Journal of the Korean Statistical Society (1999), **28**: *3, pp 371–387*

ARMA Modeling for Nonstationary Time Series Data without Differencing†

DongWan Shin[1] and YouSung Park[2]

Abstract

For possibly nonstationary autoregressive moving average, modeling based on the original observations rather than the differenced observations is considered. Under this scheme, sample autocorrelation functions, parameter estimates, model diagnostic statistics, and prediction are all computed from the original data instead of the differenced data. The methods and results established under stationarity of data are shown to naturally extend to the nonstationarity of one autoregressive unit root. The sample ACF and PACF can be used for ARMA order determination. The BIC order is strongly consistent. The parameter estimates are asymptotically normal. The portmanteau statistic has chi-square distribution. The predictor is asymptotically equivalent to that based on the differenced data.

Keywords: ACF; BIC; Information criteria; Maximum likelihood estimator; PACF; Prediction; Unit root.

1. INTRODUCTION

We consider the autoregressive moving average (ARMA) time series model defined by

$$y_t+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p} = e_t+\beta_1 e_{t-1}+\cdots+\beta_q e_{t-q}, \tag{1.1}$$

$t = 1, 2, \ldots, n$, where y_t, $t = 1, \ldots, n$, are the observations, $\pi = (\phi', \beta')' = (\phi_1, \ldots, \phi_p, \beta_1, \ldots, \beta_q)'$ are vectors of unknown coefficients and e_t is an independent identically distributed ($i.i.d.$) error sequence with mean zero and variance σ^2. When y_t is stationary, the autoregressive characteristic equation

$$\phi(L) = 1 + \phi_1 L + \ldots + \phi_p L^p = 0$$

†This research was supported by a directed research grant from Korea Sceince and Engineering Foundation.

[1] Dept. of Statistics, Ewha Womans University, Seoul, 120-750, Korea

[2] Dept. of Statistics, Korea University, Seoul, 136-701, Korea

has all roots outside the unit circle. Many real time series reveal nonstationarity in such a way that one root of $\phi(L) = 0$ is unity and the other roots lie outside the unit circle. In order to express the possible unit root explicitly, we reparametrize model (1.1) into

$$(1-\rho L)A(L)y_t = B(L)e_t, \tag{1.2}$$

where ρ is a real value either $|\rho| < 1$ or $\rho = 1$, L is the lag operator such that $Ly_t = y_{t-1}$,

$$A(L) = 1 + \alpha_1 L + \cdots + \alpha_{p-1}L^{p-1} \text{ and } \quad B(L) = 1 + \beta_1 L + \cdots + \beta_q L^q.$$

The coefficient vectors ϕ and $\alpha = (\alpha_1, ..., \alpha_{p-1})'$ are related with each other by the identity $\phi(L) = (1 - \rho L)A(L)$, which is equivalent to

$$\phi_i = -\alpha_{i-1}+\alpha_i,\ i = 1, ..., p, \tag{1.3}$$

under $\rho = 1$, where $\alpha_0 = 1$ and $\alpha_p = 0$. Conditions required for our study are stated in Assumption 1.1 below.

Assumption 1.1. *The errors e_t are independent identically distributed with mean zero and variance σ^2, where σ^2 is positive. The roots of the equations $A(L)$ and $B(L)$, when evaluated at the true parameter, lie strictly outside the unit circle. Also $A(L)$ and $B(L)$ have no common roots. The true value of ρ is one.*

When $\rho = 1$ is known, many statistical procedures such as identification, estimation, diagnostic checking, and prediction are based on the differenced observations $z_t = y_t - y_{t-1}$. However, in the real world, the true value of ρ may be unknown. Also when one uses the differenced data z_t, the data y_t which are tentatively assumed to be nonstationary may in fact be stationary causing a model misspecification.

Statistical procedures based on y_t rather than the differenced data z_t are worth being investigated. One reason is that, when one does not make differencing, the data y_t which is now tentatively assumed to be stationary, may in fact be nonstationary. This may cause a model misspecification. We need to know whether the classical statistical results established under stationarity assumption are still valid or not when y_t is nonstationary.

The aim of this study is to extend the statistical methods of model identification, estimation, model diagnostic checking, and prediction based on the stationary differenced data to the cases in which the original nonstationary data are used instead of the differenced data. In Section 2, statistical properties of the sample autocorrelation functions and the partial autocorrelation functions are established. Also, the ARMA order selected by the BIC is shown to be strongly consistent. In Section 3, strong consistency and asymptotic normality of the maximum likelihood estimator are established. In Section 4, the portmanteau statistic for model diagnostic checking is shown to have approximate chi-square distribution. In Section 5, properties of the predictor are established. In Section 6, a concluding remark is given. In the Appendix, proofs of the theoretical results are given.

2. IDENTIFICATION

2.1. Properties of the sample ACF and PACF

The sample autocorrelation function (ACF) and the partial autocorrelation function (PACF) are the major tools for identification, determination of $(p,\ q)$, in the Box-Jenkins approach, in which the ARMA order $(p,\ q)$ is determined by investigating the declining patterns of the ACF and PACF. The statistical properties of the statistics are extensively established in the literature under the assumption of stationarity. The results are well summarized in the books of Fuller(1996, Sec. 6.2), Box and Jenkins(1976, Sec. 2), Brockwell and Davis(1991, Sec. 7), and many others. However, only few results have been established under the nonstationarity of $\rho = 1$. Hasza(1980) derived the limiting distribution of the sample ACF when data come from a nonstationary ARMA process with an AR unit root.

Note that the sample ACF $\hat{r}_k$ and PACF $\hat{\phi}_{kk}$ are defined by

$$\hat{r}_k = \sum_{t=k+1}^{n} y_t y_{t-k} / \sum_{t=1}^{n} y_t^2$$

and

$$\hat{\phi}_{kk} = \text{ regressioncoefficient of } y_{t-k} \text{ in the regression of } y_t \text{ on } y_{t-1},\ \ldots,\ y_{t-k}.$$

We show that, when y_t is nonstationary, the statistics ACF and PACF can be used for model identification in the same manner as they are used for stationary ARMA in the Box-Jenkins methods. Even though the population ACF is not defined when $\rho = 1$, we have a heuristic interpretation that $r_k \cong corr(y_t,\ y_{t+k}) \cong 1$.

Also, we know that, when y_t is nonstationary, the sample ACF $\hat{r}_k$ decays slowly in k as k increases. See, for example, Box and Jenkins(1976, pp. 200 — 201). The high significance of the ACF implies that the process y_t has highly significant AR part. On the other hand, population PACF

$$\phi_{kk} = corr(y_t,\ y_{t+k}|y_{t+1},\ \ldots,\ y_{t+k-1})$$

is well defined. Let $\phi_{z,kk} = corr(z_t,\ z_{t+k}|z_{t+1},\ \ldots,\ z_{t+k-1})$ be the PACF of the differenced data z_t. Then we have, as is shown in the Appendix,

$$\phi_{kk} \cong -\phi_{z,k-1,k-1}. \tag{2.1}$$

The fact (2.1) is also justified by the fact that the sample PACF $\hat{\phi}_{kk}$ converges in probability to ϕ_{kk} as is shown in Theorem 2.1 below. By (2.1), we see that the declining pattern of the PACF of lag k for y_t is just the same as that of the PACF of lag $(k-1)$ for z_t if $\rho = 1$. One implication is that, when the true model for y_t is nonstationary $AR(p)$ with $\rho = 1$, $\phi_{kk} = -\phi_{z,k-1,k-1}$ are all zero for $k > p$.

Theorem 2.1 below shows that the sample ACF and PACF are consistent. Theorem 2.2 below establishes limiting distributions. Proofs of all of the theorems are given in the Appendix.

Theorem 2.1. *Let Assumption 1.1 holds. Then, as* $n \to \infty$
(i) $\hat{r}_k \xrightarrow{p} 1,\ k = 1,\ 2,\ \ldots$, *(ii)* $\hat{\phi}_{kk} \xrightarrow{p} \phi_{kk},\ k = 2,\ 3,\ \ldots,$
where $\xrightarrow{p}$ *denotes convergence in probability.*

Theorem 2.2. *Let Assumption 1.1 holds. Then*
(i) $n(\hat{r}_k - 1)$ *converges in distribution to a nondegenerate distribution,* $k = 1,\ 2,\ \ldots$.
(ii) $n^{1/2}(\hat{\phi}_{kk} - \phi_{kk})$ *and* $n^{1/2}(\hat{\phi}_{z,k-1,k-1} - \phi_{z,k-1,k-1})$ *have the same normal limiting distribution,* $k = 2,\ 3,\ \ldots$.

The sample ACF converges rapidly to one at the rate of order n. The limiting distribution of the ACF is function of the standard Brownian motion and is given in Hasza(1980). The sample PACF converges at the usual rate of $n^{1/2}$. It is interesting to observe that limiting distribution of the normalized sample PACF is normal. This imply that the normal test $\hat{\phi}_{kk}/\ se(\hat{\phi}_{kk})$ for significance of ϕ_{kk}, $k = 2,\ 3,\ \ldots$ is still valid in the case of $\rho = 1$, where $se(\hat{\phi}_{kk})$ is the standard error of $\hat{\phi}_{kk}$.

2.2. Consistency of order defined by BIC

For determining the ARMA order (p, q), one can use the information criteria. In this subsection, we study the order estimator $(\hat{p}, \hat{q})$ which minimizes BIC given by

$$I(p, q) = \log \hat{\sigma}^2_{p,q} + (p+q)n^{-1} \log n$$

where

$$\hat{\sigma}^2_{p,q} = n^{-1} \sum_{t=\max(p,q)+1}^{n} \hat{e}^2_{t,p,q}$$

is the estimated error variance, $\hat{e}_{t,p,q}$ are the residuals computed from the maximum likelihood estimation of model (1.1) with ARMA order (p, q).

For order estimators in stationary ARMA times series models, see Fuller(1996, pp. 438-439), Brockwell and Davis(1991, Sec. 9.3) and references therein. Regarding strong consistency of order estimators for nonstationary AR time series are the works of Paulsen(1984), Tsay(1984), Potscher(1989), Huang(1990) and Wei(1992). However, no strong consistency of order estimator for nonstationary ARMA with an autoregressive unit root has been yet established. Now, in Theorem 2.3 below, the strong consistency of the order estimator which minimizes the information criterion BIC is established.

Theorem 2.3. *Let Assumption 1.1 holds and let (p_0, q_0) denote the true value of (p, q). Then, as $n \to \infty$, $(\hat{p}, \hat{q}) \to (p_0, q_0)$ a.s.*

Theorem 2.3 tells us the practical point that the information criteria BIC still works under the nonstationarity of $\rho = 1$ in the sense that it gives strongly consistent order estimator. Therefore, for identifying the ARMA order (p, q), we can use the information criteria even when the data come from a nonstationary time series.

3. ESTIMATION

In this section, we study the distributional properties of the maximum likelihood estimator and the least squares estimator for model (1.1) under the condition of nonstationarity, $\rho = 1$. We show that the distributions of the parameter estimates of the ARMA coefficients $\pi = (\phi_1, ..., \phi_p, \beta_1, ..., \beta_q)'$ are normal regardless of $|\rho| < 1$ or $\rho = 1$ except for ϕ_1 with $p = 1$.

For stationary ARMA models, we can find consistency and asymptotic normality of estimators in Fuller(1996, Theorems 8.4.1, 8.4.2) and Brockwell and

Davis(1991, Theorems 10.8.1, 10.8.2). For nonstationary AR process with a unit root, Fuller(1996, *p.* 556) showed that the ordinary least squares estimator $\tilde{\alpha}$ of α computed from y_t and that $\tilde{\alpha}_z$ computed from z_t are asymptotically equivalent in the sense $(\tilde{\alpha} - \tilde{\alpha}_z) = O_p(n^{-1})$. The estimator $(\tilde{\rho},\, \tilde{\alpha}')'$ is obtained by regressing y_t on $y_{t-1},\, z_{t-1},\, ...,\, z_{t-p+1},\; t = p+1,\, ...,\, n$, and the estimator $\tilde{\alpha}_z$ is obtained by regressing z_t on $z_{t-1},\, ...,\, z_{t-p+1},\; t = p,\, ...,\, n$. Therefore, $n^{1/2}(\tilde{\alpha} - \alpha)$ and $n^{1/2}(\tilde{\alpha}_z - \alpha)$ have the same normal limiting distribution. Chan and Wei(1988) derived the limiting distribution of the normalized ordinary least squares estimator when the autoregressive characteristic equation has multiple roots on the unit circle. Strong consistency results for nonstationary AR processes are established by authors such as Lai and Wei(1982, 1983), Pantula(1986), and others. We extend the asymptotic analysis of the estimators to our ARMA model (1.1).

Let $Y = (y_1,\, ...,\, y_n)'$. Let $\Lambda_n(\pi) = Var(\sigma^{-1}Y)$, which is defined under the condition of stationarity $|\rho| < 1$. Then the unconditional stationary Gaussian log-likelihood is, except for a constant,

$$L_n(\pi, \sigma) = -2^{-1}[\sigma^{-2}Y'\Lambda_n^{-1}(\pi)Y + \log\{\det\ \Lambda_n(\pi)\} + nlog\sigma^2].$$

Frequently, the generalized sum of squares error $Y'\Lambda_n^{-1}(\pi)Y$ is approximated by the conditional sum of squares error $S_n(\pi) = \sum_{t=1}^{n} e_t^2(\pi)$, where $e_t = e_t(\pi)$ are defined by the recursion

$$e_t = y_t + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} - \beta_1 e_{t-1} - \beta_2 e_{t-2} - ... - \beta_q e_{t-q}, \qquad (3.1)$$

$e_t = y_t = 0$ for $t \le 0$.

Note that $L_n(\pi,\, \sigma)$ is not defined when $\rho = 1$. However, the estimator $(\hat{\pi},\, \hat{\sigma})$ maximizing $L_n(\pi,\, \sigma)$ is defined regardless of the true value of ρ. The estimator is called the unconditional maximum likelihood estimator. On the other hand, $S_n(\pi)$ is well defined when $\rho = 1$. The estimator $\tilde{\pi}$ which maximizes $S_n(\pi)$ is called the conditional least squares estimator. Strong consistency of estimators and asymptotic normality of estimators are given below.

Theorem 3.1 (Shin and Fuller, 1996) *Let Assumption 1.1 hold. Then* $\hat{\pi}$ *and* $\tilde{\pi}$ *are strongly consistent.*

Theorem 3.2. *Let Assumption 1.1 hold. Then, for each* $\bar{\pi} = \hat{\pi}$ *and* $\tilde{\pi}$,

$$n^{1/2}(\bar{\pi} - \pi) \xrightarrow{d} N(0,\, PV^{-1}(\theta)P'), \qquad (3.2)$$

where $V(\theta)$ *is the Fisher information matrix of* $\theta = (\alpha',\, \beta')'$, $\xrightarrow{d}$ *denotes convergence in distribution,* $P = diag(P_\phi,\, I_q)$, $P_\phi = 0$ *for* $p = 1$,

$$P_\phi = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ .. & .. & .. & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}$$

is a $p \times (p-1)$ matrix for $p \geq 2$, and I_q is the $q \times q$ identity matrix.

Since the marginal distributions of (3.2) are normal, the normal tests using the pivotal statistics $\bar{\phi}_i / \ se(\bar{\phi}_i)$ and $\bar{\beta}_i / \ se(\bar{\beta}_i)$ for significance of ϕ_i and β_i are still valid even under the nonstationarity of $\rho = 1$, where $se(\bar{\phi}_i)$ and $se(\bar{\beta}_i)$ are standard errors of $\bar{\phi}_i$ and $\bar{\beta}_i$, respectively. Specifically, the normal tests $\bar{\phi}_p / \ se(\bar{\phi}_p)$ and $\bar{\beta}_q / \ se(\bar{\beta}_q)$ for significances of the last elements ϕ_p and β_q, respectively, are valid. This means that we don't need to bother about checking stationarity in using the pivotal statistics for significance of the ARMA parameters. One exception for the asymptotic normality of the estimates of autoregressive parameter is for $p = 1$, in which $\phi_1 = -\rho = -1$ and $(\bar{\phi}_1 + 1) / \ se(\bar{\phi}_1)$ converges in distribution to $\int_0^1 W dW / (\int_0^1 W^2)^{1/2}$, where W is a standard Brownian motion. The limiting distribution is not normal and is well described in Fuller(1996, Section 10.1.1). Note that the limiting distribution in Theorem 3.2 is a singular $(p+q)$ dimensional multivariate normal distribution because the $(p+q) \times (p+q)$ covariance matrix $PV^{-1}(\theta_0)P'$ is of rank $(p+q-1)$. A special linear function of $\bar{\phi}$ given by $n^{1/2}(\bar{\phi}_1 + \dots \ + \bar{\phi}_p + 1)$ has singular limiting distribution $N(0,\ 0)$. According to Shin and Fuller(1996), the limiting distribution of $n(\bar{\phi}_1 + \dots \ + \bar{\phi}_p + 1)$ is $A(1) \int_0^1 W dW / \int_0^1 W^2$.

4. MODEL DIAGNOSTIC

Validity of the estimated model is usually checked by investigating the residuals. Of widely used statistics are the residual ACF

$$\hat{r}_{ek} = \sum \hat{e}_t \hat{e}_{t-k} / \ \sum \hat{e}_t^2$$

and the adjusted portmanteau statistic

$$Q_m = n(n+2) \sum_{k=1}^{m} \hat{r}_{ek}^2 / (n-k)$$

where $\hat{e}_t$ are the residuals and m is a given positive integer. When y_t is stationary, authors such as Box and Pierce(1970) and McLeod(1978) established the limiting distributions of the statistics. When y_t is a nonstationary AR process having multiple roots on the unit circle, Shin and Lee(1996) derived the limiting distributions of the statistics. When y_t is a nonstationary ARMA with one autoregressive unit root, limiting distribution of Q_m is established by Park and Shin(1996) which is restated below for completeness.

Theorem 4.1 (Park and Shin, 1996) *Let Assumption 1.1 holds. Then, Q_m has approximately chi-square distribution with degrees of freedom $m-(p-1+q)$.*

Note that, loss of degrees of freedom for Q_m is $p-1+q$ even though $(p+q)$ parameters ϕ_1, ..., ϕ_p, β_1, ..., β_q are estimated. We see that estimating the unit root does not decrease the degrees of freedom.

5. PREDICTION

In this section, we study behaviors of predictors. Harvey(1981) showed that overdifferencing does not cause inflation for the prediction mean squared error. His analysis is based on the true model instead of estimated model. For stationary ARMA model, Fuller(1996, Section 8.5) showed that predictions based on true model and estimated model are equivalent to each other up to $O_p(n^{-1/2})$. He also derived mean squared error of predictions based on estimated model. Regarding nonstationary AR process with a unit root, Fuller and Hasza(1981) derived an expression for the prediction mean squared error when the predictor is computed from estimated model. In the same model, Shin and Lee(1997) investigated the effects of order misspecification on prediction.

The two estimation schemes using y_t and z_t provide two different prediction procedures. Let $\bar{\pi} = (\bar{\phi}', \bar{\beta}')$ be the maximum likelihood estimator $\hat{\pi}$ or the conditional least squares estimators $\tilde{\pi}$ based on y_t. The h-step ahead predictor based on the estimated parameter $\bar{\pi}$ is obtained from the recursion

$$\bar{y}_{n+h} = -\bar{\phi}_1\bar{y}_{n+h-1}-\cdots-\bar{\phi}_p\bar{y}_{n+h-p}+\bar{\beta}_1\bar{e}_{n+h-1}+\bar{\beta}_2\bar{e}_{n+h-2}+\cdots+\bar{\beta}_q\bar{e}_{n+h-q}, \quad (5.1)$$

$h = 1, 2, ...,$ where $\bar{y}_{n-k} = y_{n-k}$, $k = 1, 2, ..., p$, $\bar{e}_{n+k} = 0$, $k = 1, 2, ...,$ and $\bar{e}_{n-k}$ are estimators of e_{n-k}, $k = 1, 2, ...$ One set of estimators $\{\bar{e}_k\}_{k=1}^{n}$ is computed from the recursion (3.1) with estimated parameters. The predictors based on estimated models for z_t are computed in analogous ways. Let $\bar{\theta}_z = (\bar{\alpha}_z', \bar{\beta}_z')'$ be the maximum likelihood estimator $\hat{\theta}_z = (\hat{\alpha}_z', \hat{\beta}_z')'$ or the conditional least squares estimator $\tilde{\theta}_z = (\tilde{\alpha}_z', \tilde{\beta}_z')'$ computed from z_t. The corresponding estimators $\bar{\pi}_z$ are constructed from (1.3) using $(1, \bar{\alpha}_z')'$. The h-step ahead prediction $\bar{y}_{z,n+h}$

based on the differenced observations z_t are defined by using $\bar{\pi}_z$ instead of $\bar{\pi}$ in computing the predictions (5.1). In Theorem 5.1 below, we show that $\bar{y}_{n+h}$ and $\bar{y}_{z,n+h}$ are equivalent up to $O_p(n^{-1/2})$.

Theorem 5.1. *Let Assumption 1.1 hold. Then*

$$\bar{y}_{n+h} = \bar{y}_{z,n+h} - (\bar{\rho} - 1) \textstyle\sum_{j=0}^{\infty} d_j y_{n+h-1-j} + O_p(n^{-1}), \tag{5.2}$$

where d_j is defined by the power series expansion

$$\sum_{j=0}^{\infty} d_j L^j = [B(L)]^{-1}[A(L)]. \tag{5.3}$$

Note that the second term in (5.2) is $O_p(n^{-1/2})$ because $(\bar{\rho} - 1) = O_p(n^{-1})$ and $\sum_{j=0}^{\infty} d_j y_{n+h-1-j} = O_p(n^{1/2})$. This term is a result of estimating the unit root. Theorem 5.1 implies that the prediction interval computed under the estimated model for y_t has the preassigned coverage probability even when y_t is nonstationary.

6. CONCLUSION

We have investigated the large sample behaviors of statistical procedures for nonstationary ARMA modeling. One major result is that the standard asymptotic results established for stationary ARMA model extend to the nonstationary ARMA model with one AR unit root. For model identification, even when the data come from a nonstationary process, one can investigate the declining pattern of the sample ACF and PACF to determine model order in the same way as for stationary ARMA model. Also, the information criterion BIC gives strongly consistent model order. The normalized coefficient estimates have limiting normal distribution and the normal tests for the significance of ARMA coefficient is still valid even when the true process is nonstationary. The portmanteau statistic for model diagnostic for ARMA(p, q) has also the limiting distribution of the chi-square distribution. However, the degrees of freedom is $m - (p - 1 + q)$ instead of $(m - p - q)$, where m is the number of ACF whose squares are summed in the portmanteau statistic. The predictors based on the estimated model for the original data are equivalent to those based on the differenced data up to $O_p(n^{-1/2})$.

Proof of (2.1): Observe that

$$\phi_{kk} = corr(y_t,\, y_{t+k}|y_{t+1},\, \ldots,\, y_{t+k-1})$$

$$= corr(y_{t+1} - z_{t+1},\ y_{t+k-1} + z_{t+k} | y_{t+1},\ \dots,\ y_{t+k-1})$$

$$= -corr(z_{t+1},\ z_{t+k} | y_{t+1},\ \dots,\ y_{t+k-1})$$

$$= -corr(z_{t+1},\ z_{t+k} | y_{t+1},\ z_{t+2},\ \dots,\ z_{t+k-1}) \tag{A.1}$$

$$\cong -corr(z_{t+1}, z_{t+k} | z_{t+2}, \dots, z_{t+k-1}) = -\phi_{z,k-1,k-1}$$

where, in the transition (A.1), we have used the heuristic argument that y_{t+1} and $\{z_{t+1},\ z_{t+k}\}$ are almost uncorrelated because y_{t+1} is a sum of z_s for all $s \le t+1$ and the correlation of y_{t+1} with the one observation z_{t+1} is negligible.

Proof of (i) of Theorem 2.1: See Hasza(1980).

Proof of (ii) of Theorem 2.1: Let $Y_{t,k} = (y_{t-1},\ \dots,\ y_{t-k})'$ and $Z_{t,k} = (z_{t-1},\ \dots,\ z_{t-k})'$, $k = 2,\ 3,\ \dots$. Note that $\hat{\phi}_{kk}$ is the last element of the ordinary least squares estimator

$$\hat{\phi}_k = (\hat{\phi}_{k1},\ \dots,\ \hat{\phi}_{kk})' = [\sum\ Y_{t,k}Y'_{t,k}]^{-1}[\sum\ Y_{t,k}y_t]. \tag{A.2}$$

By Theorem 1 of Shin and Lee(1997), $\hat{\phi}_k \xrightarrow{p} \phi_k(k)$, where

$$\phi_k(k) = Q_k \begin{Bmatrix} 1 \\ \phi_{z,k-1}(k-1) \end{Bmatrix},\ k \ge 2,$$

$$\phi_{z,k-1}(k-1) = \Gamma_{k-1}^{-1}\gamma_{k-1},\ \Gamma_k = E(Z_{t,k}Z'_{t,k}),\ \gamma_k = E(Z_{t,k}z_t),$$

$k = 2,\ 3,\ \dots$, and

$$Q_k = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ 0 & 0 & -1 & \dots & 0 \\ .. & .. & .. & \dots & \dots \\ 0 & 0 & 0 & & -1 \end{pmatrix},\ k = 2,\ 3,\ \dots.$$

Since the last element of $\phi_k(k)$ is the last element $of - \phi_{z,k-1}(k-1)$ and hence is $-\phi_{z,k-1,k-1}$, we get the result.

Proof of (i) of Theorem 2.2: This is due to Hasza(1980).

Proof of (ii) of Theorem 2.2: Let $W_{t,k} = (y_{t-1},\ Z'_{t,k-1})'$. Using $W'_{t,k} = Y'_{t,k}Q_k$ and $Q_k^{-1}\phi_k(k) = (1,\ \phi'_{z,k-1}(k-1))'$, together with (A.2), we get, for $k = 2,\ 3,\ \dots$,

$$K_n^{-1}Q_k^{-1}(\hat{\phi}_k - \phi_k(k)) = [K_n \sum\ W_{t,k}W'_{t,k}K_n]^{-1}[K_n \sum\ W_{t,k}v_t]$$

where

$$v_t = y_t - Y'_{t,k}\phi_k(k) = y_t - W'_{t,k}\left\{\begin{smallmatrix}1\\ \phi_{z,k-1}(k-1)\end{smallmatrix}\right\} = z_t - Z'_{t,k-1}\phi_{z,k-1}(k-1),$$

$K_n = diag(n^{-1}, n^{-1/2}I_{k-1})$, and I_{k-1} is the $(k-1)$ dimensional identity matrix. By Lemma 1 – (b) of Pantula and Hall(1991), $\sum Z_{t,k-1}y_t = O_p(n)$. Therefore, $n^{-3/2}\sum Z_{t,k-1}y_t \xrightarrow{P} 0$ and the limiting distribution of $K_n \sum W_{t,k}W'_{t,k}K_n$ is block diagonal. So the last element of $K_n^{-1}Q_k^{-1}(\hat{\phi}_k - \phi_k(k))$ which $is - n^{1/2}(\hat{\phi}_{kk} - \phi_{kk})$ has the same limiting distribution as that of $[n^{-1}\sum Z_{t,k-1}Z'_{t,k-1}]^{-1}[n^{-1/2} \sum Z_{t,k-1}v_t]$ which is $n^{1/2}(\hat{\phi}_{z,k-1,k-1} - \phi_{z,k-1,k-1})$ and we get the result.

Proof of Theorem 2.3: Let (p, q) be the ARMA order under which model (1.1) is estimated. Let $p \geq p_0$ and $q \geq q_0$. Observe that $\hat{\sigma}^2_{p,q}$ and $\hat{\sigma}^2_{p_0,q_0}$ are strongly consistent by Theorem 3.1. Hence

$$I(p, q) = \log(\hat{\sigma}^2_{p_0,q_0}) + (p+q)n^{-1}\log\ n\ + o(1)\ \ a.s.$$

is an increasing function of p and q. Therefore, $P[\hat{p} \leq p_0,\ \hat{q} \leq q_0, \text{eventually}] = 1$. Hence, the proof is complete if we show

$$P[I_{p,q} > I_{p_0,q_0}, \text{eventually}] = 1 \text{ for } p < p_0 \text{ or } q < q_0. \tag{A.3}$$

Let $p < p_0$ or $q < q_0$. We let $S_{p,q}(\pi) = S_n(\pi)$ to denote the order (p, q) explicitly in the sum of squares $S_n(\pi)$. Then $\hat{\sigma}^2_{p,q} = n^{-1}S_{p,q}(\hat{\pi})$. We give another representation for $S_n(\pi)$ which is useful in our asymptotic analysis. Let $\theta = (\alpha', \beta')'$ be the vector of the ARMA coefficients of z_t. Note that the residuals computed from the recursion (3.1) has the following representation

$$e_t(\pi) = \sum_{j=0}^{t-1} d_j(\theta)z_{t-j}(\rho),\ t = 1, 2, ..., n \tag{A.4}$$

where $z_t(\rho) = y_t - \rho y_{t-1}$, $y_0 = 0$, and $d_j = d_j(\theta)$ is defined by (5.3). Now,

$$S_{p,q}(\pi) = \sum_{t=1}^{n} e_t^2(\psi) = (Y - \rho Y_0)'D'_n(\theta)D_n(\theta)(Y - \rho Y_0),$$

where $Y_0 = (y_0, ..., y_{n-1})'$ and the matrix $D_n(\theta)$ is an $n \times n$ lower triangular matrix with (i, j) element $d_{i-j}(\theta)$, $i \geq j$.

By (2.9) of Shin and Fuller(1996), for all $p \geq 0$ and $q \geq 0$ and uniformly in θ,

$$S_{p,q}(\pi) = Z'D'_n(\theta)D_n(\theta)Z + O((\text{loglog}\ \ n)^3)\ \ a.s.,$$

where $Z = (z_1, ..., z_n)'$. Hence,

$$n\hat{\sigma}^2_{p,q} = S_{p,q}(\hat{\pi}) = n\hat{\sigma}^2_{z,p,q} + O((\text{loglog } n)^3) \ a.s., \tag{A.5}$$

where $\hat{\sigma}^2_{z,p,q} = n^{-1}Z'D_n'(\hat{\theta})D_n(\hat{\theta})Z$. Let $\tilde{\theta}_z$ be the minimizer of $Z'D_n'(\theta)D_n(\theta)Z$. Then

$$\hat{\sigma}^2_{z,p,q} = n^{-1}Z'D_n'(\hat{\theta})D_n(\hat{\theta})Z \geq \tilde{\sigma}^2_{z,p,q} \tag{A.6}$$

where $\tilde{\sigma}^2_{z,p,q} = n^{-1}Z'D_n'(\tilde{\theta}_z)D_n(\tilde{\theta}_z)Z$. Now, by the consistency of $\hat{\theta}$ under $(p, q) = (p_0, q_0)$ of Theorem 3.1 and that of $\tilde{\theta}_z$ of Brockwell and Davis(1991, *p*. 384) under $(p, q) = (p_0, q_0)$, we have

$$\tilde{\sigma}^2_{z,p_0,q_0} = \hat{\sigma}^2_{p_0,q_0} + o(1) \ a.s. \tag{A.7}$$

Let

$$H_{p,q}(\sigma^2) = \log \ \sigma^2 + (p+q)n^{-1}\log \ n \ .$$

Now, by (A.5),

$$I_{p,q} = H_{p,q}(\hat{\sigma}^2_{p,q}) = H_{p,q}(\hat{\sigma}^2_{z,p,q}) + o(1) \ \ a.s.$$

and by (A.6),

$$H_{p,q}(\hat{\sigma}^2_{z,p,q}) \geq H_{p,q}(\tilde{\sigma}^2_{z,p,q}).$$

Also, by the arguments of Hannan and Rissanen(1982),

$$P[H_{p,q}(\tilde{\sigma}^2_{z,p,q}) > H_{p_0,q_0}(\tilde{\sigma}^2_{z,p_0,q_0}), \text{ eventually}] = 1.$$

Finally, by (A.7),

$$H_{p_0,q_0}(\tilde{\sigma}^2_{z,p_0,q_0}) = H_{p_0,q_0}(\hat{\sigma}^2_{p_0,q_0}) + o(1) a.s. \ = I_{p_0,q_0} + o(1) \ \ a.s.$$

and we get (A.3).

Lemma A.1. *Let $\theta = (\alpha', \beta')'$ be the vector of the ARMA coefficients of z_t. Let $\bar{\theta} = \hat{\theta}$ or $\tilde{\theta}$ be the maximum likelihood estimator or the conditional least squares estimator of θ computed from $\bar{\pi} = \hat{\pi}$ or $\tilde{\pi}$ through*

$$(1-\bar{\phi}_1 L - ... -\bar{\phi}_p L^p) = (1-\bar{\rho}L)(1-\bar{\alpha}_1 L - ... -\bar{\alpha}_{p-1}L^{p-1}) \tag{A.8}$$

and from the fact that $\bar{\rho}$ is the largest root of $(1-\bar{\phi}_1 L - ... -\bar{\phi}_p L^p) = 0$. Then

$$(\bar{\theta} - \theta) \xrightarrow{d} N(0, V^{-1}(\theta)).$$

Proof: Let $\psi = (\rho,\, \theta')'$ and let $\psi_0 = (1,\, \theta_0')'$ be the true value. Throughout this proof, we use $S_n(\psi)$ and $L_n(\psi,\, \sigma)$ for the conditional sum of squares error and the Gaussian log-likelihood, respectively, as functions of ψ.

We first derive the limiting distribution for the conditional least squares estimator $\tilde{\theta}$. For η and τ, subvectors of ψ, let $S_\eta = \partial S_n(\psi)/\partial\eta$ and let $S_{\eta\tau} = \partial^2 S_n(\psi)/\partial\eta\partial\tau'$. Now Taylor expansion of the score equation $\{S_\psi(\tilde{\psi}) = 0\}$ at $\psi = \psi_0$ gives

$$0 = S_\theta(\psi_0) + S_{\rho\theta}(\psi^*)(\tilde{\rho} - 1) + S_{\theta\theta}(\psi^*)(\tilde{\theta} - \theta_0)$$

for some ψ^* between $\tilde{\psi}$ and ψ_0. By Remark 1 of Shin and Fuller(1996),

$$(\tilde{\rho}-1) = O_p(n^{-1}) \qquad (A.9)$$

and $S_{\rho\theta}(\psi^*) = O_p(n)$. Therefore,

$$n^{1/2}(\tilde{\theta} - \theta_0) = -[S_{\theta\theta}(\psi^*)]^{-1}[S_\theta(\psi_0) + O_p(1)].$$

Now, by Brockwell and Davis(1991, *p.* 391),

$$n^{-1}S_{\theta\theta}(\psi^*) \xrightarrow{p} (\sigma_0)^2 V(\theta_0) \text{ and } n^{-1/2}S_\theta(\psi_0) \xrightarrow{d} N[0,\, (\sigma_0)^2 V(\theta_0)] \qquad (A.10)$$

and asymptotic normality of $\tilde{\theta}$ follows.

We next derive the asymptotic normality of the maximum likelihood estimator $\hat{\theta}$. Note that the log-likelihood $L(\psi,\, \sigma)$ is a sum of the log-likelihood of y_1 and the conditional log-likelihood of $Y_2 = (y_2,\, ...,\, y_n)'$ given y_1. Hence, the Gaussian log-likelihood is

$$L(\psi,\, \sigma) = -2^{-1}[(1-\rho^2)y_1^2/\{\sigma^2 a(\psi)\} + \log\{a(\psi)/(1-\rho^2)\} + \log\ \sigma^2$$

$$+\sigma^{-2}Q_n(\psi) + \log\ \det\Gamma_n(\theta) + (n-1)\log\ \sigma^2]$$

where

$$Q_n(\psi) = (Y_1 - \rho Y_2)'\Gamma_n^{-1}(\theta)(Y_1 - \rho Y_2),$$

$Y_1 = (y_1,\, ...,\, y_{n-1})'$, $\Gamma_n(\theta) = var(\sigma^{-2}(Y_1 - Y_2))$, and $a(\psi) = (1-\rho^2)var(y_1)/\ \sigma^2$. Now, by Brockwell and Davis(1991, *p.* 393),

$$n^{-1}\partial^2 \log\ \det\Gamma_n(\hat{\theta})/\ \partial\theta\partial\theta' = o_p(1)$$

and

$$n^{-1/2}\partial log\ \det\Gamma_n(\hat{\theta})/\ \partial\theta = o_p(1).$$

Also, by Shin and Fuller(1996, Lemma A.7), $|Q_n(\psi) - S_n(\psi)|$ and derivatives of $|Q_n(\psi) - S_n(\psi)|$ with respect to θ are all $O_p(1)$ uniformly in ψ. Moreover, by the same argument for (2.15) of Shin and Fuller(1996), $\partial a(\psi)/\partial\theta$ is bounded. Therefore, together with consistency of $(\hat{\psi}, \hat{\sigma})$,

$L_\theta(\hat{\psi}, \hat{\sigma}) = -2^{-1}S_\theta(\hat{\psi})/\hat{\sigma} + o_p(n^{1/2})$ and $L_{\theta\theta}(\hat{\psi}, \hat{\sigma}) = -2^{-1}S_{\theta\theta}(\hat{\psi})/\hat{\sigma} + o_p(n)$, (A.11)

where $L_\theta(\psi, \sigma) = \partial L(\psi, \sigma)/\ \partial\theta$ and let $L_{\theta\theta}(\psi, \sigma) = \partial^2 L(\psi, \sigma)/\ \partial\theta\partial\theta'$. Now, the asymptotic normality of $\hat{\theta}$ follows from (A.10) and (A.11).

Proof of Theorem 3.2: Note that $\bar{\phi}$ and $(\bar{\rho}, \bar{\alpha})$ are related with each other through (A.8) which gives

$$\bar{\phi}_i = -\bar{\rho}\bar{\alpha}_{i-1} + \bar{\alpha}_i,\ i = 1, \ldots, p,$$

where $\bar{\alpha}_0 = 0$ and $\bar{\alpha}_p = 0$. Now,

$$\bar{\phi}_i - \phi_i = -\bar{\rho}\bar{\alpha}_{i-1} + \bar{\alpha}_i - (-\alpha_{i-1} + \alpha_i)$$

$$= -(\bar{\alpha}_{i-1} - \alpha_{i-1}) + (\bar{\alpha}_i - \alpha_i) - (\bar{\rho} - 1)\bar{\alpha}_{i-1}$$

Hence, by (A.9),

$$n^{1/2}(\bar{\phi} - \phi) = P_\phi n^{1/2}(\bar{\alpha} - \alpha) + O_p(n^{-1/2}).$$

Now the limiting distribution follows from Lemma A.1 and

$$n^{1/2}(\bar{\pi} - \pi) = P n^{1/2}(\bar{\theta} - \theta) + O_p(n^{-1/2}).$$

Proof of Theorem 5.1: Let $\bar{e}_{n+h} = \bar{y}_{n+h} - y_{n+h}$ and let $\bar{e}_{z,n+h} = \bar{y}_{z,n+h} - y_{n+h}$. A Taylor expansion gives

$$\bar{e}_{n+h} = \bar{e}_{z,n+h} + (\bar{\rho} - 1)\partial e_{n+h}/\partial\rho + (\bar{\theta} - \bar{\theta}_z)'\partial e_{n+h}/\partial\theta.$$

By differentiating (A.4) with respect to ρ, we get $\partial e_{n+h}/\partial\rho = -\sum_{j=0}^{\infty} d_j y_{n+h-1-j}$. By equations (10.8.28) and (10.8.29) of Brockwell and Davis(1991), for each i, $\partial e_t/\partial\theta_i$ is a stationary autoregressive process and hence $\partial e_{n+h}/\partial\theta = O_p(1)$. Now, it remains to show $(\bar{\theta} - \bar{\theta}_z) = O_p(n^{-1})$. Since $\bar{\psi}$ and $\bar{\psi}_z$ minimize $S_n(\psi)$ and $S_n(1, \theta)$, respectively, we have $S_\theta(\bar{\rho}, \bar{\theta}) = 0$ and $S_\theta(1, \bar{\theta}_z) = 0$. Thus, Taylor expansion of $\{0 = S_\theta(\bar{\rho}, \bar{\theta}) - S_\theta(1, \bar{\theta}_z)\}$ gives us

$$0 = S_{\rho\theta}(\bar{\rho} - 1) + S_{\theta\theta}(\bar{\theta} - \bar{\theta}_z)$$

and hence

$$(\bar{\theta} - \bar{\theta}_z) = S_{\theta\theta}^{-1} S_{\rho\theta}(\bar{\rho} - 1) = O_p(n^{-1}) O_p(n) O_p(n^{-1}) = O_p(n^{-1}).$$

We have used the facts that $n^{-1} S_{\theta\theta}$ converges in probability to a positive definite matrix (Brockwell and Davis, 1991, *p.* 291), $S_{\rho\theta} = O_p(n)$ and $(\bar{\rho} - 1) = O_p(n^{-1})$.

References

Box, G. E. P. and Jenkins, G. M.(1976). *Time series analysis: forecasting and control,* 2nd edition, Holden-Day, San francisco.

Box, G. E. P. and Pierce, D. A. (1970). "Distribution of the residual autocorrelations in autoregressive integrated moving average time series models," *Journal of American Statistical Association,* **65**, 1509-26.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods,* 2nd-edition, Springer-Verlag, New York.

Chan, N. H. and Wei, C. Z.(1988). "Limiting distributions of least squares estimates of unstable autoregressive processes," *Annals of Statistics,* **16**, 367-401.

Fuller, W. A. (1996). *Introduction to Statistical Time Series,* 2nd edition, John Wiley & Sons, New York.

Fuller, W. A. and Hasza, H. P. (1981). "Properties of predictors for autoregressive time series," *Journal of American Statistical Association,* **76**, 155-161.

Hannan, E. J. and Rissanen, J. (1982). "Recursive estimation of ARMA order," *Biometrika,* **69**, 81-94.

Harvey, A. C.(1981). "Finite sample prediction and overdifferencing," *Journal of Time Series Analysis,* **2**, 221-232.

Hasza, D. P.(1980). "The asymptotic distribution of the sample autocorrelations for an integrated ARMA process," *Journal of American Statistical Association,* **75**, 349 — 352.

Huang, D.(1990). "Selecting order for general autoregressive models by minimum description length," *Journal of Time Series Analysis,* **11**, 107-118.

Lai, T. L. and Wei, C. Z. (1982). "Asymptotic properties of projections with applications to stochastic regression problems," *Journal of Multivariate Analysis,* **12**, 346-370.

Lai, T. L. and Wei, C. Z. (1983). "Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters," *Journal of Multivariate Analysis,* **13**, 1-23.

McLeod, A. I. (1978) "On distribution of residual autocorrelations in Box-Jenkins models," *Journal of Royal Statistical Society, Ser. B,* **40**, 296-302.

Pantula, S. G. (1986). "On asymptotic properties of the least squares estimators for autoregressive time series with a unit root," *Sankhya B,* **48**, 208-218.

Pantula, S. G. and Hall, A. (1991). "Testing for unit roots in autoregressive moving average models: an instrumental variable approach," *Journal of Econometrics,* **48**, 325-53.

Park, C. G. and Shin, D. W. (1996). "On the asymptotics of residuals in autoregressive moving average processes with one autoregressive unit root," *Statistics & Probability Letters,* **27**, 341-346.

Paulsen, J.(1984). "Order determination of multivariate autoregressive time series with unit roots," *Journal of Time Series Analysis,* **5**, 115-127.

Potscher, B. M.(1989). "Model selection under nonstationary autoregressive models and stochastic linear regression models," *Annals of Statistics* **17**, 1257-1274.

Shin, D. W. and Fuller, W. A. (1996). "Strong consistency of the maximum likelihood estimators for autoregressive moving average with an autoregressive unit root," Unpublished manuscript, Iowa State University, Ames, Iowa, USA.

Shin, D. W. and Lee, J. H.(1996). "Distribution of residual autocorrelations in nonstationary autoregressive processes," *Journal of Time Series Analysis,* **17**, 105-109.

Shin, D. W. and Lee, Y. D.(1997). "A study on misspecified nonstationary autoregressive time series with a unit root," *Journal of Time Series Analysis,* **18**, 475-484.

Tsay(1984). "Order selection in nonstationary autoregressive models," *Annals of Statistics,* **12**, 1425-1433.

Wei, C. Z. (1992). "On predictive least squares principles," *Annals of Statistics,* **20**, 1-42.