THE UNIVERSITY of York

Discussion Papers in Economics

No. 2005/02

A Measure of Distance for the Unit Root Hypothesis

by

Patrick Marsh

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

A Measure of Distance for the Unit Root Hypothesis[1]

Patrick Marsh[2]
Department of Economics
University of York
Heslington, York
YO10 5DD
Tel. No. +44 (0)1904 433084
e-mail: pwnm1@york.ac.uk

March 11^{th} 2005

[1] Thanks are due to Francesco Bravo, Giovanni Forchini, Peter Phillips, Robert Taylor, two anonymous referees and participants at seminars at the Universities of Birmingham, Manchester, York and Queen Mary College.

[2] Financial support in the form of a Leverhulme Special Reseach Fellowship is gratefully acknowledged.

Abstract

This paper proposes and analyses a measure of distance for the unit root hypothesis tested against stochastic stationarity. It applies over a family of distributions, for any sample size, for any specification of deterministic components and under additional autocorrelation, here parameterised by a finite order moving-average. The measure is shown to obey a set of inequalities involving the measures of distance of Gibbs and Su (2002) which are also extended to include power. It is also shown to be a convex function of both the degree of a time polynomial regressors and the moving average parameters. Thus it is minimisable with respect to either. Implicitly, therefore, we find that linear trends and innovations having a moving average negative unit root will necessarily make power small. In the context of the Nelson and Plosser (1982) data, the distance is used to measure the impact that specification of the deterministic trend has on our ability to make unit root inferences. For certain series it highlights how imposition of a linear trend can lead to estimated models indistinguishable from unit root processes while freely estimating the degree of the trend yields a model very different in character.

1 Introduction

Over the past two decades our progress in understanding of unit root processes and our ability to model nonstationary time series has been tremendous. Despite this, analytic results in closed form still remain relatively scarce. Some noteworthy exceptions are distributional results due to Abadir (1993), Phillips and Ploberger (1994), as well as related results by these and other some other authors, such as the distribution given in Forchini (2002). With few exceptions asymptotic analysis involves finding representations for the limiting process, rather than its distribution, which then has to be approximated via Monte Carlo.

To understand why detailing analytic properties might be important consider the following model for a time series $(y_t)_{t=1}^{N}$,

$$y_t = d_t + u_t \quad ; \quad u_t = \alpha u_{t-1} + \zeta_t, \quad t = 1, 2, ..., N, \tag{1}$$

where d_t is a deterministic component and ζ_t is an innovation process. We wish to test $H_0 : \alpha = 1$ against $H_1 : |\alpha| < 1$. Numerical evidence suggests that the trending characteristics of d_t and the correlation properties of the ζ_t can dramatically affect the performance, specifically power, of all recommended test procedures. See, amongst many others, Durlauf and Phillips (1988), Phillips and Perron (1988), Perron (1989), DeJong *et al* (1991), Zivot and Andrews (1992), Elliott, Rothenberg and Stock (1996), Leybourne, Mills and Newbold (1998) and Phillips and Xiao (1998, §4).

Generally, authors providing such numerical evidence have tended to consider only a few parameterisations of d_t (a constant, a trend with possibly specified breaks in each), in practice any d_t might be appropriate, such as the slowly varying trends of Phillips (2001). The asymptotic representation of any given test statistic will differ for different d_t and so relying only on simulation of partial sum processes implies a practical limit on our capacity to investigate this dependence.

This paper proposes a measure of distance of H_0 to H_1. Since the distance is analytic the investigation of the effects of both the deterministics and the autocorrelation is made more feasible. This distance, called Statistical Entropic Complexity, seems new to the econometrics literature, but is well established elsewhere, see Poskitt (1987) and Bozdogan (1990) and is also related to Shannon Entropy as used in a similar context by Phillips and Ploberger (2003). It is a distance on the space of density functions applied, in this case, to the density of the maximal invariant, of which all invariant tests are functions, see Dufour and King (1991). The distance therefore applies over a family of sample distributions, for any deterministic component and for any stationary ζ_t.

For the distance several results are demonstrated. First it is shown that is belongs to the family of measures satisfying the inequalities of Gibbs and Su (2002). In addition, the Power Upper Bound of Würtz (1997) is added to the set of inequalities, and thus we may have confidence that the proposed distance is measuring the same phenomenon as the power of a test. To understand the impact of trends in d_t and of the autocorrelation in ζ_t we suppose that the deterministics are parameterised as a polynomial in time and the innovations as a moving average. We then find that the distance is a convex, and therefore minimisable, function of the polynomial degree and moving average parameters. The special cases of a linear time trend and a moving average with a negative unit root give distances virtually indistinguishable from the minimum. Thus, through the inequalities, the presence of these will necessarily lower the likelihood of our correctly rejecting a false unit root null hypothesis. This result gives analytic confirmation to the wealth of numerical evidence, on this point, accumulated in the papers mentioned above.

The practical usefulness of the distance measure is illustrated via application to the Nelson and Plosser (1982) dataset. For certain series and depending upon the precise specification of the model authors have reached differing conclusions about the presence of unit roots. Compare, for instance, Dejong and Whiteman (1991) and Phillips (1991). Here we will estimate two specifications, which differ only in the specification of a trend in d_t. An unrestricted model has $d_t = \beta_1 + \beta_2 t^p$, so that the degree of trend can be estimated while a restricted version imposes $p = 1$, i.e. a linear trend is assumed. Several authors, for example Bhargava (1986) and Campbell and Perron (1991), amongst others, have suggested testing should take place in the presence of a maintained linear trend, in any event. This is justified as a modelling strategy as a conservative reaction to the general uncertainty about whether a trend is necessary. Alternatively, Phillips (2001) argues that for some series, negative values for p, implying an evaporating trend, may be more appropriate.

This paper is able to address the question of the affect that imposing a linear trend has on ability to make inferences about the unit root. For several of the series p is found to significantly differ from 1. Amongst the series for which the effect is most striking are Real Wages and Money Velocity. In both cases when the linear trend is imposed the estimated model is, according to the distance measure (and indirectly therefore by power), indistinguishable from a unit root process. On the contrary, for the unrestricted case p is found to be far from 1 and the model very much distinguishable from a unit root process. Merely looking at the respective estimated autoregressive coefficients the scale of this difference might well be missed and moreover could not be easily inferred from accompanying simulation evidence.

The plan for the rest of the paper is as follows. The next section details the model specification, proposed the measure of distance and details its relationship with other measures, in terms of inequalities which are then numerically highlighted. Section 3 demonstrates how the distance depends, analytically, upon the trending behaviour of d_t and the autocorrelation structure of ζ_t. Section 4 presents the application of the distance measure for the Nelson and Plosser data and Section 5 concludes. Appendices contain all of the proofs and derivations along with tables and graphs illustrating the numerical results.

2 Some Preliminary Results

In this section we formalise the class of models under consideration and derive the measure of distance, (Statistical) Entropic Complexity. To do so define the following $N \times 1$ vectors,

$$y = (y_t)_{t=1}^N \quad ; \quad d = (d_t)_{t=1}^N \quad ; \quad \zeta = (\zeta_t)_{t=1}^N \quad \text{and } \varepsilon = (\varepsilon_t)_{t=1}^N,$$

let $L^{(i)}$ define a lower triangular matrix with $1^{'}s$ on the i^{th} lower diagonal and $0's$ elsewhere, so that

$$\left(I_N - \alpha L^{(1)}\right) y = \zeta = \sigma K_\phi \varepsilon, \tag{2}$$

where $V[\varepsilon] = I_N$ and $V[\zeta] = \sigma^2 K_\phi K_\phi'$, with K_ϕ a $N \times N$ matrix depending on some set of parameters $\phi = (\phi_j)_{j=1}^m$ and σ^2 a scalar variance. Formally, we will consider models of the form (2) satisfying:

Assumption 1 (i) Let the density of y be $f(y; d, \sigma^2, \alpha, \phi) = f(y) \in \mathcal{F}(d, \sigma^2\Omega_{\alpha,\phi})$, with

$$\mathcal{F}(\Omega_{\alpha,\phi}) = \left\{ f : f(y; d, \sigma^2, \alpha) = \left|\sigma^2\right|^{-1/2} \det \Omega_{\alpha,\phi}^{-1/2}\, q\left[\sigma^{-2}(y-d)'\Omega_{\alpha,\phi}^{-1}(y-d)\right] \right\},$$

where q is a nonincreasing convex function on $[0, \infty)$ and

$$V[y] = \sigma^2\Omega_{\alpha,\phi} = T_\alpha^{-1} K_\phi K_\phi' (T_\alpha^{-1})' \quad ; \quad T_\alpha = \left(I_N - L^{(1)}\alpha\right).$$

(ii) The mean and covariance structure of y are determined by

$$d = X\beta \quad \text{and} \quad K_\phi = (I_N + \sum\nolimits_{i=1}^{m} \phi_i L^{(i)}),$$

where $X = (x_1', .., x_t')'$ is a $N \times k$ matrix of regressors with rank k, i.e. $d_t = x_t'\beta$, with β a $k \times 1$ vector of parameters and the ϕ_j. ■

Assumption 1 implies consideration of a special case of (1), in which

$$(1 - \alpha l)(y_t - d_t) = \left(1 + \sum\nolimits_{i=1}^{m} \phi_i l^i\right) \varepsilon_t, \tag{3}$$

where l is the lag-operator. That is (3) specifies an $ARIMA(0, 1, m)$ model in $y_t - d_t$, when $\alpha = 1$. The restrictions imposed are that (a) the mean is linear in the x_t, (b) the autocorrelation in the error term may be modelled by a finite (invertible) moving average and (c) the underlying distribution is within the elliptically symmetric family $\mathcal{F}(d, \sigma^2\Omega_{\alpha,\phi})$ which contains, for example, contaminated Normal distributions, the multivariate t-distribution, including as a limit the multivariate Cauchy. We will write the joint sample density of y as,

$$y \sim \mathcal{F}(X\beta, \sigma^2 T_\alpha^{-1} K_\phi K_\phi' (T_\alpha^{-1})').$$

Let Ξ denote the space $\theta = (\alpha, \beta', \sigma^2, \phi')' = (\alpha \in (-1, 1], \beta \in \mathbb{R}^k, \sigma^2 \in \mathbb{R}^+, \phi \in \mathbb{R}^m)$, and furthermore $\Xi_0 = (1, \beta', \sigma^2, \phi')'$ and $\Xi_1 = \Xi - \Xi_0$, then the unit root hypothesis is, formally

$$H_0 : \theta \in \Xi_0 \quad \text{vs.} \quad H_1 : \theta \in \Xi_1,$$

which is a classical nuisance parameter problem, (Cox and Hinkley (1974)), with nuisance parameters $(\beta', \sigma^2, \phi')'$. Consequently, noting that $\det K_\phi = 1$, let $z = K_\phi^{-1} T_1 y$ and $W = K_\phi^{-1} T_1 X$, so that

$$z \sim \mathcal{F}(W\beta, \sigma^2 \Sigma_{\alpha,\phi}), \tag{4}$$

where $\Sigma_{\alpha,\phi} = K_\phi^{-1} T_1 T_\alpha^{-1} K_\phi K_\phi' (T_\alpha^{-1})' T_1' (K_\phi^{-1})'$. Estimation in equation (4) implies an (albeit unfeasible) GLS problem, so that if we let $M_W = I_N - W(W'W)^{-1}W'$, and decompose M_W, with $CC' = M_W$ and $C'C = I_{N-k}$, then we may transform z according to

$$z \to \begin{pmatrix} \hat{\beta} = (W'W)^{-1}W'z \\ w = C'z \end{pmatrix} \quad \text{and} \quad w \to \begin{pmatrix} s^2 = w'w = z'M_W z \\ v = w/||w|| = C'z/s \end{pmatrix}. \tag{5}$$

Notice that although $\hat{\beta}$ is not a feasible estimator of β we have two options. First we may assume that $\phi = (\phi_1, .., \phi_m)'$ is known or second, replace ϕ, wherever it appears by some consistent estimator, say $\tilde{\phi}$. If we then denote any object depending upon ϕ, evaluated at $\tilde{\phi}$ by, e.g. $\tilde{x} = K_{\tilde{\phi}}^{-1} T_1 y$, then we have similar relations in the 'tilded' quantities, but interpret the results asymptotically, since $\operatorname{plim} \tilde{\phi} = \phi$. In either case, the interpretation of the results to follow will be the same. Consequently, we shall not distinguish between the cases notationally.

Whether the interpretation is asymptotic or not, under Assumption 1, the density of v, defined with respect to Normalised Haar Measure on the unit sphere $\mathbb{S}_{N-k} = \left\{ v \in \mathbb{R}^{N-k} : v'v = 1 \right\}$, is

$$f_v(\rho) = \det A^{-1/2} \left(v' A^{-1} v \right)^{-\frac{N-k}{2}}, \tag{6}$$

see Kariya (1980), where $A = C' \Sigma_{\alpha,\phi} C$ is a $N - k \times N - k$ positive definite symmetric matrix. We call v the maximal invariant for testing H_0 and has uniform distribution on $\mathbb{S}_{N-k}$ when H_0 is true. Thus v characterises the class of invariant tests for H_0, see Dufour and King (1991). In addition, if $\tilde{\phi}$ is consistent for an unknown ϕ, v characterises a class of asymptotically pivotal tests, while if $\mathcal{F}$ is Gaussian, v characterises the class of (asymptotically) similar tests, see Hillier (1987) for more details.

Since all invariant tests are functions of v we will define a measure of distance for the unit root hypothesis, based upon the density of v given in (6).

2.1 Measures of Distance

There are many measures of distance on the space of density functions. Gibbs and Su (2002) detail the relationships between, i.e. the inequalities satisfied by, these distances. These relationships are important since, depending upon the nature of the model, one may be more readily calculable than another. For example, in order to demonstrate convergence with respect to relative entropy, or Kullback-Leibler divergence, it is sufficient to demonstrate it for the Chi-Square distance. Of course these measures are also fundamental in the sense that each also yields an associated testing procedure. For example, the former yields the Likelihood Ratio and the latter, Pearson's Chi-Square goodness-of-fit.

The measure proposed in this paper, Entropic Complexity is defined by

$$\Delta_{EC} = \frac{1}{2} \ln |A| + \frac{(N-k)}{2} \ln \int_{v'v=1} \left(v' A^{-1} v \right) (dv)$$

$$= \frac{1}{2}\ln|A| + \frac{(N-k)}{2}\ln\left[\frac{Tr(A^{-1})}{N-k}\right]. \tag{7}$$

As a statistical measure it has been used by Poskitt (1987) in the context of Bayesian model selection and Bozdogan (1990) in a wider modelling context. To ensure this measure is not arbitrary we must examine its relationship with the distances of Gibbs and Su (2002). Pertinent to the results of this paper are the Total Variation, Kullback-Leibler and Chi-Square measures of distance, defined respectively by;

$$\begin{aligned}
\Delta_{TV} &= \frac{1}{2}\int_{v'v=1} |f_v(1) - f_v(\alpha)|(dv) = \frac{1}{2}\int_{v'v=1} |1 - |A|^{-1/2}\left(v'A^{-1}v\right)^{-(N-k)/2}|(dv), \\
\Delta_{KL} &= \int_{v'v=1} f_v(1)\ln\left(\frac{f_v(1)}{f_v(\alpha)}\right)(dv) = -\frac{1}{2}\ln|A| + \frac{(N-k)}{2}\int_{v'v=1}\ln\left(v'A^{-1}v\right)(dv) \\
\Delta_{\chi^2} &= \int_{v'v=1} \frac{\left(f_v(1) - f_v(\alpha)\right)^2}{f_v(\alpha)}(dv) = -1 + |A|^{1/2}\int_{v'v=1}\left(v'A^{-1}v\right)^{(N-k)/2}(dv).
\end{aligned} \tag{8}$$

In addition we will consider the power gain, that is the power minus size, of the most powerful invariant test, characterised by a region $\omega^* \in \mathbb{S}_{N-k}$, which is given by

$$\Delta_{\omega^*} = \sup_{\omega\in\mathbb{S}_{N-k}}\int_{\omega} f_v(1) - f_v(\alpha)(dv) = \sup_{\omega\in\mathbb{S}_{N-k}} \delta - |A|^{-1/2}\int_{\omega}\left(v'A^{-1}v\right)^{-(N-k)/2}(dv). \tag{9}$$

The following Theorem gives the relationship between these five measures, extending the set of inequalities in Gibbs and Su (2002).

Theorem 1 *The Power Gain, Total Variation, Kullback-Leibler, Entropic Complexity and Chi-Square measures for the distance between the unit root null and fixed alternative* $|\alpha| < 1$ *satisfy the following set of inequalities*

$$i)\quad \Delta_{\omega} \leq \Delta_{TV} \quad ; \quad ii)\quad \Delta_{TV} \leq \sqrt{\frac{\Delta_{KL}}{2}}$$

$$iii)\quad \Delta_{KL} \leq \Delta_{EC} \quad ; \quad iv)\quad \Delta_{EC} \leq \ln(\Delta_{\chi 2} + 1) \leq \Delta_{\chi 2}. \quad \blacksquare$$

Theorem 1 establishes both power and Entropic Complexity within a well defined class of distance measures. In some respects power could be thought of as fundamental, in the sense that all other distances bound it above. In fact all measure precisely the same thing, that is how far the null density is from the alternative, Moreover all can seen to be expectations of a particular function taken with respect to the null hypothesis (since $f_v(1) = 1$). For those given in (8) that function is obvious, while for power we can write

$$\Delta_{\omega^*} = \int_{v'v=1} I_v\left(\omega^*\right)\left(f_v(1) - f_v(\alpha)\right)(dv),$$

where $I_v(\omega^*)$ is the indicator function taking a value 1 if $v \in \omega^*$ and 0 otherwise. In addition, notice that the inequalities given in Theorem 1 apply for any situation in which the density of the maximal invariant is uniform. Consequently, we should expect Δ_{EC} to be useful in other circumstances as well.

In order to illustrate the bounds given in Theorem 1, consider the following time series regression;

$$(1-\alpha l)(y_t - \beta_1 - I_\tau(\tau T)\,\beta_2 t) = \varepsilon_t \quad ; \quad t = 1,..,T, \tag{10}$$

where $I(\tau T)$ is the indicator function taking values 1 if $t \geq \tau T$ and 0 otherwise. Thus τ indexes the timing of a break in the linear trend in the regression. The values $\tau = 0, 1$ indicate respectively the cases of a full trend and no trend. Zivot and Andrews (1992) and Leybourne, Mills and Newbold (1998) have also numerically analysed the impact of the timing of breaks in a possible trend. Generally, the earlier the trend starts the lower the power against a fixed value of α under the alternative.

Although the measures of distance given in (8) and (9) are not available in closed form we may numerically evaluate the unresolved integrals via Monte Carlo. For $T = 100$ and with $100,000$ Monte Carlo replications each of these measures of distance was simulated for values of $\tau = (0, .25, .5, .75, 1)$ and $\alpha = (.9, .92, .94, .96, .98)$. Entropic Complexity was also evaluated from (7) for these values. Presented in Table 1 in Appendix II are the functions of all of these measures given in the bounds in Theorem 1 along with, for later reference, Δ_{EC} itself. Although all of the measures are nonlinear in α it is clear that they are measuring the same distance, in more-or-less the same way. Notice, also, how tight the bound $\Delta_{KL} \leq \Delta_{EC}$ is.

In addition to it being often very time consuming having to simulate those other measures of distance, having no closed form makes it impossible to determine any analytic properties. In the next section we will explore the analytic properties of Δ_{EC}, specifically how it depends upon the trending behaviour of regressors and upon the structure of serial correlation in the innovations.

3 Properties of Entropic Complexity

We seek to establish analytic links between the deterministic component and/or the autocorrelation of the errors and Δ_{EC}. Notice that these two influences enter Δ_{EC}

via the following route; A is defined by

$$A = C'\Sigma_{\alpha,\phi}C \quad ; \quad \Sigma_{\alpha,\phi} = K_{\phi}^{-1}T_1T_{\alpha}^{-1}K_{\phi}K_{\phi}'(T_{\alpha}^{-1})'T_1'(K_{\phi}^{-1})',$$

while C (defined by $M_W = CC'$ and $C'C = I_{N-k}$, where $W = K_{\phi}^{-1}T_1X$) is the singular value decomposition of the symmetric idempotent M_W. While the role of ϕ, therefore, is relatively transparent, that of d_t is less so. Hence we will capture the trending properties of the deterministics, under the following assumption:

Assumption 2 With $d = (d_1, .., d_N) = X\beta$, we assume that the i^{th} column of X is

$$X_i(p) = (1, 2^p, .., t^p, .., N^p)',$$

so that the set of regressors includes a polynomial time trend indexed by the scalar parameter p, satisfying:

(i) For every $p > 0$, X has full column rank,

(ii) No column of X, X_j with $j \neq i$, grows faster than $X_i(p)$ in t. ■

Under Assumption 2, we can focus upon the impact of polynomial time trends upon the distance. In particular, we will examine the impact of the most strongly trending regressor (for the sake of interpreting the result rather than any mathematical imperative), but must exclude $p = 0$, since we will assume the presence of a constant, in any case.

Thus we can parameterise Δ_{EC} as a function of both p and ϕ (as well as the autocorrelation coefficient α) as $\Delta_{EC}(\alpha, p, \phi)$. To fix the properties of Δ_{EC} we then require the slopes and Hessians of $\Delta_{EC}(\alpha, p, \phi)$ in both the p and ϕ directions. Before proceeding, note that A is a function of p and ϕ, through C, and in general the singular value decomposition is *not* a differentiable function. However, in this special case, we are able to prove a new result, crucial for our analysis here.

Theorem 2 *Let C be the singular value decomposition of the symmetric idempotent $CC' = M_W = I - W(W'W)^{-1}W'$ and let C_0 and W_0 define points in $\mathbb{R}^{N\times(N-k)}$ and $\mathbb{R}^{N\times k}$, then*

(i) if W is differentiable in a neighbourhood of W_0, C is differentiable in a neighbourhood of C_0,

(ii) defining the respective derivatives with respect to p and any element of ϕ, ϕ_j say,

by $\partial_p(.)$ and $\partial_{\phi_j}(.)$,we have the expressions

$$\partial_p C = W(W'W)^{-1}(d_p W)'C$$
$$\partial_{\phi_j} C = W(W'W)^{-1}(d_{\phi_j} W)'C. \quad \blacksquare \tag{11}$$

To proceed we now need to establish that $\Delta_{EC}(\alpha, p, \phi)$ is a differentiable function of both p and ϕ and then find those derivatives. By then looking at the second derivatives we find that $\Delta_{EC}(\alpha, p, \phi)$ is a (quasi) convex function over p and ϕ. Thus it is possible to find values of p and ϕ which minimise the distance, and therefore implicitly through the bounds given in Theorem 1, will ensure that power is also small. The results are presented in the following theorem.

Theorem 3 *Let Δ_{EC} be defined as in (7) and assume that Assumption 2 holds, then (i) Δ_{EC} is differentiable, and therefore continuous, with respect to p, with derivative given by*

$$\partial_p \Delta_{EC}(\alpha, p, \phi) = \frac{1}{\lambda}\left\{\operatorname{tr}\left[\left(-\lambda I_{N-k} + (N-k)A^{-1}\right)\left(C'DC\right)A^{-1}\right]\right\}, \tag{12}$$

where $\lambda = \operatorname{tr} A^{-1}$, $D = (\partial_p W)(W'W)^{-1}W'\Sigma_{\alpha,\phi}$, and

$$(\partial_p W) = T_1(\partial_p X) = T_1\left(\underline{0}, .., \underline{0}, \partial_p X_i(p), \underline{0}, .., \underline{0}\right).$$

(ii) Δ_{EC} is differentiable, and therefore continuous, with respect to $\phi = \{\phi_j\}_{j=1}^m$, with derivatives given by

$$\partial_{\phi_j} \Delta_{EC}(\alpha, p, \phi) = \frac{1}{\lambda}\left\{\operatorname{tr}\left[\left(-\lambda I_{N-k} + (N-k)A^{-1}\right)\left(C'HC\right)A^{-1}\right]\right\}, \tag{13}$$

where $H = K_\phi^{-1}L^{(i)}P_W\Sigma_{\alpha,\phi}$, and $P_W = W(W'W)^{-1}W'$.
(iii) $\Delta_{EC}(\alpha, p, \phi)$ is quasi-convex over both p and ϕ and therefore solutions, p^*, to $\partial_p \Delta_{EC}(\alpha, p, \phi) = 0$ and ϕ^* to $\partial_\phi \Delta_{EC}(\alpha, p, \phi)$ are at a minimum. $\blacksquare$

Theorem 3 implies that the distance Δ_{EC} is minimisable with respect to the degree of trending of the regressors and the autocorrelation of the innovations. Thus, via the bounds given in Theorem 1, we may also conclude that power can be made small by both these model features. Although this result has genuine theoretical significance, to illustrate the tangible effects of the different model properties we will examine each in turn.

3.1 Numerical Effects of the Polynomial Trend Degree

From Theorem 3 for any set of deterministic components d_t, including t^p, and given a particular form of moving average error autocorrelation, it is possible to obtain a $p^* = p^*(\alpha, \phi, N)$ which minimises Δ_{EC}. It does not, however, depend upon the coefficients β, in d_t, nor the variance σ^2. Since, p^* is an implicit function, we may find its slope via,

$$\frac{dp^*}{d\alpha} = -\frac{\partial^2 \Delta_{EC}}{\partial p \partial \alpha} \left(\frac{\partial^2 \Delta_{EC}}{\partial^2 p} \right)^{-1} \Bigg|_{p=p^*} . \tag{14}$$

However, (14) does not have a constant sign over $\alpha \in (-1, 1)$, and so p^* is not a monotone function of α. To illustrate, suppose that we consider a simplified version of (3) with no error autocorrelation, viz.

$$(1 - \alpha l)(y_t - \beta_1 - \beta_2 t^p) = \varepsilon_t \quad ; \quad \varepsilon_t \sim N(0, \sigma^2), \tag{15}$$

for $t = 1, .., N$. We may solve $\partial_p \Delta_{EC} = 0$, and plot the solution p^* for different sample sizes ($N = 10, 20$, and 40), giving Figure 1, in Appendix II. Notice, that for moderate sample sizes, and for alternatives 'close' to the null, Δ_{EC} is *not quite* minimised by a linear time trend.

In practice there seems little a-priori rationale for including as a regressor $t^{0.8}$, for instance. Consequently, we calculate Δ_{EC}, for models characterised by

$$(1 - \alpha l)(y_t - \beta_1 - \beta d_t^*) = \varepsilon_t \quad ; \quad \varepsilon_t \sim iid(0, \sigma^2), \tag{16}$$

and consider the following cases: (i) $d_t^* = t^{p^*}$, (where p^* is found by solving $d_p \Delta_{EC} = 0$); (ii) $d_t^* = t$ (linear trend); (iii) $d_t^* = \ln t$ (logarithmic trend); (iv) $d_t^* = t^2$ (quadratic trend) and (v) $d_t^* = 0$ (no trend). Table 2 in Appendix II, gives values for $\Delta_{EC}(\alpha)$ as α varies, for each model configuration and for sample sizes of 20 and 40.

Numerically, $p = p^*$ and $p = 1$ are barely distinguishable. While having no trend ($p = 0$) gives us the greatest ability to discriminate. These two facts simply mirror previous studies of the power of unit root tests, for example in DeJong *et al* (1992). Of some interest is that the 'ranking', in terms of the measure, is not uniform over all values of α. In summary these results compliment, and allow slightly more detailed analysis than, the related results of Phillips (1998) and Phillips and Ploberger (2003).

3.2 Numerical Effects of Innovation Autocorrelation

From an applied perspective the deterministics d_t are a choice made by the modeler to attempt to capture the trending behaviour of the data, specifically to ensure invariance with respect to those trends. On the other hand, the correlation structure of the innovations are a property of the underlying statistical process. That does not mean, however, that understanding the effect that particular autocorrelation structures have is not important.

For the purposes of numerical analysis, we again consider a simplified version of (3), namely

$$(1-\alpha l)(y_t-\beta_1-\beta_2 d_t^{**}) = (1+\phi_1 l)\varepsilon_t \quad ; \quad \varepsilon_t \sim iid(0,\sigma^2), \tag{17}$$

so that the de-trended y_t follows an $ARIMA(0,1,1)$ process. As α varies we can calculate the minimum argument ϕ_1^* for sample sizes of $N = 10, 20$ and 40 for model (17), with $d_t^{**} = t$. These values are plotted in Figure 2, in Appendix II. As we should expect it is large negative values of ϕ_1, which make the distance small. Again, the result is that it is *not quite* an MA(1) with a negative unit root which minimises the distance. Although, as in the case with a linear trend, there is some uniformity in that the value of ϕ_1^* is not particularly sensitive with respect to α. That is, we are not merely measuring a common factor effect.

To highlight the effect that different first order innovation autocorrelation has in the context of (17), we calculate Δ_{EC} for α, and for different values of ϕ_1 (namely, $\phi_1 = \phi_1^*, -1, -0.5, 0.5, 1$) and for two versions of (17) with $d_t^{**} = t$ and $d_t^{**} = 0$. The results are recorded in Table 3, for both. These tables strongly reinforce the experimental Monte Carlo evidence cited in the introduction. In addition, it is clear that an MA(1) with a negative unit root implies distances, and thus indirectly, powers exceedingly close to their minimum value.

To summarise the theoretical and numerical properties of Δ_{EC}; it is analytic and minimisable in the model features as parametrised here. Moreover, the numerical results are strongly supportive of current numerical studies, in that it is, more-or-less, linear trends and negative unit root moving averages which minimise our distance, and thus power. In the following section we'll use this knowledge to examine how model specification affects distance in practice.

4 Illustration (Nelson & Plosser Data)

We have established the validity of Δ_{EC} as a distance measure, in terms of the Gibbs and Su (2002) family and detailed two key analytic properties. In this section we will demonstrate the practical usefulness of the measure within the context of testing for a unit root in the Nelson and Plosser (1982) series of 14 macroeconomic time series. We will consider two model specifications,

$$M_1 \;:\; (1-\alpha l)\left(y_t - \beta_1 - \beta_2 t^p\right) = u_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \varepsilon_t \tag{18}$$

$$M_2 \;:\; (1-\alpha l)\left(y_t - \beta_1 - \beta_2 t\right) = u_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \varepsilon_t, \tag{19}$$

where $\varepsilon_t \sim iid(0, \sigma^2)$, l is the lag-operator and $t = 1, ..., N$. Estimation of these two models, and evaluation of Δ_{EC} at the estimated parameter values will highlight the effect that imposition of a linear trend has on our ability to perform unit root inferences. In order to estimate both M_1 and M_2 we will also need to additionally assume that the u_t are outcomes of an invertible MA(2).

The Nelson and Plosser data has been much analysed with in the literature with authors coming to different conclusions about the existence of unit roots within some of the series, for example the differing perspectives of Phillips (1991) and Dejong and Whiteman (1991). Heuristically, it seems that altering the trending behaviour of the regressors, for example the inclusion of a linear trend, the timing of any breaks in that trend, can alter the outcome of a test.

Here we characterise model M_2 as a restriction of M_1. That is in M_1 we can estimate freely, via non-linear least squares, the degree of an included time trend, while in M_2 the trend is restricted to be linear. Both models were estimated by via a combination of least squares and the Hannan-Rissanen procedure to estimate the moving average coefficients. M_2 is really the standard model estimated within this context, except that we are choosing to estimate the transfer function of the innovation sequence (u_t) with a short moving average rather than autoregression.

Full results of the estimation of M_1 for all 14 series are presented in Table 5 in the appendix. The figures below the estimated values are the estimated standard errors, obtained from the Gaussian Hessian. Noteworthy from the results are that the several of series (5 of 14) have estimated trend degrees more than two standard errors from 1. These are Real GNP, Real Wages, Unemployment, Velocity and Consumer Prices. Both Unemployment (with $\hat{p} = -.161$) and Velocity (with $\hat{p} = -.566$) would seem to have negatively powered, or evaporating trends, as detailed in Phillips (2001).

Since M_2 is a standard model the results will not be reported in full. However, it is clear that at least for some of the series imposition of $p = 1$ is not necessarily supported by the data, i.e. those mentioned in the previous paragraph. The purpose here though is to measure the impact that imposing a linear trend on the data has on our distance measure. In general entropic complexity will be a function of α, p, and $\phi = (\phi_1, \phi_2)'$. All of these parameters may be consistently estimated and since Theorem 3 ensures that Δ_{EC} is differentiable in its arguments we may consistently estimate Δ_{EC} as well, i.e.

$$\Delta_{EC}(\hat{\alpha}, \hat{p}, \hat{\phi}) \rightarrow_p \Delta_{EC}(\alpha, p, \phi).$$

We call $\hat{\alpha}_1$ and $\hat{\alpha}_2$ the estimated autoregressive coefficients for models M_1 and M_2 respectively, and similarly $\hat{\phi}_1$ and $\hat{\phi}_2$ for the estimated moving average parameters. These are given in Table 5, in Appendix II. In terms only of the estimated autoregressive parameter, with the exception of Real Wages, the effect of restricting the model to a linear trend seems negligible. However, the effect on distance, specifically the estimated distances $\Delta_{EC}(\hat{\alpha}_1, \hat{p}, \hat{\phi}_1)$ and $\Delta_{EC}(\hat{\alpha}_2, 1, \hat{\phi}_2)$ is generally much greater.

For three series, Real GNP, Real Wages and Velocity the effect of imposing a linear trend is to significantly reduce the distance of the fitted model from the unit root. From the bounds in Theorem 1 and highlighted in Table 1 we can be confident that power behaves similarly we can suggest that for these series a linear trend has a similarly dramatic negative effect on the power of unit root tests. For some series, Unemployment, the Standard & Poor 500 and Industrial Production the opposite is true, although much less dramatically. For Unemployment although imposition of a linear trend is clearly inappropriate, doing so does not seem to have serious implications for unit root testing.

The most telling individual result is that for Real Wages. The unrestricted model estimates, see Table 5, suggest values for the autoregressive coefficient and trend degree both far from unity. Imposing a linear trend though yields what appears to be a unit root. That is, far from the deterministic and stochastic trends 'competing' to explain the trending behaviour of series they can in fact combine to give an illusion of trending behaviour, when none exists. Notice that the value of $\Delta_{EC}(\hat{\alpha}_1, \hat{p}, \hat{\phi}) \approx 3.45$ corresponds, via Table 1, to situations in which power minus size (at the 5% level) is approximately 0.3, whereas imposing the linear trend yields a distance comparable to having no power at all. Qualitatively, the same can be inferred for Velocity, albeit to a slightly lesser degree.

5 Conclusions

This paper has presented an analytic closed form measure of distance for the unit root hypothesis applicable in a relatively general class of models. The link between this measure of distance and others considered by Gibbs and Su (2002) as well as power is established, so that we can be confident that is measuring exactly the same thing as, for instance, power. In addition, how the measure depends upon the key features of our time series regression; deterministic trending and autocorrelation structure, is completely transparent.

Perhaps more importantly the distance can be used to highlight exactly how sensitive our unit root inferences may be to the precise specification of the deterministic trend. It is seen that for certain series in the Nelson and Plosser (1982) Data, most strikingly for Real Wages and Velocity, constraining the trend to be linear implies an estimated model very close to a unit root process. On the other hand, freely estimating the degree of the trend implies a model very different in character.

That is, two important features have been highlighted. First, for macroeconomic series trends other than linear ones seem to have statistical relevance. Being analytic the proposed measure is more suited to handling the implied complexity than current Monte Carlo based results. Second, imposition of an inappropriate linear trend can have serious consequences in terms of our ability to perform unit root inferences.

References

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

Bhargava, A. (1986): On the theory of testing for unit roots in observed time series. *Review of Economic Studies*, **53**, 369–384.

Bozdogan, H. (1990): On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, **19**, 221-278.

Campbell, J.Y. and Perron, P. (1991): Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots. *NBER Technical Working Papers,* 0100.

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

Dejong, D.N. and Whiteman, C.H. (1991): The case for trend-stationarity is stronger than we thought. *Journal of Applied Econometrics,* **6**, 413-421.

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

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

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

Forchini, G.F. (2002): The exact cumulative distribution function of a ratio of quadratic forms in normal variables, with application to the AR(1) model. *Econometric Theory*, **18**, 823–852.

Gibbs, A.L. and Su, F.E. (2002): Choosing and bounding probability metrics. *International Statistical Review*, **70**, 419-435.

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

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

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

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

Nelson, C.R. and Plosser, C.I. (1982): Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, **10**, 139-162.

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. (1991): To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics*, **6**, 333-364.

Phillips, P.C.B. (1998): New tools for understanding spurious regressions. *Econometrica,* **66**, 1299–1325.

Phillips, P.C.B. (2001): Regression with slowly varying regressors. *CFDP*, no. **1310**, Yale University.

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

Phillips, P.C.B. and Ploberger, W. (1994): Posterior odds testing for a unit root with data-based model selection. *Econometric Theory*, **10**, 774–808.

Phillips, P.C.B. and Ploberger, W. (2003): Empirical limits for time series econometric models. *Econometrica* **71**, 627–673.

Phillips, P.C.B & Xiao, Z. (1998): A primer on unit root testing. *Journal of Economic Surveys*, **12**, 423-470.

Poskitt, D. S. (1987): Precision, complexity and Bayesian model determination. *Journal of the Royal Statistical Society B*, **49**, 199–208.

Wolkowicz, H. and Styan, G.P.H. (1980): Bounds for eigenvalues using traces. *Linear Algebra and its Applications*, **29**, 471-506.

Würtz, A. (1997): A universal upper bound on power functions. *University of New South Wales, Discussion Paper,* **97/17**.

Zivot, E. and Andrews, D.W.K. (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

I. Proofs

Proof of Theorem 1:

The first inequality is established in Würtz (1997) and the second is well known, see Gibbs and Su (2002). For the third inequality we have

$$\begin{aligned}
\Delta_{KL} &= \int_{v'v=1} \ln\left(f_v(\rho)^{-1}\right)(dv) = \int_{v'v=1} \ln\left(|A|^{1/2}(v'A^{-1}v)^{(N-k)/2}\right)(dv) \\
&= \frac{1}{2}\ln\det A + \frac{(N-k)}{2}\int_{v'v=1} \ln\left(v'A^{-1}v\right)(dv) \\
&\leq \frac{1}{2}\ln\det A + \frac{(N-k)}{2}\ln\int_{v'v=1} \left(v'A^{-1}v\right)(dv)
\end{aligned}$$

then by Jensen's inequality since ln(.) is concave, and since

$$\int_{v'v=1} \left(v'A^{-1}v\right)(dv) = \frac{Tr[A^{-1}]}{N-k}$$

the inequality is proved. For the fourth we have

$$\Delta_{\chi^2} = -1 + \int_{v'v=1} |A|^{1/2} \left(v'A^{-1}v\right)^{(N-k)/2} (dv).$$

Considering just the integral then, for $N - k > 2$

$$\int_{v'v=1} \left(v'A^{-1}v\right)^{(N-k)/2} (dv) \geq \left(\int_{v'v=1} \left(v'A^{-1}v\right) (dv)\right)^{(N-k)/2},$$

again by Jensen's inequality, the fourth inequality follows via

$$\begin{aligned}
\ln(\Delta_{\chi^2} + 1) &= \frac{1}{2}\ln\det A + \ln\left(\int_{v'v=1} \left(v'A^{-1}v\right)^{(N-k)/2} (dv)\right) \\
&\geq \frac{1}{2}\ln\det A + \frac{(N-k)}{2}\ln\int_{v'v=1} \left(v'A^{-1}v\right)(dv), \\
&= \Delta_{EC},
\end{aligned}$$

and $\ln(r+1) \leq r$ gives the final inequality. ■

Proof of Theorem 2

Since $W = K_\phi^{-1}T_1X$, then immediately W is differentiable with respect to ϕ_j. Now under Assumption 1, since $p > 0$, then the rank of X is constant, and so X is differentiable with respect to p. Consequently, the rank of W is constant and therefore W is also differentiable with respect to p, with differential $\partial W = K_\phi^{-1}T_1(\partial X)$. In fact W is an analytic (matrix) function of both p and ϕ.

To establish differentiability of C (with respect to either parameter) we note that C is defined as the singular value decomposition of $M_W = I_N - W(W'W)^{-1}W'$, and is therefore the unique solution (up to orthogonal transformation), in $R^{N\times(N-k)}$, to the equations

$$M_W = CC' \quad \text{and} \quad C'C = I_{N-k}. \tag{20}$$

We first show that (20) implies and is implied by

$$W'C = \mathbf{0} \quad \text{and} \quad C'C = I_{N-k}. \tag{21}$$

To do this note that

$$M_W = CC' \Longleftrightarrow (I_N - M_W)C = \mathbf{0}, \tag{22}$$

and define

$$P_W = I - M_W = W(W'W)^{-1}W' = WW^+ = \left(W^+\right)' W',$$

where W^+ denotes the Moore-Penrose inverse of which exists and is unique since the rank of W is constant under Assumption 2. Rewriting (22) as $(W^+)' W'C = \mathbf{0}$, then since

$$W^+ = (W'W)^+W' \quad \text{and} \quad W = W(W'W)^+(W'W),$$

we have

$$(W^+)' W'C = \mathbf{0} \Longleftrightarrow (W'W)^+W'C = \mathbf{0},$$

which leads to

$$(W'W)^+W'C = \mathbf{0} \Longleftrightarrow (W'W)(W'W)^+W'C = \mathbf{0} \Longleftrightarrow W'C = \mathbf{0},$$

as required.

To continue, define the matrix valued function $h : \mathbb{R}^{N\times k} \times \mathbb{R}^{N\times(N-k)} \to \mathbb{R}^{N\times(N-k)}$ of C and W by

$$h(C,W) = \begin{pmatrix} W'C \\ C'C - I_{N-k} \end{pmatrix},$$

then following a similar argument to Magnus and Neudecker (1988), Theorem 8.7, h is differentiable on $\mathbb{R}^{N\times k} \times \mathbb{R}^{N\times(N-k)}$. Letting the point C_0, W_0 in $\mathbb{R}^{N\times k} \times \mathbb{R}^{N\times(N-k)}$ satisfy

$$h(C_0, W_0) = \mathbf{0},$$

and further

$$\det[J_0] = \det\left[\left.\frac{h(C,W)}{dC}\right|_{C_0,W_0} \right] = \det\begin{bmatrix} W_0' \\ 2C_0' \end{bmatrix} \neq 0,$$

since by definition $W'C = 0$, then the conditions for the Implicit Function Theorem are met (see Theorem A.3, Section 7, Magnus & Neudecker (1988)). Consequently, there exists a neighbourhood in $\mathbb{R}^{N\times k}$, $V(W_0)$ and a unique (up to orthogonal transformation) matrix valued function $C : V(W_0) \to \mathbb{R}^{N\times(N-k)}$ for which the following statements hold:

(a) C is differentiable on $V(W_0)$

(b) $C(W_0) = C_0$, and

(c) $W'C = 0$ and $C'C = I_{N-k}$ for all $W \in V(W_0)$,

which concludes the proof of part (i).

For part (ii) we require an explicit relationship between the differential of C and that of W. From (21) we have

$$W'C = \mathbf{0},$$

so that denoting the differentials of W and C by ∂W and ∂C respectively (suppressing for the moment which variable we are differentiating with respect to), we have

$$(\partial W)'C + W'(\partial C) = \mathbf{0},$$

giving

$$(W')^{+}W(\partial C) = (W')^{+}(\partial W)'C.$$

Consider the matrix defined by

$$P = (W')^{+}W + CC' = P_W + M_W = I_{N-k},$$

and so

$$\partial C = P(\partial C) = \left((W')^{+}W + CC'\right)(\partial C) = (W')^{+}W(\partial C),$$

since $C'(\partial C) = 0$. Consequently, the relevant expression for the differential of C is

$$\partial C = (W')^{+}(\partial W)'C = W(W'W)^{-1}(\partial W)C,$$

which then gives the expressions in (11). ■

Proof of Theorem 3

For part (i) we have

$$\Delta_{EC} = \frac{1}{2}\ln\det A + \frac{(N-k)}{2}\ln\frac{\operatorname{tr} A^{-1}}{(N-k)},$$

where A is a function of p. In order to establish differentiability we utilise *Cauchy's rule of invariance* for (possibly) matrix valued functions of matrix arguments. If F is differentiable at D and G is differentiable at $E = F(D)$, then the composite function, defined by

$$H(D, U) = G \circ F,$$

is differentiable for all $n \times m$ matrices U and

$$\partial H(D, U) = \partial G(E; \partial F(D; U)).$$

From Theorem 2, C is differentiable with respect to p and so differentiability of A immediately follows, and consequently of $\Delta_{EC}(\alpha)$. Since also $A = C'\Sigma_{\alpha,\phi}C$, we have

$$\partial_p A = \left[(\partial_p C)'\Sigma_{\alpha,\phi}C + C'\Sigma_{\alpha,\phi}(\partial_p C)\right], \tag{23}$$

so that substitution of (11) into (23), yields

$$\partial_p A = -C'[D + D']C, \tag{24}$$

where $D = (\partial_p W)(W'W)^{-1}W'\Sigma_{\alpha,\phi}$. Finally, noting the following standard differentials,

$$\partial_p \ln\det A = \mathrm{tr}[A^{-1}\partial_p A], \quad \partial_p \ln(\mathrm{tr}\, A^{-1}) = \frac{\mathrm{tr}\left(\partial_p A^{-1}\right)}{\mathrm{tr}\, A^{-1}} \quad \& \quad \partial_p A^{-1} = -A^{-1}(\partial_p A)A^{-1}$$

so that

$$\partial_p \Delta_{EC} = \frac{1}{2}\mathrm{tr}[A^{-1}\partial_p A] - \frac{(N-k)}{2}\frac{\mathrm{tr}\left(A^{-1}(\partial_p A)A^{-1}\right)}{\mathrm{tr}\, A^{-1}}, \tag{25}$$

and letting $\lambda = \mathrm{tr}\, A^{-1}$, substituting (24) into (25) and rearranging proves part (i).

For part (ii) differentiability is established in exactly the same way as in part (i). The required derivative of $\Delta_{EC}(\alpha)$ is

$$\partial_{\phi_j}\Delta_{EC} = \frac{1}{2}\mathrm{tr}(A^{-1}(\partial_{\phi_j} A)) - \frac{(N-k)}{2\lambda}\mathrm{tr}\left(A^{-1}(\partial_{\phi_j} A)A^{-1}\right), \tag{26}$$

where again $\lambda = \mathrm{tr}\, A^{-1}$. For this case the derivative of A is

$$\partial_{\phi_j} A = (\partial_{\phi_j} C)'\Sigma_{\alpha,\phi}C + C'(\partial_{\phi_j}\Sigma_{\alpha,\phi})C + C'\Sigma_{\alpha,\phi}(\partial_{\phi_j} C), \tag{27}$$

however, from the definition of $\Sigma_{\alpha,\phi}$, $\partial_{\phi_j}\Sigma_{\alpha,\phi} = 0$, so that the second term in (27) vanishes. From (11), we have

$$\partial_{\phi_j} C = W(W'W)^{-1}(\partial_{\phi_j} W)C,$$

where

$$\begin{aligned} \partial_{\phi_j} W &= \partial_{\phi_j}(K_\phi^{-1}T_1X) = -K_\phi^{-1}L^{(i)}K_\phi^{-1}T_1X \\ &= -K_\phi^{-1}L^{(i)}W, \end{aligned}$$

so that

$$\partial_{\phi_j} C = -P_W(L^{(i)})'(K_\phi^{-1})'C,$$

and hence

$$\partial_{\phi_j} A = C'(H+H')C, \tag{28}$$

where $H = K_\phi^{-1}L^{(i)}P_W\Sigma_{\alpha,\phi}$, so that substituting (28) into (26) and rearranging gives the required derivative.

For part (iii), consider first the derivatives with respect to p. Let $\gamma_i = 1/\lambda_i$, so that $0 < \gamma_1 < \gamma_2 < .. < \gamma_{N-k}$ are the ordered eigenvalues of A^{-1}, and let $c_N = -(N-k)/2\ \ln(N-k)$, so that we may write

$$\Delta_{EC} = -\frac{1}{2}\sum_{i=1}^{N-k}\ln\gamma_i + \frac{(N-k)}{2}\ln\sum_{i=1}^{N-k}\gamma_i + c_N \tag{29}$$

Further, letting $\Delta_{EC} = \Delta(\gamma_1(p), .., \gamma_{N-k}(p))$, so that Δ_{EC} is a function of p only through the eigenvalues of A^{-1} and so

$$\frac{\partial^2 \Delta_{EC}}{\partial p^2} = \sum_{i=1}^{N-k} \left(\frac{\partial^2 \Delta[\gamma]}{\partial \gamma_i^2} \left(\frac{\partial \gamma_i}{\partial p} \right)^2 + \frac{\partial \Delta[\gamma]}{\partial \gamma_i} \frac{\partial^2 \gamma_i}{\partial p^2} \right). \tag{30}$$

The relevant partial derivatives in (30) are given by

$$\frac{\partial \Delta[\gamma]}{\partial \gamma_i} = \frac{-1}{2\gamma_i} + \frac{(N-k)}{2\sum_{i=1}^{N-k} \gamma_i}, \tag{31}$$

$$\frac{\partial^2 \Delta[\gamma]}{\partial \gamma_i^2} = \frac{1}{2\gamma_i^2} - \frac{(N-k)}{2(\sum_{i=1}^{N-k} \gamma_i)^2}, \tag{32}$$

and if we define γ_i and r_i, with $r_i' r_i = 1$, as the $N-k$ solutions to $A^{-1} r = \gamma r$, then applying Theorems 8.7 and 8.10 of Magnus and Neudecker (1999), we have

$$\frac{\partial \gamma_i}{\partial p} = r_i'(\partial_p A^{-1}) r_i = r_i' A^{-1}(-\partial_p A) A^{-1} r_i = \gamma_i r_i'(-\partial_p A) A^{-1} r_i, \tag{33}$$

$$\begin{aligned} \frac{\partial^2 \gamma_i}{\partial p^2} &= 2 r_i'(\partial_p A^{-1})(\gamma_i I - A^{-1})^+ (\partial_p A^{-1}) r_i \\ &= 2\gamma_i^2 r_i'(-\partial_p A) A^{-1} (\gamma_i I - A^{-1})^+ A^{-1}(-\partial_p A) r_i, \end{aligned} \tag{34}$$

where $(\gamma_i I - A^{-1})^+$ is the Moore-Penrose inverse of the rank $N-k-1$ matrix $\gamma_i I - A^{-1}$.

Consequently, substituting (31), (32), (33) and (34) into (30), and noting that $\sum_{i=1}^{N-k} \gamma_i = \operatorname{tr} A^{-1} = \lambda$, as in the statement of part (i), the second derivative is

$$\begin{aligned} \frac{\partial^2 \Delta_{EC}(\alpha)}{\partial p^2} = \sum_{i=1}^{N-k} \Bigg[& \left(\frac{1}{2} - \frac{\gamma_i^2 (N-k)}{2\lambda^2} \right) \left(r_i'(-\partial_p A) A^{-1} r_i \right)^2 \\ & + \left(-1 + \frac{\gamma_i (N-k)}{\lambda} \right) \gamma_i h_i' (\gamma_i I - A^{-1})^+ h_i \Bigg], \end{aligned} \tag{35}$$

where $h_i = A^{-1}(-\partial_p A) r_i$.

We can write (35) as

$$\frac{d^2 \Delta_{EC}(\alpha)}{dp^2} = F + G, \tag{36}$$

and consider F and G separately. Write F as

$$F = \frac{1}{\lambda^2} \left[\lambda^2 \sum_{i=1}^{N-k} \left(r_i'(-\partial_p A) A^{-1} r_i \right)^2 - (N-k) \sum_{i=1}^{N-k} \gamma_i^2 \left(r_i'(-\partial_p A) A^{-1} r_i \right)^2 \right],$$

so that $F \geq 0$ if $\gamma_i \leq \lambda/(N-k)^{1/2}$ for every i. From Wolkowicz and Styan (1980) the maximum eigenvalue of A^{-1} satisfies

$$\frac{\operatorname{tr}(A^{-1})}{N-k} \leq \gamma_{N-k} \leq \frac{\operatorname{tr}(A^{-1})}{N-k} + \left(\frac{N-k-1}{N-k} \right)^{1/2} \left(\operatorname{tr}(A^{-2}) - \frac{\operatorname{tr}(A^{-1})}{N-k} \right)^{1/2},$$

which gives,

$$\left(\gamma_{N-k} - \frac{\mathrm{tr}(A^{-1})}{N-k}\right)^2 \leq \left(\frac{N-k-1}{N-k}\right)\left(\mathrm{tr}(A^{-2}) - \frac{\mathrm{tr}(A^{-1})}{N-k}\right). \tag{37}$$

Using the inequalities

$$\ln(\det A^{-1}) \leq \mathrm{tr}(A^{-1}) - (N-k) \quad ; \quad \ln(\det A^{-2}) \leq \mathrm{tr}(A^{-2}) - (N-k),$$

and noting $\det A^{-1} \leq 1$, we have $\mathrm{tr}(A^{-2}) \leq \mathrm{tr}(A^{-1})$, which upon substitution into (37) then gives

$$\begin{aligned}\left(\gamma_{N-k} - \frac{\mathrm{tr}(A^{-1})}{N-k}\right)^2 &\leq \left(\frac{N-k-1}{N-k}\right)\mathrm{tr}(A^{-1}) - (N-k-1)\frac{\mathrm{tr}(A^{-1})^2}{(N-k)^2} \\ &\leq \left(\frac{N-k-1}{(N-k)^2}\right)\left(\frac{(N-k)-\mathrm{tr}(A^{-1})}{\mathrm{tr}(A^{-1})}\right)(\mathrm{tr}(A^{-1}))^2. \end{aligned} \tag{38}$$

Consider now the inequalities,

$$\ln(\det A^{-1}) \leq \mathrm{tr}(A^{-1}) - (N-k) \quad ; \quad \ln(\det A) \leq \mathrm{tr}(A) - (N-k),$$

which together imply

$$\mathrm{tr}(A^{-1}) + \mathrm{tr}(A) \geq 2(N-k),$$

and moreover

$$\begin{aligned}\mathrm{tr}(A) &= \mathrm{tr}(C'\Sigma_{\alpha,\phi}C) = \mathrm{tr}(\Sigma_{\alpha,\phi}M_W) \\ &\leq \mathrm{tr}(\Sigma_{\alpha,\phi}) \leq \mathrm{tr}(T_1 T_\alpha^{-1}(T_\alpha^{-1})' T_1) \\ &= N\left(\frac{2+\alpha}{1+\alpha^2}\right) - \frac{1-\alpha^{2N}}{1+\alpha^2} \\ &\leq N\left(\frac{2+\alpha}{1+\alpha^2}\right), \end{aligned} \tag{39}$$

since $\alpha \in (-1, 1]$. As a consequence of (39), we have, for A^{-1}

$$\mathrm{tr}(A^{-1}) \geq 2(N-k) - N\left(\frac{2+\alpha}{1+\alpha^2}\right),$$

which implies that the inequality in (37) can be replaced with

$$\left(\gamma_{N-k} - \frac{\mathrm{tr}(A^{-1})}{N-k}\right)^2 \leq \left(\frac{N-k-1}{(N-k)^2}\right)\left(\frac{(N-k) - N\left(\frac{2+\alpha}{1+\alpha^2}\right)}{2(N-k) - N\left(\frac{2+\alpha}{1+\alpha^2}\right)}\right)(\mathrm{tr}(A^{-1}))^2,$$

and again since $\alpha \in (-1, 1]$

$$\frac{(N-k) - N\left(\frac{2+\alpha}{1+\alpha^2}\right)}{2(N-k) - N\left(\frac{2+\alpha}{1+\alpha^2}\right)} \leq 1.$$

Finally, since

$$\left(\frac{N-k-1}{(N-k)^2}\right) \leq \left(\frac{\sqrt{N-k}-1}{N-k}\right)^2,$$

then

$$\left(\gamma_{N-k} - \frac{\mathrm{tr}(A^{-1})}{N-k}\right) \leq \left(\frac{\sqrt{N-k}-1}{N-k}\right) \mathrm{tr}(A^{-1}),$$

and rearranging this inequality gives, as is required,

$$\gamma_{N-k} \leq \frac{\mathrm{tr}(A^{-1})}{\sqrt{N-k}},$$

so that in (36) $F \geq 0$. Equally, we may write G as

$$G = \frac{1}{\lambda}\left[-\lambda \sum_{i=1}^{N-k} h_i'(\gamma_i I - A^{-1})^+ h_i + \sum_{i=1}^{N-k} \gamma_i h_i'(\gamma_i I - A^{-1})^+ h_i\right],$$

and let

$$R'A^{-1}R = \Lambda = \mathrm{diag}(\gamma_i) \quad ; \quad R'R = I_{N-k},$$

so that

$$\begin{aligned} h_i'(\gamma_i I - A^{-1})^+ h_i &= (Rh_i)'(\gamma_i I - \Lambda)^+ Rh_i \\ &\geq 0 \quad \text{if } \gamma_i \geq \lambda/(N-k) \\ &< 0 \quad \text{otherwise.} \end{aligned}$$

If we let t^* be such that $\gamma_i \leq \lambda/(N-k)$ for $i \leq t^*$, then

$$\begin{aligned} G &= \frac{1}{\lambda}\left[-\lambda \sum_{i=1}^{t^*} h_i'(\gamma_i I - A^{-1})^+ h_i + \sum_{i=1}^{t^*} \gamma_i h_i'(\gamma_i I - A^{-1})^+ h_i\right] \\ &\quad + \frac{1}{\lambda}\left[-\lambda \sum_{i=t^*+1}^{N-k} h_i'(\gamma_i I - A^{-1})^+ h_i + \sum_{i=t^*+1}^{N-k} \gamma_i h_i'(\gamma_i I - A^{-1})^+ h_i\right] \\ &\geq 0. \end{aligned}$$

Hence Δ_{EC} is quasi-convex over p, and so any solution to part (iii) is proved. Since (35) depends only on the square of the derivative of A, then so do both F and G as defined above. Consequently, exactly the same result holds for the second derivative with respect to any ϕ_j. That is Δ_{EC} is also quasi-convex over ϕ. Hence any solutions, p^*, to $\partial_p \Delta_{EC}(\alpha, p, \phi) = 0$ and ϕ^* to $\partial_\phi \Delta_{EC}(\alpha, p, \phi)$ must be at a minimum. ■

II. Tables and Graphs

Table 1: Illustration of the Bounds for the distances given in Theorem 1.
Simulation of Δ_ω, Δ_{TV}, Δ_{KL} and Δ_{χ_2} is based on 100000 replications of regression (10).

τ	α	Δ_ω	Δ_{TV}	$\sqrt{\frac{\Delta_{KL}}{2}}$	$\sqrt{\frac{\Delta_{EC}}{2}}$	$\sqrt{\frac{\ln(\Delta_{\chi_2}+1)}{2}}$	Δ_{EC}
	.90	.271	.553	1.05	1.07	3.66	2.294
	.92	.183	.450	.816	.822	2.74	1.351
0.00	.94	.096	.338	.572	.570	2.14	0.649
	.96	.046	.207	.331	.325	.866	0.211
	.98	.017	.075	.118	.110	.604	0.024
	.90	.319	.588	1.22	1.25	4.68	3.135
	.92	.206	.500	.989	1.00	3.85	2.011
0.25	.94	.121	.395	.734	.739	3.30	1.092
	.96	.060	.271	.476	.466	2.03	0.435
	.98	.019	.131	.213	.205	.715	0.084
	.90	.402	.650	1.52	1.61	5.73	5.184
	.92	.279	.568	1.28	1.34	4.75	3.590
0.50	.94	.168	.477	1.01	1.03	4.08	2.156
	.96	.089	.357	.694	.704	3.13	0.993
	.98	.031	.204	.347	.350	1.54	0.245
	.90	.520	.725	1.62	1.75	6.04	8.236
	.92	.373	.659	1.41	1.51	5.67	5.966
0.75	.94	.246	.567	1.17	1.37	5.01	3.801
	.96	.127	.450	.947	.972	3.75	1.891
	.98	.048	.276	.501	.507	2.29	0.515
	.90	.696	.793	1.90	2.02	4.98	12.33
	.92	.546	.733	1.62	1.72	4.39	9.199
1.00	.94	.370	.649	1.31	1.23	4.30	6.088
	.96	.193	.534	.848	.889	3.19	3.167
	.98	.070	.347	.463	.474	1.93	0.901

Fig.1: p^* derived for model (15) and

for $T = 10$ (—), $T = 20$ ($\cdots$) and $T = 40$ (- - -).

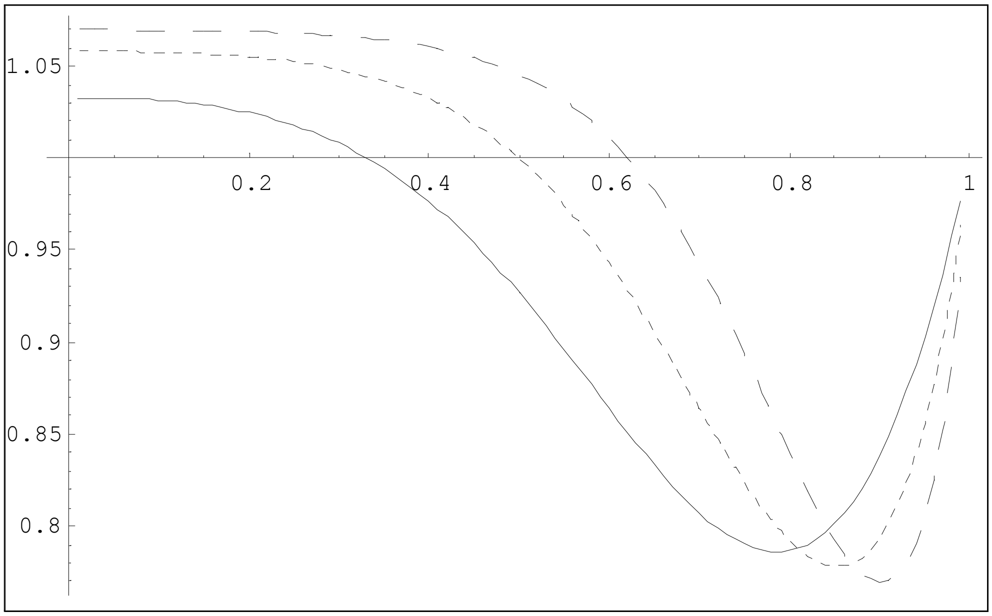

Table 2: Values for $\Delta_{EC}(\alpha)$, given model (16)

for $N = 20$ and 40 and across the different configurations.

	α	$d_t^* = t^{p^*}$	$d_t^* = t$	$d_t^* = \ln t$	$d_t^* = t^2$	$d_t^* = 0$
$N = 20$	0.800	0.091	0.096	0.179	0.208	1.203
	0.850	0.037	0.040	0.094	0.114	0.815
	0.900	0.009	0.011	0.043	0.046	0.439
	0.950	0.001	0.001	0.015	0.010	0.131
	α	$d_t^* = t^{p^*}$	$d_t^* = t$	$d_t^* = \ln t$	$d_t^* = t^2$	$d_t^* = 0$
$N = 40$	0.800	0.735	0.751	1.214	1.175	4.408
	0.850	0.345	0.361	0.644	0.677	3.070
	0.900	0.107	0.116	0.277	0.293	1.751
	0.950	0.011	0.013	0.088	0.062	0.559

Fig 2: ϕ_1^* derived for model (17) with $d^{**} = t$ and

for $T = 10$ (—), $T = 20$ ($\cdots$) and $T = 40$ (- - -).

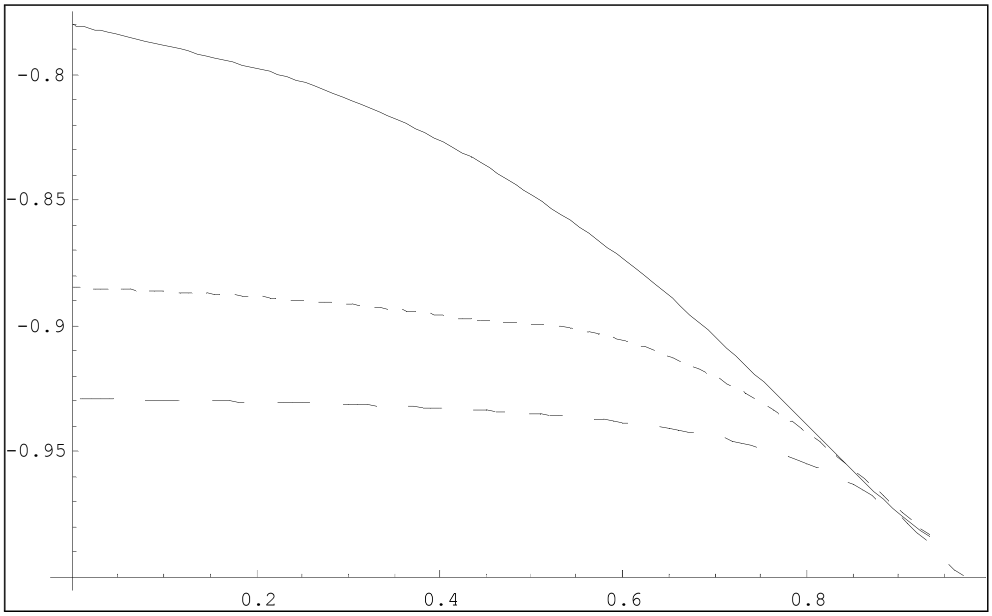

Table 3: Values for $\Delta_{EC}(\alpha)$, given model (17)

for $N = 20$ and 40, $d_t^{**} = t$, and for different MA(1) parameter values

	α	$\phi_1 = \phi_1^*$	$\phi_1 = -1$	$\phi_1 = -0.5$	$\phi_1 = 0.5$	$\phi_1 = 1$
$N = 20$	0.800	0.032	0.033	0.081	0.174	0.199
	0.850	0.011	0.011	0.034	0.072	0.081
	0.900	0.002	0.002	0.010	0.019	0.022
	0.950	0.000	0.000	0.001	0.002	0.002
	α	$\phi_1 = \phi_1^*$	$\phi_1 = -1$	$\phi_1 = -0.5$	$\phi_1 = 0.5$	$\phi_1 = 1$
$N = 40$	0.800	0.322	0.334	0.673	1.130	1.275
	0.850	0.123	0.127	0.325	0.538	0.590
	0.900	0.029	0.030	0.108	0.170	0.182
	0.950	0.002	0.002	0.012	0.019	0.019

Table 4: Estimated values for the parameters in (18) for the Nelson & Plosser data set. Figures in parentheses are estimated standard errors.

Series\Estimators	$\hat{\alpha}_1$	$\hat{p}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$
Real GNP	1.005	-0.011	0.572	0.172	0.399	-0.227
$N = 80$	(.048)	(.430)	(1.45)	(.797)	(.097)	(.073)
Nom. GNP	0.982	0.825	0.532	0.340	0.470	-0.265
$N = 80$	(.091)	(.103)	(1.03)	(.046)	(.095)	(.076)
GNP per ca.	1.003	0.518	0.560	-0.053	0.411	-0.238
$N = 80$	(.047)	(.406)	(1.13)	(1.44)	(.098)	(.070)
Bond Yield	0.954	1.466	0.317	0.012	0.136	-0.106
$N = 89$	(.077)	(.805)	(1.637)	(.737)	(.095)	(.087)
Nom. Wage	0.980	0.813	0.577	0.226	0.450	-0.201
$N = 89$	(.053)	(.066)	(.027)	(.054)	(.087)	(.064)
Real Wage	0.880	0.305	0.807	0.909	0.288	-0.094
$N = 89$	(.020)	(.024)	(.052)	(.084)	(.097)	(.079)
Unemploy.	0.750	-0.161	1.157	1.118	0.177	-0.168
$N = 99$	(.017)	(.105)	(.141)	(.146)	(.093)	(.072)
Employ.	1.001	0.810	0.537	0.010	0.405	-0.229
$N = 99$	(.043)	(.175)	(.702)	(.234)	(.085)	(.065)
GNP Defl.	0.978	0.972	0.400	0.059	0.412	-0.151
$N = 100$	(.040)	(.130)	(.053)	(.029)	(.079)	(.066)
Money	0.979	0.999	0.274	0.072	0.555	-0.216
$N = 100$	(.035)	(.104)	(.302)	(.034)	(.067)	(.065)
S&P500	0.946	1.168	0.335	0.019	0.200	-0.133
$N = 118$	(.046)	(.211)	(.053)	(.022)	(.078)	(.068)
Velocity	0.962	-0.566	0.556	0.084	0.123	-0.079
$N = 120$	(.036)	(.483)	(.272)	(.408)	(.086)	(.079)
Ind. Prod.	0.884	1.104	0.344	0.024	0.080	-0.071
$N = 129$	(.048)	(.084)	(.028)	(.008)	(.087)	(.077)
C.P.I.	0.975	0.571	-3.580	0.563	0.489	-0.188
$N = 129$	(.032)	(.080)	(.020)	(.101)	(.068)	(.057)

Table 5: $\Delta_{EC}\left(\hat{\alpha}_1, \hat{p}, \hat{\phi}_1\right)$ and $\Delta_{EC}\left(\hat{\alpha}_2, 1, \hat{\phi}_2\right)$ derived from estimating (18) and (19) for each of the series in the Nelson and Plosser data.

	R.GNP	N.GNP	GNP.p.c.	I.R.	N.Wage	R.Wage	Unemp.
$\Delta_{EC}\left(\hat{\alpha}_1, \hat{p}, \hat{\phi}_1\right)$	.0725	.0070	.0011	.3537	.0146	3.4545	23.457
$\Delta_{EC}\left(\hat{\alpha}_2, 1, \hat{\phi}_2\right)$	.0006	.0009	.0002	.2876	.0020	2.6×10^{-6}	14.397
$\hat{\alpha}_1$	1.005	0.982	1.003	0.954	0.980	0.880	0.750
$\hat{\alpha}_2$	0.991	0.990	0.993	0.950	0.990	0.998	0.751

	Employ.	GNP Defl.	Money	S&P500	Velocity	Ind. Prod.	C.P.I.
$\Delta_{EC}\left(\hat{\alpha}_1, \hat{p}, \hat{\phi}_1\right)$	2.4×10^{-6}	0.0346	0.0297	0.8733	2.0719	5.9330	0.1196
$\Delta_{EC}\left(\hat{\alpha}_2, 1, \hat{\phi}_2\right)$	4.1×10^{-6}	0.0322	0.0522	1.5254	0.0035	9.0406	0.0342
$\hat{\alpha}_1$	1.001	0.9778	0.979	0.946	0.962	0.884	0.975
$\hat{\alpha}_2$	0.998	0.9784	0.975	0.930	1.011	0.854	0.983