THE UNIVERSITY of York

Discussion Papers in Economics

No. 2005/03

The Available Information for Invariant Tests of a Unit Root

by

Patrick Marsh

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

The Available Information for Invariant Tests of a Unit Root[1]

Patrick Marsh
Department of Economics
University of York
Heslington, York
YO10 5DD
tel. +44 1904 433084
fax +44 1904 433071
e-mail: pwnm1@york.ac.uk

April 4, 2005

[1]Thanks are due to Francesco Bravo, Giovanni Forchini, Les Godfrey, Robert Taylor, participants at seminars at the Universities of Birmingham, York and at the ESRC Econometric study group conference, Bristol 2004, for comments on this and related work. Revisions of this paper have greatly benefited from comments and suggestions from Grant Hillier, Peter Phillips, Joel Horowitz and two anonymous referees.

Abstract

This paper considers the information available to invariant unit root tests at and near the unit root. Since 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 bound upon the power of all invariant unit root tests is shown to depend solely upon the information. This bound is illustrated via comparison with the local-to-unity power envelope and a brief simulation study illustrates the impact that the requirements of invariance 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 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.

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 neighbourhood 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 neighbourhood. On the other hand if c_N is $O(1)$ and the model trend stationary, then the 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 affect 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, Rothenberg and Stock (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.

This paper also derives a much stronger link between Fisher information and the power of unit root tests. Specifically, it is shown that the power of any invariant unit root test will be bounded above by its size plus a linear function of the available information. Therefore the results of this paper are shown to have a direct bearing on the power of any of the commonly used unit root tests.

There has been much success in terms of characterising the asymptotic distribution of unit root tests. Despite this success, excepting the simplest zero mean AR(1) case, see 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.

It does not follow from the results that linear trends should be excluded from deterministic components in the unit root context. Instead, it 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). In particular, that they are robust with respect to possibly nonstationary errors is crucial in the context of their use as a pre-unit root specification test on the deterministic component.

The plan for the rest of the paper is as follows, the next section details the models, 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 a conclusion follows.

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

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

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 which 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, Rothenberg and Stock (1996). Asymptotically the resultant tests are invariant and thus our results will have some baring upon all of these procedures.

Let $y = (y_1, ..., y_N)'$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)'$, $e = (1, 1, .., 1)'$, $\tau = (1, 2, ..., N)'$, $Z = (z_1, ..., z_N)'$, $X = (e : \tau : Z)$ 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 one's on the first lower off-diagonal and zero's elsewhere.

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

Assumption 1 *(i) Let the density of* y, given Z, *be* $f(y;\beta,\rho,\sigma^2\Omega\,|\,Z) = f(y) \in \mathcal{F}$ *with*

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

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

(ii) Ω *is such that, a)* $\Sigma_\rho(\Omega)$ *depends on* ρ *for all* $\rho \in \mathbb{R}$ *and b)* $||\Omega||_1 = \sup_j \sum_{i=1}^{N} |\Omega_{i,j}| = \sum_{k=0}^{N-1} |\omega(k)| < M < \infty$, *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 upon the initial condition which 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 non-zero initial condition, y_0, can be incorporated, suppose there are no other regressors, so

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

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

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 + \beta_3 y_0 + u_1 \\ y_2 &= \rho y_1 + (1-\rho)\beta_1 + (2-\rho)\beta_2 - \rho\beta_3 y_0 + u_2, \end{aligned}$$

and no other y_i depends upon y_0 explicitly. Although inference on β_3 would not be possible in this set-up, 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 power of any suitable test will depend upon this nuisance parameter.

Part (ii) a) ensures that there are no common factors in the covariance structure of the data, so ρ may be identified, in principle. Part (ii) b) 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 since ε is covariance stationary then Ω is Toeplitz.

Assumption 1 also implies the data has 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 limiting 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

$$CC' = M_{T_1X} = I - T_1X((T_1X)'T_1X)^{-1}(T_1X)' \quad ; \quad C'C = I_{N-k},$$

so that C is the singular value decomposition of M_{T_1X}. Let $w = C'T_1y$, and consider the group $G = (a, g)$, with $a \in \mathbb{R}$ and $g \in \mathbb{R}^k$ and with action

$$T_1y \to aT_1y + T_1Xg\,, \tag{4}$$

then the maximal invariant, under G, for testing H_0 in (3) is

$$v = w/|w| = \frac{C'T_1y}{(y'T_1'M_{T_1X}T_1y)^{1/2}}, \tag{5}$$

see Kariya (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 normalised Haar measure, is*

$$f(v;\rho) = \det A^{-1/2} \left(vA^{-1}v\right)^{-(N-k)/2}, \tag{6}$$

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

(ii) The Fisher information in v about ρ is

$$I_v(\rho) = \frac{(N-k)Tr\left[\left(A^{-1}\bar{A}\right)^2\right] - \left[Tr(A^{-1}\bar{A})\right]^2}{2\,(N-k+2)}, \tag{7}$$

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

To summarise; the density of v depends upon neither β or σ^2, nor, given Assumption 1, the distribution of the data. Thus the expression for $I_v(\rho)$ is constant over the family $\mathcal{F}$. Generally, though, it will depend upon the structure of $\sigma^2\Omega$. We will thus refer to (7) 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 other point information is non-zero and in any asymptotic neighbourhood, generally it will grow as the square of the sample size. Asymptotic neighbourhoods 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 neighbourhood. These results are contained in the following theorem, again 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(ii), the Fisher information in v is zero at $\rho = 1$, i.e. $I_v(1) = 0$.*

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

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

(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 $c^* \in (0, c_N)$,

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

3 Discussion and Analysis

Theorem 1 demonstrates that the available information is zero at a unit root, but generally non-zero in any asymptotic neighbourhood. Moreover, an explicit relationship between the power of invariant tests is derived. In this section the implication of 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, which leads to all of the results in this paper is that the derivative of the autoregressive covariance matrix is matrix collinear with the outer product of the trend, vis.

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

where $\Sigma_\rho = T_\rho^{-1}\left(T_\rho^{-1}\right)'$. Consequently, any projection of the derivative in the space orthogonal to a trend yields a covariance derivative at unity which 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 'trend-like' regressors will have a negative impact upon available information. This follows since the projection orthogonal to such regressors will be 'close' to being orthogonal to τ, and thus the quantity $A^{-1}\bar{A}$ in (7) 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) also points to another valuable prediction. Efron's (1975) statistical curvature is inversely proportional to the information, while 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 locally most powerful testing procedures. Indeed this is explicitly borne out in numerical work contained in Forchini and Marsh (2000).

Later in this section we will discuss the information in an asymptotic neighbourhood of 1. Before getting there though it is worth pointing out that since $I_v(\rho)$ is differentiable in ρ (in fact it is analytic in ρ, since $\Sigma(\rho)$ as defined is infinitely differentiable for every ρ), then we can find arbitrarily small and positive κ_1 and κ_2 so that

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

That is information is vanishingly small in non-asymptotic neighbourhoods of unity. In asymptotic neighbourhoods the story is quite different, as is apparent from part (ii) of the Theorem.

Suppose that we take the density in (6) as a likelihood for ρ and with a score given by

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

Then since $Var[S_v(\rho)] = I_v(\rho)$ it follows that $S_v(1) = 0$ for all v. That is the score is constant at the unit root. Although certainly unusual the fundamental implication of the score being constant is the implication that the maximum likelihood estimator for ρ based upon (6) will not converge uniformly. However, since there is no uniform convergence in autoregression, in any case, in fact this curiosity does not seem to have serious inferential consequences.

3.2 Characterisation 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

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

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

$$y \to \begin{pmatrix} \hat{\beta} = ((T_1X)'T_1X)^{-1}T_1y \\ w = C'T_1y \end{pmatrix} \quad \text{and} \quad w \to \begin{pmatrix} s^2 = w'w \\ v = w/|w| \end{pmatrix}. \tag{10}$$

Within the context of (9) more restrictive models can be characterised by our knowing the values of the parameters. For example, if we knew all three values then y itself (or T_1y) would be the maximal invariant (in the trivial sense) and the information in T_1y is well known to be

$$I_{T_1y}(\rho) = I_y(\rho) = E\left[\sum_{i=2}^{N} y_{i-1}^2\right] = \sum_{j=1}^{N-1}\sum_{i=1}^{j}\rho^{2i}.$$

On the other hand, for example as in Müller and Elliott (2003), we could assume that the variance in (1) of ε satisfies $\sigma^2 = I_N$ (see their Condition 1), i.e. σ^2 is known. Thus length invariance is not required and only the first transformation in (10) is necessary, giving a maximal invariant $w = C'T_1y \sim \mathcal{F}(C'\Sigma_\rho(\Omega)C)$. Fisher information in w is readily found to be

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

Thus for three cases for (9) we can evaluate the available information at unity, vis.

Known Parameters	Maximal Invariant	Information at Unity
β_1, β_2 and σ^2	T_1y	$I_{T_1y}(1) = \frac{N(N-1)}{2} = O(N^2)$
σ^2	$w = C'T_1y$	$I_w(1) = \frac{(N-k)}{2} = O(N)$
none	$v = w/\lvert w\rvert$	$I_v(1) = 0$

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.

3.3 Characterisation of information over (non)stationarity regimes.

Since 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 neighbourhood of unity. Specifically, for model (1) we find that for $1-\rho = O(c_N)$, where $c_N \to 0$ as $N \to \infty$, then $I_v(\rho)$ 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 Elliott, Rothenberg and Stock (1996) who derive 'optimal' tests for $c = 0$ against $c > 0$.

However, other neighbourhoods 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 neighbourhood, at least in the simplest autoregression, asymptotic Gaussianity of estimators may obtain, rather than the Weiner process driven asymptotics of the former case. Theorem 1 demonstrates that for any asymptotic neighbourhood of unity the rate of increase in available information is $O(N^2)$. That is information is zero only at the point $\rho = 1$.

As a consequence of the results of Theorem 1 we can asymptotically quantify the available information over the three regimes of interest for unit root inferences. Letting the autoregressive parameter be $\rho = 1 - O(c_N)$ then as $N \to \infty$ we can characterise 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

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

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

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

Since information is non-negative we have $I_v(\rho) \leq I_y(\rho) = O(N)$ for $|\rho| < 1$.

We can also, therefore, quantify the rate at which information will accrue in different regimes. To summarise, supposing that the autoregressive parameter is $\rho = 1 - O(c_N) \in (-1, 1]$, then as $N \to \infty$;

$$\begin{aligned} &\text{if } c_N = 0 \text{ then } I_v(\rho) = 0 \\ &\text{if } c_N = o(1) \text{ then } I_v(\rho) = O(N^2) \\ &\text{if } c_N = O(1) \text{ then } I_v(\rho) = O(N). \end{aligned}$$

3.4 Information and power.

Part (iii) of the Theorem gives a direct link between Fisher information and the power of any trend invariant unit root test. Specifically, power is bounded by the size of the test plus a linear function of the information. Therefore, information being zero at the unit root implies directly that power will be small in a neighbourhood of unity. In general the bound on the power of any invariant test for (2) is given by

$$P_{\mathbb{V}}(\rho) \leq \delta + \frac{1}{2}\left\{(1-\rho)E\left[|S_v(1)|\right] + (1-\rho)^2 I_v(\rho^*)\right\}.$$

This bound is specifically tightened by the inclusion of a linear trend in that the first term vanishes due to the constancy of the score. In order to illustrate the efficacy of the bound in this case, consider testing $\rho = 1$ in model (9).

Based upon a sample of size 500 and putting $\rho = 1 - c/N$, with $c = O(1)$, the asymptotic power envelope (the power of the set of point optimal tests of size δ) can be approximated via the experiments described in Section 4 of Elliott, Rothenberg and Stock (1996) and which were precisely replicated here. Let $\Pi(c)$ denote the asymptotic power envelope at the point c which is approximated via 100,000 Monte Carlo replications, with standard error in the fourth decimal place. From part (iii) of Theorem 1 it must be that

$$\Pi(c) \leq PUB(c, c^*) = \delta + \frac{1}{2}\left(\frac{c}{500}\right)^2 I_v\left(1 - c^*/500\right), \tag{11}$$

at each point c and for some $c^* \in (0, c)$. Below in Figure 1, both $\Pi(c)$ and $PUB(c, c/2)$ with $\delta = 0.05$ are plotted over the range $c \in (0, 25)$. Very close to unity this particular bound is very tight, less so for larger c. However, it should be borne in mind that what is graphed is a numerical bound with a uniform value of $c^* = c/2$. Point-wise it would be possible to improve upon this, perhaps significantly so for large c.

As was seen above in section (3.2) 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 characterise six different cases. For the purposes of experiment these can be represented by data $(y_t)_1^N$ generated by

$$
\begin{aligned}
M_1 &: \; y_t = u_t \\
M_2 &: \; (y_t - \beta_1) = u_t \\
M_3 &: \; (y_t - \beta_1 - \beta_2 t) = u_t,
\end{aligned} \tag{12}
$$

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 test can easily be derived. If we also define a matrix C_1, so that

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

then for each case;

Model		Maximal Invariant	Point Optimal (PO) test
M_1 :	σ^2 known	$T_1 y$	$y' T_1' \left(I - \Sigma(\rho)\right) T_1 y$
	σ^2 unknown	$v_0 = T_1 y / \lvert T_1 y \rvert$	$v_0' \Sigma(\rho)^{-1} v_0$
M_2 :	σ^2 known	$w_1 = C_1' T_1 y$	$w_1' \left(I - C_1' T_1 \Sigma(\rho) T_1 C_1\right) w_1$
	σ^2 unknown	$v_1 = w_1 / \lvert w_1 \rvert$	$v_1' \left(C_1' T_1 \Sigma(\rho) T_1 C_1\right)^{-1} v_1$
M_3 :	σ^2 known	$w = C' T_1 y$	$w' \left(I - C' T_1 \Sigma(\rho) T_1 C\right) w$
	σ^2 unknown	$v = w / \lvert w \rvert$	$v \left(C' T_1 \Sigma(\rho) T_1 C\right)^{-1} v$

. (13)

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 asymptotic power envelope for each case was approximated by simulations like those described above. The results are presented in Table 1 below.

The most restrictive case (M_1, σ^2 known) provides an absolute power upper bound, in the sense that in this case the test satisfies the Neyman-Pearson lemma 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, at which point Elliott, Rothenberg and Stock (1996) define their nearly efficient test. For the most restrictive case this is approximately 6.5 whereas for the most general it is 17.5 (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 assumption about the trend.

4 Conclusions

This paper has proven that inclusion of a trend in the Sargan and Bhargava (1983) formulation implies zero available information at the unit root. Both the explicit bound on power by information and the numerical work suggest that low power is an inevitable consequence of the requirement of invariance to a trend. 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.

From an applied perspective, determining the necessity of a trend becomes of paramount importance. Fortunately, recent work by Vogelsang (1998) and Bunzel and Vogelsang (2003) provides procedures which seem to fit this requirement. In particular procedures for testing the significance of the trend which are robust to non-stationarity of the errors.

A Figure and Table

Figure 1: Plot of $PUB(c, c/2)$ (Solid) $vs.$ $\Pi(c)$ (Dashed)

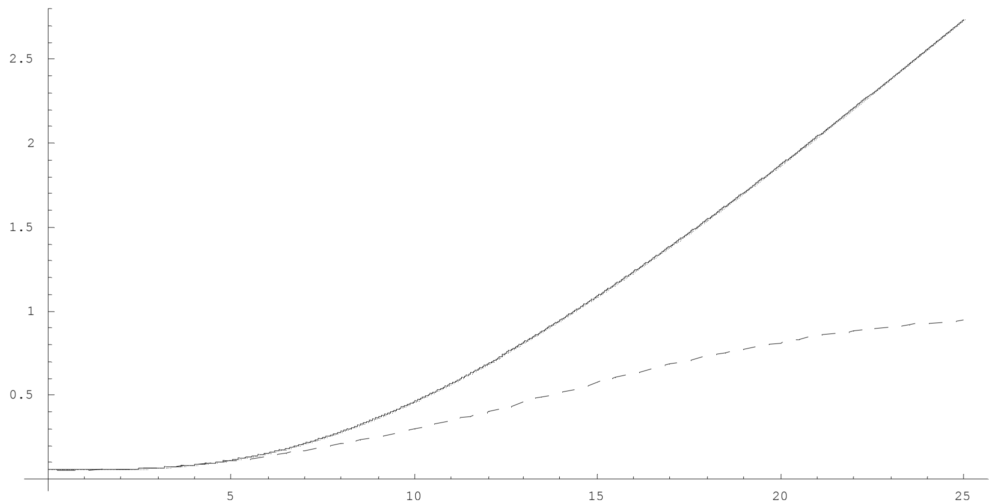

Table 1: Powers of the point optimal tests given in (13),

for the models in (12) with cases A (σ^2 known) and B (σ^2 unknown).

c	M_1^A	M_1^B	M_2^A	M_2^B	M_3^A	M_3^B
1	.083	.079	.061	.060	.058	.050
2	.130	.117	.070	.061	.066	.052
3	.197	.193	.095	.079	.087	.061
5	.346	.328	.144	.122	.099	.072
7	.538	.524	.196	.176	.146	.104
10	.793	.783	.353	.248	.235	.157
15	.964	.964	.671	.624	.473	.399
20	.997	.996	.870	.867	.700	.629
25	1.00	1.00	.972	.972	.877	.839

References

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

Bunzel, H. and 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 and 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 and C.H. Whiteman (1992) Integration versus trend stationarity in time series. *Econometrica,* **60**, 423-433.

Dufour, J-M. and M.L. King (1991) Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *Journal of Econometrics,* **47**, 115-143.

Durlauf, S.N. and 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 and J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica*, **64**, 813-836.

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

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

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

Lehmann, E.L. (1997) *Testing Statistical Hypotheses.* (2nd ed.), Springer texts in statistics.

Leybourne, S.J., T.C. Mills, and P. Newbold (1998) Spurious rejections by Dickey–Fuller tests in the presence of a break under the null. *Journal Of Econometrics*, **87**, 191-203.

Müller, U.K. and 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).

Ploberger, W. and P.C.B. Phillips (2002) Rissanen's theorem and econometric time series. In A. Zellner, H. A. Keuzenkamp and M. McAleer (eds.), *Simplicity, Inference and Modelling*, Cambridge University Press: Cambridge, 165-180.

——————— (2003) Empirical limits for time series econometric models. *Econometrica* **71**, 627-673.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**, 535-547.

——————— (1998) New tools for understanding spurious regressions. *Econometrica,* **66**, 1299-1325.

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. and 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. *University of New South Wales, Discussion Paper,* **97/17**.

Zivot, E. and D.W.K. Andrews (1992) Further evidence on the great crash, the oil price shock, and the unit root hypothesis. *Journal of Business and Economic Statistics*, **10**, 251-270.

Appendix:

Proof of Lemma 1: Part (i) follows immediately from the results of Kariya (1980), although it is worth noting that the result does not depend in any way upon the particular form of covariance matrix chosen in that paper.

For part (ii) we have that the log-likelihood is

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

and hence the score is

$$S_v(\rho) = \frac{dL_v(\rho)}{d\rho} = -\frac{Tr[A^{-1}\bar{A}]}{2} + \frac{N-k}{2}\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}, \tag{14}$$

from which it is possible to obtain

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

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

$$E\left[\frac{v'A^{-1}\bar{A}A^{-1}v}{v'A^{-1}v}\right] = \frac{Tr[A^{-1}\bar{A}]}{N-k}, \tag{15}$$

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 0 and $Var[w] = A$, then

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

where

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

and $z = A^{-1/2}w$. Further, since 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

$$\begin{aligned} E[r_z^2] &= \frac{E[(z'A^{-1/2}\bar{A}A^{-1/2}z)^2]}{E[(z'z)^2]} = \frac{Var[(z'A^{-1/2}\bar{A}A^{-1/2}z)] + E\left[(z'A^{-1/2}\bar{A}A^{-1/2}z)\right]^2}{E[(z'z)^2]} \\ &= \frac{2Tr\left[(A^{-1}\bar{A})^2\right] + Tr\left[A^{-1}\bar{A}\right]^2}{2(N-k)+(N-k)^2}. \end{aligned} \tag{16}$$

The expression for the information follows from squaring (14), utilising the expectations given in (15) and (16) and rearranging. ■

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

$$I_v(1) = \frac{(N-k)Tr\left[\left(A_1^{-1}\bar{A}_1\right)^2\right] - \left[Tr(A_1^{-1}\bar{A}_1)\right]^2}{2\,(N-k+2)}, \tag{17}$$

where

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

while 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,

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

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

$$d\Sigma_1 = SL^{(1)}SS' + SS'(L^{(1)})'S'. \tag{18}$$

From the identity,

$$S + S' - I_N = (T_1\tau)(T_1\tau)', \tag{19}$$

where $\tau = (1, 2, 3, ..., N)'$ and substituting $L^{(1)} = I_N - T_1$ into (18) we obtain

$$d\Sigma_1 = \tau\tau' - SS', \tag{20}$$

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

Since Ω is Toeplitz and symmetric we can write

$$\Omega = \sum\nolimits_{k=0}^{N-1} \omega(k) F_k \quad ; \quad F_k = \left(L_k + L_k'\right),$$

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

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

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

$$\begin{aligned} \bar{A}_1 &= \sum\nolimits_{k=0}^{N-1} \omega(k) C' \left((S - I_N) F_k + F_k (S' - I_N)\right) C \\ &= \sum\nolimits_{k=0}^{N-1} \omega(k) C' \left((S - I_N) F_k + F_k S'\right) C - C' \Omega C. \end{aligned} \tag{21}$$

The last term in (21) is just $-A_1$, and the quadratic forms in C' can be rewritten as in

$$\begin{aligned} \bar{A}_1 &= \sum\nolimits_{k=0}^{N-1} \omega(k) C' \left((S - I_N) L_k + (S - I_N) L_k' + L_k S' + L_k' S'\right)\right) C - A_1 \\ &= \sum\nolimits_{k=0}^{N-1} \omega(k) \left\{C' L_k \left(S + S' - I_N\right) C + C' \left(S + S' - I_N\right) L_k' C\right\} - A_1, \end{aligned}$$

where we have exploited $(S - I_N) L_k = L_k (S - I_N)$. Thus using (19), we have

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

and finally since $(T_1 \tau)' C = 0 = C' (T_1 \tau)$, as X includes a trend, then $\bar{A}_1 = -A_1$, and substitution into (17) gives, as required $I_v(1) = 0$. ∎

For part (ii) we require an expansion of the matrices A^{-1} and $\bar{A}$ appearing in (7). To proceed consider the matrix defined by

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

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

$$\begin{aligned} T_c^{-1} &= S - c_N S^*, \\ S^* &= \sum_{k=1}^{N-1} L_k \left(k e^{-k c_k^*}\right), \end{aligned} \tag{22}$$

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

Application of the expansion given in (22) yields

$$\begin{aligned} A &= A_1 - c_N A_2 \\ \bar{A} &= -A_1 - c_N Q_1, \end{aligned} \tag{23}$$

where

$$\begin{aligned} A_2 &= C'T_1\left(S\Omega S^{*'} + S^*\Omega S' + c_N S^*\Omega S^{*'}\right)T_1'C = A^* + c_N A_3, \\ Q_1 &= C'T_1\left(3(W_1+W_1') - 3c_N(W_2+W_2') + (c_N)^2\left(W_3+W_3'\right)\right)T_1'C \\ &= Q_2 + c_N Q_3, \end{aligned} \tag{24}$$

with $W_1 = SL^{(1)}S\Omega S^{*'}$, $W_2 = SL^{(1)}S^*\Omega S^{*'}$ and $W_3 = S^*L^{(1)}S^*\Omega S^{*'}$. 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

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

Consequently, and since Q_2 is symmetric, we have

$$Q_2 = \frac{3}{2}A^*, \tag{25}$$

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

We require an expansion for A^{-1}, and so note that the matrices A and A_1 are symmetric and positive definite, while $c_N A_2$ is symmetric and positive semi-definite. Consequently, both

$$AA_1^{-1} = \left(I - c_N A_2 A_1^{-1}\right) \quad \text{and} \quad c_N A_2 A_1^{-1}$$

are positive semi-definite, which implies that

$$0 \leq \frac{\xi'\left[c_N^* A_2 A_1^{-1}\right]\xi}{\xi'\xi} \leq 1, \tag{26}$$

for any $\xi \in \mathbb{R}^{N-k}$. Hence the spectrum of $c_N A_2 A_1^{-1}$ lies between 0 and 1 and so we can write

$$\begin{aligned} A^{-1} &= A_1^{-1}\left[I - c_N A_2 A_1^{-1}\right]^{-1} \\ &= A_1^{-1}\left[I + \sum_{j=1}^{\infty}\left(A_2A_1^{-1}\right)^j\right] = A_1^{-1}\left[I + c_N Q_4\right], \end{aligned}$$

where $Q_4 = \sum_{j=1}^{\infty} (c_N)^{j-1} \left(A_2 A_1^{-1}\right)^j$.

Therefore the crucial quantity appearing in (17) may be written as

$$\bar{A}A^{-1} = -\left(A_1 + c_N Q_1\right) A_1^{-1} \left(I + c_N Q_2\right). \tag{27}$$

Squaring (27) and taking the trace we have

$$\begin{aligned} Tr\left[\left(\bar{A}A^{-1}\right)^2\right] &= N + 2c_N Tr[Q_1 A_1^{-1} + Q_4] + (c_N)^2\, Tr[(Q_1 A_1^{-1} + Q_4)^2] \\ &\quad + 2\,(c_N)^2\, Tr[Q_1 A_1^{-1} Q_4] + 2\,(c_N)^3\, Tr[(Q_1 A_1^{-1} + Q_4) Q_1 A_1^{-1} Q_4] \\ &\quad + (c_N)^4\, Tr[(Q_1 A_1^{-1} Q_4)^2], \end{aligned} \tag{28}$$

while taking the trace of (27) and then squaring we obtain

$$\begin{aligned} Tr\left[\bar{A}A^{-1}\right]^2 &= N^2 + 2Nc_N Tr\left[Q_1 A_1^{-1} + Q_4\right] + \\ &\quad + (c_N)^2 \left(Tr\left[Q_1 A_1^{-1} + Q_2\right]^2 + 2N Tr\left[Q_1 A_1^{-1} Q_2\right]\right) \\ &\quad + 2\,(c_N)^3\, Tr[Q_1 A_1^{-1} + Q_2] Tr[Q_1 A_1^{-1} Q_2] + (c_N)^4\, Tr[Q_1 A_1^{-1} Q_2]^2. \end{aligned} \tag{29}$$

Thus, for sufficiently large N, for $1 - \rho = O(c_N)$ and from (28) and (29), we have

$$\begin{aligned} I_v(1 - O(c_N)) &\sim \frac{Tr\left[\left(\bar{A}A^{-1}\right)^2\right]}{2} - \frac{Tr\left[\bar{A}A^{-1}\right]^2}{2N} \\ &= 0 + \frac{1}{2}\,(c_N)^2 \left\{ Tr\left[(Q_1 A_1^{-1} + Q_4)^2\right] - \frac{Tr\left[Q_1 A_1^{-1} + Q_4\right]^2}{N} \right\} \\ &\quad + (c_N)^3 \left\{ Tr\left[(Q_1 A_1^{-1} + Q_4) Q_1 A_1^{-1} Q_4\right] - \frac{Tr\left[Q_1 A_1^{-1} + Q_4\right] Tr[Q_1 A_1^{-1} Q_4]}{N} \right\} \\ &\quad + \frac{1}{2}\,(c_N)^4 \left\{ Tr\left[(Q_1 A_1^{-1} Q_4)^2\right] - \frac{Tr[Q_1 A_1^{-1} Q_4]^2}{N} \right\}. \end{aligned} \tag{30}$$

If we utilise the expansions given in (24) and the fact that the leading terms in A_2 and Q_1 are proportional, then the information has an asymptotic leading term of

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

From the inequalities

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

and noting that (26) implies for $c_N > 0$ that the eigenvalues of $c_N A_2 A_1^{-1}$ are $O(1)$, then we have

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

which subsequently implies that

$$\underset{c_N \to 0^+,\, N\to\infty}{Lim} I_v(1 - O(c_N)) = O(N^2). \quad \blacksquare$$

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

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

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

$$\Delta\pi \leq \Delta(1,\rho) = \frac{1}{2}\|f(v;1) - f(v;\rho)\|_1 = \frac{1}{2}\int_{v'v=1} |f(v;1) - f(v;\rho)|\, dv, \tag{31}$$

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 where we have written

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

Further we can write

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

where

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

so that

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

Thus we have,

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

and we can define ρ^* so that

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

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

$$\begin{aligned}\|f(v;1)-f(v;\rho)\|_1 &\leq \|(1-\rho)S_v(1)f(v;1)\|_1\\ &\quad+\frac{(1-\rho)^2}{2}\left\{\int_{v'v=1}\left[-H_v(\rho^*)+|H_v(\rho^*)|\right]f(v;\rho^*)dv\right\}.\end{aligned}$$

Now partition the unit sphere into two regions

$$s_1(v)\cup s_2(v)=\left\{v:v'v=1\right\},$$

in which

$$\begin{aligned}s_1(v)&=\{v:H_v(\rho^*)<0\}\\ s_2(v)&=\{v:H_v(\rho^*)\geq 0\},\end{aligned}$$

so that

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

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

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

since both $H_v(\rho^*)$ and $f(v;\rho^*)$ are positive on $s_2(v)$ and so we have

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

where $I_v(\rho^*)$ is Fisher information evaluated at $\rho = \rho^*$. Thus, irrespective of the particular set of regressors, from (31) it follows that

$$P_{\mathbb{V}}(\rho) \leq \delta + \frac{1}{2}\left\{(1-\rho)E_{\rho=1}\left[|S_v(1)|\right] + (1-\rho)^2 I_v(\rho^*)\right\}. \tag{32}$$

When the set of regressors includes a linear trend we have

$$f(v;\rho) \propto |A|^{-1/2}\left(v'A^{-1}v\right)^{-(N-k)/2}$$

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

where A and $\bar{A}$ (respectively B and $\bar{B}$ in the case of v_2) are defined above. Thus at $\rho = 1$ we have

$$\begin{aligned} d\ln(f(v;1)) &= \left.\frac{d\ln(f(v;\rho))}{d\rho}\right|_{\rho=1} = \frac{Tr\left[A_1^{-1}\bar{A}_1\right]}{2} - \frac{(N-k)}{2}\frac{v'A_1^{-1}\bar{A}_1A_1^{-1}v}{v'A_1^{-1}v} \\ &= 0, \end{aligned} \tag{33}$$

since $\bar{A}_1 = (dA)_{\rho=1} = -A_1$. Substituting (33) into (32) and putting $c_N = 1-\rho$ and $c^* = 1-\rho^*$ gives the result. ■