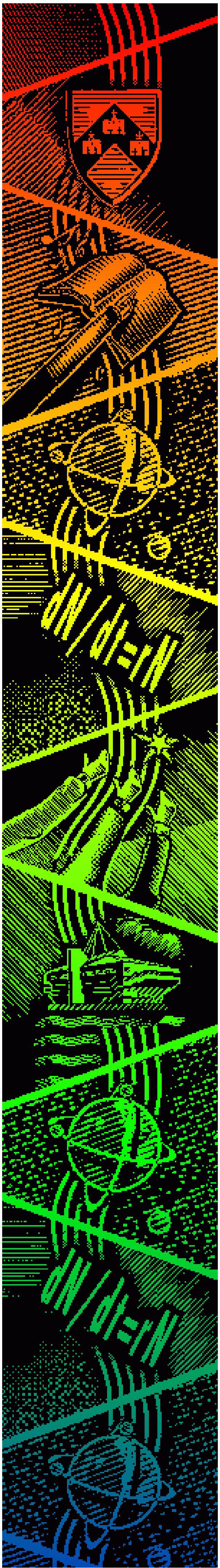

THE UNIVERSITY of York

Discussion Papers in Economics

No. 09/31

Saddlepoint Approximations for Optimal Unit Root Tests

By

Patrick Marsh, University of York

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

Saddlepoint Approximations for Optimal Unit Root Tests[*†]

Patrick Marsh[‡]

Department of Economics, University of York

October 26^{th} 2009

Abstract

This paper provides a (saddlepoint) tail probability approximation for the distribution of an optimal unit root test. Under restrictive assumptions, Gaussianity and known covariance structure, the order of error of the approximation is given. More generally, when innovations are a linear process in martingale differences, the estimated saddlepoint is proven to yield valid asymptotic inference. Numerical evidence demonstrates superiority over approximations for a directly comparable test based on simulation of its limiting stochastic representation. In addition, because the saddlepoint offers an explicit representation P-value sensitivity to model specification is easily analyzed, here in the context of the Nelson and Plosser data.

[*]**Submitted for the Granger Centre 2008 Conference Special Issue of Econometric Theory on Bootstrap and Numerical Methods in Time Series.**

[†]Thanks are due to Francesco Bravo, Giovanni Forchini, Peter Phillips, Robert Taylor, three anonymous referees and participants at seminars at the Universities of York and at the Conference on Bootstrap and Numerical Methods in Time Series at the University of Nottingham, Granger Centre for Time Series Econometrics.

[‡]Address: Department of Economics, University of York, YO105DD, tel. 01904433084, e-mail: pwnm1@york.ac.uk

1 Introduction

Critical values for unit root tests are most commonly found by simulation of numerical approximations to the limiting forms of those tests. Those limiting forms are themselves usually characterized by functionals of Brownian motion using the techniques pioneered in Phillips (1987a,b). Indeed most significant innovations in the literature, for example Phillips and Perron (1988), Elliott, Rothenberg and Stock (1996) and Ng and Perron (2001), yield tests whose properties are evaluated in this way. Some very noteworthy exceptions are Nabeya and Tanaka (1990), Abadir (1993a,b) and Juhl and Xiao (2003).

This paper uses a test based on the optimal procedures of Dufour and King (1991) and the marginal likelihood approach of Francke and de Vos (2007). Since the optimal test statistic takes the form of a ratio of quadratic forms its asymptotic distribution (which is constant under general assumptions on the innovation process) is approximated directly with the Lugannani and Rice (1980) tail probability (saddlepoint) approximation, and as previously employed in Lieberman (1994) and Marsh (1998). See also Phillips (1978), Lieberman (1996) and Larsson (1998) for applications closely related to the current context. Under general assumptions, the test is non-pivotal. Thus we provide an asymptotic order of error ($O\left(T^{-1}\right)$, for sample size T) for the tail probability and then prove that plugging in a consistent estimator for the covariance structure of the innovations yields asymptotically valid inference. Such estimated saddlepoint procedures have some history, for example see Butler and Paolella (2002).

From Francke and de Vos (2007), Dufour and King's (1991) point optimal test is the ratio (rather than the difference) of the two residual sum of squares detailed in Elliott, Rothenberg and Stock (1996, page 817) and forming their L_T and P_T^* tests. The aims of this paper are thus twofold. First, we establish that an explicit approximation, such as the saddlepoint, can yield asymptotically valid inference with the kind of generality expected in the literature. Second, we compare the finite sample properties of the procedures developed here with the method detailed above applied to the L_T and P_T^* test when the innovations follow a simple moving average process.

For the P_T^* tests (and also Dickey and Fuller (1979) type tests), the asymptotic nuisance parameter is the long run variance of the innovations process. In the context of a Dickey-Fuller regression it is consistently estimable via the autoregressive spectral density estimator, see Ng and Perron (2001). It has been conjectured, Seo (2006), that using more efficient estimators (such as estimating a moving average parameter via maximum likelihood) would improve the finite sample performance of such tests. The results of this paper directly undermine this idea. Using the true value (the perfect estimator) implies tests having zero size for small and moderate sample sizes while using efficient estimators gives tests which are more undersized at higher sample sizes than lower.

The saddlepoint based procedure does not suffer from these issues. In terms of finite sample size and power it outperforms the L_T or P_T^* tests whether the nuisance parameters are known, efficiently parametrically estimated or even non-parametrically estimated. Moreover, because the implied distributional approximation is explicit it can be exploited to give P-values for the test applied to the (extended) Nelson and Plosser (1982) data set. This allows a direct examination of the influence of the specification of the deterministic component on the outcome of the test.

The plan for the rest of the paper is as follows. The next section describes the optimal unit root test, Section 3 approximates its distribution via a saddlepoint approximation under the assumption the covariance structure is known. Section 4 proves that under a general innovation assumption the estimated saddlepoint will asymptotically valid inference. Section 5 presents the numerical analysis of the estimated saddlepoint test as well as applying it to the Nelson and Plosser data. Following the conclusion and references an appendix contains proofs of the main results.

2 The Model, Assumptions and Tests

2.1 Optimal unit root tests

We consider the following specification for the generation of time series $(y_t)_{t=1}^T$,

$$y_t = x_t'\beta + u_t \quad : \quad u_t = \rho u_{t-1} + \xi_t \,, \tag{1}$$

and tests of the hypotheses,

$$H_0 : \rho = 1 \quad vs. \quad H_1 : |\rho| < 1. \tag{2}$$

In (1) x_t is a $k \times 1$ deterministic regressor, β a $k \times 1$ unknown parameter and $(\xi_t)_{t=1}^T$ is a random innovation. Before giving the most general assumptions on both the distribution of ξ_t, and also the initial condition u_0, we first derive a point optimal test under more restrictive assumptions. The asymptotic properties of the resulting test will then be detailed under conditions very similar to those which have become standard in the unit root testing literature.

Suppose that the innovations are independent Gaussian, i.e. $\xi_t \sim iidN(0, \sigma^2)$, with finite variance σ^2. Define the following vectors and matrices; $y = (y_1, .., y_T)'$ and $\xi = (\xi_1, .., \xi_T)'$, and let $X = (x_1, ..., x_T)'$ and $\beta = (\beta_1, ..., \beta_k)'$. With $u_0 = 0$, (1) defines the following generalized Gaussian linear regression model,

$$y = X\beta + \Delta_\rho^{-1}\xi \sim N\left(X\beta, \sigma^2\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'\right), \tag{3}$$

where $\Delta_\rho = I - \rho L^{(1)}$ and $L^{(j)}$ is the lower triangular matrix with $1's$ on the j^{th} lower diagonal and $0's$ elsewhere.

The problem of testing H_0 is invariant under the group of transformations $G = (\sigma, \beta)$, with $\sigma \in \mathbb{R}$ and $\beta \in \mathbb{R}^k$ and with action, $y \to \sigma y + X\beta$ and so the optimal tests we seek must also be invariant under this group. Such tests are fully characterized through transformation of the data to the maximal invariant, see Dufour and King (1991), with properties, such as Information and Entropy, detailed in Marsh (2007, 2009). Francke and de Vos (2007) instead derive such tests via the alternative formulation of marginal likelihood. Following the latter shows that optimal tests may be constructed though a ratio of residual sums of squares from two simple regressions.

To proceed, quasi-difference the regression in (3),

$$y_{\bar{\rho}} = X_{\bar{\rho}}\beta + \xi_{\bar{\rho}}, \tag{4}$$

where $y_{\bar{\rho}} = \Delta_{\bar{\rho}}y$, $X_{\bar{\rho}} = \Delta_{\bar{\rho}}X$ and $\xi_{\bar{\rho}} = \Delta_{\bar{\rho}}\xi$. The residual sum of squares from the regression in (4) is,

$$RSS_{\bar{\rho}} = y_{\bar{\rho}}'M_{\bar{\rho}}y_{\bar{\rho}} \quad ; \quad M_{\bar{\rho}} = I - X_{\bar{\rho}}\left(X_{\bar{\rho}}'X_{\bar{\rho}}\right)^{-1}X_{\bar{\rho}}'. \tag{5}$$

Consider the singular value decomposition of $M_{\bar{\rho}}$,

$$C_{\bar{\rho}}C_{\bar{\rho}}' = M_{\bar{\rho}} \quad ; \quad C_{\bar{\rho}}'C_{\bar{\rho}} = I_n, \tag{6}$$

where $n = T - k$, so that $RSS_{\bar{\rho}} = w_{\bar{\rho}}'w_{\bar{\rho}}$, where $w_{\bar{\rho}} = C_{\bar{\rho}}'y_{\bar{\rho}}$ has density function (marginal likelihood),

$$f\left(w_{\bar{\rho}}, \rho, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-n/2}\sqrt{\frac{\det\left[X'X\right]}{\det\left[X'\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'X\right]}}\exp\left\{-\frac{y'\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'M_Xy}{2\sigma^2}\right\},$$

where $M_X = I - X\left(X'\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'X\right)X'$, see Francke and de Vos (2007, eqn 4), and notice that $\det\left[\Delta_\rho^{-1}\right] = 1$.

To remove dependence on σ^2, let $v_{\bar{\rho}} = w_{\bar{\rho}}/\sqrt{w_{\bar{\rho}}'w_{\bar{\rho}}}$, having density

$$f\left(v_{\bar{\rho}}, \rho\right) = \frac{\Gamma\left(\frac{n}{2}\right)|X'X|^{1/2}}{2\pi^{n/2}\det\left[X'\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'X\right]}\left(\frac{y'\Delta_\rho^{-1}\left(\Delta_\rho^{-1}\right)'M_X y}{y_{\bar{\rho}}'M_{\bar{\rho}}y_{\bar{\rho}}}\right)^{-n/2}.$$

The Neyman-Pearson Point Optimal invariant test of size α for (2) is to reject H_0 if,

$$NP = \frac{f\left(v_{\bar{\rho}}, \rho\right)}{f\left(v_{\bar{\rho}}, 1\right)} > \bar{k}_\alpha, \tag{7}$$

where $\bar{k}_\alpha$ is a constant chosen so that $\Pr\left[NP > \bar{k}_\alpha \mid H_0\right] = \alpha$. To focus on a specific test, here we follow Elliott, Rothenberg and Stock (1996) and choose as an 'optimal' test the Point Optimal test against the alternative, i.e. $\bar{\rho}$, where the (asymptotic) power envelope reaches 0.5,so

$$PO_T = \frac{y_{\bar{\rho}}'M_{\bar{\rho}}y_{\bar{\rho}}}{y_1'M_1 y} = \frac{RSS_{\bar{\rho}}}{RSS_1}. \tag{8}$$

The test is thus the ratio of the residual sums of squares of two regressions, one involving quasi-differenced data and one involving first-differenced data.

The optimal test in (8) is very similar to those given in Elliott, Rothenberg and Stock (1996) and before deriving fully feasible tests based on PO_T we will contrast its properties with those of their test,

$$L_T = y_{\bar{\rho}}'M_{\bar{\rho}}y_{\bar{\rho}} - \bar{\rho}y_1'M_1 y = RSS_{\bar{\rho}} - \bar{\rho}RSS_1. \tag{9}$$

As in that paper we will ultimately seek to detail the asymptotic properties of our test under much more general assumptions than those under which the test was derived. First assume;

Assumption 1 In the model defined by (1) assume that;

(i) The initial condition u_0 random and is such that $E[u_0^2] \leq \sigma_0^2 < \infty$.

(ii) The innovation process (ξ_t) is such that for all t, $E\left[\xi_t\right] = 0$,

$$\sum_{j=-\infty}^{\infty}\gamma_j > 0 \quad \text{and} \quad \sum_{j=-\infty}^{\infty}\left|\gamma_j\right| = m_\gamma < \infty, \tag{10}$$

where $\gamma_j = E\left[\xi_t \xi_{t-j}\right]$, and

$$T^{-1/2}\omega^{-1}\sum_{t=1}^{[rT]}\xi_t \Rightarrow B(r), \tag{11}$$

where $B(r)$ is standard Brownian motion on $[0,1]$ and $\omega^2 = \sum_{j=-\infty}^{\infty}\gamma_j$. $\blacksquare$

Asymptotic representations for the majority of unit root test statistics are constant under Assumption 1, at least when x_t is either a constant or a constant and a trend, and under either the null or local alternatives parameterized by $\rho = 1 - c/T$, $c > 0$. Such follow from the invariance principles detailed in, for example Phillips (1987a,b). Indeed, immediate from Elliott, Rothenberg and Stock (1996, Theorems 1 and 2) and Francke and de Vos (2007, Section 3.3) is that both L_T and PO_T converge almost surely to continuous random variables as the sample size becomes infinite. Before proceeding note that (10) implies that the long run variance is finite, but is not necessarily implied by it.

3 Saddlepoint Approximation for the Distribution of PO_T

Since the asymptotic distribution of PO_T is constant under Assumption 1 a representation for it may be found from any specific process satisfying it. Here we will derive a formal saddlepoint approximation under the following;

Condition 1A In addition to Assumption 1, suppose also that

(i) The initial condition is $u_0 = 0$.

(ii) The innovation sequence $(\xi_t)_1^T$ is a stationary Gaussian random variable, with

$$E[\xi_t] = 0 \quad ; \quad E\left[\xi_t\xi_{t-j}\right] = \gamma_j \quad ; \quad \gamma_0 = \sigma^2. \quad \blacksquare$$

Under Condition 1A, we have $\xi \sim N(0, \Gamma)$, where Γ is $T \times T$ Toeplitz and notice that the condition on the long run variance (10) ensures that Γ is bounded with respect to the matrix p-norms on the space of square matrices, with $\|\Gamma\|_1 = \|\Gamma\|_\infty = \sum_{j=0}^{T}\left|\gamma_j\right| \leq m_\gamma$, where $\|\Gamma\|_p$ is the L_p norm on $\mathbb{R}^T \times \mathbb{R}^T$.

Define $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$, where $\varepsilon_t \sim iidN(0,1)$, then we can write,

$$y = X\beta + \Delta_\rho^{-1} K\varepsilon,$$

where K is any matrix satisfying $KK' = \Gamma$. Consequently, and since by construction $M_{\bar{\rho}} X_{\bar{\rho}} = 0$,

$$PO_T = \frac{\varepsilon' K' \left(\Delta_\rho^{-1}\right)' \Delta_{\bar{\rho}}' M_{\bar{\rho}} \Delta_{\bar{\rho}} \Delta_\rho^{-1} K\varepsilon}{\varepsilon' K' M_{\bar{\rho}} K\varepsilon},$$

with $\varepsilon \sim N(0, I_T)$. Denote the distribution function of PO_T and its limit for $c = T(1-\rho)$, by

$$F_{\Gamma,\rho}^T(\kappa) = \Pr\left[PO_T < \kappa\right] \quad ; \quad \lim_{T\to\infty} F_{\Gamma,\rho}^T(\kappa) = F_{\Gamma,1-c}(\kappa),$$

then the purpose here is to provide an explicit asymptotic approximation for the distribution function, initially assuming that Γ is known.

Consider the rank $n = T - k$ matrix,

$$B(\rho) = K' \left(\Delta_\rho^{-1}\right)' \left[\Delta_{\bar{\rho}}' M_{\bar{\rho}} \Delta_{\bar{\rho}} - \kappa' \Delta_1' M_1 \Delta_1\right] \Delta_\rho^{-1} K, \tag{12}$$

and its n non-zero eigenvalues $(\lambda_t)_{t=1}^n$, and the functions,

$$\begin{aligned} R_n(\theta) &= -\frac{1}{2n}\sum_{t=1}^n \log\left(1 - 2\theta\lambda_t\right), \\ \eta_\rho(\theta) &= sign[\theta]\sqrt{-2nR_n(\theta)} \quad \text{and} \quad \delta_\rho(\theta) = \theta\sqrt{n\frac{d^2 R_n(\theta)}{d\theta^2}}, \end{aligned} \tag{13}$$

then the following theorem presents a saddlepoint approximation, with order of error $O(T^{-1})$, for the distribution of PO_T.

Theorem 1 *Under Condition 1A,*

$$F_{\Gamma,\rho}^T(\kappa) = \Phi(\tilde{\eta}_\rho) - \varphi(\tilde{\eta}_\rho)\left(\frac{1}{\tilde{\eta}_\rho} - \frac{1}{\tilde{\delta}_\rho}\right) + O(T^{-1}), \tag{14}$$

where $\Phi(.)$ and $\varphi(.)$ are the standard normal CDF and PDF,

$$\tilde{\eta}_\rho = \gamma_\rho(\tilde{\theta}_\rho) \quad \textit{and} \quad \tilde{\delta}_\rho = \delta_\rho(\tilde{\theta}_\rho),$$

and the saddlepoint $\tilde{\theta}_\rho$ is the unique solution to,

$$\sum_{t=1}^n \frac{\lambda_t}{1 - 2\tilde{\theta}_\rho\lambda_t} = 0 \quad ; \quad \frac{1}{2\lambda_n} \le \tilde{\theta}_\rho \le \frac{1}{2\lambda_1}. \quad \blacksquare$$

Theorem 1 presents a standard leading term saddlepoint approximation for the distribution of PO_T, which is, under Condition 1A, a ratio of quadratic forms in normal variables. The Theorem compliments previous results, in particular Lieberman (1994) and Marsh (1998), in that the order of error, $O(T^{-1})$, is established for the for the distribution function. This result is crucial here for two reasons. Later it will allow for consistent asymptotic inference in case where the correlations structure in Γ is not known, and must be estimated. First, the following corollary (which follows trivially from the transformation in Jing and Robinson (1994) and the invariance principle under Assumption 1) details how critical values obtained from a Gaussian approximation to the leading term in (14) are asymptotically correctly sized.

Corollary 1 *Under Assumption 1 and the null hypothesis,* $H_0 : \rho = 1$, *denote the leading term approximation by*

$$\tilde{F}_{\Gamma,1}(\kappa) = \Phi(\tilde{r}_1(\kappa)), \quad ; \quad \hat{r}_1(\kappa) = \tilde{\eta}_1 + \frac{1}{\tilde{\eta}_1}\ln\left(\frac{\tilde{\delta}_1}{\tilde{\eta}_1}\right), \tag{15}$$

where $\tilde{\eta}_1$ *and* $\tilde{\delta}_1$ *are defined above and the limiting distribution of* PO_T *is* $F_{\Gamma,1}(\kappa)$, *then for all* κ,

$$\lim_{T\to\infty} \tilde{F}_{\Gamma,1}(\kappa) = F_{\Gamma,1}(\kappa) + o(1). \quad \blacksquare$$

Before detailing how Corollary 1 generalizes to give fully feasible unit root tests, we first compare the properties of the critical values obtained for PO_T from (15), with asymptotic critical values for the L_T test of Elliott, Rothenberg and Stock (1996). Throughout this paper experiments will be performed using data generated according to the following:

Experimental Design Data $(y_t)_{t=1}^T$ is generated via,

$$\begin{aligned} M_1 &: \; y_t = \beta_1 + u_t \quad ; \quad u_t = (1 - c/T)u_{t-1} + \xi_t, \\ M_2 &: \; y_t = \beta_1 + \beta_2 t + u_t \quad ; \quad u_t = (1 - c/T)u_{t-1} + \xi_t, \end{aligned} \tag{16}$$

with $u_0 = 0$ and the innovation sequence $(\xi_t)_1^T$ is generated according to the MA(1) process,

$$\xi_t = \psi\varepsilon_t + \varepsilon_{t-1} \quad ; \quad \psi = \{-0.8, -0.5, 0, 0.5, 0.8\}, \tag{17}$$

with two possible error distributions for $(\varepsilon_t)_1^T$,

$$D_1 : \varepsilon_t \sim iidN(0,1) \quad ; \quad D_2 : \varepsilon_t \sim \frac{iid\chi^2(1) - 1}{\sqrt{2}}. \quad \blacksquare \tag{18}$$

Under H_0 L_T has the asymptotic representation,

$$(L_T - \bar{c}) \Rightarrow \omega^2 \bar{c}^2 \int_0^1 V_0^2 (r, \bar{c})\, dr + (1 + \bar{c})\, V_0^2(1, \bar{c}), \tag{19}$$

where $V_0 (r, \bar{c})$ is defined in Elliott, Rothenberg and Stock (1996, eqn 7). Critical values for L_T are then usually obtained via partial sum approximations to the stochastic integrals in (19). Based on 5000 steps in the partial sum and 20000 replications, Ng and Perron (2001) give,

$$\begin{aligned} M_1 &: cv_5 = 3.15 \quad \& \quad cv_{10} = 4.45, \\ M_2 &: cv_5 = 5.48 \quad \& \quad cv_{10} = 6.67. \end{aligned} \tag{20}$$

H_0 is to be rejected by L_T or PO_T, respectively, if

$$\omega^{-2} L_T < cv_\alpha \quad \text{or} \quad \tilde{F}_{\Gamma,1}(PO_T) < \alpha \tag{21}$$

In the first series of experiments, the accuracy of the asymptotic rejection rules in (21), is examined. Experiments involve 10000 Monte Carlo replications for both M_1 and M_2, over all MA(1) parameter values, with samples sizes, $T = 50, 100, 200$ and for both distributional assumptions D_1 and D_2. Although no author explicitly recommends use of such approximations in sample sizes this small it is worth noting that the range of sample sizes in the Nelson and Plosser (1982) data sets, as examined below, falls within that considered here. The results are presented in Tables 1 to 4 (in the Appendix). For both tests the *known* values of ψ were used to compute $\omega^2 = (1 + \psi)^2$ and $\Gamma = \left(I_T + \psi L^{(1)}\right)\left(I_T + \psi L^{(1)}\right)'$

To briefly detail the findings, Table 1 confirms the known high accuracy of the saddlepoint approximation for ratios of quadratic forms in normal variables, see Lieberman (1994, 1996) and Marsh (1998) for additional evidence. Table 3 confirms this accuracy has some robustness, in that for all sample sizes and over all ψ, this accuracy is retained even for non-normal, here skewed, variables. The accuracy is seen to improve with the sample size. In Tables 2 and 4, the standard asymptotic approximation for L_T fairs less well, particularly when $\psi < 0$. For large sample sizes and

$\psi \geq 0$ the results are comparable with the saddlepoint approximation, but to a much lesser extent for D_2, the shifted Chi-squared. In summary, the saddlepoint seems to offer a more uniformly accurate method of approximating the critical values of these very similar optimal unit root tests.

Analysis of the dramatically poor performance of L_T for $\psi < 0$ is necessary in understanding both why the standard asymptotic representations for unit root tests may not always provide accurate inference. Condition 1 requires that $\omega^2 > 0$, which here implies that $\psi > -1$. Asymptotic implications of the failure of this requirement have been explored in the literature, see Pantula (1991) and Nabeya and Perron (1994). Suppose the innovations were generated from,

$$\xi_t = \left(-1 + \frac{d}{T^\epsilon}\right)\varepsilon_t + \varepsilon_{t-1}, \quad \epsilon > 0 \tag{22}$$

so that

$$\omega^2 = \omega_T^2 = \sigma^2 \frac{d}{T^\epsilon},$$

and hence rather than part (ii) of Condition 1 holding, we instead have,

$$\left(\frac{T^\epsilon}{T}\right)^{1/2} \sum_1^{[rT]} \xi_t \Rightarrow \sqrt{d}\sigma B(r). \tag{23}$$

Limit theory, for $\epsilon = 1/2$, for autoregressive estimators and Dickey-Fuller tests under such processes are detailed in Pantula (1991) under the unit root null and in Nabeya and Perron (1994) under local to unity alternatives, see also Theorem 1 of Seo (2006).

Deriving asymptotic distributions for test statistics based on the limit in (23) will not prove useful for two reasons. First such asymptotic distributions will not be pivotal with respect to the local nuisance parameter d which, in turn, is not consistently estimable. Second, any approximate critical values derived from such distributions will not be accurate for processes not having a moving average root near -1. In any case, the results of Seo (2006) do not suggest a dramatic increase in accuracy for values which are. In addition, it has been an implicit assumption that the properties of unit root tests will improve if, in the case where ω^2 is unknown, more accurate estimators of the long run variance are used. This assumption is directly contradicted by these results. Knowing ω^2 is equivalent to having a perfect estimator and does not yield accurate critical values.

4 Estimated Saddlepoint Tests

Corollary 1 essentially defines a saddlepoint point optimal (SPO) test, as in;

Definition 1 *The SPO test at size α, consists of rejecting $H_0 : \rho = 1$, if*

$$\tilde{F}_{\Gamma,1}(PO_T) < \alpha. \quad \blacksquare \tag{24}$$

In practice this is not feasible. To construct a feasible test let $\hat{\Gamma}$ be an estimator of the covariance structure of ξ and define the matrix,

$$\hat{B}_1 = \hat{K}' \left(\Delta_1^{-1}\right)' \left[\Delta_{\bar{\rho}}' M_{\bar{\rho}} \Delta_{\bar{\rho}} - \kappa' \Delta_1' M_1 \Delta_1\right] \Delta_1^{-1} \hat{K},$$

where $\hat{\Gamma} = \hat{K}\hat{K}'$, and let the non-zero eigenvalues of $\hat{B}_1$ be $\left\{\hat{\lambda}_t\right\}_{t=1}^{n}$. Now define

$$\hat{F}_{\hat{\Gamma},1}(\kappa) = \Phi\left(\hat{r}_1(\kappa)\right) \quad ; \quad \hat{r}_1(\kappa) = \hat{\eta}_1 + \frac{1}{\hat{\eta}_1} \ln\left(\frac{\hat{\delta}_1}{\hat{\eta}_1}\right), \tag{25}$$

where

$$\hat{\eta}_1 = sign\left(\hat{\omega}_1\right) \sqrt{\sum_{t=1}^{n} \log\left(1 - 2\omega_1 \hat{\lambda}_t\right)} \quad ; \quad \hat{\delta}_1 = \hat{\omega}_1 \sqrt{2 \sum_{t=1}^{n} \left(\frac{\hat{\lambda}_t}{1 - 2\hat{\omega}_1 \hat{\lambda}_t}\right)^2}, \tag{26}$$

and the saddlepoint solves

$$\sum_{t=1}^{n} \left(\frac{\hat{\lambda}_t}{1 - 2\hat{\omega}_1 \hat{\lambda}_t}\right) = 0, \tag{27}$$

then the estimated saddlepoint point optimal test ($ESPO$) is defined by;

Definition 2 *The $ESPO$ test at size α, consists of rejecting $H_0 : \rho = 1$, if*

$$\hat{F}_{\hat{\Gamma},1}(PO_T) < \alpha. \quad \blacksquare \tag{28}$$

In essence the $ESPO$ test involves nothing more than substituting an estimator for the nuisance parameters in Γ. The most general assumptions under which Γ can be consistently estimated are detailed in Assumption A of Bühlmann (1995) and Assumptions 1 and 2 of Chang and Park (2002), i.e.

Assumption 2 In addition to Assumption 1, assume also that

(i) $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence with filtration $\mathcal{F}_t$, and such that a) $\mathrm{E}\left[\varepsilon_t^2\right] = \sigma^2$, b) $\operatorname{plim} T^{-1} \sum_1^T \varepsilon_t^2 = \sigma^2$, c) $\mathrm{E}\left[|\varepsilon_t|^s\right] \leq m_\varepsilon < \infty$ for $s \geq 4$,
(ii) $\xi_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} |j|^r \left|\psi_j\right| \leq m_\psi < \infty$, $\psi_0 = 1$, for $r \geq 1$, and
(iii) $u_0 = 0$. $\blacksquare$

In the context of an Augmented Dickey-Fuller (ADF) regression Perron and Ng (2001) exploit a modified model selection criteria to deliver ADF (and other) tests which are asymptotically correctly sized. Such, along with the work of Chang and Park (2002), are based on an autoregressive approximation to the inverse of the transfer function of the (ξ_t),

$$\Xi(z) = \sum_{j=0}^{\infty} \alpha_j z^j = \frac{1}{\Psi(z)} = \left(\sum_{j=0}^{\infty} \psi_j z^j\right)^{-1}.$$

Instead, and following Bühlmann (1995 & 1998), a consistent estimator for $\Xi(z)$ is,

$$\lim_{p_T \to \infty} \hat{\Xi}(z) = \sum_{j=1}^{p_T} \hat{\alpha}_{j,T} z^j, \quad \text{for } |z| \leq 1,$$

where $p_T = o\left(T^{1/2}\right)$ where $\hat{\alpha}_{j,T}$ are Yule-Walker autoregressive estimators obtained from the autocovariance function of the residuals from the Dickey-Fuller regression

$$\Delta\tilde{y}_t = \hat{\alpha}\tilde{y}_{t-1} + \hat{\upsilon}_t. \tag{29}$$

That is, defining $\hat{\alpha}_p = \{\hat{\alpha}_{j,T}\}_{j=1}^{p_T}$, $\hat{\gamma}_p = \left\{\hat{\gamma}_{j,T}\right\}_{j=1}^{p_T}$ where $\hat{\gamma}_{j,T} = (T - |j|)^{-1} \sum_{1+|j|}^{T} \hat{\upsilon}_t \hat{\upsilon}_{t-j}$ and the matrix $\hat{\Gamma}_p = \left\{\hat{\gamma}_{|i-j|,T}\right\}_{i,j=1}^{p_T}$, then $\hat{\alpha}_p$ satisfies

$$\hat{\Gamma}_p \hat{\alpha}_p = -\hat{\gamma}_p. \tag{30}$$

Here we employ this Yule-Walker approach, reparameterized in terms of the linear process coefficients, to prove asymptotic validity of the $ESPO$ procedure, as follows.

Theorem 2 *Under Assumption 2, let $h_T \to \infty$ and $h_T = o(p_T)$ and let $\left(\hat{\psi}_{j,T}\right)_{j=1}^{h_T}$ be the first h_T coefficients in the expansion $\left(\hat{\Xi}(z)\right)^{-1} = \sum_{j=1}^{\infty} \hat{\psi}_{j,T} z^j$, and write*

$$\hat{\Gamma}_h = \left(I_T + \sum_{j=1}^{h_T} \hat{\psi}_{j,T} L^j\right)\left(I_T + \sum_{j=1}^{h_T} \hat{\psi}_{j,T} L^j\right)',$$

then for all κ,

$$\lim_{T \to \infty} \left|\tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - F_{\Gamma,1}(\kappa)\right| \to 0, \quad \textit{almost surely.} \quad \blacksquare$$

Theorem 2 generalizes Corollary 1 such that inference based on the saddlepoint approximation remains asymptotically valid when the unknown covariance structure is estimated. For parametric assumptions, such as an explicit ARMA model, an obvious simplification of the Theorem would prove validity for inference based on a more efficient estimator, such as the maximum likelihood. In the next section we will explore the accuracy of the $ESPO$ procedure, again based upon the framework described in Section 3, and also use the approximation to give approximate P-values for optimal tests applied in the context of the (extended) Nelson and Plosser (1981) data.

5 Illustration and application of the $ESPO$ procedure

5.1 Finite sample performance

Here we assess the finite sample properties, both size and power, of the $ESPO$ test defined in (28) and compare with those of the P_T^* test of Elliot, Rothenberg and Stock (1996). The latter involves rejecting H_0 if

$$P_T^* = \hat{\omega}^{-2} L_T < cv_\alpha,$$

where $\hat{\omega}$ is a consistent estimator for ω and the critical values cv_α are given in (20). The experiments are repeated for both models in (16) and the MA parameter values in (17), but only the Gaussian distribution, with two sample sizes $T = 100$, 200 and results are presented on the basis of 10000 Monte Carlo replications.

We consider three different estimators for the nuisance parameters Γ (for the $ESPO$ test) and ω^2 (for the P_T^* test). The first two are based on the parametric Gaussian $ARMA(p,q)$ likelihood, defined (up to constants not depending on parameters) by,

$$l(\beta, \rho, a, b) = -\sum_1^T \varepsilon_t (\beta, \rho, a, b)^2, \tag{31}$$

where $a = (a_1, .., a_{p-1})$ and $b = (b_1, .., b_q)$, and

$$\varepsilon_t (\beta, \rho, a, b) = (1 - \rho L) \left(1 - \sum_{j=1}^{p-1} a_j l^{(j)} \right)^{-1} \left(1 + \sum_{j=1}^{q} b_j l^{(j)} \right) \left(y_t - \mu_{k,t}(\beta) \right), \quad k = 1, 2, \tag{32}$$

where $l^{(j)} y_t = y_{t-j}$ and for model M_1, $\mu_{1,t}(\beta) = \beta_1$ while for model M_2, $\mu_{2,t}(\beta) = \beta_1 + \beta_2 t$. In (32) it is assumed $\varepsilon_t = 0$ for all $t \leq 0$.

First take the parametric case of b_1 being unknown but it is known that $p = 1$ and $q = 1$. This yields the estimators

$$\text{I}): \hat{\Gamma}_I = \hat{K}_I \hat{K}_I', \quad K_I = \left(I + \hat{b}_1 L^{(1)} \right) \quad ; \quad \hat{\omega}_I^2 = \left(1 + \hat{b}_1 \right)^2,$$

required for the $ESPO$ and P_T^* tests respectively. Alternatively, suppose that p and q are unknown, and estimated, for model M_k, via the BIC,

$$\hat{p}, \hat{q} = \arg\min_{p,q} \left(\ln \left[T^{-1} \sum_1^T \varepsilon_t \left(\hat{\beta}, \hat{\rho}, \hat{a}, \hat{b} \right) \right]^2 + \frac{\ln(T)}{T} (p + q + k) \right), \tag{33}$$

where $1 \leq p \leq 2$ and $q \leq 2$, this yields the estimators,

$$\text{II}) \quad : \quad \begin{array}{l} \hat{\Gamma}_{II} = \hat{K}_{II} \hat{K}_{II}', \quad \hat{K}_{II} = \left(I_T - \sum_{j=1}^{\hat{p}-1} \hat{a}_j L^{(j)} \right)^{-1} \left(I_T + \sum_1^{\hat{q}} \hat{b}_j L^{(j)} \right), \\ \hat{\omega}_{II}^2 = \dfrac{\left(1 + \sum_{j=1}^{\hat{q}} \hat{b}_j \right)^2}{\left(1 - \sum_{j=1}^{\hat{p}-1} \hat{a}_j \right)^2}. \end{array}$$

Lastly we take the residuals from the Dickey-Fuller regression in (29) and run the auxiliary regression, also using the BIC,

$$\hat{\upsilon}_t = \sum_1^{\hat{p}^*} \alpha_j \hat{\upsilon}_{t-j} + \epsilon_t, \quad \hat{p}^* = \underset{\alpha_1, .., \alpha_{p^*}, p_L \leq p^* \leq p^U}{\arg\min} \ln \left[T^{-1} \sum_1^T \hat{\epsilon}_t^2 \right] + \frac{\ln T}{T} (p^* + k),$$

for each M_k, $k = 1, 2$,where the $\hat{\epsilon}_t$ are the auxiliary residuals and we chose $p_L = 3$ and $p^U = 6$, as in Elliott, Rothenberg and Stock (1996). Using first $\hat{p}^*$ residual autocovariances we construct

$$\text{III}): \hat{\Gamma}_{III} = \hat{\Gamma}_{\hat{p}^*} \quad ; \quad \hat{\omega}_{III}^2 = 1 + 2 \sum_{j=1}^{\hat{p}^*} \hat{\gamma}_{j,T},$$

and note that since an invertible $MA(\hat{p}^*)$ is uniquely defined by its covariances we may simply take $\hat{\Gamma}_{\hat{p}^*}$ to be the residual autocovariance matrix.

The results for the $ESPO$ test are given in Table 5, while those for the P_T^* test are in Table 6. For the $ESPO$ test the results using the maximum likelihood estimator $\hat{\Gamma}_I$ are very similar to those obtained when the parameter is assumed known, it delivers stable and accurate inference across the range of experiments. When an ARMA model, with lag length selected by (33), is used to estimate the covariance structure the results are slightly less accurate, although still within around 1.5% of nominal at the 5% level and within 2% at the 10% level. Simply plugging in the sample autocovariance matrix is less accurate, particularly for the case of $\psi = -0.8$, although unlike with the parametric estimators the is clearly improved for the larger sample size.

The results for the P_T^* test highlight the difficulties with using pure asymptotic results, particularly when ψ is large and negative. In such cases, when the efficient estimator is used $\hat{\omega}_I$ the test is severely undersized and is worse for the larger sample size. The slightly less efficient estimator gives a test which is slightly more accurate, but again worsens as the sample size increases. Using the sample autocovariance estimator (not employing the modified criteria of Ng and Perron (2001)) gives oversized tests although for the larger sample size they are closer to nominal. On the other hand when ψ is not large and negative the size of the P_T^* test is quite close to nominal, regardless of how the long run variance is estimated.

In direct comparisons the $ESPO$ procedure is generally more accurate over the whole range of experiments. Using parametric estimators gives results which are very similar to those obtained in the known parameter case. Using a nonparametric estimator does imply less accuracy, although still better than the P_T^* procedure and indeed those presented in Perron and Qu (2007) using a refined selection criteria. As an alternative to pure asymptotic results recent developments in the literature have exploited sieve bootstrap methods for processes satisfying Condition 1B, above. For model M_1 and $T = 100$, Cavaliere and Taylor (2008) report a size of 3.8% for an AR-based sieve applied to a GLS detrended Dickey-Fuller test, while Richard (2007) reports much less favorable results for an MA-based sieve. The latter results are surprising given the experimental design is also based on an $MA(1)$.

The $ESPO$ test, implemented with the maximum likelihood estimator, has very high accuracy when the model is correctly or over-specified. However, it is also important to demonstrate some robustness to mis-specification. To this end the results in Table 7 give sizes when the $ESPO$ test is applied by fitting an $MA(1)$, to model M_2, when in fact the innovations are generated by either an $ARMA(1,1)$ or an $MA(2)$, i.e. $\xi_t = \psi(l)\varepsilon_t$, where either,

$$E_4 : \psi(l) = (1 - \vartheta l)^{-1} (1 - 0.8l) \quad \text{or} \quad E_5 : \psi(l) = (1 + \vartheta l)(1 - 0.8L)\,. \tag{34}$$

In each case values of $\vartheta = \{\pm 0.15, \pm 0.1 \pm 0.05\}$ were used and the null rejection frequencies recorded. In both cases the effect of small, negative values of ϑ is negligible. With small positive values the effect is more significant, the size never exceeds that of the best asymptotic tests, such as those in Perron and Qu (2007). In practice such mis-specification can either be tested for, or other versions of the test used instead.

In terms of power both PO_T and L_T are constructed using power optimality criteria. Under Condition 1A both are point optimal invariant, against the alternative $H_1 : \rho = \bar{\rho}$, with only the invariance groups differing; PO_T is also scale invariant. Numerical evidence in the literature, e.g. in Perron and Qu (2007) and Seo (2006), suggests that P_T^* is, in practice, slightly more powerful than Dickey-Fuller type tests.

Here we compare the size-corrected power of the feasible versions $ESPO$ and P_T^* tests in both M_1 and M_2 in (16) and for all MA parameters in (17), of both tests for

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_1 : \rho = 0.95, 0.90, ..., 0.70,$$

with $T = 100$ and for 10000 replications. The nuisance parameters were estimated using the residual auto-covariance estimators, i.e. $\hat{\Gamma}_{III}$ and $\hat{\omega}_{III}$, although, because we are size-correcting, powers obtained from using either of the other two estimators are almost identical. The results are presented in Tables 8 (for $ESPO$) and 9 (P_T^*). As expected the powers of these two procedures are very similar. The $ESPO$ test has a slender advantage when the MA parameter is large and negative, particularly in the constant model, M_1.

5.2 Application: P-values for the Nelson and Plosser data

Here we exploit the explicit characterization of probabilities given by the saddlepoint method to report the P-values obtained from the outcome of the $ESPO$ test, when applied to real data.

Each Nelson and Plosser (1982) series was estimated for three different models, M_1 and M_2 as above, and also;

$$M_3 : y_t = \beta_1 + \beta_2 t^{\nu} + u_t \quad ; \quad u_t = (1 - c/T)u_{t-1} + \xi_t,$$

where the value of ν, for each series, was taken from the nonlinear regressions presented in Marsh (2009). All models were estimated via maximum likelihood, as in (32) (with $\mu_{3,t} = \beta_1 + \beta_2 t^{\nu}$), using the BIC in (33) to determine the orders p and q. Here, however, the focus is solely to examine the sensitivity of the $ESPO$ test to the specification of the mean function.

For models M_1 and M_2 the P-value is found from just $\tilde{F}_{\hat{\Gamma}_{II},1}(PO_T)$ as recorded in Table 10. For M_3, this is less straight forward. First, the value $\bar{\rho} = 1 - \bar{c}/T$ at which asymptotic power is 0.5 will be a function of ν if $\nu \geq 0.5$. For all values of $\nu < 0.5$, the value for $\bar{\rho}$ used in M_1 was employed, otherwise we approximate this value by utilizing the saddlepoint approximation in Theorem 1, and solving $\tilde{F}_{I_T,\bar{\rho}}(\kappa_5) = 0.5$.

Once the relevant value for $\bar{\rho}$ has been found for each series, the P-values are then obtained as described for M_1 and M_2. These values, along with the trend parameter ν are also given in Table 10. This process is vastly more straight forward than having to first characterize the asymptotic distribution in each such case and then simulate the power envelope to find the relevant value of $\bar{\rho}$. The results are mostly self-explanatory. For most of the series we do not reject a unit root for any of the mean specifications, but the P-value is generally lower when a linear trend is included. Although for the majority of series the $ESPO$ tests are not particularly sensitive to the mean function, the exceptions make the exercise worthwhile.

For Unemployment when there is just a constant the P-value is 0.1%, while including a trend the P-value becomes 1%. Potentially, therefore, if we were to test at the 1% level, whether or not a trend is included (and if so whether to impose a linear trend) could influence the outcome of the test. Bond Yields and Velocity are unusual, in the sense that the P-value is lower without a trend, of any kind. Most interesting of all is Industrial Production. When no trend is included the P-value is in the far-right tail (93%) while if a trend is included the P-value is in the far-left tail (7%). Using a nonlinear trend instead gives a P-value of 32%. Potentially, therefore, each specification could lead a practitioner to a different conclusion, M_1 giving an outcome consistent with an explosive root, M_2 with a stationary root and M_3 with a unit root.

6 Conclusions

This paper has provided a feasible method for obtaining approximate critical values for an optimal unit root test. Under restrictive assumptions the order of error of the approximation is proven to be $O(T^{-1})$ while under a general assumption on the innovation process, inference based on the feasible estimated saddlepoint approximation is shown to be asymptotically valid. In addition, because the saddlepoint provides an explicit distributional approximation, unlike numerical approximations to the distribution of limiting forms involving stochastic integrals, the former may be employed to examine the sensitivity of P-values under deterministic components.

The test statistic is that previously considered in Dufour and King (1991) and, via Francke and de Vos (2007), is directly related to that of Elliott, Rothenberg and Stock (1996). Numerical comparisons between the estimated saddlepoint procedure and pure asymptotic approximations demonstrate clear superiority for the former, demonstrated over different distributional assumptions and different methods of estimating the nuisance parameters involved in the innovation covariance structure.

Although the performance of the saddlepoint are comparable with the few given for bootstrap based inference they are not, as yet, available with quite the same generality, for example the wild bootstrap of Cavaliere and Taylor (2009). Importantly here there is no compromise in terms of power, such has yet to be clearly demonstrated for bootstrap based methods. Although beyond the scope of this paper extending the analysis to bootstrap based saddlepoint inference, i.e. bootstrapping the estimated saddlepoint P-value, for comparison with sieve based methods for Dickey-Fuller type tests is an obvious avenue for future research.

References

Abadir, K.M. (1993a) On the asymptotic power of unit-root tests. *Econometric Theory*, 9, 189-221.

Abadir, K.M. (1993b) The limiting distribution of the autocorrelation coefficient under a unit root. *The Annals of Statistics*, 21,1058-1070.

Bühlmann, P. (1995) Moving-average representation of autoregressive approximations. *Stochastic Processes and their Applications*, 60, 331-342.

Bühlmann, P. (1998) Sieve bootstrap for smoothing in nonstationary time series. *The Annals of Statistics*, 26, 48-83.

Butler, R.W. and M.S. Paolella (2002) Saddlepoint approximation and bootstrap inference for the Satterthwaite class of ratios. *Journal of the American Statistical Association*, 97, 836-846.

Cavaliere, G. and A.M.R. Taylor (2008) Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory*, 24, 43-71.

Cavaliere, G. and A.M.R. Taylor (2009) Heteroskedastic time series with a unit root. *Econometric Theory*, forthcoming.

Chang, Y. and J.Y. Park (2002) On the asymptotics of ADF tests for unit roots. Econometric Reviews, 21, 431-447.

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

Daniels, H.E. (1987) Tail Probability Approximations. *International Statistical Review,* 55, 37-48.

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.

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

Francke, M.K. and A.F. de Vos (2007) Marginal likelihood and unit roots. *Journal of Econometrics*, 137, 708-728.

Goutis, C. and G. Casella (1999) Explaining the saddlepoint approximation. *The American Statistician*, 53, 216-224.

Jing, B-Y. and J. Robinson (1994) Saddlepoint approximations for marginal and conditional probabilities of transformed variables. *Annals of Statistics*, 22, 1115–1132.

Juhl, T. and Z. Xiao (2003) Power functions and envelopes for unit root tests. *Econometric Theory*, 19, 240–253.

Larsson, R. (1998) Distribution approximation of unit root tests in autoregressive models. *Econometrics Journal*, 1, 10-26.

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

Lieberman, O. (1996) Saddlepoint approximation for the least squares estimator in first-order autoregression. *Biometrika*, 81, 807-811.

Lugannani, R. and S. Rice (1980) Saddlepoint approximations for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12, 475-490.

Magnus, J.R. and H. Neudecker (1999) *Matrix Differential Calculus, with Applications in Statistics and Econometrics* (revised edition), Chichester, Wiley.

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

Marsh, P. (2007) The available information for invariant tests of a unit root. *Econometric Theory*, 23, 686-710.

Marsh, P. (2009) The properties of Kullback-Leibler divergence for the unit root hypothesis. *Econometric Theory*, forthcoming.

Nabeya, S. and P. Perron (1994) Local asymptotic distribution related to the AR(1) model with dependent errors, *Journal of Econometrics*, 62, 29-264.

Nabeya, S. and K. Tanaka (1990) Limiting power of unit-root tests in time-series regression. *Journal of Econometrics*, 46, 247–271.

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, 139-162.

Ng, S. and P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica,* 69, 1519–1554.

Pantula, S.G. (1991) Asymptotic distributions of unit root tests when the process is nearly stationary. *Journal of Business and Economic Statistics* 9, 63-72.

Perron, P. and Z. Qu (2007) A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters,* 94, 12-19.

Phillips, P.C.B. (1978) Edgeworth and saddlepoint approximations in a first-order autoregression. *Biometrika*, 65, 91-98.

Phillips, P.C.B. (1987a) Time Series Regression with a Unit Root. *Econometrica*, 55, 277-301.
Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression. *Biometrika*, 74, 535-547.
Phillips, P.C.B. and P. Perron (1988) Testing for a unit root in time series regression. *Biometrika,* 75, 335-346.
Richard, P. (2007) Sieve bootstrap unit root tests. *Cahiers de recherche 07-05*, Departement d'Economique de la Faculte d'administration à l'Universite de Sherbrooke.
Seo, M.H. (2006) Improving Unit Root Testing with a New Long Run Variance Estimator, mimeo London School of Economics, UK.

Appendix I

(1) Proof of Theorem 1

The leading term saddlepoint approximation under Condition 1A have previously been detailed, see Lieberman (1994 & 1996) and Marsh (1998). Here what is relevant, since in general the covariances γ_j will be unknown and will have to be consistently estimated, is establishing the asymptotic order of error. To do so we demonstrate that the integral defining the probabilities $F^T_{\Gamma,1-\rho}(\kappa)$ satisfies the conditions required for the asymptotic analysis of Lugannani and Rice (1980) and Daniels (1987).

To proceed, define for any $\bar{\rho}$,

$$w_{\bar{\rho}} = C'_{\bar{\rho}}\Delta_{\bar{\rho}}y \sim N\left(0, A_{\bar{\rho}}\right), \quad A_{\bar{\rho}} = C'_{\bar{\rho}}\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\Gamma\left(\Delta_{\rho}^{-1}\right)'\Delta'_{\bar{\rho}}C_{\bar{\rho}}.$$

where $C_{\bar{\rho}}$ is the singular value decomposition in (6). Standardizing, so that $z = A_{\bar{\rho}}^{-1/2}w_{\bar{\rho}} \sim N\left(0, I_n\right)$ and $n = T-k$, then the test statistic can be written as the ratio of quadratic forms in standard normal variables, as in

$$PO_T = \frac{z'A_{\bar{\rho}}z}{z'A_1z}.$$

As a consequence,

$$F^T_{\Gamma,1-\rho}(\kappa) = \Pr\left[\frac{z'A_{\bar{\rho}}z}{z'A_1z} < \kappa\right] = \Pr\left[z'\left(A_{\bar{\rho}} - \kappa A_1\right)z < 0\right] = \Pr\left[\sum_{t=1}^{n}\lambda_t z_t^2 < 0\right], \quad (35)$$

where $z_t \sim N(0,1)$ and the non-zero eigenvalues of $B(\rho)$ as defined in (12) are also the n eigenvalues of the matrix $A = A_{\bar{\rho}} - \kappa A_1$, with

$$Ar_t = \lambda_t r_t \quad ; \quad r'_t r_t = 1, \quad r'_t r_s = 0 \text{ if } t \neq s. \quad (36)$$

Let $Q=\sum_{t=1}^{n}\lambda_t z_t^2$,then $F_{\Gamma,1-\rho}^{T}(\kappa)$ can be evaluated via,

$$F_{\Gamma,1-\rho}^{T}(\kappa)=F_Q(0)\quad;\quad F_Q(q)=\Pr\left[Q<q\right].$$

The mean moment generating and cumulant generating functions of Q are, respectively

$$\bar{\varphi}_n(\chi)=\prod_{t=1}^{n}(1-2\chi\lambda_t)^{-\frac{1}{2n}}\quad\text{and}\quad R_n(\chi)=-\frac{1}{2n}\sum_{t=1}^{n}\ln(1-2\chi\lambda_t).$$

Let $\theta=i\chi$, then from Goutis and Casella (1999, eqn 18) the density of Q is,

$$f_Q(q)=\frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty}\exp\left\{nR_n(\theta)-\theta q\right\}d\theta,$$

for some small τ. The distribution is,

$$\begin{aligned}F_Q(q)&=1-\int_q^{\infty}f_Q(w)dw=1-\int_q^{\infty}\frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty}\exp\left\{nR_n(\theta)-\theta q\right\}d\theta dw\\&=1-\frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty}\frac{\exp\left\{nR_n(\theta)-\theta q\right\}}{\theta}d\theta,\end{aligned}$$

and since we only require evaluation of the distribution at $q=0$, we set $x=n^{-1}q$ and evaluate the integral,

$$F_Q(x)=1-\frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty}\frac{\exp\left\{n\left(R_n(\theta)-\theta x\right)\right\}}{\theta}d\theta. \tag{37}$$

Following the development of Lugannani and Rice (1980) and also Daniels (1987) a formal asymptotic series approximation in powers of n^{-1} for (37) applies under the following conditions on the mean characteristic function of Q: let $\tilde{\theta}_{\rho,x}$ solve

$$\frac{dR_n(\theta)}{d\theta}=\sum_{t=1}^{n}\left(\frac{2\lambda_t}{1-2\tilde{\theta}_{\rho,x}\lambda_t}\right)=x,\quad\frac{1}{2\lambda_n}\leq\tilde{\theta}_{\rho,x}\leq\frac{1}{2\lambda_1} \tag{38}$$

and let ε,α,c_0 and c_1 be positive constants, then we require

$$\begin{array}{l}\text{a) } \bar{\varphi}_n(\theta) \text{ is analytic on the strip } \Theta=\left\{\theta:\operatorname{Im}(\theta)\leq\pm\left(\tilde{\theta}_{\rho,x}+\varepsilon\right)\right\},\\ \text{b) } \left|\bar{\varphi}_n(\theta)\right|<\frac{c_0}{|\theta|^{\alpha}} \text{ when } |\theta|>c, \text{ and}\\ \text{c) The derivatives of } R_n(\theta) \text{ are } O(1).\end{array}. \tag{39}$$

In (39) a) and b) ensure the path of integration in (37) can be deformed as in,

$$F_Q(x)=1-\int_q^{\infty}\frac{1}{2\pi i}\int_{\tilde{\theta}_{\rho,x}-i\infty}^{\tilde{\theta}_{\rho,x}+i\infty}\frac{\exp\left\{n\left(R_n(\theta)-\theta x\right)\right\}}{\theta}d\theta,$$

while c) ensures that the resulting asymptotic series approximation is in powers of T^{-1}.

Whether or not they hold depends entirely on the properties of the eigenvalues of $A_{\bar{\rho}}$, say $\{\mu_t\}_{t=1}^{n}$. To proceed, notice that the critical value κ is determined by,

$$\Pr\left[PO_T < \kappa\right] = \Pr\left[\frac{z'A_{\bar{\rho}}z}{z'A_1 z} < \kappa\right] = \alpha, \tag{40}$$

for $\alpha < 1$. Since $A_{\bar{\rho}}$ is symmetric, the eigenvalues of $A'_{\bar{\rho}}A_{\bar{\rho}} = A^2_{\bar{\rho}}$ are $\{\mu_t^2\}_{t=1}^{n}$,and consequently the maximum eigenvalue norm is given by

$$\begin{aligned}\sup_t |\mu_t| &= \|A_{\bar{\rho}}\|_2 = \left\|C'_{\bar{\rho}}\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\Gamma\left(\Delta_{\rho}^{-1}\right)'\Delta'_{\bar{\rho}}C_{\bar{\rho}}\right\|_2 \\ &\leq \|M_{\bar{\rho}}\|_2 \left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_2^2 \|\Gamma\|_2,\end{aligned}$$

since all $\|.\|_p$-norms are submultiplicative. Since $M_{\bar{\rho}}$ is idempotent then $\|M_{\bar{\rho}}\|_2 = 1$, while exploiting the matrix norm inequality $\|R\|_2^2 \leq \|R\|_1 \|R\|_\infty$ for any $n \times n$ matrix R, then

$$\|\Gamma\|_2 \leq m_\gamma < \infty \quad \text{and} \quad \left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_2^2 \leq \left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_1 \left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_\infty.$$

Since $\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}$ is Toeplitz then $\left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_\infty = \left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_1$, and

$$\begin{aligned}\left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_1 &= \sum_{j=0}^{n-1}\rho^j - \bar{\rho}\sum_{j=1}^{n-1}\rho^{j-1} = (1-\bar{\rho})\sum_{j=0}^{n-1}\rho^j + \rho^{n-1} \\ &= \frac{(1-\bar{\rho})}{(1-\rho)}\left(1-\rho^{n-1}\right) + \rho^{n-1} = \frac{\bar{c}}{c} - \frac{\bar{c}-c}{c}\left(1-\frac{c}{T}\right)^{(T-k)-1},\end{aligned}$$

so that

$$\lim_{T\to\infty}\left\|\Delta_{\bar{\rho}}\Delta_{\rho}^{-1}\right\|_1 = 1.$$

Consequently,

$$\lim_{T\to\infty}\|A_{\bar{\rho}}\|_2 \leq m_\gamma < \infty, \tag{41}$$

moreover, since both $A_{\bar{\rho}}$ and A_1 are positive definite then $PO_T = O_p(1)$ and so for any $\alpha < 1$, κ defined by (40) is positive and finite. Now note that the $\{\lambda_t\}_{t=1}^{n}$ are the eigenvalues of the symmetric $n \times n$ matrix $A = A_{\bar{\rho}} - \kappa A_1$, and so λ_t is real for all t. Moreover, the eigenvalues of $A'A = A^2$ are $\left\{\lambda_t^2\right\}_{t=1}^{n}$ and hence

$$\sup_t |\lambda_t| = \|A\|_2 = \|A_{\bar{\rho}} - \kappa A_1\|_2 \leq \|A_{\bar{\rho}}\|_2 + \kappa\|A_1\|_2,$$

using (41), which also holds for $\bar{\rho} = 1$, we find

$$\sup_t |\lambda_t| \leq (1+\kappa)m_\gamma = O(1).$$

Since the eigenvalues $\{\lambda_t\}_{t=1}^n$ are all finite for any n,then $\bar{\varphi}_n(\theta)$ is continuous (and infinitely differentiable) in θ and bounded away from zero as $|\theta| \to 0$. Hence $\bar{\varphi}_n(\theta)$ is analytic on Θ and condition (a) in (39) holds, see also Marsh (1998, Theorem 1(ii)).

To continue, since

$$|\bar{\varphi}_n(\theta)| \leq |1 - 2\theta\lambda_{t^*}|^{-\frac{1}{2}}, \quad t^* = \arg\max\sup_t |1 - 2\theta\lambda_t|^{-\frac{1}{2}}$$

so that for any $\alpha < 1/2$, $c_1 = \frac{1}{2\lambda_{t^*}-1}$, and $c_0 = 1$, we have

$$|\theta|^\alpha |\bar{\varphi}_n(\theta)| < c_0 \quad \text{for all } |\theta| > c_1,$$

and hence the condition b) also holds in this case. Under these conditions the leading term saddlepoint approximation evaluated at $x = 0$ is,

$$\tilde{F}_Q(0) = \Phi\left(\tilde{\eta}_\rho\right) - \varphi(\tilde{\eta}_\rho)\left(\frac{1}{\tilde{\eta}_\rho} - \frac{1}{\tilde{\delta}_\rho}\right), \tag{42}$$

where $\tilde{\eta}_\rho$ and $\tilde{\delta}_\rho$ are defined as in the statement of the Theorem and the saddlepoint $\tilde{\theta}_\rho = \tilde{\theta}_{\rho,0}$ satisfies,

$$\sum_{t=1}^n \left(\frac{2\lambda_t}{1 - 2\tilde{\theta}_\rho\lambda_t}\right) = 0 \quad ; \quad \frac{1}{2\lambda_n} \leq \tilde{\theta}_\rho \leq \frac{1}{2\lambda_1}.$$

Details on the construction of $\tilde{F}_Q(q)$ are given in Section 4 of Daniels (1987) and applied for the problem of a ratio quadratic forms in Lieberman (1994) and Marsh (1998).

The order of error of the leading term (and subsequent corrections to it) is determined by condition (c). The s^{th} derivative of $R_n(\theta)$ is

$$R_n^s(\theta) = \frac{d^s R_n(\theta)}{d\theta^s} = \frac{(-1)^s (s-1)!}{n} \sum_{t=1}^n \left(\frac{2\lambda_t}{1 - 2\theta\lambda_t}\right)^s,$$

so that $R_n^s(\theta) = O(1)$, as required. As an immediate consequence we have that

$$F_Q(0) = \tilde{F}_Q(0) + O(n^{-1}),$$

which then gives the order of error as required. ■

(2) Proof of Theorem 2

It is required to show that the Estimated Saddlepoint P-value is a consistent estimator for the asymptotic P-value of the PO_T test. To do so we parameterize the covariance matrix as

$$\Gamma = K_{\psi} K_{\psi}', \quad K_{\psi} = I_T + \sum_{j=1}^{T} \psi_j L^j,$$

Now similar to Richard (2007) let $h_T = o(p_T)$ and define the estimators $\hat{\psi}_h = \left\{ \hat{\psi}_{j,T} \right\}_{j=1}^{h_T}$, where

$$\left(\hat{\Xi}_h(z) \right)^{-1} = \sum_{j=1}^{\infty} \hat{\psi}_{j,T} z^j, \tag{43}$$

then according to the proof of Theorem 2 of Bühlmann (1995),

$$\lim_{h_T \to \infty} \sum_{j=1}^{h_T} \left| \hat{\psi}_{j,T} - \psi_j \right| \to 0, \quad \text{almost surely.} \tag{44}$$

Letting $\hat{\Gamma}_h = K_{\hat{\psi}_T} K_{\hat{\psi}_T}'$, then we can write

$$\begin{aligned} \left| \tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - F_{\Gamma,1}(\kappa) \right| &= \left| \left(\tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - \tilde{F}_{\Gamma,1}(\kappa) \right) + \left(\tilde{F}_{\Gamma,1}(\kappa) - F_{\Gamma,1}(\kappa) \right) \right| \\ &\leq \left| \left(\tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - \tilde{F}_{\Gamma,1}(\kappa) \right) \right| + \left| \left(\tilde{F}_{\Gamma,1}(\kappa) - F_{\Gamma,1}(\kappa) \right) \right|, \end{aligned} \tag{45}$$

via the triangle inequality. Immediately from Corollary 1 the second term in (45) satisfies

$$\left| \tilde{F}_{\Gamma,1}(\kappa) - F_{\Gamma,1}(\kappa) \right| = o(1),$$

for all κ, and can thus be neglected.

For the first term, write the estimator for $\psi_T = \{\psi_j\}_{j=1}^{T}$ as $\hat{\psi}_T = \left\{ \hat{\psi}_j \right\}_{j=1}^{T}$,

$$\hat{\psi}_j = \begin{cases} \hat{\psi}_{j,T} & \text{if } j \leq h_T \\ 0 & \text{otherwise} \end{cases},$$

and finally define the mean value ψ_T^* which lies on a line segment joining $\hat{\psi}_T$ and ψ_T. Indexing the distributions by the parameters, rather than the covariance matrix, we have,

$$\tilde{F}_{\hat{\psi}_T,1}(\kappa) - \tilde{F}_{\psi_T,1}(\kappa) = \left(\hat{\psi}_T - \psi_T \right)' \frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_T},$$

where the first derivative is evaluated at ψ_T^*,and so

$$\tilde{F}_{\hat{\psi}_T,1}(\kappa) - \tilde{F}_{\psi_T,1}(\kappa) = \sum_{j=1}^{h_T} \left(\hat{\psi}_{j,T} - \psi_j\right) \frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_j} - \sum_{h_T+1}^{T} \psi_j \frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_j}. \tag{46}$$

To proceed, from Corollary 1 we constructed the leading term by $\tilde{F}_{\Gamma,1}(\kappa) = \Phi(\hat{r}_1(\kappa))$, with $\hat{r}_1(\kappa)$ defined by (25), (26) and (27). Immediately, therefore, $\Phi(\hat{r}_1(\kappa))$ is a differentiable function of the eigenvalues $(\lambda_t)_{t=1}^{T}$. Moreover, from Magnus and Neudecker (1999, Theorem 7, p. 158), the eigenvalues are differentiable (with respect to ψ_j), with first derivative

$$\frac{\partial \lambda_t}{\partial \psi_j} = r_t' \left(\frac{\partial A}{\partial \varphi_j}\right) r_t, \quad A = A_{\bar{\rho}} - \kappa A_1.$$

The derivative of $A_{\bar{\rho}}$ (in the current parametrization of Γ) is

$$\frac{\partial A_{\bar{\rho}}}{\partial \psi_j} = C_{\bar{\rho}}' \Delta_{\bar{\rho}} \Delta_\rho^{-1} \left(\frac{\partial K_\psi K_\psi'}{\partial \psi_j}\right) \left(\Delta_\rho^{-1}\right)' \Delta_{\bar{\rho}}' C_{\bar{\rho}} = 2 C_{\bar{\rho}}' \Delta_{\bar{\rho}} \Delta_\rho^{-1} L^{(j)} K_\psi' \left(\Delta_\rho^{-1}\right)' \Delta_{\bar{\rho}}' C_{\bar{\rho}}.$$

Since $r_t' r_t = 1$, then

$$\lim_{T\to\infty} \left|\frac{\partial \lambda_t}{\partial \psi_j}\right| \leq \lim_{T\to\infty} \left\|\frac{\partial \left(A_{\bar{\rho}} - \kappa A_1\right)}{\partial \varphi_j}\right\|_2 = O(1),$$

by arguments identical to those proving that the maximum eigenvalue of A, itself, is finite. As a consequence $\tilde{F}_{\Gamma,1}(\kappa)$ has bounded derivatives with respect to the parameters ψ_j, i.e. we have

$$\lim_{T\to\infty} \max_j \left|\frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_j}\right| = O(1). \tag{47}$$

Hence from (46) and the triangle inequality, we have for all finite κ,

$$\lim_{T\to\infty} \left|\tilde{F}_{\hat{\psi}_T,1}(\kappa) - \tilde{F}_{\psi_T,1}(\kappa)\right| \leq \lim_{T\to\infty} \max_j \left|\frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_j}\right| \left(\sum_{j=1}^{h_T} \left|\hat{\psi}_{j,T} - \psi_j\right| + \sum_{h_T+1}^{T} \left|\psi_j\right|\right).$$

The first term in the parentheses is $o_p(1)$ almost surely, as in (44), while

$$\sum_{h_T+1}^{T} \left|\psi_j\right| = o\left(h_T^{-r}\right),$$

and so since κ is finite for any significance level $\alpha < 1$ then,

$$\lim_{T\to\infty} \left|\tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - F_{\Gamma,1}(\kappa)\right| \to 0, \quad \text{almost surely},$$

as required. ■

Appendix II Tables

Table 1): Nominal size of critical values for the PO_T test, $\xi_t \sim iidN(0,1)$.

T	α \ ψ	M_1 -0.8	-05	0.0	0.5	0.8	M_2 -0.8	-0.5	0.0	0.5	0.8
50	5%	5.10	5.41	4.36	4.16	4.49	5.59	4.56	4.71	4.74	4.71
50	10%	9.23	9.70	9.51	9.48	9.34	10.6	9.48	9.70	9.65	9.68
100	5%	5.16	4.67	4.89	4.64	4.46	5.50	5.22	5.13	5.69	5.36
100	10%	9.86	9.69	10.1	9.36	9.45	10.4	9.84	9.86	10.4	10.3
200	5%	5.10	5.21	4.86	5.31	5.23	5.00	4.95	4.83	5.28	5.28
200	10%	9.92	9.95	10.1	10.5	10.2	10.1	10.1	9.84	10.3	10.2

Table 2): Nominal size of critical values for the L_T test, $\xi_t \sim iidN(0,1)$.

T	α \ ψ	M_1 -0.8	-05	0.0	0.5	0.8	M_2 -0.8	-0.5	0.0	0.5	0.8
50	5%	0.00	0.00	7.42	8.62	8.44	0.00	0.00	6.69	11.4	12.0
50	10%	0.00	0.00	15.2	16.2	16.7	0.00	0.00	13.5	18.7	19.9
100	5%	0.00	0.00	6.92	6.58	6.85	0.00	0.00	5.92	7.84	7.25
100	10%	0.00	0.00	13.7	12.7	13.5	0.00	0.00	12.4	14.3	13.6
200	5%	0.00	0.00	5.62	5.63	5.08	0.00	0.00	5.28	5.79	5.67
200	10%	0.00	0.00	11.5	11.7	11.3	0.00	0.00	10.3	11.6	10.9

Table 3): Nominal size of critical values for the PO_T test, $\xi_t \sim \frac{iid\chi^2(1)-1}{\sqrt{2}}$.

T	α \ ψ	M_1 -0.8	-0.5	0.0	0.5	0.8	M_2 -0.8	-0.5	0.0	0.5	0.8
50	5%	4.15	3.88	4.18	4.27	4.22	4.15	3.88	4.18	4.29	4.21
50	10%	8.92	8.90	8.26	8.64	8.87	8.72	8.61	8.72	8.63	8.69
100	5%	4.52	3.87	4.02	5.33	4.23	3.97	4.63	4.32	4.37	4.17
100	10%	9.31	8.85	9.00	11.3	8.98	8.75	9.16	9.11	8.85	8.91
200	5%	4.78	5.33	4.32	4.68	4.47	4.28	4.73	4.88	4.63	5.78
200	10%	9.66	9.81	9.11	9.31	9.24	9.71	9.15	10.3	8.88	11.0

Table 4): Nominal size of critical values for the L_T test, $\xi_t \sim \frac{iid\chi^2(1)-1}{\sqrt{2}}$.

T	α \ ψ	M_1 -0.8	-0.5	0.0	0.5	0.8	M_2 -0.8	-0.5	0.0	0.5	0.8
50	5%	0.00	0.00	8.16	9.56	9.17	0.00	0.00	15.1	18.3	17.9
50	10%	0.00	0.00	16.0	17.7	17.9	0.00	0.00	22.1	26.0	25.4
100	5%	0.00	0.00	6.16	7.30	7.27	0.00	0.00	9.72	11.5	11.8
100	10%	0.00	0.00	12.5	13.9	13.8	0.00	0.00	15.8	18.2	18.5
200	5%	0.00	0.00	5.92	5.56	5.65	0.00	0.00	6.87	7.94	8.21
200	10%	0.00	0.00	11.7	11.0	11.8	0.00	0.00	12.6	13.2	13.7

Table 5): Size of $ESPO$ procedures, nominal size α :
(a) ψ estimated via MA(1),
(b) ψ estimated via BIC selection, (c) Γ estimated by sample autocovariances.

i) $T = 100$

	ψ / α	M_1					M_2				
		-0.8	-0.5	0.0	0.5	0.8	-0.8	-0.5	0.0	0.5	0.8
(a)	5%	4.87	3.92	4.56	5.72	5.81	5.64	4.13	4.32	5.26	5.04
	10%	9.43	8.36	10.1	9.67	10.9	11.0	8.28	9.17	10.9	10.7
(b)	5%	5.54	6.55	4.93	5.69	5.41	5.63	5.49	4.40	5.41	4.96
	10%	11.1	11.9	10.5	10.8	10.6	11.0	11.5	9.09	11.1	10.6
(c)	5%	8.72	5.79	4.30	6.29	6.56	8.01	5.19	3.38	6.39	5.87
	10%	13.7	11.5	8.65	11.3	12.0	13.1	10.1	8.37	10.8	11.6

ii) $T = 200$

	ψ / α	M_1					M_2				
		-0.8	-0.5	0.0	0.5	0.8	-0.8	-0.5	0.0	0.5	0.8
(a)	5%	4.23	4.25	5.17	5.89	5.71	4.39	3.97	5.10	5.61	5.73
	10%	7.87	8.35	10.5	11.3	10.8	9.67	8.86	9.43	10.8	11.3
(b)	5%	6.01	5.71	5.30	5.79	5.44	5.64	5.51	5.31	6.04	4.76
	10%	10.8	11.3	9.79	10.6	10.4	10.3	10.6	10.7	11.1	10.1
(c)	5%	8.08	5.51	4.83	6.01	5.44	4.65	5.59	4.28	6.15	6.74
	10%	13.4	10.2	8.75	11.1	9.24	11.0	10.5	8.79	8.93	11.7

Table 6): Size of P_T^* test, nominal size α :

(a) ω known, (b) ω estimated via MA(1),

(c) ω estimated via BIC selection, (d) ω estimatred by sample autocovariances.

i) $T = 100$

	ψ / α	M_1 -0.8	-0.5	0.0	0.5	0.8	M_2 -0.8	-0.5	0.0	0.5	0.8
(a)	5%	1.23	1.40	6.44	6.67	6.66	0.74	0.98	5.55	7.28	7.52
	10%	1.68	1.99	12.8	12.9	13.2	0.99	1.22	11.6	14.1	14.4
(b)	5%	2.09	2.18	6.34	6.27	7.23	1.23	1.39	4.75	6.63	6.97
	10%	2.88	3.46	13.2	13.6	14.3	1.44	1.68	8.99	13.4	14.7
(c)	5%	15.2	13.3	5.23	5.83	4.58	6.25	6.29	3.46	4.83	3.66
	10%	21.2	17.8	10.5	11.6	10.8	9.56	10.2	7.17	11.3	7.89

ii) $T = 200$

	ψ / α	M_1 -0.8	-0.5	0.0	0.5	0.8	M_2 -0.8	-0.5	0.0	0.5	0.8
(a)	5%	0.42	0.72	6.05	5.85	5.61	0.33	0.48	4.92	6.83	6.41
	10%	0.59	0.94	12.2	11.9	11.8	0.36	0.56	10.2	11.7	11.5
(b)	5%	0.89	1.15	6.18	5.86	6.38	0.38	0.50	4.32	3.81	8.54
	10%	1.77	2.66	12.3	11.7	13.1	0.51	0.65	11.1	10.7	14.5
(c)	5%	18.6	15.1	4.62	5.58	4.92	10.0	9.11	3.96	5.85	4.65
	10%	25.7	22.4	10.1	11.7	10.8	14.8	13.3	8.97	11.3	9.44

Table 7): Size of $ESPO$ test in model M_2 with $\psi = -0.8$, estimated by MA(1), under the mis-specification in (34); $T = 100$.

i) ξ_t generated by E_4

	α	5%	10%
	-0.15	5.42	11.0
	-0.10	5.90	11.1
ϑ	-0.05	6.42	12.1
	0.05	8.58	15.0
	0.10	9.52	16.2
	0.15	11.2	18.3

ii) ξ_t generated by E_5

	α	5%	10%
	-0.15	5.84	10.9
	-0.10	6.06	11.6
ϑ	-0.05	6.17	11.8
	0.05	7.88	14.1
	0.10	9.34	16.4
	0.15	10.9	17.9

Table 8): Size corrected power of the $ESPO$ test; $T = 100$, size $\alpha = 5\%$

		M_1					M_2				
	ψ	-0.8	-0.5	0.0	0.5	0.8	-0.8	-0.5	0.0	0.5	0.8
	.95	18.2	22.2	28.3	26.1	25.8	10.8	10.5	11.3	11.2	11.3
	.90	38.5	48.9	64.8	61.7	58.5	21.0	29.2	31.6	28.3	27.3
ρ	.85	57.2	75.9	86.4	89.3	86.2	40.8	60.8	61.1	55.1	54.7
	.80	68.4	81.7	97.0	98.6	97.1	57.5	74.3	77.5	80.4	75.6
	.75	75.3	88.4	98.3	99.8	98.2	65.2	82.4	89.9	91.0	84.8
	.70	81.8	91.7	99.9	99.9	99.9	78.6	88.1	95.7	95.3	91.2

Table 9): Size corrected power of the P_T^* test; $T = 100$, size $\alpha = 5\%$

i) Model M_1

		M_1					M_2				
	ψ	-0.8	-0.5	0.0	0.5	0.8	-0.8	-0.5	0.0	0.5	0.8
	.95	18.1	21.6	27.1	26.0	24.6	10.3	10.5	11.4	10.8	10.1
	.90	38.5	47.7	58.9	57.4	58.1	17.4	28.4	30.2	24.6	23.5
ρ	.85	52.0	71.5	81.9	79.1	80.2	36.8	54.2	56.4	43.7	40.6
	.80	64.6	78.7	90.4	88.9	89.7	51.6	65.2	72.7	59.2	55.4
	.75	67.4	80.8	94.5	93.0	94.3	64.3	78.1	85.4	73.7	68.3
	.70	68.2	81.5	96.3	95.5	95.4	72.3	82.8	94.6	82.8	79.6

Table 10): P-values for the $ESPO$ test applied to the Nelson and Plosser data

Series	T	M_1 P-val.	M_2 P-val.	M_3 P-val.	:	ν
Real GNP	80	95.8	41.7	99.9	:	-0.01
Nom. GNP	80	97.8	90.5	98.4	:	0.825
GNP Per.Cap.	80	97.7	43.1	89.8	:	0.518
Bond	89	73.3	93.0	85.6	:	1.47
Real Wage	89	99.5	70.6	99.5	:	0.813
Nom. Wage	89	97.7	88.8	98.6	:	0.305
Unemp.	99	0.10	1.00	0.31	:	-0.16
Employ.	99	93.2	26.5	25.2	:	0.81
Money	100	96.6	82.5	82.6	:	0.999
S&P500	118	93.6	70.4	53.8	:	1.17
Velocity	120	55.9	94.3	78.7	:	-0.57
I. Prod.	129	92.9	6.61	32.2	:	1.10
CPI	129	99.6	99.8	99.9	:	0.571

Table n): Powers of the SPO and L_T tests against the alternative $H_1 : \rho = 1 - \bar{c}/T$

T	ψ / Test	M_1					M_2				
		-0.8	-0.5	0.0	0.5	0.8	-0.8	-0.5	0.0	0.5	0.8
50	SPO	22.4	29.5	41.7	44.4	45.6	30.4	35.1	49.9	50.2	50.8
50	L_T	13.1	18.4	32.0	37.1	37.6	5.54	6.67	25.5	30.3	32.6
100	SPO	27.9	34.1	44.9	47.0	47.6	33.6	39.1	50.2	48.6	49.2
100	L_T	23.5	29.4	41.5	43.7	45.3	15.6	22.4	38.6	38.7	41.8
200	SPO	33.1	38.0	47.6	48.7	47.3	38.1	41.6	49.8	49.9	48.8
200	L_T	31.2	35.4	45.0	47.2	46.8	32.2	35.4	48.1	47.2	46.9