The Annals of Statistics
1994, Vol. 22, No. 4, 2051–2061

WEAK CONVERGENCE OF THE SEQUENTIAL EMPIRICAL PROCESSES OF RESIDUALS IN ARMA MODELS

BY JUSHAN BAI

Massachusetts Institute of Technology

This paper studies the weak convergence of the sequential empirical process $\widehat{K}_n$ of the estimated residuals in ARMA(p,q) models when the errors are independent and identically distributed. It is shown that, under some mild conditions, $\widehat{K}_n$ converges weakly to a Kiefer process. The weak convergence is discussed for both finite and infinite variance time series models. An application to a change-point problem is considered.

1. Introduction, notation and main results. Empirical processes based on estimated residuals have been studied by many authors for a variety of models. Koul (1969, 1984), Mukantseva (1977), Loynes (1980) and Miller (1989), for example, examined the residual empirical processes for various linear regression models. Boldin (1982, 1989), Koul and Levental (1989), Koul (1991) and Kreiss (1991) investigated their weak convergence for some ARMA(p,q) models. The literature to date has focused largely on goodness-of-fit testing. Recently, Koul (1991) demonstrated that the weak convergence result can have many important applications in robust estimation. This paper extends the above literature by considering the sequential empirical process of residuals and its weak convergence for ARMA models with an aim to test for and to identify an unknown change point.

Consider the following ARMA(p,q) time series model:

$$X_t = \rho_1 X_{t-1} + \cdots + \rho_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{1}$$

where $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) according to a distribution function F on the real line R. Assume that X_t is strictly stationary and invertible [Brockwell and Davis (1987)]. In the ARMA(1, 1) case, stationarity and invertibility restrict $|\rho_1| < 1$ and $|\theta_1| < 1$.

Given $n+p$ observations, $X_{-p+1}, X_{-p+2}, \ldots, X_0, X_1, \ldots, X_n$, one can calculate n residuals via the recursion

$$\widehat{\varepsilon}_t = X_t - \widehat{\rho}_1 X_{t-1} - \cdots - \widehat{\rho}_p X_{t-p} - \widehat{\theta}_1 \widehat{\varepsilon}_{t-1} - \cdots - \widehat{\theta}_q \widehat{\varepsilon}_{t-q}, \qquad t = 1, 2, \ldots, n, \tag{2}$$

where $(\widehat{\rho}_1, \ldots, \widehat{\rho}_q)$ and $(\widehat{\theta}_1, \ldots, \widehat{\theta}_q)$ are the estimators for $(\rho_1, \ldots, \rho_p)$ and $(\theta_1, \ldots, \theta_q)$, respectively. Let $I(A)$ be the indicator function of the event A. Define the

Received August 1991; revised July 1993.

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

Key words and phrases. Time series models, residual analysis, sequential empirical process, weak convergence, Kiefer process, change-point problem.

empirical distribution function (e.d.f.) constructed from the first $[ns]$ residuals:

$$\widehat{F}_{[ns]}(x) = \frac{1}{[ns]} \sum_{t=1}^{[ns]} I(\widehat{\varepsilon}_t \le x), \qquad 0 < s < 1, x \in R, \tag{3}$$

with $\widehat{F}_{[ns]}(\cdot) = 0$ for $s = 0$. When $s = 1$, the usual empirical process of residuals $\widehat{F}_n(x)$ is obtained. The purpose of this paper is to study the weak convergence of the process $\widehat{K}_n(s,x)$ defined as follows:

$$\widehat{K}_n(s,x) = [ns]n^{-1/2}\big(\widehat{F}_{[ns]}(x) - F(x)\big) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\widehat{\varepsilon}_t \le x) - F(x)\} \tag{4}$$

for $0 \le s \le 1$ and $x \in R$. The process K_n given by

$$K_n(s,x) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\varepsilon_t \le x) - F(x)\}$$

is called the sequential empirical process (s.e.p.); see Shorack and Wellner (1986), page 131. Thus $\widehat{K}_n$ may be called the sequential empirical process of residuals. Our main results are presented in the following two theorems.

THEOREM 1. *Assume that the following conditions hold*:

(a.1) *The ε_i are i.i.d. with zero mean, finite variance and d.f. F.*
(a.2) *F admits a uniformly continuous density function f, $f > 0$ a.e.*
(a.3) $\sqrt{n}(\widehat{\rho}_i - \rho_i) = O_p(1)$ *and* $\sqrt{n}(\widehat{\theta}_j - \theta_j) = O_p(1)$, $i = 1,\dots,p$, $j = 1,\dots,q$.

Then

$$\sup_{s \in [0,1],\, x \in R} |\widehat{K}_n(s,x) - K_n(s,x)| = o_p(1).$$

The proof of Theorem 1 is given in Section 3. From the results of Bickel and Wichura (1971), $K_n(\cdot,\cdot)$ converges weakly to a Kiefer process $K(\cdot, F(\cdot))$, a two-parameter Gaussian process with zero mean and covariance function $\mathrm{cov}(K(s_1,t_1),K(s_2,t_2)) = (s_1 \wedge s_2)(t_1 \wedge t_2 - t_1 t_2)$. Theorem 1 implies that $\widehat{K}_n$ also converges weakly to a Kiefer process. An application to a change point problem is discussed in the next section.

REMARKS. Assumption (a.1) is conventional for time series models. Assumption (a.2) is also made in Koul (1991) and is weaker than that of Boldin (1982) and Kreiss (1991). Assumption (a.3) holds with the usual estimation procedures such as the conditional least squares under (a.1).

The result of Theorem 1 holds for infinite variance ARMA models as well. We have the following result.

THEOREM 2. *Assume that the following conditions hold*:

(b.1) *The* ε_1 *are i.i.d., with d.f. F belonging to the domain of attraction of a stable law with an index* α $(0 < \alpha < 2)$.

(b.2) *The d.f. F admits a bounded derivative* f, $f > 0$ *a.e.*

(b.3) $n^{\gamma}(\widehat{\rho}_i - \rho_i) = o_p(1)$ *and* $n^{\gamma}(\widehat{\theta}_j - \theta_j) = o_p(1)$, *where* $\gamma = (1/2)I(\alpha > 1) + (1/\alpha - 1/4)I(\alpha < 1)$.

Then

$$\sup_{s \in [0,1],\, x \in R} |\widehat{K}_n(s,x) - K_n(s,x)| = o_p(1).$$

Under assumption (b.1), the estimated parameters have a faster than root n rate of convergence. Kanter and Hannan (1977) showed that, for autoregressive models, $n^{\gamma}(\widehat{\rho}_i - \rho_i) \to 0$, a.s. for any $\gamma < 1/\alpha$, where the $\widehat{\rho}_i$ are the least squares estimates. Bhansali (1988) obtained analogous results for moving average models. Using this fact as assumed in (b.3), one can prove Theorem 2 in a much similar way to the proof of Theorem 1. Details can be found in Bai (1991). Note that the uniform continuity in (a.2) is weakened to boundedness in this case.

2. An application to a change-point problem. Let $Z_1, Z_2, \dots, Z_{[n\tau]}, Z_{[n\tau]+1}, \dots, Z_n$ be n random variables. Suppose that the first $[n\tau]$ r.v.'s are i.i.d. with d.f. F_1 and the last $n - [n\tau]$ are i.i.d. with d.f. F_2, where $\tau \in (0,1)$ unknown. The objective is to test the null hypothesis (H_0) that $F_1 = F_2$. Nonparametric tests used by Picard (1985) and Carlstein (1988) are based on sequential e.d.f.'s. Let $F_{[ns]}$ and $F^*_{n-[ns]}$ be the e.d.f.'s constructed from the first $[ns]$ and the last $n - [ns]$ observations, respectively. Consider the process

$$T_n(s,x) = \sqrt{n}\frac{[ns]}{n}\left(1 - \frac{[ns]}{n}\right)\left(F_{[ns]}(x) - F^*_{n-[ns]}(x)\right)$$

and the test statistic $M_n = \sup_{s \in [0,1],\, x \in R} |T_n(s,x)|$. One rejects H_0 when M_n is too large. This test has many desirable properties as discussed in Carlstein (1988).

The result of Theorem 1 allows one to test whether there is a distributional change in the innovations ε_t. Since the ε_t are unobservable, it is natural to use the estimated residuals instead. Define

$$\widehat{T}_n(s,x) = \sqrt{n}\frac{[ns]}{n}\left(1 - \frac{[ns]}{n}\right)\left(\widehat{F}_{[ns]}(x) - \widehat{F}^*_{n-[ns]}(x)\right), \tag{5}$$

where $\widehat{F}_{[ns]}$ and $\widehat{F}^*_{n-[ns]}$ are e.d.f.'s based on the residuals. Define $\widehat{M}_n$ correspondingly. Note that T_n and $\widehat{T}_n$ can be written as $T_n(s,x) = K_n(s,x) - n^{-1}[ns]K_n(1,x)$ and $\widehat{T}_n(s,x) = \widehat{K}_n(s,x) - n^{-1}[ns]\widehat{K}_n(1,x)$, respectively. Thus Theorem 1 implies that T_n and $\widehat{T}_n$ have the same limiting null distribution. Furthermore, from

Bickel and Wichura (1971), $T_n(\cdot,\cdot)$ and hence $\widehat{T}_n(\cdot,\cdot)$ converge weakly under the null hypothesis to a Gaussian process $B(\cdot, F(\cdot))$ with zero mean and covariance function $EB(s,u)B(t,u) = (s \wedge t - st)(u \wedge v - uv)$, where F denotes $F_1 = F_2$. Accordingly, $\widehat{M}_n \to_d \sup_{0\le s\le 1} \sup_{0\le t\le 1} |B(s,t)|$ whose d.f. is tabulated in Picard (1985). Needless to say, many other tests based on $\widehat{T}_n(s,x)$, such as the Cramér–von Mises type, have the same limiting distributions as those based on $T_n(s,x)$.

3. Proofs. We prove Theorem 1 for the case of $p = 1$ and $q = 1$. The proof for general p and q and the proof of Theorem 2 are similar and can be found in Bai (1991). The proof extends some ideas of Koul and Levental (1989). Omit the subscripts on the parameters and rewrite the ARMA(1, 1) as $\varepsilon_t = X_t - \rho X_{t-1} - \theta\varepsilon_{t-1}$ and the residuals as $\widehat{\varepsilon}_t = X_t - \widehat{\rho}X_{t-1} - \widehat{\theta}\widehat{\varepsilon}_{t-1}$. Subtract the first equation from the second on both sides to obtain

$$\widehat{\varepsilon}_t - \varepsilon_t = -\widehat{\theta}(\widehat{\varepsilon}_{t-1} - \varepsilon_{t-1}) - (\widehat{\rho} - \rho)X_{t-1} - (\widehat{\theta} - \theta)\varepsilon_{t-1}. \tag{6}$$

By repeated substitution and making use of $\widehat{\varepsilon}_0 = 0$, we have

$$\begin{aligned}\widehat{\varepsilon}_t - \varepsilon_t = (-1)^{t-1}\widehat{\theta}^t\varepsilon_0 - (\widehat{\rho} - \rho)\sum_{j=0}^{t-1}(-1)^j\widehat{\theta}^jX_{t-1-j}\\ - (\widehat{\theta} - \theta)\sum_{j=0}^{t-1}(-1)^j\widehat{\theta}^j\varepsilon_{t-1-j}.\end{aligned} \tag{7}$$

Denote $\widehat{\phi} = (-\widehat{\theta}, \sqrt{n}(\widehat{\rho} - \rho), \sqrt{n}(\widehat{\theta} - \theta))$ and $\phi = (u, v, w) \in R^3$. Define

$$\Lambda_{\phi t} = u^t\varepsilon_0 + n^{-1/2}\left(v\sum_{j=0}^{t-1}u^jX_{t-1-j} + w\sum_{j=0}^{t-1}u^j\varepsilon_{t-1-j}\right) = u^t\varepsilon_0 + n^{-1/2}\xi_{\phi t}. \tag{8}$$

It follows from (7) and (8) and its definition that $\widehat{F}_{[ns]}(x)$ can be written as

$$\widehat{F}_{[ns]}(x) = \frac{1}{[ns]}\sum_{t=1}^{[ns]} I(\varepsilon_t \le x + \Lambda_{\hat{\phi}t}), \tag{9}$$

where $\Lambda_{\hat{\phi}t}$ is $\Lambda_{\phi t}$ with ϕ replaced by $\widehat{\phi}$. Thus

$$\widehat{K}_n(s,x) - K_n(s,x) = n^{-1/2}\sum_{t=1}^{[ns]}\left\{I(\varepsilon_t \le x + \Lambda_{\hat{\phi}t}) - I(\varepsilon_t \le x)\right\}. \tag{10}$$

To study the process $\widehat{K}_n(s,x) - K_n(s,x)$, it suffices to study the auxiliary process

$$G_n(s,x,\phi) = n^{-1/2}\sum_{t=1}^{[ns]}\left\{I(\varepsilon_t \le x + \Lambda_{\phi t}) - I(\varepsilon_t \le x)\right\}. \tag{11}$$

Since $|\theta| < 1$, there is $\overline{\theta} > 0$ such that $|\theta| < \overline{\theta} < 1$. Define $D_b = [-\overline{\theta}, \overline{\theta}] \times [-b, b]^2$ for $b > 0$. In view of assumption (a.3), Theorem 1 is implied by the following:

$$\sup_{\phi \in D_b} \sup_{s \in [0,1],\, x \in R} |G_n(s, x, \phi)| = o_p(1) \quad \text{for every } b > 0. \tag{12}$$

Next, define

$$\begin{aligned} Z_n(s, x, \phi) &= n^{-1/2} \sum_{t=1}^{[ns]} \{ I(\varepsilon_t \le x + \Lambda_{\phi t}) - F(x + \Lambda_{\phi t}) - I(\varepsilon_t \le x) + F(x) \}, \\ H_n(s, x, \phi) &= n^{-1/2} \sum_{t=1}^{[ns]} \{ F(x + \Lambda_{\phi t}) - F(x) \}. \end{aligned}$$

Then it is easy to see that $|G_n(s, x, \phi)| \le |Z_n(s, x, \phi)| + |H_n(s, x, \phi)|$. Therefore, to prove Theorem 1, it suffices to prove the following two propositions.

PROPOSITION 1. *Under the assumptions of Theorem* 1, *we have*

$$\sup_{\phi \in D_b} \sup_{s \in [0,1],\, x \in R} |Z_n(s, x, \phi)| = o_p(1) \quad \textit{for every } b > 0. \tag{13}$$

PROPOSITION 2. *If the assumptions in Theorem* 1 *hold*, *then*

$$\sup_{\phi \in D_b} \sup_{s \in [0,1],\, x \in R} |H_n(s, x, \phi)| = o_p(1) \quad \textit{for every } b > 0. \tag{14}$$

PROOF OF PROPOSITION 1. Let $\eta_t = C \Sigma_{j=0}^{t-1} \tau^j (|X_{t-1-j}| + |\epsilon_{t-1-j}|)$ for some $C > 0$ and $\tau \in (0, 1)$. Define for every $\lambda \in R$,

$$\begin{aligned} &\widetilde{Z}_n(s, x, \phi, \lambda) \\ &\quad = n^{-1/2} \sum_{t=1}^{[ns]} \Big\{ I\big(\varepsilon_t \le x + \Gamma_t(\phi, \lambda)\big) - F\big(x + \Gamma_t(\phi, \lambda)\big) - I(\varepsilon_t \le x) + F(x) \Big\}, \end{aligned}$$

where $\Gamma_t(\phi, \lambda) = u^t \varepsilon_0 + \lambda t \overline{\theta}^{t-1} |\varepsilon_0| + n^{-1/2} \xi_{\phi t} + \lambda n^{-1/2} \eta_t$. Since $\Gamma_t(\phi, 0) = \Lambda_{\phi t}$, it follows that $\widetilde{Z}_n(s, x, \phi, 0) = Z_n(s, x, \phi)$. As in Koul (1991), we shall argue that Proposition 1 is a consequence of the following:

$$\sup_{s \in [0,1],\, x \in R} \widetilde{Z}_n(s, x, \phi, \lambda) = o_p(1) \quad \text{for every given } \phi \text{ and } \lambda. \tag{15}$$

For any $\delta > 0$, due to its compactness, the set D_b can be partitioned into a finite number of subsets such that the diameter of each subset is not greater than δ. Denote these subsets by $\Delta_1, \Delta_2, \ldots, \Delta_{m(\delta)}$. Fix r and consider Δ_r. Pick

$\phi_r = (u_r, v_r, w_r) \in \Delta_r$. For all $\phi = (u, v, w) \in \Delta_r$, we will find an upper and a lower bound for $\Lambda_{\phi t}$ in terms of $\Lambda_{\phi_r t}$ and random variables not varying with ϕ and r. To this end, use the inequality

$$|u^j - u'^j| \le |u - u'|\, j\bar{\theta}^{j-1} \quad \text{for all } j \ge 0 \text{ if } u, u' \in [-\bar{\theta}, \bar{\theta}], \tag{16}$$

to obtain $|u^t\varepsilon_0 - u_r^t\varepsilon_0| \le \delta t\bar{\theta}^{t-1}|\varepsilon_0|$ and for $Z_t = X_t$ and ε_t to obtain

$$\left| w\sum_{j=0}^{t-1} u^j Z_{t-1-j} - w_r \sum_{j=0}^{t-1} u_r^j Z_{t-1-j}\right| \le \delta\left\{\sum_{j=0}^{t-1}\left(\bar{\theta}^j + bj\bar{\theta}^{j-1}\right)|Z_{t-1-j}|\right\}.$$

Choose $\tau \in (0,1)$ and C large enough to assure $\bar{\theta}^j + bj\bar{\theta}^{j-1} < C\tau^j$. Thus

$$|\Lambda_{\phi t} - \Lambda_{\phi_r t}| \le \delta t\bar{\theta}^{t-1}|\varepsilon_0| + \delta n^{-1/2}\eta_t \quad \text{for all } \phi \in \Delta_r. \tag{17}$$

By the monotonicity of the indicator function and inequality (17), we have

$$Z_n(s, x, \phi) \le \widetilde{Z}_n(s, x, \phi_r, \delta) + n^{-1/2}\sum_{t=1}^{[ns]}\left\{F\big(x + \Gamma_t(\phi_r, \delta)\big) - F(x + \Lambda_{\phi t})\right\}$$

and a reverse inequality with δ replaced by $-\delta$, for all $\phi \in \Delta_r$. But

$$\begin{aligned} n^{-1/2}&\left|\sum_{t=1}^{[ns]}\left\{F\big(x + \Gamma_t(\phi_r, \pm\delta)\big) - F(x + \Lambda_{\phi t})\right\}\right| \\ &\le n^{-1/2}\sum_{t=1}^{n}\left|F\big(x + \Gamma_t(\phi_r, \pm\delta)\big) - F(x + \Lambda_{\phi t})\right| \\ &\le 2\delta\|f\| n^{-1/2}\sum_{t=1}^{n}\left(t\bar{\theta}^{t-1}|\varepsilon_0| + n^{-1/2}\eta_t\right) = \delta O_p(1) \quad \text{by Lemma 1 below,} \end{aligned}$$

where the $O_p(1)$ is uniform for all $s \in [0,1]$, all $x \in R$ and all $\phi \in D_b$. Therefore,

$$\begin{aligned} \sup_{\phi \in D_b}\ \sup_{s \in [0,1],\, x \in R} |Z_n(s, x, \phi)| \le{} & \max_{r \le m(\delta)}\ \sup_{s \in [0,1],\, x \in R} |\widetilde{Z}_n(s, x, \phi_r, \delta)| \\ & + \max_{r \le m(\delta)}\ \sup_{s \in [0,1],\, x \in R} |\widetilde{Z}_n(s, x, \phi_r, -\delta)| + \delta O_p(1). \end{aligned}$$

The term $\delta O_p(1)$ can be made arbitrarily small in probability by choosing a small enough δ. Once δ is fixed, the first two terms on the right are $o_p(1)$ due to (15), thus leading to Proposition 1.

To prove (15), we need the following two lemmas.

LEMMA 1. *Under assumption* (a.1), *for every given* $\phi = (u, v, w) \in D_b$ *and* *every* $\lambda \in R$, *we have*

(a) $$n^{-1/2} \sum_{t=1}^{n} \left(|u^t \varepsilon_0| + t\bar{\theta}^{t-1} |\lambda \varepsilon_0| \right) = o_p(1),$$

(b) $$n^{-1/2} \max_{1 \le t \le n} \left(|\xi_{\phi t}| + |\lambda \eta_t| \right) = o_p(1),$$

(c) $$n^{-1} \sum_{t=1}^{n} \left(|\xi_{\phi t}| + |\lambda \eta_t| \right) = O_p(1).$$

PROOF. The proofs of (a) and (c) are trivial since $|u|$, $\bar{\theta}$ and τ all are in $(0, 1)$. Thus consider (b). From its definition, $|\xi_{\phi t}| \le b(1 - |u|)^{-1} \max_{0 \le j \le n-1}(|X_j| + |\varepsilon_j|)$ for all $t \le n$ and similarly $|\eta_t| \le C(1 - \tau)^{-1} \max_{0 \le j \le n-1}(|X_j| + |\varepsilon_j|)$ for all $t \le n$. Now (b) follows from the fact that $n^{-1/2} \max_{1 \le j \le n} |Z_j| = o_p(1)$ for arbitrary identically distributed r.v.'s $\{Z_j\}$ with finite variance [see Chung (1968), page 93]. $\square$

LEMMA 2. *For every* $d \in (0, 1/2)$, *every* $\phi = (u, v, w) \in D_b$ *and every* $\lambda \in R$,

$$\sup_{(x,y) \in B_{n,d}} n^{-1/2} \sum_{t=1}^{n} \left| F\left(y + \Gamma_t(\phi, \lambda)\right) - F\left(x + \Gamma_t(\phi, \lambda)\right) \right| = o_p(1),$$

where $B_{n,d} = \{(x, y) \in R \times R;\ |F(x) - F(y)| \le n^{-1/2 - d}\}$.

The proof of this lemma is analogous to that of Lemma 2.1 of Koul (1991) and is thus omitted. However, the use of the $n^{-1/2-d}$-grid instead of Koul's $n^{-1/2}$ is similar to Boldin (1982).

We are now in the position to prove (15). Let $N(n)$ be an integer such that $N(n) = [n^{1/2+d}] + 1$ where d is as in Lemma 2. Following the idea of Boldin (1982), we divide the real line into $N(n)$ parts by points $-\infty = x_0 < x_1 < \cdots < x_{N(n)} = \infty$ with $F(x_i) = iN(n)^{-1}$. Write Γ_t for $\Gamma_t(\phi, \lambda)$. With $x_r < x < x_{r+1}$, since $I(\varepsilon_t \le x)$ and $F(x)$ are nondecreasing, we have

$$\begin{aligned} \widetilde{Z}_n(s, x, \phi, \lambda) \le \widetilde{Z}_n(s, x_{r+1}, \phi, \lambda) &+ n^{-1/2} \sum_{t=1}^{[ns]} \left\{ F(x_{r+1} + \Gamma_t) - F(x + \Gamma_t) \right\} \\ &+ n^{-1/2} \sum_{t=1}^{[ns]} \left\{ I(\varepsilon_t \le x_{r+1}) - F(x_{r+1}) - I(\varepsilon_t \le x) + F(x) \right\} \end{aligned}$$

and a reverse inequality with x_{r+1} replaced by x_r. Therefore,

$$\sup_{s,x} |\widetilde{Z}_n(s, x, \phi, \lambda)| \le \max_r \sup_s |\widetilde{Z}_n(s, x_r, \phi, \lambda)|$$

$$+ \max_r \sup_s n^{-1/2} \left| \sum_{t=1}^{[ns]} \left\{ F(x_{r+1} + \Gamma_t) - F(x_r + \Gamma_t) \right\} \right| \tag{18}$$

$$+ \sup_{s, |g-h| \le N(n)^{-1}} n^{-1/2} \left| \sum_{t=1}^{[ns]} \left\{ I\left(\varepsilon_t \le F^{-1}(g)\right) - g - I\left(\varepsilon_t \le F^{-1}(h)\right) + h \right\} \right|. \tag{19}$$

That expression (19) is $o_p(1)$ follows from the tightness of sequential empirical processes based on i.i.d. random variables and $N(n)^{-1} = o(1)$ [Bickel and Wichura (1971)]. Convergence to 0 in probability for (18) follows from Lemma 2,

$$\max_r \sup_s n^{-1/2} \left| \sum_{t=1}^{[ns]} \{F(x_{r+1} + \Gamma_t) - F(x_r + \Gamma_t)\} \right|$$
$$\leq \max_r n^{-1/2} \sum_{t=1}^{n} |F(x_{r+1} + \Gamma_t) - F(x_r + \Gamma_t)| = o_p(1),$$

because $(x_r, x_{r+1}) \in B_{n,d}$. It remains to show

$$\max_{1 \leq r \leq N(n)} \max_{1 \leq j \leq n} \left| \widetilde{Z}_n(j/n, x_r, \phi, \lambda) \right| = o_p(1). \tag{20}$$

Notice that

$$\begin{aligned} &P\Big(\max_r \max_j \left|\widetilde{Z}_n(j/n, x_r, \phi, \lambda)\right| > \varepsilon\Big) \\ &\qquad \leq N(n) \max_r P\Big(\max_j \left|\widetilde{Z}_n(j/n, x_r, \phi, \lambda)\right| > \varepsilon\Big). \end{aligned} \tag{21}$$

We shall bound the probability in the right-hand side above. Let

$$d_{nt} = I(\varepsilon_t \leq x + \Gamma_t) - F(x + \Gamma_t) - I(\varepsilon_t \leq x) + F(x), \qquad 1 \leq t \leq n;$$
$$S_{nk} = \sum_{t=1}^{k} d_{nt}, \qquad \mathcal{F}_k = \sigma\text{-field}\{\varepsilon_i,\ i \leq k\}, \qquad 1 \leq k \leq n.$$

By construction, $\{(S_{nj}, \mathcal{F}_j);\ 1 \leq j \leq n\}$ is a martingale array and

$$\widetilde{Z}_n(j/n, x, \phi, \lambda) = n^{-1/2} S_{nj}.$$

Therefore, by the Doob inequality,

$$P\Big(\max_{1 \leq j \leq n} \left|\widetilde{Z}_n(j/n, x, \phi, \lambda)\right| > \varepsilon\Big) \leq \varepsilon^{-4} n^{-2} E(S_{nn}^4).$$

Next, by the Rosenthal inequality [Hall and Heyde (1980), page 23],

$$E(S_{nn}^4) \leq CE\left\{ \sum_{t=1}^{n} E(d_{nt}^2 \mid \mathcal{F}_{t-1}) \right\}^2 + C \sum_{t=1}^{n} E(d_{nt}^4)$$

for some $C < \infty$. Because Γ_t is measurable with respect to $\mathcal{F}_{t-1}$, we have $E(d_{nt}^2 \mid \mathcal{F}_{t-1}) \leq |F(x + \Gamma_t) - F(x)| \leq \|f\| |\Gamma_t|$, where $\|f\| = \sup_x |f(x)|$. Therefore,

$$E\left\{ \sum_{t=1}^{n} E(d_{nt}^2 \mid \mathcal{F}_{t-1}) \right\}^2 \leq \|f\|^2 E\left\{ \sum_{t=1}^{n} |\Gamma_t| \right\}^2 \leq \|f\|^2 n \sum_{t=1}^{n} E(\Gamma_t^2) \tag{22}$$

by the Cauchy–Schwarz inequality. From the definition of Γ_t,

$$|\Gamma_t| \le Mt\bar{\theta}^{t-1}|\varepsilon_0| + n^{-1/2}\{|\xi_{\phi t}| + |\lambda|\eta_t\}, \qquad t \ge 1, \text{ for some } M < \infty.$$

Moreover, it is not difficult to show that, for some $C < \infty$ and for all $t \ge 1$,

$$E(\xi_{\phi t}^2) \le C \quad \text{and} \quad E(\eta_t^2) \le C.$$

Thus $\Sigma_{t=1}^n E(\Gamma_t^2) = O(1)$ and by (22), $E\{\Sigma_{t=1}^n E(d_{nt}^2 \mid \mathcal{F}_{t-1})\}^2 = O(n)$. Next, because $|d_{nt}| \le 2$, $\Sigma_{t=1}^n E(d_{nt}^4) \le 16n$. Combining these results, we obtain

$$n^{-2}E(S_{nn}^4) = O(n^{-1}).$$

The above rate does not depend on x. Thus

$$N(n)\max_r P\Big(\max_{1 \le j \le n} |\widetilde{Z}_n(j/n, x_r, \phi, \lambda)| > \varepsilon\Big) \le \varepsilon^{-4}n^{1/2+d}O(n^{-1}) = o(1)$$

for $d \in (0, 1/2)$. The proof of (20) and thus Proposition 1 is now complete. □

PROOF OF PROPOSITION 2. Let us first show

$$\sup_{s \in [0,1],\, x \in R} |H_n(s, x, \phi)| = o_p(1) \quad \text{for every given } \phi. \tag{23}$$

Apply the mean value theorem twice to obtain

$$\begin{aligned} |H_n(s,x,\phi)| &= n^{-1/2}\left|\sum_{t=1}^{[ns]}\Big\{F(x + u^t\varepsilon_0 + n^{-1/2}\xi_{\phi t}) - F(x)\Big\}\right| \\ &\le \frac{1}{n}\left|\sum_{t=1}^{[ns]} f(\gamma_t)\xi_{\phi t}\right| + \|f\|n^{-1/2}\sum_{t=1}^{n}|u^t\varepsilon_0|, \end{aligned}$$

where γ_t is between x and $x + n^{-1/2}\xi_{\phi t}$. The second term on the right is $o_p(1)$ uniformly in s and x by Lemma 1(a). Now $\max_i|\gamma_i - x| \le n^{-1/2}\max_i|\xi_{\phi i}| = o_p(1)$ uniformly in x by Lemma 1(b). Therefore, by the uniform continuity, $f(\gamma_i) = f(x) + e_i$ with $\max_i|e_i| = o_p(1)$ uniformly in x. Thus

$$\begin{aligned} \sup_{s \in [0,1],\, x \in R} \frac{1}{n}\left|\sum_{t=1}^{[ns]} f(\gamma_t)\xi_{\phi t}\right| &\le \sup_{s \in [0,1],\, x \in R}\left(\|f\|\frac{1}{n}\left|\sum_{t=1}^{[ns]}\xi_{\phi t}\right| + \max_i|e_i|\frac{1}{n}\sum_{t=1}^{n}|\xi_{\phi t}|\right) \\ &= \sup_{s \in [0,1]}\left(\|f\|\frac{1}{n}\left|\sum_{t=1}^{[ns]}\xi_{\phi t}\right|\right) + o_p(1)O_p(1). \end{aligned}$$

It remains to show $\sup_s n^{-1}|\Sigma_{t=1}^{[ns]}\xi_{\phi t}| = o_p(1)$. However, using an invariance principle for linear processes [Billingsley (1968), page 191], one can even obtain the stronger result $\sup_s n^{-1}|\Sigma_{t=1}^{[ns]}\xi_{\phi t}| = O_p(n^{-1/2})$. Details can be found in Bai (1993).

We next argue that (23) holds uniformly in $\phi \in D_b$. Partition D_b as before and consider $\phi \in \Delta_r$. By the monotonicity of F and the inequality of (17),

$$\begin{aligned} H_n(s,x,\phi) &\le n^{-1/2} \sum_{t=1}^{[ns]} \left\{ F\left(x + \Lambda_{\phi_r t} + \delta t \bar{\theta}^{t-1} |\varepsilon_0| + \delta n^{-1/2} \eta_t\right) - F(x) \right\} \\ &\le H_n(s,x,\phi_r) + \delta \|f\| \left(n^{-1/2} \sum_{t=1}^{n} t \bar{\theta}^{t-1} |\varepsilon_0| + \frac{1}{n} \sum_{t=1}^{n} \eta_t \right), \end{aligned}$$

where the second inequality follows from the mean value theorem. A reverse inequality holds when δ is replaced by $-\delta$. Moreover, the last term in the above inequality is $\delta O_p(1)$ by Lemma 1. Therefore,

$$\sup_{s \in [0,1],\, x \in R} \ \sup_{\phi \in D_b} |H_n(s,x,\phi)| \le \max_{r \le m(\delta)} \ \sup_{s \in [0,1],\, x \in R} |H_n(s,x,\phi_r)| + \delta O_p(1),$$

which implies Proposition 2 in view of (23). □

Acknowledgments. I am grateful to the referees for their valuable comments which led to a substantial improvement in the presentation of the paper. The proof of (20) is based on the suggestions of Professor Hira Koul. I thank Professors Peter Bickel, David Brillinger, Thomas Rothenberg and James Stock for helpful comments and suggestions.

REFERENCES

BAI, J. (1991). Weak convergence of sequential empirical processes of ARMA residuals. Unpublished manuscript, Dept. Economics, Univ. California, Berkeley.

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

BHANSALI, R. J. (1988). Consistent order determination for processes with infinite variance. *J. Roy. Statist. Soc. Ser. B* **50** 46–60.

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 autoregression scheme. *Theory Probab. Appl.* **27** 866–871.

BOLDIN, M. V. (1989). On testing hypotheses in sliding average scheme by the Kolmogorov–Smirnov and ω^2 tests. *Theory Probab. Appl.* **34** 699–704.

BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series: Theory and Method*. Springer, New York.

CARLSTEIN, E. (1988). Nonparametric change point estimation. *Ann. Statist.* **16** 188–197.

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, San Diego.

KANTER, M. and HANNAN, E. J. (1977). Autoregressive processes with infinite variance. *J. Appl. Probab.* **14** 411–415.

KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40** 1950–1979.

KOUL, H. L. (1984). Tests of goodness-of-fit in linear regression. *Colloq. Math. Soc. János Bolyai* **45** 279–315.

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.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

MILLER, S. M. (1989). Empirical processes based upon residuals from errors-in-variables regressions. *Ann. Statist.* **17** 282–292.

MUKANTSEVA, L. A. (1977). Testing normality in one-dimensional and multidimensional linear regression. *Theory Probab. Appl.* **22** 591–602.

PICARD, D. (1985). Testing and estimating change-point in time series. *Adv. in Appl. Probab.* **17** 841–867.

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

DEPARTMENT OF ECONOMICS
E52-274B
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139