High Moment Partial Sum Processes of Residuals in ARMA Models and their Applications

Hao Yu*

Abstract

In this paper we study high moment partial sum processes based on residuals of a stationary ARMA model with or without a unknown mean parameter. We show that they can be approximated in probability by the analogous processes which are obtained from the independent and identically distributed (iid) errors of the ARMA model. However, if a unknown mean parameter is used, there will be an additional term that depends on model parameters and a mean estimator. But, when properly normalized, this additional term will be cancelled out. Thus they converge weakly to the same Gaussian processes as if the residuals were iid. Applications to change-point problems and goodness-of-fit are considered, in particular CUSUM statistics for testing ARMA model structure changes and the Jarque-Bera omnibus statistic for testing normality of the unobservable error distribution of an ARMA model.

Keywords: ARMA, residuals, high moment partial sum process, weak convergence, CUSUM, omnibus, skewness, kurtosis, $\sqrt{n}$ consistency.

1 Introduction and results

Statistics or stochastic processes constructed from residuals of stationary autoregressive moving-average ARMA(p, q) models have been studied extensively in literature. For examples, Boldin and Arie (1982), Boldin (1990), Koul (1991), Kreiss (1991), Bai (1994), and Yu (2003) study the weak convergence of (sequential) empirical processes. Yu (2003) shows that the standard Kolmogorov-Smirnov goodness-of-fit test based on residuals of stationary ARMA models with unknown mean parameter is not applicable. Kulpeger

*Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada. hyu@stats.uwo.ca.
Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

(1985) and Bai (1993) investigate the partial sum process of residuals in autoregressive AP(p) models and ARMA(p, q) models respectively. On the other hand, the so-called Jarque and Bera (1980,1987) test for the normality of the error distribution has been popular among economists. It is an omnibus test based on the standardized sample skewness and sample kurtosis of residuals which has been known among statisticians since the work of Bowman and Shenton (1975). So far the asymptotic validity of the Jarque and Bera test has been proved for AR models only (see Lütkepohl (1993)).

Recently Kulperger and Yu (2003) construct and study high moment partial sum processes based on residuals of GARCH models. They show that partial sum processes and Jarque and Bera test statistics are two special cases of high moment partial sum processes. In addition, CUSUM statistics can be constructed to test various GARCH model structure changes such as variance change in errors. Another important feature is that, when properly normalized, high moment partial sum processes will cancel out terms that are related to model parameters. Thus any statistics constructed from high moment partial sum processes of residuals will behave as if residuals were iid errors. In this paper we study high moment partial sum processes based on residuals of a stationary ARMA model with or without a unknown mean parameter. Applications to change-point problems and goodness-of-fit are considered, in particular CUSUM statistics for testing ARMA model structure changes and the Jarque-Bera omnibus statistic for testing normality of the unobservable error distribution of an ARMA model.

An ARMA(p, q) time series model with a unknown mean parameter is defined as

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}, \tag{1.1}$$

where the errors $\{\epsilon_t\}$ are i.i.d. with zero mean and a unknown distribution function (d.f.) F on the real line $\mathbb{R}$, and $\mu, \phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are unknown parameters. Let $X_t = Y_t - \mu$. Then $\{X_t\}$ will be the usual ARMA(p, q) process with zero mean, i.e., if we set $\mu = 0$, the $\{Y_t\}$ will be the same as $\{X_t\}$.

Let $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ and $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. According to

Brockwell and Davis (1991), if

(A1) $\Phi(z)$ and $\Theta(z)$ do not have common roots

(A2) All roots of $\Phi(z)$ and $\Theta(z)$ lie outside the unit circle of the complex plane,

then $\{Y_t\}$ (and $\{X_t\}$) is strictly stationary and invertible. In particular, the invertibility implies that

$$\Theta(1) = 1 + \theta_1 + \cdots + \theta_q \neq 0. \tag{1.2}$$

Given $n+p$ observations $\{Y_t, -p+1 \leq t \leq n\}$, the residuals are calculated by the recursion formula

$$\hat{\epsilon}_t = \hat{X}_t - \hat{\phi}_1 \hat{X}_{t-1} - \cdots - \hat{\phi}_p \hat{X}_{t-p} - \hat{\theta}_1 \hat{\epsilon}_{t-1} - \cdots - \hat{\theta}_q \hat{\epsilon}_{t-q}, \ 1 \leq t \leq n, \tag{1.3}$$

where $\hat{X}_t = Y_t - \hat{\mu}$, and $\hat{\mu}$, $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)$ are the estimators for μ, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$, respectively. The initial values of $\hat{\epsilon}_{-q+1}, \ldots, \hat{\epsilon}_0$ are set to zero if $q > 0$. In case we consider an ARMA(p,q) model without the mean parameter ($\mu \equiv 0$), then the above construction of residuals is still valid except that $\hat{X}_t = X_t = Y_t$.

The kth ($k = 1, 2, 3, 4, \ldots$) order high moment partial sum process of residuals is defined as

$$\hat{S}_n^{(k)}(x) = \sum_{t=1}^{[nx]} \hat{\epsilon}_t^k, \ 0 \leq x \leq 1, \tag{1.4}$$

where, for any real number a, $[a]$ denotes the largest integer $\leq a$. Its counterpart based on iid errors is defined as

$$S_n^{(k)}(x) = \sum_{t=1}^{[nx]} \epsilon_t^k, \ 0 \leq x \leq 1. \tag{1.5}$$

In order to present our first result, in addition to the conditions (A1) and (A2) on $\{Y_t\}$, we need the following assumptions which are similar to those given by Bai (1993):

(A3) $\{\epsilon_t\}$ are i.i.d. with zero mean, finite variance and d.f. F.

(A4) $\sqrt{n}(\hat{\mu} - \mu) = O_P(1)$, $\sqrt{n}(\hat{\phi}_i - \phi_i) = O_P(1)$, $i = 1, \ldots, p$, and $\sqrt{n}(\hat{\theta}_j - \theta_j) = O_P(1)$, $j = 1, \ldots, q$.

Theorem 1.1 *We assume that the assumptions* (A1) to (A4) *hold. Then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 1$ *implies that*

$$\sup_{0 \leq x \leq 1} \left| \frac{1}{\sqrt{n}} \left(\hat{S}_n^{(k)}(x) - S_n^{(k)}(x) \right) + \frac{k\mu_{k-1}[nx]}{n} \frac{1 - \sum_{i=1}^{p} \phi_i}{1 + \sum_{j=1}^{q} \theta_j} \sqrt{n}(\hat{\mu} - \mu) \right| = o_P(1) \qquad (1.6)$$

and

$$\sup_{0 \leq x \leq 1} \left| \frac{1}{\sqrt{n}} \left(\frac{\hat{S}_n^{(k)}(x)}{\hat{\gamma}_{(n)}^k} - \frac{S_n^{(k)}(x)}{\gamma_{(n)}^k} \right) + \frac{k\mu_{k-1}[nx]}{\mu_2^{k/2} n} \frac{1 - \sum_{i=1}^{p} \phi_i}{1 + \sum_{j=1}^{q} \theta_j} \sqrt{n}(\hat{\mu} - \mu) \right| = o_P(1), \qquad (1.7)$$

where $\mu_{k-1} = E\epsilon_0^{k-1}$, $\hat{\gamma}_{(n)}^2 = \hat{S}_n^{(2)}(1)/n$ *and* $\gamma_{(n)}^2 = S_n^{(2)}(1)/n$.

Remark 1.1 *We note that if* (A1), (A2), *and* (A3) *are assumed, then the conditional least square estimators for* μ, $\phi_1, \ldots, \phi_p$, *and* $\theta_1, \ldots, \theta_q$ *satisfy* (A4).

Remark 1.2 *Obviously, with* $k = 1$ *and a unknown mean parameter* μ *introduced and estimated, Theorem 1.1 differs from Theorem 1 of Bai (1993) where there is no an extra term in (1.6) that is related to model parameters and the estimator* $\hat{\mu}$. *Since* $\mu_1 = 0$ *by* (A3), *for* $\hat{\gamma}_{(n)}^2$ *and* $\gamma_{(n)}^2$ *defined in Theorem 1.1, (1.6) implies*

$$\sqrt{n} \left| \hat{\gamma}_{(n)}^2 - \gamma_{(n)}^2 \right| = o_P(1). \qquad (1.8)$$

Hence $\hat{\gamma}_{(n)}^2$ *is an estimator of the variance* μ_2, *i.e.,* $\hat{\gamma}_{(n)}^2 \to \mu_2$ *in probability under the minimum condition* $\mu_2 < \infty$. *Similarly, if* $\mu_{k-1} = 0$ *for some even number* $k \geq 4$, *the extra terms in (1.6) and (1.7) disappear. Notice that the standard deviation scale* $\hat{\gamma}_{(n)}$ *in (1.6) does not help to cancel the extra term out.*

To get similar results of Theorem 1.1 without a unknown mean parameter, we need to modify the assumption (A4) as

(A4') $\mu \equiv 0$, $\sqrt{n}(\hat{\phi}_i - \phi_i) = O_P(1)$, $i = 1, \ldots, p$, and $\sqrt{n}(\hat{\theta}_j - \theta_j) = O_P(1)$, $j = 1, \ldots, q$.

Theorem 1.2 *We assume that the assumptions* (A1) to (A3) *and* (A4') *hold. Then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 1$ *implies that*

$$\sup_{0 \leq x \leq 1} \frac{1}{\sqrt{n}} \left| \hat{S}_n^{(k)}(x) - S_n^{(k)}(x) \right| = o_P(1)$$

and

$$\sup_{0 \leq x \leq 1} \frac{1}{\sqrt{n}} \left| \frac{\hat{S}_n^{(k)}(x)}{\hat{\gamma}_{(n)}^k} - \frac{S_n^{(k)}(x)}{\gamma_{(n)}^k} \right| = o_P(1).$$

Theorem 1.2 extends Theorem 1 of Bai (1993) to high moment partial sum processes. As expected, there is no an extra term since there is no a mean parameter. By Theorems 1.1 and 1.2, we immediately obtain the following result after using CUSUM normalization.

Theorem 1.3 *If* (A1) to (A3) *and* (A4) or (A4') *hold, then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 1$ *implies that*

$$\sup_{0 \leq x \leq 1} \frac{1}{\sqrt{n}} \left| \left(\hat{S}_n^{(k)}(x) - \frac{[nx]}{n} \hat{S}_n^{(k)}(1) \right) - \left(S_n^{(k)}(x) - \frac{[nx]}{n} S_n^{(k)}(1) \right) \right| = o_P(1).$$

Theorem 1.3 implies that the CUSUM normalized high moment partial sum process $\{(\hat{S}_n^{(k)}(x) - x\hat{S}_n^{(k)}(1))/\sqrt{n},\ 0 \leq x \leq 1\}$ has the same Gaussian limit as that of $\{(S_n^{(k)}(x) - xS_n^{(k)}(1))/\sqrt{n},\ 0 \leq x \leq 1\}$ and the extra term in (1.6) cancels.

Let $\nu_k^2 = E(\epsilon_0^k - \mu_k)^2 < \infty$. Then the invariance principle for partial sums of iid sequence $\{\epsilon_t^k\}$ (cf. Billingsley (1999)) implies that

$$\left\{ \frac{S_n^{(k)}(x) - [nx]S_n^{(k)}(1)/n}{\nu_k \sqrt{n}},\ 0 \leq x \leq 1 \right\}$$

converges weakly in the Skorokhod space $D[0,\ 1]$ to a Brownian bridge $\{B(x),\ 0 \leq x \leq 1\}$. Hence the following result follows from Theorem 1.3.

Corollary 1.1 *If* (A1) to (A3) *and* (A4) or (A4') *hold, then* $E|\epsilon_0|^{2k} < \infty$ *for some integer* $k \geq 1$ *implies*

$$\left\{ \frac{\hat{S}_n^{(k)}(x) - [nx]\hat{S}_n^{(k)}(1)/n}{\nu_k \sqrt{n}},\ 0 \leq x \leq 1 \right\}$$

converges weakly in the Skorokhod space $D[0,\ 1]$ *to a Brownian bridge* $\{B(x),\ 0 \leq x \leq 1\}$.

Remark 1.3 To use Corollary 1.1 for CUSUM tests of structure change of stationary ARMA models, one needs to estimate ν_k. The details are left to the next section.

Before we give the next result, we need to redefine the high moment partial processes of (1.4) and (1.5). The kth order high moment centered partial sum process of residuals is defined as

$$\hat{T}_n^{(k)}(x) = \sum_{t=1}^{[nx]} \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}}\right)^k, \; 0 \leq x \leq 1, \tag{1.9}$$

where $\bar{\hat{\epsilon}}$ is the sample mean of residuals. Its counterpart based on iid errors is defined as

$$T_n^{(k)}(x) = \sum_{t=1}^{[nx]} \left(\epsilon_t - \bar{\epsilon}\right)^k, \; 0 \leq x \leq 1, \tag{1.10}$$

where $\bar{\epsilon}$ is the sample mean of errors.

Theorem 1.4 *If* (A1) to (A3) *and* (A4) or (A4') *hold, then $E|\epsilon_0|^k < \infty$ for an integer $k \geq 1$ implies that*

$$\sup_{0 \leq x \leq 1} \frac{1}{\sqrt{n}} \left| \hat{T}_n^{(k)}(x) - T_n^{(k)}(x) \right| = o_P(1) \tag{1.11}$$

and

$$\sup_{0 \leq x \leq 1} \frac{1}{\sqrt{n}} \left| \frac{\hat{T}_n^{(k)}(x)}{\hat{\sigma}_{(n)}^k} - \frac{T_n^{(k)}(x)}{\sigma_{(n)}^k} \right| = o_P(1), \tag{1.12}$$

where $\hat{\sigma}_{(n)}^2 = \hat{T}_n^{(2)}(1)/n$ and $\sigma_{(n)}^2 = T_n^{(2)}(1)/n$.

Remark 1.4 *Obviously, $\hat{\sigma}_{(n)}^2$ is the usual sample variance estimator. In fact, Theorem 1.4 implies that $\hat{\sigma}_{(n)}^2 \to \mu_2$ in probability under the minimum condition $\mu_2 < \infty$. Although the estimator $\hat{\gamma}_{(n)}^2$ in Theorem 1.1 does not use sample mean centering, both $\hat{\gamma}_{(n)}^2$ and $\hat{\sigma}_{(n)}^2$ estimate the variance μ_2. In addition, by (1.8) and Theorems 1.2 and 1.4, they have the same limiting distribution regardless whether there is a mean parameter or not.*

Remark 1.5 *By comparing Theorem 1.4 with Theorem 1.1, one can notice that, by merely sample mean centering, the extra terms in Theorem 1.1 cancel in Theorem 1.4. This is quite in contrary to the result obtained by Kuperger and Yu (2003) for GARCH* *models where $\hat{\sigma}_{(n)}$ and $\sigma_{(n)}$ must be used in order to cancel a term that is related to GARCH parameters.*

Theorem 1.4 implies that $\{(\hat{T}_n^{(k)}(x) - nx\mu_k)/\sqrt{n},\ 0 \le x \le 1\}$ has the same Gaussian limit as that of $\{(T_n^{(k)}(x) - nx\mu_k)/\sqrt{n}, 0 \le x \le 1\}$. So does $\{(\hat{T}_n^{(k)}(x)/\sigma_{(n)}^k - nx\lambda_k)/\sqrt{n},\ 0 \le x \le 1\}$, where $\lambda_k = \mu_k/\mu_2^{k/2}$. However, except for the cases $k = 1, 2$, those Gaussian limits depend on the moments of the error distribution and cannot be identified to specific processes such as Brownian motions or Brownian bridges. The details can be found in Kulperger and Yu (2003). Here we just give the following two corollaries that will be used to construct a CUSUM statistic and the Jarque-Bera test statistic given in the next section.

Corollary 1.2 *Assume that* (A1) to (A3) and (A4) or (A4') *hold. Then $E\epsilon_0^4 < \infty$ implies that*

$$\left\{ \frac{1}{\sqrt{(\lambda_4 - 1)n}} \left(\frac{\hat{T}_n^{(2)}(x)}{\hat{\sigma}_{(n)}^2} - nx \right),\ 0 \le x \le 1 \right\}$$

converges weakly in the Skorokhod space $D[0,\ 1]$ to a Brownian bridge $\{B(x),\ 0 \le x \le 1\}$.

Corollary 1.3 *Assume that* (A1) to (A3) and (A4) or (A4') *hold. Assume also that $k \ge 1$ is a odd number and $\mu_3 = \mu_k = \mu_{k+2} = \mu_{2k+1} = 0$. Then $E|\epsilon_0|^{2(k+1)} < \infty$ implies that*

$$\left\{ \frac{1}{\sqrt{n}} \left(\frac{\hat{T}_n^{(k)}(x)}{\hat{\sigma}_{(n)}^k} - nx\lambda_k,\ \frac{\hat{T}_n^{(k+1)}(y)}{\hat{\sigma}_{(n)}^{k+1}} - ny\lambda_{k+1} \right),\ 0 \le x, y \le 1 \right\}$$

converges weakly in the Skorokhod space $D^2[0,\ 1]$ to a two dimensional Gaussian process $\{\left(B^{(k)}(x),\ B^{(k+1)}(y)\right)\ 0 \le x, y \le 1\}$, where $\{B^{(k)}(x),\ 0 \le x \le 1\}$ and $\{B^{(k+1)}(y),\ 0 \le y \le 1\}$ are two independent zero mean Gaussian processes defined by

$$\begin{aligned} EB^{(i)}(x)B^{(i)}(y) &= (\lambda_{2i} - \lambda_i^2)(x \wedge y) + i\lambda_{i-1}(i\lambda_{i-1} + i\lambda_i\lambda_3 - 2\lambda_{i+1})xy \\ &\quad + i\lambda_i\left((1 - i/4)\lambda_i + i\lambda_i\lambda_4/4 - \lambda_{i+2}\right)xy,\ i = k, k+1, \end{aligned} \tag{1.13}$$

for any $0 \le x, y \le 1$ and $x \wedge y = \min(x, y)$.

Applications for change-point problems and goodness-of-fit tests are given in the next section, alone with a discussion of using the residuals to construct a kernel density estimation of the error distribution. All proofs are presented in Section 3.

2 Applications

Intuitively, the adequacy or inadequacy of the fitted model is reflected through model residuals. It includes if model parameters are properly chosen and if parameters change over time. One of the motivations to construct high moment partial sum processes of residuals is to capture as much information as possible of model parameters through different moments of residuals. On the other hand, identifying the distribution of the error distribution and to know if it has a constant variance are two important aspects of model diagnostic checking. Although normality is not necessary for many statistical procedures, its tests are useful for such tests as serial correlation in model residuals and for autoregressive conditional heteroscedasticity (ARCH). In this section we discuss two applications of high moment partial sum processes. One is to construct statistics for testing the presence of change-point in ARMA models, including if the variance of error terms changes over time. The other is to construct the popular Jarque-Bera test for the normality of the error distribution. In addition, the uniform consistency of a kernel density estimator constructed from the residuals is discussed.

2.1 Change-point Problem

The change-point problem related to ARMA models can be formulated to test the hypothesis (null) of no ARMA parameters change over time versus the hypothesis (alternative) that parameters change at unknown time. MacNeill (1978) proposes a test statistic for linear regression models. His test has been applied to AR models by Kulperger (1985) and ARMA models without a mean parameter by Bai (1993). However, Theorem 1.1 shows that, once a unknown mean parameter is introduced, the test statistic is not valid since the limiting process is no longer to be a Brownian motion. Based on Remark 1.2 one can still use the squared residuals to construct the test statistic. But one needs to verify if it performs as required. We will not pursue along this line in this paper. Rather we will propose in the following the standard CUSUM test introduced by Brown, Durbin and Evans (1975). It was one of the first tests on structural change with unknown break point.

Firstly, a change-point problem for ARMA models is to test the mean change. We can formulate it in the following hypothesis tests. The null hypothesis is "no-change in the mean"

$$H_0 : \mu = \text{constant},\ t = 1, 2, \ldots, n$$

against the "one change in the mean" alternative

$$H_a : \begin{cases} \mu = \mu', & t = 1, \ldots, [nx^*] \\ \mu = \mu'', & t = [nx^*] + 1, \ldots, n, \end{cases}$$

where $\mu' \neq \mu''$ and $0 < x^* < 1$. To test the above hypothesis, we use the standard CUSUM tests constructed from residuals as

$$CUSUM^{(1)} = \max_{1 \le i < n} \frac{\left| \sum_{t=1}^{i} \hat{\epsilon}_t - i\bar{\hat{\epsilon}} \right|}{\hat{\sigma}_{(n)} \sqrt{n}}.$$

By a straight calculation, it is easy to verify that

$$CUSUM^{(1)} = \sup_{0 \le x \le 1} \frac{\left| \hat{S}_n^{(1)}(x) - [nx]\hat{S}_n^{(1)}(1)/n \right|}{\hat{\sigma}_{(n)} \sqrt{n}} + o_P(1)$$

provided that $E\epsilon_0^2 < \infty$. Therefore, by Corollary 1.1 and Remark 1.4, under H_0,

$$CUSUM^{(1)} \xrightarrow{\mathcal{D}} \sup_{0 \le x \le 1} |B(x)|,$$

where $\{B(x),\ 0 \le x \le 1\}$ is a Brownian bridge. Hence we can reject the H_0 in favor of H_a if $CUSUM^{(1)}$ is large.

Remark 2.1 The statistic $CUSUM^{(1)}$ involves the estimation of $\sqrt{\mu_2}$ with $\hat{\sigma}_{(n)}$ being used. Based on Remark 1.4, one can use $\hat{\gamma}_{(n)}$ as well. Probably a pooled estimator of $\sqrt{\mu_2}$ should be used in $CUSUM^{(1)}$ which may result in better power.

To test the error variance change of an ARMA model, we use the null hypothesis for "no-change in the error variance"

$$H_0' : \mu_2 = \text{constant},\ t = 1, 2, \ldots, n$$

against the "one change in the error variance" alternative

$$H_a' : \begin{cases} \mu_2 = \mu_2', & t = 1, \ldots, [nx^*] \\ \mu_2 = \mu_2'', & t = [nx^*] + 1, \ldots, n, \end{cases}$$

where $\mu_2' \neq \mu_2''$ and $0 < x^* < 1$. In the following we propose two CUSUM statistics. The first one is defined as

$$CUSUM_1^{(2)} = \max_{1 \leq i < n} \frac{\left| \sum_{t=1}^{i} \hat{\epsilon}_t^2 - i \sum_{t=1}^{n} \hat{\epsilon}_t^2 / n \right|}{\hat{\nu}_2 \sqrt{n}},$$

where

$$\hat{\nu}_2^2 = \frac{1}{n} \sum_{t=1}^{n} \left(\left(\hat{\epsilon}_t - \bar{\hat{\epsilon}} \right)^2 - \hat{\sigma}_{(n)}^2 \right)^2$$

is an estimator of $\nu_2 = E(\epsilon_0^2 - \mu_2)^2 = \mu_2^2(\lambda_4 - 1)$. The second one is defined as

$$CUSUM_2^{(2)} = \max_{1 \leq i < n} \frac{\left| \sum_{t=1}^{i} \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}} \right)^2 - i \sum_{t=1}^{n} \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}} \right)^2 / n \right|}{\hat{\nu}_2 \sqrt{n}},$$

that is, $CUSUM_2^{(2)}$ is centered about the sample mean $\bar{\hat{\epsilon}}$ in contrast to no centering $CUSUM_1^{(2)}$. Again, by straight calculations, it is easy to show that

$$CUSUM_1^{(2)} = \sup_{0 \leq x \leq 1} \frac{\left| \hat{S}_n^{(2)}(x) - [nx] \hat{S}_n^{(2)}(1)/n \right|}{\hat{\nu}_2 \sqrt{n}} + o_P(1)$$

and

$$CUSUM_2^{(2)} = \sup_{0 \leq x \leq 1} \left| \frac{\hat{\sigma}_{(n)}^2}{\hat{\nu}_2 \sqrt{n}} \left(\frac{\hat{T}_n^{(2)}(x)}{\hat{\sigma}_{(n)}^2} - nx \right) \right| + o_P(1),$$

provided that $E\epsilon_0^4 < \infty$. Therefore, by Corollaries 1.1 and 1.2, under H_0',

$$CUSUM_i^{(2)} \xrightarrow{\mathcal{D}} \sup_{0 \leq u \leq 1} |B(x)|,\ i = 1, 2,$$

where $\{B(x),\ 0 \leq x \leq 1\}$ is a Brownian bridge. Hence we can reject the H_0' in favor of H_a' if $CUSUM_i^{(2)}$ $(i = 1, 2)$ is large.

2.2 Jarque-Bera normality test

Omnibus statistics based on sample skewness and kurtosis have been used to test normality. Bowman and Shenton (1975) and Gasser (1975) give details of this method. Later Jarque and Bera (1980,1987) populate it among economists. It is related to the sample skewness partial sum process and the sample kurtosis process defined in (1.9) for $k = 3$ and $k = 4$. They correspond to

$$\hat{\rho}_n(x) = \frac{\hat{T}_n^{(3)}(x)/n}{\hat{\sigma}^3_{(n)}}, \ 0 \le x \le 1$$

and

$$\hat{\kappa}_n(x) = \frac{\hat{T}_n^{(4)}(x)/n}{\hat{\sigma}^4_{(n)}}, \ 0 \le x \le 1,$$

respectively. By Corollary 1.3,

$$\frac{n}{\sigma_\rho^2}\left(\hat{\rho}_n(1) - \lambda_3\right)^2 + \frac{n}{\sigma_\kappa^2}\left(\hat{\kappa}_n(1) - \lambda_4\right)^2 \xrightarrow{\mathcal{D}} \chi^2(2), \tag{2.14}$$

where, by (1.13),

$$\sigma_\rho^2 = E(B^{(3)}(1))^2 = (\lambda_6 - \lambda_3^2) + 3(3 + 3\lambda_3^2 - 2\lambda_4) + 3\lambda_3(\lambda_3/4 + 3\lambda_3\lambda_4/4 - \lambda_5)$$

and

$$\sigma_\kappa^2 = E(B^{(4)}(1))^2 = (\lambda_8 - \lambda_4^2) + 4\lambda_3(4\lambda_3 + 4\lambda_3\lambda_4 - 2\lambda_5) + 4\lambda_4(\lambda_4^2 - \lambda_6).$$

If the error distribution F is a normal distribution which is symmetric about 0, then $\lambda_3 = 0, \ \lambda_4 = 3, \ \sigma_\rho^2 = 6$ and $\sigma_\kappa^2 = 24$. (2.14) becomes

$$JB = \frac{n}{6}\hat{\rho}_n^2(1) + \frac{n}{24}\left(\hat{\kappa}_n(1) - 3\right)^2 \xrightarrow{\mathcal{D}} \chi^2(2). \tag{2.15}$$

Jarque and Bera (1987) prove that the omnibus test based on the JB statistic can be interpreted as a Lagrange muliplier (LM) test within the Pearson family of dstributions. They point out that it is asymptotically equivalent to the likelihood ratio test, implying it has the same asymptotic power characteristics including maximum local asymptotic power (Cox and Hinkley (1974)). Hence a test based on JB is asymptotically locally most powerful and (2.15) shows that JB is asymptotically distributed as $\chi^2(2)$. The hypothesis of normality is rejected for large sample size, if the computed value of JB is greater than the appropriate critical value of a $\chi^2(2)$.

2.3 Nonparametric density estimation

Assume that the error distribution F has a uniformly continuous density function $f(x)$ which is unknown. Let h_n be a sequence of positive numbers and $K(x)$ be a probability density function (kernel). Then the kernel density estimation of $f(x)$ based on the residuals is defined as

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{x-\hat{\epsilon}_t}{h_n}\right), \ x \in \mathbb{R}.$$

Its counterpart based on iid errors is defined as

$$f_n(x) = \frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{x-\epsilon_t}{h_n}\right), \ x \in \mathbb{R}.$$

Bai (1993) obtains the following uniform consistency for a stationary ARMA model without a mean parameter

$$\sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f(x)| = o_P(1)$$

under the assumptions

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

(ii) $\sup_{|x|>b} \ |x|K(x) \to 0$ as $b \to \infty$,

(iii) K is Lipschitz, i.e., there exists a constant C such that

$$|K(x) - K(y)| \le C|x-y|, \ \forall \ x, y \in \mathbb{R}.$$

and (A1) to (A3) and (A4').

From the proof of Theorem 1.1, it is easy to see that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} |\hat{\epsilon}_t - \epsilon_t| = O_P(1)$$

which implies

$$\sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f_n(x)| = o_P(1).$$

Thus we are able to extend Bai's result to stationary ARMA models with a unknown mean parameter. The detail is omitted.

3 Proofs

First we give two technique lemmas which will be used frequently in proofs. By (A2), $1/\Theta(z)$ and $1/\Phi(z)$ have power series expansions as

$$\frac{1}{\Theta(z)} = \sum_{i=0}^{\infty} \psi_i(\boldsymbol{\theta}) z^i$$

and

$$\frac{1}{\Phi(z)} = \sum_{i=0}^{\infty} \pi_i(\boldsymbol{\phi}) z^i.$$

Lemma 3.1 *If* (A2) *holds, the there are* $\epsilon > 0,\ 0 < \beta < 1$ *and* $M > 0$ *such that*

(i) $|\psi_i(\mathbf{u})| \le M\beta^i,\ 0 \le i < \infty$ *for all* $|\mathbf{u} - \boldsymbol{\theta}| \le \epsilon$,

(ii) $|\psi_i(\mathbf{u_1}) - \psi_i(\mathbf{u_2})| \le M|\mathbf{u_1} - \mathbf{u_2}| i\beta^{i-1},\ 0 \le i < \infty$ *for all* $|\mathbf{u_1} - \boldsymbol{\theta}| \le \epsilon$ *and* $|\mathbf{u_2} - \boldsymbol{\theta}| \le \epsilon$,

(iii) $|\pi_i(\mathbf{v})| \le M\beta^i,\ 0 \le i < \infty$ *for all* $|\mathbf{v} - \boldsymbol{\phi}| \le \epsilon$,

where $\mathbf{u} = (u_1, \ldots, u_q) \in \mathbb{R}^q$, $\mathbf{v} = (v_1, \ldots, v_p) \in \mathbb{R}^p$, *and we use* $|\cdot|$ *to denote the maximum norm of vectors.*

Proof: We refer to Bai (1993).

Lemma 3.2 *Suppose that* $E|\epsilon_0|^{k+\delta} < \infty$ *for an integer* $k \ge 1$ *and some* $\delta > 0$. *Let* $\zeta_t = h(\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$ *be* $\mathcal{F}_{t-1}$ *adapted with* $E\zeta_0^2 < \infty$, *where* $\mathcal{F}_t = \sigma(\epsilon_s : s \le t)$ *is the sigma field generated by the sequence* $\{\epsilon_t, \epsilon_{t-1}, \ldots\}$. *Then*

$$\sup_{0\le x\le 1} \left| \frac{1}{n} \sum_{t=1}^{[nx]} \epsilon_t^l \zeta_t - \frac{\mu_l [nx]}{n} E\zeta_0 \right| = o_P(1),\ 0 \le l \le k.$$

Proof: We refer to Lemma 3.6 of Kulperger and Yu (2003).

In the rest of this section, we will use the well known C_r inequality in many occasions without mentioning it. It is of

$$|x_1 + \cdots + x_m|^r \leq m^{r-1} \left(|x_1|^r + \cdots + |x_m|^r\right)$$

for any integer $m \geq 2$, $r > 1$, and $x_i \in \mathbb{R},\ i = 1, \ldots, m$.

To simplify the proof of Theorem 1.1 and others as well, we need to define a few notations. It follows from the definitions of $\hat{\epsilon}_t$ that

$$\begin{aligned}
\hat{\epsilon}_t - \epsilon_t &= -\sum_{i=1}^{p} \left(\hat{\phi}_i - \phi_i\right) X_{t-i} - \sum_{i=1}^{q} \left(\hat{\theta}_i - \theta_i\right) \epsilon_{t-i} - \sum_{i=1}^{q} \hat{\theta}_i \left(\hat{\epsilon}_{t-i} - \epsilon_{t-i}\right) \\
&\quad - \left(1 - \sum_{i=1}^{p} \hat{\phi}_i\right) (\hat{\mu} - \mu).
\end{aligned}$$

By repeated substitution and using the initial values $\hat{\epsilon}_0 = \hat{\epsilon}_{-1} = \cdots = \hat{\epsilon}_{-q+1} = 0$ we obtain

$$\begin{aligned}
\hat{\epsilon}_t - \epsilon_t &= Y_t(\hat{\boldsymbol{\theta}}) - \sum_{i=1}^{p} \left(\hat{\phi}_i - \phi_i\right) \sum_{j=0}^{t-1} \psi_j(\hat{\boldsymbol{\theta}}) X_{t-i-j} - \sum_{i=1}^{q} \left(\hat{\theta}_i - \theta_i\right) \sum_{j=0}^{t-1} \psi_j(\hat{\boldsymbol{\theta}}) \epsilon_{t-i-j} \\
&\quad - \left(1 - \sum_{i=1}^{p} \hat{\phi}_i\right) (\hat{\mu} - \mu) \sum_{j=0}^{t-1} \psi_j(\hat{\boldsymbol{\theta}}),
\end{aligned} \tag{3.16}$$

where

$$\begin{aligned}
Y_t(\hat{\boldsymbol{\theta}}) &= -\psi_t(\hat{\boldsymbol{\theta}})\epsilon_0 - \left\{\psi_{t+1}(\hat{\boldsymbol{\theta}}) + \psi_t(\hat{\boldsymbol{\theta}})\hat{\theta}_1\right\} \epsilon_{-1} - \cdots \\
&\quad - \left\{\psi_{t+q-1}(\hat{\boldsymbol{\theta}}) + \psi_{t+q-2}(\hat{\boldsymbol{\theta}})\hat{\theta}_1 + \cdots + \psi_t(\hat{\boldsymbol{\theta}})\hat{\theta}_{q-1}\right\} \epsilon_{-q+1}.
\end{aligned}$$

Let

$$\begin{aligned}
\xi_t(\mathbf{u}, \mathbf{v}) &= \sum_{i=1}^{p} v_i \sum_{j=0}^{t-1} \psi_j \left(\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\mathbf{u}\right) X_{t-i-j} \\
&\quad + \sum_{i=1}^{q} u_i \sum_{j=0}^{t-1} \psi_j \left(\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\mathbf{u}\right) \epsilon_{t-i-j}
\end{aligned}$$

and

$$Z_t(\mathbf{u},\mathbf{v},w) = -\left(1-\sum_{i=1}^{p}\phi_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{p} v_i\right) w \sum_{j=0}^{t-1}\psi_j\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right),$$

where $w \in \mathbb{R}$. Then by (3.16), we have

$$\hat{\epsilon}_t = \epsilon_t\left(\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}), \sqrt{n}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}), \sqrt{n}(\hat{\mu}-\mu)\right), \tag{3.17}$$

where

$$\begin{aligned} \epsilon_t(\mathbf{u},\mathbf{v},w) &= \epsilon_t + Y_t\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right) - \frac{1}{\sqrt{n}}\xi_t(\mathbf{u},\mathbf{v}) + \frac{1}{\sqrt{n}}Z_t(\mathbf{u},\mathbf{v},w) \\ &= \epsilon_t + \Lambda_t(\mathbf{u},\mathbf{v},w). \end{aligned} \tag{3.18}$$

Notice that, in the case where there is no a mean parameter, one can drop the term $Z_t(\mathbf{u},\mathbf{v},w)/\sqrt{n}$ in (3.18).

Proof of Theorem 1.1. First by (3.18) we have

$$\sum_{t=1}^{[nx]}\epsilon_t^k(\mathbf{u},\mathbf{v},w) = \sum_{t=1}^{[nx]}\epsilon_t^k + k\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\Lambda_t(\mathbf{u},\mathbf{v},w) + \sum_{l=2}^{k}\binom{k}{l}\sum_{t=1}^{[nx]}\epsilon_t^{k-l}\Lambda_t^l(\mathbf{u},\mathbf{v},w).$$

In case that $k=1$, there is no last term in the above expression. By (A4) for any $\delta>0$, there exists $b>0$ and n_0 such that

$$P\left(\sqrt{n}|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}|>b\right)\le\delta,\ P\left(\sqrt{n}|\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}|>b\right)\le\delta,\ P\left(\sqrt{n}|\hat{\mu}-\mu|>b\right)\le\delta$$

if $n\ge n_0$. Thus, by (3.17), we can prove (1.6) if we can show that

$$\sup_{0\le x\le 1}\ \sup_{|\mathbf{u}|\le b,|\mathbf{v}|\le b,|w|\le b}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\Lambda_t(\mathbf{u},\mathbf{v},w) + \frac{\mu_{k-1}[nx]}{n}\frac{1-\sum_{i=1}^{p}\phi_i}{1+\sum_{j=1}^{q}\theta_j}w\right| = o_P(1) \tag{3.19}$$

and

$$\sup_{|\mathbf{u}|\le b,|\mathbf{v}|\le b,|w|\le b}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|\epsilon_t|^{k-l}|\Lambda_t(\mathbf{u},\mathbf{v},w)|^l = o_P(1),\ l=2,\ldots,k. \tag{3.20}$$

To simplify the proofs of (3.19) and (3.20), we break them down into Lemmas 3.3 to 3.7 which are given in the back of this section. Thus (3.19) follows easily from (3.18), Lemmas 3.3, 3.5 and 3.6, while (3.18), Lemmas 3.3, 3.4 and 3.7 yield (3.20). This finishes the proof of (1.6).

By (1.6), it is easy to prove (1.7) if we can show that

$$\sqrt{n}\left|\hat{\gamma}^k_{(n)} - \gamma^k_{(n)}\right| = o_P(1)$$

which follows by the usual Δ method and (1.8). This completes the proof of Theorem 1.1.

Proof of Theorem 1.2. Along the line in proving Theorem 1.1, the proof of Theorem 1.2 should be trivial since there is no $Z_t(\mathbf{u}, \mathbf{v}, w)$ term in (3.18) and hence is omitted.

Proof of Theorem 1.4. By (3.18)

$$\bar{\epsilon}(\mathbf{u}, \mathbf{v}, w) = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t(\mathbf{u}, \mathbf{v}, w) = \bar{\epsilon}_t + \frac{1}{n}\sum_{t=1}^{n} \Lambda_t(\mathbf{u}, \mathbf{v}, w) = \bar{\epsilon}_t + \bar{\Lambda}(\mathbf{u}, \mathbf{v}, w).$$

Hence

$$\begin{aligned}
&\sum_{t=1}^{[nx]} \left(\epsilon_t(\mathbf{u}, \mathbf{v}, w) - \bar{\epsilon}(\mathbf{u}, \mathbf{v}, w)\right)^k \\
&= \sum_{t=1}^{[nx]} (\epsilon_t - \bar{\epsilon})^k + k\sum_{t=1}^{[nx]} (\epsilon_t - \bar{\epsilon})^{k-1}(\Lambda_t(\mathbf{u}, \mathbf{v}, w) - \bar{\Lambda}(\mathbf{u}, \mathbf{v}, w)) \\
&\quad + \sum_{l=2}^{k} \binom{k}{l} \sum_{t=1}^{[nx]} (\epsilon_t - \bar{\epsilon})^{k-l}(\Lambda_t(\mathbf{u}, \mathbf{v}, w) - \bar{\Lambda}(\mathbf{u}, \mathbf{v}, w))^l.
\end{aligned}$$

Thus we can prove (1.11) if we can show that

$$\sup_{0\le x\le 1} \sup_{|\mathbf{u}|\le b, |\mathbf{v}|\le b, |w|\le b} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nx]} (\epsilon_t - \bar{\epsilon})^{k-1} \left(\Lambda_t(\mathbf{u}, \mathbf{v}, w) - \bar{\Lambda}(\mathbf{u}, \mathbf{v}, w)\right) \right| = o_P(1) \qquad (3.21)$$

and

$$\sup_{|\mathbf{u}|\le b, |\mathbf{v}|\le b, |w|\le b} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\epsilon_t - \bar{\epsilon}|^{k-l} |\Lambda_t(\mathbf{u}, \mathbf{v}, w) - \bar{\Lambda}(\mathbf{u}, \mathbf{v}, w)|^l = o_P(1),\ \ l = 2, \ldots, k. \qquad (3.22)$$

In the following, k used in Lemmas 3.2 to 3.7 is different from k used in proving (3.21) and (3.22). It will be any integer between 1 and k of (3.21) and (3.22).

We first prove (3.22) for the case $k \geq 2$. Obviously (A3) and CLT imply that $\bar{\epsilon} = O_P(1/\sqrt{n})$. Letting $k = 1$ and $x = 1$ in Lemmas 3.3, 3.5 and 3.6, we obtain

$$\sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \left|\bar{\Lambda}(\mathbf{u}, \mathbf{v}, w)\right| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Again, letting $k = l \geq 2$ and $x = 1$ in Lemmas 3.3, 3.4 and 3.7, we have

$$\sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \sum_{t=1}^{n} |\Lambda_t(\mathbf{u}, \mathbf{v}, w)|^l = O_P(1), \ \ l = 2, \ldots, k.$$

Putting all above, together with (3.20), proves (3.22).

By using the binominal formula and $\bar{\epsilon} = O_P(1/\sqrt{n})$, (3.21) is reduced to

$$\sup_{0\leq x\leq 1} \sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nx]} \epsilon_t^l \left(\Lambda_t(\mathbf{u}, \mathbf{v}, w) - \bar{\Lambda}(\mathbf{u}, \mathbf{v}, w)\right) \right| = o_P(1), \ \ l = 0, \ldots, k-1,$$

which is true by Lemmas 3.3 and 3.5 under the assumption $(A4')$ and otherwise is further reduced to

$$\sup_{0\leq x\leq 1} \sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \left| \frac{1}{n} \sum_{t=1}^{[nx]} \epsilon_t^l \left(Z_t(\mathbf{u}, \mathbf{v}, w) - \bar{Z}(\mathbf{u}, \mathbf{v}, w)\right) \right| = o_P(1), \ \ l = 0, \ldots, k-1,$$

where $\bar{Z}(\mathbf{u}, \mathbf{v}, w) = \sum_{t=1}^{n} Z_t(\mathbf{u}, \mathbf{v}, w)/n$. Finally, by Lemmas 3.2 and 3.6, the above is reduced to

$$\sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \left| \bar{Z}(\mathbf{u}, \mathbf{v}, w) + \frac{1 - \sum_{i=1}^{p} \phi_i}{1 + \sum_{j=1}^{q} \theta_j} w \right| = o(1).$$

By the definition of $Z_t(\mathbf{u}, \mathbf{v}, w)$ and (i) and (ii) of Lemma 3.1, the above expression can be again reduced to

$$\left| \frac{1}{n} \sum_{t=1}^{n} \sum_{j=0}^{t-1} \psi_j(\boldsymbol{\theta}) - \frac{1}{1 + \sum_{j=1}^{q} \theta_j} \right| = o(1)$$

which follows easily by (i) of Lemma 3.1 and the fact that

$$\sum_{j=0}^{\infty} \psi_j\left(\boldsymbol{\theta}\right) = \frac{1}{\Theta(1)} = \frac{1}{1 + \sum_{j=1}^{q} \theta_j}. \tag{3.23}$$

This completes the proof of (1.11).

(1.11) implies that

$$\sqrt{n}\left|\hat{\sigma}^2_{(n)} - \sigma^2_{(n)}\right| = o_P(1).$$

Thus (1.12) follows easily by (1.11) and Δ method. Now we finish the proof of Theorem 1.4.

Lemma 3.3 *If* (A2) and (A3) *hold, then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 1$ *implies that*

$$\sup_{|\mathbf{u}|\leq b} \sum_{t=1}^{n} \left|\epsilon_t^{k-l} Y_t^l \left(\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\mathbf{u}\right)\right| = O_P(1) \text{ for any fixed } b > 0 \text{ and } l = 1, \ldots, k.$$

Proof: For the ϵ given in Lemma 3.1, when n is large enough, we have $b/\sqrt{n} \leq \epsilon$. Hence, by Lemma 3.1, we obtain

$$\begin{aligned}
\sup_{|\mathbf{u}|\leq b}\left|Y_t\left(\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\mathbf{u}\right)\right| &\leq M\beta^t|\epsilon_0| + \{M\beta^{t+1} + M\beta^t(|\boldsymbol{\theta}| + \epsilon)\}|\epsilon_{-1}| + \cdots \\
&\quad + \{M\beta^{t+q-1} + M\beta^{t+q-2}(|\boldsymbol{\theta}| + \epsilon) + \cdots + M\beta^t(|\boldsymbol{\theta}| + \epsilon)\}|\epsilon_{-q+1}| \\
&\leq \frac{M\max(|\boldsymbol{\theta}| + \epsilon, 1)}{1-\beta}\beta^t(|\epsilon_0| + \cdots + |\epsilon_{-q+1}|).
\end{aligned}$$

Thus

$$\begin{aligned}
\sup_{|\mathbf{u}|\leq b} \sum_{t=1}^{n} \left|\epsilon_t^{k-l} Y_t^l \left(\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\mathbf{u}\right)\right| &\leq \frac{M^l \max((|\boldsymbol{\theta}| + \epsilon)^l, 1)}{(1-\beta)^l} \sum_{t=1}^{\infty} |\epsilon_t|^{k-l}\beta^{tl}(|\epsilon_0| + \cdots + |\epsilon_{-q+1}|)^l \\
&= O_P(1)
\end{aligned}$$

since

$$\begin{aligned}
&E\left(\sum_{t=1}^{\infty} |\epsilon_t|^{k-l}\beta^{tl}(|\epsilon_0| + \cdots + |\epsilon_{-q+1}|)^l\right) \\
&= E|\epsilon_1|^{k-l} E(|\epsilon_0| + \cdots + |\epsilon_{-q+1}|)^l \sum_{t=1}^{\infty}\beta^{tl} < \infty.
\end{aligned}$$

This proves Lemma 3.3.

Lemma 3.4 *If* (A2) and (A3) *hold, then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 2$ *implies that*

$$\sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b} \sum_{t=1}^{n} \left|\epsilon_t^{k-l}\xi_t^l(\mathbf{u}, \mathbf{v})\right| = O_P(n) \text{ for any fixed } b > 0 \text{ and } l = 2, \ldots, k.$$

Proof: By Lemma 3.1

$$
\begin{aligned}
&\sup_{|\mathbf{u}|\le b,|\mathbf{v}|\le b} \sum_{t=1}^{n} \left|\epsilon_t^{k-l}\xi_t^l(\mathbf{u},\mathbf{v})\right| \\
\le\ & (2p)^{l-1}b^l \sum_{i=1}^{p}\sum_{t=1}^{n} |\epsilon_t|^{k-l} \sup_{|\mathbf{u}|\le b} \left|\sum_{j=0}^{t-1} \psi_j\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right) X_{t-i-j}\right|^l \\
& +(2q)^{l-1}b^l \sum_{i=1}^{q}\sum_{t=1}^{n} |\epsilon_t|^{k-l} \sup_{|\mathbf{u}|\le b} \left|\sum_{j=0}^{t-1} \psi_j\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right) \epsilon_{t-i-j}\right|^l \\
\le\ & (2p)^{l-1}(bM)^l \sum_{i=1}^{p}\sum_{t=1}^{n} |\epsilon_t|^{k-l} \left(\sum_{j=0}^{t-1}\beta^j |X_{t-i-j}|\right)^l \\
& +(2q)^{l-1}(bM)^l \sum_{i=1}^{q}\sum_{t=1}^{n} |\epsilon_t|^{k-l} \left(\sum_{j=0}^{t-1}\beta^j |\epsilon_{t-i-j}|\right)^l .
\end{aligned}
$$

Hence Lemma 3.4 follows if we can show that

$$
\max_{2\le l\le k}\max_{1\le i\le p}\sup_{t\ge 1} E|\epsilon_t|^{k-l}\left(\sum_{j=0}^{t-1}\beta^j|X_{t-i-j}|\right)^l < \infty
$$

and

$$
\max_{2\le l\le k}\max_{1\le i\le q}\sup_{t\ge 1} E|\epsilon_t|^{k-l}\left(\sum_{j=0}^{t-1}\beta^j|\epsilon_{t-i-j}|\right)^l < \infty.
$$

We just need to verify the first one since the second one follows similarly. Since ϵ_t and X_{t-i-j} are independent and $E|\epsilon_t|^{k-l} \equiv E|\epsilon_0|^{k-l}$, we only need to show

$$
\max_{1\le i\le p}\sup_{t\ge 1} E\left(\sum_{j=0}^{t-1}\beta^j|X_{t-i-j}|\right)^k < \infty.
$$

From $E|\epsilon_0|^k < \infty$, we have $E|X_0|^k < \infty$. On the other hand, it is easy to verify that for any integers $0 \le j_1, \cdots, j_k \le k,\ j_1+\cdots+j_k = k$

$$
E|X_{t_1}|^{j_1}\cdots|X_{t_k}|^{j_k} \le E|X_0|^k.
$$

Thus by multinominal expansion, we obtain

$$
E\left(\sum_{j=0}^{t-1}\beta^j|X_{t-i-j}|\right)^k \le E|X_0|^k\left(\sum_{j=0}^{t-1}\beta^j\right)^k \le \frac{E|X_0|^k}{(1-\beta)^k} < \infty.
$$

This proves Lemma 3.4

Lemma 3.5 *If* (A2) and (A3) *hold, then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 1$ *implies that*

$$\sup_{0\leq x\leq 1}\sup_{|\mathbf{u}|\leq b,|\mathbf{v}|\leq b}\frac{1}{n}\left|\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\xi_t(\mathbf{u},\mathbf{v})\right| = o_P(1) \text{ for any fixed } b>0.$$

Proof: By the definition of $\xi_t(\mathbf{u},\mathbf{v})$, to prove Lemma 3.5, it suffices to show that

$$\sup_{0\leq x\leq 1}\sup_{|\mathbf{u}|\leq b}\frac{1}{n}\left|\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\sum_{j=0}^{t-1}\psi_j\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right)X_{t-i-j}\right| = o_P(1),\ \ i=1,\ldots,p \tag{3.24}$$

and

$$\sup_{0\leq x\leq 1}\sup_{|\mathbf{u}|\leq b}\frac{1}{n}\left|\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\sum_{j=0}^{t-1}\psi_j\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right)\epsilon_{t-i-j}\right| = o_P(1),\ \ i=1,\ldots,q. \tag{3.25}$$

By Lemma 3.1,

$$\begin{aligned}
&\sup_{0\leq x\leq 1}\sup_{|\mathbf{u}|\leq b}\frac{1}{n}\left|\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\sum_{j=0}^{t-1}\psi_j\left(\boldsymbol{\theta}+\frac{1}{\sqrt{n}}\mathbf{u}\right)X_{t-i-j}\right| \\
\leq\ & \sup_{0\leq x\leq 1}\frac{1}{n}\left|\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\sum_{j=0}^{\infty}\psi_j\left(\boldsymbol{\theta}\right)X_{t-i-j}\right| + \frac{1}{n}\sum_{t=1}^{n}|\epsilon_t|^{k-1}\sum_{j=t}^{\infty}|\psi_j\left(\boldsymbol{\theta}\right)X_{t-i-j}| \\
&+\frac{bM}{n^{3/2}}\sum_{t=1}^{n}|\epsilon_t|^{k-1}\sum_{j=1}^{t-1}j\beta^{j-1}|X_{t-i-j}|.
\end{aligned}$$

Since

$$E\left(\sum_{t=1}^{\infty}|\epsilon_t|^{k-1}\sum_{j=t}^{\infty}|\psi_j\left(\boldsymbol{\theta}\right)X_{t-i-j}|\right) \leq E|\epsilon_0|^{k-1}E|X_0|\sum_{t=1}^{\infty}\sum_{j=t}^{\infty}M\beta^j < \infty$$

and

$$E\left(\sum_{t=1}^{n}|\epsilon_t|^{k-1}\sum_{j=1}^{t-1}j\beta^{j-1}|X_{t-i-j}|\right) = O(n),$$

to prove (3.24), it suffices to prove

$$\sup_{0\leq x\leq 1}\frac{1}{n}\left|\sum_{t=1}^{[nx]}\epsilon_t^{k-1}\sum_{j=0}^{\infty}\psi_j\left(\boldsymbol{\theta}\right)X_{t-i-j}\right| = o_P(1),\ \ i=1,\ldots,p. \tag{3.26}$$

Adapting the common backshift operator B for ARMA models, we have by (1.1) and (A2)

$$\Phi(B)X_{t-i} = \Theta(B)\epsilon_{t-i} \Longrightarrow \frac{1}{\Theta(B)}X_{t-i} = \frac{1}{\Phi(B)}\epsilon_{t-i},$$

which is

$$\sum_{j=0}^{\infty} \psi_j\left(\boldsymbol{\theta}\right) X_{t-i-j} = \sum_{j=0}^{\infty} \pi_j\left(\boldsymbol{\phi}\right) \epsilon_{t-i-j} = \zeta_t(\epsilon_{t-i}, \epsilon_{t-i-1}, \ldots).$$

Now (3.26) follows from Lemmas 3.1 and 3.2. Similarly to (3.24) one can prove (3.25). This finishes the proof of Lemma 3.5.

Lemma 3.6 *If* (A2) and (A3) *hold, then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 1$ *implies that*

$$\sup_{0\leq x\leq 1} \sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \left| \frac{1}{n}\sum_{t=1}^{[nx]} \epsilon_t^{k-1} Z_t(\mathbf{u}, \mathbf{v}, w) + \frac{\mu_{k-1}[nx]}{n} \frac{1 - \sum_{i=1}^{p}\phi_i}{1 + \sum_{j=1}^{q}\theta_j} w \right| = o_P(1)$$

for any fixed $b > 0$.

Proof: The proof is similar to that of Lemma 3.5. We skip some details. We only need to show that

$$\sup_{0\leq x\leq 1} \left| \frac{1}{n}\sum_{t=1}^{[nx]} \epsilon_t^{k-1} \sum_{j=0}^{\infty} \psi_j\left(\boldsymbol{\theta}\right) - \frac{\mu_{k-1}[nx]}{n} \frac{1}{1 + \sum_{j=1}^{q}\theta_j} \right| = o_P(1)$$

which follows easily from Lemma 3.2 and (3.23). This completes the proof of Lemma 3.6.

Lemma 3.7 *If* (A2) and (A3) *hold, then* $E|\epsilon_0|^k < \infty$ *for an integer* $k \geq 2$ *implies that*

$$\sup_{|\mathbf{u}|\leq b, |\mathbf{v}|\leq b, |w|\leq b} \sum_{t=1}^{n} |\epsilon_t|^{k-l} \left|Z_t(\mathbf{u}, \mathbf{v}, w)\right|^l = O_P(n) \text{ for any fixed } b > 0 \text{ and } l = 2, \ldots, k.$$

Proof: The proof is similar to that of Lemma 3.4 and hence is omitted.

References

[1] J. Bai. (1993). On the partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis.* **14** 247–260.

[2] J. Bai. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *The Annals of Statistics.* **22** 2051–2061.

[3] P. Billingsley. (1999). *Convergence of probability measures.* John Wiley & Sons.

[4] M. V. Boldin. (1990). On testing hypotheses in the sliding average scheme by the Kolmogorov-Smirnov and ω^2 tests. *Theory of Probability and its Applications (Transl of Teorija Verojatnostei i ee Primenenija).* **34** 699–704.

[5] M. V. (ed) Boldin and A. B. (ed) Aries. (1982). Estimation of the distribution of noise in an autoregression scheme. *Theory of Probability and its Applications (Transl of Teorija Verojatnostei i ee Primenenija).* **27** 866–871.

[6] K. O. Bowman and L. R. Shenton. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika.* **62** 243–250.

[7] P. J. Brockwell and R. A. Davis. (1991). *Time series. Theory and methods (Second edition).* Springer-Verlag Inc.

[8] R. L. Brown, J. Durbin, and J. M. Evans. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B, Methodological.* **37** 149–192.

[9] D. R. Cox and D. V. Hinkley. (1974). *Theoretical Statistics.* Wiley, New York.

[10] T. Gasser. (1975). Goodness-of-fit tests for correlated data. *Biometrika.* **62** 563–570.

[11] C. M. Jarque and A. K. Bera. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters.* **6** 255–259.

[12] C. M. Jarque and A. K. Bera. (1987). A test for normality of observations and regression residuals. *International Statistical Review.* **55** 163–172.

[13] H. L. Koul. (1991). A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference*. **29** 291–308.

[14] J.-P. Kreiss. (1991). Estimation of the distribution function of noise in stationary processes. *Metrika*. **38** 285–297.

[15] R. Kulperger and H. Yu. (2003). High moment partial sum processes of residuals in garch models and their applications. *To appear in Annals of Statistics*.

[16] R. J. Kulperger. (1985). On the residuals of autoregressive processes and polynomial regression. *Stochastic Processes and their Applications*. **21** 107–118.

[17] H. Lütkepohl. (1993). *Introduction to multiple time series analysis*. Springer-Verlag Inc.

[18] I. B. MacNeill. (1978). Limit process for sequences of partial sums of regression residuals. *The Annals of Probability*. **6** 695–698.

[19] I. B. MacNeill. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics*. **6** 422–433.

[20] H. Yu. (2003). Weak convergence of the sequential empirical processes of residuals of arma models with unknown mean. Submitted.