\delta_n^TX_{t-1}\\
&\qquad\qquad -f(x)u^T\delta_n^TX_{t-1}\big]\\
&\quad\le H_n(x,s,u_r)+\frac{2\kappa\sup_x|f(x)|}{\sqrt{n}}\sum_{t=1}^{n}\big\|\delta_n^TX_{t-1}\big\|\\
&\quad= H_n(x,s,u_r)+\kappa O_p(1),
\end{aligned}
\tag{2.50}
$$

where the last equation holds by Lemma 2.1(e) and $O_p(1)$ uniformly holds for all $s \in [0,1]$, all $x \in R$, all $u \in \tilde{\Delta}_r$ and all $r \in \{1,\dots,m\}$. A reverse inequality holds as κ is replaced by $-\kappa$ in (2.50).

Given any small $\varepsilon > 0$ and $\eta > 0$, similar to (2.49), using (2.50) and Lemma 2.3(b), we can also show that there exists $n_0 = n_0(\varepsilon,\eta)$ such that, when $n > n_0$,

$$P\left\{\sup_{u\in D_\Delta}\ \sup_{x\in R,\, s\in[0,1]} |H_n(x,s,u)| \geq \varepsilon\right\} < \eta.$$

Thus (2.42) holds. This completes the proof of the theorem. □

Acknowledgments. I am very grateful to my supervisor Professor W. K. Li for his encouragements and advice. I am also very grateful to Professor J. Bai, two referees, an Associate Editor and the Editor for their very useful comments, which led to a great improvement in this paper. The proof of (2.47)–(2.49) is based on the suggestion of an Associate Editor.

Note. I was informed by a referee and an Associate Editor that the results for the residual empirical processes for the unstable processes have already appeared in the Ph.D. dissertation by Lee (1991).

REFERENCES

BAI, J. (1993). On the partial sums of residuals in autoregressive and moving average models. *J. Time. Ser. Anal.* **14** 247–260.

BAI, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.* **22** 2051–2061.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BOLDIN, M. V. (1982). Estimation of the distribution of noise in an autoregressive scheme. *Theory Probab. Appl.* **27** 866–871.

CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401.

CHUNG, K. L. (1968). *A Course in Probability Theory*. Harcourt Brace and World, New York.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, San Diego.

JEGANATHAN, P. (1991). On the asymptotic behavior of least squares estimators in AR time series with roots near the unit circle. *Econometric Theory* **7** 269–306.

KOUL, H. L. (1991). A weak convergence result useful in robust autoregression. *J. Statist. Plann. Inference* **29** 1291–1308.

KOUL, H. L. and LEVENTAL, S. (1989). Weak convergence of the residual empirical process in explosive autoregression. *Ann. Statist.* **17** 1784–1794.

KREISS, P. (1991). Estimation of the distribution of noise in stationary processes. *Metrika* **38** 285–297.

LEE, S. (1991). Testing whether a time series is Gaussian. Ph.D. dissertation, Dept. Mathematics, Univ. Maryland.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirial Processes with Applications to Statistics*. Wiley, New York.

STRAF, M. J. (1970). Weak convergence of stochastic processes with several parameters. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* 187–221. Univ. California Press, Berkeley.

DEPARTMENT OF STATISTICS
UNIVERSITY OF HONG KONG
POKFULAM ROAD
HONG KONG

DEPARTMENT OF ECONOMICS
UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS, PERTH
WESTERN AUSTRALIA