A STRONGLY CONSISTENT CRITERION TO DECIDE BETWEEN I (1) AND I (0) PROCESSES BASED ON REGRESSION PROCEDURES

*Victor Gómez Enríquez**

D-2011-04

June, 2011

*Ministerio de Economía y Hacienda

Address for correspondence: Dirección General de Presupuestos. Subdirección General de Análisis y Programación Económica. Alberto Alcocer 2, 1-P, D-34. 28046 Madrid.
e-mail: vgomez@sgpg.meh.es

This document is available at: http://www.sgpg.pap.meh.es/SITIOS/SGPG/ES-ES/PRESUPUESTOS/DOCUMENTACION/Paginas/Documentación.aspx

Abstract

The usual procedure to determine whether a univariate time series is stationary or first–difference stationary is to perform some unit root test. In this article, an alternative methodology is presented that leads to a strongly consistent two–step criterion to estimate the number of unit roots. The criterion is based on estimating some autoregressive polynomials using regression procedures and exploiting the fact that the nonstationary roots converge at a faster rate than the stationary ones. The proposed procedure requires at most four regressions and is easy to implement. A simulation study demonstrates that it can perform significantly better in practice than the Dickey–Fuller and the GLS de–trended Dickey–Fuller tests.

KEY WORDS: Unit Root Tests; GLS de–trending; Least Squares Estimator; Hannan–Rissanen Procedure; Strong Consistency; Nonstationarity.

1. INTRODUCTION

Suppose a time series $\{y_t\}$, $t = 1, \ldots, n$, is available such that it is either stationary or stationary after first differencing, that is, $\nabla y_t = y_t - y_{t-1}$ is stationary. A problem usually found in practice is to determine whether the series should be differenced. To address this problem, a number of tests for unit roots have been developed. See, for example, Hamilton (1994) and the references therein. Probably, the most widely used unit root tests are those referred to as Dickey–Fuller tests. See Dickey, Bell and Miller (1986) for a review and list of references. Subsequently, Elliott, Rothenberg and Stock (1996) proposed a modification of these unit root tests in which the data are subject to generalized least squares (GLS) de–trending prior to the estimation of an autoregression and the computation of the test statistic. Based upon these last tests, Ng and Perron (2001) developed a new family of tests that were supposed to have better size and power than all previously mentioned unit root tests.

Unfortunately, evidence presented by Schwert (1989), Pantula (1991), DeJong, Nankervis, Savin and Whiteman (1992), or Perron and Ng (1996) indicates that the most usually applied test procedures do not do very well in practice, particularly when large moving average roots are present in nonstationary series or when the root of the autoregressive polynomial is close to but less than unity in stationary series. The situation is worsened by the inclusion in the regression of constant terms or time trends.

It is well known that when one estimates an autoregression by least squares and there is a unit root in the process, the unit root estimator converges at a faster rate than the estimators of the stationary roots (Hamilton, 1994). In fact, almost sure bounds for the different terms in the formula for the least squares estimator exist. See, for example, Lemma A.1 of Poskitt (2000).

Rather than couching the determination of unit roots in a hypothesis testing framework, we build on the rate of convergence of the roots of at most two autoregressive polynomials estimated by regression procedures to construct a strongly consistent criterion to select the number of unit roots. The criterion is similar to the differenced information criterion (DIC) proposed by Akaike (1976) to determine whether a canonical correlation is zero in time series model specification.

The definitions of I(0) or I(1) processes considered in this article are not as general as the ones considered, for example, in Stock (1994), Műller (2008) or Davidson (2009). However, we believe that they are sufficiently general to cover most of the cases usually encountered in practice.

Other strongly consistent rules to decide whether a time series is I(1) or I(0) have also been proposed in the literature (Corradi, 1999). However, to the best of our knowledge, there is no simulation study available comparing any of these last rules with the Dickey–Fuller tests or any of their modifications.

The proposed procedure is easy to implement and does not require the use of tables based on simulations, like those of the Dickey–Fuller tests. Another advantage of the proposed procedure is that its application does not depend on the distribution of the true model. Thus, for the case in which a constant is included in the regression, which is the case considered in this article, the procedure is the same whether the true model is

I(0), I(0) with a mean included, I(1), or I(1) with a drift term included. This contrasts with the Dickey–Fuller test, where it is well known that the test statistic depends on the assumptions about the true model.

The results of a simulation study presented later in this article show that the proposed criterion works better than the augmented Dickey–Fuller test in many important cases in practice. Another interesting feature that emerges from the simulation study is the almost complete lack of power of the GLS de–trended Dickey–Fuller test when it is applied to stationary series with a nonzero mean in the model. This serious drawback seems to have passed unnoticed in the literature.

It is to be noticed that, since the Ng and Perron (2001) tests also use GLS de–trending, these tests will probably be affected by the previously mentioned drawback.

Given a time series $\{y_t\}$, $t = 1, \ldots, n$, the proposed methodology consists of the following two steps:

- Step 1

  1. Estimate by OLS an autoregression $y_t + \beta_1 y_{t-1} + \cdots + \beta_k y_{t-k} = u_t$.
  2. Compute the inverse roots $\hat{\lambda}_i$, $i = 1, \ldots, k$, of the polynomial $\hat{\beta}(z) = 1 + \hat{\beta}_1 z + \cdots + \hat{\beta}_k z^k$, where $\hat{\beta}_i$ is the OLS estimator of β_i, $i = 1, \ldots, k$. That is, $\hat{\beta}(z) = \prod_{i=1}^{k}(1 - \hat{\lambda}_i z)$.
  3. Accept $\{y_t\}$ as I(1) if $\text{Real}(\hat{\lambda}_i) > 1 - h_n$ and $|\text{Imag}(\hat{\lambda}_i)| < h_n$ for some $i = 1, \ldots, k$, where $\text{Real}(\hat{\lambda}_i)$ and $\text{Imag}(\hat{\lambda}_i)$ are the real and the imaginary parts of $\hat{\lambda}_i$, $h_n = 1/n^{\alpha_n}$, and $\{\alpha_n\}$ is an increasing sequence of positive integers such that $\lim_{n\to\infty} \alpha_n = \alpha < 0.5$, to be specified later. Accept $\{y_t\}$ as I(0) otherwise. If the series is accepted as I(1), stop. Otherwise, go to Step 2.

- Step 2

  1. Estimate by the Hannan–Rissanen (1982) method an ARMA$(1, 1)$ model, $y_t + \phi y_{t-1} = a_t + \theta a_{t-1}$. Let $\hat{\phi}$ and $\hat{\theta}$ be the estimated ϕ and θ and let $\hat{\lambda} = -\hat{\phi}$.
  2. Accept $\{y_t\}$ as I(1) if $\hat{\lambda} > 1 - j_n$ and $|\hat{\phi} - \hat{\theta}| > c$, where c is a small value used to detect cancelation, $j_n = 1/n^{\beta_n}$ and $\{\beta_n\}$ is an increasing sequence of positive integers satisfying $\lim_{n\to\infty} \beta_n = \beta < 0.5$. Both, c and $\{\beta_n\}$ will be specified later. Accept $\{y_t\}$ as I(0) otherwise.

For completeness, the augmented Dickey–Fuller (ADF) test and the Dickey–Fuller test with GLS de–trending (DFGLS) are briefly described in Section 2. In Section 3, the assumptions needed to develop the asymptotic properties of the proposed procedure are made. Also in Section 3, the estimation method for Step 2, based on the three stage procedure proposed by Hannan and Rissanen (1982), is described. The details of Step 1, as well as suggestions on how to select the sequence $\{\alpha_n\}$ and the number of lags k, will be given in Section 4. Section 5 will describe the details of Step 2. Section 6 discusses the impact of applying mean corrections, and Section 7 presents the results of a simulation study. Section 8 presents the conclusions.

2. THE ADF AND DFGLS TESTS

The ADF test is carried out by performing a regression of the form

$$\nabla y_t = \alpha y_{t-1} + x_t'\delta + \beta_1 \nabla y_{t-1} + \beta_2 \nabla y_{t-2} + \cdots + \beta_p \nabla y_{t-p} + v_t, \tag{1}$$

where x_t are optional exogenous regressors which may consist of a constant, or a constant and trend, and the $\{v_t\}$ are assumed to be white noise. The null and alternative hypotheses may be written as

$$\begin{array}{ll} H_0: & \alpha = 0 \\ H_1: & \alpha < 0, \end{array} \tag{2}$$

and evaluated using the conventional t–ratio for α, $t_\alpha = \hat{\alpha}/(se(\hat{\alpha}))$, where $\hat{\alpha}$ is the estimate of α and $se(\hat{\alpha})$ is the coefficient standard error. An important result obtained by Fuller is that the asymptotic distribution of the t–ratio for α is independent of the number of lagged first differences included in the regression. It is known that including irrelevant regressors in the regression will reduce the power of the test to reject the null of a unit root.

Strictly speaking, when x_t is not zero, the null hypothesis H_0 in (2) should include some statement about x_t. For example, when $x_t = \{1\}$, which is the case considered in this article, H_0 should be $\alpha = 0$ *and* $\delta = 0$. This is so because the distribution of the test statistic depends on whether a nonzero drift term is included in the true model or not (Hamilton, 1994). However, it has become standard practice in econometrics to use $H_0 : \alpha = 0$ (Hayashi, 2000, pp. 580, 595).

For the cases in which a constant or a constant and a linear trend are included in the ADF test regression, Elliott, Rothenberg and Stock (1996) propose a simple modification in which the data are de–trended so that explanatory variables are "taken out" of the data prior to running the test regression. These authors first define a quasi–difference of y_t that depends on the value a representing the specific point alternative against which the null is to be tested:

$$d(y_t|a) = \begin{cases} y_t & \text{if} \quad t = 1 \\ y_t - a y_{t-1} & \text{if} \quad t > 1. \end{cases}$$

They then consider an ordinary least squares (OLS) regression of the quasi–differenced data $d(y_t|a)$ on the quasi–differenced data $d(x_t|a)$

$$d(y_t|a) = d(x_t|a)'\delta(a) + w_t,$$

where x_t contains either a constant or a constant and trend. Letting $\hat{\delta}(a)$ be the OLS estimate from this regression and defining the GLS de–trended data y_t^d as

$$y_t^d = y_t - x_t'\hat{\delta}(a),$$

the DFGLS test involves estimating the equation

$$\nabla y_t^d = \alpha y_{t-1}^d + \beta_1 \nabla y_{t-1}^d + \beta_2 \nabla y_{t-2}^d + \cdots + \beta_p \nabla y_{t-p}^d + v_t,$$

where now the exogenous variables x_t are not included. As with the ADF test, the test statistic is the t–ratio for $\hat{\alpha}$ from the test equation. The values for a recommended by Elliott, Rothenberg and Stock (1996) are

$$a = \begin{array}{ll} 1 - 7/n & \text{if} \quad x_t = \{1\} \\ 1 - 13.5/n & \text{if} \quad x_t = \{1, t\}, \end{array}$$

where n is the sample size.

3. ASSUMPTIONS AND THE HANNAN–RISSANEN METHOD

To develop the statistical properties of the proposed procedure, the following assumptions concerning the process $\{y_t\}$ will be made:

A1 The process $\{y_t\}$ is either stationary (I(0)) or first–difference stationary (I(1)). That is, $w_t = \delta(B)y_t$ is stationary, where the polynomial $\delta(B)$ in the backshift operator $By_t = y_{t-1}$, is either 1 or $\nabla = 1 - B$. Moreover, $w_t = \sum_{i \geq 0} \psi_i a_{t-i}$.

A2 The series $\psi(z) = \sum_{i \geq 0} \psi_i z^i$ converges for $|z| < 1$ and satisfies $\sum_{i \geq 0} i|\psi_i| < \infty$ and $\sum_{i \geq 0} \psi_i = \psi(1) \neq 0$.

A3 The process $\{a_t\}$ constitutes a sequence of independent and identically distributed random variables with zero mean and variance $\sigma^2 > 0$ satisfying the moment condition $E|a_t|^{2+\alpha} < \infty$ for some $\alpha > 2$.

As noted by Poskitt (2000), Assumption A2 ensures that the Wold representation of z_t is invertible and that the limiting distribution of the associated partial sum process is not degenerate, and Assumption A3 implies that mean–squared error projections will be linear and rules out conditional heteroscedasticity. It is further noted by Poskitt (2000) that Assumption A3 is not critical for the derivations and extension to more general processes satisfying appropriate martingale or mixing conditions could be made using established theory. See, for example, Corradi (1999).

The estimation method used in Step 2 of the proposed procedure is based on the method proposed by Hannan and Rissanen (1982). It consists of the following three stages:

1. A long autoregressive model is fitted to the series $\{y_t\}$ using the Durbin–Levinson algorithm to obtain estimates $\hat{a}_t$ of the innovations a_t defined earlier in A1–A3. That is, given an adequate positive integer N, the $\hat{a}_t$ are computed using

$$\hat{a}_t = y_t + \sum_{j=1}^{N} \hat{\phi}_N(j) y_{t-j}, \quad 1 \leq t \leq n, \tag{3}$$

where $y_t = 0$ if $t \leq 0$ to start the recursion, and the $\hat{\phi}_N(j)$, $j = 1, \ldots, N$, are the coefficients estimated by the Durbin–Levinson algorithm. The value of N is selected to be $N = \{\ln^2(n)\}$, where $\{\ln^2(n)\}$ is the integer part of $\ln^2(n)$. This choice is based on the fact that Hannan and Rissanen (1982), p. 88, assume that n is greater than $\ln(n)$, but not greater than $\ln^b(n)$, for some $b < \infty$.

2. Minimize

$$S = \sum_{t=2}^{n} \left\{ y_t + \phi y_{t-1} - \theta \hat{a}_{t-1} \right\}^2. \tag{4}$$

The use of an efficient numerical method, like, for example, the application of the QR algorithm based on Housholder transformations, to minimize (4) is important to avoid singularity problems when the model is overspecified.

3. Let $\tilde{\phi}$ and $\tilde{\theta}$ be the estimators of ϕ and θ obtained by minimizing (4). To obtain the final estimators, first form

$$\tilde{a}_t = -\tilde{\theta}\tilde{a}_{t-1} + y_t + \tilde{\phi} y_{t-1}, \quad t \geq 1, \tag{5}$$

where $\tilde{a}_t = 0$ and $y_t = 0$ if $t \leq 0$. Then put

$$\eta_t = -\tilde{\phi}\eta_{t-1} + \tilde{a}_t, \quad \xi_t = -\tilde{\theta}\xi_{t-1} + \tilde{a}_t, \quad t \geq 1,$$

where $\eta_t = 0$ and $\xi_t = 0$ if $t \leq 0$. Finally, regress $\tilde{a}_t$ on $-\eta_{t-1}$, and ξ_{t-1}. The estimated regression coefficients are added to the estimators $\tilde{\phi}$ and $\tilde{\theta}$ to obtain the desired estimators $\hat{\phi}$ and $\hat{\theta}$.

The third step can be considered as a modified Gauss–Newton step in which $\eta_{t-1} = \partial a_t/\partial\phi$, $\xi_{t-1} = -\partial a_t/\partial\theta$, and the modification consists of using starting values for y_t that are zero. It is to be noted that the recursion (5) can explode if the estimate $\tilde{\theta}$ is greater than one in absolute value. For this reason, if this happens in the proposed procedure, the third stage is omitted and the estimates used are those given by the first two stages of the Hannan–Rissanen method.

4. IDENTIFICATION OF UNIT ROOTS IN STEP 1

Let $\beta(B) = 1 + \beta_1 B + \cdots + \beta_k B^k = \varphi(B)\delta(B)$, where $\delta(B)$ is the polynomial defined in A1, that is $\delta(B) = 1 - B$ if $\{y_t\}$ is difference stationary and $\delta(B) = 1$ if $\{y_t\}$ is stationary, and $\varphi(B)$ is the polynomial defined by the autoregression $\varphi(B)w_t = u_t$, $w_t = \delta(B)y_t$. Then, w_t is stationary and, letting $\varphi(B) = 1 + \varphi_1 B + \cdots + \varphi_p B^p = \prod_{i=1}^{p}(1 - \lambda_i B)$, where $p = k$ or $p = k - 1$, this last autoregression can also be expressed as

$$w_t = W_{t-1}\varphi + u_t, \tag{6}$$

where $\varphi = (-\varphi_p, \ldots, -\varphi_1)$ and $W_t = (w_{t-p+1}, \ldots, w_t)$. Note that $W_{t-1}\varphi$ in (6) is the orthogonal projection of w_t onto the space generated by $\{w_{t-p+1}, \ldots, w_t\}$ and that, therefore, u_t in (6) is orthogonal to the random variables in $\{w_{t-p+1}, \ldots, w_t\}$.

It will be convenient to express (6) into vector AR(1) form. To this end, define $N = [0, \ldots, 0, 1]$ and

$$M = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\varphi_p & -\varphi_{p-1} & \cdots & -\varphi_1 \end{bmatrix}.$$

Then,

$$W_t = MW_{t-1} + Nu_t, \tag{7}$$

and it is well known that the eigenvalues of M coincide with the inverse roots λ_i, $i = 1, \ldots, p$, of $\varphi(B)$. In addition, $|\lambda_i| < 1$ for all $i = 1, \ldots, p$. To see this, let λ and v be an eigenvalue of M and a left eigenvector of M corresponding to λ. Then, letting $V = \mathrm{Var}(W_t)$, it is obtained from (7) that $V = MVM\ + NN\,\sigma_u^2$, where $\sigma_u^2 = \mathrm{Var}(u_t)$. Premultiplying this last expression by v and postmultiplying it by $\bar{v}$, where $\bar{v}$ denotes the complex conjugate of v, it is not difficult to verify that $(1 - |\lambda|^2)v\,V\bar{v} = v\,NN\,\bar{v}\sigma_u^2 > 0$.

In terms of the original variables y_t, the autoregression (6) can be expressed as

$$y_t = x_{t-1}\beta + u_t, \tag{8}$$

where $\beta = (-\beta_k, \ldots, -\beta_1)$ and $x_t = (y_{t-k+1}, \ldots, y_t)$. Defining $G = [0, \ldots, 0, 1]$ and

$$F = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\beta_k & -\beta_{k-1} & \cdots & -\beta_1 \end{bmatrix},$$

the autoregression (8) can be expressed as

$$x_t = Fx_{t-1} + Gu_t. \tag{9}$$

To estimate β by OLS, we can use either (8) or (9) . To see this, transpose (9) to give x_t $= x_{t-1}F\ + G\,u_t$. Then, $\hat{F}\ = G_{11}^{-1}G_{10}$, where $\hat{F}$ is the OLS estimator of F ,

$$G_{ij} = \frac{1}{m}\sum_{t=k+1}^{n} x_{t-i}x_{t-j}, \quad i, j = 1, 0,$$

and $m = n - k$. On the other hand, the OLS estimator of β obtained from (8) is $\hat{\beta} = G_{11}^{-1}\sum_{t=k+1}^{n} x_{t-1}y_t$, which coincides with the last column of $\hat{F}$.

In the following, we will work with (9) to study the convergence rate of the OLS estimator $\hat{\beta}$ of β. Let T be a non singular transformation such that premultiplying (9) by T, it is obtained that $z_t = Jz_{t-1} + G_T u_t$, where $z_t = Tx_t$, G_T=TG, $J = \mathrm{diag}(J_S, J_N)$, J_S is a matrix with eigenvalues equal to those of M in (7), and $J_N = \lambda_k = 1$ if the $\{y_t\}$ process is difference stationary, and J_N is absent in J if $\{y_t\}$ is stationary. Partitioning $z_t = (z_{st}, z_{nt})$ conforming to $J = \mathrm{diag}(J_S, J_N)$, it is clear that z_{st} is a purely stationary process and z_{nt} is a difference stationary process. To obtain the OLS estimator $\hat{J}$ of J, transpose first $z_t = Jz_{t-1} + G_T u_t$ to give $z_t = z_{t-1}J\ + G_T u_t$. Then,

$$\begin{aligned} \hat{J} &= \left(\sum_{t=k+1}^{n} z_{t-1}z_{t-1}\right)^{-1} \sum_{t=k+1}^{n} z_{t-1}z_t \\ &= J + \left(\sum_{t=k+1}^{n} z_{t-1}z_{t-1}\right)^{-1} \sum_{t=k+1}^{n} z_{t-1}G_T u_t. \end{aligned} \tag{10}$$

Let $\hat{\lambda}_i$, $i = 1, \ldots, k$, be the eigenvalues of $\hat{J}$, which are also the inverse roots of the polynomial $1 + \hat{\beta}_1 z + \cdots + \hat{\beta}_k z^k$ estimated by OLS in the autoregression (8). The following lemma shows that the eigenvalues $\hat{\lambda}_i$ converge almost surely to the eigenvalues λ_i of J and, at the same time, gives the rates of this convergence. The proof is in the Appendix.

Lemma 1. *Let $m = n - k$. Then, under the assumptions A1–A3, for each eigenvalue λ_i, $i = 1, \ldots, k$, of J there exists an eigenvalue $\hat{\lambda}_i$ of $\hat{J}$ verifying with probability one $\hat{\lambda}_i = \lambda_i + O((\ln(m)/m)^{1/2})$ if $|\lambda_i| < 1$ and $\hat{\lambda}_i = 1 + O((\ln\ln(m)/m)^{1/2})$ if $\lambda_i = 1$.*

Lemma 1 implies that the rate of convergence of $\hat{\lambda}_i$ to λ_i is faster if $\lambda_i = 1$ than if $|\lambda_i| < 1$, and this fact is exploited in the construction of the statistic $r_i = \text{Real}(\hat{\lambda}_i) - 1 + h_n$. In this statistic, h_n should satisfy $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty}(h_n/LL_n) = \infty$, where $LL_n = [\ln\ln(n)/n]^{1/2}$. Also, the rate of convergence of h_n to zero should be similar to that of $(\ln(n)/n)^{1/2}$. These considerations lead us to propose $h_n = 1/n^{\alpha_n}$, where $\{\alpha_n\}$ is an increasing sequence of positive integers such that $\lim_{n\to\infty} \alpha_n = \alpha < 0.5$. The following lemma gives the asymptotic behavior of $R_i = n^{1/2}\left[\text{Real}(\hat{\lambda}_i) - 1 + h_n\right]$ and $S_i = n^{1/2}\left[-|\text{Imag}(\hat{\lambda}_i)| + h_n\right]$, $i = 1, \ldots, k$, as $n \to \infty$. The proof is in the Appendix.

Theorem 1. *Let $m = n - k$, $L_m = [\ln(m)/m]^{1/2}$, $LL_m = [\ln\ln(m)/m]^{1/2}$ and $h_m = 1/m^{\alpha_n}$, where $\{\alpha_m\}$ is an increasing sequence of positive integers such that $\lim_{m\to\infty} \alpha_m = \alpha < 0.5$. Then, under the assumptions A1–A3, as $m \to \infty$, for each estimated eigenvalue $\hat{\lambda}_i$, $i = 1, \ldots, k$, of $\hat{J}$, the statistic $R_i = m^{1/2}\left[Real(\hat{\lambda}_i) - 1 + h_m\right]$ tends to $-\infty$ almost surely if $\hat{\lambda}_i = \lambda_i + O(L_m)$ with $|\lambda_i| < 1$, and tends to $+\infty$ almost surely if $\hat{\lambda}_i = 1 + O(LL_m)$. Also, if $\hat{\lambda}_i = 1 + O(LL_m)$, the statistic $S_i = m^{1/2}\left[-|Imag(\hat{\lambda}_i)| + h_m\right]$ tends to $+\infty$ almost surely as $m \to \infty$.*

Theorem 1 provides a simple strongly consistent rule to discriminate between I(0) and I(1) processes. Firstly, choose a lag k for the autoregression (8). Secondly, estimate β by OLS and, for each estimated $\hat{\lambda}_i$, $i = 1, \ldots, k$, construct the statistics $r_i = \text{Real}(\hat{\lambda}_i) - 1 + h_n$ and $s_i = -|\text{Imag}(\hat{\lambda}_i)| + h_n$. Finally, accept the process as I(1) if $r_i > 0$ and $s_i > 0$ for some i. Otherwise, accept it as I(0).

A possible way to select the lag k in (8) is to use model selection criteria such as AIC (Akaike, 1974a, 1976) or BIC (Schwarz, 1978). Although these criteria were in principle designed for stationary series, they have been extended to the nonstationary situation (Tsay, 1984) for autoregressive processes. However, both AIC and BIC tend to select a lag k that is too small, particularly when there is a moving average in the process close to non–invertibility. For this reason, and based on the simulation experiments presented later, we select a fixed value $k = 6$. It should be emphasized that the only condition that Lemma 1 imposes to the lag k of the autoregression in (8) is that it should be at least one.

In Theorem 1, the sequence $\{\alpha_n\}$ can in principle be arbitrary. It will be shown later in the simulations that some appropriate values for α_n are: $\alpha_{30} = 0.345$, $\alpha_{50} = 0.405$ and $\alpha_{100} = 0.499$. For $n \geq 100$, one can set $\alpha_n = 0.499$.

A justification for the previous formula for α_n is that α_n should be as close to the upper bound of 0.5 as possible for each sample size n.

5. IDENTIFICATION OF UNIT ROOTS IN STEP 2

Although one can think that the rule described in the previous section is sufficient to discriminate between I(1) and I(0) processes, the fact is that, as the simulations can show, this is not the case. If the data follow ARIMA models with moving average roots or autoregressive roots with modulus close to one, the rule fails to discriminate correctly. For this reason, the second step is introduced in the proposed procedure. The rationale behind this second step is that an ARIMA$(1,1)$ model should be able to discriminate better than a pure autoregressive model when the true model has autoregressive or moving average roots with modulus close to one. It should also work better when there is an autoregressive term that tends to cancel with some moving average term, like for example in the model $y_t - y_{t-1} = a_t - .8a_{t-1}$.

Assume we fit the model $y_t + \phi y_{t-1} = u_t + \theta u_{t-1}$ to the data and we apply the Hannan-Rissanen method described in Section 3 to estimate ϕ and θ. Letting $\hat{\phi}$ be the estimator of ϕ obtained in this way, the following lemma shows that $\hat{\phi}$ converges almost surely to minus one if the process is I(1) and that it converges almost surely to ϕ with $|\phi| < 1$ if the process is I(0). The lemma also gives the rates of this convergence. The proof is in the Appendix.

Lemma 2. *Let $m = n - 1$ and let $N = \{\ln^a(n)\}$, $a > 1$, be the order of the long autoregression in the first stage of the Hannan–Rissanen method, where $\{\ln^a(n)\}$ is the integer part of $\ln^a(n)$. Then, under the assumptions A1–A3, with probability one, $\hat{\phi} = \phi + O((\ln^{a+1}(m)/m)^{1/2})$ with $|\phi| < 1$ if the process is I(0) and $\hat{\phi} = -1 + O(LL_m)$ if the process is I(1).*

The following lemma is similar to Theorem 1. The proof is omitted because it parallels the proof of that theorem.

Theorem 2. *Let $m = n-1$, $\hat{\lambda} = -\hat{\phi}$, $L_{m,a} = (\ln^{a+1}(m)/m)^{1/2}$, and $h_m = 1/m^{\beta_m}$, where $\{\beta_m\}$ is an increasing sequence of positive integers such that $\lim_{m\to\infty} \beta_m = \beta < 0.5$. Then, under the assumptions A1–A3, as $m \to \infty$, the statistic $R = m^{1/2}\left[Real(\hat{\lambda}) - 1 + h_m\right]$ tends to $-\infty$ almost surely if $\hat{\lambda} = -\phi + O(L_{m,a})$ with $|\phi| < 1$, and tends to $+\infty$ almost surely if $\hat{\lambda} = 1 + O(LL_m)$.*

Lemma 2 and Theorem 2 provide a justification for the strongly consistent rule of Step 2 of the proposed procedure to discriminate between I(1) and I(0) processes.

The expression $|\hat{\phi} - \hat{\theta}| < c$ in Step 2, where c is a small value, is used when the autoregressive and moving–average terms in the estimated ARMA$(1,1)$ model are near to cancelation. A value of $c = 0.11$ is found to be satisfactory.

6. MEAN CORRECTIONS

In this section, we consider the possibility that the true process $\{y_t\}$ may be stationary with nonzero mean or difference–stationary with a nonzero drift term. The results of Sections 4 and 5 can be extended to cover this situation, but first assumption A1 has to be modified into

A1' The process $\{y_t\}$ is either stationary or first–difference stationary. That is, $w_t = \delta(B)y_t$ is stationary, where the polynomial $\delta(B)$ is either 1 or $\nabla = 1 - B$. Moreover, $w_t = \mu + \sum_{i\geq 0} \psi_i a_{t-i}$, where μ is a constant.

To allow for a mean or drift term, the autoregressions (8) and (9) are changed into

$$y_t = \gamma + x_{t-1}\beta + u_t, \tag{11}$$

and

$$x_t = G\gamma + Fx_{t-1} + Gu_t, \tag{12}$$

where γ is a constant. Let T be, as in Section 4, a nonsingular transformation such that $TFT^{-1} = J = \text{diag}(J_S, J_N)$, where J_S and J_N are defined as before. Then, premultiplying (12) by T, it is obtained that $z_t = G_T\gamma + Jz_{t-1} + G_T u_t$, where $z_t = Tx_t$ and $G_T{=}TG$.

It is well known that, to estimate β by OLS in (11), we can either work with this equation directly or with the equation $\tilde{y}_t = \tilde{x}_{t-1}\beta + \tilde{u}_t$, where $\tilde{y}_t = y_t - \bar{y}$, $\bar{y} = \sum_{t=k+1}^{n} y_t/m$, $\tilde{x}_t = x_t - \bar{x}$, $\bar{x} = \sum_{t=k}^{n-1} x_t/m$, $\tilde{u}_t = u_t - \bar{u}$, and $\bar{u} = \sum_{t=k+1}^{n} u_t/m$. In the following, we will work with the de–meaned data. If we incorporate mean adjustments into the autoregression $z_t = G_T\gamma + Jz_{t-1} + G_T u_t$, we get $z_t - \bar{z}_0 = J(z_{t-1} - \bar{z}_1) + G_T(u_t - \bar{u})$, where $\bar{z}_i = \sum_{t=k+1}^{n} z_{t-i}/m$, $i = 0, 1$, and $\bar{u} = \sum_{t=k+1}^{n} u_t/m$, and the OLS estimator $\hat{J}$ of J is

$$\begin{aligned}
\hat{J} &= \left[\sum_{t=k+1}^{n} (z_{t-1} - \bar{z}_1)(z_{t-1} - \bar{z}_1)\right]^{-1} \sum_{t=k+1}^{n} (z_{t-1} - \bar{z}_1)(z_t - \bar{z}_0) \\
&= J + \left[\sum_{t=k+1}^{n} (z_{t-1} - \bar{z}_1)(z_{t-1} - \bar{z}_1)\right]^{-1} \sum_{t=k+1}^{n} (z_{t-1} - \bar{z}_1) G_T(u_t - \bar{u}).
\end{aligned} \tag{13}$$

Let, as in Section 4, $\hat{\lambda}_i$, $i = 1, \ldots, k$, be the eigenvalues of $\hat{J}$. The following lemma shows that, in the present situation, the eigenvalues $\hat{\lambda}_i$ also converge almost surely to the eigenvalues λ_i of J, and with the same convergence rate. The proof is in the Appendix.

Lemma 3. *Let $m = n - k$. Then, under the assumptions A1', A2 and A3, for each eigenvalue λ_i, $i = 1, \ldots, k$, of J there exists an eigenvalue $\hat{\lambda}_i$ of $\hat{J}$ verifying with probability one $\hat{\lambda}_i = \lambda_i + O(L_m)$ if $|\lambda_i| < 1$, and $\hat{\lambda}_i = 1 + O(LL_m)$ if $\lambda_i = 1$.*

The previous result justifies the extension of the strongly consistent rule of Step 1 to discriminate between I(0) and I(1) processes to the present situation. The only change that has to be made is to replace in that rule the raw data by the de–meaned data.

In a similar way, we can extend the strongly consistent rule of Step 2 to the nonzero mean or drift case considered in this section. Again, the rule is the same except that the raw data are replaced by the de–meaned data. The proof of this result is a generalization of the proof of the result of the previous section along the lines of this section and is omitted.

7. SIMULATION EVIDENCE

Table 1: Nonstationary Models for the Simulation Experiment

	p	d	q	m	r_1	r_2	r_3	s_1	s_2	s_3
N_1	0	1	1	0	.0	.0	.0	-.9	.0	.0
N_2	0	1	1	0	.0	.0	.0	-.5	.0	.0
N_3	0	1	0	0	.0	.0	.0	.0	.0	.0
N_4	0	1	1	0	.0	.0	.0	.5	.0	.0
N_5	0	1	1	0	.0	.0	.0	.8	.0	.0
N_6	0	1	1	0	.0	.0	.0	.85	.0	.0
N_7	0	1	1	0	.0	.0	.0	.9	.0	.0
N_8	0	1	1	0	.0	.0	.0	.95	.0	.0
N_9	1	1	1	0	.8	.0	.0	.4	.0	.0
N_{10}	2	1	2	0	.7	.6	.0	.3	.4	.0
N_{11}	3	1	3	0	.7	.6	.5	.3	.2	.25
N_{12}	0	1	1	1	.0	.0	.0	-.9	.0	.0
N_{13}	0	1	1	1	.0	.0	.0	-.5	.0	.0
N_{14}	0	1	0	1	.0	.0	.0	.0	.0	.0
N_{15}	0	1	1	1	.0	.0	.0	.5	.0	.0
N_{16}	0	1	1	1	.0	.0	.0	.8	.0	.0
N_{17}	0	1	1	1	.0	.0	.0	.85	.0	.0
N_{18}	0	1	1	1	.0	.0	.0	.9	.0	.0
N_{19}	0	1	1	1	.0	.0	.0	.95	.0	.0
N_{20}	1	1	1	1	.8	.0	.0	.4	.0	.0
N_{21}	2	1	2	1	.7	.6	.0	.3	.4	.0
N_{22}	3	1	3	1	.7	.6	.5	.3	.2	.25

In this section, we will provide some indication of the performance of the proposed criterion based on the rate of convergence (CRC) of the least squares estimators. As mentioned earlier, a selection of $k = 6$ in the autoregression (8) of Step 1 and a specification of $c = 0.11$ for the cancelation of terms in Step 2 seems to work well in the simulations for all models usually found in practice.

Given a sample size n, if $\{\alpha_n\}$ and $\{\beta_n\}$ denote the sequences corresponding to the first and the second step of the proposed procedure, the values selected should be as close to the upper bound of 0.5 as possible. After some experimentation, some of the values we have found appropriate for α_n and β_n are $\alpha_{30} = 0.345$, $\alpha_{50} = 0.405$, $\alpha_{100} = 0.499$, $\alpha_{150} = 0.499$, $\alpha_{200} = .499$, $\alpha_{500} = .499$, $\beta_{30} = 0.265$, $\beta_{50} = 0.311$, $\beta_{100} = 0.407$, $\beta_{150} = 0.4703$, $\beta_{200} = 0.499$ and $\beta_{500} = 0.499$.

In the simulation experiment, we start by selecting a wide range of ARIMA(p, d, q) models, where $0 \leq p \leq 3$, $0 \leq d \leq 1$ and $0 \leq q \leq 1$. The models we use are listed in Table 1. The symbols r_i and s_i, $i = 1, 2, 3$, refer to the inverse roots of the autoregressive and moving average polynomials. Thus, if $p = 3$, the autoregressive polynomial is $1+\phi_1 B+\phi_2 B^2+\phi_3 B^3 = (1 - r_1 B)(1 - r_2 B)(1 - r_3 B)$ and, if $q = 1$, the moving average polynomial is $1 + \theta B = 1 - s_1 B$. The symbol m refers to the mean of the process. If $m = 0$, there is no mean term and if $m = 1$, there is a nonzero mean if the process is stationary and a nonzero drift term if it is nonstationary. Each model has the symbol S_i or N_i for future reference, where S_i denotes stationary model number i and N_i denotes nonstationary model number i, $i = 1, \ldots, 22$.

Table 2: Stationary Models for the Simulation Experiment

	p	d	q	m	r_1	r_2	r_3	s_1	s_2	s_3
S_1	1	0	1	0	-.9	.0	.0	-.3	.0	.0
S_2	1	0	1	0	-.5	.0	.0	-.2	.0	.0
S_3	0	0	0	0	.0	.0	.0	.0	.0	.0
S_4	1	0	1	0	.6	.0	.0	.2	.0	.0
S_5	1	0	0	0	.8	.0	.0	.0	.0	.0
S_6	1	0	0	0	.85	.0	.0	.0	.0	.0
S_7	1	0	0	0	.9	.0	.0	.0	.0	.0
S_8	0	0	2	0	.0	.0	.0	.6	.4	.0
S_9	2	0	2	0	.7	.6	.0	.4	.3	.0
S_{10}	2	0	0	0	.4	.6	.0	.0	.0	.0
S_{11}	3	0	3	0	.7	.6	.7	.9	.3	.2
S_{12}	1	0	1	1	-.9	.0	.0	-.3	.0	.0
S_{13}	1	0	1	1	-.5	.0	.0	-.2	.0	.0
S_{14}	0	0	0	1	.0	.0	.0	.0	.0	.0
S_{15}	1	0	1	1	.6	.0	.0	.2	.0	.0
S_{16}	1	0	0	1	.8	.0	.0	.0	.0	.0
S_{17}	1	0	0	1	.85	.0	.0	.0	.0	.0
S_{18}	1	0	0	1	.9	.0	.0	.0	.0	.0
S_{19}	0	0	2	1	.0	.0	.0	.6	.4	.0
S_{20}	2	0	2	1	.7	.6	.0	.4	.3	.0
S_{21}	2	0	0	1	.4	.6	.0	.0	.0	.0
S_{22}	3	0	3	1	.7	.6	.7	.9	.3	.2

Each ARIMA(p,d,q) model, $(1+\phi_1 B+\cdots+\phi_p B^p)(\nabla^d y_t-\mu)=(1+\theta_1 B+\cdots+\theta_q B^q)\epsilon_t$, where $\{\epsilon_t\}$ is an i.i.d. sequence of $N(0,\sigma^2)$ random variables, is generated using first the state space representation

$$y_t = Z\alpha_t \tag{14}$$

$$\alpha_{t+1} = T\alpha_t + H\epsilon_t, \tag{15}$$

and adding later the nonzero mean or drift term if necessary. The state space representation (14) and (15) is as extension to nonstationary models of the one originally proposed by Akaike (1974b) for stationary models. More specifically,

$$T=\begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\phi_r^* & -\phi_{r-1}^* & \cdots & -\phi_1^* \end{bmatrix},$$

$Z=(1,0,\ldots,0)$, $H=(1,\psi_1,\ldots,\psi_{r-1})$, $\alpha_t=(y_t,y_{t+1|t},\ldots,y_{t+r-1|t})$, $r=\max\{p+d,q+1\}$, $y_{t+i|t}=y_t-\epsilon_t-\psi_1\epsilon_{t-1}-\cdots-\psi_{i-1}\epsilon_{t-i+1}$, $i=1,\ldots,r-1$, $\sum_{i=0}^{\infty}\psi_i B^i=(1+\theta_1 B+\cdots+\theta_q B^q)/\phi^*(B)$, $\phi^*(B)=(1+\phi_1 B+\cdots+\phi_p B^p)\nabla^d$, and $\phi_i^*=0$ for $i>p+d$. It is not difficult to verify that the elements of the state vector are the series and its first $r-1$ periods ahead function and that this state space representation is correct.

If $d=0$, the initial state vector α_1 is $\alpha_1=U=(u_1,u_{2|1},\ldots,u_{r|1})$, where $u_t=y_t$, $u_{i|1}=\mathrm{E}(u_i|u_s : s\le 1)$ and E(.|.) denotes conditional expectation. Thus, the mean

Table 3: Sample Size n=30, $\alpha_{30} = .345$, $\beta_{30} = .265$

	Nonstationary Models				Stationary Models		
Model	ADF	DFGLS	CRC	Model	ADF	DFGLS	CRC
N_1	87.82	95.96	95.04	S_1	52.47	54.49	83.30
N_2	90.98	95.29	94.91	S_2	51.22	84.32	76.28
N_3	92.03	95.48	92.10	S_3	36.40	80.53	65.88
N_4	91.24	97.24	81.31	S_4	26.22	67.96	56.76
N_5	76.11	99.86	53.44	S_5	14.43	42.77	27.57
N_6	70.09	99.98	46.35	S_6	12.49	35.33	21.01
N_7	62.68	100.00	39.78	S_7	10.63	27.19	15.22
N_8	56.05	100.00	34.94	S_8	96.87	88.02	95.91
N_9	90.88	91.55	99.33	S_9	16.49	50.04	35.36
N_{10}	89.29	92.50	99.68	S_{10}	20.39	57.17	32.76
N_{11}	87.25	89.83	99.79	S_{11}	27.00	73.51	55.46
N_{12}	98.60	98.09	99.92	S_{12}	52.22	0.50	83.20
N_{13}	98.47	97.91	99.95	S_{13}	51.11	0.60	75.89
N_{14}	98.86	97.97	99.95	S_{14}	36.39	0.66	66.03
N_{15}	99.40	97.59	99.93	S_{15}	26.35	0.70	57.10
N_{16}	99.89	97.32	99.89	S_{16}	14.50	0.95	27.30
N_{17}	99.93	97.27	99.89	S_{17}	12.20	1.27	20.80
N_{18}	99.91	97.24	99.88	S_{18}	10.34	1.77	15.01
N_{19}	99.92	97.23	99.89	S_{19}	96.54	0.58	95.88
N_{20}	96.77	97.69	99.98	S_{20}	16.35	0.85	35.37
N_{21}	97.17	97.31	99.99	S_{21}	20.24	1.08	32.61
N_{22}	95.54	95.86	99.98	S_{22}	26.82	0.83	54.78

of α_1 is zero and its covariance matrix can be easily obtained. If $d = 1$, α_1 is defined as $\alpha_1 = c_1 + A_1\delta$, where, $A_1 = (1, \ldots, 1)$, $\delta = y_0$ is a diffuse random variable, and the distribution of c_1 is well defined. In fact, it is easy to verify that for $t > 0$, $y_t = y_0 + \sum_{i=0}^{t-1} u_{t-i}$, where $u_t = \nabla y_t$. Then, A_1 is as asserted and $c_1 = \Xi U$, where Ξ is a lower triangular $r \times r$ matrix with ones in and under the main diagonal, $U = (u_1, u_{2|1}, \ldots, u_{r|1})$ and $u_{i|1} = \mathrm{E}(u_i | u_s : s \leq 1)$.

Note that in both the stationary and the nonstationary case the distribution of U depends on that of u_t, which follows an ARMA(p, q) model. Thus, to generate α_1 for the simulations, we first obtain the matrix $V = \mathrm{Cov}(U)$ and compute its Cholesky decomposition $V = LL$. Then, we generate U as $U = La$, where a is an r dimensional vector of i.i.d $N(0, \sigma^2)$ random variables. If $d > 0$, the variable δ is generated as a $N(0, 1)$ variable independent of a plus a constant equal to 100.

Finally, the series $\{y_t\}$ is generated using the state space equations (14) and (15), where the sequence $\{\epsilon_t\}$ is generated as an i.i.d. sequence of $N(0, \sigma^2)$ variables independent of a and δ. If a nonzero mean or drift term is present in the model, it is added to the series generated using the previous procedure.

A set of programs that implement the previous procedure has been written by the author in MATLAB to generate time series that follow an ARIMA(p, d, q) model. In addition, a program has also been written by the author to apply the proposed criterion to decide between I(1) and I(0) processes. These programs are available upon request. In

Table 4: Sample Size n=50, $\alpha_{50} = .405$, $\beta_{50} = .311$

	Nonstationary Models				Stationary Models		
Model	ADF	DFGLS	CRC	Model	ADF	DFGLS	CRC
N_1	94.97	93.85	96.68	S_1	75.01	55.66	96.06
N_2	93.70	95.18	96.84	S_2	72.34	93.23	92.62
N_3	93.54	95.70	95.30	S_3	56.87	93.26	84.89
N_4	93.14	97.07	89.02	S_4	41.72	85.46	72.65
N_5	78.12	99.84	57.32	S_5	19.75	58.08	37.82
N_6	68.72	99.95	43.65	S_6	15.61	47.16	25.46
N_7	55.33	100.00	29.03	S_7	11.50	34.23	15.20
N_8	39.50	100.00	18.35	S_8	99.63	87.03	99.89
N_9	92.50	93.08	99.88	S_9	25.06	68.27	47.11
N_{10}	91.54	93.29	99.92	S_{10}	30.62	75.31	54.42
N_{11}	91.25	91.24	99.97	S_{11}	49.93	92.88	82.92
N_{12}	98.33	97.93	99.98	S_{12}	75.23	0.44	96.00
N_{13}	99.89	98.13	99.99	S_{13}	72.38	0.47	92.53
N_{14}	99.18	98.14	99.97	S_{14}	56.72	0.66	84.75
N_{15}	99.48	98.03	99.98	S_{15}	41.16	0.68	72.80
N_{16}	99.95	97.93	99.97	S_{16}	19.58	0.75	37.69
N_{17}	99.94	97.92	99.95	S_{17}	15.31	0.79	25.49
N_{18}	99.95	97.89	99.94	S_{18}	11.35	0.90	15.25
N_{19}	99.96	97.90	99.95	S_{19}	99.72	0.45	99.88
N_{20}	97.65	98.56	99.95	S_{20}	24.88	0.79	47.13
N_{21}	97.98	98.41	100.00	S_{21}	30.15	1.02	54.47
N_{22}	97.39	98.30	100.00	S_{22}	50.02	0.86	82.66

the previous programs, the MATLAB function **adfreg** by Ludwig Kanzler has been used to apply the ADF test. This routine was modified to compute the critical values of the GLS-de–trended ADF tests.

In Tables 3 to 8 the performance of CRC is shown when this criterion is applied to 20,000 replications of series of length 30, 50, 100, 150, 200 and 500 that follow the models of Tables 1 and 2. The tables also show the performance of the ADF and DFGLS tests when they are applied to the same series.

Given that the methodology used by CRC differs from that of the ADF and DFGLS tests, some explanation is necessary as to how the comparison is made. For CRC, a mean term is always included in the regressions and in each table the percentages of correctly identified models are given. For the ADF and DFGLS tests, the .05 level to test H_0 against H_1 is used, where H_0 and H_1 are defined in (2) for a model (1) in which $x_t = \{1\}$ and, following Poskitt (2000, p. 81), the number of lags p is selected for each sample size n according to the formula $p = \{\ln(n) + .5\}$. In each table the percentages of correctly identified models are given when the ADF and DFGLS tests are applied in this way. Thus, the performance of these tests is judged by how close the percentages are to .95, whereas that of CRC is judged by how close the percentages are to 1.

In Table 3, CRC is applied to 20,000 replications of series of length 30 that follow the models of Tables 1 and 2. The specification $\alpha_{30} = .345$ and $\beta_{30} = .265$ is used. It is seen that the results of CRC for nonstationary models are better than those obtained with

Table 5: Sample Size n=100, $\alpha_{100} = .499$, $\beta_{100} = .407$

	Nonstationary Models				Stationary Models		
Model	ADF	DFGLS	CRC	Model	ADF	DFGLS	CRC
N_1	92.05	96.61	96.62	S_1	99.01	71.26	100.00
N_2	93.78	96.00	96.72	S_2	98.83	99.81	96.60
N_3	94.06	96.28	96.12	S_3	95.09	99.91	96.62
N_4	93.81	97.09	93.47	S_4	86.13	99.55	94.67
N_5	77.30	99.75	71.02	S_5	51.61	91.59	83.88
N_6	63.39	99.99	51.12	S_6	37.82	82.37	64.54
N_7	39.75	100.00	24.89	S_7	24.00	64.75	36.39
N_8	11.81	100.00	5.50	S_8	100.00	96.16	100.00
N_9	93.53	94.27	99.96	S_9	64.32	96.50	90.03
N_{10}	93.17	94.07	99.97	S_{10}	73.32	98.07	97.88
N_{11}	92.96	91.26	99.97	S_{11}	95.84	99.94	99.91
N_{12}	99.45	98.87	99.97	S_{12}	98.90	0.38	100.00
N_{13}	99.33	98.79	99.99	S_{13}	98.59	0.46	99.65
N_{14}	99.45	98.61	99.99	S_{14}	95.20	0.82	96.73
N_{15}	99.65	98.76	99.99	S_{15}	85.77	0.80	94.58
N_{16}	99.94	98.71	99.95	S_{16}	51.22	1.22	83.94
N_{17}	99.97	98.70	99.94	S_{17}	38.15	1.19	64.64
N_{18}	99.95	98.67	99.93	S_{18}	24.28	1.09	36.65
N_{19}	99.94	98.67	99.93	S_{19}	100.00	0.36	100.00
N_{20}	98.94	99.18	100.00	S_{20}	64.56	1.08	90.08
N_{21}	98.98	99.09	100.00	S_{21}	72.81	1.43	97.95
N_{22}	98.54	99.09	100.00	S_{22}	95.84	1.19	99.92

ADF except for models N_4 to N_8. For stationary models, CRC detects significantly more models than ADF, except for models S_8 and S_{19} where the differences are small.

As regards ADF and DFGLS, the results confirm the low power of ADF in all cases and show an increase of power of DFGLS with respect to ADF when the test is applied to stationary models with zero mean, and the almost complete lack of power of DFGLS when this test is applied to stationary models with a nonzero mean. The results confirm also that DFGLS does a much better job than ADF with nonstationary models with a moving average root close to the unit circle.

The results of CRC, together with those of the ADF and DFGLS tests, for sample sizes of $n = 50, 100, 150, 200, 500$ can be seen in Tables 4 to 8.

In Table 4, the specification $\alpha_{50} = .405$ and $\beta_{50} = .311$ is used. It is seen that CRC detects more nonstationary models than ADF, except for models N_4 to N_8. At the same time, CRC detects more stationary models than ADF for all models. With respect to DFGLS, it is to be noticed the complete lack of power for models S_{12} to S_{22}, which are the stationary models that include a nonzero mean term.

In Table 5, the specification $\alpha_{100} = .499$ and $\beta_{100} = .407$ is used. CRC detects more nonstationary models than ADF, except for models N_4, where the difference is very small, and N_5 to N_8. As in Table 4, CRC detects more stationary models than ADF, with the exception of model S_2. The difference for this last model however is not big. Again, it is to be noted the complete lack of power of DFGLS for models S_{12} to S_{22}.

Table 6: Sample Size n=150, $\alpha_{150} = .499$, $\beta_{150} = .4703$

	Nonstationary Models				Stationary Models		
Model	ADF	DFGLS	CRC	Model	ADF	DFGLS	CRC
N_1	92.28	97.03	95.69	S_1	100.00	86.06	100.00
N_2	94.00	96.47	95.94	S_2	100.00	100.00	99.99
N_3	94.57	96.31	95.48	S_3	99.95	100.00	99.43
N_4	93.86	97.32	93.57	S_4	99.28	100.00	99.69
N_5	74.38	99.86	83.75	S_5	82.93	99.37	99.68
N_6	57.05	100.00	64.71	S_6	67.84	97.45	97.74
N_7	29.04	100.00	32.58	S_7	44.16	87.53	84.87
N_8	3.18	100.00	5.38	S_8	100.00	99.67	100.00
N_9	94.08	94.61	99.95	S_9	92.66	99.88	99.96
N_{10}	93.85	94.24	99.97	S_{10}	96.28	99.94	99.99
N_{11}	93.83	91.31	99.94	S_{11}	99.96	100.00	100.00
N_{12}	99.68	99.03	99.98	S_{12}	100.00	0.58	100.00
N_{13}	99.59	98.89	99.99	S_{13}	100.00	0.63	100.00
N_{14}	99.60	98.84	99.98	S_{14}	99.90	1.01	99.37
N_{15}	99.71	98.86	99.99	S_{15}	99.24	1.02	99.73
N_{16}	99.97	98.80	99.98	S_{16}	82.86	1.59	99.66
N_{17}	99.96	98.80	99.98	S_{17}	67.96	1.61	97.81
N_{18}	99.96	98.78	99.97	S_{18}	44.36	1.64	85.27
N_{19}	99.96	98.77	99.96	S_{19}	100.00	0.46	100.00
N_{20}	99.22	99.32	100.00	S_{20}	92.77	1.42	99.98
N_{21}	99.44	99.18	100.00	S_{21}	96.20	1.77	99.99
N_{22}	99.09	99.46	100.00	S_{22}	99.97	1.44	100.00

In Table 6, the specification $\alpha_{150} = .499$ and $\beta_{150} = .4703$ is used. CRC detects more nonstationary models than ADF, except for model N_4, where the difference is very small. As regards stationary models, CRC detects less models than ADF only in models S_2 and S_3, where the differences are negligible. The differences of CRC with respect to ADF are more pronounced in the models with an autoregressive root close to one, models S_5 to S_7.

In Table 7, the specification $\alpha_{200} = .499$ and $\beta_{200} = .499$ is used. CRC detects more nonstationary models than ADF in all cases. For models N_5 to N_8, where the moving average root gets close to one, the differences between CRC and ADF are substantial. For stationary models, CRC detects more models than ADF except for models S_2, S_3 and S_{13} to S_{15} where the differences are negligible. For models with an autoregressive root close to one, models S_5 to S_7, the differences between CRC and ADF are substantial, all in favor of CRC.

Finally, in Table 8 the specification $\alpha_{500} = .499$ and $\beta_{500} = .4703$ is used. CRC detects more nonstationary models than ADF in all cases. For models N_6 to N_8 the differences are now substantial. It is seen that, for stationary models, both CRC and ADF reach almost a hundred per cent of correctly identified models.

8. CONCLUSION

The experimental results presented in this article indicate that the proposed CRC criterion can be used to decide whether a time series is stationary or first–difference stationary.

Table 7: Sample Size n=200, $\alpha_{200} = .499$, $\beta_{200} = .499$

	Nonstationary Models				Stationary Models		
Model	ADF	DFGLS	CRC	Model	ADF	DFGLS	CRC
N_1	92.33	96.69	95.81	S_1	100.00	94.23	100.00
N_2	94.28	96.22	96.25	S_2	100.00	100.00	99.61
N_3	94.43	96.40	95.80	S_3	100.00	100.00	99.08
N_4	94.16	97.25	94.28	S_4	99.99	100.00	99.69
N_5	72.62	99.88	90.13	S_5	96.71	99.98	99.68
N_6	53.85	99.99	73.45	S_6	88.63	99.78	97.74
N_7	25.14	100.00	39.16	S_7	66.37	97.20	84.87
N_8	1.51	100.00	6.39	S_8	100.00	99.99	100.00
N_9	94.27	94.28	99.96	S_9	99.40	100.00	99.96
N_{10}	94.00	94.22	99.98	S_{10}	99.81	100.00	99.99
N_{11}	94.43	91.02	99.94	S_{11}	100.00	100.00	100.00
N_{12}	99.72	99.19	99.97	S_{12}	100.00	0.65	100.00
N_{13}	99.58	99.12	99.98	S_{13}	100.00	0.77	99.60
N_{14}	99.60	98.98	99.99	S_{14}	100.00	1.09	99.16
N_{15}	99.73	99.02	100.00	S_{15}	100.00	1.22	99.72
N_{16}	99.99	98.96	100.00	S_{16}	96.93	1.78	99.66
N_{17}	99.97	98.96	100.00	S_{17}	88.97	1.89	97.81
N_{18}	99.95	98.96	99.98	S_{18}	66.51	1.92	85.27
N_{19}	99.96	98.96	99.96	S_{19}	100.00	0.62	100.00
N_{20}	99.32	99.37	100.00	S_{20}	99.43	1.77	99.98
N_{21}	99.48	99.44	100.00	S_{21}	99.82	2.09	99.99
N_{22}	99.25	99.59	100.00	S_{22}	100.00	1.81	100.00

The CRC criterion is easy to implement and performs better than ADF for nonstationary models. The improvement is substantial for models with a moving average unit root close to one. With stationary models, CRC performs in general better than ADF, with considerable improvements in the case of autoregressive roots close to one. For the stationary models for which CRC performs worse than ADF the differences seem to be small and most of them are not likely to be found in practice. As regards DFGLS, it usually works better than CRC and ADF, but its complete lack of power with stationary models which include a nonzero mean term makes it virtually useless.

One advantage of the proposed procedure with respect to the Dickey–Fuller tests is that it is easy to implement and that it is not necessary to use tables based on simulations. Another advantage is that the two steps of the procedure are always applied in the same way, regardless of the distribution of the true model. This contrasts with the Dickey–Fuller test, where the test statistic depends on the assumptions about the true model.

Finally, the procedure can be extended to nonstationary seasonal models without requiring much effort. The author has confirmed this fact using simulations, not shown in this paper for lack of space.

ACKNOWLEDGMENTS

The MATLAB function **adfreg** by Ludwig Kanzler has been used in the simulations

Table 8: Sample Size n=500, $\alpha_{500} = .499$, $\beta_{500} = .499$

	Nonstationary Models				Stationary Models		
Model	ADF	DFGLS	CRC	Model	ADF	DFGLS	CRC
N_1	95.83	94.08	99.40	S_1	100.00	99.97	100.00
N_2	94.87	95.46	99.49	S_2	100.00	100.00	100.00
N_3	94.68	96.48	99.47	S_3	100.00	100.00	100.00
N_4	94.54	96.95	99.18	S_4	100.00	100.00	100.00
N_5	77.78	99.68	99.27	S_5	100.00	100.00	100.00
N_6	59.27	99.96	92.81	S_6	100.00	100.00	100.00
N_7	27.92	100.00	50.35	S_7	99.97	100.00	99.88
N_8	1.30	100.00	6.76	S_8	100.00	100.00	100.00
N_9	94.70	93.08	100.00	S_9	100.00	100.00	100.00
N_{10}	94.80	92.80	100.00	S_{10}	100.00	100.00	100.00
N_{11}	94.70	92.18	99.99	S_{11}	100.00	100.00	100.00
N_{12}	99.54	99.60	100.00	S_{12}	100.00	0.86	100.00
N_{13}	99.70	99.56	100.00	S_{13}	100.00	1.01	100.00
N_{14}	99.69	99.51	100.00	S_{14}	100.00	1.50	99.99
N_{15}	99.76	99.49	100.00	S_{15}	100.00	1.93	100.00
N_{16}	99.99	99.44	100.00	S_{16}	100.00	2.72	100.00
N_{17}	100.00	99.44	100.00	S_{17}	100.00	2.93	100.00
N_{18}	99.99	99.44	99.99	S_{18}	99.96	3.20	99.91
N_{19}	99.99	99.42	99.99	S_{19}	100.00	0.68	100.00
N_{20}	99.60	99.74	100.00	S_{20}	100.00	2.47	100.00
N_{21}	99.57	99.81	100.00	S_{21}	100.00	3.25	100.00
N_{22}	99.44	99.83	100.00	S_{22}	100.00	2.43	100.00

to apply the ADF test. This routine was modified by the author to compute the critical values of the GLS-de–trended ADF tests. The author wants to thank Esther Ruiz of Universidad Carlos III de Madrid for many valuable comments.

APPENDIX: PROOFS

Proof of Lemma 1

Let $L_m = (\ln(m)/m)^{1/2}$ and $LL_m = (\ln\ln(m)/m)^{1/2}$. Suppose the process $\{y_t\}$ is difference stationary and define the normalization matrix $N_m = \text{diag}(m^{-1/2}I_{k-1}, m^{-1})$. Then, (10) can be expressed as

$$\hat{J} = J + N_m \left[\sum_{t=k+1}^{n} N_m z_{t-1} z_{t-1}' N_m\right]^{-1} \sum_{t=k+1}^{n} N_m z_{t-1} v_t',$$

where $v_t = G_T' u_t$. By Lemma A.1 of Poskitt (2000), $m^{-2}\sum_{t=k+1}^{n} z_{nt-1} z_{nt-1}' = O_p(1)$, $m^{-1}\sum_{t=k+1}^{n} z_{st-1} z_{st-1}' = \Gamma(0) + O(L_m)$, and $m^{-3/2}\sum_{t=k+1}^{n} z_{st-1} z_{nt-1}' = O(\ln\ln(m)^{1/2}) = O_p(m^{-1/2})$, where $\Gamma(0) = \text{Var}(z_{st})$. Thus, letting $F_{11} = \sum_{t=k+1}^{n} N_m z_{t-1} z_{t-1}' N_m$, we can write

$$F_{11} = \begin{bmatrix} \Gamma(0) + O(L_m) & O_p(m^{-1/2}) \\ O_p(m^{-1/2}) & O_p(1) \end{bmatrix}.$$

Given that, as noted by Poskitt (2000) p. 87, $m^{-2}\sum_{t=k+1}^{n} z_{nt-1} z_{nt-1}' \xrightarrow{d} W > 0$ almost surely, where W is a functional of standard Brownian motion, the previous expression for F_{11} implies

$$F_{11} \xrightarrow{d} \begin{bmatrix} \Gamma(0) & 0 \\ 0 & W \end{bmatrix} > 0 \quad \text{almost surely.}$$

Using these results and an argument similar to that used in the proof of Lemma 1.1 of Poskitt (2000), we can show that

$$F_{11}^{-1} = \begin{bmatrix} \Gamma(0)^{-1} + O(L_m) & O_p(m^{-1/2}) \\ O_p(m^{-1/2}) & O_p(1) \end{bmatrix}.$$

By Lemma A.1 of Poskitt (2000), $m^{-1}\sum_{t=k+1}^{n} z_{st-1} v_t' = O(L_m)$ and $m^{-3/2}\sum_{t=k+1}^{n} z_{nt-1} v_t' = O(\ln\ln(m)^{1/2})$, where we have used the fact that u_t is uncorrelated with z_{st-1}. Then, letting $G_{10} = \sum_{t=k+1}^{n} N_m z_{t-1} v_t'$, the following expression is obtained

$$G_{10} = \begin{bmatrix} m^{1/2} O(L_m) \\ m^{1/2} O((\ln\ln(m))^{1/2}) \end{bmatrix}.$$

From this, it follows that

$$N_m F_{11}^{-1} G_{10} = \begin{bmatrix} O(L_m) \\ O(LL_m) \end{bmatrix}$$

and

$$\hat{J}' = J' + \begin{bmatrix} O(L_m) \\ O(LL_m) \end{bmatrix} = \begin{bmatrix} J_S' + O(L_m) & O(L_m) \\ O(LL_m) & 1 + O(LL_m) \end{bmatrix}.$$

Consider the eigenvalues of $\hat{J}'$. Using the Laplace expansion of a determinant, we can write $\det(\hat{J}' - \lambda I_k)$ as

$$\det(J_S' - \lambda I_{k-1} + O(L_m))\det(1 - \lambda + O(LL_m)) + O(LL_m).$$

Let $\tilde{\lambda}_k$ verify $\det(1 - \lambda + O(LL_m)) = 0$. Then, $\tilde{\lambda}_k = 1 + O(LL_m)$ and there is an eigenvalue $\hat{\lambda}_k$ of $\hat{J}'$ such that $\hat{\lambda}_k = \tilde{\lambda}_k + O(LL_m) = 1 + O(LL_m)$. To prove this last statement, evaluate $\det(J_S' - \lambda I_{k-1} + O(L_m))$ at $\lambda = \tilde{\lambda}_k$. This gives $\det(J_S' - I_{k-1} + O(L_m))$ which, because unity

is not an eigenvalue of J_S', converges to a nonzero constant almost surely. Thus, there exists an eigenvalue $\hat{\lambda}_k$ of $\hat{J}'$ as asserted. Suppose now that $\tilde{\lambda}_i$ is a solution of $\det(J_S' - \lambda I_{k-1} + O(L_m)) = 0$. Because eigenvalues are holomorphic functions of the matrix elements, $\tilde{\lambda}_i = \lambda_i + O(L_m)$, where λ_i verifies $\det(J_S' - \lambda I_{k-1}) = 0$ and $|\lambda_i| < 1$. In addition, there is an eigenvalue $\hat{\lambda}_k$ of $\hat{J}'$ such that $\hat{\lambda}_k = \tilde{\lambda}_k + O(L_m)$. This can be proved as before, noting that evaluating $\det(1 - \lambda + O(LL_m))$ at $\lambda = \tilde{\lambda}_k$ gives an expression that converges to a nonzero constant with probability one. Thus, $\hat{\lambda}_i$ is an eigenvalue of $\hat{J}'$ that converges to λ_i. This completes the proof.

Proof of Theorem 1

Let $\hat{\lambda}$ be an eigenvalue of $\hat{J}$ such that $|\hat{\lambda}| < 1$. Then, by Lemma 1, $\hat{\lambda} = \lambda + O(L_m)$, where λ is a root of J with $|\lambda| < 1$ and $L_m = (\ln(m)/m)^{1/2}$, and the statistic $R = m^{1/2}[\text{Real}(\hat{\lambda}) - 1 + h_m]$ satisfies

$$R = m^{1/2}[\text{Real}(\lambda) - 1 + O(L_m) + h_m] \to -\infty.$$

Let now $\hat{\lambda}$ be an eigenvalue of $\hat{J}$ such that $\hat{\lambda} = 1$ and let $LL_m = (\ln\ln(m)/m)^{1/2}$. Then, by Lemma 1, $\hat{\lambda} = 1 + O(LL_m)$ and

$$\begin{aligned} R &= m^{1/2}[\text{Real}(\hat{\lambda}) - 1 + h_m] \\ &= m^{1/2}\left[O(LL_m) + h_m\right] \\ &= m^{1/2-\alpha_m}\left[O(LL_m/h_m) + 1\right] \to +\infty. \end{aligned}$$

The rest of the theorem can be proved similarly.

Proof of Lemma 2

Let $\{a_t\}$ be the sequence of innovations defined in Assumptions A1–A3 and assume the regression $y_t = -\phi y_{t-1} + \theta a_{t-1} + u_t$, where $\phi = -1$ if the process $\{y_t\}$ is I(1). Note that u_t is orthogonal to a_{t-1} in the regression. Since a_t is not observed, it is replaced in the second stage of the Hannan–Rissanen method by $\hat{a}_t$, obtained according to (3) with $N = \{\ln^a(n)\}$, $a > 1$. Assume $\{y_t\}$ is I(1), let $m = n - 1$ and define $x_t = (y_t, \hat{a}_t)'$, $\beta = (1, \theta)'$ and the normalization matrix $N_m = \text{diag}(m^{-1}, m^{-1/2})$. Then, the OLS estimator $\tilde{\beta} = (-\tilde{\phi}, \tilde{\theta})'$ of β in the second stage of the Hannan–Rissanen method is

$$\begin{aligned} \tilde{\beta} &= N_m \left[\sum_{t=2}^{n} N_m x_{t-1} x_{t-1}' N_m\right]^{-1} \sum_{t=2}^{n} N_m x_{t-1} y_t \\ &= \beta + N_m \left[\sum_{t=2}^{n} N_m x_{t-1} x_{t-1}' N_m\right]^{-1} \sum_{t=2}^{n} N_m x_{t-1}[\theta(a_{t-1} - \hat{a}_{t-1}) + u_t], \end{aligned}$$

where we have used the identity $y_t = x_{t-1}\beta + \theta(a_{t-1} - \hat{a}_{t-1}) + u_t$.

By theorem 2.1 of Huang and Guo (1990) and Lemma 3.1 of Poskitt and Lütkepohl (1995), $m^{-1}\sum_{t=2}^{n} ||\hat{a}_t - a_t||^2 = O(\ln^{a+1} m/m)$. Let $L_{m,a} = (\ln^{a+1} m/m)^{1/2}$. Then, using the previous result and Lemma A.1 of Poskitt (2000), it can be shown that $\sum_{t=2}^{n} y_{t-1}^2 = O_p(1)$, $m^{-1}\sum_{t=2}^{n} \hat{a}_t^2 = \sigma_a^2 + O(L_{m,a})$ and $m^{-3/2}\sum_{t=2}^{n} \hat{a}_{t-1} y_{t-1} = O(L_{m,a}) + O_p(m^{-1/2})$, where σ_a^2 is the variance of a_t. Thus, letting $F_{11} = \sum_{t=2}^{n} N_m x_{t-1} x_{t-1}' N_m$, we can write

$$F_{11} = \begin{bmatrix} O_p(1) & O(L_{m,a}) + O_p(m^{-1/2}) \\ O(L_{m,a}) + O_p(m^{-1/2}) & \sigma_a^2 + O(L_{m,a}) \end{bmatrix}.$$

Proceeding as in the proof of Lemma 1, we can show that

$$F_{11}^{-1} = \begin{bmatrix} O_p(1) & O(L_{m,a}) + O_p(m^{-1/2}) \\ O(L_{m,a}) + O_p(m^{-1/2}) & \sigma_a^{-2} + O(L_{m,a}) \end{bmatrix}.$$

Using Lemma A.1 of Poskitt (2000) again, we can show that $m^{-1}\sum_{t=2}^{n} y_{t-1}u_t = m^{1/2}O(\ln\ln^{1/2} m)$, $m^{-1/2}\sum_{t=2}^{n} \hat{a}_{t-1}u_t = m^{1/2}O(L_m)$, $m^{-1}\sum_{t=2}^{n}(a_{t-1}-\hat{a}_{t-1})y_{t-1} = m^{1/2}O(\ln\ln^{1/2} m)$, $m^{-1/2}\sum_{t=2}^{n}(a_{t-1}$ $-\hat{a}_{t-1})\hat{a}_{t-1} = m^{1/2}O(L_{m,a})$. The, letting $G_{10} = \sum_{t=2}^{n} N_m x_{t-1} y_t$, we obtain that

$$N_m G_{10} = \begin{bmatrix} O(LL_m) \\ O(L_{m,a}) \end{bmatrix}.$$

From this, it follows that

$$\begin{aligned} \tilde{\beta} &= \beta + N_m F_{11}^{-1} G_{10} \\ &= \begin{bmatrix} 1 + O(LL_m) \\ \theta + O(L_{m,a}) \end{bmatrix}, \end{aligned}$$

and thus $\tilde{\phi} = -1 + O(LL_m)$. Proceeding similarly, we can prove that if $\{y_t\}$ is I(0), then $\tilde{\phi} = \phi + O(L_{m,a})$ with $|\phi| < 1$.

Let us consider now the third stage regression in the Hannan–Rissanen method. To that end, define $\hat{y}_t = y_t$ if $1 \le t \le n$ and $\hat{y}_t = 0$ otherwise, and

$$\begin{array}{rclll} \tilde{a}_t &=& -\tilde{\theta}\tilde{a}_{t-1} + y_t + \tilde{\phi}y_{t-1} & 1 \le t \le n, & \tilde{a}_t = 0 \quad t \le 0 \\ \eta_t &=& -\tilde{\phi}\eta_{t-1} + \tilde{a}_t & 1 \le t \le n, & \eta_t = 0 \quad t \le 0 \\ \xi_t &=& -\tilde{\theta}\xi_{t-1} + \tilde{a}_t & 1 \le t \le n, & \eta_t = 0 \quad t \le 0 \end{array}$$

The third stage regression in the Hannan–Rissanen method consists of regressing $\tilde{a}_t$ on $-\eta_{t-1}$ and ξ_{t-1}. Letting $x_t = (\eta_t, \xi_t)'$, the OLS estimator $\bar{\beta} = (-\bar{\phi}, \bar{\theta})'$ of this last regression is given by

$$\bar{\beta} = N_m \left(\sum_{t=2}^{n} N_m x_{t-1} x_{t-1}' N_m \right)^{-1} \sum_{t=2}^{n} N_m x_{t-1} \tilde{a}_t,$$

where $N_m = \mathrm{diag}(m^{-1}, m^{-1/2})$ is the normalization matrix defined earlier. Proceeding as in Zhao–Guo (1985), we obtain that

$$\begin{aligned} \tilde{a}_t &= (1 + \tilde{\phi}B)(1 + \tilde{\theta}B)^{-1}\hat{y}_t \\ \eta_t &= (1 + \tilde{\theta}B)^{-1}\hat{y}_t \\ \xi_t &= (1 + \tilde{\phi}B)(1 + \tilde{\theta}B)^{-2}\hat{y}_t, \end{aligned}$$

where B is the backshift operator, $B\hat{y}_t = \hat{y}_{t-1}$. Using the results of the first part of the proof and proceeding as in Zhao–Guo (1985), we can show that if $\{y_t\}$ is I(1), then

$$\bar{\beta} = \begin{bmatrix} O(LL_m) \\ O(L_{m,a}) \end{bmatrix},$$

and if $\{y_t\}$ is I(0), then

$$\bar{\beta} = \begin{bmatrix} O(L_{m,a}) \\ O(L_{m,a}) \end{bmatrix}.$$

Since the final estimator of the Hannan–Rissanen method is $\hat{\beta} = \tilde{\beta} + \bar{\beta}$, the proof is complete.

Proof of Lemma 3

As in the proof of Lemma 1, let $L_m = (\ln(m)/m)^{1/2}$ and $LL_m = (\ln\ln(m)/m)^{1/2}$. Suppose the process $\{y_t\}$ is difference stationary and define the normalization matrix $N_m = \text{diag}(m^{-1/2}I_{k-1}, m^{-3/2})$ to account for the nonzero drift term. Then, (13) can be expressed as

$$\hat{J} = J + N_m \left[\sum_{t=k+1}^{n} N_m(z_{t-1} - \bar{z}_1)(z_{t-1} - \bar{z}_1)' N_m\right]^{-1} \sum_{t=k+1}^{n} N_m(z_{t-1} - \bar{z}_1)v_t',$$

where $v_t = G_T'(u_t - \bar{u})$. Letting $z_t = (z_{st}', z_{nt}')'$, where the partition is conforming to $J = \text{diag}(J_S, J_N)$, Lemma A.1 of Poskitt (2000) can be modified and extended as on p. 89 of that article to give $m^{-3}\sum_{t=k+1}^{n}(z_{nt-1} - \bar{z}_{n1})(z_{nt-1} - \bar{z}_{n1})' = O(1)$, $m^{-1}\sum_{t=k+1}^{n}(z_{st-1} - \bar{z}_{s1})(z_{st-1} - \bar{z}_{s1})' = \tilde{\Gamma}(0) + O(L_m)$, $m^{-2}\sum_{t=k+1}^{n}(z_{st-1} - \bar{z}_{s1})(z_{nt-1} - \bar{z}_{n1})' = O(1) = O_p(m^{-1/2})$ and $m^{-2}\sum_{t=k+1}^{n}(z_{nt-1} - \bar{z}_{n1})(u_t - \bar{u}) = O(1)$, where $\bar{z}_{s1} = \sum_{t=k+1}^{n} z_{st-1}/m$, $\bar{z}_{n1} = \sum_{t=k+1}^{n} z_{nt-1}/m$ and $\tilde{\Gamma}(0) = \text{Var}(z_{st})$. Thus, letting $\tilde{F}_{11} = \sum_{t=k+1}^{n} N_m(z_{t-1} - \bar{z}_1)(z_{t-1} - \bar{z}_1)N_m$, we can write

$$\tilde{F}_{11} = \begin{bmatrix} \tilde{\Gamma}(0) + O(L_m) & O_p(m^{-1/2}) \\ O_p(m^{-1/2}) & O(1) \end{bmatrix}.$$

Since $m^{-3}\sum_{t=k+1}^{n}(z_{nt-1} - \bar{z}_{n1})(z_{nt-1} - \bar{z}_{n1})' \xrightarrow{d} \tilde{W} > 0$ almost surely, where $\tilde{W}$ is a functional of standard Brownian motion, the previous expression for $\tilde{F}_{11}$ implies

$$\tilde{F}_{11} \xrightarrow{d} \begin{bmatrix} \tilde{\Gamma}(0) & 0 \\ 0 & \tilde{W} \end{bmatrix} > 0 \quad \text{almost surely.}$$

As in the proof of Lemma 1, we can use the previous results to show that

$$\tilde{F}_{11}^{-1} = \begin{bmatrix} \tilde{\Gamma}(0)^{-1} + O(L_m) & O_p(m^{-1/2}) \\ O_p(m^{-1/2}) & O_p(1) \end{bmatrix}.$$

Letting $\tilde{G}_{10} = \sum_{t=k+1}^{n} N_m(z_{t-1} - \bar{z}_1)v_t'$ and using the previous result $m^{-2}\sum_{t=k+1}^{n}(z_{nt-1} - \bar{z}_{n1})(u_t - \bar{u}) = O(1)$, it is obtained that

$$\tilde{G}_{10} = \begin{bmatrix} m^{1/2}O(L_m) \\ m^{1/2}O(1) \end{bmatrix}$$

and

$$N_m\tilde{F}_{11}^{-1}\tilde{G}_{10} = \begin{bmatrix} O(L_m) \\ O(1/m) \end{bmatrix},$$

which implies

$$\hat{J}' = J' + \begin{bmatrix} O(L_m) \\ O(1/m) \end{bmatrix} = \begin{bmatrix} J_S' + O(L_m) & O(L_m) \\ O(LL_m) & 1 + O(LL_m) \end{bmatrix}.$$

The rest of the proof is as in the proof of Lemma 1. This completes the proof.

REFERENCES

Akaike, H. (1974a), "A new Look at the Statistical Model Indentification", *IEEE Transactions on Automatic Control*, **AC–19**, 716–723.

——(1974b), "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes", *Annals of the Institute of Statistical Mathematics*, **26**, 363–387.

——(1976), "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion", in *Systems Identification: Advances and Case Studies*, eds. R. K. Mehra and D. G. Lainiotis, New York: Academic Press, pp. 27–96.

Corradi, V. (1999), "Deciding Between I(0) and I(1) via FLIL-Based Bounds", *Econometric Theory*, **15**, 643–663.

Davidson, J. (2009) in *The Methodology and Practice of Econometrics*, a festschrift in honour of F. Hendry, Jennifer Castle and Neil Shepherd Eds., Oxford: Oxford University Press.

DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992), "Integration Versus Trend Stationarity in Time Series", *Econometrica*, **60**, 423–433.

Dickey, D. A., Bell, W. R., and Miller, R. B. (1986), "Unit Roots in Time Series Models: Tests and Implications", *The American Statistician*, **40**, 12–26.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996), "Efficient tests for an Autoregressive Unit Root", *Econometrica*, **64**, 813–836.

Hannan, E. J. and Rissanen, J., "Recursive Estimation of Mixed Autoregressive-Moving Average Order", *Biometrika*, **69**, 81–94.

Hamilton, J. D. (1994), *Time Series Analysis*, NJ: Princeton University Press.

Hayashi, F. (2000), *Econometrics*, NJ: Princeton University Press.

Huang, D. and Guo, L. (1990), "Estimation of Nonstationary ARMAX Models Based on the Hannan–Rissanen Method", *Annals of Statistics*, **18**, 1729–1756.

Műller, U. K. (2008), "The Impossibility of Consistent Discrimination Between I(0) and I(1) Processes", *Econometric Theory*, **24**, 616–630.

Ng, S., and Perron, P. (1995). "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", *Econometrica*, **69**, 1519-1554.

Pantula, S. G. (1991), "Asymptotic Distributions of Unit–Root Tests When the Process Is Nearly Stationary", *Journal of Business and Economic Statistics*, **9**, 63–71.

Perron, P. and Ng., S. (1996), "Useful Modifications to Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties", *Review of Economic Studies*, **63**, 435–465.

Poskitt, D. S. (2000), "Strongly consistent Determination of Cointegrating Rank via Canonical Correlations", *Journal of Business and Economic Statistics*, **18**, 77–90.

Poskitt, D. S. and Lütkepohl, H. (1995), "Consistent Specification of Cointegrated Autoregressive Moving–Average Systems", Discussion Paper 54, Humboldt University, Berlin.

Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics*, **6**, 461–464.

Schwert, G. W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, **7**, 147–159.

Stock, J. H. (1994), "Deciding between I(1) and I(0)", *Journal of Econometrics*, **63**, 105–131.

Tsay, R. S. (1984), "Order Selection in Nonstationary autoregressive Models", *Annals of Statistics*, **12**, 1425–1433.

Zhao–Guo, C., (1985), "The Asymptotic Efficiency of a Linear Procedure of Estimation for ARMA Models", *Journal of Time Series Analysis*, **6**, 53–62.