July 22, 1988

#1924

Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach

Sastry G. Pantula* and Alastair Hall**

* Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

** Department of Economics
North Carolina State University
Raleigh, NC 27695-8110

Currently on leave at: Department of Economics
University of Pennsylvania
Philadelphia, PA 19104

Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach

Sastry G. Pantula and Alastair Hall

ABSTRACT

In this paper we propose an approach, based on an instrumental variable estimator, for testing the null hypothesis that a process Y_t is an ARIMA (p,1,q) against the alternative that it is a stationary ARIMA (p+1,0,q) process. Our approach is an extension of the procedure suggested by Hall (1988) for the case p=0. We derive the limiting distributions of the instrumental variable estimator when the estimated model is either (i) the true model, (ii) the true model with a shift in mean included or (iii) the true model with a shift in mean and a linear time trend included. The performance of the test statistics is investigated using a Monte Carlo study.

Key Words: unit root tests, instrumental variable estimation, asymptotic distributions.

1. INTRODUCTION

There has recently been a growing interest in developing tests for unit roots in time series. Early contributions to this literature concentrated on the case where the true model is an autoregressive integrated moving average process of orders (p,1,0) hereafter will be denoted by ARIMA (p,1,0), with p finite. See, for example, Fuller (1976) and Dickey and Fuller (1979, 1981). In many applications these tests are inappropriate because the time series has a moving average component. For example, Schwert (1987) presents evidence that many macroeconomic time series can be modeled as ARIMA (p,1,q) processes with $q > 0$. Several methods exist in the literature for testing the hypothesis that the process is an ARIMA (p,1,q) process against the alternative that it is an ARIMA (p+1,0,q) process. Said and Dickey (1984) suggest approximating an ARIMA (p,1,q) by an ARIMA (p_T+1,0,0) and test the hypothesis that the process is ARIMA (p_T,1,0) against the alternative that it is ARIMA (p_T+1,0,0), where p_T increases with sample size T at the rate $T^{1/4}$. This method requires estimating a large number of nuisance parameters. When p and q are known, Said and Dickey (1985) suggest a criterion based on the nonlinear least squares estimators. Phillips (1987) and Phillips and Perron (1988) develop the limiting distributions of test statistics based on the regression coefficient of the lagged value (Y_{t-1}) in the regression of Y_t on Y_{t-1}, where possibly an intercept and a linear time trend are included as regressors. Under a very weak set of assumptions on the innovation sequence, they derive the limiting distributions of the regression coefficients and suggest modifications to the coefficients such that the modified test statistics have the same asymptotic distributions as those suggested by Dickey and Fuller (1979). To apply the test criteria suggested by Phillips and Perron (1988), it is necessary to approximate the ARIMA (p,1,q) model by an ARIMA (0,1,q_T) where q_T increases with the sample size. However, intuition suggests that this method may be unsatisfactory for finite q_T, if p is large.

In a recent paper, Hall (1988) proposes a test for unit root based on an instrumental variable (IV) estimator. Hall's analysis applies only to the case where the series is generated by an ARIMA (0,1,q) process. In this paper we extend Hall's framework to consider the case where the series is generated by an ARIMA (p,1,q) model. It is demonstrated that the test statistics are easily transformed to have the limiting distributions tabulated by Dickey and Fuller (1979). While intuition suggests that Hall's analysis can be easily generalized in this fashion, we believe that the results presented here are of practical importance for several reasons. First, the tests proposed here are very general and can be applied in many more circumstances than the tests suggested by Hall (1988). Second, our test statistics are computationally simpler than those suggested by Said and Dickey (1984, 1985) and Phillips and Perron (1988). Finally, it emphasizes the convenience of the instrumental variable approach to estimating autoregressive parameters in a mixed model.

An outline of the paper is as follows. In section 2 we derive the limiting distribution of the instrumental variable estimator. Results of a Monte Carlo study that investigates the performance of the IV estimator are given in section 3. We summarize our results in section 4. The proofs of the results are given in a mathematical appendix.

2. MAIN RESULTS

Consider the model

$$Y_t = Y_{t-1} + Z_t \tag{1}$$

$$Z_t = \sum_{j=1}^{p} \theta_j Z_{t-j} + u_t ,$$

and

$$u_t = \sum_{i=1}^{q} \phi_i e_{t-i} + e_t ,$$

where $t=1,2,\ldots,T$. We assume that (i) the initial vector $(Y_0,\ldots, Y_{-p-1})'$ has a fixed distribution independent of $\{e_t\}$; (ii) the roots $m_1, \ldots, m_p$ of the autoregressive characteristic equation $m^p - \Sigma_{j=1}^{p}\theta_j m^{p-j} = 0$ and the roots $m_1^*,\ldots,m_q^*$ of the moving average characteristic equation $m^q + \Sigma_{i=1}^{q}\phi_i m^{q-i} = 0$ are such that $|m_j| < 1$, $|m_i^*| < 1$ and $m_j \neq m_i^*$ for all i and j; and (iii) $\{e_t\}$ is a sequence of iid random variables with mean zero, variance σ_e^2 and bounded $(2+\epsilon)$th moments for some $\epsilon > 0$. Note that Z_t is essentially a stationary and invertible ARIMA (p,0,q) process. We denote

$$\gamma(h) = \lim_{t\to\infty} E[Z_t Z_{t-h}] , \tag{2}$$

$$\sigma_Z^{*2} = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) ,$$

$$= \lim_{T\to\infty} \mathrm{Var}(T^{-1/2} \sum_{i=1}^{T} Z_i) ,$$

$$\sigma_u^2 = \mathrm{Var}(u_t) = \sigma_e^2(1 + \sum_{i=1}^{q} \phi_i^2) ,$$

and

$$\sigma^2 = \lim_{T\to\infty} \text{Var}(T^{-1/2} \sum_{i=1}^{T} u_i) \quad . \tag{3}$$

We also use the notation $W(t)$, $0 \le t \le 1$, for a standard Brownian Motion and " $\Rightarrow$ " for convergence in distribution.

For the case $p = 0$ in model (1), Hall (1988) considers the instrumental variable (IV) estimator $\bar{\rho}_k$ for the coefficient of Y_{t-1}, where Y_{t-k} is used as an instrument. That is,

$$\bar{\rho}_k = (\sum_{t=k+1}^{T} Y_{t-k}Y_{t-1})^{-1} (\sum_{t=k+1}^{T} Y_{t-k}Y_t) \ . \tag{4}$$

Phillips (1987) derives the limiting distribution of $\bar{\rho}_k$ for $k=1$, assuming that the process Y_t satisfies (1) with conditions (i) - (iii).

Consider the regression model

$$Y_t = \alpha Y_{t-1} + \sum_{j=1}^{p} \theta_j Z_{t-j} + u_t \tag{5}$$

where $\alpha = 1$, $Z_t = Y_t - Y_{t-1}$ and Y_t satisfies the model (1) with conditions (i) - (iii). Since u_t is a moving average process of order q, the regressor variables Y_{t-1} and Z_{t-j} $(j \le q)$ are correlated with u_t. Therefore, the ordinary least squares estimators of α and $\theta = (\theta_1, \ldots, \theta_p)'$ in (5) are in general not consistent. We therefore suggest an IV approach to obtain consistent estimators of α and θ.

The choice of the instrumental variables should be such that the instrumental variables are correlated with the regressor variables and are uncorrelated with the errors u_t in the regression. Consider Y_{t-k} and

$L_{t-k} = (Z_{t-k}, Z_{t-k-1}, \ldots, Z_{t-k-p+1})'$ as instruments for Y_{t-1} and $X_{t-1} = (Z_{t-1}, \ldots, Z_{t-p})'$. If we take $k > q$ then clearly Y_{t-k} and L_{t-1} are uncorrelated with u_t. We will see that the rank of the matrix

$$A_k = \lim_{t\to\infty} \mathrm{Cov}(L_{t-k}, X'_{t-1})$$

$$= \begin{bmatrix} \gamma(k-1) & \gamma(k-2) & \ldots & \gamma(k-p) \\ \gamma(k) & \gamma(k-1) & \ldots & \gamma(k+1-p) \\ \vdots & \vdots & & \vdots \\ \gamma(k+p-2) & \gamma(k+p-3) & \ldots & \gamma(k-1) \end{bmatrix}, \tag{6}$$

plays an important role in deriving the properties of the IV estimator,

$$\begin{bmatrix} \bar{\alpha}_k \\ \bar{\theta}_{(k)} \end{bmatrix} = \begin{bmatrix} \sum_{t=k+1}^{T} Y_{t-k}Y_{t-1} & \sum_{t=k+1}^{T} Y_{t-k}X'_{t-1} \\ \sum_{t=k+1}^{T} L_{t-k}Y_{t-1} & \sum_{t=k+1}^{T} L_{t-k}X'_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=k+1}^{T} Y_{t-k}Y_t \\ \sum_{t=k+1}^{T} L_{t-k}Y_t \end{bmatrix}. \tag{7}$$

Extensions of the estimator $\bar{\alpha}_k$ to models where a constant and/or a linear trend are included as regressors, are considered later in the paper. Notice that if k=1 then $\bar{\alpha}_1$ and $\bar{\theta}_{(1)}$ are the ordinary least squares estimators of α and θ.

To derive the limiting distributions of $T(\bar{\rho}_k - 1)$ and $T(\bar{\alpha}_k - 1)$, it is necessary to determine the limiting behavior of certain functions of Y_t. These intermediary results are presented in Lemma 1.

Lemma 1: Suppose $\{Y_t\}$ is generated by (1) and assume that the conditions (i) - (iii) are satisfied. Then, for any fixed k,

(a) $\sigma_Z^{*2} = (1 - \sum_{j=1}^{p} \Theta_j)^{-2}\sigma^2$,

(b) $T^{-1} \sum_{t=k+1}^{T} Y_{t-k}Z_t \Rightarrow \sigma_Z^{*2}\xi + \sum_{j=k}^{\infty} \gamma(j)$

(c) $T^{-3/2} \sum_{t=k+1}^{T} Y_{t-k}Z_{t-j} \Rightarrow 0$, for any fixed $j > 0$

(d) $T^{-1} \sum_{t=k+1}^{T} Y_{t-k}u_t \Rightarrow (1 - \sum_{j=1}^{p} \Theta_j)^{-1} \left[\sigma^2\xi - b_k\right]$,

and

(e) $T^{-2} \sum_{t=k+1}^{T} Y_{t-k}Y_{t-1} \Rightarrow \sigma_Z^{*2}\Gamma$

where $\Gamma = \int_0^1 W^2(t)dt$; $\xi = 0.5[W^2(1) - 1]$; $W(t)$ is a standard Brownian motion and

$$b_k = \sum_{j=1}^{p} \Theta_j \sum_{i=0}^{j-1} \lim_{t\to\infty} E(u_tZ_{t-k-i}) .$$

Note that, in Lemma 1, if $k > q$ then $b_k = 0$. The results in Lemma 1 imply the following theorem.

Theorem 1: Suppose $\{Y_t\}$ is generated by (1) and assume that the conditions (i) - (iii) are satisfied. Then:

(a) for a fixed k,

$$T(\bar{\rho}_k - 1) \Rightarrow \left[\sigma_Z^{2*} \Gamma\right]^{-1}\left[\sigma_Z^{*2}\xi + \sum_{j=k}^{\infty} \gamma(j)\right] ; \qquad (8)$$

(b) for k=1,

$$T(\bar{\alpha}_1 - 1) \Rightarrow (1 - \sum_{j=1}^{p} \theta_j)\left[\sigma^2\Gamma\right]^{-1}\left[\sigma^2\xi - b_1 - \eta\right] \quad (9)$$

where b_1 is defined in Lemma 1 and η is defined in (A.6) of the appendix;

(c) for $k > q$,

(i) $T(\bar{\alpha}_k - 1) \Rightarrow (1 - \sum_{j=1}^{p} \theta_j)\Gamma^{-1}\xi$,

(ii) $(\sum_{t=k+1}^{T} Y_{t-k}Y_{t-1})^{1/2}(\bar{\alpha}_k - 1) \Rightarrow \sigma\,\Gamma^{-1/2}\xi$

and

(iii) $\bar{\theta}_{(k)}$ is consistent for θ ,

<u>provided</u> the matrix A_k, defined in (6), is nonsingular.

The practical usefulness of the results in Theorem 1 depends on the ease with which the statistics can be modified to have limiting distributions free of nuisance parameters θ, ϕ and σ_e^2. Recall that the percentiles for the distributions of $\Gamma^{-1}\xi$ and $\Gamma^{-1/2}\xi$ are given in Chapter 8 of Fuller (1976). Phillips (1987) considers the least squares estimator $\bar{\rho}_k$ with k=1 and suggests a modification to $T(\bar{\rho}_1 - 1)$ based on a consistent estimator of $(\sigma_Z^{*2})^{-1}\,\Sigma_{j=1}^{\infty}\gamma(j)$. Similar modifications can be made to the statistic $T(\bar{\rho}_k - 1)$ so that the limiting distribution of the modified statistic has the distribution of the random variable $\Gamma^{-1}\xi$. However, we will not pursue this approach since it requires the estimation of $\Sigma_{j=k}^{\infty}\gamma(j)$ using the infinite moving average representation of the process.

Note that the limiting distribution of the ordinary least squares estimator $\bar{\alpha}_1$ of the coefficient of Y_{t-1} in (5), on the other hand, depends on b_1 and η which are not necessarily zero. It is not easy to make modifications, similar to that of Phillips (1987), to $T(\bar{\alpha}_1 - 1)$ such that the modified statistic has the desired limiting distribution.

The most important application of Theorem 1 is summarized in the following Corollary.

<u>Corollary 1</u>: Under the conditions of Theorem 1, with $k > q$ such that the matrix $\mathbf{A}_k$ is nonsingular, we have

$$(1 - \sum_{j=1}^{p} \bar{\theta}_{j,k})^{-1} \, T(\bar{\alpha}_k - 1) \Rightarrow \Gamma^{-1}\xi \, , \qquad (10)$$

where $\bar{\boldsymbol{\theta}}_{(k)} = (\bar{\theta}_{1,k}, \ldots, \bar{\theta}_{p,k})'$ and $\bar{\alpha}_k$ are defined in (7).

Note that the correction factor to the statistic $T(\bar{\alpha}_k - 1)$ in (10) is "identical" to the correction factor derived in Fuller (1976, p. 374, Theorem 8.5.1) for the case k=1 and q=0. Therefore, the result in Corollary 1 is an extension of Theorem 8.5.1 of Fuller (1976). In order to use the test statistic in (10), we must choose $k > q$ such that $\mathbf{A}_k$ is nonsingular. We now present conditions on k for which the matrix $\mathbf{A}_k$ is nonsingular.

<u>Lemma 2</u>: Assume that $\{Z_t\}$ satisfies

$$Z_t = \sum_{j=1}^{p} \theta_j Z_{t-j} + \sum_{i=1}^{q} \phi_i e_{t-i} + e_t \, ,$$

where the coefficients $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, the initial values of Z_t and $\{e_t\}$ satisfy the conditions (i) - (iii). Then, the matrix $\mathbf{A}_k$ defined in (6) is nonsingular if either

(a) $\theta_p \neq 0$ and $k > q$,

or

(b) $\phi_q \neq 0$ and $k = q + 1$.

The results of Lemma 2 follow immediately from the Lemma of Stoica (1981). In addition, Stoica (1981) shows that A_k is singular if $\theta_p = 0$ and $k > q+1$. (He also suggests the use of the matrices A_k to determine the orders p and q of the Z_t process.) If we know the <u>true</u> order p of the autoregressive equation (i.e., $\theta_p \neq 0$), then we need to know only an upper bound for the order q of the moving average equation. In this case take k to be any value greater than the upperbound for q and use the true order p in the regression (5). On the other hand, if we know the <u>true</u> order q of the moving average equation then we need to know only an upper bound for the order p of the autoregressive equation. In this case take $k=q+1$ and use the upperbound for p as the value of p in the regression (5). If we do not know either p or q and over estimate both p and q, then A_k is singular and $\bar{\theta}_k$ may not be consistent for θ. Similar problem occurs when an ARIMA $(p+\ell,0,q+j)$ model with $\ell>0$, $j>0$ is fit to an ARIMA (p,0,q) process. The extra $\ell+j$ parameters in the ARIMA $(p+\ell,0,q+j)$ are not identified. In fact, the ARIMA $(p+\ell,0,q+j)$ model has $\min(\ell,j)$ common roots shared by the autoregressive and the moving average characteristic equations, when the true model is the ARIMA (p,0,q) model. One may use the autocorrelation function and the partial autocorrelation functions of the differenced series to identify p and q. Also, the extended autocorrelation function suggested by Tsay and Tiao (1984) may be used to identify p and q.

Note that since Z_t is essentially a stationary and invertible ARIMA (p,0,q) process, we have

$$\gamma(h) = \sum_{j=1}^{p} \theta_j \gamma(h-j) \quad , \text{ for } h > q . \tag{11}$$

Writing the equation (11) for h=k, k+1,..., k+p-1, in a matrix notation, for $k > q$, we get

$$a_k = A_k \theta$$

where $a_k = (\gamma(k), \ldots, \gamma(k+p-1))'$ and A_k is defined in (6). If A_k is nonsingular then the modified Yule-Walker estimator $\hat{\theta}^*_{(k)} = \hat{A}_k^{-1}\hat{a}_k$ will be consistent for θ where $\hat{A}_k$ and $\hat{a}_k$ are obtained by replacing $\gamma(h)$ by $\hat{\gamma}(h) = T^{-1}\Sigma_{t=h+1}^{T} Z_t Z_{t-h}$ in A_k and a_k. Also, $\hat{\theta}^*_{(k)}$ and $\bar{\theta}_{(k)}$ are asymptotically equivalent.

Before we present the extensions to models with a constant term and/or a linear time trend are included, we make a few comments regarding the statistic $(\Sigma_{t=k}^{T} Y_{t-k} Y_{t-1})^{1/2}(\bar{\alpha}_k - 1)$ considered in Theorem 1. Notice that the t-statistic $\hat{t}_k$ for testing α=1 in the instrumental variable estimation of (5) is asymptotically equivalent to

$$\hat{t}_k = s_u^{-1}(\sum_{t=k}^{T} Y_{t-k} Y_{t-1})^{1/2}(\bar{\alpha}_k - 1)$$

where

$$s_u^2 = T^{-1} \sum_{t=k+1}^{T} (Y_t - \bar{\alpha}_k Y_{t-1} - \sum_{j=1}^{p} \bar{\theta}_{j,k} Z_{t-j})^2 \quad .$$

Now from the result (c) of Theorem 1, we get

$$s_u s^{-1} \hat{t}_k \Rightarrow \Gamma^{-1/2} \xi$$

where s is a consistent estimator of σ and

$$\sigma^2 = \sigma_e^2\Big[1 + \sum_{i=1}^{q} \phi_i\Big]^2 \quad .$$

A consistent estimator for σ^2 can be obtained by estimating σ_e^2 and $\boldsymbol{\phi} = (\phi_1,\ldots,\phi_q)'$ using a method suggested in Hall (1988). However, this method requires the estimation of $\boldsymbol{\phi}$ and σ^2 using nonlinear least squares and is more complicated than the method in Corollary 1. Therefore, we will not discuss the modifications to t-statistics while discussing the models with a constant and/or a linear time trend are included.

Finally, we consider the extensions of the result in Corollary 1 to the case where an intercept and/or a linear time trend are included as regressors in (5). We define the instrumental variable estimators as follows:

$$\begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_k \\ \hat{\theta}_{(k)} \end{bmatrix} = \begin{bmatrix} T-k & \Sigma Y_{t-1} & \Sigma X'_{t-1} \\ \Sigma Y_{t-k} & \Sigma Y_{t-k}Y_{t-1} & \Sigma Y_{t-k}X'_{t-1} \\ \Sigma L_{t-k} & \Sigma L_{t-k}Y_{t-1} & \Sigma L_{t-k}X'_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma Y_t \\ \Sigma Y_{t-k}Y_t \\ \Sigma L_{t-k}Y_t \end{bmatrix}, \quad (13)$$

and

$$\begin{bmatrix} \tilde{\mu} \\ \tilde{\beta} \\ \tilde{\alpha}_k \\ \tilde{\theta}_{(k)} \end{bmatrix} = \begin{bmatrix} T-k & \Sigma(t-\bar{t}) & \Sigma Y_{t-1} & \Sigma X'_{t-1} \\ \Sigma(t-\bar{t}) & \Sigma(t-\bar{t})^2 & \Sigma(t-\bar{t})Y_{t-1} & \Sigma(t-\bar{t})X'_{t-1} \\ \Sigma Y_{t-k} & \Sigma(t-\bar{t})Y_{t-k} & \Sigma Y_{t-k}Y_{t-1} & \Sigma Y_{t-k}X'_{t-1} \\ \Sigma L_{t-k} & \Sigma(t-\bar{t})L_{t-k} & \Sigma L_{t-k}Y_{t-1} & \Sigma L_{t-k}X'_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma Y_t \\ \Sigma(t-\bar{t})Y_t \\ \Sigma Y_{t-k}Y_t \\ \Sigma L_{t-k}Y_t \end{bmatrix}, \quad (14)$$

where the summations are over t from k+1 to T and $\bar{t} = (T-k)^{-1} \Sigma_{t=k+1}^{T} t$.

To derive the asymptotic distributions of the parameter estimators in (13) and (14), we require Lemma 1 combined with the following.

Lemma 3: Under the conditions of Theorem 1,

(a) $T^{-3/2} \sum_{t=k+1}^{T} Y_{t-k} \Rightarrow (1 - \sum_{j=1}^{p} \theta_j)^{-1} \sigma B$,

(b) $T^{-5/2} \sum_{t=k+1}^{T} (t-\bar{t})Y_{t-k} \Rightarrow (1 - \sum_{j=1}^{p} \theta_j)^{-1} \sigma C$,

and

(c) $T^{-2} \sum_{t=k+1}^{T} (t-\bar{t})Z_{t-j} = O_p(T^{-1/2})$,

where

$$B = \int_0^1 W(t)dt$$

and

$$C = \int_0^1 tW(t)dt \ .$$

The following theorem establishes the asymptotic distributions of the IV estimators $\hat{\alpha}_k$ and $\tilde{\alpha}_k$.

Theorem 2: Under the conditions of Theorem 1, with $k \geq q$ is such that A_k is nonsingular, then

$$\text{(a)} \quad (1 - \sum_{j=1}^{p} \hat{\theta}_{j,k})^{-1} \ T(\hat{\alpha}_k - 1) \Rightarrow \left[\Gamma - B^2\right]^{-1}\left[\xi - W(1)B\right] , \qquad (15)$$

and

$$\text{(b)} \quad (1 - \sum_{j=1}^{p} \tilde{\theta}_{j,k})^{-1} \ T(\tilde{\alpha}_k - 1) \Rightarrow D^{-1}[\xi + A_1^*] , \qquad (16)$$

where

$$A_1^* = 12[C - 0.5B][B - 0.5W(1)] - W(1)\,B$$

and

$$D = \Gamma - 12\,C^2 + 12\,BC - 4B^2.$$

The limiting distributions of the statistics in (15) and (16) are tabulated in Table 8.5.1 of Fuller (1976). Therefore, the existing tables can be used to test the hypothesis that the process is nonstationary.

3. SIMULATION RESULTS

To gain some insight into the finite sample properties of the tests we performed a simulation study. Data were generated from the following model

$$Y_t = \alpha Y_{t-1} + \theta(Y_{t-1} - \alpha Y_{t-2}) + \phi e_{t-1} + e_t\,, \qquad (17)$$

t=1,...,T where e_t are iid N(0,1). The parameter values used in the simulation are:

$$T = 100,250,500;\ \alpha = 1,0.9;\ \theta = 0,\ \pm 0.65 \text{ and } \phi = 0,\ \pm 0.5,\ \pm 0.8\ .$$

The simulations were performed using PROC MATRIX in SAS. For each parameter configuration 1000 Monte Carlo replications were generated. We considered test criteria based on $T(\bar{\alpha}_k-1)$, $T(\hat{\alpha}_k-1)$ and $T(\tilde{\alpha}_k-1)$ for k=2,3 to test the hypothesis H_0:$\alpha = 1$ in (17). Each test statistic is compared to the 0.05 and 0.10 critical values given in Table 8.5.1 of Fuller (1976). The number of times the test criteria rejected the null hypothesis is presented in Tables 1 and 2.

Several aspects of the results are worth noting. From Table 1, we observe that when $\alpha = 1$ and $\phi \geq -0.5$, the empirical level of the test criterion based on $\bar{\alpha}_2$ is very close to the theoretical level. Also, the empirical level of the test criterion based on $\hat{\alpha}_2$ is close to the theoretical level when $\alpha = 1$

and $\phi \geq 0$. On the other hand, the test based on $\tilde{\alpha}_2$ tend to reject more often than expected. (This pattern was also exhibited in the simulations when the sample size was increased to 250. However, in this case the empirical levels for the tests based on $\hat{\alpha}_2$ and $\tilde{\alpha}_2$ were closer to the expected level than in the case T=100. When T=500, the empirical levels of the test based on $\tilde{\alpha}_2$ were close to the expected level for $\phi \geq 0$. For the sake of brevity the results for T=250 and 500 are not included here.) When the intercept and a linear time trend were erroneously included, larger number of samples are needed for the asymptotic approximations to be valid. When $\alpha = 1$ and $\phi = -0.8$ none of the test criteria performed well. This is partly due to the fact that when the moving average parameter is close to -1, the process in (17) is nearly stationary. Similar results were observed by Schwert (1988) for the case $\theta=0$. (See Pantula (1988) for a theoretical properties of the unit root tests when the moving average parameter is close to negative one.)

The results in Table 2 suggest that the test criteria based on instrumental variable estimators have reasonable power against the stationary alternative when $\alpha = 0.9$.

Our analysis in the previous section indicated that the test statistics may encounter problems when the model is over fit. It is interesting to note from the case $\theta = \phi = 0$ in Table 1, that no such problems occurred in our empirical study. We also performed simulations based on $\bar{\alpha}_3$, $\hat{\alpha}_3$ and $\tilde{\alpha}_3$. The results are observed to be very similar to the ones given in Tables 1 and 2 and hence are not included in the paper. On the basis of the empirical evidence, there does not appear to be a cost to over fitting the models in these sample sizes.

Finally, it is worth noting that results similar to that given in Tables 1 and 2 are observed when we used the true correction factor $(1 - \Theta)^{-1}$ (instead of $(1-\tilde{\Theta})^{-1}$) in the computation of the test statistics. This suggests that the cost in estimating Θ is minimal.

4. SUMMARY

In this paper we have proposed a test for a unit root in autoregressive moving average time series models based on an instrumental variable estimator. The main advantage of the instrumental variable approach is that it avoids the bias in Dickey and Fuller (1979) statistics. Our approach is an extension of the results suggested in Dickey and Fuller (1979) and Hall (1988) to mixed autoregressive moving average processes. The test statistics suggested in the paper have asymptotic distributions free of nuisance parameters and the existing tables can be used to test the hypothesis that the process has a unit root. Our estimators are simple to compute and simple to use. The instrumental variable approach requires neither iterative least squares algorithms (as in Said and Dickey (1985)) nor estimating increasing number of parameters (as in Said and Dickey (1984) and Phillips and Perron (1988)). Results for estimators where an intercept and/or a linear time trend are erroneously included in the model are also derived.

Using a Monte Carlo study, the performance of our test criteria is investigated. Even though the performance of the test criterion based on $\tilde{\alpha}_K$ is not up-to-par, the remaining test criteria performed reasonably well. The test criteria seem to maintain the significance level when the sample size is 100 and have reasonably high power against the alternative where the largest root is 0.9.

MATHEMATICAL APPENDIX

In this appendix we present proofs of the main results given in section 2. Recall that the process Y_t satisfies the model (1) and we assume that the conditions (i) - (iii), given in section 2, are satisfied. Before we present the proofs, we obtain some simple expressions for Y_t. Note that

$$\begin{aligned} Y_t &= Y_{t-1} + Z_t \\ &= \sum_{i=1}^{t} Z_i + Y_0 . \end{aligned} \tag{A.1}$$

Also, since

$$Z_t = \sum_{j=1}^{p} \theta_j Z_{t-j} + u_t$$

we have

$$\begin{aligned} \sum_{i=1}^{t} u_i &= \sum_{i=1}^{t} \left[Z_i - \sum_{j=1}^{p} \theta_j Z_{i-j} \right] \\ &= (1 - \sum_{j=1}^{p} \theta_j) \sum_{i=1}^{t} Z_i - \sum_{j=1}^{p} \theta_j \sum_{i=1}^{j} Z_{i-j} + \sum_{j=1}^{p} \theta_j \sum_{i=1}^{j} Z_{t+i-j} . \end{aligned}$$

Therefore,

$$\sum_{i=1}^{t} Z_i = cS_t + R_t \tag{A.2}$$

where

$$S_t = \sum_{i=1}^{t} u_i ,$$

$$c = (1 - \sum_{j=1}^{p} \theta_j)^{-1} ,$$

and

$$R_t = c \sum_{j=1}^{p} \theta_j \sum_{i=1}^{j} (Z_{i-j} - Z_{t+i-j}) .$$

Note also that

$$T^{-1}\sum_{t=2}^{T} Y_{t-1}Z_t = 0.5T^{-1}\left[Y_T^2 - \sum_{t=1}^{T} Z_t^2\right] + o_p(1). \quad (A.3)$$

Proof of Lemma 1:

(a) Recall that

$$\sigma_Z^{*2} = \lim_{T\to\infty} \operatorname{var}(T^{-1/2}\sum_{t=1}^{T} Z_t)$$

$$= \lim_{T\to\infty} \operatorname{var}(cT^{-1/2}S_T + T^{-1/2}R_T)$$

$$= c^2 \lim_{T\to\infty} \operatorname{var}(T^{-1/2}S_T)$$

$$= c^2\sigma^2$$

since $\operatorname{var}(T^{-1/2}R_T)$ and $\operatorname{cov}(T^{-1/2}S_T, T^{-1/2}R_T)$ converge to zero,

(b) Consider $T^{-1}\Sigma_{t=k+1}^{T} Y_{t-k}Z_{t-j}$, for a fixed $j \geq 0$. For $k \leq j$

$$T^{-1}\sum_{t=k+1}^{T} Y_{t-k}Z_{t-j} = T^{-1}\sum_{t=k+1}^{T} Y_{t-j-1}Z_{t-j} + \sum_{i=k}^{j} T^{-1}\sum_{t=k+1}^{T} Z_{t-i}Z_{t-j}$$

$$= T^{-1}\sum_{t=2}^{T} Y_{t-1}Z_t + \sum_{i=k}^{j} T^{-1}\sum_{t=k+1}^{T} Z_{t-i}Z_{t-j} + o_p(1).$$

Since Z_t is essentially a stationary process we know that

$$T^{-1}\sum_{t=k+1}^{T} Z_{t-i}Z_{t-j} \overset{P}{\to} \gamma(|i-j|) .$$

From (A.2), (A.3) and Theorem 3.1 (b) of Phillips (1987) we have

$$T^{-1}\sum_{t=2}^{T} Y_{t-1}Z_t = 0.5\left[c^2T^{-1}S_{T-1}^2 - T^{-1}\sum_{t=1}^{T} Z_t^2\right] + o_p(1)$$

$$\Rightarrow c^2\sigma^2\xi + 0.5\left[c^2\sigma^2 - \gamma(0)\right] ,$$

and hence

$$T^{-1}\sum_{t=k+1}^{T} Y_{t-k}Z_{t-j} \Rightarrow c^2\sigma^2\xi + 0.5\left[c^2\sigma^2 - \gamma(0)\right] + \sum_{i=k}^{j}\gamma(j-i) . \qquad (A.4)$$

Similarly, for k > j

$$T^{-1}\sum_{t=k+1}^{T} Y_{t-k}Z_{t-j} = T^{-1}\sum_{t=k+1}^{T} Y_{t-j-1}Z_{t-j} - \sum_{i=j+1}^{k} T^{-1}\sum_{t=k+1}^{T} Z_{t-i}Z_{t-j}$$

$$\Rightarrow c^2\sigma^2\xi + 0.5\left[c^2\sigma^2 - \gamma(0)\right] - \sum_{i=j+1}^{k}\gamma(i-j) . \qquad (A.5)$$

In particular, for j = 0 in (A.5),

$$T^{-1}\sum_{t=k+1}^{T} Y_{t-k}Z_t \Rightarrow \sigma_Z^{*2}\xi + \sum_{i=k}^{\infty}\gamma(i)$$

since

$$\sigma_Z^{*2} = c^2\sigma^2$$

$$= \gamma(0) + 2\sum_{i=1}^{\infty}\gamma(i) .$$

(c) From (A.4) and (A.5) we have

$$T^{-3/2}\sum_{t=k+1}^{T} Y_{t-k}Z_{t-j} \overset{P}{\rightarrow} 0.$$

(d) Note that, from (A.1) and (A.2)

$$T^{-1}\sum_{t=k+1}^{T} Y_{t-k}u_t = T^{-1}\sum_{t=k+1}^{T}\left[c\, S_{t-k} + R_{t-k} + Y_0\right]u_t$$

$$= c\, T^{-1}\sum_{t=k+1}^{T} S_{t-k}u_t + T^{-1}\sum_{t=k+1}^{T} R_{t-k}u_t + o_p(1).$$

From Lemma 1 (iv) of Hall (1988) we know that

$$T^{-1}\sum_{t=k+1}^{T} S_{t-k}u_t \Rightarrow \sigma^2\xi .$$

Now, since Z_t satisfies a pth order autoregressive equation with errors $u_t = e_t + \Sigma_{i=1}^{q}\phi_i e_{t-i}$, we get

$$T^{-1}\sum_{t=k+1}^{T} R_{t-k}u_t = -c\sum_{j=1}^{p}\theta_j\sum_{i=1}^{j} T^{-1}\sum_{t=k+1}^{T} Z_{t-k+i-j}u_t + o_p(1)$$

$$\overset{p}{\rightarrow} -c\sum_{j=1}^{p}\theta_j\sum_{i=1}^{j}\lim_{T\to\infty} E\left[u_t Z_{t-k+i-j}\right]$$

$$= -c\, b_k .$$

Therefore,

$$T^{-1}\sum_{t=k+1}^{T} Y_{t-k}u_t \Rightarrow c\,\sigma^2\xi - c\, b_k .$$

(e) Finally, from (A.1), (A.2) and Lemma 1 (i) of Hall (1988),

$$T^{-2}\sum_{t=k+1}^{T} Y_{t-k}Y_{t-1} = T^{-2}\sum_{t=k+1}^{T} Y_{t-k}^2 + T^{-2}\sum_{t=k+1}^{T} Y_{t-k}\left(\sum_{i=1}^{k-1} Z_{t-i}\right)$$

$$= c^2 T^{-2}\sum_{t=2}^{T} S_{t-1}^2 + O_p(T^{-1/2})$$

$$\Rightarrow c^2\sigma^2\Gamma . \quad \blacksquare$$

Proof of Theorem 1:

(a) Note that from (4) and Lemma 1,

$$T(\bar{\rho}_k - 1) = \left[T^{-2}\sum_{t=k+1}^{T} Y_{t-k}Y_{t-1}\right]^{-1}\left[T^{-1}\sum_{t=k+1}^{T} Y_{t-k}Z_t\right]$$

$$\Rightarrow \left[\sigma_Z^{*2}\Gamma\right]^{-1}\left[\sigma_Z^{*2}\xi + \sum_{j=k}^{\infty}\gamma(j)\right] .$$

(b) Recall that the ordinary least squares estimator of α is given by

$$\begin{bmatrix} \bar{\alpha}_1 \\ \bar{\theta}_{(1)} \end{bmatrix} = \begin{bmatrix} \sum_{t=2}^{T} Y_{t-1}^2 & \sum_{t=2}^{T} Y_{t-1}X_{t-1}' \\ \sum_{t=2}^{T} X_{t-1}Y_{t-1} & \sum_{t=2}^{T} X_{t-1}X_{t-1}' \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=2}^{T} Y_{t-1}Y_t \\ \sum_{t=2}^{T} X_{t-1}Y_t \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ \theta \end{bmatrix} + \begin{bmatrix} \sum_{t=2}^{T} Y_{t-1}^2 & \sum_{t=2}^{T} Y_{t-1}X_{t-1}' \\ \sum_{t=2}^{T} X_{t-1}Y_{t-1} & \sum_{t=2}^{T} X_{t-1}X_{t-1}' \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=2}^{T} Y_{t-1}u_t \\ \sum_{t=2}^{T} X_{t-1}u_t \end{bmatrix} .$$

Using the formula for partitioned inverse (p. 27, Searle (1970)), Lemma A.1 and A.4 it is easy to show that

$$T(\bar{\alpha}_1 - 1) = \left[T^{-2}\sum_{t=2}^{T} Y_{t-1}^2\right]^{-1}\left[T^{-1}\sum_{t=2}^{T} Y_{t-1}u_t\right]$$

$$- \left[T^{-2}\sum_{t=2}^{T} Y_{t-1}^2\right]^{-1}\left[T^{-1}\sum_{t=2}^{T} Y_{t-1}X_{t-1}'\left(T^{-1}\sum_{t=2}^{T} X_{t-1}X_{t-1}'\right)^{-1}T^{-1}\sum_{t=2}^{T} X_{t-1}u_t\right]$$

$$+ o_p(1)$$

$$\Rightarrow \left[\sigma_Z^{*2}\Gamma\right]^{-1}\left[c\sigma^2\xi - cb_1 - \eta\right]$$

where

$$\eta = \delta' A_1^{-1} \lim_{t\to\infty} E(X_{t-1}u_t) , \tag{A.6}$$

$$\delta = c\sigma^2\xi\, 1 + (\sum_{i=1}^{\infty} \gamma(i))1 + d$$

$$d = (d_1, d_2, \ldots, d_p)' ,$$

$$d_j = \sum_{i=1}^{j} \gamma(j-i)$$

$$1 = (1, 1, \ldots, 1)' ,$$

and A_1 is defined in (6).

(c) For $k > q$ such that A_k is nonsingular, we have from Lemma 1, with $D_T = \text{diag}(T, T^{1/2}, \ldots, T^{1/2})$

$$D_T \begin{bmatrix} \bar{\alpha}_k - 1 \\ \bar{\theta}_k - \theta \end{bmatrix} = \begin{bmatrix} T^{-2}\sum_{t=k+1}^{T} Y_{t-k}Y_{t-1} & T^{-3/2}\sum_{t=k+1}^{T} Y_{t-k}X_{t-1}' \\ T^{-3/2}\sum_{t=k+1}^{T} L_{t-k}Y_{t-1} & T^{-1}\sum_{t=k+1}^{T} L_{t-k}X_{t-1}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\sum_{t=k+1}^{T} Y_{t-k}u_t \\ T^{-1/2}\sum_{t=k+1}^{T} L_{t-k}u_t \end{bmatrix}$$

$$= \begin{bmatrix} c^2T^{-2}\sum_{t=2}^{T} S_{t-1}^2 & 0 \\ 0 & A_k \end{bmatrix}^{-1} \begin{bmatrix} c\, T^{-1}\sum_{t=k+1}^{T} S_{t-k}u_t \\ T^{-1/2}\sum_{t=k+1}^{T} L_{t-k}u_t \end{bmatrix} + o_p(1) .$$

Since Z_{t-j} depends on $\{u_{t-j-i}: i \geq 0\}$ and u_t is a qth order moving average process, we have for $j > q$,

$$T^{-1/2}\sum_{t=k+1}^{T} Z_{t-j}u_t = O_p(1) .$$

(In fact, one can show that $T^{-1/2}\sum_{t=k+1}^{T} L_{t-k}u_t \Rightarrow N$, where $N \sim N(0, A_1\sigma_u^2)$ independent of $W(t)$.)

Now the results (i) - (iii) of Theorem 1 (c) follow from Lemma 1. ∎

Note that Corollary 1 follows from c(i) and c(iii) of Theorem 1. Also, Lemma 2 follows immediately from the Lemma of Stoica (1981).

Proof of Lemma 3:

(a) From (A.1) and (A.2) we have

$$T^{-3/2} \sum_{t=k+1}^{T} Y_{t-k} = cT^{-3/2} \sum_{t=k+1}^{T} S_{t-k} + o_p(1)$$

and the result follows from Lemma 1 of Hall (1988).

(b) Similarly,

$$T^{-5/2} \sum_{t=k+1}^{T} (t-\bar{t})Y_{t-k} = cT^{-5/2} \sum_{t=k+1}^{T} (t-\bar{t})S_{t-k} + o_p(1)$$

and the result follows from Lemma 1 of Hall (1988).

(c) See Dickey and Fuller (1981, p. 1066). ∎

Proof of Theorem 2: This follows directly from Lemmas 1, 2 and 3 of this paper and Theorems 1 and 2 of Phillips and Perron (1988). ∎

ACKNOWLEDGEMENT

The research of Sastry G. Pantula is partially supported by the grant NSF-DMS-8610127 for which he is grateful to the National Science Foundation.

REFERENCES

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427-431.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057-1072.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.

Hall, A. (1988). Testing for a unit root in the presence of moving average errors. *Biometrika* (forthcoming).

Pantula, S. G. (1988). Asymptotic distributions of the unit root tests when the process is nearly stationary. Institute of Statistics Mimeograph Series No. 1922, Raleigh, NC.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, **55**, 277-302.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika*, **75**, 335-346.

Searle, S. R. (1970). *Linear Models*, Wiley, New York.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika*, **71**, 599-607.

Said, S. E. and D. A. Dickey (1985). Hypothesis testing in ARIMA (p,1,q) models. *Journal of the American Statistical Association*, **80**, 369-374.

Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics*, **20**, 73-103.

Schwert, G. W. (1988). Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* (forthcoming).

Stoica, P. (1981). On a procedure for testing the orders of time series. *IEEE Transactions on Automatic Control*, Vol. Ac-26, 572-573.

Tsay, R. S. and G. C. Tiao (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*, **79**, 84-96.

Table 1. Empirical Level of the Test Criteria Based on 1000 Replications.

$(\alpha=1,\ T=100,\ Y_t=\alpha Y_{t-1} + \theta(Y_{t-1}-\alpha Y_{t-2}) + \phi e_{t-1}+e_t)$

Test Statistic	θ	Level	ϕ				
			-0.8	-0.5	0	0.5	0.8
$T(\bar{\alpha}_2-1)(1-\bar{\theta})^{-1}$	-0.65	0.05	.159	.048	.035	.036	.043
		0.10	.203	.075	.074	.076	.088
	0	0.05	.158	.051	.047	.039	.037
		0.10	.199	.075	.090	.084	.081
	0.65	0.05	.166	.045	.045	.040	.038
		0.10	.246	.079	.090	.080	.080
$T(\hat{\alpha}_2-1)(1-\hat{\theta})^{-1}$	-0.65	0.05	.291	.094	.078	.077	.086
		0.10	.325	.132	.120	.119	.145
	0	0.05	.317	.105	.084	.078	.073
		0.10	.362	.142	.131	.132	.129
	0.65	0.05	.327	.073	.074	.067	.063
		0.10	.388	.111	.130	.121	.119
$T(\tilde{\alpha}_2-1)(1-\tilde{\theta})^{-1}$	-0.65	0.05	.388	.145	.107	.109	.130
		0.10	.413	.199	.166	.170	.196
	0	0.05	.458	.180	.122	.119	.114
		0.10	.488	.218	.192	.118	.182
	0.65	0.05	.409	.094	.127	.103	.100
		0.10	.525	.132	.174	.163	.160

Table 2. Empirical Power of the Test Criteria Based on 1000 Replications.

(α=0.9, T=100, $Y_t = \alpha Y_{t-1} + \theta(Y_{t-1} - \alpha Y_{t-2}) + \phi e_{t-1} + e_t$)

Test Statistic	θ	Level	ϕ				
			-0.8	-0.5	0	0.5	0.8
$T(\bar{\alpha}_2-1)(1-\bar{\theta})^{-1}$	-0.65	0.05	.488	.544	.733	.746	.814
		0.10	.504	.638	.889	.890	.929
	0	0.05	.640	.556	.744	.761	.757
		0.10	.651	.627	.882	.918	.916
	0.65	0.05	.873	.479	.622	.629	.629
		0.10	.897	.687	.820	.816	.816
$T(\hat{\alpha}_2-1)(1-\hat{\theta})^{-1}$	-0.65	0.05	.481	.444	.486	.484	.541
		0.10	.485	.525	.650	.650	.729
	0	0.05	.656	.450	.503	.512	.504
		0.10	.666	.531	.669	.680	.679
	0.65	0.05	.845	.317	.434	.417	.420
		0.10	.881	.439	.603	.592	.590
$T(\tilde{\alpha}_2-1)(1-\tilde{\theta})^{-1}$	-0.65	0.05	.491	.376	.350	.350	.392
		0.10	.498	.438	.487	.490	.529
	0	0.05	.668	.393	.381	.368	.365
		0.10	.681	.462	.501	.502	.499
	0.65	0.05	.771	.219	.312	.295	.295
		0.10	.882	.315	.442	.413	.413