\# 1903

July 6, 1987

# Testing for Unit Roots in Time Series Data

Sastry G. Pantula
North Carolina State University
Raleigh, NC 27695-8203

ABSTRACT

Let $Y_t$ satisfy the stochastic difference equation $Y_t = \sum_{j=1}^{p} \alpha_j Y_{t-j} + e_t$, for $t = 1, 2, \ldots$, where $e_t$ are independent and identically distributed random variables and the initial conditions $(Y_{-p+1}, \ldots, Y_0)$ are fixed constants. It is assumed that the true, but unknown, roots $m_1, m_2, \ldots, m_p$ of $m^p - \sum_{j=1}^{p} \alpha_j m^{p-j} = 0$ satisfy the hypothesis $H_d$: $m_1 = \ldots = m_d = 1$ and $|m_j| < 1$ for $j = d+1, \ldots, p$. We present a reparameterization of the model for $Y_t$ that is convenient for testing the hypothesis $H_d$. We consider the asymptotic properties of (i) a likelihood ratio type "F-statistic" for testing the hypothesis $H_d$, (ii) a least squares regression t-statistic, and (iii) a likelihood ratio type t-statistic for testing the hypothesis $H_d$ against the alternative $H_{d-1}$. Using these asymptotic results, we obtain two sequential testing procedures that are asymptotically consistent. Extensions to autoregressive moving average processes are also presented.

*Key Words and Phrases: Asymptotic Distributions, Differencing, Least Squares, Reparameterizations*

AMS 1980 Subject Classification: Primary 62M10; Secondary 62J05.

1. Introduction: Autoregressive integrated moving average (ARIMA) models are used to represent economic, meteorological and physical processes. The hypothesis regarding the number of unit roots in the model has important consequences for the interpretation of economic data. Methods for detecting the presence of unit roots and their applications have appeared in various journals. For example, Altonji and Ashenfelter (1980) use the unit root tests to investigate an equilibrium hypothesis for wage movements; Meese and Singleton (1983) apply the unit root tests to exchange rates and discuss the importance of unit root testing in the theory of linearized expectations; Nelson and Plosser (1982) discuss the relationship of unit roots to the effect that monetary disturbances have on macroeconomic series; and Schwert (1987) analyses several macroeconomic variables and compares different methods for detecting the presence of a single unit root. Schwert (1987) also considers ARIMA models with two unit roots for the variables: Consumer Price Index, Gross National product, Monetary Base and Total Civilian Noninstitutional Population. (All of the variables are transformed using natural logarithm.) Models with more than two unit roots, though very rare in economics, have some applications in physical sciences. Box and Pallesen (1978) analysed the x-coordinate of a missile trajectory using an ARIMA model (ARIMA (0,3,5)) with three unit roots.

In this paper we present methods for testing the hypothesis that the process has a specified number of unit roots. We begin the discussion with pure autoregressive processes.

Let the time series $\{Y_t\}$ satisfy

$$Y_t = \Sigma_{j=1}^{p}\alpha_j Y_{t-j} + e_t \ , \ t = 1,2,\ldots, \qquad (1.1)$$

where $\{e_t\}_{t=1}^{\infty}$ is a sequence of iid random variables with mean zero and variance $\sigma^2$. It is assumed that $E[|e_t|^{4+\delta}] < M$ for some $\delta > 0$ . It is further assumed that the initial conditions $(Y_{-p+1},\ldots, Y_0)$ are known constants. The time series is said to be an autoregressive process of order p. Let

$$m^p - \Sigma_{j=1}^{p}\alpha_j m^{p-j} = 0 \qquad (1.2)$$

be the characteristic equation of the process. The roots of (1.2), denoted by $m_1, \ldots, m_p$ are called the characteristic roots of the process. Assume that $1 \geq |m_1| \geq |m_2| \geq \ldots \geq |m_p|$ .

Let the observations $Y_1, Y_2, \ldots, Y_n$ be available. It is assumed that $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ and $\sigma^2$ are unknown. The least squares estimator $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ is obtained by regressing $Y_t$ on $Y_{t-1}, \ldots, Y_{t-p}$. Lai and Wei (1983) established the strong consistency of $\hat{\boldsymbol{\alpha}}$ under a very general set of conditions, and Chan and Wei (1986) obtained the asymptotic distribution of $\hat{\boldsymbol{\alpha}}$. If $|m_j|<1$, $j = 1,\ldots,p$, then $Y_t$ converges to a weakly stationary process as $t \to \infty$. In the stationary case, the asymptotic properties of $\hat{\boldsymbol{\alpha}}$ and of the related likelihood ratio type "F-statistic" are well known. See Mann and Wald (1943), Anderson (1971) and Hannan and Heyde (1972).

Dickey and Fuller (1979) and Fuller (1979) have considered testing for a single unit root in a p-th order autoregressive process (i.e., $m_1 = 1$ and $|m_j| < 1$, $j=2,\ldots,p$). Phillips (1987) and Phillips and Perron (1986) present alternative methods for testing a single unit root in a general time series setting. Their approach is nonparametric with respect to nuisance parameters and hence allows for a wide class of ARIMA models with a single unit root. Hasza and Fuller (1979) considered a p-th order autoregressive process with $m_1 = m_2 = 1$ and $|m_j| < 1$ for $j = 3,4, \ldots, p$. They characterized the asymptotic distribution of the likelihood ratio type "F-statistic" for testing $m_1 = m_2 = 1$. Yajima (1985) proposed a method of estimating the number of unit roots using $\hat{\sigma}_n^2(k,d)$, k large and for reasonable choices of d, where $\hat{\sigma}_n^2$ (k,d) is the residual mean square error obtained by fitting a kth order autoregression to the dth difference of the data. For a second order process, Sen (1985) studied the asymptotic distribution of the regression t-statistic for testing $H_1$: $m_1 = 1$ and $|m_2|<1$, under the assumption $H_2$: $m_1 = m_2 = 1$. He observed empirically that the probability of the Dickey and Fuller (1979) criterion rejecting the hypothesis $H_1$, in favor of the hypothesis $H_0$: $|m_1| < 1$ and $|m_2| < 1$, is higher under $H_2$ than under $H_1$. That is, it is more likely to conclude incorrectly that the process is stationary when there are really two unit roots present than when there is just one unit root present. Pantula (1985) observed empirically that the probability of rejecting $H_1$ increases with the number of unit roots present. It is not clear how the modified procedure of Phillips (1987) behaves under the hypothesis $H_d$, $d > 1$.

In this paper we (i) present a reparameterization of the model (1.1) that is convenient for testing the hypothesis $H_d$: $m_1 = \ldots = m_d = 1$ and $|m_j| < 1$, $j = d+1,\ldots,p$, (ii) characterize the asymptotic distribution of the likelihood ratio type "F-statistic," the regression t-statistic and the likelihood ratio type t-statistic under the hypothesis $H_d$ for $d = 0, 1, \ldots, p$, and (iii) present two asymptotically consistent sequential procedures to test $H_d$ versus $H_{d-1}$.

The procedures are such that, asymptotically, the chance of rejecting the hypothesis $H_d$ in favor of $H_{d-1}$ is smaller than the specified level ($\alpha$) when the process has d or more unit roots and is equal to one when the process has less than d unit roots. The procedures presented here successfully answer the question "how many times should one difference the series to attain stationarity?" A proof of the main results is presented in Appendix A. Extensions of the procedures to autoregressive moving average processes is included in Appendix B.

2. Notation and the Main Results: Consider the following reparameterization of the model (1.1),

$$Y_{p,t} = \sum_{i=1}^{p} \beta_i Y_{i-1,t-1} + e_t \tag{2.1}$$

where $Y_{i,t} = (1-B)^i Y_t$ = ith difference of $Y_t$; and B is the back shift operator. Comparing (1.1) with (2.1) it is easy to show that the vector of parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ is linearly related to the parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and the relation is given by

$$\boldsymbol{\alpha} = \mathbf{T}\,\boldsymbol{\beta} + \mathbf{c} \tag{2.2}$$

where $\mathbf{T}$ is a pxp nonsingular upper triangular matrix with $T_{ij} = (-1)^{i-1}\binom{j-1}{i-1}$ for $j \geq i$ and $\mathbf{c} = (c_1, \ldots, c_p)'$ with $c_i = (-1)^{i-1}\binom{p}{i}$. We now show that the hypothesis $H_d$: $m_1 = \ldots = m_d = 1$ can be tested using a test for $\beta_1 = \ldots = \beta_d = 0$ in the model (2.1).

Lemma 1: Let $Y_t$ satisfy the model (1.1). Then, $m_1 = \ldots = m_d = 1$ if and only if $\beta_1 = \ldots = \beta_d = 0$ .

Proof: In terms of the characteristic roots, the pth order autoregressive model (1.1) can be written as

$$\pi_{i=1}^{p} [(1-m_i) + m_i(1-B)]Y_t = e_t \quad . \tag{2.3}$$

Also, from (2.1) we have

$$[(1-B)^p - \sum_{i=1}^{p} \beta_i(1-B)^{i-1}B]Y_t = e_t \ ,$$

which can also be written as

$$[(1-B)^p(1+\beta_p) - \sum_{i=1}^{p-1} (1-B)^i(\beta_i - \beta_{i+1}) - \beta_1]Y_t = e_t \ . \tag{2.4}$$

Now, comparing (2.3) with (2.4) we get

$$\beta_1 = -\pi_{i=1}^{p}(1-m_i) \ ,$$

$$\beta_k = \beta_{k-1} - \Sigma m_{i_1} \ldots m_{i_{k-1}} (1-m_{j_1}) \ldots (1-m_{j_{p-k+1}}) \ ,$$

and $\qquad j = 2, \ldots, p-1 \ .$

$$\beta_p = \pi_{i=1}^{p} m_i - 1 \ , \tag{2.5}$$

where the summation is over all possible indices $i_{k-1} = \{i_1,\ldots,i_{k-1}\}$ and $\{j_1,\ldots,j_{p-k+1}\} = \{1,\ldots,p\} - i_{k-1}$. Note that if $m_1 = \ldots = m_d = 1$, then $\beta_1 = \ldots = \beta_d = 0$ and $\beta_{d+1} = -\pi_{i=d+1}^{p}(1-m_i)$. We prove the converse by induction. Since $1 \geq |m_1| \geq \ldots \geq |m_p|$ and $\beta_1 = -\pi_{i=1}^{p}(1-m_i)$, we have that $\beta_1 = 0$ implies $m_1 = 1$. Suppose $\beta_1 = \ldots = \beta_d = 0$ implies that $m_1 = \ldots = m_d = 1$. Now, suppose $\beta_1 = \ldots = \beta_{d+1} = 0$. Then, clearly $m_1 = \ldots = m_d = 1$, $\beta_{d+1} = -\pi_{i=d+1}^{p}(1-m_i) = 0$ and hence $m_{d+1} = 1$. []

Note that the additional assumption $|m_j| < 1$ for $j = d+1,\ldots,p$ in the hypothesis $H_d$ imposes extra conditions on $\beta_j$ . In particular, $\beta_{d+1} = -\pi_{i=d+1}^{p}(1-m_i) < 0$.

Let $\hat{\boldsymbol{\beta}}$ denote the ordinary least squares estimator of $\boldsymbol{\beta}$ obtained by regressing $Y_{p,t}$ on $Y_{0,t-1}, Y_{1,t-1}, \ldots, Y_{p-1,t-1}$ . That is

$$\hat{\boldsymbol{\beta}} = (\Phi'\Phi)^{-1} \Phi' Y_{(p)} \tag{2.6}$$

where

$$\Phi_t = (Y_{0,t-1}, \ldots , Y_{p-1,t-1})' ,$$

$$\Phi = (\Phi_1, \Phi_2, \ldots , \Phi_n)' ,$$

and

$$Y_{(p)} = (Y_{p,1}, Y_{p,2}, \ldots , Y_{p,n})' .$$

Define the regression "F-statistic" for testing $\beta_1 = \beta_2 = \ldots = \beta_i = 0$ by

$$F_{i,n}(p) = (i\, \hat{\sigma}_n^2)^{-1}\, \hat{\boldsymbol{\beta}}'_{(i)}\, \mathbf{C}_{(i)}^{-1}\, \hat{\boldsymbol{\beta}}_{(i)} \tag{2.7}$$

where $\mathbf{C}_{(i)}$ is the (ixi) submatrix consisting of the first i rows and i columns of $(\Phi'\ \Phi)^{-1}$, $\hat{\boldsymbol{\beta}}_{(i)} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_i)'$; and

$$\hat{\sigma}_n^2 = (n-p)^{-1}\, \Sigma_{t=1}^{n} [Y_{p,t} - \Phi'_t\, \hat{\boldsymbol{\beta}}_{(p)}]^2 .$$

Define also the regression t-statistic for testing the hypothesis $\beta_i = 0$ by

$$t_{i,n}(p) = \left[\hat{\sigma}_n^2\, \hat{\sigma}_{p,n}^{ii}\right]^{-1/2} \hat{\beta}_i \tag{2.8}$$

where $\hat{\sigma}_{p,n}^{ii}$ is the (i,i)th element of $(\Phi'\Phi)^{-1}$ . Following Lemma 1, we use the "F-statistic" $F_{i,n}(p)$ to test the hypothesis $H_i$ . We now present a lemma that will be useful in obtaining the asymptotic distribution of $F_{i,n}(p)$ and $t_{i,n}(p)$. For the statistics considered in this paper, without loss of generality, we assume that $Y_{-p+1} = \ldots = Y_0 = 0$ and $\sigma^2 = 1$. We use the notation " $\xrightarrow{\mathcal{D}}$ " to indicate convergence in distribution and the notation " $\xrightarrow{P}$ " to indicate convergence in probability of a statistic as the sample size tends to infinity.

Lemma 2: Define,

$$S_{0,0} = 0$$

$$S_{0,t} = e_t$$

$$(1-B)S_{i,t} = S_{i-1,t}\ , \text{ for } i = 1, 2, \ldots, d,$$

where $e_t$ is a sequence of iid random variables with $E[|e_t|^{4+\delta}] < M < \infty$ for some $\delta > 0$. Then, as n tends to infinity,

$$n^{-i}\Sigma_{t=1}^{n} S_{i,t}\ S_{0,t+1} \xrightarrow{\mathcal{D}} \xi_i = \int_0^1 H_{i-1}(t)\ d\ H_0(t)$$

and

$$n^{-(i+j)}\Sigma_{t=1}^{n}\ S_{i,t}S_{j,t} \xrightarrow{\mathcal{D}} \sigma_{ij} = \int_0^1 H_{i-1}(t)\ H_{j-1}(t)\ dt$$

where $H_0(t)$ is a standard Brownian motion and

$$H_i(t) = \int_0^t H_{i-1}(s)\ ds\quad .$$

Proof: See Chan and Wei (1986) . []

Pantula (1985) obtained an alternate representation, involving quadratic forms, for the limiting distributions of Lemma 2. (See also Yajima (1985).) The representations are similar to those of the W variables presented in Hasza and Fuller (1979) and are useful in obtaining the percentiles of the limiting distribution.

We now present the joint limiting distributions of the "F-statistic" $F_{i,n}(p)$ and the regression t-statistic $t_{i,n}(p)$, under the hypothesis $H_d$ .

Theorem 1: Assume that $Y_t$ is defined by (1.1). Then, under the hypothesis $H_d$,

$$F_{i,n}(p) \xrightarrow{\mathcal{D}} \begin{cases} F_i(d) & \text{if } i \leq d \\ \infty & \text{if } i > d \end{cases} \tag{2.9}$$

$$t_{i,n}(p) \xrightarrow{\mathcal{D}} t_i(d) \qquad \text{if } i \leq d \tag{2.10}$$

and

$$n^{-1/2} t_{i,n}(p) \xrightarrow{P} f\, \beta_i \qquad \text{if } i > d \tag{2.11}$$

where f is a finite positive constant,

$$F_i(d) = i^{-1}[d\, F_d(d) - (d-i)\, F_{d-i}(d-i)],$$

$$F_k(k) = k^{-1}\, \xi_k\, \Sigma_k^{-1} \xi_k\ ,$$

$$t_i(d) = \bar{\mu}_{d-i+1,d} \quad \sigma_d^{d-i+1,d-i+1 \quad -1/2}\ ,$$

$$\bar{\mu}_{i,d} = \text{ith element of } \Sigma_d^{-1} \xi_d\ ,$$

$$\sigma_d^{ii} = (i,i)\text{th element of } \Sigma_d^{-1}\ ,$$

$$\xi_k = (\xi_1, \xi_2, \ldots, \xi_k)',$$

$$\Sigma_k = ((\sigma_{ij}))\ kxk\ ,\ k = 1,\ldots,d,$$

and $\xi_i$ and $\sigma_{ij}$ as defined in Lemma 2.

Proof: See Appendix A. []

Notice that the limiting distribution of $F_{i,n}(p)$ is independent of the order p. It depends only on i (the number of unit roots being tested) and d (the true number of unit roots present). Pantula (1985) used the quadratic form type representations to obtain the empirical percentiles of $F_i(d)$. Let

$F_i(d,\alpha)$ denote the $100(1-\alpha)\%$ percentile of the distribution of $F_i(d)$. Pantula (1985) obtained $F_i(d,\alpha)$ for $1 \le i \le d \le 5$ and $\alpha = 0.5, 0.2, 0.1, 0.05, 0.025$ and $0.01$. Extending the criterion suggested by Hasza and Fuller (1979) we may reject the hypothesis $H_d$ of exactly d unit roots (in favor of the hypothesis $H_{d-1}$ of exactly d-1 unit roots) if $F_{d,n}(p) > F_d(d,\alpha)$ . Pantula (1985) observed that the empirical percentiles $F_i(d,\alpha)$ increase in d. That is, $F_d(d,\alpha) < F_d(d+1,\alpha) < F_d(d+2,\alpha) < \ldots$ . Therefore, the probability of rejecting $H_d$ (in favor of $H_{d-1}$) increases with the number of unit roots ($\ge$ d) present. For example, if we use the criterion of rejecting the hypothesis $H_1$ that there is exactly one unit root (in favor of $H_0$: stationary process) whenever $F_{1,n}(p) > F_1(1,\alpha)$, then the higher the number of unit roots present, the more likely we are to conclude that the process is stationary. Intuitively, however, we hope that our criterion should strongly indicate nonstationarity when more than one unit root is present.

Suppose we have a prior knowledge that there are at most $s \le p$ unit roots present. (If no such prior knowledge is available then take $s = p$.) In practice $s \le 3$. We now suggest a sequential criterion to test the hypothesis $H_d$ for $d \le s$. The criterion is: "Reject the hypothesis $H_d$ (in favor of $H_{d-1}$) if $F_{i,n}(p) > F_i(i,\alpha)$ for $i = d, d+1, \ldots, s$ ." Since under the hypothesis $H_d$, $F_{i,n}(p)$ converges in probability to infinity for $i > d$ we get

$$\ell im_{n\to\infty} P_{H_j}[\text{Rejecting } H_d] = \begin{cases} \alpha & , j=d \\ \le \alpha & , j>d \\ 1 & , j<d \end{cases} .$$

That is, if we use our sequential procedure, asymptotically, the chance that we reject the hypothesis that there are exactly d unit roots is (i) $\alpha$, if there are exactly d unit roots, (ii) less than or equal to $\alpha$, if there are more than d unit roots and (iii) one, if there are less than d unit roots. Thus, for example, if we use the sequential procedure, the chance of concluding that the process is stationary is smaller than $\alpha$ when there is one or more unit roots present. In this sense, it is advantageous to use the F-statistics sequentially rather than individually.

Let us now interpret the results from the hypothesis testing problem in the context of estimating $0 \le d_0 \le s$, the unknown number of unit roots that are actually present. Suppose for a given $\alpha$ we use the estimator $\hat{d}_\alpha$ which takes the value d if $F_{d,n}(p) \le F_d(d,\alpha)$ and $F_{i,n}(p) > F_i(i,\alpha)$ for $d < i \le s$ and takes the value zero if $F_{i,n}(p) > F_i(i,\alpha)$ for $1 \le i \le s$. Then our results imply that, for $d_0 > 0$,

$$\lim_{n\to\infty} P_{H_{d_0}} [\hat{d}_\alpha = d] = \begin{cases} 1-\alpha & , \quad d = d_0 \\ \le \alpha & , \quad d < d_0 \\ 0 & , \quad d > d_0 \end{cases}$$

and for $d_0 = 0$, $P_{H_0}[\hat{d}_\alpha = 0]$ converges to one. In this sense, our procedure is asymptotically consistent. Pantula (1986) compared the power of the sequential procedure with that of the traditional methods and concluded that the loss in power is minimal even for samples of size 25.

Note that the asymptotic distribution of $t_{i,n}(p)$ also depends on the number of unit roots that are actually present. A natural question is to see whether the t-statistic $t_{i,n}(p)$ can be used in a sequential manner to test the relevant hypotheses. Consider for example an AR(3) process. As mentioned earlier, Sen (1985) observed that the Dickey and Fuller (1979) criterion based on $t_{1,n}(3)$ rejects the hypothesis $H_1$ of exactly one unit root in favor of the hypothesis $H_0$ of no unit roots more often under $H_2$ and $H_3$ than under $H_1$. So we cannot hope to obtain a consistent procedure by first testing $\beta_1 = 0$ (for a single unit root), then testing $\beta_2 = 0$ (for two unit roots) and then testing $\beta_3 = 0$. Also the sequential procedure that first tests $\beta_3 = 0$ using $t_{3,n}(3)$ and then tests $\beta_2 = 0$ using $t_{2,n}(3)$ will not be consistent. This is because it is possible to have $\beta_2 \geq 0$ even though the process is stationary. For example, if $m_1 = m_2 = m_3 = -0.5$, then the process is stationary with $\beta_2 = 0$ and the t-statistic $t_{2,n}(3)$ converges in distribution to N(0,1). Also, if $m_1 = m_2 = m_3 = -0.75$, then the process is stationary with $\beta_2 = 0.5(1.75)^2$ and the t-statistic $t_{2,n}(3)$ diverges to <u>positive</u> infinity.

Now consider the problem of testing the hypothesis $H_d$: $\beta_1 = \beta_2 = \ldots = \beta_d = 0$ against the alternative $H_{d-1}$: $\beta_1 = \beta_2 = \ldots = \beta_{d-1} = 0$, $\beta_d < 0$. Note that under the null ($H_d$) and the alternative ($H_{d-1}$) hypotheses we have $\beta_1 = \ldots = \beta_{d-1} = 0$ and the alternative hypothesis is one sided in nature ($\beta_d < 0$). This suggests that a more powerful (likelihood ratio type) test for testing the hypothesis $H_d$ against the alternative $H_{d-1}$ is obtained by regressing $Y_{p,t}$ on $Y_{d-1,t-1}, \ldots, Y_{p-1,t-1}$, (i.e., set $\beta_1 = \ldots = \beta_{d-1} = 0$ in (2.1)). Let $t^*_{d,n}(p)$ denote the regression t-statistic for testing the coefficient of $Y_{d-1,t-1}$ is zero in the regression of $Y_{p,t}$ on $Y_{d-1,t-1}, \ldots, Y_{p-1,t-1}$ . Note that $t^*_{1,n}(p) = t_{1,n}(p)$.

We now obtain the asymptotic distribution of the likelihood ratio type t-statistic $t^*_{i,n}(p)$ under the hypothesis $H_d$.

<u>Theorem 2</u>: Suppose the process $Y_t$ satisfies the conditions of Theorem 1. Then, under the hypothesis $H_d$,

$$t^*_{i,n}(p) \xrightarrow{\mathcal{D}} \begin{cases} \hat{\tau} & , \ i=d \\ t^*_i(d) & , \ i<d \\ -\infty & , \ i>d \end{cases} .$$

where

$$\hat{\tau} = t_1(1) = \xi_1 \sigma_{11}^{-1/2}$$

$$t_i^*(d) = \bar{\mu}_{d-i+1,d-i+1} \left[\sigma_{d-i+1}^{d-i+1,d-i+1}\right]^{-1/2}$$

and $\xi_1$, $\sigma_{11}$, $\bar{\mu}_{i,d}$ and $\sigma_d^{ii}$ are as defined in Theorem 1.

<u>Proof</u>: See Appendix A. []

Note that for any d, under the hypothesis $H_d$, the asymptotic distribution of the t-statistic $t_{d,n}^*(p)$ is the distribution of $\hat{\tau}$ obtained by Dickey and Fuller (1979) for the single unit root case. Therefore, we do not have to tabulate different percentiles for $t_d^*(d)$ as we have done for the F-statistic. We now suggest a sequential procedure based on the t-statistic $t_{d,n}^*(p)$. For a given $\alpha$, the second procedure is: "Reject $H_d$ in favor of $H_{d-1}$ if $t_{i,n}^*(p) < \hat{\tau}_\alpha$ for i=d, d+1, ..., s, where $\hat{\tau}_\alpha$ is the lower $\alpha$ percentile of the $\hat{\tau}$ distribution given in Table 8.5.2 of Fuller (1976)." Also, let $\tilde{d}_\alpha$ be an estimator of $d_0$ (the number of unit roots actually present) which takes the value d if $t_{d,n}^*(p) \geq \hat{\tau}_\alpha$ and $t_{i,n}^*(p) < \hat{\tau}_\alpha$, i = d+1, ..., s and takes the value zero when $t_{i,n}^*(p) < \hat{\tau}_\alpha$ for i = 1, ..., s.

From Theorem 2 it follows that

$$\ell im_{n\to\infty} P_{H_j} [\text{Rejecting } H_d] = \begin{cases} \alpha & , \ j=d \\ \leq \alpha & , \ j>d \\ 1 & , \ j<d \end{cases} .$$

Also, for $d_0 > 0$,

$$\ell im_{n\to\infty} P_{H_{d_0}} [\tilde{d}_\alpha = d] = \begin{cases} 1-\alpha & , \ d=d_0 \\ \leq \alpha & , \ d<d_0 \\ 0 & , \ d>d_0 \end{cases} .$$

and for $d_0 = 0$, $P_{H_0}[\tilde{d}_\alpha = 0]$ converges to one.

The sequential procedure based on the t-statistic $t^*_{d,n}(p)$ has several advantages over the sequential procedure based on the F-statistic $F_{d,n}(p)$. The sequential $t^*$-procedure requires only one set of tables given in Fuller (1976). Also, the $t^*$-procedure makes use of the one sided nature of the alternatives and hence is expected to be more powerful than the F-procedure. On the other hand, the F-procedure is a two-tailed test and hence is expected to be robust in the presence of mildly explosive roots. Of course, the $t^*$-procedure can be modified to a two-tailed procedure by considering $t^{*2}_{d,n}(p)$ and comparing it with the percentiles of $\hat{\tau}^2 = F_1(1)$. Using simulations, Dickey and Pantula (1987) compared the $t^*$-procedure and the F-procedure and recommend the use of the $t^*$-procedure. Dickey and Pantula (1987) also present a numerical example (Real Estate Loans) of a third order autoregressive process with two unit roots.

3. <u>Concluding Remarks</u>: We have given here a reparameterization of an autoregressive model that is convenient for testing the number of unit roots. We have derived the asymptotic distributions of the regression t-statistic $t_{i,n}(p)$ and the likelihood ratio type F-statistic, $F_{i,n}(p)$, under various hypotheses.

We have indicated why the t-statistic $t_{i,n}(p)$ may not be useful in testing or estimating the number of unit roots present. We have obtained a sequential procedure based on the F-statistic $F_{i,n}(p)$ for testing the hypothesis $H_d$ and estimating $d_0$ the actual number of unit roots present. We also obtained the asymptotic distribution of the likelihood ratio type t-statistic $t^*_{d,n}(p)$ for testing the hypothesis $H_d$ against the alternative hypothesis $H_{d-1}$. We developed a sequential procedure based on the $t^*$-statistics that is also asymptotically consistent. The $t^*$-procedure takes into account the one-sided nature of the alternatives and hence is superior to the F-procedure.

We have presented the results for an autoregressive process with mean zero. Our procedure can be extended to autoregressive processes where an intercept or a time trend is included. The percentiles for the F-statistic when an intercept is included in the regression are given in Pantula (1985). (See also Phillips and Perron (1986).) Also, the assumption of $e_t$ being a sequence of independent random variables can be weakened to $e_t$ being a sequence of martingale differences as in Chan and Wei (1986) and Phillips (1987). In Appendix B, following the approach of Said and Dickey (1984), the sequential procedures are extended to autoregressive moving average processes.

## Acknowledgements

I wish to express my thanks to Dennis Boos, David Dickey and the referees for their helpful comments. The research is partially supported by NSF Grant No. DMS-8610127.

## References

Altonji, J. and O. Ashenfelter (1980). Wage movements and the labor market equilibrium hypothesis. *Economica* **47**, 217-245.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. John Wiley and Sons: New York.

Box, G. E. P. and L. Pallesen (1978). Adaptive minimum mean square error forecast of missile trajectory using stochastic difference equation models. MRC Technical Summary Report #1821, University of Wisconsin, Madison, WI.

Chan, N. H. and C. Z. Wei (1986). Limiting distributions of least squares estimates of unstable autoregressive processes. Technical report, Department of Mathematics, University of Maryland, College Park, MD.

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 S. G. Pantula (1987). Determining the order of differencing in autoregressive processes. Forthcoming in the *Journal of Business and Economic Statistics.*

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

Fuller, W. A. (1979). Testing the autoregressive process for a unit root. Proceedings of the 42nd Session of the International Statistical Institute, Manilla, Spain.

Hannan, E. J. and C. C. Heyde (1972). On limit theorems for quadratic functions of discrete time series. *Annals of Mathematical Statistics*, **43**, 2058-2066.

Hasza. D. P. and W. A. Fuller (1979). Estimation for autoregressive processes with unit roots. *Annals of Statistics*, **7**, 1106-1120.

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

Mann, H. B. and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, **11**, 173-220.

Meese, R. A. and K. J. Singleton (1983). On unit roots and the empirical modeling of exchange rates. *International Economic Review*, **24**, 1029-1035.

Nelson, C. R. and C. I. Plosser (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, **10**, 139-162.

Pagano, M. (1973). When is an autoregressive scheme stationary? *Communications in Statistics*, **1**, 533-544.

Pantula, S. G. (1985). Estimation for autoregressive processes with several unit roots. Institute of Statistics Mimeograph Series No. 1665, North Carolina State University, Raleigh, NC.

Pantula, S. G. (1986). Determining the order of differencing in autoregressive processes. Technical report, Department of Statistics, North Carolina State University, Raleigh, NC.

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

Phillips, P. C. B. and P. Perron (1986). Testing for a unit root in time series regression. Cowles Foundation discussion paper no. 795, Yale University.

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

Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. Technical report, University of Rochester.

Searle, S. R. (1971). *Linear Models*. John Wiley and Sons: New York.

Sen, D. L. (1985). Robustness of single unit root test statistics in the presence of multiple unit roots. Ph.D. Thesis, North Carolina State University, Raleigh, NC.

Tiao, G. C. and R. S. Tsay (1983). Consistency properties of least squares estimates of autoregressive parameters in ARIMA models. *Annals of Statistics*, **11**, 856-871.

Yajima, Y. (1985). Estimation of the degree of differencing of an ARIMA process. *Annals of the Institute of Statistical Mathematics* (Part A), **37**, 387-408.

## Appendix A

We prove Theorem 1 in three steps: (i) we consider a dth order process with d unit roots and obtain the limiting representation $F_i(d)$ of the "F-statistic" $F_{i,n}(d)$ for i=1,2,..., d; (ii) for a pth order process with exactly d unit roots and remaining p-d roots inside the unit circle we will show that $F_{i,n}(p)$ converges in distribution to $F_i(d)$ for i = 1,2,..., d, and (iii) diverges to infinity for i > d . The results for the t-statistic $t_{i,n}(p)$ follow in the process. We will use the following fact from regression theory repeatedly: The F (or t) - statistic for testing a linear hypothesis involving the parameters of a linear model is identical to the F (or t) - statistic of an equivalent hypothesis under a (linearly) reparameterized model.

Consider a dth order autoregressive process with d unit roots given by

$$(1-B)^d W_t = e_t \quad , \tag{A.1}$$

where $\{e_t\}$ is a sequence of iid random variables given in Lemma 2. Notice that $W_t = S_{d,t}$ where $S_{d,t}$ is defined in Lemma 2. In the parameterization of (2.1) we can write $W_t$ as

$$W_{d,t} = \Sigma_{i=1}^{d} \beta_i W_{i-1,t-1} + e_t \tag{A.2}$$

where

$$W_{i,t} = (1-B)^i W_t = S_{d-i,t} \ ,$$

and

$$\beta_1 = \beta_2 = \dots = \beta_d = 0 \ .$$

Note that we can rewrite (A.2) in terms of $S_{i,t}$ as

$$S_{0,t} = \Sigma_{i=1}^{d} \mu_i S_{i,t-1} + e_t \tag{A.3}$$

where $\mu_i = \beta_{d-i+1}$ .

Therefore, the hypothesis $\beta_1 = \ldots = \beta_i = 0$ is the same as $\mu_{d-i+1} = \ldots = \mu_d = 0$. Consider the regression of $S_{0,t}$ on $S_{1,t-1}, S_{2,t-1}, \ldots, S_{d,t-1}$ . Then, the regression F-statistic $F_{i,n}(d)$ for testing exactly $i (\leq d)$ unit roots in a dth order process is given by

$$F_{i,n}(d) = [i\, \hat{\sigma}_n^2(d)]^{-1} [R_n(\mu_1, \ldots, \mu_d) - R_n(\mu_1, \ldots, \mu_{d-i})] \tag{A.4}$$

where $R_n(\mu_1, \ldots, \mu_i)$ denotes the regression sums of squares for the regression of $S_{0,t}$ on $S_{1,t-1}, \ldots, S_{i,t-1}$ and $\hat{\sigma}_n^2(d)$ denotes the mean square error in the regression of $S_{0,t}$ on $S_{1,t-1}, \ldots, S_{d,t-1}$. Let $t_{d-i+1,n}(d)$ denote the regression t-statistic for testing $\mu_i = 0$. Note that, for $k = 1, \ldots, d$,

$$\hat{\boldsymbol{\mu}}_d = G_{d,n}^{-1} \mathbf{g}_{d,n} = (\hat{\mu}_1, \ldots, \hat{\mu}_d)' ,$$

$$t_{d-i+1,n}(d) = [\hat{\sigma}_n^2(d)\, G_{d,n}^{ii}]^{-1/2} \hat{\mu}_i ,$$

$$R_n(\mu_1, \ldots, \mu_k) = \mathbf{g}_{k,n}' G_{k,n}^{-1} \mathbf{g}_{k,n} ,$$

where $G_{k,n} = ((G_{ij,k,n}))$ is a kxk matrix; $\mathbf{g}_{k,n} = (g_{1,k,n}, \ldots, g_{k,k,n})'$; $G_{ij,k,n} = \Sigma_{t=1}^{n} S_{i,t-1} S_{j,t-1}$; $g_{i,k,n} = \Sigma_{t=1}^{n} S_{i,t-1} S_{0,t}$ and $G_{d,n}^{ii}$ is the (i,i)th element of $G_{d,n}^{-1}$.

We now obtain the limiting distributions of $F_{i,n}(d)$ and $t_{i,n}(d)$.

<u>Lemma A.1</u>: Consider the model (A.3) where $e_t$ satisfy the conditions of Lemma 2.

Then,

(i) $R_n(\mu_1, \ldots, \mu_k) \xrightarrow{\mathcal{D}} \xi_k' \Sigma_k^{-1} \xi_k = R_k$ , $k=1,\ldots,d$

(ii) $\hat{\sigma}_n^2(d) \xrightarrow{P} 1$

(iii) $F_{i,n}(d) \xrightarrow{\mathcal{D}} F_i(d) = i^{-1}[R_d - R_{d-i}]$ , $i \le d$

and

(iv) $t_{d-i+1,n}(d) \xrightarrow{\mathcal{D}} t_{d-i+1}(d)$ , $i \le d$ .

where $F_{i,n}(d)$ is defined in (A.4), $\xi_k$, $\Sigma_k$, and $t_{d-i+1}(d)$ are as defined in Theorem 1 and $R_0 = 0$.

Proof: From Lemma 2, we know that for $k \le d$,

$$D_{k,n}^{-1}\, g_{k,n} \xrightarrow{\mathcal{D}} \xi_k \tag{A.5}$$

and

$$D_{k,n}^{-1}\, G_{k,n}\, D_{k,n}^{-1} \xrightarrow{\mathcal{D}} \Sigma_k$$

where $D_{k,n} = \text{diagonal}\ \{n, n^2, \ldots, n^k\}$ .

Pantula (1985) and Chan and Wei (1986) showed that $\Sigma_k$ is nonsingular. Therefore,

$$R_n(\mu_1,\ldots,\mu_k) \xrightarrow{\mathcal{D}} \xi_k' \Sigma_k^{-1} \xi_k \quad . \tag{A.6}$$

Now,

$$\hat{\sigma}_n^2(d) = (n-d)^{-1} \quad \Sigma_{t=1}^n\ S_{0,t}^2 - R_n(\mu_1,\ldots,\mu_d)$$

$$= (n-d)^{-1}\, \Sigma_{t=1}^n\ e_t^2 + 0_p((n-d)^{-1})$$

$$\xrightarrow{P} 1 \quad .$$

The result (iii) follows immediately from (i) and (ii). The result (iv) follows from (A.5), (A.6) and (ii). []

Note that, from Lemma A.1,

$$F_d(d) = d^{-1} \xi_d' \Sigma_d^{-1} \xi_d$$

and hence for $i \leq d$,

$$F_i(d) = i^{-1}[d\, F_d(d) - (d-i)F_{d-i}(d-i)] \quad .$$

Now we will obtain the asymptotic distribution of $F_{i,n}(p)$, the F-statistic for testing i unit roots in a pth order autoregressive process, under the hypothesis $H_d$.

Consider $Y_t$ a pth order autoregressive process with d unit roots and p-d stationary roots. Then, $Y_t$ can be written as

$$Y_{p,t} = \Sigma_{i=1}^{d} \beta_i Y_{i-1,t-1} + \Sigma_{i=d+1}^{p} \beta_i Y_{i-1,t-1} + e_t \tag{A.7}$$

where $Y_{i,t} = (1-B)^i\ Y_t$, $\beta_1 = \ldots = \beta_d = 0$ and $\beta_{d+1} < 0$ .

Let

$$W_t = Y_{p-d,t} - \Sigma_{i=d+1}^{p} \beta_i Y_{i-1-d,t-1} \tag{A.8}$$

and

$$Z_t = (1-B)^d Y_t = Y_{d,t} \ . \tag{A.9}$$

Then $W_t$ is a dth order autoregressive process with d unit roots and $Z_t$ is a (p-d)th order autoregressive process with (p-d) stationary roots. Following the arguments used in Lemma 5.1 of Hasza and Fuller (1979) and Lemmas 2.4 and 2.5 of Tiao and Tsay (1983) it can be shown that, for $0 \leq i,\ j \leq d-1$,

$$n^{-(2d-i-j)}\ \Sigma_{t=1}^{n} Y_{i,t-1} Y_{j,t-1} = \beta_{d+1}^{-2}\ n^{-(2d-i-j)}\ \Sigma_{t=1}^{n} W_{i,t-1} W_{j,t-1} + O_p(n^{-1}) \tag{A.10}$$

and

$$n^{-(d-i)}\Sigma_{t=1}^{n} Y_{i,t-1} Y_{p,t} = -\beta_{d+1}^{-1} n^{-(d-i)} \Sigma_{t=1}^{n} W_{i,t-1} e_t + O_p(n^{-1}) , \qquad (A.11)$$

where $W_{i,t} = (1-B)^i W_t$ .

(See also Lemma A.1 and Proposition A.1 of Yajima (1985).)

We now obtain the limiting distributions of $F_{i,n}(p)$ and $t_{i,n}(p)$ under the hypothesis $H_d$.

<u>Lemma A.2</u>: Let $Y_t$ be a pth order process with d unit roots and (p-d) stationary roots. Let $F_{i,n}(p)$ be the regression F-statistic for testing $\beta_1 = \ldots = \beta_i = 0$ in the model (A.7). Then, as n tends to infinity,

(i) $F_{i,n}(p) \xrightarrow{\mathcal{D}} F_i(d)$ for $i \leq d$ ,

(ii) $F_{d+1,n}(p)$ diverges to infinity,

(iii) $t_{i,n}(p) \xrightarrow{\mathcal{D}} t_i(d)$ for $i \leq d$ ,

and

(iv) $n^{-1/2} t_{i,n}(p) \xrightarrow{P} f \beta_i$ for $i > d$ ,

where $F_i(d)$ and $t_i(d)$ are given in Lemma A.1 and and f is a finite positive constant.

<u>Proof</u>: Let

$$\Phi_t = (Y_{0,t-1}, \ldots, Y_{p-1,t-1})'$$

$$= (\Psi_t' , \nu_t')' ,$$

$$\Psi_t = (Y_{0,t-1}, \ldots, Y_{d-1,t-1})' ,$$

and

$$\nu_t = (Y_{d,t-1}, \ldots, Y_{p-1,t-1})' .$$

Define, $M_n$ = diagonal $\{n^d, n^{d-1}, \ldots, n, n^{1/2}, \ldots, n^{1/2}\}$ .

Then, from the arguments used in Theorem 5.1 of Hasza and Fuller (1979), and (A.10) we get

$$M_n \left[ \Sigma_{t=1}^{n} \Phi_t \Phi_t' \right]^{-1} M_n \xrightarrow{\mathcal{D}} \begin{bmatrix} \beta_{d+1}^2 A \Sigma_{d.}^{-1} A & 0 \\ 0 & T^{-1} \Gamma_z^{-1} T^{-1\prime} \end{bmatrix} \quad \text{(A.12)}$$

where T is defined in (2.2) with (p-d) in place of p; $\Sigma_d$ defined in Lemma 2; $\Gamma_z$ is a (p-d) x (p-d) matrix with $\gamma_z(i-j)$ as the (i,j)th element, $\gamma_z(h) = \lim_{t\to\infty} E[Z_t Z_{t-h}]$; and

$$A = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \\ \vdots & & & & \\ 1 & 0 & \ldots & 0 & 0 \end{bmatrix} : d \times d .$$

Also, from Lemma 2 and (A.11) it follows that

$$M_n^{-1} \Sigma_{t=1}^{n} \Phi_t Y_{p,t} \xrightarrow{\mathcal{D}} \begin{bmatrix} -\beta_{d+1}^{-1} A \xi_d \\ T'b - T'\Gamma_z c \end{bmatrix} \quad \text{(A.13)}$$

where $\xi_d$ is defined in Lemma 2, $b = (\gamma_z(1), \ldots, \gamma_z(p-d))'$ and c is as defined in (2.2) with p-d in place of p.

Now consider $F_{i,n}(p)$, the regression F-statistic for testing $\beta_1 = \ldots = \beta_i = 0$ in (A.7) . Note that,

$$F_{i,n}(p) = [i\, \hat{\sigma}_n^2(p)]^{-1} [R_n(\beta_1, \ldots, \beta_p) - R_n(\beta_{i+1}, \ldots, \beta_p)]$$

where $\hat{\sigma}_n^2(p)$ is the residual mean square of the regression (A.7), $R_n(\beta_{i+1},\ldots,\beta_p)$ denote the regression sums of squares of the regression (A.7) with $\beta_1$, $\beta_2$, ..., $\beta_i$ set equal to zero. Since the expression (A.12) is block diagonal, it follows from (A.13) and Lemma A.1, for $i \le d$,

$$F_{i,n}(p) \xrightarrow{\mathcal{D}} F_i(d) \quad .$$

and

$$t_{i,n}(p) \xrightarrow{\mathcal{D}} t_i(d) \quad .$$

Now,

$$n^{-1}F_{d+1,n}(p) = [n(d+1)\hat{\sigma}_n^2(p)]^{-1}[R_n(\beta_1,\ldots,\beta_p) - R_n(\beta_{d+2},\ldots,\beta_p)]$$

$$\ge [n(d+1)\hat{\sigma}_n^2(p)]^{-1}[R_n(\beta_1,\ldots,\beta_p) - R_n(\beta_1,\ldots,\beta_d,\beta_{d+2},\ldots,\beta_p)]$$

$$= [(d+1)\hat{\sigma}_n^2(p)\ n\ \hat{\sigma}_{p,n}^{d+1,d+1}]^{-1}\ \hat{\beta}_{d+1}^2$$

where $\hat{\sigma}_{p,n}^{d+1,d+1}$ is the (d+1,d+1)th element of $\left[\Sigma_{t=1}^{n} \Phi_t\Phi_t'\right]^{-1}$. Note that,

$$n\ \hat{\sigma}_{p,n}^{d+1,d+1} = \epsilon_{d+1}'\ M_n[\Sigma_{t=1}^{n} \Phi_t\Phi_t']^{-1}\ M_n\ \epsilon_{d+1}$$

$$\xrightarrow{P} \text{ the first element of } T^{-1}\ \Gamma_z^{-1}\ T^{-1\prime}$$

and

$$\hat{\beta}_{d+1}^2 \xrightarrow{P} \beta_{d+1}^2 \quad ,$$

where $\epsilon_{d+1}$ is the (d+1)th column of a pxp identity matrix. Therefore, $F_{d+1,n}(p)$ diverges to infinity as n tends to infinity.

Also, note that for $i > d$,

$$t_{i,n}(p) = \left[\hat{\sigma}_n^2(p)\ \hat{\sigma}_{p,n}^{ii}\right]^{-1/2} \left[\hat{\beta}_i - \beta_i\right] + \left[\hat{\sigma}_n^2(p)\ \hat{\sigma}_{p,n}^{ii}\right]^{-1/2}\beta_i \qquad \text{(A.14)}$$

and hence from (A.12) and (A.13)

$$n^{-1/2}\, t_{i,n}(p) \xrightarrow{P} f\,\beta_i \ ,$$

where f is a finite positive constant. []

We now complete the proof of Theorem 1.

<u>Proof of Theorem 1</u>: From Lemma A.1 and Lemma A.2, we know that for $i \leq d$ ,

$$F_{i,n}(p) \xrightarrow{\mathcal{D}} F_i(d)$$

under the hypothesis $H_d$. From Lemma A.2, it follows that $F_{d+1,n}(p)$ diverges to infinity under the hypothesis $H_d$. Now, since

$$F_{d+j,n}(p) \geq (d+j)^{-1}(d+1)F_{d+1,n}(p)$$

for $j \geq 2$, we have that $F_{i,n}$ diverges to infinity for $i > d$, under the hypothesis $H_d$ . []

We now prove Theorem 2.

<u>Proof of Theorem 2</u>: Note that $t^*_{i,n}(p)$ is the regression t-statistic for testing the hypothesis that the coefficient of $V_{t-1}$ is zero in the regression of $(1-B)^{p-i+1}V_t$ on $V_{t-1}$, $(1-B)V_{t-1},\ldots,(1-B)^{p-i}V_{t-1}$, where $V_t = (1-B)^{i-1}Y_t$. Under the hypothesis $H_d$, for $i \leq d$, the process $V_t$ has (d-i+1) unit roots and (p-d) stationary roots. Therefore, from the results of Dickey and Fuller (1979) and the arguments of Lemma A.2 it follows that, under the hypothesis $H_d$,

$$t^*_{i,n}\ (p) \xrightarrow{\mathcal{D}} \begin{cases} \hat{\tau} = t_1(1) & \text{for } i = d \\ t^*_i(d) & \text{for } i < d\ , \end{cases}$$

where $t^*_i(d)$ and $\hat{\tau}$ are as defined in Theorem 2. Also, as in (A.14) it follows that under the hypothesis $H_d$, for $i > d$

$$n^{-1/2}\, t^*_{i,n}(p) \xrightarrow{P} f^* \beta^*_{i,p}$$

where $f^*$ is a finite positive constant, $\beta^*_{i,p}$ is the probability limit of $\hat{\beta}^*_{i,p}$ and $\hat{\beta}^*_{i,p}$ is the regression coefficient of $Y_{i-1,t-1} = (1-B)^{i-1} Y_{t-1}$ in the regression of $(1-B)^p Y_t$ on $(1-B)^{i-1} Y_{t-1}, \ldots, (1-B)^{p-1} Y_{t-1}$ . We will show that under the hypothesis $H_d$, $\beta^*_{i,p}$ is a finite negative constant.

Note that for $i > d$, under the hypothesis $H_d$, the process $V_t = (1-B)^{i-1} Y_t$ is essentially a stationary process. It is easy to see that $\hat{\beta}^*_{i,p}$ is also the regression coefficient of $V_{t-1}$ in the regression of $(1-B)V_t$ on $V_{t-1}$, $(1-B)V_{t-1}, \ldots, (1-B)V_{t-\ell+1}$ where $\ell = p-i+1$ . Since the dth difference of $Y_t$ is essentially a stationary (p-d)th autoregressive process and that $V_t$ is an (i-1)th difference of $Y_t$, it follows that

$$\gamma_V(h) = \lim_{t\to\infty} E[V_t V_{t-h}]$$

is a positive definite function. Define

$$\Gamma_V = \text{toeplitz}(\gamma_V(0), \ldots, \gamma_V(\ell-1)),$$

$$\gamma_0 = (\gamma_V(1), \ldots, \gamma_V(\ell))$$

and

$$\delta = \Gamma_V^{-1} \gamma_0 \; . \qquad \text{(A.15)}$$

Now, using the standard arguments it can be shown that, under the hypothesis $H_d$, the regression coefficients $\hat{\delta}$ of the regression $V_t$ on $V_{t-1}, \ldots, V_{t-\ell}$ converge in probability to $\delta$. Also, since

$$\hat{\beta}^*_{i,p} = \Sigma_{i=1}^{\ell} \hat{\delta}_i - 1$$

we get

$$\hat{\beta}^*_{i,p} \xrightarrow{P} \beta^*_{i,p} = \Sigma^{\ell}_{i=1}\delta_i - 1 .$$

(See Fuller (1976), p. 374). Now, since $\delta$ is a solution to the Yule-Walker equations in (A.15), it follows from Theorem 2.2 of Pagano (1973) that the roots of the characteristic equation

$$m^{\ell} - \Sigma^{\ell}_{i=1} \delta_i m^{\ell-i} = 0 \tag{A.16}$$

lie inside the unit circle. Therefore,

$$\beta^*_{i,p} = -(1 - \Sigma^{\ell}_{i=1} \delta_i)$$

$$= -\pi^{\ell}_{i=1}(1-m_i)$$

is negative, where $m_1,\ldots,m_\ell$ are the roots of the characteristic equation (A.16). []

## Appendix B

In this appendix we extend the testing procedures to autoregressive moving average processes. Consider the model

$$Y_t = \Sigma^p_{i=1} \alpha_i Y_{t-i} + \Sigma^q_{j=1}\kappa_j e_{t-j} + e_t \tag{B.1}$$

where we assume that

(i) $\{e_t\}$ is a sequence of iid $(0,\sigma^2)$ random variables with finite $(4+\delta)$th moment for some $\delta > 0$,

(ii) the characteristic equations

$$m^p - \Sigma_{i=1}^{p}\alpha_i m^{p-i} = 0 \ , \tag{B.2}$$

and

$$m^q + \Sigma_{j=1}^{q}\kappa_j m^{q-j} = 0 \ , \tag{B.3}$$

have no common roots,

and

(iii) the roots of (B.3) lie inside the unit circle.

Let $1 \geq |m_1| \geq \ldots \geq |m_p|$ denote the roots of (B.2). We wish to test the hypothesis $H_d$: $m_1 = \ldots = m_d = 1$ and $|m_{d+1}| < 1$. Consider the reparameterization of (B.1)

$$(1-B)^p \ Y_t = \Sigma_{i=1}^{p}\beta_i(1-B)^{i-1}Y_{t-i} + \Sigma_{j=1}^{q}\kappa_j \ e_{t-1} + e_t \ . \tag{B.4}$$

From Lemma 1 we know that, under the hypothesis $H_d$, $\beta_1 = \ldots = \beta_d = 0$ (and $\beta_{d+1} < 0$). Consider the regression models

$$(1-B)^p \ Y_t = \Sigma_{\ell=1}^{p} \ a_\ell(1-B)^{\ell-1}Y_{t-1} + \Sigma_{j=1}^{k-p} \ b_j(1-B)^p \ Y_{t-j} + \text{error} \ , \tag{B.5}$$

and, for $i = 1, \ldots, p$,

$$(1-B)^pY_t = \Sigma_{\ell=i}^{p}a_\ell^*(1-B)^{\ell-1}Y_{t-1} + \Sigma_{j=1}^{k-p} \ b_j^*(1-B)^pY_{t-j} + \text{error} \ , \tag{B.6}$$

where $k$ ($>p$) will be chosen to be a function of the sample size $n$. Let $F_{i,n}(p,k)$ denote the usual regression F-statistic for testing $a_1 = \ldots = a_i = 0$ in (B.5) and let $t_{i,n}^*(p,k)$ denote the regression t-statistic for testing $a_i^* = 0$ in (B.6). We now present the asymptotic distributions of $F_{i,n}(p,k)$ and $t_{i,n}^*(p,k)$ under the hypothesis $H_d$ for $i \geq d$. For the remainder of the paper, we <u>assume</u> that the choice of $k$ is such that $k^{-1}$ and $n^{-1}k^3$ tend to zero as $n$ tends to infinity.

Result B.1: Suppose the hypothesis $H_d$ is true. Then,

(i) $F_{d,n}(p,k) \xrightarrow{\mathcal{D}} F_d(d)$ ,

(ii) $t^*_{d,n}(p,k) \xrightarrow{\mathcal{D}} \hat{\tau}$ ,

(iii) for $i > d$, $t^*_{i,n}(p,k)$ diverges to negative infinity, and

(iv) $F_{i,n}(p,k)$ diverges to infinity for $i > d$,

where $F_d(d)$ and $\hat{\tau}$ are as defined in Theorems 1 and 2.

Based on Result B.1, we suggest two sequential procedures for testing the hypothesis $H_d$. Let $s \le p$ denote our prior belief of the maximum number of unit roots present in the process.

Procedure 1: Consider the regression (B.5) and compute $F_{i,n}(p,k)$ where k is such that $k^{-1}$ and $n^{-1}k^3$ tend to zero as n tends to infinity. Reject the hypothesis $H_d$ (in favor of $H_{d-1}$) if $F_{i,n}(p,k) > F_i(i,\alpha)$, for $i=d,d+1,\ldots,s$, where $F_i(i,\alpha)$ is the $100(1-\alpha)\%$ percentile of the distribution of $F_i(i)$. (See Theorem 1.)

Procedure 2: Consider the regressions (B.6) to compute $t^*_{i,n}(p,k)$, $i=1,\ldots,s$, where k is such that $k^{-1}$ and $n^{-1}k^3$ tend to zero as n tends to infinity. Reject the hypothesis $H_d$ (in favor of $H_{d-1}$) if $t^*_{i,n}(p,k) < \hat{\tau}_\alpha$, for $i=d, d+1, \ldots$, s, where $\hat{\tau}_\alpha$ are the percentiles of $\hat{\tau}$ given in Table 8.5.2 of Fuller (1976).

For either procedure, we get,

$$\lim_{n\to\infty} P_{H_j}[\text{Rejecting } H_d] = \begin{cases} \alpha & , j=d \\ \le\alpha & , j>d \\ 1 & , j<d \end{cases} .$$

We will now outline a proof of Result B.1. Assume that the hypothesis $H_d$ of exactly d unit roots is true.

1. Note that $Z_t = (1-B)^d Y_t$ is essentially a stationary and invertible ARMA (p-d,q) process.

2. Consider the regressions

$$Z_t = \Sigma_{i=1}^{d} a_i (1-B)^{i-1} Y_{t-1} + \Sigma_{j=1}^{k-d} \eta_j Z_{t-j} + \text{error} , \qquad (B.7)$$

$$(1-B)Z_t = \Sigma_{i=1}^{d+1} a_i (1-B)^{i-1} Y_{t-1} + \Sigma_{j=1}^{k-d-1} \eta_j^{*} (1-B)Z_{t-j} + \text{error} , \qquad (B.8)$$

$$Z_t = a_d^{*} (1-B)^{d-1} Y_{t-1} + \Sigma_{j=1}^{k-d} \eta_j^{**} Z_{t-j} + \text{error} , \qquad (B.9)$$

and

$$(1-B)^{i-d} Z_t = a_i^{*} (1-B)^{i-d-1} Z_{t-1} + \Sigma_{j=1}^{k-i} \nu_j (1-B)^{i-d} Z_{t-j} + \text{error}. \qquad (B.10)$$

Note that (B.7) and (B.8) are reparameterizations of (B.5) and (B.9) and (B.10) are reparameterizations of (B.6). Therefore, $F_{d,n}(p,k)$ is the same as the regression F-statistic for testing $a_1 = \ldots = a_d = 0$ in (B.7); $F_{d+1,n}(p,k)$ is the regression F-statistic for testing $a_1 = \ldots = a_{d+1} = 0$ in (B.8); $t_{d,n}(p,k)$ is the same as the t-statistic for testing $a_d^{*} = 0$ in (B.9); and $t_{i,n}^{*}(p,k)$ is the same as the t-statistic for testing $a_i^{*} = 0$ in (B.10). Also, $t_{i,n}^{*}(p,k)$ is the regression on t-statistic for testing $\Sigma_{j=1}^{k-i+1} \delta_j = 1$ in the regression

$$(1-B)^{i-d-1} Z_t = \Sigma_{j=1}^{k-i+1} \delta_j (1-B)^{i-d-1} Z_{t-j} + \text{error} . \qquad (B.11)$$

3. Said and Dickey (1984) showed that $t_{d,n}^{*}(p,k)$ converges in distribution to $\hat{\tau}$ as n tends to infinity. (See Theorem 6.1 and Section 7 of Said and Dickey (1984).) Also, using the arguments of Said and Dickey (1984) and the arguments used in the proof of Lemma A.2, it can be shown that $F_{d,n}(p,k)$ converges in distribution to $F_d(d)$ as n tends to infinity. The main difference for the general ARMA case is that in the expressions (A.10)

and (A.11),

$$\beta^*_{d+1} = - [\pi^q_{j=1}(1-m^*_j)]^{-1} \pi^p_{i=d+1}(1-m_i) \qquad (B.12)$$

is used in place of $\beta_{d+1}$, where $1 > |m^*_1| \geq \ldots \geq |m^*_q|$ are the roots of the characteristic equation (B.3). (See also Yajima (1985).)

4. Now, to show that $t^*_{d+1,n}(p,k)$ diverges to negative infinity, consider the regression (B.11) with i=d+1, i.e.,

$$Z_t = \Sigma^{k-d+1}_{j=1} \delta_j Z_{t-j} + \text{error} . \qquad (B.13)$$

Note that $t^*_{d+1,n}(p,k)$ is the regression t-statistic for testing the hypothesis $\Sigma^{k-d+1}_{j=1}\delta_j = 1$ in (B.13). Therefore,

$$k^{1/2}(n-k)^{-1/2}t^*_{d+1,n}(p,k) = [\hat{\sigma}^2_n(p,k)k^{-1}\mathbf{1}'\hat{R}^{-1}_Z\mathbf{1}]^{-1/2}(\mathbf{1}'\hat{\boldsymbol{\delta}}_Z - 1), \qquad (B.14)$$

where $\hat{\boldsymbol{\delta}}_Z = \hat{R}^{-1}_Z\hat{r}_Z$ , $\hat{R}_Z = (n-k)^{-1}\Sigma_t[X_t(k)X_t'(k)]$ , $\mathbf{1} = (1,\ldots,1)'$, $\hat{r}_Z = (n-k)^{-1} \Sigma_t[X_t(k)Z_t]$, $\hat{\sigma}^2_n(p,k)$ is the residual mean square error and $X_t(k) = (Z_{t-1},\ldots,Z_{t-k+d-1})'$. Since $|m^*_1| < 1$ and $|m_{d+1}| < 1$, from Theorem 1 of Berk (1974) it follows that, as n tends to infinity,

$$\hat{\sigma}^2_n(p,k) \xrightarrow{P} \sigma^2$$

$$\mathbf{1}'\hat{\boldsymbol{\delta}}_Z - 1 = \mathbf{1}'(\hat{\boldsymbol{\delta}}_Z - \boldsymbol{\delta}) + \mathbf{1}'\boldsymbol{\delta} - 1$$

$$\xrightarrow{P} \beta^*_{d+1}$$

and $k^{-1}\ \mathbf{1}'\hat{R}^{-1}_Z\mathbf{1}$ is bounded and bounded away from zero, where $\beta^*_{d+1}$ is defined in (B.12). (In Berk's notation, $\boldsymbol{\delta} = -\mathbf{a}(k)$ and $1 + \Sigma^{\infty}_{i=1}a_i = -\beta^*_{d+1}$ is positive.) Therefore, $t^*_{d+1,n}(p,k)$ diverges to negative infinity.

5. We will now show that $F_{d+1,n}(p,k)$ diverges to infinity and hence $F_{d+j,n}(p,k)$ ($\geq (d+j)^{-1}(d+1)\, F_{d+1,n}(p,k)$) diverges to infinity as n tends to infinity. Consider the regressions (B.8) and (B.10) with i = d+1. Then, in the notation of regression sums of squares introduced in (A.4),

$$F_{d+1,n}(p,k) = [(d+1)(n-2k-d)^{-1}(M_1 - M_2)]^{-1}(M_2 - M_4)$$

and

$$t^{*2}_{d+1,n}(p,k) = [(n-2k)^{-1}(M_1 - M_3)]^{-1}(M_3 - M_4)$$

where $M_1 > M_2 \geq M_3 \geq M_4$, $M_1 = \Sigma_t(Z_t - Z_{t-1})^2$ ,
$M_2 = R_n(a_1,\ldots,a_{d+1},\eta^*_1,\ldots,\eta^*_{k-d-1})$,
$M_3 = R_n(a^*_{d+1},\nu_1,\ldots,\nu_{k-d-1}) = R_n(a_{d+1},\eta^*_1,\ldots,\eta^*_{k-d-1})$, and
$M_4 = R_n(\nu_1,\ldots,\nu_{k-d-1}) = R_n(\eta^*_1,\ldots,\eta^*_{k-d-1})$.
Note that, since $(M_1 - M_4)\,(M_2 - M_3) \geq 0$ ,

$$(M_1 - M_2)^{-1}(M_2 - M_4) \geq (M_1 - M_3)^{-1}\,(M_3 - M_4)\ .$$

Therefore,

$$F_{d+1,n}(p,k) \geq [(d+1)(n-2k)]^{-1}(n-2k-d)t^{*2}_{d+1,n}(p,k)$$

and hence $F_{d+1,n}(p,k)$ diverges to infinity.

6. We will now show that $t^*_{i,n}(p,k)$ diverges to negative infinity for $i > d+1$. Recall that $t^*_{i,n}(p,k)$ is the regression t-statistic for testing $\Sigma_{j=1}^{k-i+1}\delta_j = 1$ in the regression

$$V_t = \Sigma_{j=1}^{k-i+1}\,\delta_j V_{t-j} + \text{error}\ , \tag{B.15}$$

where $V_t = (1-B)^{i-d-1}Z_t$ , i.e.,

$$(n-k)^{-1/2}t^*_{i,n}(p,k) = [\hat{\sigma}^2_n(p,k)\ 1'\hat{R}_V^{-1}1]^{-1/2}(1'\hat{\delta}_V - 1)\ ,$$

where $\hat{\delta}_V = \hat{R}_V^{-1}\hat{r}_V$, $\hat{R}_V = (n-k)^{-1} \sum_t[X_t(k)X_t'(k)]$ ,
$\hat{r}_V = (n-k)^{-1} \sum_t[X_t(k)V_t]$, $\hat{\sigma}_n^2(p,k)$ is the residual mean square error of the regression (B.15) and $X_t(k) = (V_{t-1},\dots,V_{t-k+i-1})'$. Since $V_t$ is the (i-d-1)th difference of $Z_t$, we can write

$$\hat{R}_V = C_{i-d-1}C_{i-d} \cdots C_1\hat{R}_Z C_1' \cdots C_{i-d-1}' , \tag{B.16}$$

where $\hat{R}_Z = (n-k)^{-1}\sum_t[X_t^*(k)X_t^*(k)']$ , $X_t^*(k) = (Z_{t-1},\dots,Z_{t-k+d})'$, and

$$C_j = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \quad : (k-d-j) \times (k-d-j+1) .$$

Therefore, from Theorem 6.5.5 of Anderson (1971), we get

$$\lambda_{min}(\hat{R}_V) \geq \lambda_{min}(\hat{R}_Z)2^{i-d-1} \pi_{j=1}^{i-d-1} [1 - \text{Cos } \pi(k-d-j+1)^{-1}]$$

and

$$\lambda_{max}(\hat{R}_V) \leq \lambda_{max}(\hat{R}_Z)2^{i-d-1} \pi_{j=1}^{i-d-1} [1 + \text{Cos } \pi(k-d-j+1)^{-1}] .$$

Since $[1 - \text{Cos } \pi(k+1)^{-1}]^{-1} = O(k^2)$, we get

$$\lambda_{max}(\hat{R}_V^{-1}) \leq M_1\lambda_{max}(\hat{R}_Z^{-1})\ k^{2(i-d-1)}$$

and

$$\lambda_{min}(\hat{R}_V^{-1}) \geq M_2\lambda_{min}(\hat{R}_Z^{-1}) ,$$

where $M_1$ and $M_2$ are finite constants and $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the minimum and maximum eigenvalues of A, respectively. From Berk's results, we know that $\lambda_{min}$ and $\lambda_{max}$ of $\hat{R}_Z$ are bounded and bounded away from zero.

Note that,

$$k^{1/2}(n-k)^{-1/2}|t^*_{i,n}(p,k)| \geq [M_1^{-1}(n-k)^{-1}\Sigma_t V_t^2 \lambda_{min}(\hat{R}_Z)]^{-1/2}$$

$$k^{-(i-d-1)}|\mathbf{1}'\hat{\boldsymbol{\delta}}_V - 1| .$$

We will show that $k^{-(i-d-1)}(\mathbf{1}'\hat{\boldsymbol{\delta}}_V - 1)$ converges to a negative constant and hence that $t^*_{i,n}(p,k)$ diverges to negative infinity. We will present the proof for the case $i = d+2$ and indicate how to extend for $i > d+2$. Consider the case $i = d+2$. The regression (B.15) then is

$$Z_t - Z_{t-1} = \Sigma_{j=1}^{k-d-1}\delta_j(Z_{t-j} - Z_{t-j-1}) + \text{error},$$

or, equivalently,

$$Z_t = \Sigma_{j=1}^{k-d} \theta_j Z_{t-j} + \text{error}, \tag{B.17}$$

where $\Sigma_{j=1}^{k-d}\theta_j$ is <u>restricted</u> to be equal to one. Note that $\delta_j = \Sigma_{\ell=1}^{j}\theta_\ell - 1$, for $j=1,\ldots, k-d-1$ and hence $\mathbf{1}'\boldsymbol{\delta} = \Sigma_{j=1}^{k-d-1}(k-d-j)\ \theta_j - (k-d-1)$. From equation (97) of Searle (1971, p. 206), the restricted least squares estimate of $\boldsymbol{\theta}$ in (B.17) is,

$$\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}} - \hat{R}_Z^{-1}\mathbf{1}(\mathbf{1}'\hat{R}_Z^{-1}\mathbf{1})^{-1}\ (\mathbf{1}'\hat{\boldsymbol{\theta}} - 1)\ , \tag{B.18}$$

where $\hat{\boldsymbol{\theta}} = \hat{R}_Z^{-1}\hat{\mathbf{r}}_Z$ .

Therefore,

$$(k-d)^{-1}(\mathbf{1}'\hat{\boldsymbol{\delta}}_V - 1) = (k-d)^{-1}\Sigma_{j=1}^{k-d-1}(k-d-j)\hat{\theta}^*_j - 1$$

$$= -\mathbf{b}_k'\ \hat{\boldsymbol{\theta}} + \mathbf{b}_k'\ \hat{R}_Z^{-1}\mathbf{1}(\mathbf{1}'\hat{R}_Z^{-1}\ \mathbf{1})^{-1}(\mathbf{1}'\hat{\boldsymbol{\theta}} - 1) \tag{B.19}$$

where $\mathbf{b}_k = (k-d)^{-1}(1,2,\ldots,k-d)'$. From Theorem 1 of Berk (1974) it follows that

$$b_k'(\hat{\theta} - \theta) \xrightarrow{P} 0 ,$$

$$(k-d)^{-1} b_k'(\hat{R}_Z^{-1} - R_Z^{-1}) \, 1 \xrightarrow{P} 0$$

$$(k-d)^{-1} 1'(\hat{R}_Z^{-1} - R_Z^{-1}) 1 \xrightarrow{P} 0$$

and

$$1'(\hat{\theta} - \theta) \xrightarrow{P} 0 ,$$

where $R_Z = \lim_{t\to\infty} \mathrm{Cov}(X_t^*(k))$.

Since $Z_t$ has stationary and invertible roots, we get $|\theta_j| < M\lambda^j$ for some $0 < \lambda < 1$ and $0 < M < \infty$ . Therefore, as k tends to infinity,

$$1' \theta - 1 \rightarrow \Sigma_{j=1}^{\infty} \theta_j - 1 = \beta_{d+1}^* < 0 ,$$

and

$$b_k' \theta = (k-d)^{-1} \Sigma_{j=1}^{k-d} j\theta_j \rightarrow 0 .$$

Also, from the arguments of Theorem 10.2.7 of Anderson (1971) we get, as k tends to infinity,

$$(k-d)^{-1} b_k' R_Z^{-1} 1 \rightarrow \frac{1}{2} [2\pi f_Z(0)]^{-1}$$

and

$$(k-d)^{-1} 1' R_Z^{-1} 1 \rightarrow [2\pi f_Z(0)]^{-1} ,$$

where $f_Z(\cdot)$ is the spectral density of an ARMA (p-d,q) process with $m_{d+1}, \ldots, m_p$ as the roots of the AR polynomial and $m_1^*, \ldots, m_q^*$ as the roots of the MA polynomial.