Testing for a unit root in the presence of autocorrelated errors*

Michalis P. Stamatogiannis†
School of Economics
University of Nottingham

April 3, 2009

Abstract

Most commonly used unit root tests employ test statistics whose distribution under the null hypothesis depends on nuisance parameters. The presence of nuisance parameters, especially in the case where they determine the correlation structure of the innovation errors, can seriously affect the size of the test. To address this issue, we adopt an approach based on the characterization of the class of asymptotically similar critical regions for the unit root hypothesis and the application of some new optimality criteria for the choice of a test within this class. This method is designed to address the issue of size stability right from the point of selecting a test. Related methods of Forchini and Marsh (2000) are extended to the case of a finite order moving average innovation errors. Limit theory for the resulting test statistics is developed and simulation evidence suggests that our statistics have substantially reduced size distortion while retaining good power properties.

1 Introduction

The unit root hypothesis has attracted a great deal of interest in econometrics. Nelson and Plosser (1982) provided empirical evidence that many macroeconomic series have a unit root. From the statistical point of view it

*JEL classification: C12, C22, C32. Keywords: unit root test, nuisance parameter

†Correspondence to: School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, U.K. Email: lexms9@nottingham.ac.uk

is important to know whether or not series are stationary in order to conduct valid inference. The outcome of non-stationarity introduces the possibility of differencing the series (Plosser and Schwert, 1978) cointegration (Johansen, 1988) or error-correction (Engle and Granger, 1987) models. Banerjee *et al.* (1993) and Maddala and Kim (1998) give a review of the literature for unit root tests. Fuller (1976) and Dickey and Fuller (1979, 1981) proposed a unit root test (DF) which is widely used.

As in many testing problems, the fact that the distribution of unit root test statistics under the null hypothesis depends on nuisance parameters can result in serious size distortions for the associated unit root tests. Said and Dickey (1984) showed that the "augmented" DF (ADF) test is appropriate for processes with autoregressive moving average (ARMA) errors. Phillips and Perron (1988) proposed a nonparametric testing procedure (PP) which allowed for a wider class of stationary time series in the error term. Schwert (1989) used Monte Carlo simulations to show the existence of size distortions in the ADF and PP tests. His results suggest that PP has higher power than ADF, but also much higher size distortions in the presence of negative moving average (MA) parameter in the error term. DeJong *et al.* (1992) showed that PP tests perform poorly against trend stationary alternatives and suggested the use of the Said-Dickey testing procedure.

This paper addresses the issue of unit root testing in the presence of correlated innovation errors that take the form of a finite order moving average process. Our approach is based on obtaining a characterization of the class of *similar tests*. These are tests whose size does not depend on nuisance parameters, provided that sufficient statistics for the nuisance parameters exist, under the null hypothesis. Given the fact that a sufficient statistic for the MA parameters is not available, we consistently estimate the MA parameters by maximum likelihood, and then use the above estimates to characterize the class of (asymptotically) similar tests. After the characterization of the class of similar regions we proceed to the selection of some tests within this class by the use of appropriate optimality criteria. The advantage of such an approach is that we can focus our attention on a set of tests whose size is independent of the nuisance parameters involved. In this way we can address the serious issue of size stability at the first stage of selecting a test.

An ideal test would be one which has actual size near to the nominal size and high power. However, as the above discussion suggests, in many cases there is a trade-off between loss of power and size distortion. Following Hillier (1987), we give priority to size stability and concentrate on tests which are asymptotically similar. Forchini and Marsh (2000) have chosen similar unit root tests according to the Bounded Norm Minimizing (BNM) and Bounded Estimated Point Optimal (BEPO) criteria under the assumption of i.i.d.

innovation errors.

In this paper, we construct asymptotically similar BNM and BEPO unit root tests in the presence of possibly correlated innovation errors that may take the form of a finite order MA process. The objective is to derive a testing procedure for the unit root hypothesis with fairly stable size over different MA models with varying MA parameters, and with power close to the power envelope.

The paper is organised as follows. In Section 2 we refer to the theory related to the construction of similar tests. Section 3 describes the optimality criteria for the choice test statistics proposed by Forchini and Marsh (2000). Section 4 describes the construction of similar regions in the case of correlated errors and in Section 5 we use the optimality criteria to derive the test statistics followed by the description of the method of estimation we are using. The limiting distributions of the resulting test statistics are derived in the presence of deterministics consisting of an intercept and a linear trend. In Section 6 the finite-sample performance of the statistics is assessed in the context of a simulation study. In Section 7 we provide some final concluding remarks. All proofs are included in the technical Appendix of Section 8.

2 Characterization of similar regions

The methodology we follow for the characterization of similar regions is described by Hillier (1987). Let z be a vector of random variables with density $f(z;\eta,\theta)$ depending on two vectors of parameters η, and θ. In case we want to test the null hypothesis

$$H_0 : \theta = \theta_0$$

then θ is the vector of parameters of interest and η is the vector of nuisance parameters. In general the size of any critical region ω in this context is going to be dependent on η

$$\int_\omega f(z;\eta,\theta_0) = \alpha(\eta).$$

Critical regions related to this problem which are independent to nuisance parameters

$$\int_\omega f(z;\eta,\theta_0) = \alpha$$

are called similar critical regions. In case there is a sufficient statistic t for η under H_0 the density function is

$$f(z;\eta,\theta_0) = pdf(t;\eta,\theta_0)pdf(z|t;\theta_0)$$

where $pdf(t; \eta, \theta_0)$ in the density of the sufficient statistic under the null and $pdf(z|t; \theta_0)$ in the conditional density of z given t. The latter density does not depend on the nuisance parameter η. So provided that we can have sufficient statistics for η the conditional distribution of z given these statistics will be free of nuisance parameters and as a consequence is going to result a similar critical region.

A critical region ω for which

$$\Pr\{z \in \omega | t; \theta_0, \eta\} = \alpha$$

i.e.. is constant across all t, has *Neyman structure*. Consequently for hypotheses involving nuisance parameters and there exist sufficient statistics under the null hypothesis for these parameters, then similar regions exist. These can be constructed by choosing critical region ω so that the conditional probability content of ω is α, on each surface of constant t.

A critical region with Neyman structure is a similar critical region. However the opposite does not hold. In order a similar critical region to have Neyman structure t must be *boundedly complete.* Bounded completeness means that for every bounded function $g(t)$ not depending on the parameters, $E(g(t)) = 0$ implies that $g(t) = 0$, everywhere, except perhaps on sets of zero measure (Lehmann (1986) p. 140). So if the sufficient statistic t for the nuisance parameter is boundedly complete under the null hypothesis, then every critical similar region has Neyman structure.

Hillier (1987) summarizes the procedure for constructing similar regions in Theorem 2.1

Theorem 1. *Let t be a boundedly complete sufficient statistic for the nuisance parameter η under H_0. If for almost all t there is a one-to-one transformation $z \longmapsto (t(z), \upsilon(z))$ for which under H_0 υ is independent of t, then the statistic υ characterizes the class of similar regions for testing H_0 in the sense that a region ω is similar of size α if and only if ω has size α in the distribution of υ.*

We are going to use this theorem for the characterization of similar critical tests in our case, where an MA parameter in the errors is involved as a nuisance parameter.

3 Optimality criteria

Until now the method for the characterization of the class of similar tests has been discussed. However, we need to use some optimality criteria for the selection of a specific test statistic from this class. Ideally, we would choose a *Uniformly Most Powerful* (UMP) test. A UMP test is a test which has the highest power for every η, and θ. In unit root tests the power of a test depends on the nuisance parameters η and the value of the parameter of interest θ under H_1, so it is not possible to achieve the UMP criterion. In such case we have to use weaker optimality criteria for the selection of test. Cox and Hinkley (1974) suggest some alternative optimality criteria, like the selection of a typical alternative for θ (*point optimal* (PO)) or the construction of *locally most powerful* (LMP) test, which involves the maximization of the power of the test in the neighborhood of the null hypothesis. Selecting a typical value of θ could be seen as arbitrary unless there are specific prior information for the parameter. The problem with the LMP tests is that their power is low, especially for alternatives far from the null (Zaman (1996, pp. 133-136).

Forchini and Marsh (2000) suggest the use of two alternative optimality criteria, which approximate the UMP criterion. They consider a testing problem for the case of $H_0 : y \sim N(0, \sigma^2\Omega(\theta_0))$ against $H_1 : y \sim N(0, \sigma^2\Omega(\theta))$, where y is a $N \times 1$ vector, σ^2 is unknown parameter and θ is a scalar. The test takes the form:

reject H_0 if

$$\frac{y'\Omega^{-1}(\theta)y}{y'\Omega^{-1}(\theta_0)y} < k_\alpha \tag{1}$$

where k_α is chosen so that α is the size of the test. In the case that the numerator changes with θ, we cannot have a UMP test.

3.1 Bounded Norm Minimizing Tests

Suppose that $y'\Omega^{-1}(\theta)y \leq l(\theta)'\Psi(y)l(\theta)$, where $l(\theta)$ is a vector depending only upon θ and $\Psi(y)$ is a positive definite matrix depending only upon y.

A sufficient condition for

$$\frac{l(\theta)'\Psi(y)l(\theta)}{y'\Omega^{-1}(\theta_0)y} < k_\alpha$$

is to minimize the norm

$$\left\| \frac{\Psi(y)}{y'\Omega^{-1}(\theta_0)y} \right\| < k,$$

for k such that the size of the test is α. The norm in the above display can be any matrix norm (see e.g. Horn and Johnson, 1985). Note that any norm of the matrix $\Psi(y)/y'\Omega^{-1}(\theta_0)y$ gives a *norm minimizing* (NM) test and when (1) holds with equality, or a BNM test otherwise.

3.2 Bounded Estimated Point Optimal Tests

The second optimality criterion is that of using *estimated point optimal* tests (EPO). This criterion is related to the PO tests which are discussed above. Even if the alternative is generally unknown, it is possible to estimate it with the value θ^* which minimizes

$$\frac{l(\theta)'\Psi(y)l(\theta)}{y'\Omega^{-1}(\theta_0)y}$$

for a set of observations y. In case (1) holds with equality, the EPO test is: reject H_0 if

$$\frac{l(\theta^*)'\Psi(y)l(\theta^*)}{y'\Omega^{-1}(\theta_0)y} < k, \tag{2}$$

where k is chosen such that the size of the test is α. Similarly to the case of the BNM criterion, if (1) does not hold with equality (2) is a BEPO test. Another criterion of this type is reject H_0 if

$$|\rho^* - \rho_0| > k_a, \tag{3}$$

where ρ_0 is the value of the parameter under H_0 and k_a is chosen such that the size of the test is a.

Forchini and Marsh (2000) use the above criteria for the derivation of similar unit root test statistics. Simulation results suggest that these statistics have distorted size in the presence of an MA parameter in the errors. We apply these criteria to choose statistics from the class of asymptotically similar tests and we find that these have good power properties in finite samples. Moreover, they appear to have small size distortions in comparison to the statistics given by Forchini and Marsh. This is due to the construction of these statistics which takes into consideration the presence of the MA term.

4 Testing for unit roots in the errors

Marsh (2005) considers a linear regression model with an MA term in the errors and characterizes the class of asymptotically similar tests. We use the LMP, BNM and BEPO optimality criteria for deriving tests within this class. The model is

$$y = X\beta + u, \tag{4}$$

where β is a $k \times 1$ vector of parameters, X a $N \times k$ full rank matrix of the exogenous variables, $u = (u_1, ..., u_N)'$ and

$$\begin{aligned} u_t &= \rho u_{t-1} + \zeta_t \\ \zeta_t &= \sum_{j=0}^{m} \phi_j \varepsilon_{t-j} \\ \varepsilon_t &\sim NIID(0, \sigma^2) \end{aligned}$$

for $t = 1, ..., N$, $u_0 = 0$, and $\phi_0 = 1$. In order ϕ to be identified from data, we need to impose the invertibility condition $\left|\phi_j\right| < 1$ for $j = 1, ..., m$.

Letting $L^{(i)}$ be the lower-triangular matrix with ones on the i^{th} off-diagonal element and zeros elsewhere and defining

$$T_\rho = (I_N - \rho L^{(1)}) \tag{5}$$

$$K_\phi = (I_N + \sum_{i=1}^{m} \phi_i L^{(i)}) \tag{6}$$

$$\phi = (\phi_1, ..., \phi_m)' \text{ and } \varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)',$$

(4) can be expressed as

$$T_\rho(y - X\beta) = K_\phi \varepsilon.$$

For notational simplicity, we define

$$x = K_\phi^{-1} T_1 y, \tag{7}$$

$$Z = K_\phi^{-1} T_1 X \tag{8}$$

and

$$\nu = M_Z K_\phi^{-1} T_1 y. \tag{9}$$

In this context the unit root hypothesis takes the form

$$H_0 : \rho = 1 \text{ vs. } H_1 : |\rho| < 1$$

Under the above assumptions the joint sample density of y is

$$y \sim N(X\beta, \sigma^2 T_\rho K_\phi K_\phi' \left(T_\rho^{-1}\right)')$$

Let Ξ denote the parameter space $\theta = (\rho, \beta', \sigma^2, \phi')' = (\rho \in (-1, 1], \beta \in \mathbb{R}^k, \sigma^2 \in \mathbb{R}^+, \phi \in \mathbb{R})$ and $\Xi_0 = (1, \beta', \sigma^2, \phi')$, $\Xi_1 = \Xi - \Xi_0$ then the unit root hypothesis can be stated formally as

$$H_0 : \theta \in \Xi_0 \text{ vs. } H_1 : \theta \in \Xi_1$$

which is a nuisance parameter problem, with nuisance parameters $(\beta', \sigma^2, \phi')'$.

The distribution of x is given by

$$x \sim N(Z\beta, \sigma^2 \Sigma_{\rho,\phi})$$

where

$$\Sigma_{\rho,\phi} = K_\phi^{-1} T_1 T_\rho^{-1} K_\phi K_\phi' \left(T_\rho^{-1}\right)' T_1' \left(K_\phi^{-1}\right)'. \tag{10}$$

At this point it is useful to use the following lemma before proceeding.

Lemma 1. *Matrix $\Sigma_{\rho,\phi}$ given in (10) can be expressed as*

$$\Sigma_{\rho,\phi} \equiv \Sigma_\rho = T_1 T_\rho^{-1} \left(T_\rho^{-1}\right)' T_1'.$$

Note also that x under H_0 is

$$x \sim N(Z\beta, \sigma^2 I_N)$$

For the characterization of the class of similar tests the methodology by Hillier (1987) described in section 2 is applied in this setup. The projection matrix

$$M_Z = I_N - Z\left(Z'Z\right)^{-1} Z'$$

can be decomposed as:

$$\begin{aligned} CC' &= M_Z \\ C'C &= I_{N-k} \end{aligned}$$

where C is a $N \times N - k$ matrix.

The following transformation are applied using C matrix. First x is transformed as

$$x \longmapsto \begin{pmatrix} \widehat{\beta} = \left(Z'Z\right)^{-1} Z'x \\ w = C'x \end{pmatrix}$$

and then w as

$$w \longmapsto \begin{pmatrix} s^2 = x' M_Z x \\ v = C'x/s^2 \end{pmatrix}$$

As it can be seen from the above, $\widehat{\beta}$ is not feasible due to the fact that is dependent on ϕ. It is possible however to proceed by finding a consistent estimate of ϕ.

The distribution of w is

$$w \sim N(0, \sigma^2 A) \tag{11}$$

where $A = C'\Sigma_\rho C$. A is a $N \times k$ positive semi definite matrix. We can see that $\widehat{\beta}$ and w are independent. From (11) we have the joint density of s^2 and υ

$$pdf(s^2, \upsilon) = |J| \det A^{-1/2} \exp\left(-\frac{s^2}{2\sigma^2}\upsilon' A^{-1}\upsilon\right)\left(2\pi\sigma^2\right)^{-\frac{N-k}{2}}. \qquad (12)$$

We get the marginal density of by integrating out s^2 from (12)

$$pdf(\upsilon) = \det A^{-1/2}\left(\upsilon' A^{-1}\upsilon\right)^{-\frac{N-k}{2}}. \qquad (13)$$

Under H_0, s^2 and υ are independent. So υ characterizes the class of asymptotically similar tests for H_0. Because of the fact that under H_0 $A = I_{N-k}$, υ is uniform on the surface of unit sphere in $\mathbb{R}^{N-k}$, $S_{N-k} = \{\upsilon \in \mathbb{R}^{N-k} : \upsilon'\upsilon\}$. If ω is a critical region on S_{N-k} the most powerful critical region of H_0 vs H_1, $\overline{\omega}$ satisfies:

$$\overline{\omega} = \arg\max_{\omega \in S_{N-k}}\left(P_\omega = \det A^{-1/2}\int_\omega \left(\upsilon' A^{-1}\upsilon\right)^{-\frac{N-k}{2}} d(\upsilon)\right),$$

which implies rejection region:

$$\overline{\omega} = \upsilon' A^{-1}\upsilon < k_\alpha, \qquad (14)$$

where k_α is chosen such that the size of the test is α.

We estimate the MA parameter, using conditional maximum likelihood estimation (MLE). The MLE is conditioned on the initial value ε_t being zero.

5 Similar Statistics

After the characterization of the class of asymptotically similar statistics we use the optimality criteria discussed above, in order to derive test statistics from this class. The first test statistic we construct is the LMP test.

Theorem 2. *The LMP test for $H_0 : \rho = 1$ against $H_1 : \rho < 1$ is given by the following rule: reject H_0 if*

$$LMP_m = \frac{\nu'\left(K_\phi^{-1}L_N T_1^{-1}K_\phi + \left(K_\phi^{-1}L_N T_1^{-1}K_\phi\right)'\right)\nu}{\nu'\nu} < k_\alpha$$

where $L_N = I_N - T_1$ is the first-order lag operator matrix, with one's on the lower off-diagonal and zero's elsewhere, $\nu = M_Z K_\phi^{-1} T_1 y$, and k_α is chosen such that the size of the test is α.

The LMP test for $H_0 : \rho = 1$ against $H_1 : \rho > 1$ is given by the following rule: reject H_0 if

$$LMP_m = \frac{\nu' \left(K_\phi^{-1} L_N T_1^{-1} K_\phi + \left(K_\phi^{-1} L_N T_1^{-1} K_\phi \right)' \right) \nu}{\nu' \nu} > k'_\alpha$$

where $L_N = I_N - T_1$ is the first-order lag operator matrix, with one's on the lower off-diagonal and zero's elsewhere, $\nu = M_Z K_\phi^{-1} T_1 y$, and k'_α is chosen such that the size of the test is α.

Now we use the optimality criteria given by Forchini and Marsh (2000), and described above. Let

$$\Psi_{11} = \left(T_1^{-1} \right)' T_1^{-1} \qquad (15)$$

$$\Psi_{12} = \left(T_1^{-1} \right)' \left(T_1^{-1} - I_N \right) \qquad (16)$$

$$\Psi_{22} = \left(T_1^{-1} - I_N \right)' \left(T_1^{-1} - I_N \right). \qquad (17)$$

Theorem 3. *Let $\|.\|$ denote a norm on the space 2×2 positive definite matrices, and let*

$$\Psi(\nu) = \frac{1}{\nu'\nu} \begin{pmatrix} \nu'\Psi_{11}\nu & \nu'\Psi_{12}\nu \\ \nu'\Psi_{12}\nu & \nu'\Psi_{22}\nu \end{pmatrix} \qquad (18)$$

Then a BNM test is: reject $H_0 : \rho = 1$ if

$$BNM < k_\alpha$$

where $BNM = \|\Psi(\nu)\|$ and k_α is chosen such that the size of the test is α.

Theorem 2 generates a class of BNM tests, depending upon the particular norm chosen. A statistic from this class could result from the use of the Euclidean matrix norm $\|\Psi(\nu)\| = \left\{ tr\Psi(\nu)'\Psi(\nu) \right\}^{1/2}$. The spectral norm of $\Psi(\nu)$, defined as the square root of the maximal eigenvalue of $\Psi(\nu)'\Psi(\nu)$, is also frequently used.

Theorem 4. *BEPO tests for $H_0 : \rho = 1$ against $H_1 : -1 < \rho < 1$ are given by the following rule:*
reject H_0 if

$$BEPO_1 = \frac{1}{\nu'\nu} \frac{(\nu'\Psi_{11}\nu)(\nu'\Psi_{22}\nu) - (\nu'\Psi_{12}\nu)^2}{\nu'\Psi_{22}\nu} < k_\alpha,$$

and

$$BEPO_2 = N \left| \frac{\nu'\Psi_{12}\nu - \nu'\Psi_{22}\nu}{\nu'\Psi_{22}\nu} \right| > k_\alpha$$

where k_a is such that the size of the tests is α.

The above statistics include matrix K_{ϕ}. In the absence of a sufficient statistic for ϕ, we need to find a consistent estimate of it. It has to be stressed that the choice of a good estimator for ϕ is of major importance for the good properties (stable size and high power) of the statistics. We estimate ϕ by conditional maximum likelihood or pseudo-maximum likelihood if we do not wish to maintain the normality assumption on the innovation errors. It is a well known fact that, under the invertibility assumption imposed on the moving average process, the (pseudo) maximum likelihood estimator of ϕ is $\sqrt{N}$-consistent.

The algorithm for estimating ϕ is described below. We first estimate the following model with least squares:

$$y_t = X\widehat{\beta} + \hat{u}_t. \tag{19}$$

We assume that the residuals of (19) follow an $ARMA(1, m)$

$$\widehat{u}_t = \rho\widehat{u}_{t-1} + \varepsilon_t + \sum_{i=1}^{m} \phi_i \varepsilon_{t-i},$$

for $t = 1, 2, ..., N$. We condition on the m first values of ε being zero:

$$\varepsilon_0 = \varepsilon_1 = ... = \varepsilon_m = 0.$$

From the above assumptions we can iterate on:

$$\varepsilon_t = (\widehat{u}_t - \rho\widehat{u}_{t-1}) - \sum_{i=1}^{m} \phi_i \varepsilon_{t-i},$$

for $t = 1, 2, ..., N$.

The conditional log likelihood is

$$\mathcal{L}\left(\rho, \phi, \sigma^2\right) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \sum_{t=1}^{N} \frac{\varepsilon^2}{2\sigma^2}$$

Since we assumed $\left|\phi_j\right| < 1$ for $j = 1, ..., m$ the effect of the initial condition fades out as sample size increases (Hamilton p.128). After estimating the MA parameters, we can substitute them in the sufficient statistics for (β, σ^2) and then construct the similar critical regions. It is important to note here that, asymptotically, the test statistics we derive do not depend on the nuisance parameter under H_0 since $Z = K_{\hat{\phi}}^{-1} T_1 X$ and $T_1 u = K_{\phi}\varepsilon$ which gives

$$\begin{aligned} \nu &= M_Z K_{\hat{\phi}}^{-1} T_1 \left(X\beta + u\right) = M_Z K_{\hat{\phi}}^{-1} T_1 u \\ &= M_Z K_{\hat{\phi}}^{-1} K_{\phi}\varepsilon = \left[I + o_p\left(1\right)\right] M_Z \varepsilon. \end{aligned}$$

The above result shows that the statistics we derive are asymptotically similar.

Having derived the BNM and BEPO test statistics for the unit root hypothesis, we proceed to derive their limiting distributions. To this end, we restrict the deterministic components of the data generating process to an intercept and a linear trend, i.e. we assume that the matrix of deterministics in (4) takes the form

$$X' = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & N \end{bmatrix}. \tag{20}$$

Theorem 5. *Consider the process in (4) with X satisfying (20) and let $W(\cdot)$ be standard Brownian motion on $D[0,1]$. Under the null hypothesis $H_0 : \rho = 1$, the following limit theory applies as $N \to \infty$:*

(i) *The BNM test of Theorem 3 satisfies*

$$BNM \Rightarrow \left\{ \int_0^1 W(r)^2 dr - 2W(1) \int_0^1 rW(r)dr + \frac{1}{3}W(1)^2 \right\} \left\| \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\|.$$

(ii) *The BEPO tests of Theorem 4 satisfy*

$$\begin{aligned} BEPO_1 &\Rightarrow 2W^2(1) - 2W(1)\int_0^1 W(r)dr + 1, \\ BEPO_2 &\Rightarrow \left| \frac{-W^2(1) + W(1)\int_0^1 W(r)dr - \frac{3}{2}}{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)} \right| \end{aligned}$$

Note, that the matrix on the limit distribution of part (i) has identical Euclidian and spectral norm equal to 2.

6 Numerical Study

The test statistics we develop are motivated asymptotically. This means that they are asymptotically similar with respect to the MA parameter. In order to examine their size and power properties in small samples we proceed to a Monte Carlo study. We use two models for the simulations. The first is

$$\begin{aligned} y_t &= 0.5 + u_t, \\ u_t &= \rho u_{t-1} + \varepsilon_t + \phi\varepsilon_{t-1}, \\ \varepsilon_t &\sim NIID(0,1), \end{aligned} \tag{21}$$

and the second

$$
\begin{aligned}
y_t &= 0.5 + 0.3t + u_t, \\
u_t &= \rho u_{t-1} + \varepsilon_t + \phi \varepsilon_{t-1}, \\
\varepsilon_t &\sim NIID(0,1).
\end{aligned} \tag{22}
$$

The Monte Carlo experiment was based on 5000 replications. We investigated the size distortion and power of the statistics in finite samples. For the numerical study related to size distortion we used

$$
\begin{aligned}
\phi &= -0.8, -0.7, ..., 0.8, \\
N &= 50, 100, 200, \\
\rho &= 1, \\
\alpha &= 0.05,
\end{aligned}
$$

where α is the nominal size of the test statistics. For the numerical study referring to power of the statistics we used

$$
\begin{aligned}
\rho &= 0.8, 0.82, ..., 0.98, \\
N &= 50, 100, 200, \\
\phi &= -0.5, 0, \\
\alpha &= 0.05.
\end{aligned}
$$

LMP statistic was found to have very low power in the case of constant and trend included in the mean function. For this reason simulation results for this statistic are not reported for model (22). Statistics LMP, BNM and $BEPO$ correspond to the case in which MA terms are not estimated. These are the statistics proposed by Forchini and Marsh (2000). In order to get statistics LMP, BNM and $BEPO$ we set $\phi = 0$ (i.e. $K_\phi = I_N$) in LMP_m, BNM_e and $BEPO_1$ respectively. $BEPO_1$ and $BEPO_2$ test statistics were found to have almost identical size distortion and power, so we present results only for $BEPO_1$.

Tables 1 and 2 report size distortions of the statistics for models (21) and (22) respectively. We observe that using the similar statistics we derived in this paper can result substantial decrease is size distortion in comparison to the statistics that do not take autocorrelation into account. For $N = 100$, $\phi = -0.8$ the size of BNM_e statistic is 0.157 while the one of BNM is 0.831. LMP_m is found to have very low size distortion even for values of ϕ close to -1. BNM_e, BNM_a, and $BEPO_1$ statistics have the same behaviour in terms of size distortion. Finally the size distortion of LMP_m is smaller in general than the size distortion of BNM_e, BNM_a, and $BEPO_1$.

Another crucial observation for the similar statistics is that their size distortion reduces as sample size N increases. For example, when $\phi = -0.8$, $BEPO_1$ statistic has size 0.254 for $N = 50$, 0.155 for $N = 100$ and 0.09 for $N = 200$. So we can conclude that the size of the similar statistics derived in this paper converges to the nominal value (5% in this case), as sample size increases. This is not true for the statistics LMP, BNM and $BEPO$: size distortion increases as sample size N increases. $BEPO$ statistic for example has size 0.673 for $N = 50$, 0.832 for $N = 100$ and 0.908 for $N = 200$, when $\phi = -0.8$. This suggests that empirical size of the statistics of Forchini and Marsh can go farther from nominal size as N increases in the presence of autocorrelation in the errors.

Table 1. Size of the tests for model (21)

N	ϕ	LMP	BNM	$BEPO$	LMP_m	BNM_e	BMN_a	$BEPO_1$
50	−0.8	0.188	0.669	0.673	0.101	0.255	0.255	0.254
	−0.7	0.140	0.559	0.563	0.076	0.156	0.155	0.155
	−0.6	0.113	0.464	0.466	0.059	0.099	0.099	0.099
	−0.5	0.116	0.364	0.366	0.055	0.080	0.080	0.080
	−0.4	0.101	0.255	0.259	0.051	0.079	0.079	0.077
	−0.3	0.076	0.155	0.162	0.058	0.049	0.049	0.048
	−0.2	0.067	0.135	0.135	0.063	0.056	0.055	0.055
	−0.1	0.041	0.068	0.070	0.049	0.043	0.043	0.042
	0.0	0.045	0.050	0.054	0.044	0.048	0.048	0.048
	0.1	0.043	0.023	0.024	0.055	0.038	0.037	0.039
	0.2	0.045	0.025	0.027	0.045	0.033	0.033	0.032
	0.3	0.030	0.018	0.019	0.052	0.036	0.034	0.035
	0.4	0.050	0.010	0.010	0.058	0.044	0.044	0.043
	0.5	0.030	0.011	0.010	0.037	0.034	0.034	0.033
	0.6	0.041	0.007	0.007	0.041	0.030	0.028	0.029
	0.7	0.031	0.008	0.009	0.048	0.033	0.033	0.032
	0.8	0.046	0.009	0.008	0.049	0.025	0.025	0.023
100	−0.8	0.201	0.831	0.832	0.062	0.157	0.157	0.155
	−0.7	0.150	0.680	0.681	0.055	0.105	0.105	0.103
	−0.6	0.118	0.566	0.569	0.042	0.063	0.063	0.062
	−0.5	0.106	0.435	0.436	0.049	0.078	0.078	0.077
	−0.4	0.092	0.327	0.328	0.054	0.052	0.052	0.051
	−0.3	0.094	0.210	0.213	0.039	0.054	0.055	0.054
	−0.2	0.064	0.135	0.133	0.040	0.062	0.062	0.061
	−0.1	0.062	0.071	0.071	0.043	0.058	0.059	0.057
	0.0	0.061	0.053	0.054	0.047	0.042	0.042	0.042
	0.1	0.042	0.036	0.035	0.040	0.054	0.054	0.054
	0.2	0.037	0.010	0.011	0.057	0.049	0.049	0.050
	0.3	0.044	0.021	0.021	0.037	0.044	0.044	0.044
	0.4	0.043	0.008	0.008	0.050	0.047	0.048	0.048
	0.5	0.034	0.004	0.004	0.056	0.052	0.052	0.051
	0.6	0.038	0.005	0.005	0.052	0.046	0.046	0.046
	0.7	0.049	0.006	0.006	0.043	0.040	0.040	0.040
	0.8	0.042	0.006	0.005	0.056	0.039	0.040	0.039
200	−0.8	0.247	0.908	0.908	0.061	0.091	0.091	0.090
	−0.7	0.173	0.768	0.767	0.068	0.082	0.082	0.083
	−0.6	0.130	0.598	0.599	0.061	0.078	0.078	0.079
	−0.5	0.106	0.481	0.480	0.051	0.057	0.057	0.057
	−0.4	0.106	0.346	0.346	0.034	0.052	0.052	0.052
	−0.3	0.081	0.202	0.202	0.050	0.048	0.048	0.050
	−0.2	0.055	0.151	0.151	0.049	0.060	0.060	0.060
	−0.1	0.064	0.085	0.086	0.050	0.060	0.060	0.061
	0.0	0.047	0.048	0.048	0.049	0.033	0.033	0.033
	0.1	0.035	0.033	0.033	0.055	0.044	0.044	0.045
	0.2	0.042	0.014	0.015	0.050	0.058	0.058	0.058
	0.3	0.036	0.005	0.005	0.046	0.041	0.041	0.041
	0.4	0.050	0.009	0.009	0.049	0.052	0.052	0.052
	0.5	0.029	0.006	0.006	0.056	0.055	0.055	0.055
	0.6	0.043	0.012	0.012	0.049	0.054	0.054	0.053
	0.7	0.030	0.009	0.009	0.056	0.049	0.049	0.051
	0.8	0.034	0.009	0.008	0.071	0.054	0.054	0.054

Table 2.Size of the tests for model (22)						
	ϕ	BNM	$BEPO$	BNM_e	BMN_a	$BEPO_1$
50	−0.8	0.778	0.781	0.374	0.371	0.373
	−0.7	0.717	0.717	0.297	0.292	0.296
	−0.6	0.613	0.613	0.206	0.199	0.205
	−0.5	0.456	0.457	0.144	0.139	0.143
	−0.4	0.337	0.341	0.115	0.109	0.115
	−0.3	0.224	0.225	0.081	0.077	0.080
	−0.2	0.139	0.139	0.069	0.066	0.069
	−0.1	0.069	0.071	0.056	0.055	0.056
	0.0	0.043	0.044	0.053	0.050	0.052
	0.1	0.025	0.026	0.049	0.045	0.048
	0.2	0.015	0.015	0.037	0.035	0.037
	0.3	0.003	0.003	0.043	0.038	0.043
	0.4	0.006	0.006	0.031	0.028	0.030
	0.5	0.007	0.007	0.026	0.022	0.025
	0.6	0.002	0.002	0.028	0.027	0.028
	0.7	0.002	0.002	0.023	0.022	0.023
	0.8	0.001	0.001	0.021	0.020	0.021
100	−0.8	0.931	0.931	0.294	0.294	0.296
	−0.7	0.841	0.838	0.178	0.180	0.181
	−0.6	0.743	0.742	0.118	0.118	0.120
	−0.5	0.581	0.579	0.093	0.093	0.093
	−0.4	0.380	0.378	0.097	0.097	0.097
	−0.3	0.272	0.269	0.088	0.088	0.090
	−0.2	0.172	0.167	0.060	0.060	0.061
	−0.1	0.089	0.088	0.046	0.046	0.048
	0.0	0.050	0.050	0.047	0.047	0.047
	0.1	0.028	0.027	0.043	0.043	0.043
	0.2	0.010	0.010	0.053	0.053	0.054
	0.3	0.005	0.005	0.043	0.043	0.043
	0.4	0.006	0.006	0.047	0.048	0.048
	0.5	0.004	0.004	0.039	0.039	0.040
	0.6	0.006	0.006	0.044	0.045	0.046
	0.7	0.002	0.002	0.036	0.036	0.036
	0.8	0.001	0.000	0.030	0.030	0.031
200	−0.8	0.983	0.983	0.169	0.169	0.169
	−0.7	0.921	0.921	0.107	0.107	0.107
	−0.6	0.796	0.796	0.092	0.092	0.092
	−0.5	0.625	0.624	0.069	0.069	0.068
	−0.4	0.458	0.457	0.066	0.067	0.066
	−0.3	0.284	0.283	0.050	0.050	0.049
	−0.2	0.161	0.161	0.066	0.066	0.066
	−0.1	0.086	0.085	0.057	0.057	0.057
	0.0	0.055	0.055	0.047	0.048	0.047
	0.1	0.026	0.026	0.034	0.034	0.034
	0.2	0.013	0.013	0.044	0.044	0.043
	0.3	0.011	0.011	0.050	0.050	0.050
	0.4	0.003	0.003	0.036	0.036	0.036
	0.5	0.004	0.004	0.032	0.032	0.032
	0.6	0.002	0.002	0.044	0.045	0.043
	0.7	0.000	0.000	0.037	0.037	0.037
	0.8	0.004	0.004	0.053	0.053	0.053

Tables 3 and 4 report the power of the statistics for models (21) and (22) respectively, when there is no autocorrelation in the error term ε_t ($\phi = 0$).

This is not a favourable case for the statistics we derive in this paper, since an MA term is estimated that does not exist under the data generating process. However, we observe that the power loss (in comparison to the statistics that do not assume autocorrelation of ε_t) due to this inefficiency is minimal for values of ρ near the null hypothesis and very small for values farther.

In Table 3, we observe that for $N = 50$, LMP_m outperforms BNM_e, BNM_a, and $BEPO_1$ for alternatives close to the null hypothesis ($\rho_0 = 1$). This is something that we expect, since the LMP_m is constructed to maximise power near the null hypothesis (locally) in finite samples. However, when we move to alternatives far from the null LMP_m has extremely lower power than statisticsBNM_e, BNM_a, and $BEPO_1$. For $N = 50, \rho = 0.8$, power for LMP_m is 0.187 and for BNM_e is 0.436. For higher sample sizes ($N = 100, 200$) we see that BNM_e, BNM_a, and $BEPO_1$ have higher power than LMP_m for any alternative.

Table 3. Power of the tests for model (21), $\phi = 0$

N	ρ	LMP	BNM	$BEPO$	LMP_m	BNM_e	BNM_a	$BEPO_1$
50	0.80	0.182	0.550	0.566	0.187	0.436	0.434	0.433
	0.82	0.179	0.486	0.495	0.182	0.391	0.388	0.385
	0.84	0.181	0.430	0.440	0.204	0.379	0.379	0.375
	0.86	0.163	0.350	0.358	0.158	0.303	0.300	0.299
	0.88	0.156	0.282	0.292	0.155	0.264	0.262	0.258
	0.90	0.129	0.211	0.222	0.152	0.209	0.205	0.206
	0.92	0.111	0.184	0.187	0.126	0.168	0.167	0.167
	0.94	0.094	0.125	0.128	0.127	0.126	0.124	0.124
	0.96	0.089	0.100	0.103	0.092	0.087	0.086	0.087
	0.98	0.073	0.064	0.068	0.072	0.066	0.065	0.064
100	0.80	0.260	0.926	0.930	0.226	0.889	0.889	0.889
	0.82	0.243	0.884	0.887	0.231	0.869	0.869	0.869
	0.84	0.232	0.851	0.853	0.208	0.800	0.800	0.801
	0.86	0.243	0.819	0.823	0.196	0.752	0.752	0.753
	0.88	0.250	0.739	0.747	0.173	0.674	0.674	0.676
	0.90	0.199	0.626	0.631	0.150	0.557	0.557	0.561
	0.92	0.174	0.514	0.519	0.143	0.446	0.447	0.447
	0.94	0.158	0.328	0.336	0.152	0.339	0.340	0.338
	0.96	0.146	0.237	0.240	0.135	0.197	0.197	0.197
	0.98	0.113	0.097	0.098	0.082	0.102	0.103	0.102
200	0.80	0.367	0.999	0.999	0.380	0.994	0.994	0.994
	0.82	0.341	0.996	0.996	0.345	0.994	0.994	0.994
	0.84	0.341	0.996	0.996	0.331	0.991	0.991	0.991
	0.86	0.324	0.994	0.994	0.325	0.981	0.981	0.981
	0.88	0.289	0.988	0.988	0.305	0.975	0.975	0.974
	0.90	0.281	0.970	0.971	0.276	0.965	0.965	0.965
	0.92	0.226	0.903	0.905	0.275	0.913	0.913	0.917
	0.94	0.225	0.810	0.811	0.212	0.769	0.769	0.772
	0.96	0.185	0.545	0.548	0.188	0.507	0.507	0.511
	0.98	0.158	0.223	0.224	0.130	0.216	0.216	0.218

Table 4. Power of the tests for model (22), $\phi = 0$

N	ρ	BNM	$BEPO$	BNM_e	BNM_a	$BEPO_1$
50	0.80	0.272	0.275	0.230	0.225	0.230
	0.82	0.217	0.218	0.224	0.217	0.224
	0.84	0.210	0.210	0.168	0.162	0.168
	0.86	0.151	0.153	0.146	0.138	0.144
	0.88	0.115	0.116	0.103	0.100	0.103
	0.90	0.092	0.092	0.087	0.083	0.086
	0.92	0.092	0.093	0.074	0.070	0.074
	0.94	0.064	0.064	0.063	0.060	0.063
	0.96	0.045	0.045	0.058	0.055	0.057
	0.98	0.052	0.053	0.040	0.039	0.040
100	0.80	0.684	0.683	0.643	0.646	0.651
	0.82	0.623	0.615	0.554	0.557	0.562
	0.84	0.570	0.564	0.499	0.503	0.503
	0.86	0.478	0.471	0.422	0.422	0.425
	0.88	0.356	0.355	0.335	0.338	0.340
	0.90	0.281	0.273	0.221	0.223	0.223
	0.92	0.198	0.192	0.180	0.182	0.183
	0.94	0.123	0.121	0.132	0.133	0.135
	0.96	0.092	0.091	0.086	0.089	0.090
	0.98	0.074	0.072	0.062	0.062	0.062
200	0.80	0.964	0.964	0.965	0.965	0.965
	0.82	0.960	0.959	0.950	0.950	0.949
	0.84	0.926	0.926	0.922	0.922	0.922
	0.86	0.885	0.883	0.887	0.887	0.886
	0.88	0.849	0.848	0.818	0.818	0.817
	0.90	0.718	0.718	0.704	0.704	0.704
	0.92	0.566	0.564	0.575	0.576	0.574
	0.94	0.374	0.371	0.377	0.377	0.375
	0.96	0.183	0.182	0.208	0.208	0.207
	0.98	0.070	0.069	0.089	0.089	0.089

Tables 5 and 6 present the size adjusted power of the statistics when there is negative autocorrelation of ε_t ($\phi = -0.5$). In this case we observe that the statistics we derive (LMP_m, BNM_e, BNM_a, and $BEPO_1$) have superior power to their counterparts that do not assume autocorrelation of the error term (LMP, BNM, and $BEPO$) for $\rho \leq 0.94$ and $N = 200$. For smaller sample sizes power of our statistics is very close (almost indistinguishable) to the power of LMP, BNM, and $BEPO$.

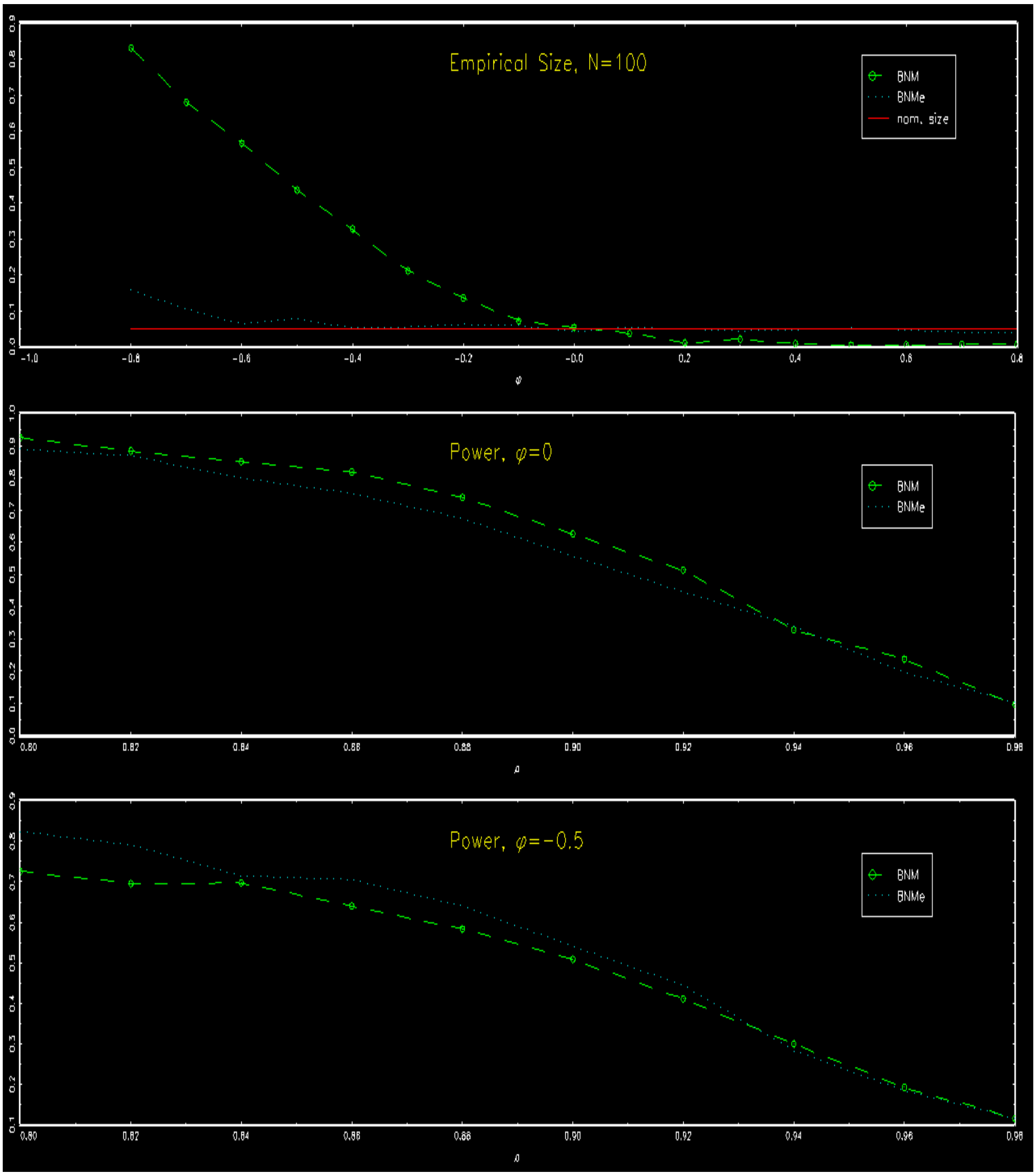
Table 5. Size-adjusted power of the tests for model (21), $\phi = -0.5$

N	ρ	LMP	BNM	$BEPO$	LMP_m	BNM_e	BNM_a	$BEPO_1$
50	0.80	0.150	0.389	0.394	0.162	0.406	0.406	0.403
	0.82	0.147	0.399	0.406	0.140	0.368	0.367	0.366
	0.84	0.116	0.340	0.348	0.154	0.309	0.308	0.307
	0.86	0.129	0.256	0.269	0.119	0.292	0.288	0.288
	0.88	0.110	0.221	0.226	0.131	0.247	0.244	0.244
	0.90	0.130	0.201	0.205	0.126	0.196	0.196	0.196
	0.92	0.116	0.139	0.144	0.112	0.166	0.165	0.164
	0.94	0.109	0.118	0.119	0.092	0.121	0.121	0.119
	0.96	0.103	0.093	0.100	0.084	0.094	0.094	0.091
	0.98	0.073	0.055	0.056	0.058	0.086	0.085	0.085
100	0.80	0.188	0.726	0.725	0.234	0.824	0.827	0.823
	0.82	0.180	0.695	0.693	0.247	0.792	0.793	0.791
	0.84	0.197	0.697	0.695	0.196	0.715	0.719	0.717
	0.86	0.197	0.640	0.637	0.215	0.706	0.712	0.707
	0.88	0.146	0.583	0.579	0.188	0.640	0.643	0.641
	0.90	0.186	0.508	0.501	0.202	0.541	0.548	0.542
	0.92	0.177	0.410	0.405	0.170	0.444	0.449	0.445
	0.94	0.132	0.299	0.297	0.154	0.282	0.287	0.283
	0.96	0.142	0.192	0.191	0.113	0.183	0.187	0.185
	0.98	0.081	0.115	0.111	0.085	0.116	0.117	0.117
200	0.80	0.289	0.914	0.914	0.348	0.985	0.985	0.987
	0.82	0.273	0.911	0.911	0.343	0.980	0.980	0.980
	0.84	0.291	0.897	0.897	0.291	0.973	0.973	0.973
	0.86	0.258	0.886	0.886	0.289	0.972	0.972	0.974
	0.88	0.243	0.872	0.873	0.261	0.956	0.956	0.956
	0.90	0.219	0.831	0.832	0.234	0.906	0.906	0.906
	0.92	0.201	0.788	0.789	0.245	0.844	0.844	0.844
	0.94	0.209	0.677	0.680	0.197	0.712	0.712	0.715
	0.96	0.167	0.501	0.504	0.177	0.490	0.490	0.491
	0.98	0.127	0.237	0.239	0.104	0.211	0.211	0.212

Table 6. Size-adjusted power of the tests for model (22), $\phi = -0.5$						
N	ρ	BNM	$BEPO$	BNM_e	BNM_a	$BEPO_1$
50	0.80	0.211	0.209	0.181	0.178	0.167
	0.82	0.196	0.193	0.182	0.178	0.167
	0.84	0.169	0.166	0.159	0.156	0.144
	0.86	0.124	0.120	0.123	0.122	0.109
	0.88	0.097	0.097	0.109	0.108	0.102
	0.90	0.081	0.076	0.091	0.089	0.085
	0.92	0.074	0.073	0.078	0.078	0.073
	0.94	0.074	0.074	0.079	0.077	0.070
	0.96	0.062	0.061	0.071	0.069	0.063
	0.98	0.053	0.051	0.055	0.055	0.050
100	0.80	0.603	0.585	0.646	0.645	0.638
	0.82	0.552	0.541	0.578	0.572	0.569
	0.84	0.532	0.517	0.510	0.509	0.505
	0.86	0.445	0.432	0.415	0.410	0.406
	0.88	0.365	0.348	0.376	0.375	0.374
	0.90	0.296	0.284	0.286	0.284	0.276
	0.92	0.218	0.208	0.215	0.214	0.210
	0.94	0.138	0.129	0.153	0.153	0.148
	0.96	0.104	0.102	0.098	0.097	0.095
	0.98	0.081	0.076	0.072	0.072	0.070
200	0.80	0.873	0.873	0.939	0.939	0.939
	0.82	0.843	0.843	0.910	0.910	0.910
	0.84	0.834	0.836	0.877	0.877	0.878
	0.86	0.767	0.770	0.824	0.824	0.825
	0.88	0.701	0.703	0.750	0.750	0.756
	0.90	0.621	0.625	0.660	0.660	0.660
	0.92	0.465	0.468	0.496	0.496	0.497
	0.94	0.308	0.310	0.355	0.357	0.361
	0.96	0.186	0.187	0.180	0.182	0.182
	0.98	0.096	0.097	0.096	0.096	0.098

Observing the results in tables 1-6 we notice that BNM_e and BNM_a have very similar finite sample properties, which suggests that a BNM statistic is robust to different norms. Figure (1) summarizes the argument about choosing the modified statistics (the ones which account for autocorrelation in the error term) instead of the ones that do not account for autocorrelation in the error term. In this figure BNM_e and BNM statistics are compared for $T = 100$. The first panel presents the empirical size of the statistics for $-0.8 \leq \phi \leq 0.8$. We observe that the comparative performance of BNM_e is impressive. Its empirical size is very close to the nominal one, while the empirical size of BNM is much higher for negative values of ϕ, and smaller for positive values. This is expected, since BNM_e is derived in order to reduce size distortion. In the second panel the power of BNM_e and BNM are given in the case of no autocorrelation. We note that this is a case which is not the one our statistic BNM_e was designed for, i.e. there is no nuisance parameter to be accounted for. Even in this case the power loss in comparison to BNM is small. Finally, in the 3rd panel the size-adjusted power of the

Figure 1: Comparison of BNM_e and BNM statistics for $T = 100$



two statistics is given for $\phi = -0.5$. For the most alternative

7 Conclusion

In this paper we derive statistics for testing the unit root hypothesis in the presence of autocorrelated errors. Based on optimality criteria proposed by Forchini and Marsh (2000) we generalise the test statistics in order to take into consideration possible autocorrelation in the error term. Our simulation study suggests that these statistics have substantially smaller size distortions. Additionally they have very good power properties. Even in the case that there is no autocorrelation in the error term we find that our statistics have power very close to the power of the statistics that do not assume autocorrelation in the error term. In the presence of a negative MA parameter, we find that our statistics have higher size adjusted power in most cases. Summarising the above arguments, we can claim that we improve a lot on controlling the size of the statistics without having substantial loss in terms of power.

Finally, we find that statistics BNM_e, BNM_a, and $BEPO_1$ have very similar size distortion and power, under different specifications. LMP_m statistic exhibits comparatively very small size distortion, but its power is extremely low, so we would favour the use of BNM_e, BNM_a, and $BEPO_1$.

8 Technical Appendix and Proofs

Proposition A1. *Lag matrix $L^{(i)}$ commutes with any other lag matrix of different or same order $L^{(j)}$ and*

$$\begin{aligned} K_\phi T_\rho &= T_\rho K_\phi, \\ T_1^{-1} K_\phi^{-1} &= K_\phi^{-1} T_1^{-1}, \\ T_1' K_\phi' &= K_\phi' T_1', \\ \left(T_1^{-1}\right)' \left(K_\phi^{-1}\right)' &= \left(K_\phi^{-1}\right)' \left(T_1^{-1}\right)', \\ K_\phi^{-1} T_\rho &= T_\rho K_\phi^{-1}, \\ \left(K_\phi^{-1}\right)' T_\rho' &= T_\rho' \left(K_\phi^{-1}\right)', \end{aligned}$$

given that K_ϕ and T_ρ are invertible.

Proof. Lag matrix $L^{(i)}$ commutes with any other lag matrix of the same or different order $L^{(j)}$ and:

$$L^{(i)} L^{(j)} = L^{(j)} L^{(i)} = \begin{cases} L^{(i+j)}, & \text{for } i + j \leq N - 1 \\ \mathbf{0}, & \text{for } i + j > N - 1. \end{cases} \quad (23)$$

Noting the definitions in (5) and (6) and the commutative property of lag matrix $L^{(i)}$ (23) we have:

$$\begin{aligned} K_\phi T_\rho &= \left(I_N + \sum_{i=1}^{q} \phi_i L^{(i)}\right)\left(I_N - \rho L^{(1)}\right) = I_N - \rho L^{(1)} + \sum_{i=1}^{q} \phi_i L^{(i)} - \left(\sum_{i=1}^{q} \phi_i L^{(i)}\right) \rho L^{(1)} \\ &= I_N - \rho L^{(1)} + \sum_{i=1}^{q} \phi_i L^{(i)} - \rho \sum_{i=1}^{q} \phi_i L^{(i)} L^{(1)} = I_N - \rho L^{(1)} + \sum_{i=1}^{q} \phi_i L^{(i)} - \rho L^{(1)} \sum_{i=1}^{q} \phi_i L^{(i)} \\ &= I_N - \rho L^{(1)} + \left(I_N - \rho L^{(1)}\right) \sum_{i=1}^{q} \phi_i L^{(i)} = \left(I_N - \rho L^{(1)}\right)\left(I_N + \sum_{i=1}^{q} \phi_i L^{(i)}\right) = T_\rho K_\phi. \quad (24) \end{aligned}$$

Result in (24) means that K_ϕ commutes with T_ρ (and with T_1 which is a special case of T_ρ). Given that K_ϕ and T_ρ are nonsingular matrices, we can easily show that their respective inverse and transpose matrices commute with each other as well:

$$K_\phi T_\rho = T_\rho K_\phi \Leftrightarrow (K_\phi T_\rho)^{-1} = (T_\rho K_\phi)^{-1} \Leftrightarrow T_\rho^{-1} K_\phi^{-1} = K_\phi^{-1} T_\rho^{-1}, \quad (25)$$

$$K_\phi T_\rho = T_\rho K_\phi \Leftrightarrow (K_\phi T_\rho)' = (T_\rho K_\phi)' \Leftrightarrow T_\rho' K_\phi' = K_\phi' T_\rho', \quad (26)$$

and combining (25) and (26) we get

$$\left(T_1^{-1}\right)' \left(K_\phi^{-1}\right)' = \left(K_\phi^{-1}\right)' \left(T_1^{-1}\right)'. \quad (27)$$

Finally, using (24) we show that T_ρ commutes with K_ϕ^{-1}

$$K_\phi T_\rho = T_\rho K_\phi \Rightarrow T_\rho = K_\phi^{-1} T_\rho K_\phi \Rightarrow T_\rho K_\phi^{-1} = K_\phi^{-1} T_\rho, \quad (28)$$

and transposing both sides of (24) we can show that $\left(K_\phi^{-1}\right)' T_\rho' = T_\rho' \left(K_\phi^{-1}\right)'$.

Proposition A2. *Let $S = T_1^{-1}\varepsilon$ and $\sigma^2 = E\left(\varepsilon_1^2\right)$. Under the assumptions of Theorem 5, the following limit theory applies under the null hypothesis $H_0 : \rho = 1$ as $N \to \infty$:*

(i) $N^{-1} S'\varepsilon \Rightarrow \frac{1}{2}\sigma^2 \left[W^2(1) + 1\right]$

(ii) $N^{-1} S' P_Z \varepsilon \Rightarrow \sigma^2 W(1) \int_0^1 W(r)dr$

(iii) $N^{-2} S' T_1^{-1} P_Z \varepsilon \Rightarrow \sigma^2 W(1) \int_0^1 rW(r)dr$

(iv) $N^{-2} \left(T_1^{-1} P_Z \varepsilon\right)' T_1^{-1} P_Z \varepsilon \Rightarrow \frac{1}{3}\sigma^2 W(1)^2$

(v) $N^{-1} \left(T_1^{-1} P_Z \varepsilon\right)' P_Z \varepsilon \Rightarrow \frac{1}{2}\sigma^2 W(1)^2$

(vi) $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon \Rightarrow \sigma^2 W(1)\left(W(1) - \int_0^1 W(r)dr\right)$

(vii) $N^{-1}\nu'\nu \rightarrow_p \sigma^2$

where $W(\cdot)$ denotes standard Brownian motion on $D[0,1]$.

Proof. By definition of the matrix T_1^{-1}, S_t is a unit root process with i.i.d. innovations ε_t. Also, using the particular form of the matrix X of deterministics, it is easy to obtain the following identities:

$$P_Z\varepsilon = \frac{1}{N-1}\left[(N-1)\varepsilon_1, u_N - \varepsilon_1, ..., u_N - \varepsilon_1\right]'$$

and

$$T_1^{-1}P_Z\varepsilon = \frac{1}{N-1}\left[(N-1)\varepsilon_1, u_{N-1} + (N-1)\varepsilon_1, ..., (N-1)u_{N-1} + (N-1)\varepsilon_1\right]'.$$

In what follows, we make use of standard unit root asymptotics, see e.g. Phillips (1987) and Phillips and Perron (1988).

For part (i), we have

$$\begin{aligned}
N^{-1}S'\varepsilon &= N^{-1}\sum_{i=1}^{N} S_i\varepsilon_i = N^{-1}\left(\sum_{i=1}^{N} S_{i-1}\varepsilon_i + \sum_{i=1}^{N}\varepsilon_i^2\right) \\
&= N^{-1}\sum_{i=1}^{N} S_{i-1}\varepsilon_i + N^{-1}\sum_{i=1}^{N}\varepsilon_i^2 \\
&\Rightarrow \frac{1}{2}\sigma^2\left\{[W(1)]^2 - 1\right\} + \sigma^2 \\
&= \frac{1}{2}\sigma^2\left[W^2(1) + 1\right].
\end{aligned}$$

For part (ii),

$$\begin{aligned}
\frac{1}{N}S'P_Z\varepsilon &= \frac{1}{N-1}\left[S_1(N-1)\varepsilon_1 + \sum_{i=2}^{N} S_i(S_N - \varepsilon_1)\right] \\
&= \frac{1}{N(N-1)}S_N\sum_{i=2}^{N} S_i + O_p\left(\frac{1}{N^2}\sum_{i=2}^{N} S_i\right) \\
&= \frac{1}{N^{1/2}}S_N\frac{1}{N^{3/2}}\sum_{i=2}^{N} S_i + O_p(N^{-1/2}) \\
&\Rightarrow \sigma^2 W(1)\int_0^1 W(r)dr.
\end{aligned}$$

For part (iii),

$$
\begin{aligned}
\frac{1}{N^2}S'T_1^{-1}P_Z\varepsilon &= \frac{1}{N^2}\frac{1}{N-1}\sum_{i=1}^{N}\left\{S_i\left[(i-1)\,S_{N-1}+(N-1)\,\varepsilon_1\right]\right\} \\
&= \frac{1}{N^2}\frac{1}{N-1}S_{N-1}\sum_{i=1}^{N}S_i i+O_p\left(\frac{1}{N^2}\sum_{i=1}^{N}S_i\right) \\
&= \frac{1}{N^{1/2}}S_{N-1}\frac{1}{N^{5/2}}\sum_{i=1}^{N}S_i i+O_p\left(N^{-1/2}\right) \\
&\Rightarrow \sigma^2 W(1)\int_0^1 rW(r)dr.
\end{aligned}
$$

For part (iv),

$$
\begin{aligned}
\frac{1}{N^2}\left(T_1^{-1}P_Z\varepsilon\right)'T_1^{-1}P_Z\varepsilon &= \frac{1}{N^2}\left(\frac{1}{N-1}\right)^2\sum_{i=1}^{N}\left[(i-1)\,S_{N-1}+(N-1)\,\varepsilon_1\right]^2 \\
&= \frac{S_{N-1}^2}{(N-1)^2}\frac{1}{N^2}\sum_{i=1}^{N}(i-1)^2+O_p\left(\frac{1}{N}S_{N-1}\right) \\
&= \left[1+o\left(1\right)\right]\frac{S_{N-1}^2}{3N}+O_p\left(N^{-1/2}\right) \\
&\Rightarrow \frac{1}{3}\sigma^2 W^2(1).
\end{aligned}
$$

For part (v),

$$
\begin{aligned}
\frac{1}{N}\left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon &= \frac{1}{N}\left(\frac{1}{N-1}\right)^2\left\{(N-1)^2\varepsilon_1^2+\sum_{i=1}^{N-1}\left[\left(iS_{N-1}+(N-1)\,\varepsilon_1\right)\left(S_N-\varepsilon_1\right)\right]\right\} \\
&= \frac{1}{N\left(N-1\right)^2}S_{N-1}S_N\sum_{i=1}^{N-1}i+O_p\left(\frac{1}{N^3}S_{N-1}\sum_{i=1}^{N-1}i\right) \\
&= \left[1+o\left(1\right)\right]\frac{1}{2}\frac{S_{N-1}}{N^{1/2}}\frac{S_N}{N^{1/2}}+O_p\left(N^{-1/2}\right) \\
&\Rightarrow \frac{1}{2}\sigma^2 W^2(1)
\end{aligned}
$$

For part (vi),

$$\begin{aligned}\frac{1}{N}\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon &= \frac{1}{N}\left\{\frac{1}{N-1}S_{N-1}\sum_{i=1}^{N}i\varepsilon_i-\frac{1}{N-1}S_{N-1}S_N+\varepsilon_1S_N\right\}\\ &= \left[1+o\left(1\right)\right]\frac{S_{N-1}}{N^{1/2}}\frac{1}{N^{3/2}}\sum_{i=1}^{N}i\varepsilon_i+O_p\left(N^{-1/2}\right)\\ &\Rightarrow \sigma W(1)\left(\sigma W(1)-\sigma\int_0^1 W(r)dr\right).\end{aligned}$$

For part (vii), recall that, under H_0, $Z=K_{\hat{\phi}}^{-1}T_1X$ and $T_1u=K_\phi\varepsilon$ which gives

$$\begin{aligned}\nu &= M_ZK_{\hat{\phi}}^{-1}T_1\left(X\beta+u\right)=M_ZK_{\hat{\phi}}^{-1}T_1u\\ &= M_ZK_{\hat{\phi}}^{-1}K_\phi\varepsilon=\left[I+o_p\left(1\right)\right]M_Z\varepsilon\end{aligned}$$

using the fact that $\hat{\phi}-\phi=o_p\left(1\right)$. Therefore, since

$$\varepsilon'P_Z\varepsilon=\varepsilon_1^2+\frac{1}{N-1}\left(u_N-\varepsilon_1\right)^2=O_p\left(1\right),$$

the weak law of large numbers yields

$$\begin{aligned}\frac{1}{N}\nu'\nu &= \left[I+o_p\left(1\right)\right]\frac{1}{N}\varepsilon'M_Z\varepsilon\\ &= \left[I+o_p\left(1\right)\right]\left\{\frac{1}{N}\varepsilon'\varepsilon+O_p\left(N^{-1}\right)\right\}\rightarrow_p\sigma^2.\end{aligned}$$

Proof of Lemma 1. Using the commutation results given in Proposition A1 we get

$$\begin{aligned}\Sigma_{\rho,\phi} &= K_\phi^{-1}T_1T_\rho^{-1}K_\phi K_\phi'\left(T_\rho^{-1}\right)'T_1'\left(K_\phi^{-1}\right)'=T_1K_\phi^{-1}K_\phi T_\rho^{-1}\left(T_\rho^{-1}\right)'K_\phi'\left(K_\phi^{-1}\right)'T_1'\\ &= T_1T_\rho^{-1}\left(T_\rho^{-1}\right)'T_1'.\end{aligned}$$

Proof of Theorem 2. Let $\rho=1-\gamma$, so for $\gamma>0$ $H_1:\rho<1$, and for $\gamma<0$ $H_1:\rho>1$. We differentiate (13) with respect to $\gamma=0$ and then set $\gamma=0:$

$$\begin{aligned}sign(\gamma)\left.\frac{dpdf(\boldsymbol{v})}{d\gamma}\right|_{\gamma=0} &= \left.\frac{d\left[|A|^{-\frac{1}{2}}\left(\boldsymbol{v}'A^{-1}\boldsymbol{v}\right)^{-\frac{N-k}{2}}\right]}{d\gamma}\right|_{\gamma=0}\\ &= d(N,k)+\frac{N-k}{2}\left.\frac{\partial\left(\boldsymbol{v}'A^{-1}\boldsymbol{v}\right)}{\partial\gamma}\right|_{\gamma=0},\end{aligned}$$

where the first part of the sum does not depend upon $\boldsymbol{v}$.

$$\left.\frac{\partial\left(\boldsymbol{v}'A^{-1}\boldsymbol{v}\right)}{\partial\gamma}\right|_{\gamma=0} = \boldsymbol{v}'C'\left(K_\phi^{-1}L^{(1)}T_1^{-1}K_\phi + K_\phi'\left(T_1^{-1}\right)'\left(L^{(1)}\right)'\left(K_\phi^{-1}\right)'\right)C\boldsymbol{v}$$
$$= \frac{\nu'\left(K_\phi^{-1}L_NT_1^{-1}K_\phi + \left(K_\phi^{-1}L_NT_1^{-1}K_\phi\right)'\right)\nu}{\nu'\nu}$$

where $\nu = M_ZK_\phi^{-1}T_1y$. Noting the definition of the most powerful similar test in (14), the first part of the theorem is proved. The second part is proved analogously.

Proof of Theorem 3. The most powerful similar test of size α is given by (14) which can be rewritten as:

$$\frac{y'T_1'\left(K_\phi^{-1}\right)'C\left(C'\Sigma_\rho C\right)^{-1}C'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} < k_\alpha$$

Forchini and Marsh (2000) in lemma 3 show that the matrix

$$Q = C'B^{-1}C - \left(C'BC\right)^{-1}$$

is positive semi-definite. Applying this in our case gives the following result:

$$\frac{y'T_1'\left(K_\phi^{-1}\right)'C\left(C'\Sigma_\rho C\right)^{-1}C'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} \leq \frac{y'T_1'\left(K_\phi^{-1}\right)'CC'\Sigma_\rho^{-1}CC'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y}$$
$$\frac{y'T_1'\left(K_\phi^{-1}\right)'C\left(C'\Sigma_\rho C\right)^{-1}C'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} \leq \frac{\nu'\Sigma_\rho^{-1}\nu}{\nu'\nu},$$

where ν is defined above. So (14) is bounded above by the ratio of quadratic forms in ν. Inverting Σ_ρ and expressing T_ρ as $T_\rho = I_N - \rho L^{(1)}$:

$$\begin{gathered}\Sigma_\rho^{-1} = \left[T_1T_\rho^{-1}\left(T_\rho^{-1}\right)'T_1'\right]^{-1} = \left(T_1^{-1}\right)'T_\rho'T_\rho T_1^{-1} = \\ \left(T_1^{-1}\right)'\left(I_N-\rho L^{(1)}\right)'\left(I_N-\rho L^{(1)}\right)T_1^{-1} = \\ \left(T_1^{-1}\right)'\left(I_N-\rho L^{(1)}\right)T_1^{-1} - \rho\left(T_1^{-1}\right)'L^{(1)\prime}\left(I_N-\rho L^{(1)}\right)T_1^{-1} = \\ \left(T_1^{-1}\right)'T_1^{-1} - \rho\left(T_1^{-1}\right)'L^{(1)}T_1^{-1} - \rho\left(T_1^{-1}\right)'L^{(1)\prime}T_1^{-1} + \rho^2\left(T_1^{-1}\right)'L^{(1)\prime}L^{(1)}T_1^{-1}.\end{gathered} \tag{29}$$

From equation (29) and the definition of the matrix $\Psi(\nu)$ we obtain:

$$\frac{\nu'\Sigma_\rho^{-1}\nu}{\nu'\nu} = \begin{pmatrix} 1 & -\rho \end{pmatrix}\Psi(\nu)\begin{pmatrix} 1 \\ -\rho \end{pmatrix} \tag{30}$$

So a sufficient condition for (14) to hold is that the positive definite matrix $\Psi(\nu)$ is small with respect to some norm. We can find statistics such that $\Pr\{\|\Psi(\nu)\| < k_\alpha | H_0\} = a$.

Proof of Theorem 4. BEPO criterion is:

$$\frac{l(\rho^*)'\Psi(y)l(\rho^*)}{y'\Omega^{-1}(\rho_0)y} < k_a \tag{31}$$

where k_α is such that the size of the test is α and ρ^* is the value of ρ which minimises (30). We differentiate (30) with respect to parameter ρ and set it equal to zero. From equations (18) and (30) we get:

$$\begin{aligned}
\begin{pmatrix} 1 & -\rho \end{pmatrix} \Psi(\nu) \begin{pmatrix} 1 \\ -\rho \end{pmatrix} &= \left(\rho^2\nu'\Psi_{22}\nu - 2\rho\nu'\Psi_{12}\nu + \nu'\Psi_{11}\nu\right). \\
\frac{1}{\nu'\nu}\frac{\partial\left(\rho^{*2}\nu'\Psi_{22}\nu - 2\rho^*\nu'\Psi_{12}\nu + \nu'\Psi_{11}\nu\right)}{\partial\rho^*} &= 0 \Rightarrow \\
\frac{2}{\nu'\nu}\left(\rho^*\psi_{22} - \psi_{12}\right) &= 0 \Rightarrow \rho^* = \frac{\psi_{12}}{\psi_{22}}. \qquad (32)
\end{aligned}$$

Combining condition (31) with (32) and values given by (16) and (17) we get the BEPO statistic. Also we need to note that $\psi_{22} \geq 0$ since Ψ_{22} is a positive semi-definite matrix, so $\frac{\partial^2\left(\rho^2\psi_{22} - 2\rho\psi_{12} + \psi_{11}\right)}{\partial\rho^2} \geq 0$.

The second part of the theorem is proved by substituting (32) and in (3).

Proof of Theorem 5. We make repeated use of the limit theory established in Proposition A2. For notational simplicity, define

$$\psi_{11} = \nu'\Psi_{11}\nu,\ \psi_{22} = \nu'\Psi_{22}\nu \text{ and } \psi_{12} = \nu'\Psi_{12}\nu$$

and note that

$$\begin{aligned}
\psi_{22} &= \psi_{11} - 2\left(T_1^{-1}\nu\right)'\nu + \nu'\nu \\
&= \psi_{11} - 2\left[S'\varepsilon - S'P_Z\varepsilon + \left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon\right] - 2\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon + \nu'\nu \qquad (33)
\end{aligned}$$

and

$$\psi_{12} = \psi_{11} - S'\varepsilon + S'P_Z\varepsilon - \left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon. \tag{34}$$

For part (i), it is clear Proposition A2 and (33) and (34) we obtain that $\psi_{22} = \psi_{11} + O_p(N)$ and $\psi_{12} = \psi_{11} + O_p(N)$. Now by Proposition A2,

$$\begin{aligned}
\frac{1}{N^2}\psi_{11} &= \frac{1}{N^2}S'S - \frac{2}{N^2}S'T_1^{-1}P_Z\varepsilon + \frac{1}{N^2}\left(T_1^{-1}P_Z\varepsilon\right)'T_1^{-1}P_Z\varepsilon \\
&\Rightarrow \sigma^2\left\{\int_0^1 W(r)^2dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W(1)^2\right\}. \qquad (35)
\end{aligned}$$

The BNM test statistic is given by

$$\begin{aligned}
\frac{1}{N}\|\Psi(\nu)\| &= \frac{1}{N^{-1}\nu'\nu}\left\|\frac{1}{N^2}\begin{bmatrix}\psi_{11} & \psi_{12}\\ \psi_{12} & \psi_{22}\end{bmatrix}\right\| \\
&= \frac{1}{N^{-1}\nu'\nu}\left\|\frac{1}{N^2}\begin{bmatrix}\psi_{11} & \psi_{12}\\ \psi_{12} & \psi_{22}\end{bmatrix}\right\| \\
&= \frac{1}{N^{-1}\nu'\nu}\left\|\frac{\psi_{11}}{N^2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\right\| + O_p\left(N^{-1}\right)
\end{aligned}$$

and the result follows from Proposition A2(vi) and (35).

For part (ii),note that

$$\begin{aligned}
\psi_{11}\psi_{22}-\psi_{12}^2 &= \psi_{11}\left\{\psi_{11}-2\left[S'\varepsilon - S'P_Z\varepsilon + \left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon\right] + 2\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon + \nu'\nu\right\} \\
&\quad -\left\{\psi_{11}-\left[S'\varepsilon - S'P_Z\varepsilon + \left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon\right]\right\}^2 \\
&= \psi_{11}\left\{2\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon + \nu'\nu\right\} + O_p\left(N^2\right),
\end{aligned}$$

which yields

$$\frac{1}{N^3}\left(\psi_{11}\psi_{22}-\psi_{12}^2\right) = \frac{\psi_{11}}{N^2}\left\{\frac{2\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon}{N} + \frac{\nu'\nu}{N}\right\} + O_p(N^{-1}).$$

As before, $N^{-2}\psi_{22} = N^{-2}\psi_{11} + O_p(N^{-1})$ so the asymptotic distribution of the BEPO_1 statistic is given by

$$\begin{aligned}
BEPO_1 = \frac{1}{\nu'\nu}\frac{\psi_{11}\psi_{22}-\psi_{12}^2}{\psi_{22}} &= \frac{1}{N^{-1}\nu'\nu}\left\{\frac{2\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon}{N} + \frac{\nu'\nu}{N}\right\} + O_p(N^{-1}) \\
&\xrightarrow{L} \frac{1}{\sigma^2}\left\{2\sigma^2 W(1)\left(W(1)-\int_0^1 W(r)dr\right)+\sigma^2\right\} \\
&= 2W^2(1) - 2W(1)\int_0^1 W(r)dr + 1.
\end{aligned}$$

Finally, using result

$$\begin{aligned}
\frac{1}{N}\left(\psi_{12}-\psi_{22}\right) &= \frac{1}{N}\left(S'\varepsilon - S'P_Z\varepsilon + \left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon - \nu'\nu\right) \\
&\xrightarrow{L} \frac{1}{2}\sigma^2\left[W^2(1)+1\right] - \sigma^2 W(1)\int_0^1 W(r)dr \\
&\quad + \frac{1}{2}\sigma^2 W^2(1) - 2\sigma^2 W(1)\left(W(1)-\int_0^1 rW(r)dr\right) - \sigma^2 \\
&= \sigma^2\left\{\begin{array}{c}\frac{1}{2}\left[W^2(1)-1\right] - W(1)\int_0^1 W(r)dr \\ +\frac{1}{2}W^2(1) - 2W(1)\left(W(1)-\int_0^1 rW(r)dr\right) - 1\end{array}\right\}.
\end{aligned}$$

We get the asymptotic distribution of BEPO_2 statistic which is given by

$$
\begin{aligned}
BEPO_2 &= N\left|\frac{\psi_{12}-\psi_{22}}{\psi_{22}}\right| = \left|\frac{N^{-1}(\psi_{12}-\psi_{22})}{N^{-2}\psi_{22}}\right| \\
&= \left|\frac{\frac{1}{2}\left[W^2(1)-1\right] - W(1)\int_0^1 W(r)dr + \frac{1}{2}W^2(1) - 2W(1)\left(W(1)-\int_0^1 W(r)dr\right) - 1}{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)}\right| \\
&= \left|\frac{\frac{1}{2}W^2(1) - \frac{1}{2} - W(1)\int_0^1 W(r)dr + \frac{1}{2}W^2(1) - 2W^2(1) + 2W(1)\int_0^1 W(r)dr - 1}{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)}\right| \\
&= \left|\frac{-W^2(1) + W(1)\int_0^1 W(r)dr - \frac{3}{2}}{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)}\right|.
\end{aligned}
$$

9 References

Banerjee, A., Dolado, J. J., Galbraith, J. W. and D. F. Hendry (1993), *Co-integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data,* Oxford University Press.

Cox, D. R. and D. V. Hinkley (1974), *Theoretical Statistics,* Chapman and Hall, London.

DeJong, D. N., Nankervis, J. C., Savin, N.E. and C. H. Whiteman (1992), The power problems of unit root tests in time series with autocorrelated errors, *Journal of Econometrics*, 53, pp. 323-343.

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

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

Engle, R. F. and C. W. J. Granger (1987), Cointegration and error correction: Representation, estimation and testing, *Econometrica,* 51, pp. 251-276.

Forchini, G. F. and P. W. Marsh (2000), Exact inference for the unit root hypothesis, Discussion Paper 00/54, University of York .

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

Granger, C. W. J. and P. Newbold (1974), Spurious regressions in econometrics, *Journal of Econometrics*, 2, pp. 111-120.

Hamilton, J. D., (1994), *Time Series Analysis,* Princeton University Press, New Jersey.

Hillier, G. H., (1987), Classes of similar regions and their power properties for some econometric testing problems, *Econometric Theory,* 3, pp. 1-44.

Horn R.A. and C.R. Johnson, (1985), *Matrix Analysis.* Cambridge University Press, Cambridge

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, pp. 231-254.

Lehmann, E. L., (1986), *Testing Statistical Hypotheses,* Wiley, New York.

Lehmann, E. L. and C. Stein, (1948), Most powerful tests of composite hypotheses, *The Annals of Mathematical Statistics*, 19, pp. 495-516.

Maddala, G. S. and I. M. Kim (1998), *Unit Roots, Cointegration, and Structural Change,* Cambridge University Press, Cambridge.

Marsh, P. (2005), A measure of discrimination for the unit root hypothesis, Discussion Paper 05/02, University of York .

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

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

Phillips, P. C. B. and P. Perron (1988), Testing for a unit root in time series regression, *Biometrika*, 75:2, pp. 335-346.

Plosser, C. I. and Schwert, G. W., (1978), Money, income, and sunspots: Measuring economic relationships and the effects of differencing, *Journal of Monetary Economics*, 4, pp. 637-660.

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

Schwert, G. W., (1989), Tests for unit roots: A Monte Carlo investigation, *Journal of Business & Economic Statistics*, 7:2, pp. 147-159.

Zaman, A., (1996), *Statistical Foundations for Econometric Techniques,* Academic Press, London.