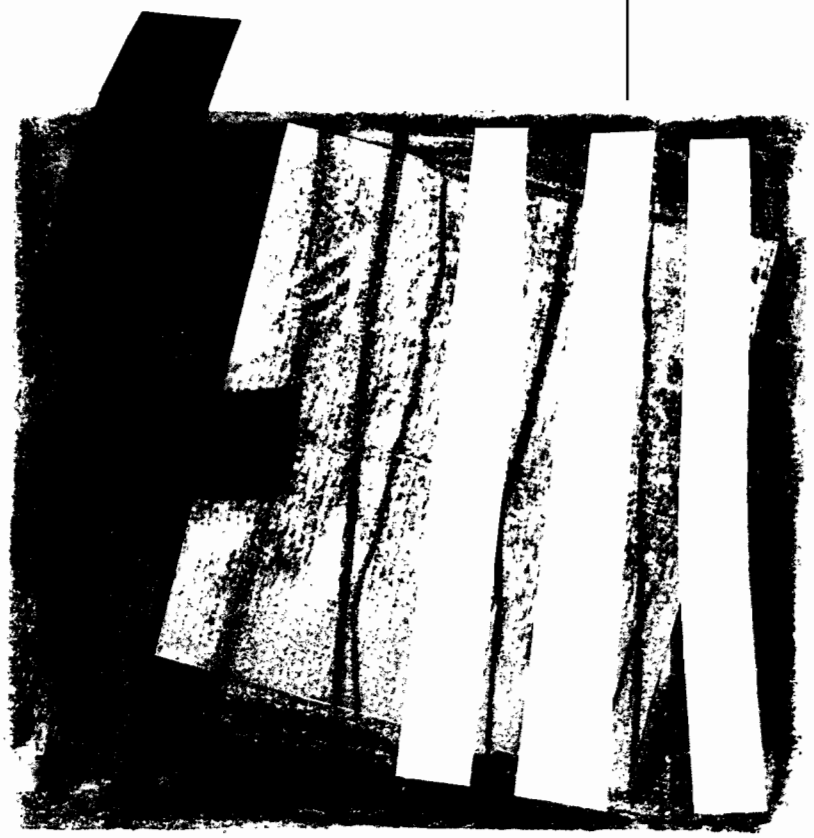
EFFICIENT TESTS FOR UNIT ROOTS WITH PREDICTION ERRORS

Ismael Sánchez

00-80



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-80
Statistics and Econometrics Series 39
November 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

EFFICIENT TESTS FOR UNIT ROOTS WITH PREDICTION ERRORS

Ismael Sánchez*

Abstract

It is well known that the main difference between a stationary (or trend-stationary) process and a process with a unit root is to be observed in their long-term behaviour. This paper exploits this idea and shows that nearly optimal unit-root tests can admit an interpretation based on prediction performance.

This result is not only useful in understanding how efficient tests use the information, but it can also be used to construct new unit-root tests based on prediction errors. A Monte Carlo experiment for the autoregressive moving-average of order (1,1) indicates that the proposed tests have desirable size and power properties.

Keywords: Optimal tests, predictive mean squared error, unit roots.

*Department of Statistics and Econometrics; Universidad Carlos III de Madrid, Avda. de la Universidad, 30, 28911 Leganés (Madrid); e-mail: ismael@est-econ.uc3m.es. This research was supported in part by CICYT, grant PB96-0339.

1 Introduction

1.1 General considerations

This article analyses the relationship between the prediction errors of a predictor that assumes the presence of a unit root and the efficient detection of such a root. The motivation for this analysis is the intuitive concept that the main difference between a stationary (or trend-stationary) process and a process with a unit root is to be observed in their long-term behaviour (see, e.g., Hamilton, 1994, ch. 15). In spite of this well documented result, there is not yet a theory that relates the optimal detection of a unit root with the long-term behaviour of a process. This article fills this gap, and proves that nearly optimal unit root tests can be build using the information of prediction errors. Therefore, prediction errors do contain valuable information for the efficient detection of a unit root.

Since the work of Dickey and Fuller (1979), there has been an abundance of literature devoted to the detection of autoregressive unit roots. A possible criterion for classifying the existing literature is the approach to the detection of the unit root. The original and most commonly-employed approach is that of Dickey and Fuller (1979), which is based on the asymptotic properties of the ordinary least squares (OLS) estimator. Important variations of Dickey-Fuller tests (hereafter, DF tests) are their extensions to other estimation methods, such as maximum likelihood (e.g., Pantula *et al.*, 1994; Yap and Reinsel, 1995; Shin and Fuller, 1998; Shin and Lee, 2000), the weighted symmetric estimator (hereafter, τ_W test) (Park and Fuller, 1995; Fuller, 1996, p. 568) and the generalized least-squares (GLS) estimator under a fixed local alternative (hereafter, τ_{GLS} test), developed by Elliott *et al.* (1996) (see also Hwang and Schmidt, 1996, and Xiao and Phillips, 1998).

A very important approach is the construction of tests observing certain optimal criteria (under normality). Works following this so-called optimal approach are Sargan and Bhargava (1983) and Bhargava (1986). These authors extend the basis of optimal serial correlation tests to construct approximate uniformly most powerful invariant (UMPI) tests and approximate locally best invariant (LBI) tests. Important extensions of the Sargan and Bhargava tests, and also closely related to DF tests, are the M tests of Stock (1990) (see

also Perron and Ng, 1996). Recently, Ng and Perron (2000) have improved the performance of these M tests using GLS detrending under a fixed local alternative (hereafter M^{GLS} tests). Important contributions to this optimal approach are the (infeasible) point optimal invariant (POI) tests and LBI tests developed by Dufour and King (1991). A sequence of these infeasible POI tests (that use the true value of the root as the alternative) provide the upper attainable power (Gaussian power envelope) at each alternative. This power envelope can be used as a benchmark to evaluate the performance of any feasible test. Dufour and King (1991) found that POI tests perform well in practice if a fixed alternative is properly chosen. According to this idea, Elliott *et al.* (1996) and Elliott (1999) proposed a feasible POI test (hereafter P_T test) using a fixed local-to-unity alternative.

In this article, a new approach, based on prediction errors, is introduced. The usefulness of this new approach is twofold. First, it proves to be very useful to give an intuitive interpretation of the use of the information by efficient tests. Second, it allows the construction of new unit-root tests, based on prediction errors. The new tests are shown to be asymptotically nearly optimal in the vicinity of one. Therefore, this approach is also related to the mentioned optimal approach. Hence, the intuitive notion that the nonstationarity is related with the long-term behaviour of the process has optimal properties for unit root detection. The new tests also have good finite sample properties. In the AR(1) case, the empirical power is similar to the POI test. The proposed tests can be extended to the general ARMA case by fitting an ARMA model. Therefore, they need not rely on autoregressive approximations. A Monte Carlo experiment for an ARMA(1,1) model shows that the proposed tests still have high power with very small size distortion.

The paper proceeds as follows: After setting up the notation and the model in the next subsection, Section 2 introduces test statistics based on the mean squared prediction error (MSPE) of a random walk predictor. These tests are the basis for the remaining tests of the paper. Section 3 derives tests that are nearly optimal in the vicinity of unity and shows that they have a prediction-error interpretation. The resulting near-optimal tests happen to be functions of the statistics proposed in Section 2. Section 4 uses the prediction-error interpretation to modify the near-optimal tests in order to achieve high power in regions far from the unit circle. Section 5 compares the proposed tests with some other existing in the literature through a Monte Carlo experiment in the AR(1) case. Section 6 extends the tests to the ARMA case and compares them

with other tests in a simulation exercise with an ARMA(1,1) model. Section 7 concludes. Mathematical proofs are given in the Appendix.

1.2 Notation and the model

Let $\{y_t\}$ be a discrete stochastic process. We assume that this process contains a deterministic component d_t and a pure stochastic component x_t; namely, $y_t = d_t + x_t$. It is assumed that the deterministic component can be a mean, $d_t = \mu$, and a deterministic trend, $d_t = \mu + \delta t$. The pure stochastic part has the following structure: $x_t = \rho x_{t-1} + u_t$, satisfying the following conditions:

Assumption A: *x_t is initialized at $t = 0$ by x_0, a random variable with finite variance.*

Assumption B: *u_t is a stationary and invertible ARMA(p,q) satisfying $\phi(B)u_t = \theta(B)a_t$, where a_t is a sequence of iid random variables with $E(a_t) = 0$ and $E(a_t^2) = \sigma^2$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in the backward-shift operator B, with no common factors and with either $\phi_p \neq 0$ or $\theta_q \neq 0$.*

Assumption C: *$T^{-1/2} \sum_{t=1}^{[Tr]} u_t \xrightarrow{d} \omega W(r)$, $T^{-1/2} \sum_{t=1}^{[Tr]} a_t \xrightarrow{d} \sigma W(r)$, where $\omega^2 = E(u_t^2) + 2\sum_{k=2}^{\infty} E(u_1 u_k) = \sigma^2 \{\theta(1)/\phi(1)\}^2$ is the long-run variance of u_t, and $W(r)$ is a standard Brownian motion defined on $C[0,1]$. The symbol $\xrightarrow{d}$ denotes weak convergence in distribution. For $0 \leq r \leq 1$, $[Tr]$ denotes the greatest integer less than or equal to Tr.*

Notice that Assumption A includes the case x_0 =constant, with probability one, as a special case. Notice also that the limits of partial sums of u_t and a_t depend on the same standard Brownian motion $W(r)$. The data-generating process is, therefore,

$$y_t = \mu + \delta t + x_t, \quad x_t = \rho x_{t-1} + u_t, \quad u_t = \psi(B)a_t, \tag{1}$$

where $\psi(B) = \phi(B)^{-1}\theta(B)$. This process can also be expressed as $y_t = \mu(1-\rho) + \delta\rho + \delta(1-\rho)t + \rho y_{t-1} + \psi(B)a_t$.

2 Tests based on the ratio of observed and expected MSPE

This section shows basic test statistics of the prediction-error approach. They arise by comparing the empirical MSPE of a predictor with the expected one, under the null hypothesis of a unit root. The proposed tests are based on the following property: if the process is stationary, as the horizon increases the empirical MSPE is bounded, although the predictor assumes there is a unit root. On the other hand, if there is a unit root, the empirical and the expected MSPEs are unbounded. We define $e_{j|i}$ as the empirical prediction error of forecasting y_j from $t = i$ $(j > i)$ under the null hypothesis of random walk $(u_t = a_t)$. Therefore, $e_{j|i} = y_j - y_i$. We first base the test statistics on the cumulative sum of squared prediction errors from the origin of the series. Then

$$\sum_{t=2}^{T} e_{t|1}^2 = \sum_{t=2}^{T} (y_t - y_1)^2. \tag{2}$$

The expected value of this statistic, under the null hypothesis of a unit root, is:

$$E\left(\sum_{t=2}^{T} e_{t|1}^2\right) = \sigma^2\{1 + 2 + \cdots + (T-1)\} = (1/2)\sigma_u^2 T(T-1). \tag{3}$$

Since σ_u^2 is not known, the consistent (under the null) estimator $\hat{\sigma}_u^2 = T^{-1}\sum_{t=2}^{T}(y_t - y_{t-1})^2$ can be used. Finally, the test statistic is obtained by dividing the cumulative sum (2) and its estimated expected value under the null (3). The constant can be dropped since it does not affect the test. Also, for the sake of simplicity, $T(T-1)$ is approximated to T^2. The proposed statistic of this prediction-error approach is as follows:

$$C_1^{\mu} = \frac{\sum_{t=2}^{T}(y_t - y_1)^2}{T\sum_{t=2}^{T}(y_t - y_{t-1})^2}, \tag{4}$$

where C stands for *cumulative* and the subscript shows the origin of the predictions. This test statistic is invariant under the group of transformations $y_t \rightarrow ay_t + b$, with a, b constants and, therefore, it is not affected by the mean value of the series. Therefore, it is also applicable to the case of a non-zero mean. It can be verified that, under the null hypothesis of a unit root, $C_1^{\mu} = O_p(1)$. Under the alternative, the numerator is, however, of lower order of magnitude than in the previous case since $T^{-2}\sum_{t=2}^{T} y_t^2 = O_p(T^{-1})$. Therefore, under the stationary alternative, $C_1^{\mu} \xrightarrow{p} 0$. Therefore, a consistent unit-root test against a stationary alternative has the rejection region $C_1^{\mu} < c_1^{\mu}$.

We now extend this test to the case of a null hypothesis of a random walk with a drift, $y_t = \delta + y_{t-1} + u_t$, where $u_t = a_t$. The prediction errors are $e_{t|1} = y_t - y_1 - (t-1)\delta$. An efficient estimator of δ, under the null, can be obtained from the regression $\Delta y_t = \delta + u_t$. Therefore, $\hat{\delta} = (T-1)^{-1}(y_T - y_1)$. A consistent estimator of σ_u^2, under the null, is $\hat{\sigma}_u^2 = T^{-1}\sum_{t=2}^{T}(y_t - y_{t-1} - \hat{\delta})^2$. This leads to the following test statistic:

$$C_1^{\tau} = \frac{\sum_{t=2}^{T}\{y_t - y_1 - \hat{\delta}(t-1)\}^2}{T\sum_{t=2}^{T}(y_t - y_{t-1} - \hat{\delta})^2} = \frac{\sum_{t=2}^{T}(\hat{y}_t - \hat{y}_1)^2}{T\sum_{t=2}^{T}(\hat{y}_t - \hat{y}_{t-1})^2}, \tag{5}$$

where $\hat{y}_t$ is the estimated detrended series. Similarly to C_1^{μ}, the rejection region is $C_1^{\tau} < c_1^{\tau}$. The statistics C_1^{μ} and C_1^{τ} are already in the literature using different justification. They correspond to $(TN_1)^{-1}$ and $(TN_2)^{-1}$, respectively, where N_1 and N_2 are in Bhargava (1986). Bhargava (1986) shows that, if the Anderson approximation is used for the inverse of the covariance of the process, and the first observation is extracted from its conditional distribution, these statistics lead to approximate LBI tests, under normality. These statistics are also derived in Tanaka (1996), with the same initial conditions, by taking the second derivative of the likelihood function, under normality (extended score tests).

Another way of evaluating the prediction errors is to also consider the series in reverse order (backward process). The use of the reversed time series can be justified by the time-reversibility of Gaussian stationary processes (see, e.g., Weiss, 1975; Box and Jenkins, 1976, p. 197). This property states that the processes in direct and reverse order have the same covariance structure. Therefore, under stationarity, better use of the information could be obtained if both processes are analyzed. The statistic that averages the prediction errors from both extremes of the series and at each horizon is:

$$C_{1T}^{\mu} = \frac{\sum_{t=2}^{T}(y_t - y_1)^2 + \sum_{t=1}^{T-1}(y_t - y_T)^2}{2T\sum_{t=2}^{T}(y_t - y_{t-1})^2}, \tag{6}$$

in which the subscript $1T$ denotes the origin of the predictions. Several authors have also used the property of time-reversibility to improve the performance of DF tests (Sen and Dickey, 1987; Pantula *et al.*, 1994; Leybourne, 1995; Park and Fuller, 1995; among others). For the case of a null hypothesis of a random walk with a drift, it can be shown that

$$C_{1T}^{\tau} = C_1^{\tau}.$$

The following theorem shows the limiting distribution of the statistics $C_1^{\mu}$and C_{1T}^{μ} when the process is $y_t = y_{t-1} + u_t$, $u_t = \psi(B)a_t$, and that of C_1^{τ} when the process is $y_t = \delta + y_{t-1} + u_t$, $u_t = \psi(B)a_t$. For

convenience, the following processes are defined: $W_B(r) \equiv \{W(1-r) - W(1)\}$, and $W^\tau(r) = W(r) - rW(1)$.

Theorem 1 *Under Assumptions A, B, and C, when* $\rho = 1$,

$$\begin{aligned}
&(i) \quad C_1^\mu \xrightarrow{d} \kappa^{-2} \int_0^1 \{W(r)\}^2 dr, \\
&(ii) \quad C_1^\tau (\equiv C_{1T}^\tau) \xrightarrow{d} \kappa^{-2} \int_0^1 \{W^\tau(r)\}^2 dr, \\
&(iii) \quad C_{1T}^\mu \xrightarrow{d} \kappa^{-2} \tfrac{1}{2} \left\{ \int_0^1 \{W(r)\}^2 dr + \int_0^1 \{W_B(r)\}^2 dr \right\},
\end{aligned}$$

where $\kappa^2 = \omega^{-2}\sigma_u^2; \sigma_u^2 = E(u^2); \omega^2 = \sigma^2 \psi(1)^2$.

Remark 1 The consistency of the tests follows from noting that the limiting distributions are $O_p(1)$ and are positive with probability one, so that the critical values of the tests are positive. Since under the alternative $C_1^\mu \xrightarrow{p} 0$, $C_1^\tau \xrightarrow{p} 0$, and $C_{1T}^\mu \xrightarrow{d} 0$ as $T \to \infty$, the consistency holds.

3 Nearly-optimal tests and prediction errors

3.1 Asymptotic results

This section proposes asymptotic tests that are nearly optimal under normality of the disturbances. The resulting statistics are found to be the same as the statistics proposed in the previous section or functions of them. In all of these cases, the resulting nearly-optimal tests also allow an interpretation based on prediction errors and, therefore, are useful to explain how efficient unit-root tests use the information. Very important conclusions can be obtained from this section. First, it confirms the intuition that the behaviour of a process, in the long term, provides important information for the efficient detection of the unit root. Secondly, it is shown here that efficient tests utilize the time-reversibility property of Gaussian stationary processes to improve their asymptotic performances. The interest here is in asymptotic and invariant tests (invariant to the values of the parameters describing the deterministic terms) with high power in the vicinity of $\rho = 1$. To obtain these properties, a second-order Taylor expansion of the asymptotic POI test statistics, around the null hypothesis, is performed. Hence, the proposed tests are, asymptotically, nearly-optimal invariant, close to the null hypothesis. This section develops tests for the AR(1) case. Section 6 extends the proposed tests to the ARMA case.

We assume, first, the model (1) with $\delta = 0$ and $\psi(B) = 1$ (then, $u_t = a_t$). We also assume that u_t is

normally distributed. In this paper, two different assumptions are made about x_0:

Assumption A1 *(nonstationary case)*: *We assume that* $x_0 = 0$ *so* $x_1 = a_1$. *Therefore,* y_1 *is extracted from its conditional distribution.*

Assumption A2 *(stationary case)*: *We assume that* $x_0 = 0$ *when* $\rho = 1$ *and* x_0 *is random with zero mean and variance* $\sigma^2/(1-\rho^2)$ *when* $|\rho| < 1$. *Therefore,* y_1 *is extracted from its unconditional distribution.*

Under Assumption A1, the process x_t is nonstationary under the alternative. This assumption is denoted as the nonstationary case, although some authors also refer to this as the conditional case. The covariance matrix of the $T \times 1$ vector $\mathbf{x} = (x_1, ..., x_T)'$ is $\sigma^2\Omega_N(\rho)$, where

$$\Omega_N^{-1}(\rho) = \mathbf{I}_T - \rho(\mathbf{L}_T + \mathbf{L}_T') + \rho^2\mathbf{L}_T'\mathbf{L}_T, \tag{7}$$

with $\mathbf{I}_T$ a $T \times T$ identity matrix and $\mathbf{L}_T$ a $T \times T$ matrix with ones on the diagonal immediately below the main diagonal and zeros elsewhere. This nonstationary model is the same as that of Dufour and King (1991), with $d_1 = 1$. According to these authors, the POI test of $\rho = 1$ against $\rho = \rho_0$ rejects the null hypothesis for small values of

$$S_N(\rho_0) = \frac{\hat{\mathbf{x}}_N'\Omega_N^{-1}(\rho_0)\hat{\mathbf{x}}_N}{\hat{\mathbf{x}}_1'\Omega_N^{-1}(1)\hat{\mathbf{x}}_1}, \tag{8}$$

where $\hat{\mathbf{x}}_N$ and $\hat{\mathbf{x}}_1$ are GLS residual vectors using $\Omega_N^{-1}(\rho_0)$ and $\Omega_N^{-1}(1)$, respectively. Under Assumption A2, x_t is covariance stationary under the alternative. This assumption is denoted as the stationary case (also known in the unit root literature as the unconditional case). The covariance matrix of the vector $\mathbf{x}$ is $\sigma^2\Omega_N(1)$, under the null, and $\sigma^2\Omega_S(\rho)$, under the alternative, where

$$\Omega_S^{-1}(\rho) = \Omega_N^{-1}(\rho) - \rho^2\mathbf{e}\mathbf{e}', \tag{9}$$

with $\mathbf{e} = (1, 0, ..., 0)'$. From the Neyman-Pearson lemma, an asymptotic representation of the POI test against some alternative, $|\rho_0| < 1$, rejects the null for small values of

$$S_S(\rho_0) = \frac{\hat{\mathbf{x}}_S'\Omega_S^{-1}(\rho_0)\hat{\mathbf{x}}_S}{\hat{\mathbf{x}}_1'\Omega_N^{-1}(1)\hat{\mathbf{x}}_1}, \tag{10}$$

where $\hat{\mathbf{x}}_S$ are the GLS residual vectors using $\Omega_S^{-1}(\rho_0)$ (see, e.g., Kadiyala, 1970; or Elliott, 1999). The nonstationary case is a reasonable assumption when the time series has its origin at x_0. Potential applications of this assumption can be found, for instance, in engineering, where the beginning of processes and experiments are known and controlled. On the other hand, the stationary case reflects the situation of a time series with the origin in a very far past point. This assumption can be of interest, for instance, with some macroeconomic data. The following theorems show the Taylor-series expansion of (8) and (10) around $\rho = 1$.

Theorem 2 *Let $y_t = \mu + x_t$, where $x_t = \rho x_{t-1} + a_t$, with x_0 following Assumption A1. Then,*

$$S_N(\rho) = \frac{\hat{\mathbf{x}}_N' \Omega_N^{-1}(\rho) \hat{\mathbf{x}}_N}{\hat{\mathbf{x}}_1' \Omega_N^{-1}(1) \hat{\mathbf{x}}_1} = 1 + (\rho - 1) S_N'(1) + \frac{(\rho - 1)^2}{2} S_N''(1) + O\{(\rho - 1)^3\},$$

where

$$\begin{aligned} S_N'(1) &= \left. \frac{\partial S_N(\rho)}{\partial \rho} \right|_{\rho=1} = 1 - \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \\ S_N''(1) &= \left. \frac{\partial^2 S_N(\rho)}{\partial \rho^2} \right|_{\rho=1} = 2 \frac{\sum_{t=2}^T (y_t - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2} - 4 \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}. \end{aligned} \tag{11}$$

Theorem 3 *Let $y_t = \mu + x_t$, where $x_t = \rho x_{t-1} + a_t$, with x_0 following Assumption A2. Then,*

$$S_S(\rho) = \frac{\hat{\mathbf{x}}_S' \Omega_S^{-1}(\rho) \hat{\mathbf{x}}_S}{\hat{\mathbf{x}}_1' \Omega_N^{-1}(1) \hat{\mathbf{x}}_1} = 1 + (\rho - 1) S_S'(1) + \frac{(\rho - 1)^2}{2} S_S''(1) + O\{(\rho - 1)^3\},$$

where

$$\begin{aligned} S_S'(1) &= \left. \frac{\partial S_S(\rho)}{\partial \rho} \right|_{\rho=1} = 1 - \frac{1}{2} \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \\ S_S''(1) &= \left. \frac{\partial^2 S_S(\rho)}{\partial \rho^2} \right|_{\rho=1} = \frac{\sum_{t=2}^T (y_t - y_1)^2 + \sum_{t=1}^{T-1} (y_t - y_T)^2}{2 \sum_{t=2}^T (y_t - y_{t-1})^2} \\ &\quad - \frac{(y_T - y_1)^2 + \sum_{t=2}^{T-1} (y_t - y_1)(y_T - y_t)}{\sum_{t=2}^T (y_t - y_{t-1})^2}. \end{aligned} \tag{12}$$

These Taylor expansions suggest different kinds of asymptotic tests. At first approximation, the test statistics can be constructed by using only the first derivatives, $S_N'(1)$ and $S_S'(1)$, as expressed in (11) and (12). These derivatives are, apart from a constant, identical for both stationary and nonstationary cases. By the Generalized Neyman-Pearson Lemma (see, e.g., Ferguson, 1967), these first derivatives lead to asymptotically LBI tests. Therefore, these LBI tests reject the null for small values of

$$E_N^\mu \equiv E_S^\mu = \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \tag{13}$$

where E stands for *extreme* points of the series and the subscript N or S denotes the nonstationary case or the stationary case, respectively. This statistic was previously obtained by Nabeya and Tanaka (1990) and Tanaka (1996), but only under Assumption A1 (corresponding to E_N^{μ}). Statistic (13) also has an interpretation in terms of prediction errors. It is the ratio of the empirical prediction squared error of the last observation, calculated from the origin of the series, assuming a unit root, to the consistent estimator of its expectation. When $\mu = 0$ (or μ is known, and the series is previously demeaned with that value), following the same arguments as in the proof of theorems 2 and 3, the corresponding first derivatives lead to the test statistics:

$$E_S^0 = \frac{y_T^2 + y_1^2}{\sum_{t=2}^{T}(y_t - y_{t-1})^2}, \tag{14}$$

$$E_N^0 = \frac{y_T^2}{\sum_{t=2}^{T}(y_t - y_{t-1})^2}, \tag{15}$$

for the stationary and nonstationary cases, respectively. These asymptotic LBI tests are only invariant under transformations of the type $y_t \rightarrow ay_t$. This last statistic, E_N^0, was also previously obtained by Nabeya and Tanaka (1990), Stock (1994), and Tanaka (1996). Tests based on these E statistics only have optimal power when the alternative is very close to the null (say, $\rho = 0.999...$). The power is, however, very low for alternatives of practical interest. Given the low usefulness of the first derivatives, one could discard this part of the Taylor expansion and consider test statistics based only on the second derivatives. To obtain random variables with non-degenerate limiting distributions, it is necessary to divide the second derivatives by T. This correction makes the second term both in $S_N''(1)$ and $S_S''(1)$ converge in probability to zero, under the null hypothesis of a unit root. In the nonstationary case, it is verified that

$$\frac{(y_T - y_1)^2}{T\sum_{t=2}^{T}(y_t - y_{t-1})^2} = O_p(T^{-1}).$$

In the stationary case, it holds that

$$\frac{\sum_{t=2}^{T-1}(y_t - y_1)(y_T - y_t)}{T\sum_{t=2}^{T}(y_t - y_{t-1})^2} = \frac{\sum_{t=2}^{T} t\hat{\gamma}_t}{(T-1)\hat{\sigma}_u^2} = O_p(T^{-1}), \tag{16}$$

where $\hat{\gamma}_t = \sum_{j=1}^{T-t} u_j u_{j+t}/T$. Note that (16) still holds if u_t is a general stationary ARMA model. With these results, the second derivatives satisfy:

$$\frac{1}{2T}S_N''(1) = C_1^{\mu} + O_p(T^{-1}), \tag{17}$$

$$\frac{1}{2T}S_S''(1) = C_{1T}^{\mu} + O_p(T^{-1}), \tag{18}$$

where C_1^μ and C_{1T}^μ are derived in the previous section using the prediction-error interpretation. Therefore, the second derivatives lead to the asymptotic tests C_1^μ and C_{1T}^μ, respectively.

An important conclusion can be extracted from these tests based on the first or second derivatives: in the stationary case, unit-root tests seem to use the time-reversibility property of Gaussian stationary processes to obtain a more efficient use of the information. This can be seen, e.g., in the asymptotic equivalence between a test based on $S_S''(1)$ and on C_{1T}^μ, where prediction errors are evaluated from both ends of the series. However, in the nonstationary case, a test based on $S_N''(1)$ is asymptotically equivalent to C_1^μ, where only one extreme of the series is considered. Statistics (13) and (14) also admit this time-reversibility interpretation: when $\mu = 0$, expression (15) is a measure of how far the last observation has deviated from the mark, using zero as a reference. Under Assumption A2, however, if the time-reversal series is also considered, the corresponding asymptotically LBI test should be based on both extremes, as does (14), since the first observation will be the last one for the reverse process. Similarly, when μ is unknown, the time-reversibility interpretation can also explain the fact that the corresponding asymptotically LBI tests use the same test statistic (13), both in the stationary and nonstationary cases. This optimal property of the time-reversibility is also important to understand why modifications of DF tests, based on the use of the series in reverse order as well, improve their performance (e.g., Sen and Dickey, 1987; Leybourne, 1995; Park and Fuller, 1995; Shin and So, 1997).

If $\mu = 0$, following the same arguments as in the proof of theorems 2 and 3, the asymptotic tests based on the second derivatives use the statistics

$$C_1^0 = C_{1T}^0 = \frac{\sum_{t=2}^{T-1} y_t^2}{T \sum_{t=2}^{T} (y_t - y_{t-1})^2}.$$

for the nonstationary and stationary cases, respectively.

A third type of asymptotic test, using this nearly-optimal approach, can be based on both first and second derivatives of the Taylor expansion. Hence, they also have a prediction-error interpretation. These tests are asymptotically nearly-optimal invariant in the vicinity of one, and are denoted as N tests. For convenience, the value of ρ that determines the neighbourhood around unity is parameterized as $\rho_c = 1 - c/T$, with

$c > 0$. This formulation is more appropriate than the use of a fixed value of ρ, since the power for a given alternative depends on T. In the nonstationary case, the second-order Taylor expansion leads to $S_N(\rho_c) \approx 1 - c/T + (c/T)\left(E^{\mu} + cC_1^{\mu}\right)$. Therefore, the corresponding N test rejects the null hypothesis for small values of

$$N_N^{\mu} = E_N^{\mu} + cC_1^{\mu}. \tag{19}$$

Similarly, for the stationary case, $S_S(\rho_c) \approx 1 - c/T + \{c/(2T)\}\left(E^{\mu} + cC_{1T}^{\mu}\right)$, and the corresponding N test will be

$$N_S^{\mu} = E_S^{\mu} + cC_{1T}^{\mu}. \tag{20}$$

The extension to the $\mu = 0$ case leads to:

$$N_N^{0} = E_N^{0} + cC_1^{0}, \tag{21}$$

$$N_S^{0} = E_S^{0} + cC_{1T}^{0}. \tag{22}$$

The following theorem extends the previous results to the case with deterministic linear trend. Since the theorem holds for both stationary and nonstationary cases, differences in notation are omitted. This theorem extends the results of Nabeya and Tanaka (1990) and Tanaka (1996) to the stationary case.

Theorem 4 *Let $y_t = \mu + \delta t + x_t$, where $x_t = \rho x_{t-1} + a_t$, with x_0 either fixed or following a random variable with zero mean and variance $\sigma^2/(1-\rho^2)$. Then,*

$$S(\rho) = \frac{\hat{\mathbf{x}}'\Omega^{-1}(\rho)\hat{\mathbf{x}}}{\hat{\mathbf{x}}_1'\Omega_N^{-1}\hat{\mathbf{x}}_1} = 1 + (\rho-1)S'(1) + \frac{(\rho-1)^2}{2}S''(1) + O\{(\rho-1)^3\},$$

where

$$S'(1) = \left.\frac{\partial S(\rho)}{\partial \rho}\right|_{\rho=1} = 1,$$

$$S''(1) = \left.\frac{\partial^2 S(\rho)}{\partial \rho^2}\right|_{\rho=1} = 2\frac{\sum_{t=2}^{T}\{y_t - y_1 - \hat{\delta}(t-1)\}^2}{\sum_{t=2}^{T}(y_t - y_{t-1} - \hat{\delta})^2} = 2\frac{\sum_{t=2}^{T}(\hat{y}_t - \hat{y}_1)^2}{\sum_{t=2}^{T}(\hat{y}_t - \hat{y}_{t-1})^2}.$$

Since the first derivative is a constant, an LBI test would be of no practical interest. By the Generalized Neyman-Pearson lemma, the second derivative can be used to construct asymptotically LBIU tests for both the stationary and nonstationary cases. It can be verified, by dividing the second derivative by T, that the asymptotically LBIU tests reject the null for small values of the statistics C_1^{τ}, derived in section 2 using

the prediction-error interpretation. The fact that the test statistic is the same in both the stationary and nonstationary cases, is also compatible with the time-reversibility interpretation of the stationary case since, as seen in Section 2, $C^{\tau}_{1T} = C^{\tau}_{1}$. Therefore, the N tests are:

$$N^{\tau}_{N} = N^{\tau}_{S} = C^{\tau}_{1}. \tag{23}$$

The limiting distributions of the statistics N^{μ}_{N} and N^{μ}_{S}, for the general case of $u_t = \psi(B)a_t$, are shown in the following theorem. The limiting distribution of N^{0}_{N} and N^{0}_{S} tests is the same as for the N^{μ}_{N} test. The proof is a straightforward application of the proof of theorem 1 and is omitted.

Theorem 5 *Let y_t be the process (1) with $\rho = 1$ and $\delta = 0$. Then, under Assumptions A, B, and C*

$$(i) \quad N^{\mu}_{N} \xrightarrow{d} \kappa^{-2}\left\{W^2(1) + c\int_0^1 \{W(r)\}^2 dr\right\},$$
$$(ii) \quad N^{\mu}_{S} \xrightarrow{d} \kappa^{-2}\left[W^2(1) + c\tfrac{1}{2}\left\{\int_0^1 \{W(r)\}^2 dr + \int_0^1 \{W^0_B(r)\}^2 dr\right\}\right].$$

Remark 2 *The consistency of the tests can be established using the same arguments as in remark 1*

3.2 Relation with existing tests

The N tests are closely related to the P_T tests in Elliott *et al.* (1996) and Elliott (1999). These P_T tests are asymptotically POI tests at the specific alternative $\rho = \rho_c$, whereas N tests come from an approximation of asymptotically POI tests around unity. Therefore, both N and P_T tests are alternative ways to use the optimal theory to construct tests with high power when ρ is close to the null, although only the N tests have an explicit prediction-error interpretation. When there are no deterministic components, these tests (denoted as P^{0}_{T} and N^{0} tests) verify, if $\rho = 1$,

$$P^{0}_{T\text{-}N} = cN^{0}_{N} + o_p(1),$$
$$P^{0}_{T\text{-}S} = cN^{0}_{S} + o_p(1),$$

where the subscript N or S is added to the P_T tests to distinguish the initial-condition assumption. The extension of P^{0}_{T} and tests N^{0} to the case with deterministic components can be interpreted as replacing y_t, in the numerator of the statistics, with its detrended counterpart. The tests differ, however, in the detrending method. Whereas in P_T tests, detrending is made by GLS under the alternative ρ_c, in N tests it is made under the null. This detrending method gives N tests a clear interpretation in terms of prediction errors.

The N_N^0 test is also related with the MZ_α test, proposed by Stock (1990) (see also Perron and Ng, 1996,2000), since it can be written that

$$\mathrm{MZ}_\alpha = \left(E_N^0 - 1\right)\left(2C_1^0\right)^{-1}. \tag{24}$$

Therefore, the test also has information about the first and second derivatives of the Taylor expansion of the asymptotically POI in the non-deterministic case. It can be seen from (24) that the MZ_α is designed assuming Assumption A1. Under Assumption A2, these tests could have been written as $\mathrm{MZ}_\alpha = \left(E_S^0 - 1\right)\left(2C_1^0\right)^{-1}$, instead of (24). Analogously, the MSB test of Stock (1990), based on Bhargava's (1986) R_1 statistics (see also Sargan and Bhargava, 1983; and Perron and Ng, 1996), can be expressed as

$$\mathrm{MSB} = \sqrt{C_1^0}. \tag{25}$$

Similarly to P_T tests, Ng and Perron (2000) extend the MZ_α and MSB tests to the case with deterministic components by detrending by GLS under the alternative ρ_c. The resulting tests are denoted by MZ_α^{GLS} and MSB^{GLS}, respectively.

Notice that, when $d_t = \mu$, it has been proved here that an efficient use of the information leads to different test statistics, depending on the initial conditions. Therefore, the classic unit-root tests that assume a specific assumption on the initial values need not have comparable performances under different assumptions. However, when $d_t = \mu + \delta t$, this comment does not apply.

4 Modified tests with predictors based on a local alternative

The previous section shows how the prediction-error approach leads to (asymptotically nearly) efficient tests in the vicinity of unity. This section also uses the prediction-error interpretation to modify the proposed statistics, to achieve greater power at alternatives far from the unit circle. The modified statistics use predictors constructed under a fixed local alternative instead of under the null of a unit root. It should be noted that the information regarding the alternative ρ_c has already been used in the definition of the N tests in the last section. This information is summarized in the parameter c of the statistics (19) to (22). In this section, the information of this local alternative is extended to the predictors also. All previous tests have

greater empirical power with this approach. The best performance is obtained by the modifications of the N tests (denoted as N^{GLS} tests). Our attention, therefore, is restricted to these tests.

We suppose that the process follows the model (1) with $\psi(B) = 1$. Under the null hypothesis of a unit root, the h-steps ahead prediction of y_{t+h} from y_t is $\hat{y}_{t+h|t} = \delta h + y_t$. Test statistics proposed in previous sections use this predictor. It is now considered that the predictor assumes that the process is nearly nonstationary with $\rho = \rho_c$. Therefore, the new predictor is $\hat{y}^c_{t+h|t} = \hat{\mu}^c(1-\rho_c^h) + \hat{\delta}^c t(1-\rho_c^h) + \hat{\delta}^c h + \rho_c^h y_t$, where $\hat{\mu}^c$ and $\hat{\delta}^c$ are obtained by GLS estimation using the covariance matrix $\Omega_N(\rho_c)$ or $\Omega_S(\rho_c)$. In the nonstationary case, these estimators can also be obtained by regressing $[y_1,(1-\rho_c B)y_2, ..., (1-\rho_c B)y_T]$ on $[z_1,(1-\rho_c B)z_2, ..., (1-\rho_c B)z_T]$, where $z_t = 1$ if $d_t = \mu$ and $z_t = (1,t)$ if $d_t = \mu + \delta t$. In the stationary case, the same procedure can be used, but y_1 is replaced with $(1-\rho_c^2)^{1/2}y_1$, and z_1 is replaced with $(1-\rho_c^2)^{1/2}z_1$. We denote $y_t^\mu = y_t - \hat{\mu}^c$ and $y_t^\tau = y_t - \hat{\mu}^c - \hat{\delta}^c t$. Then, the empirical prediction error of forecasting y^μ_{t+h} from y_t^μ can be expressed as $y^\mu_{t+h} - \hat{y}^\mu_{t+h|t} = y^\mu_{t+h} - \rho_c^h y_t^\mu$ if $d_t = \mu$; or $y^\tau_{t+h} - \hat{y}^\tau_{t+h|t} = y^\tau_{t+h} - \rho_c^h y_t^\tau$ if $d_t = \mu + \delta t$. In the nonstationary case, the modified N test statistics are:

$$N_N^{GLS-\mu} = \frac{(y_T^\mu - \rho_c^{T-1}y_1^\mu)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2} + c\frac{\sum_{t=2}^T (y_t^\mu - \rho_c^{t-1}y_1^\mu)^2}{T\sum_{t=2}^T (y_t - y_{t-1})^2} \tag{26}$$

and

$$N_N^{GLS-\tau} = \frac{\sum_{t=2}^T (y_t^\tau - \rho_c^{t-1}y_1^\tau)^2}{T\sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}. \tag{27}$$

In the stationary case, the modified N test statistics are

$$N_S^{GLS-\mu} = \frac{(y_T^\mu - \rho_c^{T-1}y_1^\mu)^2 + (y_1^\mu - \rho_c^{T-1}y_T^\mu)^2}{2\sum_{t=2}^T (y_t - y_{t-1})^2} + c\frac{\sum_{t=2}^T (y_t^\mu - \rho_c^{t-1}y_1^\mu)^2 + \sum_{t=1}^{T-1}(y_{T-t}^\mu - \rho_c^{t-1}y_T^\mu)^2}{2T\sum_{t=2}^T (y_t - y_{t-1})^2} \tag{28}$$

and

$$N_S^{GLS-\tau} = \frac{\sum_{t=2}^T (y_t^\tau - \rho_c^{t-1}y_1^\tau)^2}{T\sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}. \tag{29}$$

The motivation for these modified tests is clear. If the process is such that $\rho = \rho_c$, it can be expected that the numerators of the N^{GLS} tests are smaller than those based on a misspecified random walk predictor, as in N tests, and, hence, it is easier to reject the null of a unit root. Besides, it is also reasonable to foresee a similar effect if the process is such that $\rho \approx \rho_c$. This modification, however, can alter the limiting distributions, as stated in the following theorem.

Theorem 6 *Let y_t be the process (1) with $\rho = 1$ and let $\rho_c = 1 - c/T$. Then, under assumptions A, B, and C,*

$$
\begin{aligned}
&(i) \quad N_N^{GLS-\mu} \xrightarrow{d} \kappa^{-2}\left\{W^2(1) + c\int_0^1 \{W(r)\}^2 dr\right\},\\
&(ii) \quad N_N^{GLS-\tau} \xrightarrow{d} \kappa^{-2}\left\{\int_0^1 \{U^c(r)\}^2 dr\right\},\\
&(iii) \quad N_S^{GLS-\mu} \xrightarrow{d} \kappa^{-2}\tfrac{1}{2}\left\{\left[\{W_F^c(1)\}^2 + \{W_B^c(1)\}^2\right] + c\left[\int_0^1 \{W_F^c(r)\}^2 dr + \int_0^1 \{W_B^c(r)\}^2 dr\right]\right\},\\
&(iv) \quad N_S^{GLS-\tau} \xrightarrow{d} \kappa^{-2}\left\{\int_0^1 \{K^c(r)\}^2 dr\right\},
\end{aligned}
$$

where

$$U^c(r) = W(r) - rb, b = \lambda W(1) + 3(1-\lambda)\int_0^1 rW(r)dr, \lambda = (1+c)/(1+c+c^2/3);$$

$$W_F^c(r) = V_0^\mu(r) - e^{-cr}V_0^\mu(0), V_0^\mu(r) = W(r) - \beta_0^\mu, \beta_0^\mu = \{1/(2+c)\}\left\{W(1) + c\int_0^1 W(r)dr\right\};$$

$$W_B^c(r) = V_0^\mu(1-r) - e^{-cr}V_0^\mu(1);$$

$$K^c(r) = V_0^\tau(r) - e^{-cr}V_0^\tau(0), V_0^\tau(r) = \{W(r) - \beta_0^\tau - r\beta_1^\tau\},$$

$$\begin{bmatrix} \beta_0^\tau \\ \beta_1^\tau \end{bmatrix} = \begin{bmatrix} c^2+2c & c+c^2/2 \\ c+c^2/2 & 1+c+c^2/3 \end{bmatrix}^{-1} \begin{bmatrix} cW(1) + c^2\int_0^1 W(r)dr \\ (1+c)W(1) + c^2\int_0^1 rW(r)dr \end{bmatrix}.$$

5 Finite sample performance. AR(1) case

This section reports on a Monte Carlo experiment, for the AR(1) case, to compare the empirical power of the N and N^{GLS} tests with some others that appear in the literature. The main conclusion of this section is that the proposed N^{GLS} tests, based on prediction errors, have empirical powers very close to those of the POI tests, used as a benchmark. As a secondary result, we conclude that the results of Section 3 can be used to explain the relative behaviour of the competing tests. Elliott *et al.* (1996) showed that, in absence of deterministic components, currently used tests (apart from the LBI test) have asymptotic power functions very close to the asymptotic Gaussian power envelope. However, their performance in the presence of deterministic components can be very different, even asymptotically. Therefore, only processes with some deterministic component are considered.

The proposed tests are compared with the pivotal τ_{DF}, proposed by Dickey and Fuller (1979); the P_T tests and τ_{GLS} tests, proposed by Elliott *et al.* (1996) for the nonstationary case, and Elliott (1999) for the stationary case (denoted by Q_T and DF-GLSu, respectively, in Elliott, 1999); the weighted symmetric

estimator tests (τ_W), proposed by Park and Fuller (1995) (see also Pantula *et al.*, 1994; and Fuller, 1996); and the MZ_α^{GLS} and MSB^{GLS} tests of Ng and Perron (1999). The MZ_α^{GLS} and MSB^{GLS} tests have only been developed for the nonstationary case, but they have been applied here to the stationary case using the corresponding covariance matrix in the GLS detrending model. The asymptotically POI tests ($S_S(\rho)$ and $S_N(\rho)$), where the true parameter ρ is used as the alternative, are also included as a benchmark. For the sake of clarity, the subscript N is added to those statistics that use GLS estimation with the covariance matrix $\Omega_N(\rho_c)$, and the subscript S when the matrix $\Omega_S(\rho_c)$ is used. The Monte Carlo experiment was conducted for sample sizes $T = 50, 100$; but only $T = 100$ is reported. Conclusions are identical in both sample sizes.

Critical values for some of the tests depend on the value of c in $\rho_c = 1 - c/T$. Since we are interested in tests with high power in a real situation, the selection of c is based on the empirical performance in finite samples. Table 1 summarizes the selected value of c for each test. To construct this table, critical values for each test and a given value of c were obtained from 100,000 replications of the model $y_t = y_{t-1} + a_t$, with $a_t \sim N(0,1)$ and $y_1 = a_1$. Then, the empirical power was evaluated from 100,000 replications of the model $y_t = \rho y_{t-1} + a_t$, with $a_t \sim N(0,1)$ and $y_1 = a_1$ in the nonstationary case, and $y_1 = a_1/\sqrt{1-\rho^2}$ in the stationary case. The selected values of c are those that provide empirical power closer to that of the $S_N(\rho)$ and $S_S(\rho)$ tests for the set of values $\rho = 0.97, 0.95, 0.90, 0.80$. The critical values and empirical power of POI tests are obtained, for each value of ρ, using a similar Monte Carlo experiment.

It can be observed in Table 1 that, in the nonstationary case, the values of c for the $P_{T\text{-}N}$, $\tau_{GLS\text{-}N}$, $\mathrm{MZ}_{\alpha\text{-}N}^{GLS}$, and MSB_N^{GLS} tests agree with the values proposed by the respective authors, using asymptotic criteria. In the stationary case, the values of c, proposed in Elliott (1999) for the $P_{T\text{-}S}$ and $\tau_{GLS\text{-}S}$ tests, also take finite sample performance into consideration and they agree with the values obtained in the present experiment. There were no reported values for $\mathrm{MZ}_{\alpha\text{-}S}^{GLS}$ and MSB_S^{GLS} tests in the stationary case. It should be noted, however, that the values of c for the MSB_S^{GLS} test are different from the remaining competing tests, whereas in the nonstationary case, they coincide with those of $P_{T\text{-}N}$, $\tau_{GLS\text{-}N}$, and $\mathrm{MZ}_{\alpha\text{-}N}^{GLS}$ tests. The N_N^{GLS} test is also included in the stationary case because of its excellent performance in that setting. This N_N^{GLS} test is always calculated as in (27) and the estimation is made with the matrix $\Omega_N(\rho_c)$. To facilitate comparison,

the N tests use the same values of c as the N^{GLS} tests.

[Table 1 about here]

Finite sample critical values for DF tests are in Fuller (1996). Finite sample critical values for $\tau_{GLS\text{-}N}$ tests (linear trend case) and P_T tests are in Elliott *et al.* (1996) for the nonstationary case. Finite sample critical values for the remaining tests have been obtained through 100,000 Monte Carlo replications. All these critical values, for $T = 100$, can be found in Table 2. Tables 3 and 4 show the empirical powers of the competing tests. From the tables, we observe the following:

1. As suggested by the asymptotic results, the N test has high power (similar to the benchmark) when the alternative is close to the unit root (ρ lower than 0.9). For smaller values of ρ, however, this test tends to have lower power than the benchmark.

2. The N tests are, in general, surpassed by the N^{GLS} tests. These N^{GLS} tests have empirical power very close to the benchmark, both for the stationary and nonstationary cases. Although N_N^{GLS} was constructed to have high power for the nonstationary case, it also has very high power for the stationary case.

3. The N^{GLS}, P_T, and MZ_α^{GLS} tests have similar behaviour, with empirical power identical or very close to the benchmark.

4. The MSB^{GLS} test has lower power than the N^{GLS}, P_T, and MZ_α^{GLS} tests for the case with $d_t = \mu$, but it has similar performance when $d_t = \mu + \delta t$. This result is in accordance with the asymptotic theory. As can be seen in (25), the MSB^{GLS} test only uses the information of the second derivative of the Taylor expansion of the asymptotically POI test. This information might be enough to construct an efficient test if $d_t = \mu + \delta t$ (see theorem 4). When $d_t = \mu$, however, more efficient tests might need the use of both first and second derivatives (see theorems 2 and 3) as N^{GLS}, P_T, and MZ_α^{GLS} do.

5. The asymptotic results of Section 3 also help to explain the relative behaviour of the tests based on least squares, τ_W and τ_{GLS}. When $d_t = \mu$, Table 3 shows that the τ_W test has better performance than the τ_{GLS} test in the stationary case, but is worse in the nonstationary case. This result can

be explained by the different efficiency (near the unit root) of the information in each case shown in Section 3. In the stationary case, it has been seen that an efficient use of the information leads to the use of the series, both in direct and reverse order, as τ_W does. On the other hand, in the nonstationary case, it would be more efficient to use only the series in direct order, just as τ_{GLS} does. This result suggests that a modification of the τ_{GLS} test, based on the use of a symmetric estimator, could improve its performance in the stationary case.

6. Regarding the case $d_t = \mu + \delta t$, it has been seen in Section 3 that both the stationary and the non-stationary cases use the same nearly-optimal test. This means that, in the presence of a deterministic linear trend and as far as unit root detection is concerned, the series in reverse order contains the same information as in direct order. This can explain the fact that the performances of τ_W and τ_{GLS} are very similar in this setting.

[Table 2 about here]

[Tables 3 and 4 about here]

6 The general ARMA case

This section extends the proposed tests to a general ARMA case. For the sake of brevity, only N^{GLS} tests are considered. The extension of the proposed tests requires the consistent estimation of κ^2, where $\kappa^2 = \omega^{-2}\sigma_u^2$, with $\omega^2 = \sigma^2\psi(1)^2$. As a consistent estimator of σ_u^2 we can use $\hat{\sigma}_u^2 = \sum_{t=2}^{T}(y_t - y_{t-1})^2/T$ if $d_t = \mu$ or $\hat{\sigma}_u^2 = \sum_{t=2}^{T}(y_t - y_{t-1} - \hat{\delta})^2/T$ if $d_t = \mu + \delta t$. It can be seen that the numerator of $\hat{\sigma}_u^2$ is part of the N^{GLS} test statistic. Therefore, the generalization to a general ARMA case is straightforward. For example, the generalization of $N_N^{GLS-\mu}$ is

$$N_N^{GLS-\mu} = \frac{(y_T^{\mu} - \rho_c^{T-1}y_1^{\mu})^2}{T\hat{\omega}^2} + c\frac{\sum_{t=2}^{T}(y_t^{\mu} - \rho_c^{t-1}y_1^{\mu})^2}{T^2\hat{\omega}^2}, \tag{30}$$

with $\hat{\omega}^2$ a consistent estimator of ω^2. In the same fashion, the remaining statistics of previous sections can be generalized. A second option for the extension to an ARMA case would consist of also incorporating the information of the model into the predictions in the numerator of the test statistics. For example, if u_t is an MA(1), and considering $d_t = \mu$, and $a_t = 0$ for $t < 1$, the prediction of y_t^{μ} from the first observation y_1^{μ} would

be $\hat{y}^{\mu}_{t|1} = \rho_c^{t-2}(\rho_c - \hat{\theta})y_1^{\mu}$, with $\hat{\theta}$ an estimator of θ, instead of $\rho_c^{t-1}y_1^{\mu}$ used in (30). A limited experiment has revealed that this second approach does not have better empirical performance than the previous one. Therefore, for the sake of brevity, further details are omitted.

6.1 Finite sample performance

The finite sample properties of N^{GLS}, P_T, $\mathrm{MZ}_{\alpha}^{GLS}$, τ_{GLS}, and τ_W are compared. The model considered is $y_t = \rho y_{t-1} + a_t - \theta a_{t-1}$, where a_t are independent, identically distributed, standard normal errors. Experiments have been devised for both the nonstationary case ($y_1 = a_1$) and the stationary case, with sample sizes $T = 100$. In the stationary case, and for $\rho < 1$, a set of 50 initial observations were generated and discarded to avoid the effect of initial values. The selection of the autoregressive truncation lag k for τ_{GLS} and τ_W has been made with the BIC and also with the MIC procedures, proposed in Ng and Perron (2000). The MIC procedure is designed to diminish the well-known size distortion in the presence of MA components. The remaining competing tests need the estimation of ω^2. In the literature, there are several well-known estimators of ω^2. The most popular is the autoregressive estimator of the spectral density at frequency zero, defined as $\hat{\omega}^2_{AR} = \hat{\sigma}^2_{AR}(1 - \sum_{j=1}^{k}\hat{\phi}^*_j)^{-2}$, where $\hat{\sigma}^2_{AR}$ is the residual variance obtained from the OLS estimation of $\Delta y_t = d_t + \alpha y_{t-1} + \sum_{j=1}^{k}\phi^*_j \Delta y_{t-j} + \varepsilon_t$. There are also several procedures for the choice of the truncation lag k (see, e.g., Ng and Perron, 1995, 2000). In this experiment, we consider a modification of this estimator, proposed in Ng and Perron (2000), involving detrended data from using local-to-unity GLS. This estimator of ω^2 is $\hat{\omega}^2_{GLS-AR} = \tilde{\sigma}^2_{AR}(1 - \sum_{j=1}^{k}\tilde{\phi}^*_j)^{-2}$, where $\tilde{\sigma}^2_{AR}$ is the residual variance obtained from the OLS estimation of $\Delta y^c_t = \alpha y^c_{t-1} + \sum_{j=1}^{k}\phi^*_j \Delta y^c_{t-j} + v_t$; where y^c_t are GLS detrended data using $\rho_c = 1 - c/T$, with $c = 7$ if $d_t = \mu$, and $c = 13.5$ if $d_t = \mu + \delta t$. Ng and Perron (2000) show that $\hat{\omega}^2_{GLS-AR}$ improves the performance of unit-root tests with respect to $\hat{\omega}^2_{OLS-AR}$. The choice of the truncation lag k in $\hat{\omega}^2_{GLS-AR}$ was also made by the BIC and MIC. When the BIC is used, k is restricted to be $3 \leq k \leq 8$, as in Elliott *et al.* (1996). For the MIC, the restriction $0 \leq k \leq 10(T/100)^{1/4}$ is used, as in Ng and Perron (2000). In this paper, we also consider the estimator, proposed in Sánchez (2000), defined as

$$\hat{\omega}^2_{ARMA} = \hat{\sigma}^2 \frac{(1 - \sum_{i=1}^{q}\hat{\theta}_i)^2}{(1 - \sum_{j=1}^{p}\hat{\phi}^*_j)^2}. \tag{31}$$

where $\hat{\theta}_j$, $\hat{\phi}_i$ are estimates from the error correction regression

$$\Delta y_t^c = \alpha y_{t-1}^c + \sum_{j=1}^{p} \phi_j^* \Delta y_{t-j}^c + a_t + \sum_{j=1}^{q} \theta_i a_{t-j}. \tag{32}$$

The advantage of $\hat{\omega}^2_{ARMA}$ is in the possibility of fitting an ARMA model, whereas $\hat{\omega}^2_{AR}$ and $\hat{\omega}^2_{GLS-AR}$ should always rely on autoregressive approximations. There are, besides, many procedures for the identification of ARMA models (see Koreisha and Pukkila, 1995; and references therein). As shown in Sánchez (2000), the possibility of estimating an ARMA model significantly reduces the size distortion induced by moving average polynomials. The estimation method is exact maximum likelihood when $d_t = \mu$ and marginal likelihood when $d_t = \mu + \delta t$. The computations have been done with the NAG routine G13bef. Since the estimation methods are nonlinear, a set of initial values were needed. The performance of the tests when $\theta \leq 0$ is very robust to these initial values. However, when $\theta > 0$, the tests are highly sensitive to the initial values. Sánchez (2000) analyzes several sets of initial values, both random and fixed. The best performance is obtained with initial values close to the unit circle for both parameters: ρ and θ. The initial values used in this experiment are: $\rho_0 = 0.95$ and $\theta_0 = 0.90$. For each parameter configuration, all tests are based on the same seeds. Empirical size and power, based on 10,000 replications, are summarized in Tables 5 and 6. From these tables, we observe the following:

1. The advantage of the estimator $\hat{\omega}^2_{ARMA}$, shown in Sánchez (2000), is also observed with the proposed N^{GLS} tests. Tests based on BIC tend to have inflated size, specially at $\theta = 0.8$. Tests based on MIC, apart from τ_W and τ_{GLS}, have reasonable size at $\theta = 0.8$ in the stationary case. However, they are undersize at $\theta = 0$ and can have inflated size at negative values of θ. Besides, tests based on MIC can suffer from a severe loss of power when compared with the AR(1) case.

2. The N_N^{GLS} test has similar performance to P_T test and, in the stationary case, to N_S^{GLS} test. These tests also have better performance than the remaining competing tests. As mentioned above, the N^{GLS} and P_T tests can be considered as alternative ways to use the optimal theory to build efficient tests, although only N^{GLS} tests have a prediction error interpretation.

We conclude that, as expected from the intuition, the use of prediction errors allows the construction of unit-root tests with excellent properties of size and power, even with large moving average roots. These two characteristics: intuitive interpretation and good performance, make the N^{GLS} test of potential interest to

practitioners. A more detailed set of critical values can be found in Table 7.Since the performance of N_N^{GLS} and N_S^{GLS} is similar only critical values for N_N^{GLS} are supplied.

[Tables 5 to 7 about here]

7 Concluding remarks

This paper shows a new approach for unit root detection. This so-called prediction-error approach exploits the intuitive relationship between the long-run behaviour of a process and the presence of a unit root, and shows the link between our intuition and the optimal detection of such a root.

From a theoretical point of view, this prediction-error approach appears very useful to understand how (nearly) optimal tests use the information. It is seen in this paper that nearly-optimal tests in the vicinity of the null can be constructed with the prediction errors of a predictor that assumes the presence of a unit root, where the predictions are evaluated from the origin of the series. When the series is covariance-stationary under the alternative, a more efficient use of the information (in the vicinity of the null) is obtained if the time-reversal series is also considered. In that case, both ends of the series are used as origin of the predictions.

From a practical point of view, the proposed approach allows the construction of tests for a general ARMA setting with excellent properties of size and power. These N^{GLS} tests are based on a predictor constructed under a local alternative. A Monte Carlo experiment for the ARMA(1,1) model shows that the N^{GLS} tests have an excellent relative performance among a set of tests currently in common use.

The present approach, not only have an intuitive interpretation, but good theoretical and empirical properties. Then, it can be used to design nonstationarity tests under different circumstances than those exposed here. Outliers robust unit root tests, or nonparametric tests, could be some examples of promising directions for future research.

References

[1] Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Rev. Econ. Stud.* 53, 369-384.

[2] Box, G.E.P. and G.M. Jenkins (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

[3] Dickey, D.A. and W.A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* 74, 427-431.

[4] Dufour, J.M. and M.L. King (1991). Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *J. Econom.* 47, 115-143.

[5] Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *Intern. Econ. Rev.* 40, 767-783.

[6] Elliott, G., T.J. Rothenberg and J.H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813-836.

[7] Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

[8] Fuller, W.A. (1996). *Introduction to Statistical Time Series.* Second edition. John Wiley, New York.

[9] Hamilton, J.D. (1994). *Time Series Analysis.* Princeton University Press, New York.

[10] Hwang, J. and P. Schmidt (1996). Alternative methods of detrending and the power of unit-root tests. *J. Econom.* 71, 227-248.

[11] Kadiyala, K.R. (1970). Testing for the independence of regression disturbances. *Econometrica* 38, 97-117.

[12] Koreisha, S.G. and T. Pukkila (1995). A comparison between different order-determination criteria for identification of ARIMA models. *J. Busin. Econom. Statis.* 13, 127-131.

[13] Leybourne, S.J. (1995). Testing for unit roots using forward and reverse Dickey-Fuller regressions. *Oxford Bull. Econ. Stat.* 57, 559-571.

[14] Nabeya, S. and K. Tanaka (1990). Limiting power of unit-root tests in time-series regression. *J. Economet.* 46, 247-271.

[15] Ng, S. and P. Perron (1995). Unit-root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *J. Amer. Statist. Assoc.* 90, 268-281.

[16] Ng, S. and P. Perron (2000). Lag length selection and the construction of unit-root tests with good size and power. *Econometrica.* Forthcoming.

[17] Pantula, S.G., G. Gonzalez-Farias and W.A. Fuller (1994). A comparison of unit-root test criteria. *J. Busin. Econom. Statis.* 12, 449-459.

[18] Park, H.J. and W.A. Fuller (1995). Alternative estimators and unit root tests for the autoregressive process. *J. Time Ser. Anal.* 16, 415-429.

[19] Perron, P. and S. Ng (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Rev. Econ. Stud.* 63, 435-463.

[20] Sánchez, I. (2000). Spectral density estimators at frequency zero for nonstationary tests in ARMA models. Working Paper. Universidad Carlos III de Madrid.

[21] Sargan, J.D. and A. Bhargava (1983). Testing the residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 41, 153-174.

[22] Sen, D.L. and D.A. Dickey (1987). Symmetric test for second differencing in univariate time series. *J. Bus. Econ. Stat.* 5, 463-473.

[23] Shin, D.W. and W.A. Fuller (1998). Unit-root tests based on unconditional maximum likelihood estimation for the autoregressive moving average. *J. Time Ser. Anal.* 19, 591-599.

[24] Shin, D.W. and J.H. Lee (2000). Consistency of the maximum likelihood estimators for nonstationary ARMA regressions with time trends. *J. Statist. Plann. Inference*, 87, 55-68.

[25] Shin, D.W. and B.S. So (1997). Semiparametric unit-root tests based on symmetric estimators. *Statist. Probab. Lett.* 33, 177-184.

[26] Stock. J.H. (1990). A class of tests for integration and cointegration. Manuscript. Harvard University.

[27] Stock, J.H. (1994). Unit roots, structural breaks, and trends. In: R. F. Engle and D.L. McFadden, Eds., *Handbook of Econometrics.* Vol. IV. Elsevier, France.

[28] Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory.* John Wiley, New York.

[29] Weiss, G. (1975). Time-reversibility of linear stochastic processes. *J. Appl. Probab.* 12, 831-836.

[30] Xiao, Z. and P.C.B. Phillips (1998). An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy. *Econometric Journal* 1, 27-43.

[31] Yap, S.F. and G.C. Reinsel (1995). Results on estimation and testing for a unit root in the nonstationary autoregressive moving-average model. *J. Time Ser. Anal.* 16, 339-353.

APPENDIX

Proof of theorem 1:

(i) The limiting distribution of C_1^{μ} is a straightforward application of the functional central limit theorem (FCLT) to $S_t = \sum_{j=2}^{t} u_j$.

(ii) To obtain the limiting distribution of C_1^{τ}, it is convenient to rewrite $y_t - y_1 - \hat{\delta}(t-1) = x_t - x_1 - (\hat{\delta} - \delta)(t-1)$. Since $\hat{\delta} = (T-1)^{-1}(y_T - y_1) = \delta + (T-1)^{-1}\sum_{t=2}^{T} u_j$, it can be shown that $\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} \sigma\psi(1)W(1)$. By the FCLT and using $T^{-(k+1)}\sum_{t=1}^{T} t^k \to (k+1)^{-1}$, $k = 0, 1, ...$ as $T \to \infty$, it is verified that: $T^{-1/2}\left\{y_t - y_1 - \hat{\delta}(t-1)\right\} = T^{-1/2}\left\{x_t - x_1 - (\hat{\delta} - \delta)(t-1)\right\} \xrightarrow{d} \omega\{W(r) - rW(1)\} \equiv \omega W^{\tau}(r)$. Then, from the continuous mapping theorem (CMT), the distribution can be obtained.

(iii) By the FCLT, $T^{-1/2}(y_{T-t} - y_T) \xrightarrow{d} \omega\{W(1-r) - W(1)\} \equiv \omega W_B^0(r)$. From the CMT, the limiting distribution is obtained. □

Proof of theorem 2:

The GLS residuals of the numerator of S_N are $\hat{x}_{N,t} = y_t - \hat{\mu}_N$, where

$$\hat{\mu}_N = \frac{y_1 + (1-\rho)\sum_{t=2}^{T}(y_t - \rho y_{t-1})^2}{1 + (T-1)(1-\rho)^2}. \tag{A.1}$$

Likewise, the residuals of the denominator of S_N are $\hat{x}_{1,t} = y_t - \hat{\mu}_1$, where $\hat{\mu}_1$ is the GLS estimator using $\Omega_N^{-1}(1)$. Then, $\hat{\mu}_1 = y_1$. The first derivative is $S_N'(1) = \left(\hat{\mathbf{x}}_1'\Omega_N^{-1}(1)\hat{\mathbf{x}}_1\right)^{-1} \partial\left(\hat{\mathbf{x}}_N'\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N\right)/\partial\rho\big|_{\rho=1}$, where it can be verified that $\hat{\mathbf{x}}_1'\Omega_N^{-1}(1)\hat{\mathbf{x}}_1 = \sum_{t=2}^{T}(y_t - y_{t-1})^2$. Then, the numerator of $S_N'(1)$ can be written as $\partial\left(\hat{\mathbf{x}}_N'\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N\right)/\partial\rho\big|_{\rho=1} = 2\left(\partial\hat{\mathbf{x}}_N'/\partial\rho\right)\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N\big|_{\rho=1} + \hat{\mathbf{x}}_N'\left(\partial\Omega_N^{-1}(\rho)/\partial\rho\right)\hat{\mathbf{x}}_N\big|_{\rho=1}$. Since $(\partial\hat{\mathbf{x}}_N'/\partial\rho) = -(\partial\hat{\mu}_N/\partial\rho)\,\mathbf{e}'$and also

$$\mathbf{e}'\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N\big|_{\rho=1} = 0, \tag{A.2}$$

it can be shown that $\partial\left(\hat{\mathbf{x}}_N'\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N\right)/\partial\rho\big|_{\rho=1} = \hat{\mathbf{x}}_N'\left(\partial\Omega_N^{-1}(\rho)/\partial\rho\right)\hat{\mathbf{x}}_N\big|_{\rho=1}$. From (7) it can be shown that

$$\left.\frac{\partial\Omega_N^{-1}(\rho)}{\partial\rho}\right|_{\rho=1} = 2\mathbf{L}'\mathbf{L} - (\mathbf{L} + \mathbf{L}') = \Omega_N^{-1}(1) - diag(0, ..., 0, 1), \tag{A.3}$$

where *diag* expresses a diagonal matrix. Also, from (A.1), it holds that $\hat{\mathbf{x}}_{N,t}(1) = (y_t - y_1)$. Therefore, it can be obtained that $S'_N(1) = 1 - \left\{\sum_{t=2}^{T}(y_t - y_{t-1})^2\right\}^{-1}(y_T - y_1)^2$. For the second derivative, it can be written that

$$\frac{\partial^2}{\partial\rho^2}\hat{\mathbf{x}}'_N\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N = 2\frac{\partial^2\hat{\mathbf{x}}'_N}{\partial\rho^2}\Omega_N^{-1}(\rho)\hat{\mathbf{x}}_N + 4\frac{\partial\hat{\mathbf{x}}'_N}{\partial\rho}\frac{\partial\Omega_N^{-1}(\rho)}{\partial\rho}\hat{\mathbf{x}}_N + 2\frac{\partial\hat{\mathbf{x}}'_N}{\partial\rho}\Omega_N^{-1}(\rho)\frac{\partial\hat{\mathbf{x}}_N}{\partial\rho} + \hat{\mathbf{x}}'_N\frac{\partial^2\Omega_N^{-1}(\rho)}{\partial\rho^2}\hat{\mathbf{x}}_N. \quad \text{(A.4)}$$

Applying (A.2), the first term in (A.4) is null at $\rho = 1$. Also, by (A.3) and applying that $\partial\hat{\mu}_N/\partial\rho|_{\rho=1} = (y_1 - y_T)$, it can be obtained that $4\left(\partial\hat{\mathbf{x}}'_N/\partial\rho\right)\left(\partial\Omega_N^{-1}(\rho)/\partial\rho\right)\hat{\mathbf{x}}_N\big|_{\rho=1} = -4(y_T - y_1)^2$. Similarly, since $\mathbf{e}'\Omega_N^{-1}(1)\mathbf{e} = 1$ it holds that $\left(\partial\hat{\mathbf{x}}'_N/\partial\rho\right)\Omega_N^{-1}(\rho)\left(\partial\hat{\mathbf{x}}_N/\partial\rho\right)\big|_{\rho=1} = (y_T - y_1)^2$. To solve the last term in (A.4), it can be applied that

$$\left.\frac{\partial^2\Omega_N^{-1}(\rho)}{\partial\rho^2}\right|_{\rho=1} = 2\mathbf{L}'\mathbf{L}, \quad \text{(A.5)}$$

and, therefore, $\hat{\mathbf{x}}'_N\left(\partial^2\Omega_N^{-1}(\rho)\partial\rho^2\right)\hat{\mathbf{x}}_N\big|_{\rho=1} = 2\sum_{t=2}^{T-1}(y_t - y_1)^2$ and the theorem holds. □

Proof of theorem 3:

The GLS residuals of the numerator of S_S are $\hat{x}_{S,t}(\rho) = y_t - \hat{\mu}_S$. Then,

$$\hat{\mu}_S = \frac{(1-\rho)y_1 + (1-\rho)y_T + \sum_{t=2}^{T-1}(1-\rho)^2 y_t}{2(1-\rho) + (T-2)(1-\rho)^2} = \frac{y_1 + y_T + \sum_{t=2}^{T-1}(1-\rho)y_t}{2 + (T-2)(1-\rho)}. \quad \text{(A.6)}$$

Following the same arguments as in theorem 2, it holds that

$$\mathbf{e}'\Omega_S^{-1}(\rho)\hat{\mathbf{x}}_S\big|_{\rho=1} = 0. \quad \text{(A.7)}$$

Also, by (9), it holds that

$$\left.\frac{\partial\Omega_S^{-1}(\rho)}{\partial\rho}\right|_{\rho=1} = 2\mathbf{L}'\mathbf{L} - (\mathbf{L} + \mathbf{L}') - 2\mathbf{e}\mathbf{e}' = \Omega_S^{-1}(1) - diag(1, 0, ..., 0, 1). \quad \text{(A.8)}$$

Then, it can be obtained that $S'_S(1) = 1 - \left\{2\sum_{t=2}^{T}(y_t - y_{t-1})^2\right\}^{-1}(y_T - y_1)^2$. Likewise, in the second derivative, it can be obtained that

$$\frac{\partial^2}{\partial\rho^2}\hat{\mathbf{x}}'_S\Omega_S^{-1}(\rho)\hat{\mathbf{x}}_S = 2\frac{\partial^2\hat{\mathbf{x}}'_S}{\partial\rho^2}\Omega_S^{-1}(\rho)\hat{\mathbf{x}}_S + 4\frac{\partial\hat{\mathbf{x}}'_S}{\partial\rho}\frac{\partial\Omega_S^{-1}(\rho)}{\partial\rho}\hat{\mathbf{x}}_S + 2\frac{\partial\hat{\mathbf{x}}'_S}{\partial\rho}\Omega_S^{-1}(\rho)\frac{\partial\hat{\mathbf{x}}_S}{\partial\rho} + \hat{\mathbf{x}}'_S\frac{\partial^2\Omega_S^{-1}(\rho)}{\partial\rho^2}\hat{\mathbf{x}}_S. \quad \text{(A.9)}$$

Similarly to the nonstationary case, it can be obtained that the first term in (A.9) is null at $\rho = 1$. Applying (A.8) it can also be seen that the second term in (A.9) is null in $\rho = 1$. Since $\mathbf{e}'\Omega_S^{-1}(\mathbf{1})\mathbf{e} = 0$, the third term

is also null at unity. To solve the fourth term in (A.9), it can be verified that

$$\left.\frac{\partial^2 \Omega_S^{-1}(\rho)}{\partial \rho^2}\right|_{\rho=1} = 2\mathbf{L}'\mathbf{L} - 2\mathbf{e}\mathbf{e}' = diag(0,2,2,...,2,0). \tag{A.10}$$

Hence, $\hat{\mathbf{x}}_S' \left(\partial^2 \Omega_S^{-1}(\rho)\partial\rho^2\right) \hat{\mathbf{x}}_S\big|_{\rho=1} = (1/2)\sum_{t=2}^{T-1}(y_t - y_1)^2 + (1/2)\sum_{t=2}^{T-1}(y_T - y_t)^2 - \sum_{t=2}^{T-1}(y_t - y_1)(y_T - y_t)$, and the theorem holds. □

Proof of theorem 4:

The proof is given firstly for the stationary case and then for the nonstationary case. The GLS residuals in the numerator of S_S are $\hat{x}_{S,t} = y_t - \hat{\mu}_\rho - \hat{\delta}_\rho t$, where $\hat{\mu}_\rho$ and $\hat{\delta}_\rho$ are the GLS estimator using the matrix $\Omega_S^{-1}(\rho)$. Likewise, the residuals of the denominator are $\hat{x}_{1,t} = y_t - \hat{\mu}_1 - \hat{\delta} t$, where $\hat{\mu}_1$ and $\hat{\delta}$ are the GLS estimator using the matrix $\Omega^{-1}(1)$. Then, $\hat{\mu}_1 = y_1 - \hat{\delta}$ and $\hat{\delta} = (y_T - y_1)(T-1)^{-1}$. It can, then, be verified that $\hat{\mathbf{x}}_1'\Omega^{-1}(1)\hat{\mathbf{x}}_1 = \sum_{t=2}^{T}(y_t - y_{t-1} - \hat{\delta})^2$. The numerator of $S_S'(1)$ can be written as

$$\left.\frac{\partial}{\partial\rho}\hat{\mathbf{x}}_S'\Omega_S^{-1}(\rho)\hat{\mathbf{x}}_S\right|_{\rho=1} = 2\left.\frac{\partial\hat{\mathbf{x}}_S'}{\partial\rho}\Omega_S^{-1}(\rho)\hat{\mathbf{x}}_S\right|_{\rho=1} + \left.\hat{\mathbf{x}}_S'\frac{\partial\Omega_S^{-1}(\rho)}{\partial\rho}\hat{\mathbf{x}}_S\right|_{\rho=1}. \tag{A.11}$$

After some algebraic manipulation, it can be obtained that

$$\begin{bmatrix} \hat{\mu}_\rho \\ \hat{\delta}_\rho \end{bmatrix} = \begin{bmatrix} \dfrac{d_3 n_1 - d_2 n_2}{D} \\ \dfrac{d_1 n_2 - (1-\rho) d_2 n_1}{D} \end{bmatrix},$$

where $n_1 = y_T + y_1 + (1-\rho)\sum_{t=2}^{T-1} y_t$, $n_2 = y_1(1-2\rho) + \{T - (T-1)\rho\}\, y_T + (1-\rho)^2 \sum_{t=2}^{T-1} t y_t$, $d_1 = (1+\rho) + (T-1)(1-\rho)$, $d_2 = (1+\rho) + \sum_{t=2}^{T}\{t - (t-1)\rho\}$, $d_3 = (1-\rho^2) + \sum_{t=2}^{T}\{t - (t-1)\rho\}^2$, $D = d_1 d_3 - (1-\rho) d_2^2$. It can then be obtained that, when $\rho = 1$, $\hat{\mu}_\rho|_{\rho=1} = y_1 - (T-1)^{-1}(y_T - y_1)$, $\hat{\delta}_\rho|_{\rho=1} = (T-1)^{-1}(y_T - y_1) \equiv \hat{\delta}$, $\partial\hat{\mu}_\rho/\partial\rho|_{\rho=1} = (y_T - y_1)(T/4 - 1/2) - (1/2)\sum_{t=2}^{T-1} y_t$, and $\partial\hat{\delta}_\rho/\partial\rho|_{\rho=1} = 0$. Then, when $\rho = 1$, $\hat{x}_{S,t}|_{\rho=1} = y_t - y_1 - \hat{\delta}(t-1)$. After some algebra, it can be seen that $(\partial\hat{\mathbf{x}}_S'/\partial\rho)\,\Omega_S^{-1}(\rho)\big|_{\rho=1} = (0,0,...,0)$. Then, the first term in (A.11) is null. Applying the result (A.8), it can be seen that the second term in (A.11) is equal to one. For the second derivative, we can make a decomposition similar to (A.9). Following the same arguments as in the first derivative, it can be seen that the second and third terms of this decomposition are zero. It is easy to verify that the term $(\partial^2\hat{\mathbf{x}}_S'/\partial\rho^2)\Omega_S^{-1}(\rho)\big|_{\rho=1}$ is a vector with all elements equal to zero, except for the first and the last one. The term $\hat{\mathbf{x}}_S|_{\rho=1}$, however, is a vector with first and last elements equal to zero. Therefore, the first term in this decomposition is also null. By (A.10), we obtain that the fourth term of this decomposition is $2\sum_{t=2}^{T-1}\left\{y_t - y_1 - \hat{\delta}(t-1)\right\}^2$ and, then, the theorem holds in the stationary case. In

the nonstationary case, we also have a decomposition like (A.11). The GLS residuals in the numerator of S_N are $\hat{x}_{N,t} = y_t - \hat{\mu}_\rho - \hat{\delta}_\rho t$, where $\hat{\mu}_\rho$ and $\hat{\delta}_\rho$ are now the GLS estimator using the matrix $\Omega_N^{-1}(\rho)$. After some algebra, we obtain that

$$\begin{bmatrix} \hat{\mu}_\rho \\ \hat{\delta}_\rho \end{bmatrix} = \begin{bmatrix} \dfrac{b_3 m_1 - b_2 m_2}{F} \\ \dfrac{b_1 m_2 - b_2 m_1}{F} \end{bmatrix},$$

where $m_1 = (1-\rho)y_T + (1-\rho+\rho^2)y_1 + (1-\rho)^2 \sum_{t=2}^{T-1} y_t$, $m_2 = (1-\rho)^2 y_1 + \{T-(T-1)\rho\} y_T + (1-\rho)^2 \sum_{t=2}^{T-1} t y_t$, $b_1 = 1 + (T-1)(1-\rho)^2$, $b_2 = 1 + (1-\rho)\left\{(T-\rho) + (1-\rho)\sum_{t=2}^{T-1} t\right\}$, $b_3 = (1-\rho^2) + \sum_{t=2}^{T} \{t-(t-1)\rho\}^2$, $F = b_1 b_3 - b_2^2$. Then, it can be obtained that, when $\rho = 1$, $\hat{\mu}_\rho|_{\rho=1} = y_1 - (T-1)^{-1}(y_T - y_1)$, $\hat{\delta}_\rho|_{\rho=1} = (T-1)^{-1}(y_T - y_1) \equiv \hat{\delta}$, $\partial\hat{\mu}_\rho/\partial\rho|_{\rho=1} = 0$, and $\partial\hat{\delta}_\rho/\partial\rho|_{\rho=1} = 0$. Then, when $\rho = 1$, $\hat{x}_{N,t}|_{\rho=1} = y_t - y_1 - \hat{\delta}(t-1)$. Therefore, $\partial\hat{\mathbf{x}}_N'/\partial\rho|_{\rho=1}$ is a vector of zeros. Using the same arguments as in the stationary case, we have $S_N'(1) = 1$. For the second derivative, we have a decomposition similar to (A.4), where it is easy to check that the second and third terms are null. It can also be verified, in this nonstationary case, that $(\partial^2\hat{\mathbf{x}}_N'/\partial\rho^2)\Omega_N^{-1}(\rho)\big|_{\rho=1}$ is a vector with all elements equal to zero, except the first and the last ones, and that $\hat{\mathbf{x}}_N|_{\rho=1}$ is a vector with first and last elements equal to zero. Therefore, the first term of this decomposition is null. By (A.5), we obtain that the fourth term of this decomposition is $2\sum_{t=2}^{T-1}\left\{y_t - y_1 - \hat{\delta}(t-1)\right\}^2$ and, then, the theorem also holds in the nonstationary case. □

Proof of theorem 6

(*i*) In the $N_N^{GLS-\mu}$ statistic, $y_t^\mu = y_t - \hat{\mu}^c$, where $\hat{\mu}^c$ is estimated by GLS using the matrix $\Omega_N^{-1}(\rho_c)$, $\rho_c = 1 - c/T$. From Elliott *et al.* (1996), it can be obtained that $T^{-1/2}y_t^\mu \xrightarrow{d} \omega W(r)$. Then, applying that $\lim_{T\to\infty}\rho_c^T = e^{-c}$, $\lim_{T\to\infty}\rho_c^{[Tr]} = e^{-rc}$, and the CMT, it can be verified that $T^{-1}\hat{\sigma}_u^{-2}(y_T^\mu - \rho_c^{T-1}y_1^\mu)^2 \xrightarrow{d} \kappa^{-2}\{W(1)\}^2$ and $T^{-2}\hat{\sigma}_u^{-2}\sum_{t=2}^{T}(y_t^\mu - \rho_c^{t-1}y_1^\mu)^2 \xrightarrow{d} \kappa^{-2}\left[\int_0^1 \{W(r)\}^2 dr\right]$.

(*ii*) In the $N_N^{GLS-\tau}$ test, $y_t^\tau = y_t - \hat{\mu}^c - \hat{\delta}^c t$, where the estimation is by GLS using the matrix $\Omega_N^{-1}(\rho_c)$. From Elliott *et al.* (1996), $T^{-1/2}y_t^\tau \xrightarrow{d} \omega U^c(r)$. It can be checked that $U^c(0) = 0$. Then, $T^{-1/2}(y_t^\tau - \rho_c^{t-1}y_1^\tau) \xrightarrow{d} \omega U^c(r)$. From the CMT, $T^{-2}\hat{\sigma}_u^{-2}\sum_{t=2}^{T}(y_t^\tau - \rho_c^{t-1}y_1^\tau)^2 \xrightarrow{d} \kappa^{-2}\left[\int_0^1 \{U^c(r)\}^2 dr\right]$

(*iii*) In the $N_S^{GLS-\mu}$ test, $y_t^\mu = y_t - \hat{\mu}^c$, where $\hat{\mu}^c$ is estimated by GLS using $\Omega_S^{-1}(\rho_c)$. From Lemma 3 in Elliott (1999), $T^{-1/2}y_t^\mu \xrightarrow{d} \omega V_0^\mu(r)$. Then, $T^{-1/2}(y_T^\mu - \rho_c^{T-1}y_1^\mu) \xrightarrow{d} \omega\{V_0^\mu(1) - e^{-c}V_0^\mu(0)\} \equiv \omega W_F^c(1)$ and

$T^{-1/2}(y_1^\mu - \rho_c^{T-1} y_T^\mu) \xrightarrow{d} \omega \{V_0^\mu(0) - e^{-c} V_0^\mu(1)\} \equiv \omega W_B^c(1)$. Therefore, from the CMT,

$$\frac{(y_T^\mu - \rho_c^{T-1} y_1^\mu)^2}{2T\hat{\sigma}_u^2} + \frac{(y_1^\mu - \rho_c^{T-1} y_T^\mu)^2}{2T\hat{\sigma}_u^2} \xrightarrow{d} \frac{1}{2}\kappa^{-2}\left[\{W_F^c(1)\}^2 + \{W_B^c(1)\}^2\right].$$

Similarly, it can also be verified that $T^{-1/2}(y_t^\mu - \rho_c^{t-1} y_1^\mu) \xrightarrow{d} \omega \{V_0^\mu(r) - e^{-cr} V_0^\mu(0)\} \equiv \omega W_F^c(r)$ and $T^{-1/2}(y_{T-t}^\mu - \rho_c^{t-1} y_T^\mu) \xrightarrow{d} \omega \{V_0^\mu(1-r) - e^{-cr} V_0^\mu(1)\} \equiv \omega W_B^c(r)$. Hence,

$$\frac{\sum_{t=2}^T (y_t^\mu - \rho_c^{t-1} y_1^\mu)^2}{2T^2\hat{\sigma}_u^2} + \frac{(y_{T-t}^\mu - \rho_c^{t-1} y_T^\mu)^2}{2T^2\hat{\sigma}_u^2} \xrightarrow{d} \kappa^{-2}\frac{1}{2}\left[\int_0^1 \{W_F^c(r)\}^2\, dr + \int_0^1 \{W_B^c(r)\}^2\, dr\right].$$

(iv) In the $N_S^{GLS-\tau}$ tests, $y_t^\tau = y_t - \hat{\mu}^c - \hat{\delta}^c t$, where the estimation is by GLS with $\Omega_S^{-1}(\rho_c)$. From Elliott (1999), $T^{-1/2} y_t^\tau \xrightarrow{d} \omega V_0^\tau(r)$. Therefore, $T^{-1/2}(y_t^\tau - \rho_c^{t-1} y_1^\tau) \xrightarrow{d} \omega \{V_0^\tau(r) - e^{-cr} V_0^\tau(0)\} \equiv \omega K^c(r)$. Then, $T^{-2}\hat{\sigma}_u^{-2} \sum_{t=1}^{T-1} (y_t^\tau - \rho_c^{t-1} y_1^\tau)^2 \xrightarrow{d} \kappa^{-2} \int_0^1 \{K^c(r)\}^2\, dr$. □

Table 1: Selected values of c for different unit-root tests

Stationary case

	N_S^{GLS}	N_N^{GLS}	$P_{T\text{-}S}$	$\tau_{GLS\text{-}S}$	$MZ_{\alpha\text{-}S}^{GLS}$	MSB_S^{GLS}
$d_t = \mu$	10	13.5	10	10	10	7
$d_t = \mu + \delta t$	10	20	10	10	10	7

Non-stationary case

	N_N^{GLS}	$P_{T\text{-}N}$	$\tau_{GLS\text{-}N}$	$MZ_{\alpha\text{-}N}^{GLS}$	MSB_N^{GLS}
$d_t = \mu$	8	7	7	7	7
$d_t = \mu + \delta t$	8	13.5	13.5	13.5	13.5

Table 2: Empirical 5% critical values. $T = 100$. 100,000 Monte Carlo rep.

Stationary and Non-stationary case

	τ_W	τ_{DF}
$d_t = \mu$	-2.56	-2.90
$d_t = \mu + \delta t$	-3.28	-3.45

Stationary case

	N_S	N_S^{GLS}	N_N	N_N^{GLS}	$P_{T\text{-}S}$	$MZ_{\alpha\text{-}S}^{GLS}$	MSB_S^{GLS}	$\tau_{GLS\text{-}S}$
$d_t = \mu$	0.7605	0.4292	0.8702	0.6406	4.71	-11.29	0.2060	-2.77
$d_t = \mu + \delta t$	0.0380	0.0269	0.0380	0.0268	2.94	-16.72	0.1736	-3.21

Non-stationary case

	N_N	N_N^{GLS}	$P_{T\text{-}N}$	$MZ_{\alpha\text{-}N}^{GLS}$	MSB_N^{GLS}	$\tau_{GLS\text{-}N}$
$d_t = \mu$	0.5472	0.4581	3.11	-8.89	0.2286	-2.14
$d_t = \mu + \delta t$	0.0380	0.0305	5.64	-15.76	0.1759	-3.03

Table 3: AR(1) case. Empirical power at size 5%. $T = 100$. $d_t = \mu$. 100,000 rep.

Stationary case

ρ	$S_S(\rho)$	N_S	N_S^{GLS}	N_N	N_N^{GLS}	τ_W	$P_{T\text{-}S}$	$\tau_{GLS\text{-}S}$	$MZ_{\alpha\text{-}S}^{GLS}$	MSB_S^{GLS}	τ_{DF}
0.97	0.12	0.12	0.12	0.11	0.12	0.12	0.12	0.09	0.12	0.11	0.08
0.95	0.20	0.19	0.20	0.18	0.20	0.20	0.20	0.15	0.20	0.19	0.13
0.90	0.52	0.45	0.52	0.43	0.52	0.52	0.52	0.39	0.52	0.50	0.34
0.80	0.97	0.82	0.97	0.76	0.97	0.97	0.97	0.92	0.97	0.97	0.87

Non-stationary case

ρ	$S_N(\rho)$	N_N	N_N^{GLS}	τ_W	$P_{T\text{-}N}$	$\tau_{GLS\text{-}N}$	$MZ_{\alpha\text{-}N}^{GLS}$	MSB_N^{GLS}	τ_{DF}
0.97	0.17	0.17	0.16	0.15	0.17	0.17	0.16	0.16	0.08
0.95	0.29	0.29	0.29	0.26	0.29	0.28	0.28	0.27	0.11
0.90	0.67	0.65	0.67	0.61	0.66	0.66	0.66	0.64	0.29
0.80	0.98	0.93	0.98	0.98	0.96	0.98	0.98	0.97	0.86

Table 4: AR(1) case. Empirical power at size 5%. $T = 100$. $d_t = \mu + \delta t$. 100,000 rep.

Stationary case

ρ	$S_S(\rho)$	N_S	N_S^{GLS}	N_N^{GLS}	τ_W	$P_{T\text{-}S}$	τ_{GLS}	$MZ_{\alpha\text{-}S}^{GLS}$	MSB_S^{GLS}	τ_{DF}
0.97	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06
0.95	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.09
0.90	0.26	0.24	0.25	0.25	0.25	0.26	0.24	0.26	0.26	0.20
0.80	0.78	0.64	0.77	0.77	0.78	0.77	0.76	0.78	0.78	0.66

Non-stationary case

ρ	$S_N(\rho)$	N_N	N_N^{GLS}	τ_W	$P_{T\text{-}N}$	$\tau_{GLS\text{-}N}$	$MZ_{\alpha\text{-}N}^{GLS}$	MSB_N^{GLS}	τ_{DF}
0.97	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
0.95	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.08
0.90	0.30	0.29	0.30	0.29	0.30	0.30	0.30	0.30	0.19
0.80	0.82	0.72	0.82	0.81	0.82	0.82	0.82	0.82	0.64

Table 5: Empirical size and power for 5% level. T=100. Model: $(1-\rho B)(y_t - dt) = (1-\theta B)a_t$. Stationary case. 5000 rep.

Model	Test	ρ	$d_t = \mu$					$d_t = \mu + \delta t$				
			θ					θ				
			-0.8	-0.5	0	0.5	0.8	-0.8	-0.5	0	0.5	0.8
ARMA	N_N^{GLS}	1.00	0.058	0.059	0.057	0.037	0.050	0.069	0.068	0.059	0.029	0.066
		0.90	0.571	0.568	0.526	0.286	0.228	0.301	0.297	0.244	0.103	0.159
	N_S^{GLS}	1.00	0.059	0.060	0.058	0.036	0.049	0.068	0.067	0.058	0.028	0.065
		0.90	0.571	0.568	0.518	0.261	0.220	0.301	0.300	0.245	0.103	0.158
	MZ_α^{GLS}	1.00	0.072	0.073	0.072	0.044	0.055	0.098	0.097	0.087	0.041	0.073
		0.90	0.623	0.619	0.574	0.314	0.238	0.407	0.398	0.335	0.137	0.171
	P_T	1.00	0.055	0.056	0.056	0.037	0.051	0.061	0.064	0.058	0.030	0.068
		0.90	0.564	0.563	0.521	0.287	0.232	0.299	0.294	0.250	0.110	0.164
AR(MIC)	N_N^{GLS}	1.00	0.085	0.050	0.033	0.041	0.056	0.088	0.051	0.028	0.027	0.059
		0.90	0.295	0.273	0.270	0.235	0.296	0.137	0.107	0.077	0.097	0.205
	N_S^{GLS}	1.00	0.087	0.048	0.033	0.036	0.050	0.089	0.049	0.028	0.028	0.061
		0.90	0.277	0.255	0.244	0.210	0.279	0.141	0.106	0.077	0.096	0.205
	MZ_α^{GLS}	1.00	0.096	0.060	0.042	0.045	0.058	0.115	0.073	0.043	0.045	0.077
		0.90	0.317	0.303	0.300	0.251	0.311	0.193	0.160	0.128	0.149	0.233
	P_T	1.00	0.080	0.047	0.034	0.039	0.055	0.079	0.048	0.026	0.029	0.065
		0.90	0.294	0.270	0.263	0.237	0.301	0.141	0.112	0.083	0.111	0.218
	τ_{GLS}	1.00	0.027	0.024	0.021	0.032	0.082	0.030	0.029	0.028	0.033	0.109
		0.90	0.169	0.173	0.192	0.212	0.401	0.118	0.119	0.127	0.142	0.299
	τ_W	1.00	0.020	0.024	0.022	0.036	0.106	0.007	0.016	0.024	0.045	0.181
		0.90	0.132	0.181	0.241	0.273	0.515	0.033	0.060	0.121	0.169	0.449
AR(BIC)	N_N^{GLS}	1.00	0.142	0.106	0.086	0.086	0.357	0.255	0.166	0.127	0.095	0.375
		0.90	0.580	0.556	0.536	0.528	0.894	0.479	0.414	0.340	0.294	0.711
	N_S^{GLS}	1.00	0.150	0.112	0.085	0.085	0.351	0.255	0.169	0.126	0.094	0.373
		0.90	0.582	0.559	0.526	0.503	0.889	0.479	0.415	0.341	0.293	0.706
	MZ_α^{GLS}	1.00	0.158	0.127	0.099	0.099	0.391	0.293	0.207	0.157	0.131	0.455
		0.90	0.609	0.597	0.574	0.561	0.924	0.542	0.487	0.421	0.382	0.785
	P_T	1.00	0.139	0.104	0.085	0.085	0.359	0.241	0.155	0.122	0.101	0.388
		0.90	0.575	0.553	0.532	0.529	0.904	0.473	0.417	0.341	0.308	0.727
	τ_{GLS}	1.00	0.065	0.052	0.052	0.074	0.437	0.079	0.059	0.049	0.083	0.564
		0.90	0.348	0.331	0.313	0.442	0.965	0.245	0.220	0.189	0.303	0.896
	τ_W	1.00	0.067	0.037	0.051	0.122	0.690	0.076	0.033	0.052	0.148	0.845
		0.90	0.410	0.346	0.456	0.749	0.998	0.242	0.148	0.211	0.495	0.995

The tests N^{GLS}, P_T, and MZ_α^{GLS} use the spectral estimator $\hat{\omega}^2_{ARMA}$ when an ARMA model is fitted, and the spectral estimator $\hat{\omega}^2_{GLS-AR}$ when an AR approximation is used.

Table 6: Empirical size and power for 5% level. T=100. Model: $(1-\rho B)(y_t - d_t) = (1-\theta B)a_t$. Non-Stationary case. 5000 rep.

Model	Test	ρ	$d_t = \mu$ θ					$d_t = \mu + \delta t$ θ				
			-0.8	-0.5	0	0.5	0.8	-0.8	-0.5	0	0.5	0.8
ARMA	N_N^{GLS}	1.00	0.055	0.057	0.055	0.042	0.050	0.064	0.064	0.060	0.031	0.067
		0.90	0.732	0.717	0.643	0.339	0.232	0.352	0.343	0.285	0.123	0.164
	MZ_α^{GLS}	1.00	0.060	0.063	0.062	0.045	0.053	0.088	0.090	0.082	0.039	0.072
		0.90	0.758	0.741	0.670	0.353	0.236	0.437	0.428	0.359	0.150	0.174
	P_T	1.00	0.055	0.056	0.057	0.043	0.050	0.062	0.063	0.059	0.031	0.068
		0.90	0.747	0.732	0.649	0.347	0.231	0.351	0.344	0.287	0.126	0.165
AR(MIC)	N_N^{GLS}	1.00	0.085	0.050	0.033	0.041	0.056	0.088	0.051	0.028	0.027	0.059
		0.90	0.295	0.273	0.270	0.235	0.296	0.137	0.107	0.077	0.097	0.205
	MZ_α^{GLS}	1.00	0.096	0.060	0.042	0.045	0.058	0.115	0.073	0.043	0.045	0.077
		0.90	0.317	0.303	0.300	0.251	0.311	0.193	0.160	0.128	0.149	0.233
	P_T	1.00	0.080	0.047	0.034	0.039	0.055	0.079	0.048	0.026	0.029	0.065
		0.90	0.294	0.270	0.263	0.237	0.301	0.141	0.112	0.083	0.111	0.218
	τ_{GLS}	1.00	0.036	0.038	0.031	0.039	0.079	0.034	0.033	0.029	0.037	0.098
		0.90	0.358	0.279	0.357	0.279	0.263	0.143	0.146	0.147	0.159	0.221
	τ_W	1.00	0.020	0.024	0.022	0.036	0.106	0.007	0.016	0.024	0.045	0.183
		0.90	0.175	0.232	0.299	0.299	0.522	0.042	0.066	0.129	0.173	0.441
AR(BIC)	N_N^{GLS}	1.00	0.142	0.106	0.086	0.086	0.357	0.255	0.166	0.127	0.095	0.375
		0.90	0.580	0.556	0.536	0.528	0.894	0.479	0.414	0.340	0.294	0.711
	MZ_α^{GLS}	1.00	0.158	0.127	0.099	0.099	0.391	0.293	0.207	0.157	0.131	0.455
		0.90	0.609	0.597	0.574	0.561	0.924	0.542	0.487	0.421	0.382	0.785
	P_T	1.00	0.139	0.104	0.085	0.085	0.359	0.241	0.155	0.122	0.101	0.388
		0.90	0.575	0.553	0.532	0.529	0.904	0.473	0.417	0.341	0.308	0.727
	τ_{GLS}	1.00	0.063	0.062	0.050	0.075	0.355	0.079	0.064	0.049	0.080	0.506
		0.90	0.563	0.599	0.536	0.560	0.727	0.275	0.262	0.219	0.334	0.802
	τ_W	1.00	0.066	0.041	0.051	0.126	0.695	0.078	0.038	0.052	0.142	0.857
		0.90	0.480	0.444	0.527	0.791	0.999	0.264	0.171	0.230	0.521	0.997

The tests N^{GLS}, P_T, and MZ_α^{GLS} use the spectral estimator $\hat{\omega}^2_{ARMA}$ when an ARMA model is fitted, and the spectral estimator $\hat{\omega}^2_{GLS-AR}$ when an AR approximation is used.

Table 7: Critical values

Stationary case					Non-stationary case				
T	1%	2.5%	5%	10%	T	1%	2.5%	5%	10%
$N_N^{GLS-\mu}$ with $c = 13.5$					$N_N^{GLS-\mu}$ with $c = 8$				
25	0.4207	0.4943	0.5782	0.7081	25	0.2843	0.3418	0.4056	0.4993
50	0.4146	0.4987	0.5917	0.7318	50	0.2835	0.3486	0.4251	0.5440
100	0.4285	0.5249	0.6406	0.8045	100	0.2930	0.3699	0.4581	0.5988
150	0.4426	0.5518	0.6728	0.8587	150	0.2988	0.3787	0.4743	0.6268
250	0.4555	0.5770	0.7120	0.9284	250	0.3027	0.3920	0.4945	0.6650
500	0.4780	0.6105	0.7674	1.0180	500	0.3103	0.3989	0.5088	0.6928
∞	0.5165	0.6697	0.8551	1.1726	∞	0.3152	0.4120	0.5303	0.7301
$N_N^{GLS-\tau}$ with $c = 20$					$N_N^{GLS-\tau}$ with $c = 8$				
25	0.0221	0.0248	0.0277	0.0319	25	0.0253	0.0289	0.0326	0.0380
50	0.0204	0.0234	0.0266	0.0311	50	0.0231	0.0270	0.0311	0.0369
100	0.0201	0.0234	0.0268	0.0317	100	0.0221	0.0262	0.0305	0.0367
150	0.0201	0.0234	0.0272	0.0324	150	0.0219	0.0260	0.0303	0.0366
250	0.0199	0.0236	0.0276	0.0331	250	0.0217	0.0258	0.0302	0.0366
500	0.0203	0.0242	0.0282	0.0341	500	0.0215	0.0257	0.0302	0.0367
∞	0.0206	0.0246	0.0290	0.0354	∞	0.0212	0.0255	0.0299	0.0366