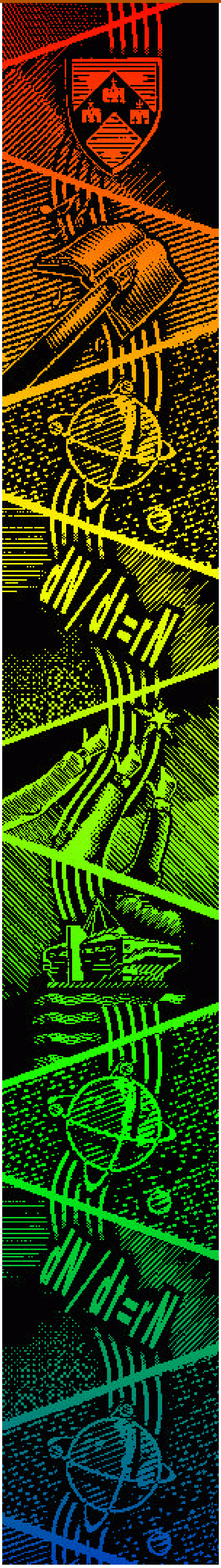

THE UNIVERSITY of York

Discussion Papers in Economics

No. 2000/54

Exact Inference for the Unit Root Hypothesis

by

Giovanni Forchini and Patrick Marsh

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Exact Inference for the Unit Root Hypothesis*

Giovanni Forchini†
Department of Economics
University of York
Heslington, York
YO10 5DD
Tel. No. +44 (0)1904 433076
e-mail: gf7@york.ac.uk

Patrick Marsh‡
Departments of Economics and Mathematics
University of York
Heslington, York
YO10 5DD
Tel. No. +44 (0)1904 433084
e-mail: pwnm1@york.ac.uk

July 1998
This version: December 1999

Abstract

This paper suggests operational procedures to construct similar tests for the autoregressive parameter for both the dynamic regression model and the linear regression model with autocorrelated errors. For both models, we characterize similar tests for the null hypothesis that the autoregressive parameter is unity, and construct "optimal" tests in cases where uniformly most powerful similar tests do not exist. We thus show that classical statistical principles can be successfully used even in areas of econometrics and statistics where they have seldom been applied before (i.e. time series models). Moreover, we derive saddlepoint approximations for the density and distribution functions of some of the tests statistics derived, and thus provide an alternative to numerical or asymptotic methods. By focusing upon the finite sample, rather than asymptotic, properties, key features of the testing problem become much more apparent.

*Thanks are due to Karim Abadir, Grant Hillier, Chris Orme, Simon Peters, Jan Podivinsky, Robert Taylor and participants at a University of York internal seminar and at the ESRC study group meeting, Bristol, U.K., July 1998. Comments by Peter Phillips and two anonymous referees have helped improve the exposition of this paper, and were gratefully received.

†Supported by ESRC research fellowship under grant no. R000236627.

‡Leverhulme research fellow in Economics and Mathematics.

1 Introduction

Testing for a unit root in time series has become the focus of attention for much of applied and theoretical econometrics. The problem has its own intrinsic interest to theorists due to the various asymptotic implications on the properties of estimators and tests. More importantly, for applied work establishing the presence of unit roots remains the cornerstone of the analysis of possibly cointegrated systems: first as a preliminary tool to test for integration in the variables of interest, and second as a test for cointegration itself.

In this paper we develop an optimality theory for tests for a unit root in finite samples. This is achieved upon application of classical testing procedures in the presence of nuisance parameters (the construction of similar tests), and of the Neyman-Pearson lemma. These ideas have been successfully applied elsewhere, but not yet fully in time series analysis. Heuristically, what we seek is a procedure which provides *control* over the size of the test, and delivers *large* power under the alternative.

Two commonly used models are the *dynamic linear regression model*, and the *regression model with autocorrelated errors*. The former, considered by amongst others, Dickey and Fuller (1979) and (1981), Evans and Savin (1981) and (1984), Phillips and Perron (1988), Faust (1996) and Dufour and Kiviet (1998), takes the form

$$\begin{aligned} y_i &= z_i'\beta + \rho y_{i-1} + \varepsilon_i \\ i &= 1,..,N, \end{aligned}$$

where, z_i is a $k \times 1$ vector of explanatory variables and the ε_i are usually $NID(0, \sigma^2)$. For asymptotic treatments of the problem the ε_i may follow some more general process, as in Phillips and Perron (1988). The hypothesis of interest is $\rho = 1$, often with the restriction $\beta = \beta_0$ (usually 0). Tests for this hypothesis have generally been asymptotically motivated. Monte Carlo studies (for instance Dickey and Fuller (1981)) have shown that they retain reasonable power properties even in small samples (although often this is because the simulated models have far more simple structure, than the model under which the tests were developed). However, Faust (1996) has shown that the size of certain testing procedures in this model may tend to one, if the number of nuisance parameters grows asymptotically. Recently, the finite sample properties of tests in dynamic models have been investigated by Dufour and Kiviet (1998). However, their analysis was limited to the demonstration that a particular test statistic has distribution free of nuisance parameters under the null hypothesis. No attempt to characterize the class of similar tests was made, nor to choose an 'optimal' test from within this class.

Regressions with autocorrelated errors, considered by Sargan and Bhargava (1983), Dufour and King (1991) and Elliott, Rothenberg and Stock (1996), can be written as

$$\begin{aligned} y_i &= z_i'\beta + v_i \\ v_i &= \rho v_{i-1} + \varepsilon_i, \end{aligned}$$

with the same specifications as above. The null hypothesis is again, $\rho = 1$. Notice that this model may be written as a dynamic regression model with restrictions on the coefficients of the explanatory variables (see Dejong *et al* (1992)). In this case small sample procedures have been established, generally following those of Sargan and Bhargava (1983), but rely upon approximating the process $\{v\}_i$, by a circular autoregression (the 'Anderson (1948) approximation'). For this model, Dufour and King (1991) derive locally best invariant and point-optimal invariant tests. Alternatively, Elliott, Rothenberg and Stock (1996), develop asymptotic optimality criteria for this class of models, and suggest tests having asymptotic power close to the asymptotic power envelope. However, there is no guarantee that the tests they propose retain their 'optimal' properties in finite samples.

A recent development in the theory of testing for a unit root is due to Ploberger (1999). There, under the assumption of a locally quadratic likelihood, a "complete class" of tests is characterised, i.e. those tests not having power functions dominated by any other test. Unfortunately, as yet, it has not been established whether any of the standard unit root tests, in particular the likelihood ratio, are members of this class.

In broad outline our strategy is as follows. Utilising the results of Cox and Hinkley (1974) and Hillier (1987) we characterise the class of similar tests for the unit root hypothesis. Since, as is well known, there are no *uniformly most powerful* (UMP) tests we weaken the optimality criteria in two distinct ways. Firstly, we consider the construction of point optimal (PO) tests and *locally most powerful* (LMP) tests, and secondly, suggest approximating the criterion for the UMP tests. This latter class of tests satisfies weaker, sufficient conditions for the power to be large, rather than the stringent necessary and sufficient conditions required for UMP tests. Monte Carlo simulations show that these alternative tests have power very close to the power envelope across a range of parameter values. Many of the tests we propose have a very simple structure in terms of the data, and often coincide with previously suggested tests. Moreover, the simple structure of the test statistics facilitates the construction of analytic approximations to their finite sample distributions via the saddlepoint algorithm. Thus, we provide an alternative to numerical techniques used elsewhere. The analysis of the null and alternative distributions of a test yields some insight into the statistical nature of the problem itself.

The plan of the paper is as follows. The next section summarises the construction of similar tests, and reviews the optimality criteria which will be applied later. Section 3 derives optimal similar tests for a unit root for the linear regression model with autocorrelated errors. Section 4 deals with analogous tests for the dynamic regression model. Two cases are considered according to the coefficients of the exogenous variables being unrestricted (Section 4.1) or restricted under the null (4.2). Section 5 gives a saddlepoint approximation for the densities and distributions of some of the tests statistics proposed. Monte Carlo simulations directed to assess the performance of the tests and of the approximations are contained in Section 6. Finally Section 7 concludes.

2 Test criteria

This section summarises the concept of similarity (exact inference) and discusses the optimality criteria which will be of use later. In the Neyman-Pearson approach we need to find procedures for which the size of the test is known (or at least it is bounded by a constant) and power is large. This involves two distinct stages which are described below.

2.1 Test criteria under the null hypothesis

In order to control the size of a test, the criterion of similarity will be used (for further exposition see, for instance, Cox and Hinkley (1974) or Hillier (1987)).

Let $f(y;\theta_1,\theta_2)$ denote the joint density of some $N\times 1$ vector of observations y, depending upon the parameters θ_1 and θ_2, $k_1\times 1$ and $k_2\times 1$ vectors, respectively. We wish to test the hypothesis $H_0:\theta_1=\theta_1^0$ against $H_1:\theta_1\neq\theta_1^0$, for a fixed vector θ_1^0. In this setup θ_2 is a nuisance parameter, and, in general the size of a test (critical region) for H_0 will depend upon θ_2. Any critical region ω with size independent of θ_2 is called a *similar critical region*.

A critical region ω has *Neyman structure* if under H_0 there exists a sufficient statistic t for θ_2 such that $\Pr(y\in\omega\,|\,t;\theta_1^0,\theta_2)=\alpha$ is constant for all t. Thus ω is composed of a fraction α of the probability content of each contour of constant t. If a critical region of size α has Neyman structure it must be similar. However, for a similar critical region to have Neyman structure, t must be boundedly complete. That is for every bounded function $h(t)$, not depending on the parameters, $E(h(t))=0$ implies that $h(t)=0$ everywhere, except possibly on sets of zero measure. Consequently, if under H_0, t is sufficient for θ_2, and is boundedly complete, then every similar critical region has Neyman structure.

If complete sufficient statistics for the nuisance parameters exist under the null, the problem of selecting a most powerful test is reduced considerably by restricting our attention to similar regions. For a given testing situation, there might be a nonsimilar test, having size (dependent on the nuisance parameters) no larger than a fixed value α, which is uniformly more powerful than any similar test of size α (Lehmann and Stein (1948)). In practice, though, this criticism is not very persuasive, because this result is usually achieved by restricting the class of alternatives, and seldom is there enough prior information to specify such a restrictive alternative hypothesis. Moreover, even if prior information is available, it might still be difficult to find such a test.

Many interesting econometric models are in the class of (curved) exponential models (see van Garderen (1998)), with joint density

$$f(y;\theta) = \exp\left\{\sum\nolimits_{j=1}^{p} t_j \eta_j(\theta) - K_N(\eta(\theta)) + h(y)\right\},$$

where $\theta = (\theta_1', \theta_2')'$ is the $k_1 + k_2 \times 1$ vector of *natural* parameters, the t_j are the p *canonical* (minimal sufficient) *statistics*, and the $\eta_j(\theta)$ are the *canonical parameters*; $K_N(\eta(\theta))$ is the *cumulant function* and $h(y)$ is a normalising constant. For curved exponential models $p > k_1 + k_2$. When $p = k_1 + k_2$, we have a *full* exponential model. For exponential models similar tests can be characterised as follows.

Theorem 1 *Let $f(y;\theta_1,\theta_2)$ denote a curved exponential model. If*
(i) under $H_0 : \theta_1 = \theta_1^0$ there exists a sufficient statistic t for θ_2, of dimension k_2; and
(ii) there is a one-to-one transformation $y \to (t, v)$ such that under H_0, v is independent of t;
then a critical region ω is similar of size α if and only if it has size α in the distribution of v.

Proof. An immediate consequence of Theorem 1, p.142 of Lehmann (1986), is that (i) guarantees bounded completeness, and then Theorem 2.1 of Hillier (1987) establishes the result. ■

2.2 Test criteria under the alternative hypothesis

Although Theorem 1 characterises the class of similar tests for H_0, it is of no help in choosing an optimal test. Ideally, this will be the one having greatest (unconditional) power, P_ω, for every (θ_1, θ_2). In general though, P_ω will depend upon both the nuisance parameters and the value of the interest parameter under the alternative. Hence no UMP test exists. The unit root hypothesis turns out to be such a case. In these cases weaker criteria of optimality must be used. For instance Cox and Hinkley (1974) suggest to maximise the power against a 'typical' alternative (giving *point optimal* tests), or to maximize a weighted average of the power, or finally to construct LMP tests (which maximise the slope of the power function in a neighbourhood of the null hypothesis).

In this paper we suggest two alternative criteria which approximate the most powerful test criterion. Our method may be best illustrated by the following simple example. Consider the problem of testing $H_0 : y \sim N(0, \sigma^2\Omega(\theta_0))$ against $H_1 : y \sim N(0, \sigma^2\Omega(\theta))$, $\theta \neq \theta_0$, where y is an $n \times 1$ random vector and θ is a scalar parameter. Upon application of the Neyman-Pearson lemma, the most powerful similar 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 the size is α. Since the numerator depends upon the value of θ under the alternative, no uniformly most powerful test exists.

A PO test is constructed by choosing a 'representative' value of θ under the alternative and substituting it in (1). The PO test is the most powerful test only if the representative value chosen coincides with the true value of the unknown parameter. The power envelope (PE) is the function describing the power of the PO test when the value of the parameter appearing in the numerator happens to be the true parameter θ. In the absence of uniformly most powerful tests, there is no test having power equal to the power envelope for all values of θ. Therefore, we need to find criteria for the construction of tests having power close to the power envelope for a whole range of θ. In what follows we suggest two alternative criteria.

Bounded Norm Minimising Tests

Suppose that

$$y'\Omega^{-1}(\theta)y \leq l(\theta)'\Psi(y)l(\theta), \tag{2}$$

where $l(\theta)$ is a vector depending only upon θ, and $\Psi(y)$ is a positive definite matrix depending only upon y. If such an inequality holds, then we can approximate the most powerful critical region with the set in which $\Psi(y)$ is 'small' (in some sense to be defined later) for all θ.

Define a norm, $\|.\|$, on the space of positive definite matrices in the usual way, i.e. as a mapping $\|.\| : S \rightarrow \mathbb{R}$, satisfying (i) $\|M\| \geq 0$ for all $M \in S$ and $\|M\| = 0$ if and only if $M = 0$, (ii) $\|\alpha M\| = |\alpha| \, \|M\|$ for all $\alpha \in \mathbb{R}$ and $M \in S$, and (iii) $\|M_1 + M_2\| \leq \|M_1\| + \|M_2\|$, M_1, $M_2 \in S$. Thus a sufficient condition for

$$\frac{l(\theta)'\Psi(y)l(\theta)}{y'\Omega^{-1}(\theta_0)y} < k_\alpha \tag{3}$$

is

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

for a suitable choice of k. That is, for any such norm we may find a k such that (3) holds. Thus, any norm of the matrix $\Psi(y)/y'\Omega^{-1}(\theta_0)y$ delivers a *norm minimising* (NM) test, when (2) holds with equality, or a *bounded norm minimising* (BNM) test otherwise.

Bounded Estimated Point Optimal Tests

Another option we have is that of considering *estimated point optimal* (EPO) tests. For point optimal (PO) tests θ is taken as known under the alternative. In general, θ under the alternative is unknown, however, we can estimate it with the value of θ^* which minimises $l(\theta)'\Psi(y)l(\theta)/\left[y'\Omega^{-1}(\theta_0)y\right]$ for a fixed y. Then, supposing (2) holds with equality, the EPO test consists in rejecting H_0 if

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

where k is chosen so that the size of the test is α. Alternatively, we have the test: reject H_0 if $|\theta^* - \theta_0| < k$, where again k is chosen so that the size is α (if the alternative is in a particular direction this last criterion is modified accordingly). Bounded estimated point optimal (BEPO) tests are defined similarly, but apply when (2) does not hold with equality.

For the models under consideration these criteria yield tests with powers close to the power envelope. In general, though, their power properties will have to be checked on a case by case basis.

3 Testing for unit roots in the errors

The first situation we consider is that of testing for a unit root in the errors in a linear regression model,

$$y = Z\beta + v, \tag{4}$$

where β is a $k \times 1$ vector of parameters, Z an $N \times k$ full rank matrix containing observations on the exogenous variables, $v = (v_1, .., v_N)'$ and

$$\begin{aligned} v_i &= \rho v_{i-1} + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2), \end{aligned} \tag{5}$$

for $i = 1, .., N$, and $v_0 = 0$. We consider the sample size as fixed, and make no assumption on the asymptotic behaviour of the regressors. Given (4) and (5), $y \sim N(Z\beta, \sigma^2 \Sigma_N(\rho))$, where

$$\Sigma_N^{-1}(\rho) = T_\rho' T_\rho \tag{6}$$

and

$$T_\rho = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\rho & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & -\rho & 1 \end{pmatrix}. \tag{7}$$

In this setup the unit root hypothesis, $H_0 : \rho = 1$ against $H_1 : \rho \neq 1$, may be written as

$$H_0 : y \sim N(Z\beta, \sigma^2 \Sigma_N(1)) \tag{8}$$

against

$$H_1 : y \sim N(Z\beta, \sigma^2 \Sigma_N(\rho)). \tag{9}$$

In order to simplify the notation, we transform the model so that $x = T_1 y$, and, letting $W = T_1 Z$, $H_0 : x \sim N(W\beta, \sigma^2 I_N)$ and $H_1 : x \sim N(W\beta, \sigma^2 T_1 \Sigma_N(\rho) T_1')$. Note that for this model only the covariance, not the mean, depends upon the parameter of interest.

In order to characterise the class of similar tests for H_0 we follow Hillier (1987, Section 4). Note that under H_0 the statistics

$$\begin{aligned} \hat{\beta} &= (W'W)^{-1} W'x \\ s^2 &= x' M_W x, \end{aligned}$$

where $M_W = I_N - W(W'W)^{-1}W$, are jointly sufficient for the nuisance parameters (β, σ^2), and that the conditions of Theorem 1 are met. Therefore, all similar critical regions of size α are a fraction α of the surface $\left(\hat{\beta}, s^2\right)$ = constant. To characterise the class of similar regions for H_0 we need to find a transformation of y to $\left(v, \hat{\beta}, s^2\right)$, such that v is independent of $\left(\hat{\beta}, s^2\right)$ under the null hypothesis. First we transform x to $\left(\hat{\beta}, w_1\right)$, where $w_1 = C'x$ and C is a left-orthogonal matrix of dimension $N \times N-k$ and $CC' = M_W$. Note that $\hat{\beta}$ and w_1 are independent, and

$$\hat{\beta} \sim N\left(\beta, \sigma^2 \left(W'W\right)^{-1} W'T_1\Sigma_N(\rho)T_1'W\left(W'W\right)^{-1}\right)$$

$$w_1 \sim N\left(0, \sigma^2 C'T_1\Sigma_N(\rho)T_1'C\right).$$

Then, we let $w_1 = \left(s^2\right)^{1/2} v$, $s^2 = w_1'w_1$, and $v = w_1/(w_1'w_1)^{1/2}$ (the Jacobian is $\frac{1}{2}(s^2)^{\frac{N-k}{2}-1}$) so that the joint density of $\left(v, \hat{\beta}, s^2\right)$ is

$$\begin{aligned} f(v, \hat{\beta}, s^2; \rho, \sigma^2, \beta) &= f\left(\hat{\beta}; \rho, \sigma^2, \beta\right) \frac{1}{2}(2\pi\sigma^2)^{-\frac{N-k}{2}} \left(s^2\right)^{\frac{N-k}{2}-1} \\ & \quad \left|C'T_1\Sigma_N(\rho)T_1'C\right|^{-1/2} \exp\left\{-\frac{s^2}{2\sigma^2} v' \left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1} v\right\} \end{aligned} \tag{10}$$

Therefore, $(N-k)$ vector

$$v = \frac{C'T_1y}{(y'T_1'M_WT_1y)^{1/2}}, \tag{11}$$

characterises the class of similar tests for H_0. Since, under H_0, v has uniform distribution over the unit $(N-k)$-sphere $\left\{v \in \mathbb{R}^{N-k} : v'v = 1\right\}$, every similar critical region ω of size α, will consist of the fraction α of the surface of the unit $(N-k)$-sphere. Notice also that v characterises the class of similar tests for H_0 against every alternative.

Within the class of similar tests, we now must identify those having optimal power properties. The unconditional power of any similar critical region ω, for H_0, is

$$P_\omega = \int_{s^2>0} \int_{\hat{\beta}\in\mathbb{R}^k} \int_\omega f\left(\hat{\beta}; \rho, \sigma^2, \beta\right) f(v, s^2; \rho, \sigma^2, \beta)\, ds^2 d\beta\, (dv)$$

$$= \int_\omega \int_{s^2>0} f(v, s^2; \rho, \sigma^2, \beta)\, ds^2 (dv)$$

$$= \frac{\Gamma(\frac{N-k}{2})}{2\pi^{\frac{N-k}{2}}} \left|C'T_1\Sigma_N(\rho)T_1'C\right|^{-1/2} \int_\omega \left[v' \left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1} v\right]^{-\frac{N-k}{2}} (dv),$$

where (dv) denotes the un-normalised measure on the unit $(N-k)$-sphere. Hence, the most powerful similar test of size α takes the form

$$v' \left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1} v = \frac{y'T_1'C \left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1} C'T_1y}{y'T_1'M_WT_1y} < k_\alpha. \qquad (12)$$

The optimal similar critical region defined by (12) depends upon the value of ρ under the alternative, and so no UMP test exists. Note that (12) yields a PO test for any fixed value of ρ under the alternative.

To obtain tests having power against a range of alternatives we must weaken the optimality criterion. The first option is to construct the LMP test, as given in the following theorem.

Theorem 2 *Suppose the linear regression model (4) has errors with autocorrelation structure given by (5), then the LMP test for $H_0 : \rho = 1$ against $H_1 : \rho < 1$, is: reject H_0 if*

$$\frac{\left(e_N' T_1^{-1} u\right)^2}{u'u} > k_\alpha, \qquad (13)$$

where $u = M_WT_1y$, e_N is the unit vector in $\mathbb{R}^N$ and k_α is chosen so that the size of the test is α. Analogously, the LMP test of $H_0 : \rho = 1$ against $H_1 : \rho > 1$ is: reject H_0 if

$$\frac{\left(e_N' T_1^{-1} u\right)^2}{u'u} < k_\alpha', \qquad (14)$$

where k_α' is chosen analogously to k_α.

Proof. Set $\rho = 1 - \gamma$, where $\gamma > 0$ for $H_1 : \rho < 1$, and $\gamma < 0$ for $H_1 : \rho > 1$, and differentiating (11) with respect to γ we obtain

$$sign(\gamma) \left.\frac{\partial P_\omega}{\partial \gamma}\right|_{\gamma=0} = d(N,k) + \frac{N-k}{2} \left[\frac{\partial \left(v' \left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1} v\right)}{\partial \gamma}\right]_{\gamma=0},$$

where $d(N,k)$ does not depend upon v. Moreover, since

$$\left[\frac{\partial\left(v'\left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1}v\right)}{\partial\gamma}\right]_{\gamma=0} = v'C'(T_1^{-1})'e_Ne_N'T_1^{-1}Cv$$

$$= \left(e_N'T_1^{-1}Cv\right)^2,$$

then noting the definition of v in (12) proves the first part of the theorem. The second part is proved analogously. ■

As mentioned in the previous section, maximising the slope of the power in a neighbourhood of the null is not the only criterion which may be used. We thus implement the BNM and BEPO test criteria as suggested in Section 2.2. Before proceeding, however, we will require the following lemma.

Lemma 3 *The matrix*

$$Q = C'T_1\Sigma_N(\rho)T_1'C - \left(C'\left(T_1\Sigma_N(\rho)T_1'\right)^{-1}C\right)^{-1},$$

is positive semi-definite, where the quantities C, T_1 *and* $\Sigma_N(\rho)$ *are defined above.*

Proof. Let $A = T_1\Sigma_N(\rho)T_1'$ and $H = (C : J)$, where J is any $N \times K$ matrix such that H is an $N \times N$ orthogonal matrix. Then from

$$H'A^{-1}H = (H'AH)^{-1},$$

and applying the inverse of a partitioned matrix (conformably with C and J), we have

$$C'A^{-1}C = \left(C'AC - C'AJ(J'AJ)^{-1}J'AC\right)^{-1}.$$

Thus

$$C'A^{-1}C \geq (C'AC)^{-1},$$

with equality if and only if C is orthogonal. Taking the inverse establishes the lemma. ■

As a consequence of Lemma 1 we may construct BNM and BEPO tests, whose derivations are contained in the following two theorems.

Theorem 4 *Let $\|.\|$ denote a norm on the space of $2{\times}2$ positive definite matrices, and let*

$$\Psi(u) = \frac{1}{u'u}\begin{pmatrix} u'\Sigma_N(1)u & u'(T_1^{-1})'L_N T_1^{-1}u \\ u'(T_1^{-1})'L_N T_1^{-1}u & u'(T_1^{-1})'L_N L_N' T_1^{-1}u \end{pmatrix}, \tag{15}$$

where $L_N = I_N - T_1'$ is the first-order lag operator matrix, with one's on the upper off-diagonal and zero's elsewhere, and $u = M_W T_1 y$. Then a BNM test is: reject $H_0 : \rho = 1$ if

$$\|\Psi(u)\| < k_\alpha, \tag{16}$$

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

Proof. We rewrite (12) as

$$\frac{y'T_1'C(C'AC)^{-1}C'T_1y}{y'T_1'M_WT_1y} < k_\alpha,$$

with $A = T_1\Sigma_N(\rho)T_1'$. Hence from Lemma 3,

$$\frac{y'T_1'C(C'AC)^{-1}C'T_1y}{y'T_1'M_WT_1y} \leq \frac{u'(T_1^{-1})'\Sigma_N^{-1}(\rho)T_1^{-1}u}{u'u},$$

where u is defined above, thus (12) is bounded above by the ratio of quadratic forms in u. Moreover, from

$$\Sigma_N^{-1}(\rho) = I_N - \rho(L_N + L_N)' + \rho^2 L_N L_N',$$

it follows that

$$\begin{aligned} \frac{u'(T_1^{-1})'\Sigma_N^{-1}(\rho)T_1^{-1}u}{u'u} &= \frac{1}{u'u}(1,-\rho)\begin{pmatrix} u'\Sigma_N(1)u & u'(T_1^{-1})'L_N T_1^{-1}u \\ u'(T_1^{-1})'L_N T_1^{-1}u & u'(T_1^{-1})'L_N L_N' T_1^{-1}u \end{pmatrix}\begin{pmatrix} 1 \\ -\rho \end{pmatrix} \\ &= (1,-\rho)\,\Psi(u)\begin{pmatrix} 1 \\ -\rho \end{pmatrix}. \end{aligned} \tag{17}$$

So a sufficient, but not necessary condition for (12) to hold is that the positive definite matrix $\Psi(u)$ is small with respect to some norm. Denoting all such norms by $\|\Psi(u)\|$, then for each $\|\Psi(u)\|$ we may find a k_α such that $\Pr\{\|\Psi(u)\| < k_\alpha | H_0\} = \alpha$. ■

Theorem 5 *BEPO tests for* $H_0 : \rho = 1$ *against* $H_1 : \rho \neq 1$, *are given by either of the following rules*

i) reject H_0 *if*

$$\left| \frac{u'(T_1^{-1})' L_N T_1^{-1} u}{u'(T_1^{-1})' L_N L_N' T_1^{-1} u} - 1 \right| < k_\alpha, \tag{18}$$

ii) reject H_0 *if*

$$\frac{(u'\Sigma_N(1)u)(u'(T_1^{-1})' L_N L_N' T_1^{-1} u) - (u'(T_1^{-1})' L_N T_1^{-1} u)^2}{(u'u)^2} < k_\alpha$$

where in both case k_α *is such that the size of the test is* α, *and* $u = M_W T_1 y$.

Proof. Using Lemma 1 as in the proof of Theorem 3, we obtain (17), which, as function of ρ, is a parabola. This has a minimum at

$$\rho^* = \left(u'(T_1^{-1})' L_N T_1^{-1} u \right) / \left(u'(T_1^{-1})' L_N L_N' T_1^{-1} u \right).$$

So using the BEPO criterion part i) of the theorem is proved. To prove ii), replace ρ^* in (17) and rearrange the terms. ■

Theorem 3 generates a class of BNM tests, depending upon the particular norm chosen. Since $\Psi(u)$ does not vary over the whole space of 2×2 positive definite matrices, but over a subspace defined by the inequalities

$$\begin{aligned} u'\Sigma_N(1)u &\geq u'(T_1^{-1})' L_N L_N' T_1^{-1} u \\ (u'\Sigma_N(1)u)(u'(T_1^{-1})' L_N L_N' T_1^{-1} u) &\geq \left(u'(T_1^{-1})' L_N T_1^{-1} u \right)^2, \end{aligned} \tag{19}$$

a norm yielding a simple BNM test statistic is

$$\|\Psi(u)\| = \frac{u'\Sigma_N(1)u}{u'u}. \tag{20}$$

Equally, fully exploiting the inequalities in (19), the BEPO test statistic in part ii) of Theorem 4, becomes

$$\frac{u'(\Sigma_N(1) - (T_1^{-1})' L_N T_1^{-1})u}{u'u}, \tag{21}$$

although application of (19) may entail throwing away some information contained in the sample.

4 Dynamic regression models

The dynamic regression model has the form,

$$
\begin{aligned}
y_i &= \rho y_{i-1} + z_i'\beta + \varepsilon_i \qquad (22)\\
\varepsilon_i &\sim N(0,\sigma^2),\ y_0 = 0,\ i = 1,..,N,
\end{aligned}
$$

where the z_i are $k \times 1$ vectors of exogenous variables and β is a $k \times 1$ vector of parameters. The assumption that $y_0 = 0$ simplifies the derivations later on, without constraining the behaviour of the time series $\{y_i\}$ in a significant way since we may initialise the series with any starting value via appropriate choice of z_1. Indeed we may condition upon previous values of $\{y_i\}_{i<1}$ and include those as regressors in z_1.

In order to simplify the notation, we define y and Z as before, and we wish to test the null hypothesis $H_0 : \rho = 1$ against $H_1 : \rho \neq 1$. The model in (22) implies testing

$$H_0 : y \sim N\left(T_1^{-1} Z\beta, \sigma^2 \Sigma_N(1)\right) \qquad (23)$$

against

$$H_1 : y \sim N\left(T_\rho^{-1} Z\beta, \sigma^2 \Sigma_N(\rho)\right), \qquad (24)$$

with T_ρ and $\Sigma_N(\rho)$ defined in (6) and (7). An intrinsic difficulty of constructing tests for (23), rather than (8), is that the mean of y is also a function of ρ. As a consequence, two different null hypotheses must be considered, one in which β is unknown, and is thus a nuisance parameter, and one in which β is known. We will consider these two cases separately. For either case, and as before, the transformation

$$x = T_1 y \sim N(T_1 T_\rho^{-1} Z\beta, \sigma^2 T_1 \Sigma_N(\rho) T_1'),$$

will simplify the notation.

4.1 No restrictions upon β under the null

When β is not restricted the null and alternative hypotheses take the form

$$
\begin{aligned}
H_0 &: x \sim N\left(Z\beta, \sigma^2 \Sigma_N(1)\right)\\
H_1 &: x \sim N\left(T_1 T_\rho^{-1} Z\beta, \sigma^2 \Sigma_N(\rho)\right),
\end{aligned}
$$

and so under H_0, we have a linear regression, analogous to (8). Again, under H_0 the model is full exponential and the statistics

$$\begin{aligned}\hat{\beta} &= (Z'Z)^{-1}Z'x \\ s^2 &= x'M_Z x,\end{aligned}$$

where $M_Z = I - Z(Z'Z)^{-1}Z'$, are jointly sufficient for the nuisance parameters β and σ^2, and the conditions of Theorem 1 are met. As before we transform $x \to \left(\hat{\beta}, w_1\right)$ with $w_1 = C'x$, with now $CC' = M_Z$. Again $\hat{\beta}$ and w_1 are independent under both the null and the alternative, and

$$\begin{aligned}\hat{\beta} &\sim N\left(\left(Z'Z\right)^{-1} Z'T_1T_\rho^{-1}Z\beta, \sigma^2\left(Z'Z\right)^{-1} Z'T_1\Sigma_N(\rho)T_1'Z\left(Z'Z\right)^{-1}\right) \\ w_1 &\sim N\left(C'T_1T_\rho^{-1}Z\beta, \sigma^2 C'T_1\Sigma_N(\rho)T_1'C\right).\end{aligned}$$

Following the previous section we transform to polar coordinates, $w_1 \to (v, s^2)$, where $v = w_1/(w_1'w_1)^{1/2}$, so that $v'v = 1$ (with Jacobian $\frac{1}{2}(s^2)^{\frac{N-k}{2}-1}$). The joint density of $\left(v, \hat{\beta}, s^2\right)$ is

$$\begin{aligned}f(v, \hat{\beta}, s^2; \rho, \sigma^2, \beta) &= f\left(\hat{\beta}; \rho, \sigma^2, \beta\right)\frac{1}{2}(2\pi\sigma^2)^{-\frac{N-k}{2}}\left(s^2\right)^{\frac{N-k}{2}-1} \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}\xi'\left(C'T_1\Sigma_N(\rho)T_1'C\right)\xi\right\} \\ &\quad \exp\left\{-\frac{s^2}{2\sigma^2}v'\left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1}v + \frac{\left(s^2\right)^{1/2}}{\sigma^2}v'\xi\right\}\end{aligned}$$

where $\xi = \left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1}C'T_1T_\rho^{-1}Z\beta$, and so, under H_0, v is uniformly distributed over the surface of the unit $(N-k)$-sphere. Therefore every similar test is characterised by the vector

$$v = \frac{C'T_1 y}{(y'T_1'M_Z T_1 y)^{1/2}}.$$

In this case, however, the unconditional power of every similar critical region ω

$$P_\omega = \int_\omega \int_{s^2>0} f(v, s^2; \rho, \beta, \sigma^2) ds^2 (dv)$$

depends not only upon the value of ρ under the alternative, but also upon the unknown β, and there exists no *UMP* test. Since β is unknown under the alternative, it is natural to maximise some weighted average of the power (Wald

(1943), Hillier (1987) and Andrews and Ploberger (1994)). To do this, we note that the $(N-k)\times 1$ vector ξ is zero under H_0, and that, under H_1, it spans a k-dimensional subspace of $\mathbb{R}^{N-k}$ as β varies in $\mathbb{R}^k$ and ρ is fixed. However, as ρ varies, this subspace also varies in $\mathbb{R}^{N-k}$. It is thus natural to choose a weighting function which takes into account of all these possibilities. A possible choice is

$$\mu(\xi)=(2\pi c\sigma^2)^{-(T-k)/2}\left|C'T_1\Sigma_N(\rho)T_1'C\right|^{\frac{1}{2}}\exp\left\{-\frac{1}{2c\sigma^2}\xi'C'T_1\Sigma_N(\rho)T_1'C\xi\right\}, \quad (25)$$

where $c>0$ is an arbitrary constant scaling the magnitude of the changes in ξ for which we want the test to be powerful. Note that the weighting function is defined over the whole of $\mathbb{R}^{N-k}$, and thus averages not only over ξ but also over all possible subspaces in which it may lie.

The weighting function (25) is chosen for its simplicity and flexibility. By choosing a Gaussian weight we retain the ability to derive tractable solutions. However, the particular choice allows flexibility in that it describes any intermediate case between no prior information on the location of β under the alternative (i.e. $c\to\infty$ implies almost uniform weight) and certainty (i.e. $c\to 0$ implies Dirac weight on a singleton).

The (unconditional) weighted power of any similar region ω is thus

$$\bar{P}_\omega=\int_\omega\int_{s^2>0}\int_{\xi\in\mathbb{R}^{T-k}}f(v,s^2;\rho,\sigma^2,\xi)\mu(\xi)\,d\xi ds^2(dv).$$

Evaluating the integrals over $s^2>0$ and $\xi\in\mathbb{R}^{T-k}$, $\bar{P}_\omega$ can be written as

$$\bar{P}_\omega=\frac{\Gamma\left(\frac{N-k}{2}\right)}{2\pi^{\frac{N-k}{2}}}(1+c)^{-\frac{N-k}{2}}\int_\omega\left(v'\left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1}v\right)^{-\frac{N-k}{2}}(dv). \quad (26)$$

Therefore, the weighted most powerful similar test for H_0 is given by small values of the statistic

$$v'\left(C'T_1\Sigma_N(\rho)T_1'C\right)^{-1}v, \quad (27)$$

and does not depend on c. However, since (27) depends upon the value of ρ under the alternative, no *UMP* exists, even in terms of the weighted power $\bar{P}_\omega$. Weighted point optimal tests are directly obtainable from (27) by substituting fixed values of ρ under the alternative. In order, though, to obtain tests independent of the alternative, we must weaken the optimality criteria further. Since (27) is directly analogous to (12), excepting the change of definition for C, the following theorems follow directly from Theorems 2, 3 and 4.

Theorem 6 *For the dynamic regression model (22) the weighted LMP test for* $H_0 : \rho = 1$ *against* $H_1 : \rho < 1$*, given the weighting function (25), is: reject* H_0 *if*

$$\frac{\left(e_N' T_1^{-1} u\right)^2}{u'u} > k_\alpha, \tag{28}$$

where now $u = M_Z T_1 y$ *and* k_α *is chosen so that the size of the test is* α*. Analogously, the weighted LMP test of* $H_0 : \rho = 1$ *against* $H_1 : \rho > 1$ *is obtained by reversing the inequality in (28).*

Theorem 7 *Let* $\|.\|$ *denote some measure on the space of* 2×2 *positive definite matrices, and let*

$$\Psi(u) = \frac{1}{u'u} \begin{pmatrix} u'\Sigma_N(1)u & u'(T_1^{-1})'L_N T_1^{-1}u \\ u'(T_1^{-1})'L_N T_1^{-1}u & u'(T_1^{-1})'L_N L_N' T_1^{-1}u \end{pmatrix}, \tag{29}$$

where $u = M_Z T_1 y$*. Then weighted BNM tests for* $H_0 : \rho = 1$ *against* $H_1 : \rho \neq 1$*, given the weighting function (25), are: reject* H_0 *if*

$$\|\Psi(u)\| < k_\alpha,$$

where the k_α *is chosen so that the size of the test is* α*.*

Theorem 8 *Weighted BEPO tests for* $H_0 : \rho = 1$ *against* $H_1 : \rho \neq 1$ *given the weighting function (25), are either:*
i) reject H_0 *if*

$$\left| \frac{u'(T_1^{-1})'L_N T_1^{-1}u}{u'(T_1^{-1})'L_N L_N' T_1^{-1}u} - 1 \right| < k_\alpha$$

ii) reject H_0 *if*

$$\frac{(u'\Sigma_N(1)u)(u'(T_1^{-1})'L_N L_N' T_1^{-1}u) - (u'(T_1^{-1})'L_N T_1^{-1}u)^2}{(u'u)^2} < k_\alpha$$

where in both case k_α *is such that the size of the test is* α*, and* $u = M_Z T_1 y$*.*

Note that weighted LMP, BNM and BEPO tests for the dynamic regression model differ from the corresponding tests for the linear regression with autocorrelation, only for the use of M_Z rather than M_W. Consequently, again fully exploiting the inequalities in (19), which hold for any vector u, we find that for the dynamic regression model (22), when we do not restrict β under H_0 that the BNM and BEPO tests coincide, giving

$$BNM = BEPO = \frac{u'\Sigma_N(1)u}{u'u}. \tag{30}$$

4.2 Restricting β under the null

Instead of letting β vary freely in (22), we may wish to restrict β under the null. For example, suppose we wish to test that y follows a simple random walk against alternatives such as trend stationary models (Dickey and Fuller (1981)). For this case we want to test

$$\begin{aligned} H_0 &: x \sim N\left(0, \sigma^2 I_N\right), \\ H_1 &: x \sim N\left(T_1 T_\rho^{-1} Z\beta, \sigma^2 T_1 \Sigma_N(\rho) T_1'\right), \ |\rho| < 1. \end{aligned} \tag{31}$$

We need to find a transformation $y \to (v, s^2)$, with v independent of s^2. Transforming to polar coordinates through $s^2 = x'x$ (which is sufficient for the scalar nuisance parameter σ^2) and $v = x(x'x)^{-1/2}$, $v'v = 1$ (with Jacobian $\frac{1}{2}\left(s^2\right)^{\frac{N}{2}-1}$), the joint density of s^2 and v is

$$\begin{aligned} f(v, s^2; \rho, \sigma^2, \beta) &= \frac{1}{2}(2\pi\sigma^2)^{-N/2}\left(s^2\right)^{\frac{N}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\beta' Z' Z\beta\right\} \\ &\quad \exp\left\{-\frac{s^2}{2\sigma^2} v'(T_1^{-1})'\Sigma_N^{-1}(\rho) T_1^{-1} v + \frac{\left(s^2\right)^{1/2}}{\sigma^2}\beta' Z' T_\rho T_1^{-1} v\right\}. \end{aligned}$$

Thus, under H_0, v is uniformly distributed, on the surface of the unit N-sphere, and is independent of s^2. Therefore, every similar critical region of size α for H_0 is characterised by the vector

$$v = \frac{T_1 y}{\sqrt{y'\Sigma_N^{-1}(1)y}}. \tag{32}$$

There is no *UMP* test and since β is unknown under the alternative, we average the power over all directions of β according to the weighting function

$$\mu_c(\beta) = (2\pi c\sigma^2)^{-k/2}|Z'Z|^{1/2}\exp\left\{-\frac{1}{2c\sigma^2}\beta'(Z'Z)\beta\right\}, \tag{33}$$

with $c > 0$ scaling the magnitude of changes in β for which we want the test to be powerful. Once again, the choice of the weighting function is chosen to simplify the derivations and to guarantee enough flexibility (see the discussion under equation (25)).

The weighted power of any similar critical similar region ω is

$$\begin{aligned} \bar{P}_\omega &= \int_\omega \int_{\beta\in\mathbb{R}^k}\int_{s^2>0} f(v, s^2; \rho, \sigma^2, \beta)\, \mu_c(\beta)\, ds^2\, d\beta\, (v'dv) \qquad (34) \\ &= \frac{\Gamma\left(\frac{N}{2}\right)}{2\pi^{N/2}}(1+c)^{-\frac{1}{2}}\int_\omega \left[v'(T_1^{-1})'T_\rho' M_Z^c T_\rho T_1^{-1} v\right]^{-\frac{N}{2}}(dv), \end{aligned}$$

where $M_Z^c = I_N + \frac{1+c}{c}Z(Z'Z)^{-1}Z'$. Therefore, the most powerful similar test for $H_0 : \rho = 1$, takes the form

$$v'(T_1^{-1})'T_\rho' M_Z^c T_\rho T_1^{-1} v < k_\alpha, \tag{35}$$

with k_α chosen as before. Of course no UMP exists, even in terms of the weighted power, and so we apply the LMP, BNM and BEPO criteria. The proof of the following theorems can be established from the proof of Theorems 2, 3 and 4.

Theorem 9 *For the dynamic regression model (22) the weighted LMP test for $H_0 : \rho = 1, \beta = 0$ against $H_1 : \rho < 1, \beta \neq 0$ which maximizes the slope of the weighted power (34) at $\rho = 1$, given the weighting function (33), is: reject H_0 if*

$$\frac{y'L_N M_Z^c L_N' y - y'L_N M_Z^c y}{(1+\frac{c}{1+c}y'T_1 Z(Z'Z)^{-1}Z'y)^{\frac{N}{2}+1}} > k_\alpha, \tag{36}$$

where k_α is chosen so that the size of the test is α. Analogously, the weighted LMP test of $H_0 : \rho = 1$ against $H_1 : \rho > 1$ is obtained by reversing the inequality in (36).

Theorem 10 *Let $\|.\|$ denote some measure on the space of 2×2 positive definite matrices, and let*

$$\Psi(y) = \frac{(1+c)^N}{y'\Sigma_N^{-1}(1)y} \begin{pmatrix} y'M_Z^c y & y'M_Z^c L_N' y \\ y'M_Z^c L_N' y & y'L_N M_Z^c L_N' y \end{pmatrix} \tag{37}$$

$u = M_Z T_1 y$. Then weighted BNM tests for $H_0 : \rho = 1$ against $H_1 : \rho \neq 1$, given the weighting function (33), are: reject H_0 if

$$\|\Psi(y)\| < k_\alpha,$$

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

Theorem 11 *Weighted BEPO tests for $H_0 : \rho = 1$ against $H_1 : \rho \neq 1$, given the weighting function (33), are either:*
i) reject H_0 if

$$\left| \frac{y'M_Z^c L_N' y}{y'L_N M_Z^c L_N' y} - 1 \right| < k_\alpha \tag{38}$$

ii) reject H_0 if

$$\frac{(y'M_Z^c y)\,(y'L_N M_Z^c L_N' y) - (y'M_Z^c L_N' y)^2}{(u'u)^2} < k_\alpha$$

where in both case k_α is such that the size of the test is α.

Noting the inequalities amongst the entries of $\Psi(y)$, corresponding to those given in (19),

$$\begin{aligned} y'M_Z^c y &\geq y'L_N M_Z^c L_N' y \\ (y'M_Z^c y)(y'L_N M_Z^c L_N' y) &\geq (y'L_N M_Z^c L_N' y)^2, \end{aligned}$$

then a candidate BNM test for this case is

$$BNM^c = (1+c)^N \frac{y'M_Z^c y}{y'\Sigma_N^{-1}(1)y} < k_\alpha. \tag{39}$$

4.3 Some remarks

(i) We have considered the unit root hypothesis in three different situations described in Sections 3, 4.1, and 4.2. The class of similar tests in each case, is characterised by a vector reflecting the appropriate transformation of the data in order to gain distributions free of the nuisance parameters under the null hypothesis. The one difference between the cases is the use of weighting functions for the dynamic regression cases.

(ii) When $c = 0$, and $\beta = 0$ under H_1, model (22) simplifies to

$$y_i = \rho y_{i-1} + \varepsilon_i,$$

and we test $H_0 : \rho = 1$ against $H_1 : |\rho| < 1$. This case is of its own intrinsic interest, see Dickey and Fuller (1979) or Evans and Savin (1981), and both the LMP, BNM and BEPO tests simplify considerably. The LMP test, after some manipulation, becomes: reject H_0 if

$$\frac{y_N^2}{\sum_{i=1}^N (y_i - y_{i-1})^2} < k_\alpha,$$

the NM (the bound is now an equality) test is: reject H_0 if

$$\frac{\sum_{i=1}^N y_i^2}{\sum_{i=1}^N (y_i - y_{i-1})^2} = \frac{1}{DW} < k_\alpha \tag{40}$$

where DW denotes the Durbin-Watson statistic and the EPO test is

$$\frac{\sum_{i=1}^N y_i y_{i-1}}{\sum_{i=1}^N y_{i-1}^2} = \hat{\rho} < k_\alpha, \tag{41}$$

where $\hat{\rho}$ is the maximum likelihood estimate for ρ and the critical values are chosen as before.

(iii) The other obvious limit is when $c \to \infty$, implying almost uniform weight on any value of β under the alternative. Hence, after normalisation, the LMP test statistic is

$$\frac{y' L_N M_Z L_N' y - y' L_N M_Z y}{(1 - y' T_1' Z (Z'Z)^{-1} Z T_1 y)^{\frac{N}{2}+1}},$$

the BNM test statistic is

$$\frac{y' M_Z y}{y' \Sigma^{-1}(1) y}$$

and the BEPO test statistic is

$$\frac{y'M_Z L_N y}{y'L_N' M_Z L_N y}$$

where $M_Z = I + Z(Z'Z)^{-1}Z$, and we reject for small values of the statistic compared with the corresponding critical values.

(iv) Consider, for example, for a finite $c > 0$, the BNM test statistic is

$$\frac{y'M_Z^c y}{y'\Sigma^{-1}(1)\,y},$$

which is simply a weighted sum of a distance in the y direction and a distance in the direction of the projection of y on the column space orthogonal to those of Z, where the weight is determined by how far from null in the β direction we want to have large power against. This clearly seems a desirable property for an optimal test in our set up.

5 Approximate distributions for the test statistics

The BNM and BEPO tests derived in the previous sections generally have the form of, or are a simple function of, a ratio of quadratic forms in normal variables, $Q = y'Ay/y'By$, where A and B are symmetric, B positive semi-definite and $y \sim N(\mu, \Sigma(\rho))$, and μ is a function of Z, β and in some cases ρ and the weighting function parameter c. The exact density of quadratic forms is known in only a few cases (see Hillier (1999)). However, in general the density of a quadratic form needs to be calculated either by numerical methods, for example Imhof's (1961) procedure (Sargan and Bhargava (1983)), or by appealing to asymptotic results (Phillips (1987)).

In this section we provide a saddlepoint approximation for the density and distribution of ratios of quadratic forms (see also Daniels (1954), Phillips (1978), Lieberman (1994) and Marsh (1998)). Applied to the statistics derived in this paper, it serves as a convenient analytic alternative to numerical methods or crude asymptotic results.

The procedure is simplified if we transform y so that the covariance matrix becomes the identity. This is achieved through transforming y to $T_\rho y$, although, to simplify notation, we will still write $Q = y'Ay/yBy$. The density and the cumulative density function of Q at a point q can be obtained using Gurland's (1948) inversion formulae

$$f(q;c) = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp\{P(\theta)\}\, Q(\theta) d\theta \tag{42}$$

$$F(q;c) = 1 - \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \frac{\exp\{P(\theta)\}}{\theta}\, d\theta, \tag{43}$$

where

$$P(\theta) = -\frac{1}{2}\left(Tr[G^{-1}\mu\mu'] - \ln|G|\right) \quad ; \quad Q(\theta) = Tr[G^{-1}B(G^{-1}\mu\mu' + I)],$$

and

$$G = I_N - 2\theta\,(A - qB)\,.$$

Note that $G = G(\theta)$ is a function of the parameter θ, although we will suppress this dependence in the notation.

Although (42) and (43) characterise the density and distribution, no closed form or convergent series representation has as yet been found for these integrals. The saddlepoint technique allows us to find an approximate solution to these integrals by expanding $\exp\{P(\theta)\}$ and $Q(\theta)$ around the saddlepoint, i.e. the arg. max. of $P(\theta)$. The resulting series is then transformed to a series of Gaussian integrals, evaluated term by term. The leading term of the expansion is the saddlepoint approximation. Although the technique is an asymptotic one, delivering an order of error of (at least) $O(N^{-1})$, the numerical accuracy in finite samples is sufficient for most purposes.

The following theorem presents the general forms of the approximations, which were obtained by Marsh (1998) by a direct application of the Daniels (1954) technique to ratios of quadratic forms.

Theorem 12 *The Saddlepoint approximations for inversions (42) and (43) at a point q are, respectively,*

$$\hat{f}(q) = \frac{\exp\left\{-\frac{\mu'\mu}{2}\right\} \text{etr}\left\{\hat{G}^{-1}\mu'\mu\right\} Tr\left[\hat{G}^{-1}B\left(\hat{G}^{-1}\mu'\mu + I\right)\right]}{\left|\hat{G}\right| \sqrt{4\pi \left|\text{tr}\left[\left(\hat{G}^{-1}(A - qB)\right)^2 \left(2\hat{G}^{-1}\mu'\mu + I\right)\right]\right|}} \quad (44)$$

$$\hat{F}(q) = \left(1 - \exp\left\{-\frac{\mu'\mu}{2}\right\}\right) + \exp\left\{-\frac{\mu'\mu}{2}\right\} \Phi\left(\hat{p} - \frac{1}{\hat{p}}\log\left[\frac{\hat{p}}{\hat{r}}\right]\right), \quad (45)$$

where

$$\hat{p} = sign(\hat{\theta})\sqrt{\left|\text{tr}[\hat{G}^{-1}\mu'\mu] - \ln\left|\hat{G}\right|\right|},$$

$$\hat{r} = 2\hat{\theta}\sqrt{\left|\text{tr}\left[(\hat{G}^{-1}(A - qB))^2(2\hat{G}^{-1}\mu\mu' + I)\right]\right|},$$

and $\hat{\theta}$ solves the saddlepoint defining equation

$$\text{tr}\left(G^{-1}(A - qB)\left(G^{-1}\mu\mu' + I\right)\right) = 0 .$$

Note that there exists a unique saddlepoint $\hat{\theta} \in \left(\frac{1}{2\lambda_N}, \frac{1}{2\lambda_1}\right)$, where the λ_1 and λ_N are the smallest and the largest eigenvalues of the matrix $A - qB$, and that the approximation is analytic and continuous in q (Daniels (1954)). Theorem 11 also allows us to approximate the distribution of *any* ratio of quadratic form in normal variables having the properties described earlier on. In some special cases Theorem 12 can be simplified considerably, as in the following corollaries.

Corollary 13 *For the dynamic regression model, under the null hypothesis H_0 : $\rho = 1$ and $\beta = 0$, the saddlepoint approximations for the density and distribution of the BEPO test (41) at a point q are*

$$\hat{f}(q) = \frac{\left(\prod_{k=1}^{N}\left[1 - 2\hat{\theta}\lambda_k\right]\right)^{-1/2} \sum_{k=1}^{N}\left[1 - 2\hat{\theta}\lambda_k\right]^{-1}}{\sqrt{4\pi\sum_{k=1}^{N}\left[\lambda_k^2\left(1 - 2\hat{\theta}\lambda_k\right)^2\right]}}, \quad (46)$$

$$\hat{F}(q) = 1 - \Phi\left(\hat{p} - \frac{1}{\hat{p}}\ln\left(\frac{\hat{p}}{\hat{r}}\right)\right), \quad (47)$$

$$
\begin{aligned}
\hat{p} &= sign(\hat{\theta})\sqrt{\left|\sum_{k=1}^{N} \ln\left(1 - 2\hat{\theta}\tau_k\right)\right|}\ , \\
\hat{r} &= 2\hat{\theta}\sqrt{\left|\sum_{k=1}^{N} \lambda_k^2 \left(1 - 2\hat{\theta}\lambda_k\right)^{-2}\right|},
\end{aligned}
$$

and the saddlepoint $\hat{\theta}$ *solves*

$$
\sum_{k=1}^{N} \left(\frac{\lambda_k}{(1 - 2\hat{\theta}\lambda_k)}\right) = 0 \ ; \ \hat{\theta} \in \left(\frac{1}{2\lambda_N}, \frac{1}{2\lambda_1}\right),
$$

where $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$ *are the ordered eigenvalues of* $A - qB$.

Corollary 14 *For the dynamic regression model and the regression with autocorrelated errors model, under the null hypothesis,* $H_0 : \rho = 1$ *and* $\beta = 0$*, the saddlepoint approximations for the density and distribution of the BNM test (40) at a point* q *are*

$$
\begin{aligned}
\hat{f}(q\,|H_0) &= \frac{2\left(\prod_{k=1}^{N}\left[1 - \frac{\hat{\theta}}{2}\left(\csc^2[\xi_k] - 4q\right)\right]\right)^{-1/2} \sum_{k=1}^{N}\left[1 - \frac{\hat{\theta}}{2}\left(\csc^2[\xi_k] - 4q\right)\right]^{-1}}{\sqrt{\pi \sum_{k=1}^{N}\left[\left(\csc^2[\xi_k] - 4q\right)^2 \left(1 - \frac{\hat{\theta}_1}{2}\left(\csc^2[\xi_k] - 4q\right)\right)^{-2}\right]}}, \\
\hat{F}(q|H_0) &= 1 - \Phi\left(\hat{p} - \frac{1}{\hat{p}}\ln\left(\frac{\hat{p}}{\hat{r}}\right)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{p} &= sign(\hat{\theta})\sqrt{\left|\sum_{k=1}^{N}\left(1 - \frac{\hat{\theta}}{2}\left(\csc^2[\xi_k] - 4q\right)\right)\right|} \\
\hat{r} &= \frac{\hat{\theta}}{2}\sqrt{\left|\sum_{k=1}^{N}\left(\csc^2[\xi_k] - 4q\right)^2 \left(1 - \frac{\hat{\theta}}{2}\left(\csc^2[\xi_k] - 4q\right)\right)^{-2}\right|} \\
\xi_k &= \frac{(2j-1)\pi}{2(2N+1)},
\end{aligned}
$$

and the saddlepoint $\hat{\theta}$ *solves*

$$
\sum_{k=1}^{N}\left(\frac{\csc^2[\xi_k] - 4q}{2 - \hat{\theta}\left(\csc^2[\xi_k] - 4q\right)}\right) = 0,
$$

with $\hat{\theta} \in \left(-\frac{2}{4q - \csc^2[\xi_N]}, \frac{2}{\csc^2[\xi_1] - 4q}\right)$.

The accuracy of these approximations will be analysed in the following section.

6 Numerical analysis

In this Section we assess numerically the power of the procedures suggested by comparing them to both competing tests and the finite sample power envelope. We also show that the saddlepoint approximations derived give an adequate characterisation of the finite sample distributions of the tests statistics proposed in Sections 3 and 4.

6.1 Evaluation of the power properties

This simulation will focus on tests for a unit root against a stationary alternative. In appendix A, Tables 1 through 4 contain a comparative power study, obtained through Monte Carlo simulation of the models with 50000 replications. Tables 1 and 2 concern the regression models with autocorrelated errors: for Table 1

$$\begin{aligned} y_i &= 0.5 + 5\left(1 - \frac{2i}{N}\right) + u_i \\ u_i &= \rho u_{i-1} + \varepsilon_i, \end{aligned} \tag{48}$$

and for Table 2

$$\begin{aligned} y_i &= \sin\left(\frac{2\pi i}{N}\right) + \frac{i}{N} + u_i \\ u_i &= \rho u_{i-1} + \varepsilon_i. \end{aligned} \tag{49}$$

In particular we compare the LMP, BNM and BEPO tests given respectively by (13), (20) and (18) and labelled LM, BN and BE, with the Sargan and Bhargava (1983) test, labelled SB and the power envelope, labelled PE, obtained from equation (12).

The BNM and the BEPO tests yield approximately the same power for all alternatives across the different specifications, and seem to yield a small but consistent advantage in terms of power over the Sargan and Bhargava test. Moreover the power of these tests lies within 1% of the power envelope everywhere. The LMP test performs poorly in comparison.

Tables 3 and 4 consider the dynamic regression models: for Table 3

$$y_i = 0.1\left(0.3 + 5\left(1 - \frac{2i}{N}\right)\right) + \rho y_{i-1} + \varepsilon_i, \tag{50}$$

and for Table 4

$$y_i = 0.1\left(\sin\left(\frac{2\pi i}{N}\right) + \frac{i}{N}\right) + \rho y_{t-1} + \varepsilon_i. \tag{51}$$

The value of $\beta = 0.1$ is chosen so that there is no 'jump' in power for the alternative 'closest' to the null hypothesis. Here we compare the LMP, BNM and BEPO tests (with two choices of c, 1 and .01), with the (studentised) Dickey-Fuller (1979) test, labelled DF. In this case values for the power envelope would be misleading, since in our framework the power envelope depends upon the value of c. However, the results suggest that, at least in these models, the choice of c has little impact upon power.

As in the previous case both the BNM and BEPO tests deliver a small, consistent power gain over the established test, although equally the BNM test seems to perform best of all. The tables also suggest that tests in the dynamic regression framework have more power, this we would of course expect, since the mean of the process also depends upon ρ in this case.

6.2 Evaluation of the saddlepoint approximation.

In order to assess the accuracy of the saddlepoint approximations derived in Section 5 for the models in (48), (49), (50) and (51) we use the following procedure:
1) the cumulative density function for the BNM and the BEPO tests is simulated for the model described in the previous Section using 50000 replications. Let $F(q)$ be such a simulated distribution.
2) The values of q_α for which $F(q_\alpha) = \alpha$ are found for some specific values of α.
3) Finally the cumulative density function at the value q_α is approximated by the saddlepoint approximation, $\hat{F}(q_\alpha)$ of Theorem 11. If the approximation is accurate we would then expect $\hat{F}(q_\alpha) \simeq \alpha$.

Tables 5, 6, 7, and 8 in Appendix B compare α and $\hat{F}(q_\alpha)$ for $\rho = 1$ and $\rho = .9$ for the models in (48), (49), (50) and (51). These tables (as well as other simulations not reported) show that the saddlepoint approximation is very close to the true distribution, for all values of α and ρ and for all tests considered.

If we accept that it forms an adequate approximation, the saddlepoint approximation can be used to plot the density of the tests statistic of interest under both the null and the alternative hypothesis. For example Figures 1 and 2 in Appendix C graph the densities of the BNM statistic (with $c = 1$), under the null hypothesis, $\rho = 1$ and $\beta = 0$ (corresponding to the solid line), and the alternatives, $\rho = 0.95$, 0.9, 0.85 and $\beta = 1$ (the length of the 'dash' increases as ρ decreases) for two sample sizes, $N = 15$, 30 for the model

$$y_i = 0.05\beta + \rho y_{i-1} + \varepsilon_i, \quad i = 1, 2, ..., N. \tag{52}$$

The vertical solid line marks the critical value in each case, specifically

$$\begin{aligned} cv &= 0.865 : N = 15 \\ cv &= 1.44 : N = 30. \end{aligned}$$

The power of the test is then the area under each graph to the left of the critical value. Some features are worth noting. First, comparing the two graphs it is apparent that the density is more sensitive to changes in the sample size under the null than under corresponding alternatives. This is, of course, precisely what we would expect, since convergence (actually in this case divergence, as we do not standardise) occurs at a faster rate under a unit root. Consequently, most of the power, as N increases, comes from changes in the null distribution, not the alternative. This asymmetry is a feature of the general problem of tests on the covariance structure of random samples, indeed if the alternative were explosive we would expect the reverse (i.e. faster divergence under the explosive alternative).

7 Conclusions

This paper has analysed tests for a unit root in both current formulations of the problem: dynamic linear regressions and linear regressions with autocorrelated errors. An optimality theory, valid for any sample size, has been developed, based on explicit use of classical statistical principles for both formulations. The main aim of the paper has been to show that analytic small sample inference for a problem that has, by and large, been analysed only asymptotically, is not only possible, but also applicable. Note that the analysis can be extended to testing any value of the autoregressive parameters and not just a unit root.

While we also construct locally most powerful tests, analogous to the feasible tests of Dufour and King (1991), it is the new criteria that we suggest may yield practical benefits. The essence of the criterion, that of replacing the search for the conditions for power to be maximised with sufficient ones, should prove applicable for other statistical and econometric problems. For the unit root problem, a Monte Carlo study has demonstrated that the procedure may lead to small but consistent power gains over other, well established tests. Although our numerical analysis is far from exhaustive, the tests which seem to perform consistently well are those based on the BNM criterion. Therefore, we would suggest the use of these tests, namely (20) for the autocorrelated errors and (30) for the dynamic regression model, although further detailed study of, for example, their robustness properties, would be needed.

A secondary aim of the paper has been to provide a distribution theory for the tests. Although approximate in nature, the saddlepoint technique, does seem to yield an adequate characterisation even in very small samples. In particular, enumeration of the approximation is exceptionally swift in comparison with other techniques. Moreover, since the functional form of the approximation is unchanged under both the null and alternative (unlike asymptotic treatments), the analysis of the problem is considerably simplified.

References

Anderson, T.W. (1948) On the theory of testing serial correlation. *Skandinavisk Aktuarietidskriet.*, 31, 88-116.

Andrews, D.W.K. and Ploberger, W. (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62, 1383-1414.

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

Daniels, H.E. (1954) Saddlepoint approximations in statistics. *Annals of Mathematics and Statistics*, 25, 631-650.

Dejong, D.N., Nankervis, J.C., Savin, N.E. and Whiteman, C.H. (1992) Integration versus trend stationarity in time series. Econometrica, 60, 423-433.

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

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

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. *Journal of Econometrics* 47, 115-143.

Dufour, J-M. and Kiviet, J.F. (1998) Exact inference methods for first-order autoregressive distributed lag models. *Econometrica*, 66, 79-104.

Durbin, J. and Watson, G.B. (1950) Testing for serial correlation in least squares regression I. *Biometrika*, 37, 409-428.

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

Evans, G.B.A. and Savin, N.E. (1981) Testing for unit roots: 1. *Econometrica*, 49, 753-779.

Evans, G.B.A. and Savin, N.E. (1984) Testing for unit roots: 2. *Econometrica*, 52, 1241-1269.

Faust, J. (1996) Near observational equivalence and theoretical size problems with unit root tests. *Econometric Theory*, 12, 724-732.

Forchini, G. (1998) Miscellaneous results for the Gaussian AR(1) model, in *Exact Distribution theory for some Econometric Problems.* Unpublished PhD thesis, University of Southampton.

Gurland, J. (1948) Inversion formulae for the distribution of ratios. *Annals of Mathematical Statistics*, (19): 228-237.

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

Hillier, G.H. (1999) The density of a quadratic form in a vector uniformly distributed on the n-sphere. To appear in *Econometric Theory.*

Imhof, J.P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419-426.

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

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

Lieberman, O. (1994), Saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables. *Journal of the American Statistical Association*, 89, 924-928.

Marsh, P.W.N. (1998) Saddlepoint approximations for non-central quadratic forms. *Econometric Theory*, 14, 539–559.

Phillips, P.C.B. (1978), Edgeworth and saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables. *Biometrika*, 65, 91-98.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica*, 55, 277-301.

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

Ploberger, W. (1999) A complete class of tests when the likelihood is locally asymptotically quadratic. Paper presented at the Cowles Foundation Econometrics Conference, "New Developments in Time Series Econometrics", Yale University, October 23-24, 1999.

Sargan, J.D. and Bhargava, A. (1983) Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica*, 51, 153-174.

van Garderen, K-J. (1997) Curved exponential models in statistics. *Econometric Theory*, 13,771-790.

Wald, A. (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society,* 54, 426-482.

Wolfram, S. (1999) *Mathematica: A System for Doing Mathematics by Computer*, 4^{th} ed. New York: Addison-Wesley.

Appendix A

This appendix contains four tables reporting the power of the tests, as detailed in section 6.1. The reported entries were obtained from a Monte Carlo study based on 50000 replications

	N=30					N=50				
ρ	LM	BN	BE	SB	PE	LM	BN	BE	SB	PE
.80	.056	.143	.141	.136	.145	.058	.321	.317	.292	.327
.82	.055	.125	.124	.119	.125	.058	.272	.269	.240	.277
.84	.054	.111	.111	.106	.111	.056	.225	.220	.201	.225
.86	.054	.098	.098	.094	.100	.055	.188	.186	.167	.188
.88	.053	.085	.083	.082	.086	.054	.150	.149	.134	.150
.90	.052	.075	.076	.074	.075	.052	.119	.118	.109	.119
.92	.051	.065	.063	.065	.066	.052	.097	.096	.091	.097
.94	.050	.059	.057	.060	.061	.051	.076	.076	.073	.076
.96	.050	.053	.054	.053	.053	.050	.061	.061	.061	.061
.98	.050	.050	.050	.050	.050	.050	.056	.056	.056	.056

Table 1. Power of the LMP (LM), the BNM (BN), the BEPO (BE), the Sargan-Bhargava (SB) tests and the power envelope (PE) for the model (48).

	N=30					N=50				
ρ	LM	BN	BE	SB	PE	LM	BN	BE	SB	PE
.80	.054	.122	.120	.120	.122	.057	.243	.237	.233	.247
.82	.054	.106	.109	.107	.108	.055	.207	.203	.202	.210
.84	.053	.095	.094	.094	.095	.055	.172	.170	.167	.172
.86	.052	.086	.084	.083	.087	.054	.147	.146	.144	.147
.88	.052	.078	.077	.077	.079	.053	.124	.124	.121	.124
.90	.051	.070	.070	.069	.070	.052	.103	.102	.102	.103
.92	.050	.064	.063	.064	.066	.051	.087	.086	.085	.087
.94	.050	.057	.056	.056	.057	.050	.073	.072	.073	.073
.96	.050	.053	.054	.054	.054	.050	.061	.061	.059	.061
.98	.050	.051	.051	.052	.052	.050	.053	.052	.052	.053

Table 2. Power of the LMP (LM), the BNM (BN), the BEPO, the Sargan-Bhargava tests and the power envelope (PE) for the model (49).

ρ	N=30 LM	N=30 BN^1	N=30 BE^1	N=30 $BN^{.01}$	N=30 $BE^{.01}$	N=30 DF	N=50 LM	N=50 BN^1	N=50 BE^1	N=50 $BN^{.01}$	N=50 $BE^{.01}$	N=50 DF
.80	.108	.371	.343	.367	.341	.331	.112	.681	.631	.687	.640	.599
.82	.108	.301	.281	.302	.279	.271	.112	.593	.524	.595	.539	.521
.84	.094	.246	.229	.238	.226	.225	.096	.492	.443	.495	.461	.430
.86	.091	.211	.201	.204	.201	.190	.084	.411	.371	.410	.377	.354
.88	.087	.166	.157	.166	.158	.152	.076	.302	.265	.305	.279	.261
.90	.085	.133	.125	.131	.125	.123	.070	.240	.215	.245	.220	.207
.92	.079	.101	.096	.100	.099	.096	.067	.171	.154	.174	.158	.148
.94	.071	.083	.078	.079	.079	.078	.066	.119	.106	.124	.109	.109
.96	.063	.065	.061	.065	.062	.064	.062	.074	.065	.078	.067	.071
.98	.055	.051	.050	.051	.053	.052	.056	.051	.050	.051	.050	.050

Table 3. Power of the LMP (LM), the BNM (BN^c) and the BEPO (BE^c) tests with $c = 1$, $c = 0.01$ and the Dickey-Fuller test (DF) for the model (50).

ρ	N=30 LM	N=30 BN^1	N=30 BE^1	N=30 $BN^{.01}$	N=30 $BE^{.01}$	N=30 DF	N=50 LM	N=50 BN^1	N=50 BE^1	N=50 $BN^{.01}$	N=50 $BE^{.01}$	N=50 DF
.80	.172	.394	.352	.398	.354	.356	.218	.728	.650	.751	.674	.693
.82	.168	.356	.318	.364	.322	.318	.211	.650	.587	.687	.597	.622
.84	.159	.292	.267	.299	.262	.268	.199	.571	.504	.603	.510	.544
.86	.147	.238	.213	.240	.210	.226	.184	.472	.415	.507	.421	.457
.88	.142	.204	.185	.205	.181	.183	.173	.392	.340	.412	.353	.364
.90	.122	.158	.153	.160	.145	.151	.154	.297	.264	.325	.264	.283
.92	.108	.135	.129	.136	.124	.127	.140	.231	.205	.245	.202	.221
.94	.095	.110	.102	.108	.101	.103	.119	.167	.151	.174	.145	.155
.96	.089	.089	.084	.085	.080	.081	.107	.113	.105	.126	.102	.119
.98	.065	.064	.064	.062	.062	.063	.076	.072	.071	.084	.069	.080

Table 4. Power of the LMP (LM), the BNM (BN^c) and the BEPO (BE^c) tests with $c = 1$, $c = 0.01$ and the Dickey-Fuller test (DF) for the model (51).

Appendix B

	$\hat{F}(q_\alpha)$	$\rho = 1$	$\hat{F}(q_\alpha)$	$\rho = .9$
α	N=20	N=40	N=20	N=40
0.05	0.044	0.045	0.042	0.047
0.10	0.091	0.089	0.086	0.091
0.20	0.184	0.187	0.181	0.188
0.30	0.278	0.283	0.277	0.285
0.40	0.374	0.384	0.376	0.386
0.50	0.477	0.485	0.478	0.490
0.60	0.579	0.588	0.580	0.591
0.70	0.681	0.691	0.683	0.695
0.80	0.786	0.798	0.788	0.799
0.90	0.890	0.904	0.892	0.904

Table 5: Saddlepoint approximations for the p-values of the the $BEPO$ test (21) statistic for model (48).

	$\hat{F}(q_\alpha)$	$\rho = 1$	$\hat{F}(q_\alpha)$	$\rho = .9$
α	N=20	N=40	N=20	N=40
0.05	0.041	0.046	0.042	0.048
0.10	0.099	0.103	0.103	0.098
0.20	0.214	0.207	0.215	0.206
0.30	0.321	0.312	0.319	0.310
0.40	0.425	0.414	0.422	0.412
0.50	0.523	0.515	0.518	0.511
0.60	0.619	0.613	0.619	0.610
0.70	0.717	0.710	0.715	0.709
0.80	0.813	0.806	0.814	0.807
0.90	0.909	0.905	0.907	0.903

Table 6: Saddlepoint approximations for the p-values of the the $BEPO$ test (21) statistic for model (49).

	$\hat{F}(q_\alpha)$	$\rho = 1$	$\hat{F}(q_\alpha)$	$\rho = .9$
α	N=20	N=40	N=20	N=40
0.05	0.053	0.051	0.043	0.051
0.10	0.105	0.103	0.089	0.102
0.20	0.211	0.204	0.184	0.204
0.30	0.321	0.306	0.279	0.307
0.40	0.417	0.409	0.378	0.405
0.50	0.518	0.511	0.482	0.505
0.60	0.620	0.613	0.585	0.605
0.70	0.712	0.709	0.686	0.702
0.80	0.810	0.807	0.788	0.801
0.90	0.908	0.904	0.895	0.901

Table 7: Saddlepoint approximations for the p-values of the the BNM test (30) statistic for model (50).

	$\hat{F}(q_\alpha)$	$\rho = 1$	$\hat{F}(q_\alpha)$	$\rho = .9$
α	N=20	N=40	N=20	N=40
0.05	0.042	0.046	0.040	0.045
0.10	0.093	0.092	0.085	0.091
0.20	0.189	0.189	0.184	0.189
0.30	0.285	0.286	0.283	0.286
0.40	0.382	0.387	0.387	0.384
0.50	0.480	0.492	0.486	0.486
0.60	0.585	0.595	0.588	0.590
0.70	0.692	0.698	0.690	0.691
0.80	0.798	0.800	0.797	0.796
0.90	0.909	0.903	0.908	0.900

Table 8: Saddlepoint approximations for the p-values of the the BNM test (30) statistic for model (51).

Appendix C

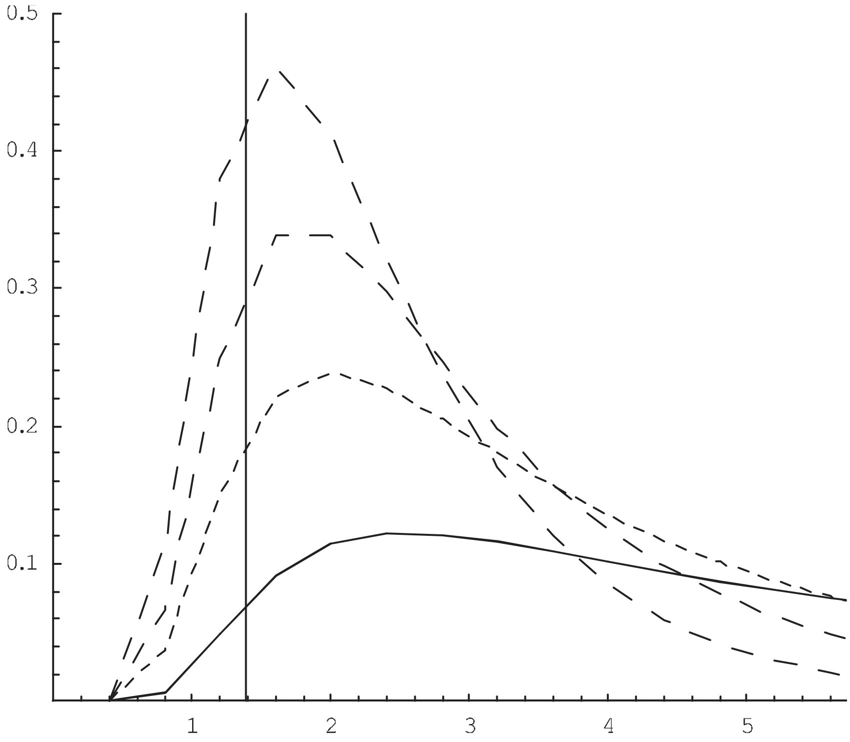

Figure 1: Densities for the BNM test statistic in the model (52) with $N = 15$, under the null; $\rho = 0$ and $\beta = 0$ (solid) and alternatives; $\rho = 0.95, 0.9, 0.85$ ($\beta = 1$) (dotted-dashed).

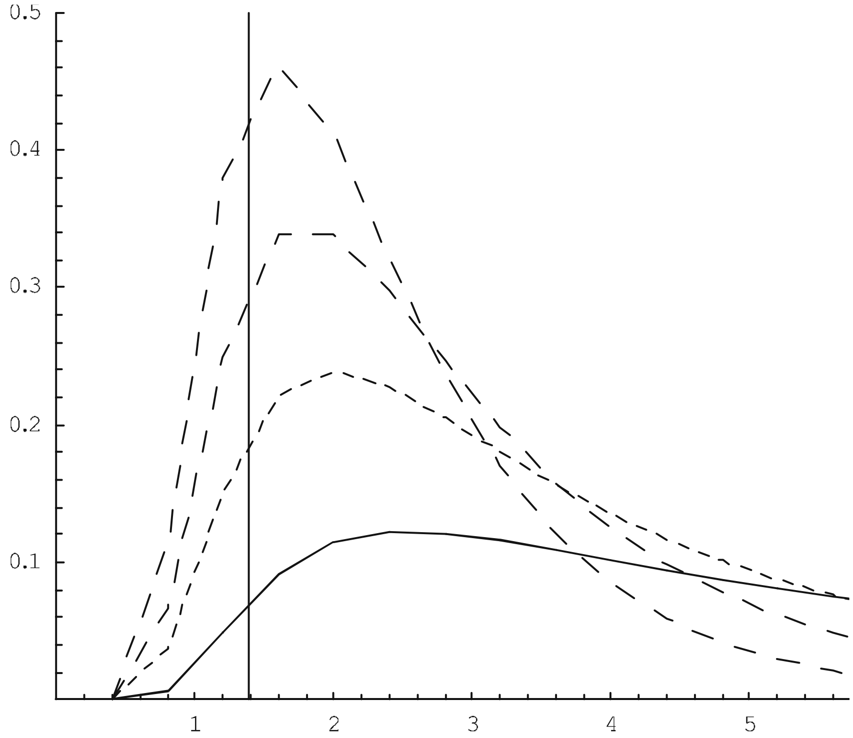

Figure 1: Figure 2: Densities for the BNM test statistic in the model (52) with $N = 30$, under the null; $\rho = 0$ and $\beta = 0$ (solid) and alternatives; $\rho = 0.95, 0.9, 0.85$ ($\beta = 1$) (dotted-dashed).