Econometric Theory, **27**, 2011, 1026–1047.
doi:10.1017/S0266466610000629

SADDLEPOINT AND ESTIMATED SADDLEPOINT APPROXIMATIONS FOR OPTIMAL UNIT ROOT TESTS

PATRICK MARSH
University of York

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 proved to yield valid asymptotic inference. Numerical evidence, considered over a range of models, demonstrates some finite-sample superiority over approximations for a directly comparable test based on simulation of its limiting stochastic representation.

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, 1987b). Indeed most significant innovations in the literature, e.g., Phillips and Perron (1988), Elliott, Rothenberg, and Stock (1996), and Ng and Perron (2001), yield tests whose properties are evaluated in this way. For some very noteworthy exceptions to this see Nabeya and Tanaka (1990), Abadir (1993a, 1993b), 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). Because 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),

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 correspondence to Patrick Marsh, Department of Economics, University of York, YO105DD, United Kingdom; e-mail: pwnm1@york.ac.uk.

 © Cambridge University Press 2011

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

and Larsson (1998) for applications closely related to the current context. Under general assumptions, the test is nonpivotal. 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; e.g., see Butler and Paolella (2002).

From Francke and de Vos (2007), the Dufour and King (1991) point optimal test is the ratio (rather than the difference) of the two residual sum of squares detailed in Elliott et al. (1996, p. 817) and forming their 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 earlier applied to the P_T test when the innovations follow simple time series processes.

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). Here we use both this estimator and a maximum likelihood estimator, under the assumption of a correctly specified Gaussian likelihood. Although the latter is more efficient, it does not necessarily produce asymptotic tests with better numerical properties in finite samples. Moreover, in common with previous findings in the literature for the P_T, and also other tests, the numerical performance does not necessarily improve as the sample size increases.

The saddlepoint-based procedure does not suffer from these issues. In terms of finite-sample size and power it generally outperforms the P_T tests whether the nuisance parameters are efficiently parametrically estimated or nonparametrically estimated. Moreover, because the implied distributional approximation is explicit it can be exploited to give P-values for the outcome of the test when applied to any data set. Because the approximation can be implemented for time series having any chosen deterministic component, approximate critical or P-values are immediately available for models involving more complicated deterministic components, such as broken trends as in Harris, Harvey, Leybourne, and Taylor (2009) or nonlinear trends as in Marsh (2009).

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, first under the Gaussianity assumption with known covariances yielding a known order of error. When an estimated covariance is instead required, the saddlepoint is then shown to provide asymptotically valid critical values under more primitive assumptions on the innovation. Section 4 details the numerical comparisons between the proposed test and its direct asymptotic competitor, P_T. Section 5 concludes, and an Appendix contains proofs of the main results.

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

2. THE MODEL AND "OPTIMAL" TESTS

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

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

and tests of the hypotheses

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

In (1) x_t is a $k \times 1$ deterministic regressor, β a $k \times 1$ unknown parameter, and $(\xi_t)_{t=1}^T$ a random innovation. Before giving the most general assumptions on the distribution of ξ_t we first derive a point optimal test under more restrictive assumptions.

Suppose that the innovations are independent and Gaussian, i.e., $\xi_t \sim iidN(0, \sigma^2)$, with finite variance σ^2. Define the following vectors and matrices, $y = (y_1, \ldots, y_T)'$ and $\xi = (\xi_1, \ldots, \xi_T)'$, and let $X = (x_1, \ldots, x_T)'$ and $\beta = (\beta_1, \ldots, \beta_k)'$. 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 jth 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 their approach 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

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

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

$$C_{\bar\rho} C_{\bar\rho}' = M_{\bar\rho}; \qquad C_{\bar\rho}' C_{\bar\rho} = I_n, \tag{5}$$

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

where $n = T - k$. We write $RSS_{\bar{\rho}} = w'_{\bar{\rho}} w_{\bar{\rho}}$, where $w_{\bar{\rho}} = C'_{\bar{\rho}} y_{\bar{\rho}}$ has density function (marginal likelihood)

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

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 $v_{\bar{\rho}} = w_{\bar{\rho}} \Big/ \sqrt{w'_{\bar{\rho}} w_{\bar{\rho}}}$, having density

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

The Neyman–Pearson point optimal invariant test of size α for (2) is to reject H_0 if

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

where $\bar{k}_\alpha$ is a constant chosen so that $\Pr\left[NP > \bar{k}_\alpha \,|\, H_0\right] = \alpha$. To focus on a specific test, here we follow Elliott et al. (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, and so

$$PO_T = \frac{y'_{\bar{\rho}} M_{\bar{\rho}} y_{\bar{\rho}}}{y'_1 M_1 y} = \frac{RSS_{\bar{\rho}}}{RSS_1}. \tag{6}$$

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 (6) and also its feasible variant described subsequently are similar to those developed in Elliott et al. (1996), which, instead, are based on the difference

$$L_T = y'_{\bar{\rho}} M_{\bar{\rho}} y_{\bar{\rho}} - y'_1 M_1 y = RSS_{\bar{\rho}} - RSS_1.$$

As in that paper we will detail the asymptotic properties of our test under much more general assumptions than those under which the test was designed. The most general assumptions under which we can provide asymptotically valid saddlepoint

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

inference are as in Assumption A of Bühlmann (1995) and Assumptions 1 and 2 of Chang and Park (2002). Here we make the following assumption.

Assumption 1.

(a) $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence with filtration $\mathcal{F}_t$ and such that (1) $\mathrm{E}\left[\varepsilon_t^2\right] = \sigma^2$, (2) $\mathrm{plim} T^{-1}\sum_1^T \varepsilon_t^2 = \sigma^2$, (3) $\mathrm{E}\left[|\varepsilon_t|^s\right] \leq m_\varepsilon < \infty$ for $s \geq 4$,
(b) $\xi_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} |j| \left|\psi_j\right| \leq m_\psi < \infty$, $\psi_0 = 1$, and
(c) $u_0 = O_p(1)$.

In common with much of the literature we exploit that Assumption 1 implies that the innovation process (ξ_t) is such that for all t, $\mathrm{E}[\xi_t] = 0$,

$$\omega^2 = \sum_{j=-\infty}^{\infty} \gamma_j > 0 \quad \text{and} \quad m_\gamma = \sum_{j=-\infty}^{\infty} \left|\gamma_j\right| < \infty, \tag{7}$$

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

$$T^{-1/2}\omega^{-1} \sum_{t=1}^{[rT]} \xi_t \Rightarrow B(r), \tag{8}$$

where $B(r)$ is standard Brownian motion on $[0, 1]$.

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 $\rho = 1 - c/T$, $c > 0$. This follows from the invariance principles detailed in, e.g., Phillips (1987a, 1987b). Indeed, immediate from Elliott et al. (1996, Thms. 1 and 2) and Francke and de Vos (2007, Sect. 3.3) is that both L_T and PO_T converge almost surely to continuous random variables as the sample size becomes infinite.

3. SADDLEPOINT APPROXIMATIONS FOR THE DISTRIBUTION OF PO_T

Because the asymptotic distribution of PO_T is constant for any innovation process satisfying Assumption 1 a representation for that distribution may be found from any specific process satisfying it. In Theorem 1 later in this section we derive a saddlepoint approximation, with given order of error, for the distribution of PO_T under the additional assumptions of Gaussianity and known innovation covariance structure. Then, exploiting the constancy of the asymptotic distribution of PO_T under Assumption 1, a corollary shows that the saddlepoint approximation remains asymptotically valid far more generally. Finally, Theorem 2 shows that substitution of a consistent estimator for the innovation covariance structure gives asymptotically correctly sized critical and probability values for PO_T.

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

To proceed, the saddlepoint approximation is first derived for an innovation process satisfying the following assumption.

Assumption 2. In addition to Assumption 1, suppose also that

(a) the initial condition is $u_0 = 0$;
(b) the innovation sequence $(\xi_t)_1^T$ is a stationary Gaussian random variable, with

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

Under Assumption 2, we have $\xi \sim N(0, \Gamma)$, where Γ is $T \times T$ Toeplitz. Notice that the implied condition on the long-run variance (7) ensures that Γ is bounded with respect to the matrix norms on the space of square Toeplitz matrices, $\|\Gamma\|_1 = \|\Gamma\|_\infty = \sum_{j=0}^{T} |\gamma_j| < m_\gamma < \infty$.

Define $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, 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 because 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

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

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,

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

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

$$R_n(\theta) = -\frac{1}{2n} \sum_{t=1}^{n} \log(1 - 2\theta\lambda_t),$$

$$\eta_\rho(\theta) = \mathrm{sign}[\theta] \sqrt{-2n R_n(\theta)} \quad \text{and} \quad \delta_\rho(\theta) = \theta \sqrt{n \frac{d^2 R_n(\theta)}{d\theta^2}}.$$

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

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

THEOREM 1. *Under Assumption 2,*

$$F^T_{\Gamma,\rho}(\kappa) = \Phi(\tilde{\eta}_\rho) - \varphi(\tilde{\eta}_\rho)\left(\frac{1}{\tilde{\eta}_\rho} - \frac{1}{\tilde{\delta}_\rho}\right) + O(T^{-1}), \tag{9}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal cumulative distribution function and probability density function,

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

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

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

Theorem 1 gives a standard leading term saddlepoint approximation for the distribution of PO_T, which under Assumption 2 is a ratio of quadratic forms in normal variables. The theorem complements previous results, in particular Lieberman (1994) and Marsh (1998), in that the order of error, $O(T^{-1})$, is explicitly established for the distribution function. This result is crucial for two reasons. Later it will allow for consistent asymptotic inference in the case where the covariance 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 that the asymptotic distribution is constant for all processes satisfying Assumption 1, demonstrates that critical values obtained from a Gaussian approximation to the leading term in (9) are asymptotically correctly sized.

COROLLARY 1. *In Assumption 2, let $\gamma_j = \mathrm{E}[\xi_t \xi_{t-j}]$ be known and under the null hypothesis, $H_0 : \rho = 1$, denote the leading term approximation by*

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

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

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

Corollary 1 essentially defines a (not yet feasible) saddlepoint point optimal (*SPO*) test, as in the following definition.

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

$$\tilde{F}_{\Gamma,1}(PO_T) < \alpha. \tag{10}$$

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

By way of comparison, under H_0, L_T has the asymptotic representation

$$L_T \Rightarrow \omega^2 \left(\bar{c}^2 \int_0^1 V_0^2 (r, \bar{c})\, dr + (1+\bar{c})\, V_0^2(1, \bar{c}) \right), \tag{11}$$

where $V_0 (r, \bar{c})$ is defined in Elliott et al. (1996, eqn. (7)) and H_0 is to be rejected by L_T if

$$\omega^{-2} L_T < c_\alpha .$$

The critical values, c_α, are then obtained via partial sum approximations to the stochastic integrals in (11).

Critical values for PO_T, delivered by Corollary 1 in (10), are not yet feasible because the covariance matrix Γ is assumed known. 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 nonzero eigenvalues of $\hat{B}_1$ be $\left\{ \hat{\lambda}_t \right\}_{t=1}^n$. Now define

$$\hat{F}_{\hat{\Gamma},1} (\kappa) = \Phi \left(\hat{r}_1 (\kappa) \right); \qquad \hat{r}_1 (\kappa) = \hat{\eta}_1 + \frac{1}{\hat{\eta}_1} \ln \left(\frac{\hat{\delta}_1}{\hat{\eta}_1} \right), \tag{12}$$

where

$$\hat{\eta}_1 = \operatorname{sign} \left(\hat{\theta}_1 \right) \sqrt{\sum_{t=1}^n \log \left(1 - 2\hat{\theta}_1 \hat{\lambda}_t \right)}; \qquad \hat{\delta}_1 = \hat{\theta}_1 \sqrt{2 \sum_{t=1}^n \left(\frac{\hat{\lambda}_t}{1 - 2\hat{\theta}_1 \hat{\lambda}_t} \right)^2} \tag{13}$$

and the saddlepoint solves

$$\sum_{t=1}^n \left(\frac{\hat{\lambda}_t}{1 - 2\hat{\theta}_1 \hat{\lambda}_t} \right) = 0. \tag{14}$$

Then the estimated saddlepoint point optimal (*ESPO*) test is defined as follows.

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

$$\hat{F}_{\hat{\Gamma},1} (PO_T) < \alpha. \tag{15}$$

In the context of an augmented Dickey–Fuller (ADF) regression Ng and Perron (2001) exploit a modified model selection criterion to deliver ADF (and other) tests that are asymptotically correctly sized. This, along with the work of Chang

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

and Park (2002), is based on an autoregressive approximation to the inverse of the transfer function of the linear process (ξ_t),

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

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

$$\hat{\Xi}(z) = \sum_{j=1}^{p_T} \hat{\alpha}_{j,T} z^j \quad \text{for } |z| \leq 1, \tag{16}$$

where $p_T = o\left(T^{1/2}\right)$ and $\hat{\alpha}_{j,T}$ are the autoregressive estimators obtained from the ADF regression

$$\Delta \tilde{y}_t = a_0 \tilde{y}_{t-1} + \sum_{j=1}^{p_T} a_{j,T} \Delta \tilde{y}_{t-j} + v_t, \tag{17}$$

where $\tilde{y}_t$ is the tth element of the quasi-detrended vector of observations $\tilde{y} = y - X\hat{\beta}_{\bar{\rho}}$, with $\hat{\beta}_{\bar{\rho}} = \left(X'_{\bar{\rho}} X_{\bar{\rho}} \right)^{-1} X'_{\bar{\rho}} y_{\bar{\rho}}$. Unlike Ng and Perron (2001) and Chang and Park (2002) here we invert the transfer function to yield an estimated covariance structure that can be plugged directly into the saddlepoint approximation given in Corollary 1. That the resulting estimated probability values of PO_T are asymptotically consistent is proved in the following theorem.

THEOREM 2. *Under Assumption 1, let* $h_T \to \infty$ *and* $h_T = o(p_T)$. *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,

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

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 autoregressive moving average model, an obvious simplification of this theorem proves validity for inference based on a more efficient estimator, such as the maximum likelihood. The numerical accuracy of the *ESPO* test is examined in the next section.

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

4. NUMERICAL COMPARISONS

The numerical experiments in this paper will be performed using data generated according to the following design.

Experimental design. Data $(y_t)_{t=1}^T$ are generated via

$$M_1 : y_t = \beta_1 + u_t; \qquad u_t = (1 - c/T)u_{t-1} + \xi_t,$$
$$M_2 : y_t = \beta_1 + \beta_2 t + u_t; \qquad u_t = (1 - c/T)u_{t-1} + \xi_t, \tag{18}$$

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

$$E_1 : \xi_t = \psi \varepsilon_{t-1} + \varepsilon_t; \qquad \psi = \{-0.8, -0.5, 0, 0.5, 0.8\}, \tag{19}$$

$$E_2 : \xi_t = \vartheta \xi_{t-1} + \varepsilon_t; \qquad \vartheta = \{-0.8, -0.5, 0, 0.5, 0.8\}, \tag{20}$$

where the underlying errors are generated by

$$\varepsilon_t \sim iidN(0, 1).$$

Here we assess the finite-sample properties, both size and power, of the *ESPO* test defined in (15) and compare them with those of the P_T test of Elliot et al. (1996). The latter involves rejecting H_0 if

$$P_T = \hat{\omega}^{-2} \left(y_{\bar{\rho}}' M_{\bar{\rho}} y_{\bar{\rho}} - y_1' M_1 y \right) < cv_\alpha, \tag{21}$$

where $\hat{\omega}$ is a consistent estimator for ω and the critical values cv_α are given in the last column in Table 1 of Ng and Perron (2001). The experiments are repeated for both models in (18) and the parameter values in (19) and (20), with two sample sizes $T = 100, 200$, and results are presented on the basis of 10,000 Monte Carlo replications.

We consider two different estimators for the nuisance parameters Γ (for the *ESPO* test) and ω^2 (for the P_T test). The first is based on the assumption of a correctly specified parametric likelihood, defined (up to constants not depending on parameters) for both models in (18) by

$$l\left(\beta, \rho, \psi, \vartheta, \sigma^2\right) = -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t (\beta, \rho, \psi, \vartheta)^2,$$

where

$$\varepsilon_t (\beta, \rho, \psi, \vartheta) = (1 - \rho l)(1 + \psi l)^{-1} (1 - \vartheta l) \left(y_t - \mu_{k,t}(\beta) \right), \qquad k = 1, 2. \tag{22}$$

In (22) $l^{(j)} y_t = y_{t-j}$, for model M_1, $\mu_{1,t}(\beta) = \beta_1$, and for model M_2, $\mu_{2,t}(\beta) = \beta_1 + \beta_2 t$. We impose $\vartheta = 0$ for the moving average innovations E_1 and $\psi = 0$

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

TABLE 1. Size of *ESPO* procedures, nominal size α, and (I) Γ estimated via maximum likelihood ($\hat{\Gamma}_I$), (II) Γ estimated by ADF regression ($\hat{\Gamma}_{II}$)

			M_1					M_2				
			$\psi =$					$\psi =$				
	T	α	−0.8	−0.5	0.0	0.5	0.8	−0.8	−0.5	0.0	0.5	0.8
(a) $MA(1)$ innovations												
(I)	100	5%	4.87	3.92	4.56	5.72	5.81	5.64	4.13	4.32	5.26	5.04
(I)	100	10%	9.43	8.36	10.1	9.67	10.9	11.0	8.28	9.17	10.9	10.7
(I)	200	5%	4.23	4.25	5.17	5.89	5.71	4.39	3.97	5.10	5.61	5.73
(I)	200	10%	7.87	8.35	10.5	11.3	10.8	9.67	8.86	9.43	10.8	11.3
(II)	100	5%	9.07	8.92	4.44	2.99	3.70	10.4	6.30	5.51	3.76	3.97
(II)	100	10%	15.8	14.9	7.83	6.91	8.06	18.3	17.6	11.2	7.78	9.36
(II)	200	5%	5.67	5.98	4.56	4.23	4.16	4.80	4.19	4.57	3.85	3.95
(II)	200	10%	11.6	12.1	8.72	7.61	7.76	9.41	8.52	9.03	7.99	9.41
			M_1					M_2				
			$\vartheta =$					$\vartheta =$				
	T	α	−0.8	−0.5	0.0	0.5	0.8	−0.8	−0.5	0.0	0.5	0.8
(b) $AR(1)$ innovations												
(I)	100	5%	4.27	4.33	4.52	4.31	4.78	3.70	4.35	5.39	3.45	3.96
(I)	100	10%	8.98	8.99	9.21	9.13	9.44	7.64	8.41	10.6	7.09	7.95
(I)	200	5%	4.83	4.29	5.06	4.73	4.67	4.90	5.15	4.66	4.13	5.31
(I)	200	10%	9.96	8.55	9.38	9.50	9.61	9.88	10.6	10.3	8.09	9.33
(II)	100	5%	10.7	9.02	—	3.75	3.70	14.2	9.40	—	3.51	3.26
(II)	100	10%	15.6	14.8	—	7.96	7.98	21.5	15.6	—	8.58	8.13
(II)	200	5%	6.35	4.67	—	4.28	3.77	4.41	3.95	—	3.85	3.75
(II)	200	10%	11.6	8.38	—	8.15	8.22	7.83	6.93	—	8.58	7.98

for the autoregressive innovations E_2, and it is assumed that $\varepsilon_t = 0$ for all $t \leq 0$, which implies $\xi_0 = 0$ and $\xi_1 = \varepsilon_1 \sim N(0, 1)$ in both E_1 and E_2.

For the *ESPO* test this yields parametric estimators of Γ,

$$\hat{\Gamma}_I = \hat{K}_I \hat{K}_I', \quad \hat{K}_I = \begin{cases} \left(I + \hat{\psi} L\right) & \text{for } E_1 \\ \left(I - \hat{\vartheta} L\right)^{-1} & \text{for } E_2 \end{cases},$$

where $\hat{\psi}$ and $\hat{\vartheta}$ are the maximum likelihood estimators for ψ and ϑ. For the P_T test the maximum likelihood estimators of ω^2 are

$$\hat{\omega}_I^2 = \begin{cases} \hat{\sigma}_1^2 \left(1 + \hat{\psi}\right)^2 & \text{for } E_1 \\ \hat{\sigma}_2^2 \left(1 - \hat{\vartheta}\right)^{-2} & \text{for } E_2 \end{cases},$$

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

where $\hat{\sigma}_j^2 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{j,t}^2$, where the residuals $\hat{\varepsilon}_{j,t}$ are obtained from the fitted likelihood with innovation process E_j.

The second estimators are based on the ADF regression given in (17). Specifically, define the ADF residuals by

$$\hat{v}_t = \hat{v}_t(p_T) = \Delta\tilde{y}_t - \hat{\alpha}_{0,T}\tilde{y}_{t-1} - \sum_{j=1}^{p_T}\hat{\alpha}_{j,T}\Delta\tilde{y}_{t-j},$$

and for each M_k the estimated lag length $\hat{p}_T$ is chosen according to the modified Akaike information criterion (MAIC) of Ng and Perron (2001).

For the *ESPO* test, for convenience, we put $h_T = \hat{p}_T$ and let $\hat{\psi} = \left(\hat{\psi}_{1,T},\ldots,\hat{\psi}_{h_T,T}\right)'$ denote the first h_T coefficients in the expansion of

$$\left(1 - \sum_{j=1}^{\hat{p}_T}\hat{\alpha}_{j,T}z^j\right)^{-1} = \sum_{j=1}^{\infty}\hat{\psi}_{j,T}z^j.$$

This yields a nonparametric estimator of the innovation covariance matrix as

$$\hat{\Gamma}_{II} = \hat{K}_{II}\hat{K}_{II}', \qquad \hat{K}_{II} = I + \sum_{j=1}^{\infty}\hat{\psi}_{j,T}L^j,$$

again defined for both M_1 and M_2. Similarly, for P_T, the autoregressive long-run variance estimator is

$$\hat{\omega}_{II}^2 = \hat{\sigma}^2\left(1 - \sum_{j=1}^{\hat{p}_T}\hat{\alpha}_{j,T}\right)^{-2},$$

where $\hat{\sigma}^2 = \left(T - \hat{p}_T\right)^{-1}\sum_{t=1}^{T}\hat{v}_t^2$.

The results for the *ESPO* test are given in Table 1, and those for the P_T test are in Table 2. Consider first the results based on the maximum likelihood estimator for the nuisance parameters. For the *ESPO* test the results using $\hat{\Gamma}_I$ are very close to nominal for both models, both innovation processes, and both sample sizes. For the case of autoregressive innovations there is also evidence of significant improvement in performance for the larger sample size. The equivalent results for the P_T test, based on the estimator $\hat{\omega}_1$, are less favorable. For both models, both innovation processes, and both sample sizes when the nuisance parameter is large and negative the test is significantly undersized; moreover for these values the performance is worse for the larger sample size. For other parameter values the size is closer to nominal, although slightly more undersized than the *ESPO* test.

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

TABLE 2. Size of P_T test, nominal size α, and (I) ω^2 estimated via maximum likelihood $\left(\hat{\omega}_I^2\right)$, (II) ω estimated via ADF regression $\left(\hat{\omega}_{II}^2\right)$

			M_1					M_2				
			$\psi =$					$\psi =$				
	T	α	−0.8	−0.5	0.0	0.5	0.8	−0.8	−0.5	0.0	0.5	0.8
(a) $MA(1)$ innovations												
(I)	100	5%	2.93	2.09	6.32	7.41	7.49	0.84	2.73	6.45	6.07	7.52
(I)	100	10%	5.88	5.47	12.7	12.9	13.2	2.23	6.21	12.8	11.8	14.4
(I)	200	5%	2.02	1.91	5.55	6.47	5.77	0.89	2.48	6.12	6.25	6.44
(I)	200	10%	3.99	4.03	12.2	11.9	11.8	2.36	2.68	11.3	11.6	11.5
(II)	100	5%	6.45	5.71	3.19	4.29	3.90	7.02	7.61	2.01	3.12	3.95
(II)	100	10%	12.4	16.1	10.4	9.11	13.9	12.8	10.7	7.53	6.94	7.11
(II)	200	5%	3.26	3.27	4.93	3.22	2.81	1.09	3.14	3.07	3.71	4.75
(II)	200	10%	7.02	7.67	10.7	6.95	6.34	2.51	5.88	10.6	7.63	3.61
			M_1					M_2				
			$\vartheta =$					$\vartheta =$				
	T	α	−0.8	−0.5	0.0	0.5	0.8	−0.8	−0.5	0.0	0.5	0.8
(b) $AR(1)$ innovations												
(I)	100	5%	2.24	4.05	5.42	5.71	3.90	1.01	1.84	4.05	4.55	3.28
(I)	100	10%	7.57	10.9	12.5	11.9	7.25	2.04	4.46	9.11	9.54	6.28
(I)	200	5%	2.90	4.48	4.60	4.93	3.36	0.71	2.86	3.86	4.35	2.44
(I)	200	10%	8.01	10.2	11.9	10.9	6.64	2.90	7.38	8.93	9.87	5.56
(II)	100	5%	1.60	2.95	—	5.89	8.02	0.64	1.09	—	3.33	6.04
(II)	100	10%	5.66	6.57	—	12.2	17.4	1.57	2.86	—	7.12	11.7
(II)	200	5%	2.89	4.38	—	5.41	6.71	0.93	2.03	—	3.91	6.01
(II)	200	10%	8.37	10.1	—	11.8	12.4	1.98	4.51	—	7.49	11.8

When the nuisance parameters are estimated by the ADF-based estimator the performance of the *ESPO* test suffers somewhat. For $T = 100$, for large negative values of the nuisance parameters the test becomes oversized, and for large positive values it becomes undersized, for both processes and models. For $T = 200$ the sizes are generally closer to nominal, particularly for the case of autoregressive innovations. As has been found in previous papers, particularly Ng and Perron (2001), for $T = 100$, use of the ADF regression with lag length selected by MAIC to provide an autoregressive long-run variance estimator has sizes close to nominal. In fact, at this sample size the P_T test performs better with the nonparametric estimator than it does with the maximum likelihood estimator. However, for $T = 200$, and as also found by Ng and Perron (2001), the test becomes significantly undersized for large negative values of the nuisance parameters in both models and innovation processes. Overall the performance of P_T is less stable than that of the *ESPO* test.

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

From the perspective of Theorem 1 and its corollary the saddlepoint procedure is in essence a parametric one, as in Lieberman (1994) and Marsh (1998). As found in those and many other papers the saddlepoint approximation provides extraordinarily accurate approximations when the underlying covariance structure is known. It is not surprising that substituting an efficient maximum likelihood estimator for this covariance structure maintains this accuracy. Of more practical relevance is that the performance of the test remains relatively stable when a nonparametric estimator is used and is generally more accurate than the P_T test and those presented in Perron and Qu (2007) using a refined selection criterion. As an alternative to pure asymptotic results recent developments in the literature have exploited sieve bootstrap methods for processes satisfying Assumption 1. For instance, in model M_1 and $T = 100$, Cavaliere and Taylor (2008) report a size of 3.8% for an autoregression-based sieve applied to a generalized least squares detrended Dickey–Fuller test, whereas Richard (2007) reports less favorable results for a moving average–based sieve.

The statistics PO_T and L_T are both constructed using power optimality criteria. Specifically, for independent Gaussian innovations, both are point optimal invariant against the alternative $H_1 : \rho = \bar{\rho}$, with only the invariance groups differing;

TABLE 3. Size corrected power of the *ESPO* test; $T = 100$, size $\alpha = 5\%$

	M_1					M_2				
	$\psi =$					$\psi =$				
ρ	−0.8	−0.5	0.0	0.5	0.8	−0.8	−0.5	0.0	0.5	0.8
(a) $MA(1)$ innovations										
0.95	12.4	25.1	30.3	31.4	30.3	7.66	9.92	11.8	11.7	12.0
0.90	30.1	48.4	67.6	71.9	70.0	14.0	22.7	31.4	27.6	30.1
0.85	39.9	64.2	89.1	91.4	90.3	20.5	42.8	59.3	55.9	58.2
0.80	48.1	73.1	96.0	97.8	97.6	29.1	62.4	83.1	80.7	80.9
0.75	51.8	77.5	97.8	99.3	99.3	37.8	68.1	94.2	93.5	93.1
0.70	55.3	84.9	98.8	99.6	99.5	44.6	73.7	97.6	98.3	98.3
	M_1					M_2				
ρ	−0.8	−0.5	0.5	0.8	−0.8	−0.5	0.5	0.8		
(b) $AR(1)$ innovations										
0.95	20.3	24.9	31.7	31.9	9.46	10.8	10.2	10.6		
0.90	43.7	56.0	72.3	71.6	23.5	28.5	29.8	27.4		
0.85	60.4	76.7	93.1	92.3	42.2	52.9	58.3	50.8		
0.80	72.0	86.3	98.7	97.9	59.0	73.7	81.7	74.5		
0.75	78.1	90.5	99.7	99.5	70.9	86.6	94.5	89.1		
0.70	83.5	92.4	99.9	99.9	80.8	94.1	98.9	96.0		

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

PO_T is also scale invariant. In practice, numerical evidence in the literature (e.g., Perron and Qu, 2007) suggests that P_T is slightly more powerful than procedures based on a test of $H_0 : a_0 = 0$ in the ADF regression (17). Here, therefore, we compare the size corrected power of the *ESPO* and P_T tests in both models in (18) and both innovation processes in (19) and (20), for testing

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

with $T = 100$ and for 10,000 replications. The nuisance parameters were estimated using the ADF-based estimators, i.e., $\hat{\Gamma}_{II}$ and $\hat{\omega}_{II}$. The results are presented in Tables 3 (for *ESPO*) and 4 (for P_T). Because the comparison is between size corrected powers, with the nuisance parameters estimated in the same way, any differences are due to the particular form of the test, i.e., the difference of quadratic forms in P_T or the ratio of those same forms in PO_T. For both models and innovation processes, when the parameter is large and negative the power of both the procedures is significantly below that for all other parameter values. In these cases the *ESPO* procedure performs slightly better and consistently over all model configurations.

TABLE 4. Size corrected power of the P_T test; $T = 100$, size $\alpha = 5\%$

	M_1					M_2				
	$\psi =$					$\psi =$				
ρ	−0.8	−0.5	0.0	0.5	0.8	−0.8	−0.5	0.0	0.5	0.8
(a) $MA(1)$ innovations										
0.95	16.6	22.3	27.4	30.1	28.6	7.02	9.78	10.5	10.4	10.8
0.90	27.4	41.1	61.2	67.7	64.2	15.7	19.5	29.6	30.3	29.7
0.85	39.8	55.3	84.8	89.3	88.5	26.5	39.1	56.3	53.9	54.4
0.80	43.9	67.4	93.3	97.5	96.8	32.7	58.8	79.8	79.4	79.2
0.75	46.4	72.1	96.2	99.0	99.3	37.5	65.4	92.2	92.7	91.8
0.70	48.7	77.8	97.4	99.6	99.8	40.8	72.1	96.5	98.0	97.9

	M_1				M_2			
	ϑ				ϑ			
ρ	−0.8	−0.5	0.5	0.8	−0.8	−0.5	0.5	0.8
(b) $AR(1)$ innovations								
0.95	19.9	24.5	25.5	32.4	7.88	8.97	9.86	9.75
0.90	39.7	51.4	60.8	72.8	16.9	21.3	27.7	24.9
0.85	62.2	68.5	83.3	91.5	30.5	42.1	50.5	51.3
0.80	67.4	81.1	92.1	98.1	44.8	68.4	75.3	75.9
0.75	71.1	84.9	95.2	99.7	56.4	82.2	87.1	91.4
0.70	75.6	88.5	97.3	99.9	66.6	92.0	94.2	97.7

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

5. CONCLUSIONS

This paper has provided a feasible method for obtaining approximate critical values for an optimal unit root test. Under Assumption 2 the order of error of the approximation is proved to be $O(T^{-1})$ whereas under Assumption 1, inference based on the feasible estimated saddlepoint approximation is shown to be asymptotically valid. 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 et al. (1996). Numerical comparisons between the estimated saddlepoint procedure and pure asymptotic approximations demonstrate clear superiority for the former, when considered across a range of model configurations, innovation processes, and estimation methods.

Although the performance of the saddlepoint is comparable with that found for bootstrap-based inference they are not, as yet, available with quite the same generality, such as the wild bootstrap procedure of Cavaliere and Taylor (2009). Although beyond the scope of this paper, extending the analysis to bootstrap-based saddlepoint inference, i.e., bootstrapping the asymptotically pivotal estimated saddlepoint *P*-values, for direct 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. *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. *Annals of Statistics* 26, 48–83.

Butler, R.W. & 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. & A.M.R. Taylor (2008) Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory* 24, 43–71.

Cavaliere, G. & A.M.R. Taylor (2009) Heteroskedastic time series with a unit root. *Econometric Theory* 25, 1228–1276.

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

Daniels, H.E. (1987) Tail probability approximations. *International Statistical Review* 55, 37–48.

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

Dufour, J.-M. & 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., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

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

Goutis, C. & G. Casella (1999) Explaining the saddlepoint approximation. *American Statistician* 53, 216–224.

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

Harris, D., D.I. Harvey, S.J. Leybourne, & A.M.R. Taylor (2009) Testing for a unit root in the presence of a possible break in trend. *Econometric Theory* 25, 1545–1588.
Jing, B.-Y. & J. Robinson (1994) Saddlepoint approximations for marginal and conditional probabilities of transformed variables. *Annals of Statistics* 22, 1115–1132.
Juhl, T. & 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. & 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. & H. Neudecker (1999) *Matrix Differential Calculus, with Applications in Statistics and Econometrics*, rev. ed. 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* 25, 1662–1681.
Nabeya, S. & K. Tanaka (1990) Limiting power of unit-root tests in time-series regression. *Journal of Econometrics* 46, 247–271.
Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
Perron, P. & 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. & 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 Faculté d'administration à l'Université de Sherbrooke.

APPENDIX

Proof of Theorem 1. The leading term saddlepoint approximations under Assumption 2 have previously been detailed; see Lieberman (1994, 1996) and Marsh (1998). Here what is relevant, because 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 $\bar{\rho}$,

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

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

where $C_{\bar{\rho}}$ is the singular value decomposition in (5). Standardizing, so that $z = A_{\bar{\rho}}^{-1/2} w_{\bar{\rho}} \sim N(0, I_n)$ and $n = T - k$, then the test statistic can be written as the ratio of quadratic forms in standard normal variables, as in

$$PO_T = \frac{z' A_{\bar{\rho}} z}{z' A_1 z}.$$

As a consequence,

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

where $z_t \sim N(0,1)$ and the nonzero eigenvalues of $B(\rho)$ are also the n eigenvalues of the matrix $A = A_{\bar{\rho}} - \kappa A_1$, $(\lambda_t)_{t=1}^n$, with

$$Ar_t = \lambda_t r_t; \qquad r_t' r_t = 1, \qquad r_t' r_s = 0 \quad \text{if } t \neq s.$$

Let $Q = \sum_{t=1}^n \lambda_t z_t^2$. Then $F^T_{\Gamma,1-\rho}(\kappa)$ can be evaluated via

$$F^T_{\Gamma,1-\rho}(\kappa) = F_Q(0); \qquad F_Q(q) = \Pr[Q < q].$$

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

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

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

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

for some small τ. The distribution is

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

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

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

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

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

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

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

$$\begin{array}{l}\text{(a) } \bar{\varphi}_n(\theta) \quad \text{is analytic on the strip } \Theta = \left\{\theta : \mathrm{Im}(\theta) \leq \pm\left(\tilde{\theta}_{\rho,x} + \varepsilon\right)\right\}, \\ \text{(b) } |\bar{\varphi}_n(\theta)| < \frac{c_0}{|\theta|^{\alpha}} \text{ when } |\theta| > c, \quad \text{and} \\ \text{(c) the derivatives of } R_n(\theta) \text{ are } O(1).\end{array} \quad . \tag{A.2}$$

In (A.2) (a) and (b) ensure that the path of integration in (A.1) can be deformed as in

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

whereas (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

$$\Pr[PO_T < \kappa] = \Pr\left[\frac{z' A_{\bar{\rho}} z}{z' A_1 z} < \kappa\right] = \alpha, \tag{A.3}$$

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

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

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

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

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

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

so that

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

Consequently,

$$\lim_{T\to\infty} \left\|A_{\bar{\rho}}\right\|_2 < m_{\gamma} < \infty. \tag{A.4}$$

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

Moreover, because 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 (A.3) 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

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

Using (A.4), which also holds for $\bar{\rho} = 1$, we find

$$\lim_{T\to\infty} \sup_{t\leq T} |\lambda_t| < (1+\kappa) m_\gamma = O(1).$$

Because the eigenvalues $\{\lambda_t\}_{t=1}^n$ are all finite $\bar{\varphi}_n(\theta)$ 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 (A.2) holds. See also Marsh (1998, Thm. 1(ii)).

To continue, because

$$|\bar{\varphi}_n(\theta)| \leq |1 - 2\theta\lambda_{t^*}|^{-(1/2)}, \qquad t^* = \arg\max_t \sup |1 - 2\theta\lambda_t|^{-(1/2)}$$

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

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

and hence condition (b) in (A.2) also holds in this case. Under these conditions the leading term saddlepoint approximation evaluated at $x = 0$ is

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

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

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

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 of 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 sth 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. ■

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

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 PO_T. To do so we parameterize the covariance matrix as

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

where L is the lag-operator matrix. 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

$$\left(\hat{\Xi}_h(z) \right)^{-1} = \sum_{j=1}^{\infty} \hat{\psi}_{j,T} z^j. \tag{A.5}$$

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

$$\lim_{h_T \to \infty} \sum_{j=1}^{h_T} \left| \hat{\psi}_{j,T} - \psi_j \right| \to 0, \quad \text{almost surely.} \tag{A.6}$$

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

$$\begin{aligned} \left| \tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - F_{\Gamma,1}(\kappa) \right| &= \left| \left(\tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - \tilde{F}_{\Gamma,1}(\kappa) \right) + \left(\tilde{F}_{\Gamma,1}(\kappa) - F_{\Gamma,1}(\kappa) \right) \right| \\ &\leq \left| \left(\tilde{F}_{\hat{\Gamma}_h,1}(\kappa) - \tilde{F}_{\Gamma,1}(\kappa) \right) \right| + \left| \left(\tilde{F}_{\Gamma,1}(\kappa) - F_{\Gamma,1}(\kappa) \right) \right|, \end{aligned} \tag{A.7}$$

via the triangle inequality. Immediately from Corollary 1 the second term in (A.7) 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

$$\tilde{F}_{\hat{\psi}_T,1}(\kappa) - \tilde{F}_{\psi_T,1}(\kappa) = \sum_{j=1}^{h_T} \left(\hat{\psi}_{j,T} - \psi_j \right) \frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_j} - \sum_{h_T+1}^{T} \psi_j \frac{\partial \tilde{F}_{\psi_T^*,1}(\kappa)}{\partial \psi_j}. \tag{A.8}$$

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press

To proceed, from Corollary 1 we constructed the leading term by $\tilde{F}_{\Gamma,1}(\kappa) = \Phi\left(\hat{r}_1(\kappa)\right)$, with $\hat{r}_1(\kappa)$ defined by (12)–(14). Immediately, therefore, $\Phi\left(\hat{r}_1(\kappa)\right)$ is a differentiable function of the eigenvalues $(\lambda_t)_{t=1}^{T}$. Moreover, from Magnus and Neudecker (1999, Thm. 7, p. 158), the eigenvalues are differentiable (with respect to ψ_j), with first derivative

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

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

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

Because $r_t' r_t = 1$, 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

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

Hence from (A.8) 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 (A.6), and

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

Therefore, because 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. ■

https://doi.org/10.1017/S0266466610000629 Published online by Cambridge University Press