Biometrika (1984), **71**, 3, *pp*. 599–607
Printed in Great Britain

Testing for unit roots in autoregressive-moving average models of unknown order

By SAID E. SAID

Department of Mathematics, East Carolina University, Greenville, North Carolina, U.S.A.

AND DAVID A. DICKEY

Department of Statistics, North Carolina State University, Raleigh, North Carolina, U.S.A.

Summary

Recently, methods for detecting unit roots in autoregressive and autoregressive-moving average time series have been proposed. The presence of a unit root indicates that the time series is not stationary but that differencing will reduce it to stationarity. The tests proposed to date require specification of the number of autoregressive and moving average coefficients in the model. In this paper we develop a test for unit roots which is based on an approximation of an autoregressive-moving average model by an autoregression. The test statistic is standard output from most regression programs and has a limit distribution whose percentiles have been tabulated. An example is provided.

Some key words: Mixed model; Nonstationary; Time series; Unit root.

1. Introduction

Box & Jenkins (1970) discuss a class of models known as ARIMA (p, d, q) models. The term ARIMA (p, d, q) refers to an autoregressive-moving average model of the form

$$Z_t - \alpha_1 Z_{t-1} - \ldots - \alpha_p Z_{t-p} = e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q},$$

where Z_t is the order d difference of the data and e_t is a white noise sequence. These have become quite popular for modelling time series data. Methods are available for testing the p autoregressive and q moving average coefficients for significance. The value of d indicates the amount of differencing necessary to reduce the series to stationarity. The number d also equals the number of unit roots in the characteristic equation for the time series.

While much has been written about estimation when $d = 0$, relatively little has been written about cases with $d > 0$. D. A. Dickey, in his Iowa State University Ph.D. thesis, Dickey & Fuller (1979) and Hasza & Fuller (1979) discuss cases with $q = 0$. Dickey & Said (1981) discuss a method for testing the hypothesis H_0: $d = 1$ when p and q are known. Unfortunately, p and q are usually unknown and it is necessary to determine d before estimating p and q if we use the approach described by Box & Jenkins (1970).

In the present paper we show that it is possible to approximate an ARIMA $(p, 1, q)$ by an autoregression whose order is a function of the number of observations n. Using least squares to estimate coefficients in this autoregressive approximation produces statistics whose limit distributions are the same as those tabulated by Dickey and listed by Fuller (1976, p. 373). Thus it is possible to test the null hypothesis H_0: $d = 1$ without knowing p or q.

The development is illustrated for $p = q = 1$ and normal errors then extended to the general p, q case with independent identically distributed errors.

2. Estimation procedure

We first consider the model described by

$$\begin{aligned} Y_t &= \rho Y_{t-1} + Z_t \quad (t = 1, 2, \ldots), \\ Z_t &= \alpha Z_{t-1} + e_t + \beta e_{t-1} \quad (t = \ldots, -2, -1, 0, 1, 2, \ldots), \end{aligned} \tag{2·1}$$

where it is assumed that $|\alpha| < 1$, $|\beta| < 1$, $Y_0 = 0$ and $\{e_t\}$ is a sequence of independent identically distributed random variables which, for the moment, we will assume to be normal. If $|\rho| < 1$ then, except for transitory start up effects, Y_t is a stationary ARMA (2, 1) series in the terminology of Box & Jenkins (1970). If $\rho = 1$ the series (2·1) is an ARIMA (1, 1, 1) process. We consider $\rho = 1$ as a null hypothesis to be tested. Notice that

$$e_t = \sum_{j=0}^{\infty} (-\beta)^j (Z_{t-j} - \alpha Z_{t-j-1})$$

and it follows that

$$Y_t - Y_{t-1} = (\rho - 1)(Y_{t-1}) + (\alpha + \beta)(Z_{t-1} - \beta Z_{t-2} + \beta^2 Z_{t-3} - \ldots) + e_t. \tag{2·2}$$

Under the null hypothesis that $\rho = 1$, we see that $Z_t = Y_t - Y_{t-1}$. This motivates us to estimate the coefficients in (2·2) by regressing the first difference $\dot{Y}_t = Y_t - Y_{t-1}$ on $Y_{t-1}, \dot{Y}_{t-1}, \ldots, \dot{Y}_{t-k}$ where k is a suitably chosen integer. To get consistent estimates of the coefficients in (2·2) it is necessary to let k be a function of n. We shall assume that $n^{-1/3} k \to 0$ and that there exist $c > 0, r > 0$ such that $ck > n^{1/r}$.

3. Definitions and notation

Let $d_0 = \rho - 1$ and $d_i = (\alpha + \beta)(-\beta)^{i-1}$ $(i > 0)$ be the coefficients in (2·2). Let

$$X'_t = (Z_{t-1}, \ldots, Z_{t-k}), \quad U'_t = (Y_{t-1} : X'_t)$$

and $d' = (d_0, d_1, \ldots, d_k)$. Now consider a truncated version of (2·2)

$$\dot{Y}_t = (\rho - 1) Y_{t-1} + (\alpha + \beta)\{Z_{t-1} - \beta Z_{t-2} + \beta^2 Z_{t-3} - \ldots + (-\beta)^{k-1} Z_{t-k}\} + e_{tk}. \tag{3·1}$$

Notice that e_{tk} is not a white noise series. In fact $e_{tk} = \dot{Y}_t - U'_t d$. Define $\hat{a} = (\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_k)$ to be the vector of estimated coefficients in the regression of $\dot{Y}_t$ on $Y_{t-1}, \dot{Y}_{t-1}, \ldots, \dot{Y}_{t-k}$, hereafter referred to as regression (3·1). Further, define a sum of squares and cross-products matrix R_0 and a normalizing matrix D_n by

$$R_0 = \sum_{t=k+1}^{n} U_t U'_t = \begin{bmatrix} \Sigma Y_{t-1}^2 & \Sigma Y_{t-1} X'_t \\ \Sigma Y_{t-1} X_t & \Sigma X_t X'_t \end{bmatrix}, \tag{3·2}$$

$$D_n = \operatorname{diag}\{(n-k)^{-1}, (n-k)^{-\frac{1}{2}}, \ldots, (n-k)^{-\frac{1}{2}}\}.$$

We shall develop the limit distribution of

$$D_n^{-1}(\hat{a} - d) = (D_n R_0 D_n)^{-1} D_n \sum_{t=k+1}^{n} U_t e_{tk}. \tag{3·3}$$

The first step will be to show that $D_n R_0 D_n - R$ converges to zero, where R is a block diagonal matrix. Let $\gamma_z(i-j) = E(Z_{t-i} Z_{t-j})$ be the autocovariance function of the stationary series Z_t, $\Gamma_{ij} = \gamma_z(i-j)$, where Γ is a $k \times k$ matrix, and $\gamma' = \{\gamma_z(1), \ldots, \gamma_z(k)\}$. It is well known (Fuller, 1976, p. 239), that

$$(n-k)^{-1} \sum_{t=k+1}^{n} X_t X_t' \rightarrow \Gamma, \quad (n-k)^{-1} \sum_{t=k+1}^{n} X_t Z_t \rightarrow \gamma,$$

where the convergence is in probability. Finally, let $W_t = e_1 + \ldots + e_t$ and define the block diagonal matrix R by

$$R = \operatorname{diag}\{(1-\alpha)^{-2}(1+\beta)^2(n-k)^{-2} \Sigma W_{t-1}^2, \Gamma\}. \tag{3·4}$$

The ideas presented in the next section follow the lines of Berk (1974). He uses the standard Euclidean norm, $\| x \| = (x'x)^{\frac{1}{2}}$, of a column vector x to define a matrix norm $\| B \|$ and defines

$$\| B \| = \sup\{\| Bx \| : \| x \| < 1\}.$$

Notice that $\| B \|^2$ is bounded by the sum of squares of the elements of B and that $\| B \|$ is bounded by the largest modulus of the eigenvalues of B. Using this norm we investigate the convergence of R_0.

4. Asymptotic equivalence of $D_n R_0 D_n$ and R

In this section, we let $Q = D_n R_0 D_n - R$ and show that $\| Q \|$ converges in probability to 0. Let the (i, j)th element of Q be denoted by q_{ij}.

Lemma 4·1. *Under the assumptions in* §3, *if* $\rho = 1$ *and* $n^{-1/3} k \rightarrow 0$ *then* $k^{\frac{1}{2}} \| Q \|$ *converges in probability to* 0.

Proof. First,

$$q_{11} = (n-k)^{-2}\{\Sigma Y_{t-1}^2 - (1-\alpha)^{-2}(1+\beta)^2 \Sigma W_{t-1}^2\}.$$

For $\rho = 1$ we have $Y_{t-1} = Z_1 + \ldots + Z_{t-1}$ so that

$$Y_{t-1} - \alpha Y_{t-2} = \sum_{j=1}^{t-1} (e_j + \beta e_{j-1}) + \alpha Z_0$$

and thus

$$(1-\alpha) Y_{t-1} + \alpha Z_{t-1} = (1+\beta) \sum_{j=1}^{t-1} e_j + \beta(e_0 - e_{t-1}) + \alpha Z_0 \tag{4·1}$$

or

$$Y_{t-1} = (1-\alpha)^{-1}(1+\beta) W_{t-1} + O_p(1).$$

Dickey & Fuller (1979) show that $\Sigma W_{t-1}^2 = O_p(n^2)$. Their arguments show that $(n-k) E(q_{11}^2) \leqslant C_1$, where C_1 is some constant.

Berk (1974) shows that for some C_2,

$$(n-k) E\{(n-k)^{-1} \Sigma Z_t Z_{t-j} - \gamma_z(j)\}^2 \leqslant C_2 \quad (j = 1, \ldots, k).$$

For $i > 1$, $j > 1$, we note that $q_{ij} = (n-k)^{-1} \Sigma Z_{t-i} Z_{t-j} - \gamma_z(i-j)$.

We now consider q_{ij} for $i = 1$, $j > 1$. The vector of these q_{ij} from (3·2) is $(n-k)^{-3/2} \Sigma Y_{t-1} X_t'$. Now for some C_3,

$$(n-k) E(q_{ij}^2) = (n-k)^{-2} \Sigma_s \Sigma_t E(Y_{t-1} Z_{t-j} Y_{s-1} Z_{s-j}) \leqslant C_3.$$

This follows from the facts that

$$|E(Y_t Z_s)| = \left|\sum_{j=1}^{t} \gamma_z(s-j)\right| \leqslant \sum_{j=-\infty}^{\infty} |\gamma_z(j)| = C_4 < \infty,$$

$$|E(Y_t Y_s)| \leqslant t \sum_{j=-\infty}^{\infty} |\gamma_z(j)| = tC_4.$$

To finish the proof, let $C = \max(C_1, C_2, C_3)$, so that $(n-k)E(q_{ij}^2) \leqslant C$ for all i, j; note that C_1, C_2 and C_3 do not depend on n or k. Since Q has dimension $(k+1)\times(k+1)$ we see $E(\|Q\|^2) \leqslant C(k+1)^2/(n-k)$ and if $k^2/n \to 0$ then $\|Q\|$ converges in probability to 0. Under our assumption that $k^3/n \to 0$ we have $k^{\frac{1}{2}}\|Q\|$ converging to 0 also.

LEMMA 4·2. *Under the assumptions of model* (2·1) *with* $\rho = 1$

$$\left\{(1-\alpha)^{-2}(1+\beta)^2(n-k)^{-2}\sum_{t=k+1}^{n} W_{t-1}^2\right\}^{-1} = O_p(1).$$

Proof. Now

$$\sum_{t=1}^{n} W_{t-1}^2 = e'Ae,$$

where $e' = (e_1, \ldots, e_{n-1})$ and the (i, j)th element of A is $n - \max(i, j)$. Thus

$$e'Ae = \sum_{i=1}^{n-1} \gamma_{in} Z_i^2,$$

where $\{Z_i\}$ is an independent $N(0, \sigma^2)$ sequence, and (Dickey & Fuller, 1979)

$$\gamma_{in} = 4^{-1}\sec^2\{(2n-1)^{-1}(n-i)\pi\} \quad (i = 1, \ldots, n).$$

Also, for any fixed i, $n^{-2}\gamma_{in} \to 4\{(2i-1)\pi\}^{-2} > 0$ as $n \to \infty$, so that, given $\varepsilon > 0$, we can find M_ε such that

$$\mathrm{pr}\left\{n^2\left(\sum_{t=2}^{n} W_{t-1}^2\right)^{-1} > M_\varepsilon\right\} < \mathrm{pr}(n^2\gamma_{1n}^{-1}Z_1^{-2} > M_\varepsilon) = \mathrm{pr}(n^{-2}\gamma_{1n}^{-1}Z_1^2 < M_\varepsilon^{-1}) < \varepsilon.$$

The ability to find M_ε follows from the fact that $n^{-2}\gamma_{1n}Z_1^2$ converges to a nonzero multiple of a χ_1^2 variate. This proves the result through a direct application of the definition of $O_p(1)$.

In the following lemma we show that the norm of R^{-1} is bounded in probability.

LEMMA 4·3. *Under the assumptions of model* (2·1) *with* $\rho = 1$, $\|R^{-1}\| = O_p(1)$.

Proof. Since the matrix R is block diagonal with invertible diagonal blocks, R^{-1} is block diagonal and $\|R^{-1}\|$ is bounded by the sum of the norms of the diagonal blocks of R^{-1}. The lower right-hand corner block of R^{-1} is Γ^{-1}, which is the inverse of the autocovariance matrix of a stationary invertible series, and thus $\|\Gamma^{-1}\|$ is bounded. By Lemma 4·2 the upper left-hand corner element of R^{-1} is also bounded in probability.

The main result of this section is proved next.

THEOREM 4·1. *Under the conditions of model* (2·1) *with* $\rho = 1$, $n^{-1/3} k \to 0$, *we have that* $k^{\frac{1}{2}} \| D_n^{-1} R_0^{-1} D_n^{-1} - R^{-1} \|$ *converges in probability to* 0.

Proof. Let $R^* = D_n R_0 D_n$, $q = \| R^{*-1} - R^{-1} \|$ and $p = \| R^{-1} \|$. Now $p = O_p(1)$ and

$$\begin{aligned} q = \| R^{*-1}(R - R^*) R^{-1} \| &\leqslant \| R^{*-1} \| \| R - R^* \| \| R^{-1} \| \\ &\leqslant (p+q) \| Q \| p \\ &= p^2 \| Q \| + qp \| Q \|, \end{aligned} \quad (4{\cdot}2)$$

and since $k^{\frac{1}{2}} \| Q \|$ converges in probability to 0 we find, upon rearranging (4·2), $k^{\frac{1}{2}} q \leqslant k^{\frac{1}{2}}(1 - p \| Q \|)^{-1} p^2 \| Q \| \to 0$. Here we have used the fact that pr ($\| Q \| > 1$) can be made arbitrarily small by choice of n since $p \| Q \| \to 0$.

5. CONSISTENCY OF LEAST SQUARES ESTIMATORS

In § 2 we propose the use of least squares autoregression to estimate ρ in model (2·1). We estimate $\rho - 1$ by the first element of a vector $\hat{a}$ given by expression (3·3). In § 4 we looked at $D_n R_0 D_n$, so it remains to consider $D_n \Sigma U_t e_{tk}$. We show that this vector converges to 0 and thus prove consistency.

LEMMA 5·1. *As* $n \to \infty$

$$\left\| D_n \sum_{t=k+1}^{n} U_t (e_{tk} - e_t) \right\| = O_p(n^{-1}).$$

Proof. We have

$$\left\| D_n \sum_{t=k+1}^{n} U_t (e_{tk} - e_t) \right\|^2 =$$

$$(n-k)^{-2} \{\Sigma_t Y_{t-1}(e_{tk} - e_t)\}^2 + (n-k)^{-1} \sum_{j=1}^{k} \{\Sigma_t Z_{t-j}(e_{tk} - e_t)\}. \quad (5{\cdot}1)$$

Since $\{Z_t\}$ is a stationary invertible ARMA process, we can use (3·1) and (3·2) to see that

$$E\left[(n-k)^{-1} \sum_{j=1}^{k} \{\Sigma_t Z_{t-j}(e_{tk} - e_t)\}^2 \right] \leqslant ck(n-k) \sum_{i=k+1}^{\infty} d_i^2,$$

where c is a constant. Further, there exists λ with $|\beta| < \lambda < 1$ such that $|d_i|$ is bounded by a constant multiple of λ^i. Thus $k(n-k)(d_{k+1}^2 + d_{k+2}^2 + \ldots) = O(n^{-2})$ under our assumption that k is bounded below by a positive multiple of $n^{1/r}$ for some $r > 0$.

The order of the first term in (5·1) is established by expressing Y_{t-1} and $e_{tk} - e_t$ in terms of $\{Z_i\}$ and the proof is thereby completed.

At this point we will establish the probability order of

$$\left\| D_n \sum_{t=k+1}^{n} U_t e_t \right\|,$$

which, in combination with Lemma 5·1 and Theorem 4·1, will establish consistency.

LEMMA 5·2. *As* $n \to \infty$,

$$\left\| D_n \sum_{t=k+1}^{n} U_t e_t \right\| = O_p(k^{\frac{1}{2}}).$$

Proof. Now

$$E \| D_n \Sigma_t U_t e_t \|^2 = (n-k)^{-2} E(\Sigma Y_{t-1} e_t)^2 + (n-k)^{-1} \sum_{j=1}^{k} E\left(\sum_{t=1}^{n-k} Z_{t-j} e_t\right)^2.$$

Since Z_t is a stationary invertible autoregressive-moving average process

$$(n-k)^{-1} \sum_{j=1}^{k} E\left(\sum_{t=1}^{n-k} Z_{t-j} e_t\right)^2 = k\gamma_z(0)\,\sigma^2.$$

Here we have used the independence of Z_{t-j} and e_t for $j > 0$. This same fact shows that

$$E(\Sigma_t Y_{t-1} e_t)^2 = \sigma^2 \Sigma_t E(Y_{t-1}^2) = O(n^2).$$

Combining Lemmas 5·1 and 5·2, we see that

$$\| D_n \Sigma_t U_t e_{tk} \| = O_p(k^{\frac{1}{2}}) \tag{5·2}$$

and we thus are in a position to prove our main result.

THEOREM 5·1. *Under the assumptions of model* (2·1) *with* $\rho = 1$, $\| \hat{a} - d \|$ *converges in probability to* 0.

Proof. We write

$$D_n^{-1}(\hat{a}-d) = (R^{*-1} - R^{-1}) D_n \Sigma_t U_t e_{tk} + R^{-1} D_n \Sigma_t U_t e_t + R^{-1} D_n \Sigma_t U_t (e_{tk} - e_t). \tag{5·3}$$

Taking norms of the three terms above we obtain, respectively, $o_p(1)$, $O_p(k^{\frac{1}{2}})$ and $O_p(n^{-1})$ so that $\| D_n^{-1}(\hat{a}-d) \| = O_p(k^{\frac{1}{2}})$. Since $\| D_n^{-1} \| = O(n-k)$ and $k = O(n^{1/3})$ the proof is complete.

6. LIMIT DISTRIBUTION OF UNIT ROOT TEST STATISTICS

Having shown that $D_n^{-1}(\hat{a}-d) = O_p(1)$ we now develop the limit distribution of $D_n^{-1}(\hat{a}-d)$ which, by (5·3), is the same as that of $R^{-1} D_n \Sigma_t U_t e_t$. The first element of $D_n^{-1}(\hat{a}-d)$ is

$$(n-k)(\hat{\rho}-1) = \{(n-k)^{-2} \Sigma Y_{t-1}^2\}^{-1} (n-k)^{-1} \Sigma Y_{t-1} e_t.$$

By Lemma 4·1,

$$(n-k)^{-2} \Sigma Y_{t-1}^2 = (1-\alpha)^{-2}(1+\beta)^2 \Sigma W_{t-1}^2 + O_p(n^{-\frac{1}{2}}). \tag{6·1}$$

By the arguments of Lemma 5·1,

$$(n-k)^{-1} \Sigma Y_{t-1} e_{tk} = (n-k)^{-1} \Sigma Y_{t-1} e_t + O_p(n^{-1}) \tag{6·2}$$

and, if we use result (4·1),

$$(n-k)^{-1} \Sigma Y_{t-1} e_t = (n-k)^{-1}(1-\alpha)^{-1}(1+\beta) \Sigma W_{t-1} e_t + O_p(n^{-\frac{1}{2}}).$$

Dickey & Fuller (1979) define random variables (Γ, ξ) and show that

$$(\sigma^{-2} n^{-2} \Sigma W_{t-1}^2,\ \sigma^{-2} n^{-1} \Sigma W_{t-1} e_t)$$

converges in law to (Γ, ξ). If we use (6·1) and (6·2), the distribution of $\Gamma^{-1}\xi$ is the same as the limit distribution of

$$(1-\alpha)^{-1}(1+\beta)(n-k)(\hat{\alpha}_0-d_0) = (1-\alpha)^{-1}(1+\beta)(n-k)(\hat{\rho}-1).$$

If we use the results of Hasza & Fuller (1979) the above limiting distribution is unchanged if the e_t sequence is taken to be independently and identically distributed with zero mean and variance σ^2. Further, since R is block diagonal, the results of Berk (1974) apply to the coefficient vector $(\hat{\alpha}_1-d_1, \ldots, \hat{\alpha}_k-d_k)$. That is, the limit distribution of this part of $a-d$ is the same whether or not we include Y_{t-1} on the right-hand side in regression (3·1). Thus, for large n, significance tests for these coefficients are not affected by including Y_{t-1} in the regression.

Since the distribution of $n(\hat{\rho}-1)$ involves the unknown parameters α and β we would be unable to use $\hat{\rho}$ as a test for unit roots at the identification stage of analysis. We now show that the limit distribution of the studentized t statistic associated with $\hat{\rho}$ does not depend on any unknown parameters.

THEOREM 6·1. *Under the assumptions of Theorem 4·1, define $\tau = (C_{11}\hat{\sigma}^2)^{-\frac{1}{2}}(\hat{\rho}-1)$, where C_{11} is the upper left-hand corner element of R_0^{-1}, $\hat{\rho}-1 = \hat{\alpha}_0-d_0$, and $\hat{\sigma}^2$ is the error mean square from regression (3·1). Then the limit distribution of τ is the same as the distribution of $\Gamma^{-\frac{1}{2}}\xi$.*

Proof. By our previous consistency results, $\hat{\sigma}^2$ converges in probability to σ^2. Using the arguments of Lemma 4·1 we see that

$$\tau - \{n^2\hat{\sigma}^2(\Sigma Y_{t-1}^2)^{-1}\}^{-\frac{1}{2}} n(\hat{\rho}-1) \rightarrow 0.$$

If we use (6·1) and (6·2) it follows that

$$\tau - \hat{\sigma}^{-1}(\Sigma W_{t-1}^2)^{-\frac{1}{2}}\Sigma W_{t-1}e_t \rightarrow 0$$

and thus $\tau \rightarrow \Gamma^{-\frac{1}{2}}\xi$.

Percentiles of the distribution of τ are given by Fuller (1976, p. 373) and can be used to test the null hypothesis of a unit root. An illustration is given in §8.

7. EXTENSIONS

The general ARIMA $(p, 1, q)$ model is defined by

$$Y_t = \rho Y_{t-1} + Z_t \quad (t = 1, 2, \ldots), \tag{7·1}$$

$$Z_t + \sum_{i=1}^{p}\alpha_i Z_{t-i} = e_t + \sum_{j=1}^{q}\beta_j e_{t-j}, \tag{7·2}$$

with $Y_0 = 0$, e_t a normal pure noise sequence and $\rho = 1$. We assume that Z_t is stationary and invertible. Thus there exists a sequence of real numbers $\{d_j\}$, a number $0 < \lambda < 1$ and a number M such that $e_t = \Sigma d_j Z_{t-j}$ and $|d_j| < M\lambda^j$ (Fuller, 1976, Theorem 2·7·2). Rearranging, we can write, in analogy to (3·1),

$$\dot{Y}_t = (\rho-1)Y_{t-j} - d_1 Z_{t-1} - \ldots - d_k Z_{t-k} + e_{tk}. \tag{7·3}$$

Notice that the consistency proofs in the ARIMA $(1, 1, 1)$ case depend only on the existence of exponentially decreasing bounds on the d_i and so they generalize immediately to the higher order case.

If $\rho = 1$, $Z_{t-j} = Y_{t-j} - Y_{t-j-1} = \dot{Y}_{t-j}$ and in regression (7·3) we will refer to the least squares coefficient on Y_{t-1} as $\hat{\rho}-1$. Summation of (7·2) on both sides as t goes from 1 to n shows that, in analogy to (4·1) with $\rho = 1$,

$$(1+\alpha_1+\ldots+\alpha_p)\, Y_t = (1+\beta_1+\ldots+\beta_q)\, W_t + V_t, \tag{7·4}$$

where V_t is a linear combination of a finite number of Z_t and e_t and $W_t = e_1+\ldots+e_t$ as before. If we use (7·4) and the arguments of §6 it follows that, for the general ARIMA $(p, 1, q)$ case, then in law

$$(1+\alpha_1+\ldots+\alpha_p)^{-1}(1+\beta_1+\ldots+\beta_q)\, n(\hat{\rho}-1) \rightarrow \Gamma^{-1}\xi, \tag{7·5}$$

as does the studentized statistic from the least squares regression (7·3).

Finally, if the series mean $\bar{Y}$ is subtracted from each Y_t prior to analysis, we replace (7·4) by

$$(1+\alpha_1+\ldots+\alpha_p)\,(Y_t - \bar{Y}) = (1+\beta_1+\ldots+\beta_q)\,(W_t - \bar{W}) + V_t, \tag{7·6}$$

where $\bar{W} = n^{-1}\Sigma\, W_t$. We note that $Z_t = (Y_t - \bar{Y}) - (Y_{t-1} - \bar{Y}) = \dot{Y}_t$ when $\rho = 1$. Using the results of Dickey & Fuller (1979) we get distributions similar to (7·5) except that the limit distribution percentiles now correspond to the $\hat{\rho}_\mu$ and τ_μ tables of Fuller (1976, p. 371–3). This follows from the fact that for fixed j, $\bar{Y}\,\Sigma\, Z_{t-j} = O_p(n^{\frac{1}{2}})$, so that the order results following (4·1) are unchanged if the series mean $\bar{Y}$ is subtracted prior to analysis. The limit matrix R is still block diagonal. The only difference from (3·3) is that the upper left-hand element of R is now, if we use (6·1),

$$(n-k)^{-2}(1+\alpha_1+\ldots+\alpha_p)^{-2}(1+\beta_1+\ldots+\beta_q)^2\,\Sigma\,(W_t - \bar{W}^2),$$

and a similar modification in (6·2) yields results like (7·5).

8. EXAMPLE

Box & Jenkins (1970, p. 525) list 197 concentration readings from a chemical process. The authors conclude that an ARMA (1, 1) or an ARIMA (0, 1, 1) model should be fitted to the data. We use our testing procedure to test the hypothesis that the model form is ARIMA $(p, 1, q)$.

Using SAS (Barr, Goodnight & Sall, 1979) we fit the model

$$\dot{Y}_t = -0{\cdot}1601(Y_{t-1} - \bar{Y}) - 0{\cdot}4941\,\dot{Y}_{t-1} - 0{\cdot}2919\,\dot{Y}_{t-2} - 0{\cdot}2640\,\dot{Y}_{t-3}$$
$$-0{\cdot}2477\,\dot{Y}_{t-4} - 0{\cdot}2682\,\dot{Y}_{t-5} - 0{\cdot}1888\,\dot{Y}_{t-6},$$

where the error mean square is 0·0938 and $\dot{Y}_t = Y_t - Y_{t-1}$. The coefficient standard errors are 0·0785, 0·0963, 0·0985, 0·0947, 0·0903, 0·0858 and 0·0726. The studentized statistic $\hat{\tau}_\mu = (0{\cdot}0785)^{-1}(-0{\cdot}1601) = -2{\cdot}04$ is compared to the $\hat{\tau}_\mu$ tables of Fuller (1976, p. 373). We do not reject the unit root hypothesis using a one sided 10% significance level test.

The limit theory, of course, does not specify the value of k for any given n. To select k, we initially fit regression (3·1) for $k = 6$, 7, 8, 9 and 10. Assuming that an autoregressive order 10 model gives a sufficient approximation to the data we used the standard regression F test that the coefficients on $\dot{Y}_{t-7}$ to $\dot{Y}_{t-10}$ are simultaneously zero. The justification for this as an approximate test is the fact from §6 that Berk's results hold for the coefficients on $\dot{Y}_{t-j}$ in regression (3·1). The sequential sums of squares for $\dot{Y}_{t-7}$ to $\dot{Y}_{t-10}$ are 0·0279, 0·0019, 0·0013 and 0·1341. The error sum of squares is

16·6423 with 174 degrees of freedom and we compute $F^4_{174} = 0{\cdot}43$. None of the individual t tests for these 4 lags was significant, even at the 20% level.

Furthermore, when regression (3·1) is fitted with $k = 7, 8, 9$ and 10, the τ_μ statistics are $-1{\cdot}931$, $-1{\cdot}830$, $-1{\cdot}796$ and $-2{\cdot}013$ so that our unit root test is not affected over this range of k.

One will now want to fit a model to the differenced data. Once this has been done, unit root tests based on known p and q are available (Dickey & Said, 1981).

References

Barr, A. J., Goodnight, J. H. & Sall, J. P. (1979). *SAS User's Guide*. Cary, North Carolina: SAS Institute.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489–502.

Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* **74**, 427–31.

Dickey, D. A. & Said, S. E. (1981). Testing ARIMA $(p, 1, q)$ versus ARMA $(p+1, q)$. *Proc. Bus. Econ. Statist. Sect., Am. Statist. Assoc.*, 318–22.

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

Hasza, D. P. & Fuller, W. A. (1979). Estimation for autoregressive processes with unit roots. *Ann. Statist.* **7**, 1106–20.

[*Received November* 1983. *Revised February* 1984]