



Munich Personal RePEc Archive

**Weak convergence of the sequential
empirical processes of residuals in
ARMA models**

Bai, Jushan

Massachusetts Institute of Technology

August 1991

Online at <https://mpra.ub.uni-muenchen.de/32915/>
MPRA Paper No. 32915, posted 20 Aug 2011 16:50 UTC

Weak Convergence of the Sequential Empirical Processes of Residuals in ARMA models

by Jushan Bai
Massachusetts Institute of Technology

August 10, 1991

Second Revision: July 6, 1993

Abstract

This paper studies the weak convergence of the sequential empirical process \hat{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, \hat{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.

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.

Running head. Sequential Empirical of ARMA Residuals.

1 Introduction, Notations, 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:

$$(1) \quad X_t = \rho_1 X_{t-1} + \dots + \rho_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

where $\{\epsilon_t\}$ 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, $X_{-p+1}, X_{-p+2}, \dots, X_0, X_1, \dots, X_n$, one can calculate n residuals via the recursion:

$$(2) \quad \hat{\epsilon}_t = X_t - \hat{\rho}_1 X_{t-1} - \dots - \hat{\rho}_p X_{t-p} - \hat{\theta}_1 \hat{\epsilon}_{t-1} - \dots - \hat{\theta}_q \hat{\epsilon}_{t-q}, \quad t = 1, 2, \dots, n$$

where $(\hat{\rho}_1, \dots, \hat{\rho}_p)$ and $(\hat{\theta}_1, \dots, \hat{\theta}_q)$ are the estimators for (ρ_1, \dots, ρ_p) and $(\theta_1, \dots, \theta_q)$ respectively. Let $I(A)$ be the indicator function of the event A. Define the empirical distribution function (e.d.f.) constructed from the first $[ns]$ residuals:

$$(3) \quad \hat{F}_{[ns]}(x) = \frac{1}{[ns]} \sum_{t=1}^{[ns]} I(\hat{\epsilon}_t \leq x), \quad 0 < s \leq 1, x \in R$$

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

$\hat{K}_n(s, x)$ defined as follows:

$$(4) \quad \hat{K}_n(s, x) = [ns]n^{-1/2}(\hat{F}_{[ns]}(x) - F(x)) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\hat{\epsilon}_t \leq x) - F(x)\},$$

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

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

is called the sequential empirical process (s.e.p.), see Shorack and Wellner [(1986), p. 131]. Thus \hat{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) *$\sqrt{n}(\hat{\rho}_i - \rho_i) = O_p(1)$ and $\sqrt{n}(\hat{\theta}_j - \theta_j) = O_p(1)$ $i = 1, \dots, p$ $j = 1, \dots, q$.*

Then

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

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 $\text{cov}(K(s_1, t_1), K(s_2, t_2)) = (s_1 \wedge s_2)(t_1 \wedge t_2 - t_1 t_2)$. Theorem 1 implies that \hat{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

Theorem 2 *Assume that the following conditions hold:*

- (b.1) *The ϵ_i 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) $n^\gamma(\hat{\rho}_i - \rho_i) = o_p(1)$, and $n^\gamma(\hat{\theta}_j - \theta_j) = o_p(1)$,

where $\gamma = (1/2)I(\alpha > 1) + (1/\alpha - 1/4)I(\alpha < 1)$.

Then

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

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, $n^\gamma(\hat{\rho}_i - \rho_i) \rightarrow 0$, a.s. for any $\gamma < 1/\alpha$, where the $\hat{\rho}_i$ 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 (1991b). 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

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

where $\hat{F}_{[ns]}$ and $\hat{F}_{n-[ns]}^*$ are e.d.f.'s based on the residuals. Define \hat{M}_n correspondingly. Note that, T_n and \hat{T}_n can be written as $T_n(s, x) = K_n(s, x) - n^{-1}[ns]K_n(1, x)$ and $\hat{T}_n(s, x) = \hat{K}_n(s, x) - n^{-1}[ns]\hat{K}_n(1, x)$ respectively. Thus Theorem 1 implies that T_n and \hat{T}_n has the same limiting null distribution. Furthermore, from Bickel and Wichura (1971), $T_n(\cdot, \cdot)$ and hence $\hat{T}_n(\cdot, \cdot)$ converge weakly under the null hypothesis to a Gaussian process $B(\cdot, F(\cdot))$ with zero mean and covariance function $EB(s, u)B(t, v) = (s \wedge t - st)(u \wedge v - uv)$, where F denotes $F_1 = F_2$. Accordingly, $\hat{M}_n \xrightarrow{d} \sup_{0 \leq s \leq 1} \sup_{0 \leq t \leq 1} |B(s, t)|$ whose d.f. is tabulated in Picard (1985). Needless to say, many other tests based on $\hat{T}_n(s, x)$, such as the Cramer-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 (1991b). The proof extends some ideas of Koul and Levental (1989). Omit the subscripts on the parameters and rewrite the ARMA(1,1) as $\epsilon_t = X_t - \rho X_{t-1} - \theta \epsilon_{t-1}$ and the residuals as $\hat{\epsilon}_t = X_t - \hat{\rho} X_{t-1} - \hat{\theta} \hat{\epsilon}_{t-1}$. Subtract the first equation from the second on both sides to obtain

$$(6) \quad \hat{\epsilon}_t - \epsilon_t = -\hat{\theta}(\hat{\epsilon}_{t-1} - \epsilon_{t-1}) - (\hat{\rho} - \rho)X_{t-1} - (\hat{\theta} - \theta)\epsilon_{t-1}.$$

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

$$(7) \quad \hat{\epsilon}_t - \epsilon_t = (-1)^{t-1} \hat{\theta}^t \epsilon_0 - (\hat{\rho} - \rho) \sum_{j=0}^{t-1} (-1)^j \hat{\theta}^j X_{t-1-j} - (\hat{\theta} - \theta) \sum_{j=0}^{t-1} (-1)^j \hat{\theta}^j \epsilon_{t-1-j}.$$

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

$$(8) \quad \Lambda_{\phi t} = u^t \epsilon_0 + n^{-1/2} \left(v \sum_{j=0}^{t-1} u^j X_{t-1-j} + w \sum_{j=0}^{t-1} u^j \epsilon_{t-1-j} \right) = u^t \epsilon_0 + n^{-1/2} \xi_{\phi t}.$$

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

$$(9) \quad \hat{F}_{[ns]}(x) = \frac{1}{[ns]} \sum_{t=1}^{[ns]} I(\epsilon_t \leq x + \Lambda_{\hat{\phi}t}),$$

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

$$(10) \quad \hat{K}_n(s, x) - K_n(s, x) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\epsilon_t \leq x + \Lambda_{\hat{\phi}t}) - I(\epsilon_t \leq x)\},$$

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

$$(11) \quad G_n(s, x, \phi) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\epsilon_t \leq x + \Lambda_{\phi t}) - I(\epsilon_t \leq x)\}.$$

Since $|\theta| < 1$, there is $\bar{\theta} > 0$ such that $|\theta| < \bar{\theta} < 1$. Define $D_b = [-\bar{\theta}, \bar{\theta}] \times [-b, b]^2$ for $b > 0$. In view of assumption (a.4), Theorem 1 is implied by the following

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

Next, define

$$Z_n(s, x, \phi) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\epsilon_t \leq x + \Lambda_{\phi t}) - F(x + \Lambda_{\phi t}) - I(\epsilon_t \leq x) + F(x)\},$$

$$H_n(s, x, \phi) = n^{-1/2} \sum_{t=1}^{[ns]} \{F(x + \Lambda_{\phi t}) - F(x)\}.$$

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

Proposition 1 *Under the assumptions of Theorem 1, we have*

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

Proposition 2 *If the assumptions in Theorem 1 hold, then*

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

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

$$\tilde{Z}_n(s, x, \phi, \lambda) = n^{-1/2} \sum_{t=1}^{[ns]} \{I(\epsilon_t \leq x + \Gamma_t(\phi, \lambda)) - F(x + \Gamma_t(\phi, \lambda)) - I(\epsilon_t \leq x) + F(x)\},$$

where $\Gamma_t(\phi, \lambda) = u^t \epsilon_0 + \lambda t \bar{\theta}^{t-1} |\epsilon_0| + n^{-1/2} \xi_{\phi t} + \lambda n^{-1/2} \eta_t$. Since $\Gamma_t(\phi, 0) = \Lambda_{\phi t}$, it follows that $\tilde{Z}_n(s, x, \phi, 0) = Z_n(s, x, \phi)$. As in Koul (1991), we shall argue that Proposition 1 is a consequence of the following:

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

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 $\Delta_1, \Delta_2, \dots, \Delta_{m(\delta)}$. 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

$$(16) \quad |u^j - u'^j| \leq |u - u'| j \bar{\theta}^{j-1} \quad \text{for all } j \geq 0, \text{ if } u, u' \in [-\bar{\theta}, \bar{\theta}],$$

to obtain $|u^t \epsilon_0 - u_r^t \epsilon_0| \leq \delta t \bar{\theta}^{t-1} |\epsilon_0|$ and for $Z_t = X_t$ and ϵ_t to obtain

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

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

$$(17) \quad |\Lambda_{\phi t} - \Lambda_{\phi_r t}| \leq \delta t \bar{\theta}^{t-1} |\epsilon_0| + \delta n^{-1/2} \eta_t, \quad \text{for all } \phi \in \Delta_r.$$

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

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

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

$$\begin{aligned} & n^{-1/2} \left| \sum_{t=1}^{[ns]} \{F(x + \Gamma_t(\phi_r, \pm \delta)) - F(x + \Lambda_{\phi t})\} \right| \\ & \leq n^{-1/2} \sum_{t=1}^n |F(x + \Gamma_t(\phi_r, \pm \delta)) - F(x + \Lambda_{\phi t})| \\ & \leq 2\delta \|f\| n^{-1/2} \sum_{t=1}^n (t \bar{\theta}^{t-1} |\epsilon_0| + n^{-1/2} \eta_t) = \delta O_p(1), \quad \text{by Lemma 1 below} \end{aligned}$$

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

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

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

$$\begin{aligned} \text{(a)} \quad &n^{-1/2} \sum_{t=1}^n (|u^t \epsilon_0| + t \bar{\theta}^{t-1} |\lambda \epsilon_0|) = o_p(1). \\ \text{(b)} \quad &n^{-1/2} \max_{1 \leq t \leq n} (|\xi_{\phi t}| + |\lambda \eta_t|) = o_p(1), \\ \text{(c)} \quad &n^{-1} \sum_{t=1}^n (|\xi_{\phi t}| + |\lambda \eta_t|) = O_p(1), \end{aligned}$$

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, $|\xi_{\phi t}| \leq b(1 - |u|)^{-1} \max_{0 \leq j \leq n-1} (|X_j| + |\epsilon_j|)$ for all $t \leq n$ and similarly $|\eta_t| \leq C(1 - \tau)^{-1} \max_{0 \leq j \leq n-1} (|X_j| + |\epsilon_j|)$ for all $t \leq n$. Now (b) follows from the fact that $n^{-1/2} \max_{1 \leq j \leq n} |Z_j| = o_p(1)$ for arbitrary identically distributed *r.v.'s* $\{Z_j\}$ with finite variance [see Chung (1968), p. 93]. \square

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

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

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

The proof of this lemma is analogous to that of Lemma 2.1 of Koul (1991) and is thus omitted. However, the use of $n^{-1/2-d}$ -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 $N(n) = \lceil n^{1/2+d} \rceil + 1$ where d is as in Lemma 2. Following the idea of Boldin (1982), we divide the real line into $N(n)$ parts by points $-\infty = x_0 < x_1 < \dots < x_{N(n)} = \infty$ with $F(x_i) = iN(n)^{-1}$. Write Γ_t for $\Gamma_t(\phi, \lambda)$. When $x_r < x < x_{r+1}$, since $I(\epsilon_t \leq x)$ and $F(x)$ are nondecreasing, we have

$$\begin{aligned} \tilde{Z}_n(s, x, \phi, \lambda) &\leq \tilde{Z}_n(s, x_{r+1}, \phi, \lambda) + n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} \{F(x_{r+1} + \Gamma_t) - F(x + \Gamma_t)\} \\ &\quad + n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} \{I(\epsilon_t \leq x_{r+1}) - F(x_{r+1}) - I(\epsilon_t \leq x) + F(x)\} \end{aligned}$$

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

$$\begin{aligned} \sup_{s,x} |\tilde{Z}_n(s, x, \phi, \lambda)| &\leq \max_r \sup_s |\tilde{Z}_n(s, x_r, \phi, \lambda)| \\ (18) \quad &+ \max_r \sup_s n^{-1/2} \left| \sum_{t=1}^{\lfloor ns \rfloor} \{F(x_{r+1} + \Gamma_t) - F(x_r + \Gamma_t)\} \right| \\ (19) \quad &+ \sup_{s, |g-h| \leq N(n)^{-1}} n^{-1/2} \left| \sum_{t=1}^{\lfloor ns \rfloor} \{I(\epsilon_t \leq F^{-1}(g)) - g - I(\epsilon_t \leq F^{-1}(h)) + h\} \right|. \end{aligned}$$

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 zero in probability for (18) follows from Lemma 2,

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

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

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

Notice that

$$(21) \quad P(\max_r \max_j |\tilde{Z}_n(j/n, x_r, \phi, \lambda)| > \epsilon) \leq N(n) \max_r P(\max_j |\tilde{Z}_n(j/n, x_r, \phi, \lambda)| > \epsilon).$$

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

$$d_{nt} = I(\epsilon_t \leq x + \Gamma_t) - F(x + \Gamma_t) - I(\epsilon_t \leq x) + F(x), \quad 1 \leq t \leq n;$$

$$S_{nk} = \sum_{t=1}^k d_{nt}, \quad \mathcal{F}_k = \sigma - \text{field}\{\epsilon_i, i \leq k\}, \quad 1 \leq k \leq n.$$

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

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

Therefore by the Doob inequality,

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

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

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

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

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

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

$$|\Gamma_t| \leq Mt\bar{\theta}^{t-1} |\epsilon_0| + n^{-1/2} \{|\xi_{\phi t}| + |\lambda|\eta_t\}, \quad t \geq 1, \quad \text{for some } M < \infty.$$

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

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

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

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

The above rate does not depend on x . Thus

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

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

Proof of Proposition 2. Let us first show

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

Apply the Mean Value Theorem twice to obtain

$$\begin{aligned} |H_n(s, x, \phi)| &= n^{-1/2} \left| \sum_{t=1}^{[ns]} \{F(x + u^t \epsilon_0 + n^{-1/2} \xi_{\phi t}) - F(x)\} \right| \\ &\leq \frac{1}{n} \left| \sum_{t=1}^{[ns]} f(\gamma_t) \xi_{\phi t} \right| + \|f\| n^{-1/2} \sum_{t=1}^n |u^t \epsilon_0|, \end{aligned}$$

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 $\max_i |\gamma_i - x| \leq n^{-1/2} \max_i |\xi_{\phi i}| = o_p(1)$ 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

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

It remains to show $\sup_s n^{-1} \left| \sum_{t=1}^{[ns]} \xi_{\phi t} \right| = o_p(1)$. However, using an invariance principle for linear processes [Billingsley (1968), p. 191], one can even obtain the stronger result $\sup_s n^{-1} \left| \sum_{t=1}^{[ns]} \xi_{\phi t} \right| = O_p(n^{-1/2})$. Details can be found in Bai (1991a).

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)

$$\begin{aligned} H_n(s, x, \phi) &\leq n^{-1/2} \sum_{t=1}^{[ns]} \{F(x + \Lambda_{\phi, t} + \delta t \bar{\theta}^{t-1} |\epsilon_0| + \delta n^{-1/2} \eta_t) - F(x)\} \\ &\leq H_n(s, x, \phi_r) + \delta \|f\| \left(n^{-1/2} \sum_{t=1}^n t \bar{\theta}^{t-1} |\epsilon_0| + \frac{1}{n} \sum_{t=1}^n \eta_t \right) \end{aligned}$$

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,

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

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

- [1] Bai, J. (1991a). On the partial sums of residuals in autoregressive and moving average models. *J. of Time Series Analysis*. To appear.
- [2] Bai, J. (1991b). Weak convergence of sequential empirical processes of ARMA residuals. Manuscript, Department of Economics, U.C. Berkeley.
- [3] Bhansali, R.J. (1988). Consistent order determination for processes with infinite variance. *J. R. Stat. Society*, 50 46-60.
- [4] Bickel, P.J. and Wichura, M.J. (1971). Convergence for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* 42 1656-1670.
- [5] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [6] Boldin, M.V. (1982). Estimation of the distribution of noise in an autoregression scheme. *Theory Probab. Appl.* 27 866-871.
- [7] 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.
- [8] Brockwell, P.J. and Davis, R.A. (1986). *Time Series: Theory and Method*. Springer-Verlag, New York.
- [9] Chung, K.L. (1968). *A Course in Probability Theory*. Harcourt, Brace & World, Inc.

- [10] Carlstein, E. (1988). Nonparametric change point estimation. *Ann. Statist.*, 16, 188-197.
- [11] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Applications*. Academic Press, San Diego.
- [12] Kanter, M. and Hannan, E.J. (1977). Autoregressive processes with infinite variance. *J. Applied Prob.* 14 411-415.
- [13] Koul, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* 40, 1950-1979.
- [14] Koul, H. L. (1984). Tests of goodness-of-fit in linear regression. *Colloquia Mathematica Societatis Janos Bolyai.*, 45. *Goodness of fit*, Debrecen, Hungary. 279-315.
- [15] Koul, H. L. (1991). A weak convergence result useful in robust autoregression. *Statist. Plann. & Inference*, 29, 1291-308.
- [16] Koul, H.L. and Levental, S. (1989). Weak convergence of the residual empirical process in explosive autoregression. *Ann. Statist.* 17 1784-1794.
- [17] Kreiss, P. (1991). Estimation of the distribution of noise in stationary processes. *Metrika*, 38, 285-297.
- [18] Loynes, R.M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* 8 285-298.
- [19] Miller, S. M. (1989). Empirical processes based upon residuals from errors-in-variables regressions. *Ann. Statist.* 17 282-292.
- [20] Mukantseva, L.A. (1977). Testing normality in one-dimensional and multidimensional linear regression. *Theory Prob. Appl.* 22 591-602.
- [21] Picard, D. (1985). Testing and estimating change-point in time series. *Adv. Appl. Prob.* 17 841-867.

- [22] 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, MA 02139