ON THE PARTIAL SUMS OF RESIDUALS IN AUTOREGRESSIVE AND MOVING AVERAGE MODELS

By Jushan Bai

University of California

First version received May 1991

Abstract. The limiting process of partial sums of residuals in stationary and invertible autoregressive moving-average models is studied. It is shown that the partial sums converge to a standard Brownian motion under the assumptions that estimators of unknown parameters are root-n consistent and that innovations are independent and identically distributed random variables with zero mean and finite variance or, more generally, are martingale differences with moment restrictions specified in Theorem 1. Applications for goodness-of-fit and change-point problems are considered. The use of residuals for constructing nonparametric density estimation is discussed.

Keywords. Time series models; residual analysis; Brownian motion; change-point problem; nonparametric density estimation.

1. INTRODUCTION

Kulperger (1985) investigated the limiting process for partial sums of the residuals in autoregressive AR(p) models showing that the partial sums converge weakly to a standard Brownian motion. This paper extends this and other related results to an important class of time series models, namely autogressive moving-average (ARMA) models. Moreover, results are obtained under weaker conditions. More specifically, it is assumed that the innovations which drive the ARMA process have finite variance (for independent and identically distributed (i.i.d.) case) instead of finite fourth moment. Also, the estimators of unknown parameters are root-n consistent; they are not necessarily from least squares estimation. With the least squares estimators for AR models, Kulperger was able to use the Skorohod representation technique to prove his theorems. In ARMA time series models, however, the residuals depend on the estimated parameters in a more complex way; therefore the Skorohod representation technique seems difficult to apply. Consequently, other techniques have to be used. It turns out that some elementary arguments lead to the desired results. In addition, the weak convergence of the residual process is proved for martingale-difference innovations.

In a separate paper, Kulperger (1987) examined the residual process of regression models with autoregressive errors. He also considered a special case of ARMA (1, 1) errors with the moving-average parameter being equal

0143-9782/93/03 247–60 JOURNAL OF TIME SERIES ANALYSIS Vol. 14, No. 3

to one (non-invertible model). Such a model yields an interesting result in that different initial values used in calculating the residuals can lead to different limiting processes. This is in contrast with the results in this paper. As will be seen, the limiting process is not affected by the initial values.

Residual partial sum processes and their applications have been studied for a variety of models. MacNeill (1978a,b) studied the residual processes of polynomial regression models and of general linear regression models, and proposed some testing statistics for parameter changes at unknown times. The residual partial sums for nonlinear regression models were studied by MacNeill and Jandhyala (1985). Recently, Jandhyala and MacNeill (1989) obtained the limit processes for partial sums of linear functions of regression residuals. They also derived asymptotic forms of some testing statistics for parameter changes.

An important topic in time series analysis is to determine the adequacy of a fitted model. Many tests have been developed to perform such tasks. Most are based on residuals, such as the Box–Pierce portmanteau test, the turning-point and the rank test (for a brief summary see Brockwell and Davis, 1986, pp. 296–304). The results of this paper allow one to use some goodness-of-fit test statistics based on the partial sums.

This paper is organized as follows. Section 2 introduces notations and assumptions and states the main results. Section 3 proves the weak convergence result for the ARMA (1, 1) model and then generalizes the proof to ARMA (p, q) models. Some analytical properties of the power series expansion of the reciprocal of the associated polynomials are crucial to the generalization. These properties are derived in Lemma 1 and may be of independent interest. Statistical applications to the change-point problem and density estimation are considered in the final section.

2. NOTATIONS, ASSUMPTIONS AND RESULTS

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

$$X_t = \rho_1 X_{t-1} + \ldots + \rho_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \qquad (t = \ldots, -1, 0, 1, \ldots),$$

where $\rho_1, \ldots, \rho_p$ and $\theta_1, \ldots, \theta_q$ are unknown parameters, $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with zero mean and finite variance or, more generally, $\{\varepsilon_t\}$ is a weakly stationary martingale-difference sequence (m.d.s.) with uniformly bounded $2+\delta$ moment ($\delta > 0$). We also assume that the roots of $1 - \rho_1 z - \rho_2 z^2 - \ldots - \rho_p z^p = 0$ and the roots of $1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q = 0$ all lie outside the unit circle, and that the two polynomials (referred to as the associated polynomials) have no common roots, so that the process is stationary (weakly stationary unless i.i.d. innovations) and invertible. In the ARMA (1, 1) case this requires $|\rho_1| < 1$ and $|\theta_1| < 1$.

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

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

where $\hat{\rho}_1, \ldots, \hat{\rho}_p$ and $\hat{\theta}_1, \ldots, \hat{\theta}_q$ are the estimators of $\rho_1, \ldots, \rho_p$ and $\theta_1, \ldots, \theta_q$ respectively. The initial value of $(\hat{\varepsilon}_{-q}, \ldots, \hat{\varepsilon}_0)$ is set to the null vector as is commonly done in practice. The choice of initial values is inconsequential to our results. It could be estimated from the data or generated according to a random law (see the remarks at the end of Section 3).

In what follows $o_p(1)$ ($O_p(1)$) represents a sequence of random variables which converges to zero (remains bounded) in probability, and $[x]$ denotes the greatest integer function of x. Moreover, R^k will denote the k-dimensional Euclidean space with $k \geq 1$; $R = R^1$; and $\|u\|_1 = \sum_{i=1}^k |u_i|$ for $u \in R^k$.

Let

$$B^{(n)}(x) = \frac{1}{\sigma n^{1/2}} \sum_{t=1}^{[nx]} \varepsilon_t \qquad \hat{B}^{(n)}(x) = \frac{1}{\sigma n^{1/2}} \sum_{t=1}^{[nx]} \hat{\varepsilon}_t \qquad 0 \leq x \leq 1,$$

so that $B^{(n)}(x)$ and $\hat{B}^{(n)}(x)$ are the partial sums based on the innovations and on the residuals respectively.

THEOREM 1. *Assume that the following conditions hold.*
(a.1) The ε_t are i.i.d. with zero mean and variance σ^2, or
(a.1′) the ε_t are martingale differences satisfying

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0 \qquad E\varepsilon_t^2 = \sigma^2 \qquad n^{-1} \sum_{t=1}^{n} E(\varepsilon_t^2 | \mathcal{F}_{t-1}) \to \sigma^2,$$

and, for some $\delta > 0$,

$$\sup_t E|\varepsilon_t|^{2+\delta} < \infty,$$

where $\mathcal{F}_t$ is the σ-field generated by ε_s, $s \leq t$.
(a.2)

$$n^{1/2}(\hat{\rho}_i - \rho_i) = O_p(1) \qquad i = 1, \ldots, p$$
$$n^{1/2}(\hat{\theta}_i - \theta_i) = O_p(1) \qquad i = 1, \ldots, q.$$

Then

$$\sup_{0 \leq x \leq 1} |\hat{B}^{(n)}(x) - B^{(n)}(x)| = o_p(1).$$

The proof of Theorem 1 is given in the next section. Note that condition (a.1′) allows the conditional variance of ε_t to be non-constant.

It is well known that under (a.1) or (a.1′) $B^{(n)}(x)$ converges weakly to

$B(x)$, where $B(x)$ is a standard Brownian motion on $[0, 1]$ and the convergence is in the sense of $D[0, 1]$, the space of right continuous real-valued functions on $[0, 1]$ endowed with the Skorohod J_1 topology. See, for example, Theorem 16.1 of Billingsley (1968) for the i.i.d. case and Theorem 3.2 of McLeish (1974) for the m.d.s. case (with weaker assumptions). Therefore, under the conditions of Theorem 1, $\hat{B}^{(n)}(x)$ also converges weakly to a standard Brownian motion. Consider as an application the problem of goodness-of-fit. A test for this problem could be based on

$$T_n = \max_{1 \le t \le n} \left| \frac{1}{\sigma n^{1/2}} \sum_{j=1}^{t} \hat{\varepsilon}_j \right| = \sup_{0 \le x \le 1} \frac{\sigma}{\hat{\sigma}} |\hat{B}^{(n)}(x)|,$$

where $\hat{\sigma}$ is a consistent estimator of σ, such as

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t^2.$$

If the conditions of Theorem 1 are satisfied, then T_n converges in distribution to $\sup_{0 \le x \le 1} |B(x)|$. High values of T_n may indicate inadequacy of the fitted model and suggest modifications. Other applications will be considered in the final section.

3. PROOF OF THEOREM 1

First consider the case with $p = 1$ and $q = 1$. The general case will be considered later. Omit the subscripts on the parameters; then (1) becomes

$$\hat{\varepsilon}_t = X_t - \hat{\rho} X_{t-1} - \hat{\theta} \hat{\varepsilon}_{t-1}.$$

Rewrite the ARMA (1, 1) model correspondingly,

$$\varepsilon_t = X_t - \rho X_{t-1} - \theta \varepsilon_{t-1},$$

and then subtract the second equation from the first to obtain

$$\hat{\varepsilon}_t - \varepsilon_t = - \hat{\theta}(\hat{\varepsilon}_{t-1} - \varepsilon_{t-1}) - (\hat{\rho} - \rho) X_{t-1} - (\hat{\theta} - \theta) \varepsilon_{t-1}. \tag{2}$$

Notice the recursiveness of $\hat{\varepsilon}_t - \varepsilon_t$. By repeated substitution and use of $\hat{\varepsilon}_0 = 0$, we have

$$\hat{\varepsilon}_t - \varepsilon_t = (-1)^{t-1} \hat{\theta}^t \varepsilon_0 - (\hat{\rho} - \rho) \sum_{j=0}^{t-1} (-1)^j \hat{\theta}^j X_{t-1-j} - (\hat{\theta} - \theta) \sum_{j=0}^{t-1} (-1)^j \hat{\theta}^j \varepsilon_{t-1-j}. \tag{3}$$

There are alternative expressions for relating $\hat{\varepsilon}_t$ and ε_t, but (3) is easier to work with. From (3) we have

$$\begin{aligned}\hat{B}^{(n)}(x) - B^{(n)}(x) &= \frac{1}{\sigma n^{1/2}} \sum_{t=1}^{[nx]} (-1)^{t-1} \hat{\theta}^t \varepsilon_0 \\ &\quad - n^{1/2} (\hat{\rho} - \rho) \frac{1}{\sigma n} \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} (-1)^j \hat{\theta}^j X_{t-1-j} \\ &\quad - n^{1/2} (\hat{\theta} - \theta) \frac{1}{\sigma n} \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} (-1)^j \hat{\theta}^j \varepsilon_{t-1-j}. \qquad (4)\end{aligned}$$

Since $|\theta| < 1$, there is a $\bar{\theta} > 0$ such that $|\theta| < \bar{\theta} < 1$. By assumption (a.2), $P(\hat{\theta} \notin [-\bar{\theta}, \bar{\theta}]) \to 0$. Consequently, to prove Theorem 1, it suffices to show that

$$\sup_{|u| \le \bar{\theta}} \sup_{0 \le x \le 1} \frac{1}{n^{1/2}} \sum_{t=1}^{[nx]} |u^t \varepsilon_0| = o_p(1), \qquad (5)$$

$$\sup_{|u| \le \bar{\theta}} \sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} u^j X_{t-1-j} \right| = o_p(1), \qquad (6)$$

$$\sup_{|u| \le \bar{\theta}} \sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} u^j \varepsilon_{t-1-j} \right| = o_p(1). \qquad (7)$$

We shall prove (6) only, since (5) is obvious and (7) is simpler than (6). The proof of (6) is completed in (22). The main arguments are (8) and (20). Let us first show that for each $u \in [-\bar{\theta}, \bar{\theta}]$,

$$\sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} u^j X_{t-1-j} \right| = o_p(1). \qquad (8)$$

Notice that

$$\frac{1}{n} \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} u^j X_{t-1-j} = \frac{1}{n} \sum_{t=1}^{[nx]} \sum_{j=0}^{\infty} u^j X_{t-1-j} - \frac{1}{n} \sum_{t=1}^{[nx]} \sum_{j=t}^{\infty} u^j X_{t-1-j}. \qquad (9)$$

The first term on the right-hand side consists of partial sums of a stationary process which are more convenient to work with, and the second is $o_p(1)$ uniformly in x since

$$E\left(\sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=t}^{\infty} u^j X_{t-1-j} \right| \right) \le E|X_1| \frac{1}{n} \sum_{t=1}^{n} \sum_{j=t}^{\infty} |u|^j \to 0.$$

Consequently, we need to be concerned with the first term on the right-hand side of (9). However,

$$X_{t-1-j} = \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-1-i-k} + \theta \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-2-j-k} = \xi_{t,j}^{(1)} + \xi_{t,j}^{(2)}, \qquad (10)$$

say. Thus, to prove (8), the problem is further reduced to showing that

$$\sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=0}^{\infty} u^j \xi_{t,j}^{(r)} \right| = o_p(1) \qquad \text{for } r = 1, 2. \qquad (11)$$

In what follows, we will consider the case for $r = 1$; the proof for $r = 2$ is the same.

Let

$$\eta_t := \sum_{j=0}^{\infty} u^j \xi_{t,j}^{(1)} = \sum_{j=0}^{\infty} u^j \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-1-j-k} = \sum_{k=0}^{\infty} a_k \varepsilon_{t-1-k} \qquad t \geq 1, \tag{12}$$

where $a_k = \sum_{j=0}^{k} u^j \rho^{k-j}$. Since

$$|a_k| = \left| \sum_{j=0}^{k} u^j \rho^{k-j} \right| \leq (k+1)\gamma^k \qquad \gamma = \max(|u|, |\rho|) < 1, \tag{13}$$

it is straightforward to show that

$$\sum_{k=1}^{\infty} \left(\sum_{l>k} a_l^2 \right)^{1/2} < \infty \tag{14}$$

and

$$\sum_{k=1}^{\infty} \left| \sum_{l>k} a_l \right| < \infty. \tag{15}$$

Condition (a.1) and (14) allow us to invoke Theorem 21.1 of Billingsley (1968, pp. 184, 191), which asserts that

$$\frac{1}{n^{1/2}} \sum_{t=1}^{[nx]} \eta_t \Rightarrow \sigma_1 B(x), \tag{16}$$

where

$$\sigma_1^2 = E(\eta_1^2) + 2E \sum_{k=2}^{\infty} (\eta_1 \eta_k) = \sigma^2 \left(\sum_{k=0}^{\infty} a_k \right)^2$$

and '$\Rightarrow$' denotes the weak convergence. Since $\sup_{0 \leq x \leq 1} |B(x)| = O_p(1)$, we see that the left-hand side of (11) is $O_p(n^{-1/2})$, which is stronger that we need. The result of (16) holds under (a.1′) and (15). The proof appears in Phillips and Solo (1989). For completeness, their proof will be sketched here. Write $\tilde{a}_j = \sum_{k \geq j} a_k$ and $\tilde{\varepsilon}_t = \sum_{j=0}^{\infty} \tilde{a}_j \varepsilon_{t-j}$; then algebra shows that

$$\eta_t := \sum_{k=0}^{\infty} a_k \varepsilon_{t-1-k} = \tilde{a}_0 \varepsilon_{t-1} - \tilde{\varepsilon}_t + \tilde{\varepsilon}_{t-1}.$$

Thus

$$\frac{1}{n^{1/2}} \left| \sum_{t=1}^{[nx]} \eta_t - \tilde{a}_0 \sum_{t=1}^{[nx]} \varepsilon_{t-1} \right| \leq \frac{1}{n^{1/2}} |\tilde{\varepsilon}_0| + \frac{1}{n^{1/2}} |\tilde{\varepsilon}_{[nx]}|.$$

It is easy to see that the first term on the right-hand side is $o_p(1)$. Thus (16) follows if

$$\max_{1 \leq r \leq n} \frac{1}{n^{1/2}} |\tilde{\varepsilon}_r| = o_p(1).$$

The above holds if the following holds (Hall and Heyde, 1980, p. 53):

$$\frac{1}{n} \sum_{k=1}^{n} E \tilde{\varepsilon}_k^2 I(\tilde{\varepsilon}_k^2 > n\varepsilon) \to 0 \qquad \text{for any } \varepsilon > 0, \tag{17}$$

where $I(\cdot)$ denotes the indicator function. The left-hand side is bounded by $(n\varepsilon)^{-\delta/2}\{(1/n)\sum_{k=1}^{n} E|\tilde{\varepsilon}_k|^{2+\delta}\}$. By applying Minkowski's inequality to $E|\tilde{\varepsilon}_k|^{2+\delta}$ and utilizing (15) and $\sup_t|\varepsilon_t|^{2+\delta} < \infty$, we can readily deduce the boundedness of $(1/n)\sum_{k=1}^{n} E|\tilde{\varepsilon}_k|^{2+\delta}$. Since $(n\varepsilon)^{-\delta/2} \to 0$, (17) is proved. In summary, we have proved (16) under (a.1) or (a.1'), which implies (11), which in turn implies (8). This completes the first step toward (6).

To conserve space, we define

$$Z_n(x, u) = \frac{1}{n} \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} u^j X_{t-1-j}, \tag{18}$$

so that (8) is equivalent to

$$\sup_{0 \le x \le 1} |Z_n(x, u)| = o_p(1) \qquad \text{for each } u \in [-\bar{\theta}, \bar{\theta}]. \tag{19}$$

Thus, to finish the proof, it suffices to show that for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\limsup_n P\left\{ \sup_{0 \le x \le 1} \sup_{[\delta]} |Z_n(x, u) - Z_n(x, u')| > \varepsilon \right\} < \varepsilon \tag{20}$$

where $[\delta] = \{(u, u');\ u, u' \in [-\bar{\theta}, \bar{\theta}],\ |u - u'| \le \delta\}$. In fact, if we fix a $\delta > 0$, choose a partition $-\bar{\theta} = u_0 < u_1 < \ldots < u_r = \bar{\theta}$ of $[-\bar{\theta}, \bar{\theta}]$ such that $\max_{1 \le i \le r}(u_i - u_{i-1}) \le \delta$, then, for every $u \in [-\bar{\theta}, \bar{\theta}]$, there exists an i such that $u \in [u_{i-1}, u_i]$ and thus $|u - u_i| \le \delta$. Consequently,

$$\begin{aligned} |Z_n(x, u)| &\le |Z_n(x, u_i)| + |Z_n(x, u) - Z_n(x, u_i)| \\ &\le |Z_n(x, u_i)| + \sup_{[\delta]} |Z_n(x, u) - Z_n(x, u')|. \end{aligned}$$

Hence

$$\begin{aligned} \sup_{|u| \le \bar{\theta}} \sup_{0 \le x \le 1} |Z_n(x, u)| \le{} & \max_{1 \le i \le r} \sup_{0 \le x \le 1} |Z_n(x, u_i)| \\ & + \sup_{0 \le x \le 1} \sup_{[\delta]} |Z_n(x, u) - Z_n(x, u')|. \end{aligned} \tag{21}$$

Combining (19), (20) and (21), we obtain

$$\sup_{|u| \le \bar{\theta}} \sup_{0 \le x \le 1} |Z_n(x, u)| = o_p(1), \tag{22}$$

which is (6), recalling the definition of $Z_n(x, u)$ in (18).

To prove (20), we use the inequality

$$|u^j - u'^j| \le \delta j \bar{\theta}^{j-1} \qquad \text{for } j \ge 0, \text{ if } u, u' \in [\delta], \tag{23}$$

and obtain

$$|Z_n(x, u) - Z_n(x, u')| \le \delta \frac{1}{n} \sum_{t=1}^{n} \sum_{j=0}^{t-1} j \bar{\theta}^{j-1} |X_{t-1-j}| = \delta O_p(1)$$

for all $u, u' \in [\delta]$ and all $x \in [0, 1]$. We now concludes that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that (20) holds. The proof of Theorem 1 is complete. ∎

REMARKS. In calculating the residuals, the initial value of $\hat{\varepsilon}_0$ is assumed to be zero. It can be set to any real number or to a realization of a random variable (r.v.) without affecting the result of Theorem 1. When $\hat{\varepsilon}_0$ is nonzero, the instances where ε_0 appears in (3)–(5) should be replaced by $\varepsilon_0 - \hat{\varepsilon}_0$. Since (5) holds as long as $\varepsilon_0 - \hat{\varepsilon}_0$ is bounded in probability, Theorem 1 holds as long as $\hat{\varepsilon}_0 = O_p(1)$. In practice, $\hat{\varepsilon}_0$ is either set to zero or estimated from the data (the backcasting method). The irrelevance of the limiting process to the initial contrasts with the results of Kulperger (1987) who investigated the residual process from regression models with autoregressive and ARMA (1, 1) errors under the assumption that the moving-average parameter is equal to unity. Because of the non invertibility, Kulperger showed that different limiting processes can result from different initial values when calculating the residuals.

Having proved Theorem 1 for the ARMA (1, 1) case it can now be demonstrated for the ARMA (p, q) case. The key to the settlement of the problem lies in some analytical properties of the power series expansion of the reciprocal of the associated polynomials. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_q)$ be the true parameter vector for the moving-average part of the model, and let $\Psi(\theta, z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q$. By the assumption of invertibility, $\Psi(\theta, z) \neq 0$ for $|z| \leqslant 1$, there exists an $\eta > 0$ such that $1/\Psi(\theta, z)$ has power series expansion

$$\frac{1}{\Psi(\theta, z)} = \sum_{k=0}^{\infty} \psi_k(\theta) z^k \qquad |z| \leqslant 1 + \eta.$$

Consequently, there exists $B > 0$ such that $|\psi_k(\theta)| < B(1 + \eta/2)^{-k}$ for all $k \geqslant 0$ (Brockwell and Davis, 1986, p. 85). This implies that $\psi_k(\theta) \to 0$ at an exponential rate and $\sum_{k=0}^{\infty} |\psi_k(\theta)| < \infty$.

Let $u = (u_1, u_2, \ldots, u_q)$, and $\psi_k(u)$ be the coefficient of z^k in the power series expansion of $1/\Psi(u, z)$. We are interested in the behaviour of $\psi_k(u)$ when u is close to θ.

LEMMA 1. *If the roots of $\Psi(\theta, z) = 1 + \theta_1 z + \ldots + \theta_q z^q = 0$ all lie outside of the unit circle, then there exists a neighborhood V_θ of θ, $M > 0$, $C > 0$ and $0 < \beta < 1$ such that*

(a) for every $u \in V_\theta$,

$$|\psi_k(u)| < M\beta^k \qquad k = 0, 1, 2, \ldots,$$

(b) for any $\delta > 0$, if $u, u' \in V_\theta$, and $\|u - u'\|_1 \leqslant \delta$, then

$$|\psi_k(u) - \psi_k(u')| \leqslant \delta C k \beta^{k-1} \qquad k = 0, 1, 2, \ldots.$$

The proof is given in the Appendix. Observe that Lemma 1(a) implies $\sup_{u \in V_\theta} \sum_{k=0}^{\infty} |\psi_k(u)| < \infty$ and Lemma 1(b) is analogous to the inequality (23).

Likewise, let π_k be the coefficient of z^k in the power series expansion of the reciprocal of the polynomial associated with the autoregressive (AR) part, namely

$$\frac{1}{1 - \rho_1 z - \rho_2 z^2 - \ldots - \rho_p z^p} = \sum_{k=0}^{\infty} \pi_k z^k.$$

Then, by the assumption of stationarity, there exists an $M_1 > 0$ and a $\beta_1 \in (0, 1)$ such that

$$|\pi_k| < M_1 \beta_1^k \qquad k = 0, 1, 2, \ldots. \tag{24}$$

We now turn to the residual analysis. Similar to (2), the relation between $\hat{\varepsilon}_t$ and ε_t is described by

$$\hat{\varepsilon}_t - \varepsilon_t = -\sum_{i=1}^{q} \hat{\theta}_i(\hat{\varepsilon}_{t-i} - \varepsilon_{t-i}) - \sum_{i=1}^{p} (\hat{\rho}_i - \rho_i) X_{t-i} - \sum_{i=1}^{q} (\hat{\theta}_i - \theta_i)\varepsilon_{t-i}.$$

By repeated substitution and using the initial condition $(\hat{\varepsilon}_0, \hat{\varepsilon}_{-1}, \ldots, \hat{\varepsilon}_{-q+1}) = 0$, we obtain the counterpart of (3):

$$\hat{\varepsilon}_t - \varepsilon_t = Y_t(\hat{\theta}) - \sum_{i=1}^{p} (\hat{\rho}_i - \rho_i) \sum_{j=0}^{t-1} \psi_j(\hat{\theta}) X_{t-i-j} - \sum_{i=1}^{q} (\hat{\theta}_i - \theta_i) \sum_{j=0}^{t-1} \psi_j(\hat{\theta}) \varepsilon_{t-i-j} \tag{25}$$

for $t = 1, 2, \ldots, n$, where $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)$ and

$$Y_t(\hat{\theta}) = -\psi_t(\hat{\theta})\varepsilon_0 - \{\psi_{t+1}(\hat{\theta}) + \psi_t \hat{\theta}_1\}\varepsilon_{-1} - \ldots$$
$$- \{\psi_{t+q-1}(\hat{\theta}) + \psi_{t+q-2}(\hat{\theta})\hat{\theta}_1 + \ldots + \psi_t(\hat{\theta})\hat{\theta}_{-q+1}\}\varepsilon_{-q+1}, \tag{26}$$

where $\psi_k(\hat{\theta})$ is $\psi_k(\theta)$ evaluated at $\hat{\theta}$ and $\psi_k(\cdot) = 0$, for $k < 0$.

Again, by assumption (a.2), $P(\hat{\theta} \notin V_\theta) \to 0$. Thus, to prove the theorem, it suffices to show that

$$\sup_{u \in V_\theta} \sup_{0 \le x \le 1} \frac{1}{n^{1/2}} \sum_{t=1}^{[nx]} |Y_t(u)| = o_p(1), \tag{27}$$

$$\sup_{u \in V_\theta} \sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} \psi_j(u) X_{t-i-j} \right| = o_p(1) \qquad i = 1, 2, \ldots, p, \tag{28}$$

$$\sup_{u \in V_\theta} \sup_{0 \le x \le 1} \frac{1}{n} \left| \sum_{t=1}^{[nx]} \sum_{j=0}^{t-1} \psi_j(u) \varepsilon_{t-i-j} \right| = o_p(1) \qquad i = 1, 2, \ldots, q. \tag{29}$$

When $p = 1$ and $q = 1$ the above equations are reduced to (5)–(7) with V_θ taken as $[-\bar{\theta}, \bar{\theta}]$. First, notice that Lemma 1(a) and the boundedness of V_θ lead immediately to (27) (V_θ can always be chosen to be bounded). To prove (28) we can proceed in almost the same way as in the preceding proof of (6), by changing u^j to $\psi_j(u)$ and ρ^k to π_k. Note that there will be $q + 1$ terms on the right-hand side of (10) instead of two, but each can be taken care of separately as we have done. Since a_k is now changed to $\sum_{j=0}^{k} \psi_j(u)\pi_{k-j}$ (see (12)), we have, by Lemma 1(a) and (24)

$$|a_k| = \left|\sum_{j=0}^{k} \psi_j(u)\pi_{k-j}\right| \leq MM_1(k+1)\tau^k \qquad \tau = \max(\beta_1, \beta) < 1.$$

The above is analogous to (13). Thus (14) and (15) hold for the newly defined a_k. Consequently, Theorem 21.1 of Billingsley (1968) and the results of Phillips and Solo (1989) are still applicable. Therefore, (16) and hence (19) are still valid under the new setting. To continue, redefine $[\delta]$ to be $[\delta] = \{(u, u');\ u, u' \in V_\theta;\ \|u - u'\|_1 \leq \delta\}$. Then, using Lemma 1(b), one readily obtains (20) for the general case. Finally, for any $\delta > 0$, since V_θ is bounded (or can be chosen to be compact), V_θ can be partitioned into a finite number of subsets such that the diameter of each subset is not greater than δ. This together with (19) and (20) is sufficient to prove (28), and (29) can be proved similarly.

4. APPLICATIONS

In this section we discuss two additional related applications. One is concerned with the asymptotic distribution of a test statistic for testing the presence of change-point in time series models. The other regards the uniform consistency of a kernel density estimator constructed from the residuals.

4.1. *Change-point problem*

One aspect of the change-point problem involves testing the hypothesis (null) of no parameter changes over time versus the hypothesis (alternative) that parameter changes do take place but with the change occurring at an unknown time. MacNeill (1978) proposed a test statistic for linear regression models. The test is constructed from regression residuals. The following discussion is a direct application of MacNeill's test to ARMA time series models. The application of the test to AR models was examined by Kulperger (1985). See also the discussion by Jandhyala and MacNeill (1989) for other models and related test statistics.

Let $\Phi(s)$ be a nonnegative function on $[0, 1]$, such that $\int_0^1 s\Phi(s)ds < \infty$. Define

$$R(j/n) = \int_{(2j-1)/2n}^{(2j+1)/2n} \Phi(s)ds \qquad j = 1, 2, \ldots, n-1$$

and the statistic S_n for testing the change-point problem:

$$S_n = \sum_{t=1}^{n-1} R\left(\frac{t}{n}\right)\left(\frac{1}{\hat{\sigma}n^{1/2}} \sum_{j=1}^{t} \hat{\varepsilon}_j\right)^2 = \frac{\sigma^2}{\hat{\sigma}^2} \sum_{t=1}^{n-1} R\left(\frac{t}{n}\right)\left\{\hat{B}^{(n)}\left(\frac{t}{n}\right)\right\}^2,$$

where $\hat{\sigma}$ is a consistent estimator for σ (under the null). Using Theorem 1 and (27)–(29), one can show that under the null hypothesis

$$S_n \to \int_0^1 \Phi(s) B^2(s) ds$$

in distribution. The proof is omitted because once (27)–(29) have been proved, it is similar to that given by Kulperger (1985, Theorem 2.2) for AR models. When $\Phi(s) \equiv 1$, the limiting distribution becomes $\int_0^1 B^2(s)ds$, and selected quantiles have been calculated by MacNeill (1978).

REMARKS. The literature on the change-point problem is voluminous. See the survey articles by Zacks (1983) and Krishnaiah and Miao (1988). Many test statistics for the change-point problem employ recursive residuals (Brown *et al.*, 1975). The test proposed by MacNeill, however, utilizes raw residuals. The test is easy to construct since many software packages compute the raw residuals when estimating models.

4.2. *Nonparametric density estimation*

This example illustrates the usefulness of Lemma 1 and of the representation of the residuals in the form of (25). Suppose the i.i.d. innovations possess a uniformly continuous density function $f(x)$ which is unknown. We will consider a nonparametric kernel estimation of $f(x)$ based on the residuals and then establish the uniform consistency for the suggested estimator. To begin, let h_n be a sequence of positive numbers tending to zero, and let $K(x)$ be a probability density function. We define

$$f_n(x) = \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{x - \varepsilon_t}{h_n}\right) \qquad \hat{f}_n(x) = \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{x - \hat{\varepsilon}_t}{h_n}\right) \qquad x \in R.$$

In addition, we assume that the following conditions hold.

(i) $h_n > 0$; $h_n \to 0$, $n^{1/2} h_n^2 \to \infty$.

(ii) $\sup_{|x|>b} |x| K(x) \to 0$ as $b \to \infty$; there exists $c < \infty$ such that $|K(x) - K(y)| \leq c|x - y|$ for all $x, y \in R$, i.e. K is Lipschitz.

(iii) (a.1) and (a.2) of Theorem 1.

Under these conditions, we have the following uniform consistency result:

$$\sup_{x \in R} |\hat{f}_n(x) - f(x)| = o_p(1). \tag{30}$$

REMARKS. This problem was considered by Robinson (1987) for a wider class of time series. It would be interesting, however, to compare the assumptions and proofs employed in his paper and in this paper. With the aid of Lemma 1, the uniform consistency is considerably easier to establish for ARMA models.

PROOF OF (30). Under the above conditions $\sup_{x \in R} |f_n(x) - f(x)| = o_p(1)$ (see Parzen (1962) for example). Thus it suffices to show that $\sup_{x \in R} |\hat{f}_n(x) - f_n(x)| = o_p(1)$. Now,

$$\sup_x |\hat{f}_n(x) - f_n(x)| \leq \sup_x \frac{1}{nh_n} \sum_{t=1}^{n} \left| K\left(\frac{x - \hat{\varepsilon}_t}{h_n}\right) - K\left(\frac{x - \varepsilon_t}{h_n}\right) \right|$$

$$\leq c \frac{1}{nh_n^2} \sum_{t=1}^{n} |\hat{\varepsilon}_t - \varepsilon_t| \qquad \text{by } K \text{ Lipschitzness}$$

$$\leq c \frac{1}{nh_n^2} \sum_{t=1}^{n} |Y_t(\hat{\theta})|$$

$$+ \frac{1}{n^{1/2} h_n^2} \left[\sum_{i=1}^{p} \left\{ n^{1/2} |\hat{\rho}_i - \rho_i| \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{t} |\psi_j(\hat{\theta}) X_{t-i-j}| \right\} \right.$$

$$\left. + \sum_{i=1}^{q} \left\{ n^{1/2} |\hat{\theta}_i - \theta_i| \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{t} |\psi_j(\hat{\theta}) \varepsilon_{t-i-j}| \right\} \right]$$

by (25). It remains to show that each of the two terms on the right converges to zero in probability. This is the case if we can show, after replacing $\hat{\theta}$ by u, that each of them is $o_p(1)$ uniformly in $u \in V_\theta$, since $P(\hat{\theta} \notin V_\theta) \to 0$ by condition (a.2) of Theorem 1. But (27) implies

$$\sup_{u \in V_\theta} \frac{1}{nh_n^2} \sum_{t=1}^{n} |Y_t(u)| = o_p(1),$$

and thus the first term is $o_p(1)$. Next consider the second term:

$$E\left\{ \sup_{u \in V_\theta} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{t} |\psi_j(u) X_{t-i-j}| \right\} \leq E\left\{ \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{t} \sup_{u \in V_\theta} |\psi_j(u)| |X_{t-i-j}| \right\}$$

$$\leq \left(M \sum_{j=1}^{\infty} \beta^j \right) E|X_1|$$

$$= M\beta(1 - \beta)^{-1} E|X_1| < \infty$$

by Lemma 1(a). The above inequalities hold when X_t is replaced by ε_t. These, together with condition (a.2) of Theorem 1, imply that the second term is equal to $(n^{1/2} h_n^2)^{-1} O_p(1)$ which coverges to zero in probability since $(n^{1/2} h_n^2)^{-1} \to 0$ by condition (i). ■

APPENDIX: PROOF OF LEMMA 1

(a) Write $\Psi(u, z) = \{1 - \lambda_1(u)z\} \{1 - \lambda_2(u)z\} \ldots \{1 - \lambda_q(u)z\}$, where the $1/\lambda_i(u)$ are the roots of $\Psi(u, z) = 0$. By definition,

$$|\psi_k(u)| = \left| \sum_{l_1 + l_2 + \ldots + l_q = k} \lambda_1(u)^{l_1} \lambda_2(u)^{l_2} \ldots \lambda_q(u)^{l_q} \right| \leq (k + 1)^q \max_{1 \leq i \leq q} |\lambda_i(u)|^k,$$

where each l_i runs from 0 to k. Since $\Psi(\theta, z) \neq 0$ for $|z \leq 1$, there is a β_0 such that $\max_i |\lambda_i(\theta)| < \beta_0 < 1$. From the well-known property that a polynomial's roots depend continuously on its coefficients (Goursat, 1933), it follows that there is a neigborhood V_θ of θ such that, for each $u \in V_\theta$, $\max_i |\lambda_i(u)| < \beta_0$. Now take β slightly larger than β_0 but less than 1, and take M large enough to assure

$$(k + 1)^q \beta_0^k \leq M\beta^k \qquad \text{for all } k \geq 0.$$

(b) The term $\psi_k(u) - \psi_k(u')$ is the coefficient of z^k in the power series expansion of $1/\Psi(u, z) - 1/\Psi(u', z)$, but

$$\frac{1}{\Psi(u, z)} - \frac{1}{\Psi(u', z)} = \frac{\sum_{i=1}^{q}(u_i' - u_i)z^i}{\Psi(u, z)\Psi(u', z)}.$$

Thus the coefficient of z^k can alternatively be expressed as $\sum_{i=1}^{q}(u_i' - u_i)a_{k-i}$ (take $a_j = 0$ for $j < 0$), where $a_j = \sum_{l=0}^{j}\psi_l(u)\psi_{j-l}(u')$ which is the coefficient of z^j in the power series expansion of the product of $1/\Psi(u, z)$ and $1/\Psi(u', z)$. Since $|a_j| \leq M^2(j+1)\beta^j$ by Lemma 1(a) and $|u_i' - u_i| \leq \|u' - u\|_1 \leq \delta$ by assumption,

$$|\psi_k(u) - \psi_k(u')| \leq \delta M^2\{k\beta^{k-1} + \ldots + (k - q + 1)\beta^{k-q}\} \qquad \text{for } k > q$$

and

$$|\psi_k(u) - \psi_k(u')| \leq \delta M^2\{k\beta^{k-1} + (k-1)\,\beta^{k-2} + \ldots + 1\} \qquad 1 \leq k \leq q.$$

Note that $\psi_0(u) - \psi_0(u') = 0$, since $\psi_0(\cdot) \equiv 1$. Thus, for all k, an upper bound for $|\psi_k(u) - \psi_k(u')|$ is $\delta(qM^2/\beta^{q-1})k\beta^{k-1}$. Lemma 1(b) follows upon choosing $C = qM^2/\beta^{q-1}$.

REMARKS. Let $D = \{u \in R^q;\ \Psi(u, z) = 1 + u_1 z + \ldots + u_q z^q \neq 0, \text{ for } |z| \leq 1\}$, i.e. D is a collection of all vectors such that the corresponding polynomial has roots all located outside the unit circle. Then D is open and

$$D \subset \prod_{i=1}^{q}\left[-\binom{q}{i}, \binom{q}{i}\right],$$

a rectangle in R^q. The former is essentially implied by the proof of (a); the latter can be verified either by induction or by directly examining the relations between roots and coefficients. The upper bound set can be used as a crude criterion for judging the stationarity and invertibility of an ARMA model with known coefficients. Finally, notice that, from the proof of (a), $V_\theta \subset D$.

ACKNOWLEDGEMENT

The author is grateful to Professor David Brillinger for suggesting this topic. He also thanks Professors Deborah Nolan, Thomas Rothenberg and James Stock for helpful comments and suggestions.

REFERENCES

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

BROCKWELL, P. J. and DAVIS, R. A. (1986) *Time Series: Theory and Method*. New York: Springer-Verlag.

BROWN, R. J., DURBIN, J. and EVANS, J. M. (1975) Techniques for testing the constancy of regression relationships over time. *J. R. Statis. Soc. Ser. B* 37, 149–92.

GOURSAT, E. (1933) *Cours d'Analyse Mathématique*, Vol. II, 5th edn. Paris: Gauthier Villars.

HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and its Applications*. New York: Academic Press.

JANDHYALA, V. K. and MACNEILL, I. B. (1989) Residual partial sum limit process for regression models with applications to detecting parameter changes at unknown times. *Stoch. Process. Appl.* 33, 309–23.

KRISHNAIAH, P. R. and MIAO, B. M. (1988) Review about estimation of change points. In *Handbook of Statistics*, Vol 7 (eds P. R. Krishnaiah and C. R. Rao). New York: Elsevier.

KULPERGER, R. J. (1985) On the residuals of autoregressive processes and polynomial regression. *Stoch. Process. Appl.* 21, 107–18.

KULPERGER, R. J. (1987) Some remarks on regression with autoregressive errors and their residual processes. *J. Appl. Probab.* 24, 668–78.

MACNEILL, I. B. (1978a) Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change in regression at unknown times. *Ann. Statis.* 6, 422–33.

MACNEILL, I. B. (1978b) Limit processes for sequences of partial sums of regression residuals. *Ann. Probab.* 6, 695–98.

MACNEILL, I. B. and JANDHYALA, V. K. (1985) The residual process for non-linear regression. *J. Appl. Probab.* 22, 957–63.

MCLEISH, D. L. (1974) Dependent central limit theorems and invariance principles. *Ann. Probab.* 2, 620–28.

PARZEN, E. (1962) On estimation of probability density function and mode. *Ann. Math. Statist.* 33, 1065–76.

PHILLIPS, P. C. B. and SOLO, V. (1989) Asymptotics for linear processes. Cowles Foundation Discussion Paper 932, Yale University.

ROBINSON, P. M. (1987) Time series residuals with application to probability density estimation. *J. Time Series Anal.* 8, 329–44.

ZACKS, S. (1983) Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. In *Recent Advances in Statistics* (eds M. H. Rivzi, J. S. Rustagi and D. Siegmund). New York: Academic Press.