Econometric Theory, **23**, 2007, 686–710. Printed in the United States of America.
DOI: 10.1017/S0266466607070296

THE AVAILABLE INFORMATION FOR INVARIANT TESTS OF A UNIT ROOT

PATRICK MARSH
University of York

This paper considers the information available to invariant unit root tests at and near the unit root. Because all invariant tests will be functions of the maximal invariant, the Fisher information in this statistic will be the available information. The main finding of the paper is that the available information for all tests invariant to a linear trend is zero at the unit root. This result applies for any sample size, over a variety of distributions and correlation structures, and is robust to the inclusion of any other deterministic component. In addition, an explicit upper bound upon the power of all invariant unit root tests is shown to depend solely upon the information. This bound is illustrated via a brief simulation study that also examines the impact that different invariance requirements have on power.

1. INTRODUCTION

It is well known that the properties, particularly the power, of unit root tests depend upon the trending characteristics of the deterministic component in the maintained model. Numerical and asymptotic evidence of this is contained in Perron (1989), DeJong, Nankervis, Savin, and Whiteman (1992), Zivot and Andrews (1992), and Leybourne, Mills, and Newbold (1998).

Analysis of this dependence appears in Durlauf and Phillips (1988), Phillips (1998), and Ploberger and Phillips (2002, 2003) with derived asymptotic relationships between nonstationary processes and deterministic trends. The first two papers concern the spurious success of regressions of unit root processes on linear trends and vice versa. The latter two demonstrate that our ability to model nonstationary series is limited by both the number and trending characteristics of deterministic components.

Despite this body of evidence a precise relationship between a unit root process and a simple linear trend has not yet been detailed. This paper provides such a relationship in the context of invariant unit root tests. The available information for such tests is defined to be Fisher information in the maximal invariant, of which all invariant tests must be functions; see Lehmann (1997). For

Thanks are due to Francesco Bravo, Giovanni Forchini, Les Godfrey, Robert Taylor, participants at seminars at the Universities of Birmingham and York and at the ESRC Econometric study group conference, Bristol, 2004, and also to Bruce Hansen, Joel Horowitz, and five anonymous referees. Revisions of this paper have greatly benefited from comments and suggestions from Grant Hillier and Peter Phillips. Address correspondence to Patrick Marsh, Department of Economics, University of York, Heslington, York YO10 5DD; e-mail: pwnm1@york.ac.uk.

 © 2007 Cambridge University Press 0266-4666/07 $15.00

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

the Sargan and Bhargava (1983) formulation it is found that requirement of invariance with respect to a linear trend implies that the available information is zero at the unit root. This result holds over a family of sample distributions, for any additional deterministic components, for any sample size, and for a general class of innovation dependence.

For the simple zero-mean Gaussian AR(1) process it is well known that the standardized observed information converges to a random variable. Moreover, for potentially nonergodic processes we do not have uniform convergence of estimators over the parameter space, nor are asymptotic distributions Gaussian. Consequently, the role of Fisher–Rao efficiency theory, having Fisher information at its heart, is at best ambiguous. Despite this, this paper is able to demonstrate a central role for information in determining the properties of tests.

It is shown that the power of any invariant test will be bounded above by the chosen size plus a simple function of the available information. Consequently, the relationship between the efficiency of a test (power minus size) and information is preserved and applies generally. In addition, as a corollary to the main theorem, it is demonstrated, via an applicable bound fundamental to the Fisher–Rao theory, that the variance of any trend invariant test is indeterminate, under a unit root.

The majority of asymptotic treatments of unit root testing adopt a local-to-unity framework, for example, that of Phillips (1987) and Elliott, Rothenberg, and Stock (1996). Supposing that the autoregressive parameter is in an asymptotic neighborhood of one, say, $1 - O(c_N)$, where $c_N \to 0$ as the sample size $N \to \infty$, then this paper finds that the information is of order $O(N^2)$ in that neighborhood. On the other hand, if c_N is $O(1)$, the model trend is stationary, and information is of order $O(N)$.

The results of this paper help explain two key findings within the literature: that the inclusion of trends has a detrimental effect on the power of tests in general and that this is specifically acute for locally most powerful tests. These findings are neatly illustrated by the numerical work of Dufour and King (1991) and Elliott et al. (1996). Following Rao (1945), Fisher information may be interpreted as a metric on the space of densities. Furthermore, Efron (1975) and Kallenberg (1981) argue that high curvature, which is inversely proportional to information, implies a lack of efficiency for linear methods, such as locally most powerful tests.

There has been much success in terms of characterizing the asymptotic distribution of unit root tests. Despite this success, excepting the simplest zero-mean AR(1) case, as in Abadir (1993), we have no analytic expressions for asymptotic moments, densities, or distributions. Without such expressions, establishing results analogous to those derived here would prove impossible. Consequently, our only knowledge of the influence of trends on, for example, the power of unit root tests comes from Monte Carlo experimentation. This paper provides an explicit link between the inclusion of a trend, its impact upon information, and thence the power of any invariant test.

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

It does not follow from the results that linear trends should be excluded from deterministic components in the unit root context. Instead, this paper highlights the importance of determining the necessity of trends prior to the unit root test being performed. Tests for precisely this purpose have recently been constructed; see Vogelsang (1998) and Bunzel and Vogelsang (2003).

The plan for the rest of the paper is as follows. The next section details the model and the assumption under which the results hold and presents all of the results in a single theorem. Section 3 discusses the implication of these results, and Section 4 concludes. An Appendix contains the proof of the theorem and of a key lemma.

2. ASSUMPTIONS AND MAIN RESULTS

We will consider the formulation of the unit root hypothesis in the time series regression

$$y_t = \beta_1 + \beta_2 t + z_t'\beta_3 + u_t; \qquad u_t = \rho u_{t-1} + \varepsilon_t; \qquad t = 1, \ldots, N. \tag{1}$$

In (1) $\beta = (\beta_1, \beta_2, \beta_3')'$ is a $k \times 1$ vector of unknowns, z_t a $(k-2) \times 1$ vector of strongly exogenous variables, and ε_t a zero-mean, stationary, and ergodic innovation sequence having covariances $E[\varepsilon_t \varepsilon_s] = \sigma^2 \omega(|t-s|)$. Regressions of this type were introduced by Sargan and Bhargava (1983) in the context of testing for a unit root. We shall consider hypothesis tests of

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

and in particular those tests that are, at least asymptotically, invariant with respect to the deterministic and exogenous components in (1) and the marginal variance of ε_t. Consistent, robust testing procedures for (2) in (1) are fully established within the literature; see, for example, Elliott et al. (1996). Asymptotically the resultant tests are invariant, and thus our results will have some bearing upon all of these procedures.

Let $y = (y_1, \ldots, y_N)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$, $e = (1, 1, \ldots, 1)'$, $\tau = (1, 2, \ldots, N)'$, $Z = (z_1, \ldots, z_N)'$, $X = (e : \tau : Z)$ and let $\sigma^2 \Omega = E[\varepsilon \varepsilon']$ be the covariance structure of the stationary innovation sequence. As a consequence (1) may be written as

$$T_\rho (y - X\beta) = \varepsilon, \tag{3}$$

where $T_\rho = I_N - \rho L^{(1)}$ and $L^{(1)}$ is the $N \times N$ matrix lag operator, having ones on the first lower off-diagonal and zeros elsewhere.

We shall proceed under the following assumption on the density of y and the family $\mathcal{F}$ of which it is a member.

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

Assumption 1.

(a) Let the density of y, given Z, be $f(y;\beta,\rho,\sigma^2\Omega|Z) = f(y) \in \mathcal{F}$ with

$$\mathcal{F} = \left\{ f : f(y;\beta,\rho,\sigma^2\Omega) = \frac{q[(y - X\beta)'(\sigma^2\Sigma_\rho(\Omega))^{-1}(y - X\beta)]}{|\sigma^2\Sigma_\rho(\Omega)|^{1/2}} \right\},$$

where X and β are defined following equations (2) and (1), respectively, σ^2 is a scalar, and Ω is an $N \times N$ covariance matrix so that $\Sigma_\rho(\Omega) = T_\rho^{-1}\Omega(T_\rho^{-1})'$. Furthermore, we assume that $q[.]$ is a nonincreasing convex function on $[0,\infty)$.

(b) Ω is such that (1) $\Sigma_\rho(\Omega)$ depends on ρ for all $\rho \in \mathbb{R}$ and (2) $\|\Omega\|_1 = \sup_j \sum_{i=1}^N |\Omega_{i,j}| = \sum_{k=0}^{N-1} |\omega(k)| < M < \infty$, for all N.

It is presumed that the regressor set always includes a constant and a trend. However, the augmented regressor set may include additional polynomial trends or any random (exogenous) regressors to be conditioned upon. Assumptions on the initial condition that are consistent with Assumption 1 are either that it is constant (without loss of generality, zero) or that it is exogenously distributed and so may be conditioned upon and included in the regressor set. To illustrate how a nonzero initial condition, y_0, can be incorporated, suppose there are no other regressors, so that

$$T_\rho(y - \beta_1 e - \beta_2 \tau - \beta_3 \bar{y}_0) = \varepsilon,$$

where $\bar{y}_0 = (y_0, 0, \ldots, 0)'$. That is, in terms of the notation of Assumption 1, $Z = \bar{y}_0$. Thus for the first two values of y,

$$y_1 = \beta_1 + \beta_2 + \beta_3 y_0 + \varepsilon_1,$$

$$y_2 = \rho y_1 + (1 - \rho)\beta_1 + (2 - \rho)\beta_2 - \rho\beta_3 y_0 + \varepsilon_2,$$

and no other y_i depends upon y_0 explicitly. Although inference on β_3 would not be possible in this setup, our inferences on ρ can be made invariant with respect to y_0. Therefore inference upon ρ will not depend upon β_3 in the sense that neither the size nor the power of any suitable test will depend upon this nuisance parameter.

Part (ii)(1) of Assumption 1 ensures that there are no common factors in the covariance structure of the data and so ρ may be identified, in principle. Part (ii)(2) guarantees the "finiteness" of Ω with respect to the absolute column sum norm. It implicitly provides for the ergodicity of the ε_t in that the columns/rows of Ω must be absolutely summable as $N \to \infty$. Notice also that because ε is covariance stationary then Ω is Toeplitz.

Assumption 1 also implies that the data have a distribution within the family of elliptically symmetric distributions. This family includes all finite mixtures of normals, including multivariate t-distributions, with the Cauchy as a special

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

case. That is, the kurtosis of the $\{y_t\}$ may be any constant not depending upon the parameters, although skewed distributions are ruled out.

To proceed define the $N \times N - k$ matrix C by

$$CC' = M_{T_1 X} = I - T_1 X((T_1 X)' T_1 X)^{-1}(T_1 X)'; \qquad C'C = I_{N-k},$$

so that C is the singular value decomposition of $M_{T_1 X}$. Let $w = C' T_1 y$ and consider the group $G = (a, g)$, with $a \in \mathbb{R}$ and $g \in \mathbb{R}^k$ and with action

$$T_1 y \to a T_1 y + T_1 X g;$$

then the maximal invariant, under G, for testing H_0 in (2) is

$$v = \frac{w}{|w|} = \frac{C' T_1 y}{(y' T_1' M_{T_1 X} T_1 y)^{1/2}};$$

see Kariya (1980) and King (1980). Furthermore the density of and Fisher information in the maximal invariant are given in the following lemma, which is proved in the Appendix.

LEMMA 1.

(i) The density of v on the surface of the unit $N - k$ sphere, with respect to normalized Haar measure, is

$$f(v; \rho) = \det A^{-1/2} (v' A^{-1} v)^{-(N-k)/2}, \tag{4}$$

where $A = C' T_1 \Sigma_\rho(\Omega) T_1' C$.

(ii) The Fisher information in v about ρ is

$$I_v(\rho) = \frac{(N-k)\mathrm{Tr}[(A^{-1}\bar{A})^2] - [\mathrm{Tr}(A^{-1}\bar{A})]^2}{2(N-k+2)}, \tag{5}$$

where A is defined following (4), $\bar{A} = C' T_1 [d_\rho \Sigma_\rho(\Omega)] T_1' C$, and $d_\rho \Sigma_\rho(\Omega)$ denotes the derivative of $\Sigma_\rho(\Omega)$ with respect to ρ.

To summarize: the density of v depends upon neither β nor σ^2, nor, given Assumption 1, the distribution of the data. Thus the expression for $I_v(\rho)$ does not depend on which member of the family $\mathcal{F}$ generates y. Generally, though, it will depend upon the structure of $\sigma^2 \Omega$. We will thus refer to (5) as the available information for invariant tests, in the sense that any invariant test, or indeed any other function of the maximal invariant, will have Fisher information bounded above by this quantity.

On the basis of Lemma 1, we are able to demonstrate the following facts about information at and near the unit root. First, information vanishes at the point of unity; that is, the available information is zero at the unit root. For any

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

other point information is nonzero and in any asymptotic neighborhood will grow as the square of the sample size. Asymptotic neighborhoods are of specific interest for unit root tests, and so we prove that the asymptotic power (minus size) of any invariant unit root test can be bounded above by a simple function of the available information and the radius of the asymptotic neighborhood. These results are contained in the following theorem, which is proved in the Appendix.

THEOREM 1.

(i) For all sample sizes N, for all sets of exogenous or fixed regressors Z, and for all Ω *that satisfy Assumption 1(b), the Fisher information in v is zero at* $\rho = 1$*; that is,*

$$I_v(1)=0.$$

(ii) Define any asymptotic neighborhood of the unit root by $1 - \rho = O(c_N) > 0$*, with* $c_N \to 0$ *as* $N \to \infty$*; then in that neighborhood,*

$$I_v(1 - O(c_N)) = O(N^2).$$

(iii) Let $P_{\mathbb{V}}(\rho)$ *be the power of any invariant unit root test, of size* δ*, against the alternative with* $1 - \rho = c_N$*; then for some* $c^* \in (0, c_N)$*,*

$$P_{\mathbb{V}}(\rho) \leq \delta + \frac{1}{2} c_N^2 I_v(1 - c^*).$$

3. DISCUSSION AND ANALYSIS

Theorem 1 demonstrates that the available information is zero at a unit root but generally nonzero in any asymptotic neighborhood. Moreover, an explicit relationship between the power of invariant tests is derived. In this section implications of the results will be discussed, in particular within the context of the inferential problem of testing for a unit root.

3.1. Available Information Is Zero at a Unit Root

The strongest result contained in Theorem 1 is that the information vanishes in model (3) at the unit root whenever a trend is included as a regressor. This result holds for all other sets of variables z_t, including higher order polynomial trends, and for any Ω defined for a stationary error sequence. The crucial relationship that leads to all of the results in this paper is that the derivative of the autoregressive covariance matrix is matrix colinear with the outer product of the trend, namely,

$$\left.\frac{d\Sigma_\rho}{d\rho}\right|_{\rho=1} = \tau\tau' - \Sigma(1), \tag{6}$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

where $\Sigma_\rho = T_\rho^{-1}(T_\rho^{-1})'$. Consequently, any projection of the derivative in the space orthogonal to a trend yields a covariance derivative at unity that is equal to minus the covariance. The properties of invariant unit root tests when a linear trend is included follow as a consequence of this relationship.

Formally this result holds only for the inclusion of a linear trend. However, it may be inferred that "trendlike" regressors will have a negative impact upon available information. This follows because the projection orthogonal to such regressors will be "close" to being orthogonal to τ and thus the quantity $A^{-1}\bar{A}$ in (5) evaluated at the unit root will be "close" to the identity. Examples would be inclusion of structural breaks in the constant, periodic functions with long periods, or the inclusion of strongly trending exogenous variables.

Part (i) of Theorem 1 also points to another valuable prediction. The statistical curvature obtained by Efron (1975) is inversely proportional to the information, whereas Kallenberg (1981) argues that large curvature implies a shortcoming of locally most powerful tests. In the extreme case of the inclusion of τ as a regressor, curvature will be unbounded. Therefore, we would expect inclusion of trending regressors to have a significant negative impact upon the power of any locally most powerful testing procedures; indeed, see the numerical work contained in Forchini and Marsh (2000).

Later in this section we will discuss the information in an asymptotic neighborhood of one. Before getting there though it is worth pointing out that because $I_v(\rho)$ is differentiable in ρ (in fact it is analytic in ρ, because $\Sigma(\rho)$ is infinitely differentiable for every ρ), then we can find arbitrarily small and positive κ_1 and κ_2 so that

$$\rho > 1 - \kappa_1 \to I_v(\rho) < \kappa_2.$$

That is, information is vanishingly small in nonasymptotic neighborhoods of unity. In asymptotic neighborhoods the story is quite different, as is apparent from part (ii) of the theorem.

Suppose that we take the density in (4) as a likelihood for ρ and with a score given by

$$S_v(\rho) = -\frac{\mathrm{Tr}[A^{-1}\bar{A}]}{2} + \frac{N-k}{2}\,\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}.$$

Then because $\mathrm{Var}[S_v(\rho)] = I_v(\rho)$ it follows that $S_v(1) = 0$ for all v. That is, the score is constant at the unit root. The fundamental implication of the score being constant is that the maximum likelihood estimator for ρ based upon (4) will not converge uniformly. However, because there is not uniform convergence in autoregression in any case, the additional results of Theorem 1 are needed to explore the consequences of this.

The fact that information is zero is preserved in two interesting special cases. First it is well known in the seasonal unit root literature that the trend $(-1)^t t$ is the analogue to the linear trend in the positive unit root case. If we call the

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

maximal invariant for this case $\tilde{v}$, then one can establish that $I_{\tilde{v}}(-1) = 0$, also. In addition, suppose that we quasi-difference the data, that is, take $T_{\rho_0} y$, where $T_{\rho_0} = I - \rho_0 L$, $\rho_0 \neq 1$. Again calling the maximal invariant for this case $v^{(\rho_0)}$, then also $I_{v^{(\rho_0)}}(1) = 0$. It therefore seems that it is only the requirement of invariance to a trend at the unit root that matters, not precisely how this is achieved.

3.2. Detrending and Differencing

That information is zero is a consequence of the algebraic identity given in (6). Related to that identity is a heuristic explanation as to why it is that the precise combination of a linear trend and unit root implies a complete loss of information.

There are two methods for trend removal: either one can project the data onto the space orthogonal to that trend, or one can difference the data. Suppose in a basic model with a constant and trend that we difference the data d times and obtain

$$(T_1)^d y = (T_1)^d(\beta_1 e + \beta_2 \tau) + (T_1)^d u; \tag{7}$$

then for any d a maximal invariant is defined by

$$v^{(d)} = \frac{C'_{(d)}(T_1)^d y}{(y'(T_1')^d M_{(d)}(T_1)^d y)^{1/2}},$$

$$C_{(d)} C'_{(d)} = M_{(d)} = I - (T_1)^d X(((T_1)^d X)'(T_1)^d X)^{-1}((T_1)^d X)'.$$

However, notice that in (7) all we have really achieved is to transform the variance of y, under the null, to

$$\mathrm{Var}[y|H_0] = (T_1)^{d-1}(T_1')^{d-1} = (T_1)^d \Sigma(1)(T_1')^d,$$

and the identity in (6) then becomes

$$(T_1)^d \left.\frac{d\Sigma_\rho}{d\rho}\right|_{\rho=1} (T_1')^d = (T_1)^d \tau\tau'(T_1')^d - (T_1)^d \Sigma(1)(T_1')^d. \tag{8}$$

Furthermore, running through the proof of part (i) of Theorem 1 in the Appendix we find that an identical argument applies. That is, the common orthogonality of $C'_{(d)}$ to both $(T_1)^d\tau$ and $(T_1)^d(d\Sigma_\rho/d\rho)|_{\rho=1}$ gives exactly the same result. Thus, at the unit root, there is a synonymity between projecting out a trend and differencing it out. Moreover, this applies for differencing an arbitrary number of times. It is well known in the standard time series literature that differencing, particularly overdifferencing, destroys signal, or here, information. To illustrate this point Figure 1 plots $I_{v^{(d)}}(\rho)$, in a neighborhood of $\rho = 1$, for values of $d = 1,3,6$ for a sample size of $N = 50$. Clearly we lose

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

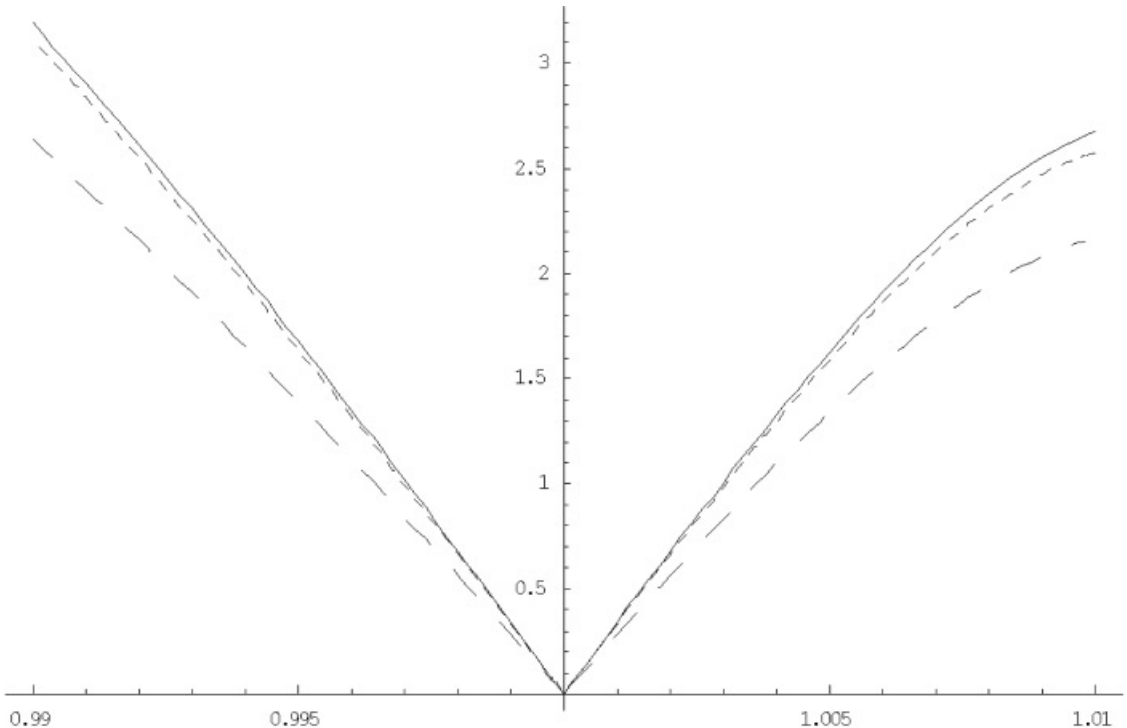


FIGURE 1. Plots of $I_{v^{(d)}}(\rho)$ based on model (7) for $N = 50$ and different d. Solid line: $d = 1$; dotted line: $d = 3$; and dashed line: $d = 6$.

information by differencing; ultimately as $d \to \infty$ we would expect to lose all information. At the unit root, via (8), projecting out the trend is equivalent to differencing to this limit.

3.3. Information and Fisher–Rao Efficiency

The general validity of the efficiency theory of Fisher–Rao is lost for problems in which there is not uniform convergence over the parameter space leading to non-Gaussian asymptotics. Indeed, for the pure random walk it is well known that the standardized observed information is random, although still having mean equal to the standardized expected information.

Despite that, absolute efficiency bounds may still be derived in the context of the current framework. Doing so yields an interesting corollary to Theorem 1. Let $t(v)$ be any invariant statistic having finite mean τ whose derivative is also finite, that is,

$$\tau = \tau(\rho) = E[t(v)]; \qquad \left| \frac{d\tau}{d\rho} \right| < \infty.$$

Now construct a statistic

$$t^*(a) = t(v) - aS_v(\rho),$$

where $S_v(\rho)$ is the score. Minimizing the variance of $t^*(a)$ with respect to a (the minimum occurs at $a = (d\tau/d\rho)' I_v^{-1}(\rho)$) and noting that by definition $\mathrm{Var}[t^*(a)] \geq 0$, it must be that

$$\mathrm{Var}[t(v)] \geq \frac{d\tau'}{d\rho} I_v^{-1}(\rho) \frac{d\tau}{d\rho}.$$

That is, the variance of any statistic $t(v)$ has a lower bound involving Fisher information. For the unit root we obtain the following corollary.

COROLLARY 1. *Let $t(v)$ be any trend invariant unit root test or estimator having finite mean $\tau(\rho)$ with finite first derivative $\tau'(\rho)$ in the autoregressive parameter ρ; then the variance of $t(v)$ is indeterminate in the limit $\rho \to 1$.*

It should, however, be noted that the indeterminacy of the variance does not imply, if $t(v)$ is a test statistic, any loss of consistency. This is because, under the null, critical values can be defined as usual, whereas power is determined under the alternative, when by definition $\rho \neq 1$.

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

3.4. Characterization of Information at Unity

The "availability" of information is determined by the invariances that we demand our statistics to obey. To illustrate, suppose that we take as a baseline model

$$T_\rho(y - \beta_1 e - \beta_2 \tau) = \varepsilon; \qquad E(\varepsilon\varepsilon') = \sigma^2 I. \tag{9}$$

In (9) there are three nuisance parameters, β_1, β_2, and σ^2. Theorem 1 deals with the information available for tests that are invariant with respect to a constant, trend, and length. The maximal invariant is derived via transformations of the data, as in

$$y \to \begin{pmatrix} \hat{\beta} = ((T_1 X)' T_1 X)^{-1} (T_1 X)' T_1 y \\ w = C' T_1 y \end{pmatrix} \quad \text{and} \quad w \to \begin{pmatrix} s^2 = w'w \\ v = w/|w| \end{pmatrix}. \tag{10}$$

Within the context of (9) more restrictive models can be characterized by our knowing the values of the parameters. For example, if we knew all three values then y itself (or $T_1 y$) would be the maximal invariant (in the trivial sense), and the information in $T_1 y$ is well known to be

$$I_{T_1 y}(\rho) = I_y(\rho) = E\left[\sum_{i=2}^{N} y_{i-1}^2\right] = \sum_{j=1}^{N-1} \sum_{i=1}^{j} \rho^{2i}.$$

On the other hand, for example, as in Müller and Elliott (2003), we could assume that the variance of ε in (1) is known (see their Condition 1). If length invariance is not required, only the first transformation in (10) is necessary, giving a maximal invariant $w = C' T_1 y \sim \mathcal{F}(C' \Sigma_\rho(\Omega) C)$. Fisher information in w is readily found to be

$$I_w(\rho) = \frac{\mathrm{Tr}[(A^{-1}\bar{A})^2]}{2}.$$

Thus for three cases in (9) we can evaluate the available information at unity, as shown in Table 1. Thus what we are able to assume is known in the model has a significant quantitative effect on the rate at which information accrues at unity.

Table 1. Required invariances in equation (9) and the information at unity

Known parameters	Maximal invariant	Information at unity
β_1, β_2 and σ^2	$T_1 y$	$I_{T_1 y}(1) = N(N-1)/2 = O(N^2)$
σ^2	$w = C' T_1 y$	$I_w(1) = (N-k)/2 = O(N)$
None	$v = w/\|w\|$	$I_v(1) = 0$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

3.5. Characterization of Information over (Non)stationarity Regimes

Because the majority of distributional approximations for unit root tests are derived via local-to-unity asymptotics (e.g., Phillips, 1987), part (ii) of Theorem 1 also gives the properties of Fisher information in any asymptotic neighborhood of unity. Specifically, for model (3) we find that for $1 - \rho = O(c_N)$, where $c_N \to 0$ as $N \to \infty$, $I_v(\rho)$ is $O(N^2)$. Asymptotic distributional approximations are most conveniently made in the regime in which $c_N = c/N$, for constant c, as in the work of Elliott et al. (1996), who derive "optimal" tests for $c = 0$ against $c > 0$.

However, other neighborhoods are of interest, for example, if $c_N = c/\sqrt{N}$, which corresponds with the limiting cases of both Chan and Wei (1987) and Phillips (1987). In this neighborhood, at least in the simplest autoregression, asymptotic normality of estimators may obtain, rather than the Weiner process driven asymptotics of the former case; see Phillips and Magdalinos (2007). Theorem 1 demonstrates that for any asymptotic neighborhood of unity the rate of increase in available information is $O(N^2)$. Information is zero only at the point $\rho = 1$.

From Theorem 1 we can asymptotically quantify the available information in the three regimes of interest. Letting the autoregressive parameter be $\rho = 1 - O(c_N)$, then as $N \to \infty$ we can characterize the unit root regime as $c_N = 0$, any local-to-unity regime as $c_N = o(1)$, and the stationary regime as $c_N = O(1)$. For the stationary case, consider the transformations in (10), so that we can write the density of y as

$$f(y;\rho,\beta,\sigma^2) \propto f(\hat{\beta},w;\rho,\beta,\sigma^2) = f(\hat{\beta}|s^2,v;\rho,\beta,\sigma^2)f(s^2|v;\rho,\beta,\sigma^2)f(v;\rho);$$

hence the information in y may be partitioned as

$$I_y(\rho) = I_{\hat{\beta}|s^2,v}(\rho) + I_{s^2|v}(\rho) + I_v(\rho),$$

where

$$I_{\hat{\beta}|s^2,v}(\rho) = -E\left[\frac{d^2 \ln(f(\hat{\beta}|s^2,v;\rho,\beta,\sigma^2)}{d\rho^2}\right],$$

$$I_{s^2|v}(\rho) = -E\left[\frac{d^2 \ln(f(s^2|v;\rho,\beta,\sigma^2)}{d\rho^2}\right].$$

Because information is nonnegative we have $I_v(\rho) \leq I_y(\rho) = O(N)$ for $|\rho| < 1$.

To summarize, suppose that the autoregressive parameter is $\rho = 1 - O(c_N) \in (-1,1]$; then in these regimes and as $N \to \infty$,

if $c_N = 0$ then $I_v(\rho) = 0$,

if $c_N = o(1)$ then $I_v(\rho) = O(N^2)$,

if $c_N = O(1)$ then $I_v(\rho) = O(N)$.

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

3.6. Information and Power

The direct effect on power of information being zero can be explored via a simple example. Suppose that we have a sample of size N on $x_i \sim iidN(\mu^k, 1)$, $k \geq 1$ and known. In this case it is trivial to establish that Fisher information is zero at $\mu = 0$ and that the maximum likelihood estimator (the global maximum of the likelihood) is $\hat{\mu} = \bar{x}^{1/k}$. The immediate implication is that uniform convergence is lost, in that

$$(\hat{\mu} - \mu) = \begin{Bmatrix} O_p(n^{-1/2}) & \text{if } \mu \neq 0 \\ O_p(n^{-1/k}) & \text{if } \mu = 0 \end{Bmatrix},$$

which in turn affects the rate of asymptotic power because all tests have to be a function of $\hat{\mu}$, the sufficient statistic in this case.

The consequences for the unit root model are not as obvious. However, part (iii) of Theorem 1 does provide an alternative link between power and information. As was seen in Section 3.4, there is an intimate relationship between what we are willing to assume as known in the model, and hence the required invariances, and information at unity. We can also, therefore, explore the impact that the invariance requirements have on the power envelope for the unit root. For model (9) we can characterize six different cases. For the purposes of experiment these can be represented by data $(y_t)_1^N$ generated by

$$\begin{aligned} M_1 &: y_t = u_t, \\ M_2 &: (y_t - \beta_1) = u_t, \\ M_3 &: (y_t - \beta_1 - \beta_2 t) = u_t, \end{aligned} \tag{11}$$

where $u_t = \rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iidN(0, \sigma^2)$ and for each model σ^2 will either be presumed known ($\sigma^2 = 1$) or unknown. For each case both the maximal invariant and the point optimal (PO) test can easily be derived. If we also define a matrix C_1, so that

$$C_1 C_1' = I - \frac{T_1 ee' T_1'}{e' T_1' T_1 e},$$

then for each case the results shown in Table 2 pertain.

In each case the PO test rejects for observed outcomes smaller than some pre-specified constant, chosen so that the size is fixed at $\delta = 0.05$. The power of all of these PO tests was simulated via experiments analogous to those giving the power envelopes in Elliott et al. (1996). Put $\rho = 1 - c/N$, fix the sample size at $N = 500$, and on the basis of 100,000 replications the critical values are calculated and then from those, the power under alternatives parameterized by $c = 1, 2, 3, \ldots, 25$.

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

TABLE 2. Maximal invariants and point optimal tests for different configurations of the models in equation (11)

Model	Maximal invariant	Point optimal (PO) test
M_1		
σ^2 known	$w_{(1)} = T_1 y$	$w'_{(1)}(I - \Sigma(\rho))w_{(1)}$
σ^2 unknown	$v_{(1)} = T_1 y/\lvert T_1 y\rvert$	$v'_{(1)}\Sigma(\rho)^{-1}v_{(1)}$
M_2		
σ^2 known	$w_{(2)} = C_1' T_1 y$	$w'_{(2)}(I - C_1' T_1 \Sigma(\rho) T_1 C_1)w_{(2)}$
σ^2 unknown	$v_{(2)} = w_{(2)}/\lvert w_{(2)}\rvert$	$v'_{(2)}(C_1' T_1 \Sigma(\rho) T_1 C_1)^{-1}v_{(2)}$
M_3		
σ^2 known	$w_{(3)} = C' T_1 y$	$w'_{(3)}(I - C' T_1 \Sigma(\rho) T_1 C)w_{(3)}$
σ^2 unknown	$v_{(3)} = w_{(3)}/\lvert w_{(3)}\rvert$	$v'_{(3)}(C' T_1 \Sigma(\rho) T_1 C)^{-1}v_{(3)}$

The most restrictive case (M_1, σ^2 known) provides an absolute power upper bound, in that the Neyman–Pearson lemma applies in the strictest sense. All other cases can be compared to this benchmark. Of interest are the values of c at which the power envelope is 0.5, the point at which Elliott et al. (1996) define their nearly efficient test. For the most restrictive case this is approximately 6.5, whereas for the most general it is 13.5 (M_3, σ^2 unknown). Collectively the results are really self-explanatory. Of note though are two things. Near the null hypothesis the effect of requiring both trend and length invariance is dramatic, implying powers close to just 20% of benchmark. Additionally, the assumption that σ^2 is known is not quite as benign as perhaps suggested in Müller and Elliott (2003). Although what is clear is that what is most crucial is the trend, because then the assumption about σ^2 has the greatest effect, it is also clear that a consistent procedure for determining the necessity of a trend, such as in Bunzel and Vogelsang (2003), may have a significant bearing on our subsequent ability to correctly detect a unit root.

Contained in Table 3 are illustrations of the power bound of part (iii) of Theorem 1. In general the bound on the power of any invariant test for (2) in the context of the models M_i (with variance unknown) of (11) is

$$P_{\mathbb{V}}(\rho) \leq P_i(\rho^*) = \delta + \frac{1}{2}\{(1-\rho)E[\lvert S_{v_{(i)}}(1)\rvert] + (1-\rho)^2 I_{v_{(i)}}(\rho^*)\}. \tag{12}$$

Notice that the bound explicitly tightens for M_3 in that the first term vanishes as a result of $S_{v_{(3)}}(1) = 0$, as indicated by part (iii) of Theorem 1. An alternative bound is obtained by applying

$$E[\lvert S_{v_{(i)}}(1)\rvert] \leq E[(S_{v_{(i)}}(1))^2]^{1/2} = \sqrt{I_{v_{(i)}}(1)}$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

TABLE 3. Powers and bounds of the point optimal tests given in Table 2, for the models in (11) with cases $A(\sigma^2$ known) and $B(\sigma^2$ unknown)

c	M_1^A	M_1^B	$P_1(c)$	M_2^A	M_2^B	$P_2(c)$	M_3^A	M_3^B	$P_3(c)$
1	0.083	0.079	0.189	0.061	0.060	0.186	0.058	0.050	0.057
2	0.130	0.117	0.413	0.070	0.061	0.404	0.066	0.052	0.109
3	0.197	0.193	0.651	0.095	0.079	0.627	0.087	0.061	0.216
5	0.346	0.328	0.894	0.144	0.122	0.849	0.099	0.072	0.363
7	0.538	0.524	1.13	0.196	0.176	1.07	0.146	0.104	0.536
10	0.793	0.783	1.64	0.353	0.248	1.49	0.235	0.157	0.726
15	0.964	0.964	2.13	0.671	0.624	1.91	0.473	0.399	1.13
20	0.997	0.996	2.64	0.870	0.867	2.32	0.700	0.629	1.34
25	1.00	1.00	3.14	0.972	0.972	2.74	0.877	0.839	1.56

Note: The power bounds $P_i(c)$ apply for each case B.

to (12). This gives a slightly weaker bound, depending only on information, but that applies in cases where the score is not identically zero.

These results are very important from a theoretical view, being the first explicit links between power and information. Because for nonergodic models the role of Fisher–Rao theory is at best ambiguous, this result retains a role for information. Practically its usefulness is less obvious. It is an existence result. Although numerically we can always find a particular ρ^* such that the bound is as tight as our numerical accuracy permits, this has to be done pointwise. For a practical bound put $\rho^* = 1 - c^*/N$, and the bounds in (12) become

$$P_i(c^*) = \delta + \frac{1}{2}\{cE[|S_{v_{(i)}}(1)|] + c^2 I_{v_{(i)}}(1 - c^*/N)\}; \qquad i = 1, 2,$$

$$P_3(c^*) = \delta + \frac{c^2 I_{v_{(3)}}(1 - c^*/N)}{2}.$$

Moreover, because $I_{v_{(3)}}(1) = 0$, then for sufficiently large N we must have $I_{v_{(3)}}(1 - c^*/N) \leq I_{v_{(3)}}(1 - c/N)$ if $c^* \leq c$. Consequently, we can utilize the following locally uniform bound for M_3:

$$P_{\mathbb{V}}(\rho) \leq P_3(c^*) \leq P_3(c).$$

In addition to the power envelopes Table 3 also contains the appropriate values for $P_3(c)$ and also, although local uniformity may not apply, for $P_1(c)$ and $P_2(c)$, for which the term $E[|S_{v_{(i)}}(1)|]$ was approximated via the simulations described previously. Numerically the importance of the vanishing of the score is highlighted, although because at each value of c it is possible to improve the bound arbitrarily, no further importance can be attached. Instead, if we seek to

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

measure efficiency in terms of how tests are able to discriminate between the null and alternative given particular mean functions, a more practical measure would be to use an explicit measure of distance, such as that proposed in Marsh (2005).

4. CONCLUSIONS

This paper has proved that inclusion of a trend in the Sargan and Bhargava (1983) formulation implies zero available information at the unit root. The explicit bound on power by information implies that low power is an inevitable consequence of trend invariance. This precise relationship between the inclusion of a trend and information at the unit root and also the explicit bound on the power of invariant unit root tests add significantly to our understanding of this problem.

Specifically, the relationship between unit root and trend is captured algebraically by (6). That is, the consequence of requiring invariance to a trend produces this identity, which in turn implies that Fisher information is zero. This identity is essentially unchanged if we difference the data any number of times. Consequently, at the unit root, there is a complete equivalence between differencing and projecting orthogonally to a trend.

REFERENCES

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

Bunzel, H. & T.J. Vogelsang (2003) Powerful Trend Function Tests That Are Robust to Strong Serial Correlation with an Application to the Prebisch-Singer Hypothesis. Mimeo, Cornell University.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

DeJong, D.N., J.C. Nankervis, N.E. Savin, & C.H. Whiteman (1992) Integration versus trend stationarity in time series. *Econometrica* 60, 423–433.

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.

Durlauf, S.N. & P.C.B. Phillips (1988) Trends versus random walks in time series analysis. *Econometrica* 56, 1333–1354.

Efron, B. (1975) Defining the curvature of a statistical problem (with applications to second order efficiency). *Annals of Statistics* 3, 1189–1242.

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

Forchini, G. & P.W. Marsh (2000) Exact Inference for the Unit Root Hypothesis. Discussion Paper in Economics 00/54, University of York.

Kallenberg, W.C.M. (1981) The shortcoming of locally most powerful tests in curved exponential families. *Annals of Statistics* 9, 673–677.

Kariya, T. (1980) Locally robust tests for serial correlation in least squares regression. *Annals of Statistics* 8, 1065–1070.

King, M.L. (1980) Robust tests for spherical symmetry and their application to least squares regression. *Annals of Statistics* 8, 1265–1271.

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

Lehmann, E.L. (1997) *Testing Statistical Hypotheses*, 2nd ed., Springer Texts in Statistics. Springer-Verlag.
Leybourne, S.J., T.C. Mills, & P. Newbold (1998) Spurious rejections by Dickey–Fuller tests in the presence of a break under the null. *Journal of Econometrics* 87, 191–203.
Marsh, P.W.N. (2005) A Measure of Distance for the Unit Root Hypothesis. Discussion Paper in Economics 05/03, University of York.
Müller, U.K. & G. Elliott (2003) Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.
Perron, P. (1989) The Great Crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361–1401 (Erratum, 61, 248–249).
Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
Phillips, P.C.B. (1998) New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.
Phillips, P.C.B. & T. Magdalinos (2007) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
Ploberger, W. & P.C.B. Phillips (2002) Rissanen's theorem and econometric time series. In A. Zellner, H.A. Keuzenkamp, & M. McAleer (eds.), *Simplicity, Inference and Modelling*, pp. 165–180. Cambridge University Press.
Ploberger, W. & P.C.B. Phillips (2003) Empirical limits for time series econometric models. *Econometrica* 71, 627–673.
Rao, C.R. (1945) Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* 37, 81–89.
Sargan, J.D. & A. Bhargava (1983) Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 51, 153–174.
Vogelsang, T.J. (1998) Trend function hypothesis testing in the presence of serial correlation parameters. *Econometrica* 65, 123–148.
Würtz, A. (1997) A Universal Upper Bound on Power Functions. Discussion Paper 97/17, University of New South Wales.
Zivot, E. & D.W.K. Andrews (1992) Further evidence on the great crash, the oil price shock, and the unit root hypothesis. *Journal of Business & Economic Statistics* 10, 251–270.

APPENDIX

Proof of Lemma 1. Part (i) follows immediately from the results of King (1980). For part (ii), we have that the log-likelihood is

$$L_v(\rho) = -\frac{1}{2}\ln\det A - \frac{N-k}{2}\ln(v'A^{-1}v),$$

and hence the score is

$$S_v(\rho) = \frac{dL_v(\rho)}{d\rho} = -\frac{\mathrm{Tr}[A^{-1}\bar{A}]}{2} + \frac{N-k}{2}\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}, \tag{A.1}$$

and so information is obtained as

$$I_v(\rho) = E[(S_v(\rho))^2].$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

By definition, $E[S_v(\rho)] = 0$ immediately implies

$$E\left[\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}\right] = \frac{\mathrm{Tr}[A^{-1}\bar{A}]}{N-k}, \tag{A.2}$$

so that the only other expectation to be calculated is

$$E\left[\left(\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}\right)^2\right].$$

Noting that we can write $v = w/|w|$, where w is spherically symmetric with mean zero and $\mathrm{Var}[w] = A$, then

$$E\left[\left(\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}\right)^2\right] = E\left[\left(\frac{w'A^{-1}\bar{A}A^{-1}w}{w'A^{-1}w}\right)^2\right] = E[r_z^2],$$

where

$$r_z = \frac{z'A^{-1/2}\bar{A}A^{-1/2}z}{z'z}$$

and $z = A^{-1/2}w$. Further, because r_z is independent of $|z| = \sqrt{z'z}$, we have

$$E[|z|^4 r_z^2] = E[(z'z)^2]E[r_z^2],$$

and hence

$$E[r_z^2] = \frac{E[(z'A^{-1/2}\bar{A}A^{-1/2}z)^2]}{E[(z'z)^2]} = \frac{\mathrm{Var}[(z'A^{-1/2}\bar{A}A^{-1/2}z)] + E[(z'A^{-1/2}\bar{A}A^{-1/2}z)]^2}{E[(z'z)^2]}$$

$$= \frac{2\mathrm{Tr}[(A^{-1}\bar{A})^2] + [\mathrm{Tr}(A^{-1}\bar{A})]^2}{2(N-k) + (N-k)^2}. \tag{A.3}$$

The expression for the information follows from squaring (A.1), utilizing the expectations given in (A.2) and (A.3), and rearranging. ■

Proof of Theorem 1. For part (i), from Lemma 1, Fisher information in v at $\rho = 1$ is given by

$$I_v(1) = \frac{(N-k)\mathrm{Tr}[(A_1^{-1}\bar{A}_1)^2] - [\mathrm{Tr}(A_1^{-1}\bar{A}_1)]^2}{2(N-k+2)}, \tag{A.4}$$

where

$$A_1 = C'T_1\Sigma_1(\Omega)T_1'C \quad \text{and} \quad \bar{A}_1 = C'T_1\left.\frac{\partial\Sigma_\rho(\Omega)}{\partial\rho}\right|_{\rho=1}T_1'C,$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

and the derivative of the covariance matrix is

$$\frac{\partial \Sigma_\rho(\Omega)}{\partial \rho} = T_\rho^{-1} L^{(1)} T_\rho^{-1} \Omega (T_\rho^{-1})' + T_\rho^{-1} \Omega (T_\rho^{-1})' (L^{(1)})' (T_\rho^{-1})',$$

so that

$$\bar{A}_1 = C'(L^{(1)} S \Omega + \Omega S' (L^{(1)})') C,$$

where $S = T_1^{-1}$, is the lower-triangular partial summation matrix. Consider the derivative of the matrix $\Sigma_\rho = \Sigma_\rho(I_N)$,

$$d\Sigma_\rho = \frac{d\Sigma_\rho}{d\rho} = T_\rho^{-1} L^{(1)} T_\rho^{-1} (T_\rho^{-1})' + T_\rho^{-1} (T_\rho^{-1})' (L^{(1)})' (T_\rho^{-1})',$$

which at the unit root evaluates to

$$d\Sigma_1 = S L^{(1)} S S' + S S' (L^{(1)})' S'. \tag{A.5}$$

From the identity

$$S + S' - I_N = (T_1 \tau)(T_1 \tau)', \tag{A.6}$$

where $\tau = (1, 2, 3, \ldots, N)'$ and substituting $L^{(1)} = I_N - T_1$ into (A.5), we obtain

$$d\Sigma_1 = \tau \tau' - S S',$$

which forms the fundamental relationship between the unit root covariance and the time trend.

Because Ω is Toeplitz and symmetric we can write

$$\Omega = \sum_{k=0}^{N-1} \omega(k) F_k \quad \text{and} \quad F_k = (L_k + L_k'),$$

where $L_k = (L^{(1)})^k$, and put $L_0 + L_0' = I$, so that

$$\bar{A}_1 = \sum_{k=0}^{N-1} \omega(k) C' (L^{(1)} S F_k + F_k S' (L^{(1)})') C.$$

Furthermore, $L^{(1)} S = (S - I_N) = S L^{(1)}$, and hence

$$\bar{A}_1 = \sum_{k=0}^{N-1} \omega(k) C' ((S - I_N) F_k + F_k (S' - I_N)) C$$

$$= \sum_{k=0}^{N-1} \omega(k) C' ((S - I_N) F_k + F_k S') C - C' \Omega C. \tag{A.7}$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

The last term in (A.7) is just $-A_1$, and the quadratic forms in C' can be rewritten, as in

$$\bar{A}_1 = \sum_{k=0}^{N-1} \omega(k) C'((S - I_N)L_k + (S - I_N)L_k' + L_k S' + L_k' S')C - A_1,$$

$$= \sum_{k=0}^{N-1} \omega(k)\{C'L_k(S + S' - I_N)C + C'(S + S' - I_N)L_k' C\} - A_1,$$

where we have exploited $(S - I_N)L_k = L_k(S - I_N)$. Thus using (A.6), we have

$$\bar{A}_1 = \sum_{k=0}^{N-1} \omega(k)\{C'L_k((T_1\tau)(T_1\tau)')C + C'((T_1\tau)(T_1\tau)')L_k' C\} - A_1,$$

and finally because $(T_1\tau)'C = 0 = C'(T_1\tau)$, as X includes a trend, then $\bar{A}_1 = -A_1$, and substitution into (A.4) gives, as required, $I_v(1) = 0$.

For part (ii), we require an expansion of the matrices A^{-1} and $\bar{A}$ appearing in (5). To proceed consider the matrix defined by

$$\{T_\rho^{-1}\}_{i,j} = \begin{Bmatrix} \rho^{|i-j|}: & \text{if } i \geq j \\ 0: & \text{otherwise} \end{Bmatrix}.$$

Letting $\rho = e^{-c_N}$, so that $1 - \rho = O(c_N)$ with $c_N \to 0$, and putting $c_k^* \in (0, c_N)$ for $k = 1, 2, \ldots, N-1$, then each lower diagonal of $T_c^{-1} = T_\rho^{-1}$ has a mean value expansion, so that

$$\begin{aligned} T_c^{-1} &= S - c_N S^*, \\ S^* &= \sum_{k=1}^{N-1} L_k(ke^{-kc_k^*}), \end{aligned} \tag{A.8}$$

where L_k is defined in the proof of part (i).

Application of the expansion given in (A.8) yields

$$A = A_1 - c_N A_2,$$

$$\bar{A} = -A_1 - c_N Q_1,$$

where

$$A_2 = C'T_1(S\Omega S^{*\prime} + S^*\Omega S' + c_N S^*\Omega S^{*\prime})T_1' C = A^* + c_N A_3,$$

$$\begin{aligned} Q_1 &= C'T_1(3(W_1 + W_1') - 3c_N(W_2 + W_2') + (c_N)^2(W_3 + W_3'))T_1' C \\ &= Q_2 + c_N Q_3, \end{aligned} \tag{A.9}$$

with $W_1 = SL^{(1)}S\Omega S^{*\prime}$, $W_2 = SL^{(1)}S^*\Omega S^{*\prime}$, and $W_3 = S^*L^{(1)}S^*\Omega S^{*\prime}$. Using $SL^{(1)} = I - S$ and the fact that S^* commutes with L_k, then similar to the steps taken toward the end of the proof of part (i) we can show that

$$C'T_1(W_1 + W_1')T_1' C = C'T_1(S\Omega S^{*\prime})T_1' C.$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

Consequently, and because Q_2 is symmetric, we have

$$Q_2 = \frac{3}{2} A^*; \tag{A.10}$$

that is, the leading terms of A_2 and Q_1 are proportional.

We require an expansion for A^{-1} and so notice that

$$AA_1^{-1} = I - c_N A_2 A_1^{-1}$$

implies that as $N \to \infty$, $c_N \to 0$, $AA_1^{-1} \to I$, and hence for any $\xi \in \mathbb{R}^{N-k}$,

$$\lim_{N\to\infty} \frac{\xi'[c_N A_2 A_1^{-1}]\xi}{\xi'\xi} \to 0.$$

Now define an N_k such that for all $N \geq N_k$,

$$\lim_{N_k\to\infty} \left| \frac{\xi'[c_N^* A_2 A_1^{-1}]\xi}{\xi'\xi} \right| \leq k = O(1). \tag{A.11}$$

Consequently, all asymptotic statements to follow are to be interpreted as $N \geq N_k$, $N_k \to \infty$, which implies $c_N \to 0$.

To proceed, we can write

$$\begin{aligned} A^{-1} &= A_1^{-1}[I - c_N A_2 A_1^{-1}]^{-1} \\ &= A_1^{-1}\left[I + \sum_{j=1}^{\infty} (A_2 A_1^{-1})^j \right] = A_1^{-1}[I + c_N Q_4], \end{aligned}$$

where $Q_4 = \sum_{j=1}^{\infty} (c_N)^{j-1} (A_2 A_1^{-1})^j$ and convergence of the summation is asymptotic, as $c_N \to 0$. The crucial quantity appearing in (5) may therefore be written as

$$\bar{A}A^{-1} = -(A_1 + c_N Q_1) A_1^{-1} (I + c_N Q_4). \tag{A.12}$$

Squaring (A.12) and taking the trace we have

$$\begin{aligned} \mathrm{Tr}[(\bar{A}A^{-1})^2] = N &+ 2c_N \mathrm{Tr}[Q_1 A_1^{-1} + Q_4] + (c_N)^2 \mathrm{Tr}[(Q_1 A_1^{-1} + Q_4)^2] \\ &+ 2(c_N)^2 \mathrm{Tr}[Q_1 A_1^{-1} Q_4] + 2(c_N)^3 \mathrm{Tr}[(Q_1 A_1^{-1} + Q_4) Q_1 A_1^{-1} Q_4] \\ &+ (c_N)^4 \mathrm{Tr}[(Q_1 A_1^{-1} Q_4)^2], \end{aligned} \tag{A.13}$$

and taking the trace of (A.12) and then squaring we obtain

$$\begin{aligned} \mathrm{Tr}[\bar{A}A^{-1}]^2 = N^2 &+ 2Nc_N \mathrm{Tr}[Q_1 A_1^{-1} + Q_4] \\ &+ (c_N)^2 (\mathrm{Tr}[Q_1 A_1^{-1} + Q_2]^2 + 2N \mathrm{Tr}[Q_1 A_1^{-1} Q_2]) \\ &+ 2(c_N)^3 \mathrm{Tr}[Q_1 A_1^{-1} + Q_2] \mathrm{Tr}[Q_1 A_1^{-1} Q_2] + (c_N)^4 \mathrm{Tr}[Q_1 A_1^{-1} Q_2]^2. \end{aligned} \tag{A.14}$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

Thus, for sufficiently large N, for $1-\rho = O(c_N)$ and from (A.13) and (A.14), we have

$$\begin{aligned} I_v(1-O(c_N)) &\sim \frac{\mathrm{Tr}[(\bar{A}A^{-1})^2]}{2} - \frac{\mathrm{Tr}[\bar{A}A^{-1}]^2}{2N} \\ &= 0 + \frac{1}{2}(c_N)^2 \left\{ \mathrm{Tr}[(Q_1 A_1^{-1} + Q_4)^2] - \frac{\mathrm{Tr}[Q_1 A_1^{-1} + Q_4]^2}{N} \right\} \\ &\quad + (c_N)^3 \left\{ \mathrm{Tr}[(Q_1 A_1^{-1} + Q_4) Q_1 A_1^{-1} Q_4] \right. \\ &\qquad \left. - \frac{\mathrm{Tr}[Q_1 A_1^{-1} + Q_4]\mathrm{Tr}[Q_1 A_1^{-1} Q_4]}{N} \right\} \\ &\quad + \frac{1}{2}(c_N)^4 \left\{ \mathrm{Tr}[(Q_1 A_1^{-1} Q_4)^2] - \frac{\mathrm{Tr}[Q_1 A_1^{-1} Q_4]^2}{N} \right\}. \end{aligned}$$

If we utilize the expansions given in (A.9) and the proportionality in (A.10), then the information has an asymptotic leading term of

$$I_v(1-O(c_N)) \sim \frac{1}{2}(c_N)^2 \left\{ \mathrm{Tr}\left[\left(\frac{5}{2} A_2 A^{-1} \right)^2 \right] - \frac{\mathrm{Tr}\left[\frac{5}{2} A_2 A^{-1} \right]^2}{2N} \right\}.$$

From the inequalities

$$\begin{aligned} \mathrm{Tr}\left[\left(\frac{5}{2} A_2 A^{-1} \right)^2 \right] - \frac{\mathrm{Tr}\left[\frac{5}{2} A_2 A^{-1} \right]^2}{2N} &\leq \mathrm{Tr}\left[\left(\frac{5}{2} A_2 A^{-1} \right)^2 \right] \\ &\leq \mathrm{Tr}\left[\frac{5}{2} A_2 A^{-1} \right]^2, \end{aligned}$$

and noting that (A.11) implies that for $N \geq N_k$ the eigenvalues of $c_N A_2 A_1^{-1}$ are $O(1)$, we have

$$\mathrm{Tr}[c_N A_2 A^{-1}] = O(N),$$

which subsequently implies that

$$\lim_{c_N \to 0^+, N \geq N_k \to \infty} I_v(1-O(c_N)) = O(N^2).$$

For part (iii), let $\mathbb{V}$ be the rejection region for any invariant test, having size δ, of $H_0: \rho = 1$ against $H_1: |\rho| < 1$. Let $P_{\mathbb{V}}(\rho)$ be the power of that region and define the power gain (i.e., power minus size) as

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

$$\Delta\pi = P_{\mathbb{V}}(\rho) - \delta = \int_{\mathbb{V}} (f(v;\rho) - f(v;1))\, dv.$$

Following Würtz (1997) the power gain satisfies the following bound:

$$\Delta\pi \leq \Delta(1,\rho) = \frac{1}{2}\,\|f(v;1) - f(v;\rho)\|_1 = \frac{1}{2}\int_{v'v=1} |f(v;1) - f(v;\rho)|\, dv, \qquad \textbf{(A.15)}$$

where $\Delta(1,\rho)$ is called the displacement function and $\|.\|_1$ is the L_1 norm.

By Taylor's theorem we may write

$$f(v;\rho) = f(v;1) - (1-\rho)df(v;1) - \frac{(1-\rho)^2}{2}\, d^2 f(v;\tilde{\rho})$$

for some $\tilde{\rho} \in (\rho, 1)$, and we have written

$$df_v(1) = \left.\frac{df(v;\rho)}{d\rho}\right|_{\rho=1} \quad \text{and} \quad d^2 f(v;\tilde{\rho}) = \left.\frac{d^2 f(v;\rho)}{d^2\rho}\right|_{\rho=\tilde{\rho}}.$$

Further, we can write

$$\frac{df(v;\rho)}{d\rho} = S_v(\rho) f(v;\rho),$$

$$\frac{d^2 f(v;\rho)}{d\rho} = [H_v(\rho) + (S_v(\rho))^2] f(v;\rho),$$

where

$$S_v(\rho) = \frac{d\ln[f(v;\rho)]}{d\rho} \quad \text{and} \quad H_v(\rho) = \frac{d^2\ln[f(v;\rho)]}{d\rho^2},$$

so that

$$f(v;1) - f(v;\rho) = (1-\rho)S_v(1)f(v;1) + \frac{(1-\rho)^2}{2}[H_v(\tilde{\rho}) + (S_v(\tilde{\rho}))^2] f(v;\tilde{\rho}).$$

Thus we have

$$\begin{aligned}\|f(v;1) - f(v;\rho)\|_1 &= \left\| (1-\rho)df(v;1) + \frac{(1-\rho)^2}{2}\, d^2 f(v;\tilde{\rho}) \right\|_1 \\ &\leq \|(1-\rho)S_v(1)f(v;1)\|_1 \\ &\quad + \frac{(1-\rho)^2}{2}\,\|[H_v(\tilde{\rho}) + (S_v(\tilde{\rho}))^2] f(v;\tilde{\rho})\|_1,\end{aligned}$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

and we can define ρ^* according to

$$\rho^* = \arg\max_{\tilde{\rho}\in(\rho,1)} \|[H_v(\tilde{\rho}) + (S_v(\tilde{\rho}))^2]f(v;\tilde{\rho})\|_1.$$

Then because $E[(S_v(\rho))^2] = E[-H_v(\rho)]$ we have

$$\|f(v;1) - f(v;\rho)\|_1 \le \|(1-\rho)S_v(1)f(v;1)\|_1 + \frac{(1-\rho)^2}{2}\left\{\int_{v'v=1}[-H_v(\rho^*) + |H_v(\rho^*)|]f(v;\rho^*)\,dv\right\}.$$

Now partition the unit sphere into two regions,

$$s_1(v) \cup s_2(v) = \{v : v'v = 1\}, \qquad s_1(v) \cap s_2(v) = \varnothing,$$

with

$$s_1(v) = \{v : H_v(\rho^*) < 0\},$$

$$s_2(v) = \{v : H_v(\rho^*) \ge 0\},$$

so that

$$\int_{v'v=1}[-H_v(\rho^*) + |H_v(\rho^*)|]f(v;\rho^*)\,dv = \int_{s_1(v)}[-2H_v(\rho^*)]f(v;\rho^*)\,dv,$$

because $-H_v(\rho^*) + |H_v(\rho^*)| = 0$ on $s_2(v)$. Consequently,

$$\begin{aligned}\int_{s_1(v)}[-2H_v(\rho^*)]f(v;\rho^*)\,dv &= \int_{v'v=1}[-2H_v(\rho^*)]f(v;\rho^*)\,dv \\ &\quad + \int_{s_2(v)}[2H_v(\rho^*)]f(v;\rho^*)\,dv \\ &\le \int_{v'v=1}[-2H_v(\rho^*)]f(v;\rho^*)\,dv,\end{aligned}$$

because both $H_v(\rho^*)$ and $f(v;\rho^*)$ are positive on $s_2(v)$. Thus we have

$$\int_{v'v=1}[-H_v(\rho^*) + |H_v(\rho^*)|]f(v;\rho^*)\,dv \le 2I_v(\rho^*),$$

where $I_v(\rho^*)$ is Fisher information evaluated at $\rho = \rho^*$.

Irrespective of the particular set of regressors, from (A.15) it follows that

$$P_{\mathbb{V}}(\rho) \le \delta + \frac{1}{2}\{(1-\rho)E_{\rho=1}[|S_v(1)|] + (1-\rho)^2 I_v(\rho^*)\}. \tag{A.16}$$

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press

When the set of regressors includes a linear trend we have

$$\frac{d\ln(f(v;\rho))}{d\rho} = -\frac{\mathrm{Tr}[A^{-1}dA]}{2} + \frac{(N-k)}{2}\,\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v},$$

where A and $\bar{A}$ are defined earlier in this proof. Thus at $\rho = 1$,

$$d\ln(f(v;1)) = \left.\frac{d\ln(f(v;\rho))}{d\rho}\right|_{\rho=1} = \frac{\mathrm{Tr}[A_1^{-1}\bar{A}_1]}{2} - \frac{(N-k)}{2}\,\frac{v'A_1^{-1}\bar{A}_1A_1^{-1}v}{v'A_1^{-1}v}$$

$$= 0, \qquad \textbf{(A.17)}$$

because $\bar{A}_1 = (dA)_{\rho=1} = -A_1$. Substituting (A.17) into (A.16) and putting $c_N = 1 - \rho$ and $c^* = 1 - \rho^*$ gives the result. ■

https://doi.org/10.1017/S0266466607070296 Published online by Cambridge University Press