The Properties of Entropy for the Unit Root Hypothesis[1]

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

September 6th 2007

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

Abstract

This paper details the differential and numeric properties of two measures of entropy, Shannon entropy and Kullback-Leibler distance, applicable for the unit root hypothesis. It is found that they are differentiable functions of the degree of trending in any included deterministic component and of the correlation of the underlying innovations. Moreover, Shannon entropy is concave in these, and thus maximisable. Kullback-Leibler is instead convex, and thus minimizable. It is explicitly confirmed, therefore, that it is approximately linear trends and negative unit root moving average innovations which minimize the efficacy of unit root inferential tools. Moreover, applied to the Nelson and Plosser macroeconomic series the effect that the inclusion, or not, of a linear trend, for example, is explicitly quantified.

1 Introduction

Despite tremendous progress in understanding the properties of unit root time series and tests thereof, analytic results in closed form are extremely rare. Exceptions are the distributional results of Abadir (1993), Phillips and Ploberger (1994) and more recently, Phillips and Magdalinos (2007). To see why detailing analytic properties may be important, consider the set-up standard in the literature, with data y_t depending upon a deterministic component d_t, an error u_t and stationary innovation ζ_t, according to

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

Extensive numerical evidence in the literature has demonstrated that both the trending characteristics of d_t and the correlation properties of ζ_t can have a significant impact upon the properties, notably the power, of standard unit root tests. See, amongst many others, Durlauf and Phillips (1988), Perron (1989), Perron (1989), Zivot and Andrews (1992), Elliott, Rothenberg and Stock (1996), Leybourne, Mills and Newbold (1998), Phillips and Xiao (1998, §4) and Harvey, Leybourne and Taylor (2007).

The breadth of plausible model configurations continually expands, for example now incorporating multiple breaks in the trend, non-linear or slowly varying trends, as in Phillips (2001). Consequently, we increasingly require analytic results, better able to cope with the increasing demands that consideration of such diverse configurations will imply. This paper details the differential and numeric properties of two measures of entropy applicable for the unit root problem. Specifically, Shannon entropy and relative entropy, i.e. the Kullback-Leibler distance, are derived for the GLS detrended data, the basis for the most often used statistical procedures. We show that these measures are differentiable functions of the degree of trending in d_t and, parametrizing ζ_t as a moving average, of the moving average parameters. Moreover, it is shown that Shannon Entropy is concave and hence maximizable and Kullback-Leibler is convex and minimizable in both these model features.

This paper seeks to complement recent findings which provide explicit measures of the efficacy of statistical inference for non-stationary or unit root time series. Phillips (1998) showed that regressing a unit root time series on a sequence of polynomial trends will be ultimately successful, generalizing the concept of the spurious regression of Granger and Newbold (1974) and Phillips (1986). Phillips and Ploberger (2003) derive a distance between a fitted empirical model and the true generating process and show that including nonstationary regressors implies a higher loss. Marsh (2007) found that the information contained in any statistic invariant to a linear trend was zero at the unit root.

The results here imply that there are model configurations which explicitly limit the discriminatory power any inferential tool has, as measured by entropy. Specifically, it is found that trends which are approximately linear and approximately a negative unit root moving average minimize relative entropy (or maximize Shannon entropy). These results thus both confirm the outcomes of the published experimental evidence and provide the possibility of predicting that of future study. In addition, by estimating both entropic quantities for the often studied Nelson and Plosser (1982) data set we can measure the precise effect that choices of model configuration has on applied studies.

Indeed it is found that for the series Unemployment, Velocity and Industrial production the decision of whether or not to include a trend, such as in the procedure detailed in Harvey, Leybourne and Taylor (2007), has a profound effect on the entropy of the empirical model. Most striking is that for industrial production if a trend is not included relative entropy is zero (and smallest amongst the series), while if a trend is included relative entropy is 9.6 (and second largest amongst the series). This is clearly useful information to have, not necessarily regarding the actual choice of model, but in ascribing confidence in the resulting outcomes of tests in the chosen model.

The plan for the paper is as follows. The next section details the assumptions under which entropy can be explicitly derived, and derives and comments upon the measures. Section 3 details the differential properties while Section 4 details the numerical properties, and applies the measures for the Nelson and Plosser (1982) series highlighting the practical importance of the results. Conclusions are followed by an appendix which contains all the proofs and all tables and graphs used in the analysis.

2 Preliminary Results

In this section we formalize the class of models under consideration and derive both Shannon Entropy (SE) and Relative Entropy, i.e. the Kullback-Leibler distance (KL). The purpose of the paper is to provide and analyze explicit representations for Entropy applicable for tests of the unit root hypothesis, formally testing in (1),

$$H_0 : \alpha = 1 \quad \text{vs.} \quad H_1 : |\alpha| < 1.$$

To do so it is assumed that the deterministic component, d_t, and the underlying innovations, ζ_t, satisfy:

Assumption 1 (i) $(\zeta_t)_{t=1}^T$ is a stationary Gaussian process generated according to

$$\zeta_t = \sum_{j=1}^{m} \phi_j \varepsilon_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim iidN(0, \sigma^2),$$

so that the coefficients ϕ_j are such that the roots of the polynomial $\phi(z) = 1 + z\phi_1 + ...z^m\phi_m$ lie outside the unit circle.

(ii) The deterministic component d_t is linear in a set of fixed or strongly exogenous variables, $x_t = (x_{1t}, ..., x_{kt})$, with

$$d_t = x_{1t}\beta_1 + x_{2t}\beta_2 + .. + x_{kt}\beta_k,$$

where the β_i, $i = 1, 2, .., k$ are unknown parameters.

Under Assumption 1 we are able to characterize the model in terms of a Generalized Linear Regression Model (GLRM). To do so define the following $T \times 1$ vectors, $y = (y_t)_{t=1}^T$ and $\varepsilon = (\varepsilon_t)_{t=1}^T$, and let $X = (x_1, x_2, ..., x_T)'$ and $\beta = (\beta_1, ..., \beta_k)'$. Now let $L^{(j)}$ define a lower triangular matrix with $1's$ on the j^{th} lower diagonal and $0's$ elsewhere, so that we can construct the matrices

$$\Delta_\alpha = I - \alpha L^{(1)} \quad \text{and} \quad K_\phi = I + \sum_{j=1}^{m} L^{(j)}\phi_j.$$

As a consequence, Assumption 1 implies the Gaussian GLRM given by

$$y = X\beta + \Delta_\alpha^{-1} K_\phi \varepsilon, \tag{2}$$

so that

$$y \sim N\left(X\beta, \sigma^2 \Sigma_\phi(\alpha)\right), \tag{3}$$

where

$$\Sigma_\phi(\alpha) = \Delta_\alpha^{-1} K_\phi K_\phi' \left(\Delta_\alpha^{-1}\right)'.$$

From (3) we have a log-likelihood for y,

$$\begin{aligned} L(\alpha) &= \log\left(f(y;\alpha)\right) \\ &= -\frac{1}{2\sigma^2}\left(y - X\beta\right)'\left(\Sigma_\phi(\alpha)\right)^{-1}\left(y - X\beta\right) - \frac{1}{2}\ln\det\Sigma_\phi(\alpha) - \frac{T}{2}\ln\left(2\pi\sigma^2\right). \end{aligned}$$

and so given the following definitions of SE and KL,

$$\begin{aligned} SE(\alpha) &= -E_\alpha[L(\alpha)] = -\int_y L(\alpha) f(y;\alpha)dy \\ &\text{and} \\ KL(\alpha) &= E_1\left[L(1) - L(\alpha)\right] = \int_y \left(L(1) - L(\alpha)\right) f(y;1)dy, \end{aligned}$$

we have the standard results for multivariate Gaussian variates, that

$$SE(\alpha) = \frac{1}{2}\left[\ln\det\Sigma_\phi(\alpha) + T\left(1 + \ln\left(2\pi\sigma^2\right)\right)\right] \tag{4}$$

and

$$KL(\alpha) = \frac{1}{2}\left[Tr\left[\left(\Sigma_\phi(\alpha)\right)^{-1}\Sigma_\phi(1)\right] + \ln\frac{|\Sigma_\phi(\alpha)|}{|\Sigma_\phi(1)|} - T\right]. \tag{5}$$

Although neither quantity depends upon the unknown vector β (nor does KL depend upon the unknown scalar variance σ^2), neither do they depend upon the deterministic component, d_t. Given the wealth of numerical evidence cited in the introduction which points to a clear dependence of the properties of unit root tests, particularly the degree of trending in d_t, measuring the entropy in the data itself cannot be informative about those properties. Here, instead, we will derive Entropy for the data after it has been detrended. Currently the favoured method of detrending is via a GLS estimation of β, see for example Elliott, Rothenberg and Stock (1996). Analogous results for OLS detrended data may also be derived via the methods described below.

To proceed, define

$$z = K_\phi^{-1}\Delta_1 y, \quad W = K_\phi^{-1}\Delta_1 X \quad \text{and} \quad M = I - W(W'W)^{-1}W', \tag{6}$$

so that

$$\hat{\beta}_{GLS} = (W'W)^{-1}W'z,$$

Entropy will then be analyzed for the quantity,

$$w = z - W(W'W)^{-1}W'z = Mz.$$

where M is the symmetric idempotent matrix of rank $T-k$ given by

$$M = I - W(W'W)^{-1}W'.$$

Although w is not a feasible statistic its properties are highly relevant for the problem, in the sense that it is precisely the statistic that we would construct if we had knowledge of the nuisance parameters, ϕ. A feasible version can be calculated via

$$\hat{w} = \hat{M}z,$$

where $\hat{M} = I - \hat{W}(\hat{W}'\hat{W})^{-1}\hat{W}'$ and $\hat{W} = K_{\hat{\phi}}^{-1}\Delta_1 X$, with $\hat{\phi}$ a consistent estimator for ϕ. The given results would then apply asymptotically to $\hat{w}$.

Defining the following matrices,

$$\begin{aligned} \Omega_\phi(\alpha) &= Cov[z] = K_\phi^{-1}T_1T_\alpha^{-1}K_\phi K_\phi'(T_\alpha^{-1})'T_1'(K_\phi^{-1})', \\ & \text{and} \\ A &= C'\Omega_\phi(\alpha)C, \end{aligned}$$

where C is an $T \times (T-k)$ matrix satisfying

$$CC' = M \quad \text{and} \quad C'C = I_{T-k}, \tag{7}$$

then in the following Lemma, the Likelihood, Shannon Entropy and Kullback-Leibler are given for the detrended data w.

Lemma 1 *Under Assumption 1; (i) the distribution of w is Singular-Normal, with Log-Likelihood*

$$L(\alpha) = -\frac{1}{2}y'C'A^{-1}Cy - \frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log\sum_{i=1}^{n}\lambda_i, \tag{8}$$

where $n = T-k$, the λ_i are the non-zero ordered eigenvalues of A.
(ii) Shannon Entropy and Kullback-Leibler are given by

$$SE = \frac{1}{2}\left[\sum_{i=1}^{n}\log\lambda_i + n\left(1+\ln\left(2\pi\sigma^2\right)\right)\right],$$

$$KL = \frac{1}{2}\left[\sum_{i=1}^{n}\left(\lambda_i^{-1} + \log\lambda_i\right) - n\right]. \quad \blacksquare$$

Remarks:

(i) Although not constructed via a formal invariance argument, KL does not depend upon either of the nuisance parameters β or σ^2. SE does, of course depend upon σ^2. For the purposes of this paper we are not treating the moving average parameters ϕ as nuisance, in the sense that we wish to explicitly measure their impact upon these measures of entropy.

(ii) Crucially, however, both SE and KL depend on both the deterministic component and also the correlation structure of the innovations. This dependence is not altogether transparent, since it follows from the properties of the eigenvalues of the singular covariance matrix $M\Sigma_\phi(\alpha)^{-1}M$. The following section will detail the differential properties of the λ_i, and thence both measures of entropy, as functions of both the parameters ϕ and also the degree of trending in d_t.

(iii) Since w is a zero-mean singular Normal random variable its distribution is characterized entirely by its covariance matrix, $M\Sigma_\phi(\alpha)^{-1}M$. Consequently any measure of distance, such as KL, or entropy measure in general, for the unit root problem, is going to be function only of that covariance. By choosing to detrend via the GLS estimator and normalizing via $z = K_\phi^{-1}\Delta_1 y$, we greatly simplify subsequent derivations and their interpretation since under the null hypothesis of a unit root we have,

$$pdf(w|H_0) = pdf(Z), \quad \text{where } Z \sim N(0, \sigma^2 I).$$

That is, we standardize so that Entropy is measured relative to that of a Gaussian random variable having scalar covariance. Although OLS detrending could easily be employed instead, the simplifications offered by the current framework would be lost.

(iv) Of some interest in the recent literature has been the impact of a non-zero initial condition on the properties of unit root tests, see for example Elliott and Müller (2007). Here a non-zero, possibly even divergent, observed value may be incorporated, via the set up in Marsh (2007), by allowing for one column of X to be the vector $(y_0, ..., 0)'$. Consequently, the observed initial value can be simply regressed out. Although this approach is not efficient, since the information about α contained in the dependence of y_1 on y_0 is discarded, it does allow us to abstract from this issue to concentrate upon the effect of the deterministic component and innovation correlation.

(v) In the following section it will be shown that KL is (quasi) convex in the parameters ϕ and also the degree of trending in d_t. Thus for a given value of α under the alternative they are minimizable with respect to these parameters. Following Gibbs and Su (2002) other measures of distance on the space of density functions are bounded above by KL. Specifically, if we define the Total Variation by

$$TV = \frac{1}{2} \int_w |pdf(w|H_0) - pdf(w|H_1)|\, dw,$$

and the Hellinger distance by

$$HD = \frac{1}{2} \int_w \left| \sqrt{pdf(w|H_1)} - \sqrt{pdf(w|H_0)} \right|^2 dw,$$

then

$$TV \leq \sqrt{\frac{KL}{2}} \quad \text{and} \quad HD_w \leq \sqrt{KL}.$$

Consequently, by minimizing KL we inevitably make both Total Variation and Hellinger distance small as well, even if we are not formally minimizing them.

3 Properties of entropy measures for unit roots

In this section we will explore both the differential and numerical properties of both measures of entropy, as functions of the salient model features. While the dependence upon the moving average parameters is explicit,it is currently only implicit for the characteristics of the deterministic component. To proceed we will parametrize the trending characteristics of the deterministic component as in the following assumption.

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

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

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. We must exclude $p = 0$, since we assume the presence of a constant, in any case.

Thus we can parameterize both SE and KL as functions of both p and ϕ (as well as the autocorrelation coefficient α, and for SE also σ^2) as $SE(\alpha, p, \phi, \sigma^2)$ and $KL(\alpha, p, \phi)$. Here we require the differential properties of these functions. The non-zero eigenvalues of the covariance of w are simply the eigenvalues of the matrix $A = C'\Sigma_\phi(\alpha)C$, where C is the singular value decomposition of M. The following Theorem establishes differentiability and gives the first differentials with respect to p and ϕ.

Theorem 1 *Let C be the singular value decomposition of the symmetric idempotent $CC' = M = I - W(W'W)^{-1}W'$ and let C_0 and W_0 define points in $\mathbb{R}^{T\times n}$ and $\mathbb{R}^{T\times k}$, 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,

$$\begin{aligned} \partial_p C &= W(W'W)^{-1}(d_p W)'C \\ \partial_{\phi_j} C &= W(W'W)^{-1}(d_{\phi_j} W)'C. \qquad \blacksquare \end{aligned} \tag{9}$$

Theorem 1 establishes that C is differentiable in both the trend parameter p and moving average parameters ϕ. We are thus in a position to state the main theorem of this paper and demonstrate that both measures of entropy are themselves differentiable functions of p and ϕ. Moreover, both are shown to be (quasi) convex in their arguments and are thus, in principle, minimizable.

Theorem 2 *Let SE and KL be defined as in Lemma 1, and assume also that Assumption 2 holds, then;*
(i) both SE and KL are differentiable, and therefore continuous, with respect to p, with derivatives given by

$$\frac{\partial KL}{\partial p} = Tr\left[A^{-1}(\partial_p W)'(W'W)^{-1}W'\Sigma_\phi(\alpha)C\left(I_n - A^{-1}\right)\right] \tag{10}$$

$$\frac{\partial SE}{\partial p} = Tr\left[A^{-1}(\partial_p W)'(W'W)^{-1}W'\Sigma_\phi(\alpha)C\right], \tag{11}$$

where

$$\partial_p W = \Delta_1(\partial_p X) = \Delta_1\left(\underline{0}, .., \underline{0}, \partial_p X_i(p), \underline{0}, .., \underline{0}\right).$$

(ii) both SE_w and KL_w are differentiable, and therefore continuous, with respect to $\phi = \{\phi_j\}_{j=1}^{m}$, with derivatives given by

$$\frac{\partial KL}{\partial \phi_j} = Tr\left[A^{-1}\left(d_{\phi_j}W\right)'(W'W)^{-1}W'\Sigma_\phi(\alpha)C\left(I_n - A^{-1}\right)\right] \tag{12}$$

$$\frac{\partial SE}{\partial \phi_j} = Tr\left[A^{-1}\left(d_{\phi_j}W\right)'(W'W)^{-1}W'\Sigma_\phi(\alpha)C\right], \tag{13}$$

where

$$d_{\phi_j}W = -K_\phi^{-1}L^{(j)}W.$$

(iii) $KL(\alpha, p, \phi)$ is quasi-convex and $SE(\alpha, p, \phi, \sigma^2)$ is quasi-concave over both p and ϕ and therefore any solutions in p, consequently any solution in p to (10) and in ϕ to (12) are at minima, while solutions to (11) and (13) are at maxima. ■

Theorem 2 establishes that the measures of entropy are continuous, differentiable and convex functions of the model parameters p and ϕ. Even before finding values of p and ϕ which minimize entropy in particular examples it is now unequivocally established that we can measure the precise effect that different model configurations have on our ability to, statistically, discriminate between processes having unit roots and those which are driven by a stationary autoregression.

In the following section these results will be both illustrated numerically and also applied to investigate the effect of model configuration for the Nelson and Plosser (1982) series of macroeconomic data.

4 Numerical Analysis

4.1 The effects of deterministic trending and innovation autocorrelation

According to Theorem 2, both SE and KL are convex functions of the degree of trending and the moving average parameters. We can therefore, in principle find values of these parameters that minimize entropy, for each value of α under the alternative. To illustrate, suppose first that we consider a simplified version of (1) with no error autocorrelation, viz.

$$(1-\alpha l)(y_t - \beta_1 - \beta_2 t^p) = \varepsilon_t \quad ; \quad \varepsilon_t \sim iidN(0,\sigma^2), \quad t = 1,..,T, \tag{14}$$

where l is the lag-operator. We may solve both

$$\left.\frac{\partial KL}{\partial p}\right|_{p=p^*_{KL}} = 0 \quad \text{and} \quad \left.\frac{\partial SE}{\partial p}\right|_{p=p^*_{SE}}$$

and plot the solutions p^*_{KL} and p^*_{SE} for different sample sizes ($T = 25, 50$, and 100), giving Figures 1a and 1b, in the Appendix. Notice, that for moderate sample sizes, and for alternatives 'close' to the null neither measures of entropy have stationary points, exactly, at a linear time trend.

In practice trends of the form $d_t = \beta_1 + \beta_2 t^{0.85}$ are seldom employed, of course. Consequently, Tables 1a and 1b, in the appendix, evaluate KL and SE in the following model,

$$(1-\alpha l)(y_t - d_t) = \varepsilon_t \quad ; \quad \varepsilon_t \sim iidN(0,\sigma^2), \tag{15}$$

with different d_t representing different deterministic components. In comparison, it is clear that a linear trend implies much smaller relative entropy (larger Shannon entropy) than other reasonable trends, such as either a square root or squared trend.

Consider now the time series regression;

$$(1-\alpha l)\left(y_t - \beta_1 - I(\tau T)\,\beta_2 t\right) = \varepsilon_t \quad ; \quad \varepsilon_t \sim iidN(0,\sigma^2) \tag{16}$$

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 analyzed 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. Those Monte Carlo findings correspond exactly with the explicit results given here. KL is evidently a decreasing function of τ while SE is increasing.

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 (1), namely

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

so that the y_t follows an $ARIMA(0,1,1)$ process under the null. As α varies we can calculate the minimum and maximum arguments of KL and SE, $\bar{\phi}_1^{KL}$ and $\bar{\phi}_1^{SE}$ respectively, for sample sizes of $T=25,50$, and 100 for model (17). These values are plotted in Figures 2a and 2b, 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 minimizes the distance. Although, as in the case with a linear trend, there is some uniformity in that the values of either $\bar{\phi}_1^{KL}$ or $\bar{\phi}_1^{SE}$ is not particularly sensitive with respect to α. That is, we are not merely measuring a common factor effect because there can either be a common factor under the null or the alternative, but not both.

To summarize the theoretical and numerical properties of KL, in particular; it is analytic and minimizable in the model features as parametrized 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 minimize our distance, and thus power. In the following sub-section we'll use this knowledge to examine how model specification affects entropy in practice.

4.2 Illustration (Nelson & Plosser Data)

Having detailed the differential and numeric properties of KL and SE, 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 macroeconomic time series. We will consider the two model specifications,

$$M_1 \;:\; (1-\alpha l)\,(y_t-\beta_1)=\phi_1\varepsilon_{t-1}+\phi_2\varepsilon_{t-2}+\varepsilon_t \tag{18}$$

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

where again $\varepsilon_t \sim iidN(0, \sigma^2)$, l is the lag-operator and $t = 1, ..., T$. Estimation of these two models, and evaluation of KL and SE 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 consistently estimate both M_1 and M_2 we will also need to additionally assume that the process $\{u_t\}$ is specifically that of an invertible MA(2).

The Nelson and Plosser data has been much analyzed 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, at least, it seems that altering the trending behaviour of the regressors, for example the inclusion of a linear trend, or the timing of any breaks in that trend, can alter the outcome of a test. Here, to illustrate the practical relevance of the results of the previous two section, we will focus upon the simple issue of the effect of whether, or not, we include a linear trend in our fitted model.

In the context of either model, M_1 or M_2, the entropic measures are functions of α, and $\phi = (\phi_1, \phi_2)'$ and also, only for Shannon entropy, σ^2. Let $\theta_1 = (\alpha, \phi_1, \phi_2)'$ and $\theta_2 = (\theta_1', \sigma^2)$, and let $\hat{\theta}_1$ and $\hat{\theta}_2$ be their respective consistent estimators. Since Theorem 2 establishes that these are (infinitely) differentiable functions of these parameters, then via the mean value theorem we have

$$\begin{aligned} KL(\theta_1) &= KL\left(\hat{\theta}_1\right) + \left(\theta_1 - \hat{\theta}_1\right)' \left.\frac{dKL(\theta_1)}{d\theta_1}\right|_{\theta_1=\bar{\theta}_1} \\ SE(\theta_2) &= SE\left(\hat{\theta}_2\right) + \left(\theta_2 - \hat{\theta}_2\right)' \left.\frac{dSE(\theta_2)}{d\theta_2}\right|_{\theta_2=\bar{\theta}_2}, \end{aligned}$$

where $\bar{\theta}_1$ and $\bar{\theta}_2$ are mean values. Consequently, we may consistently estimate KL and SE, via

$$\begin{aligned} \widehat{KL} &= KL\left(\hat{\theta}_1\right) = KL(\theta_1) + o_p(1) \\ \widehat{SE} &= SE\left(\hat{\theta}_2\right) = SE(\theta_2) + o_p(1). \end{aligned}$$

For each series in the Nelson and Plosser (1982) data set, the maximum likelihood estimator for θ_1 and θ_2 was computed and the estimated measures of entropy calculated. In order to validate, to some extent, the assumption of an MA(2) innovation process, for each model the first-order residual correlation,

$$r_m = \sqrt{T - k_m}\frac{\sum_{k_m+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}}{\sum_{k_m+1}^{T} \hat{\varepsilon}_t^2},$$

where k_m is the total number of parameters in M_m, $m = 1, 2$. The maximum value of $|r_m|$ obtained in any estimated model was 1.815, which is not significant according to standard asymptotic critical values at the 5% significance level.

Recorded in Table 3 in the Appendix, for each series, is the sample size, and for both M_1 and M_2 the estimated autoregressive coefficient, $\hat{\alpha}_m$ and the estimated entropic measures, $\widehat{KL}_m$ and $\widehat{SE}_m$, again for $m = 1, 2$. Notice that since estimated Shannon entropy depends upon the estimated variance the values for $\widehat{SE}_m$ are comparable only within the context of each individual data set. Also it should be highlighted that the purpose of Table 3 is to illustrate the practical usefulness of the measures of entropy, not to determine which of M_1 or M_2 is more appropriate for any series, nor to determine whether or not the series actually are unit root processes.

For the majority of series the effect of including a trend is actually relatively benign. Generally it is the case that including a trend implies an estimated model a little further from a unit root, whether measured by either the estimated coefficient or the Kullback-Leibler divergence. Equally, again generalizing, including a trend implies a slightly higher Shannon entropy. It is, of course, the exceptions and the more extreme results to this which are of interest. The exceptions are found for Velocity and for the CPI. For velocity inclusion of a trend implies a fitted model very close to a unit root, while not including it implies a model which ranks 4th in terms of distance from the unit root. Similar is true for CPI, although in the explosive direction when a trend is not included. As one would expect, perhaps, Unemployment does not have a unit root. Importantly though, notice that although the estimated coefficient is closer to unity in M_1, the fitted model is 'further' from a unit root process than is the case for M_2. Finally, Industrial Production has a fitted model indistinguishable from a unit root when no trend is included, in stark contrast to the fitted model for M_2.

5 Conclusions

This paper has derived and analyzed two measures of entropy applicable for the unit root problem, namely Shannon entropy and Kullback-Leibler distance for the GLS detrended data. Specifically, both are shown to be differentiable functions of parameterizations of the important model features; the degree of deterministic trending and innovation autocorrelation. Kullback-Leibler is convex, and so minimizable, in these parameters while Shannon entropy is concave, and thus maximizable. Theoretically, therefore, there are particular trends and correlations which limit the efficacy of any statistical method. It was then confirmed that, as predicted by the many experiments in the literature, it is (approximately) linear trends and negative unit root moving average innovations which do this.

The results also have practical importance. In the context of the Nelson and Plosser (1982) data set, the entropic measures can be estimated via the fitted model. By calculating them for two models, one in which only a constant is included the other with also a trend, the impact of model specification can be explicitly measured. For several series the results are dramatic, notably Velocity and Industrial Production. Consequently, in the context of actual unit root testing having some measure of the impact of the chosen model would compliment the outcome of any coefficient estimator or unit root test.

References

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

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

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.

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

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

Granger, C.W.J. and P Newbold (1974): Spurious regressions in econometrics. *Journal of Econometrics*, **2**, 111-120.

Harvey, D.I., Leybourne, S.J. and A.M.R. Taylor (2007): Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. Mimeo, University of Nottingham.

Leybourne, S.J., Mills, T.C. & P. Newbold. (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 H. Neudecker (1999): *Matrix Differential Calculus, with Applications in Statistics and Econometrics* (revised edition), Chichester, Wiley.

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

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

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 T. Magdalinos (2006) Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, **136**, 115-130.

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

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

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

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

I. Proofs

Proof of Lemma 1

Part (i): We have $w = My$, and M is square, symmetric and idempotent of rank $n = T - k$, and hence it has singular Normal distribution, with mean and variance

$$E[w] = \mu = MW = 0 \quad ; \quad V[w] = \sigma^2\Omega = \sigma^2 M\Sigma_\phi(\alpha)M.$$

The density function of w is therefore given by

$$f(w;\alpha) = \frac{\exp\left\{-\frac{1}{2\sigma^2}w'\Omega^- w\right\}}{(2\pi\sigma^2)^{n/2}\sqrt{\prod_{i=1}^n \lambda_i}}, \tag{20}$$

where the λ_i, $i = 1, 2, .., n$ are the ordered non-zero eigenvalues of $M\Sigma_\phi(\alpha)M$ and Ω^- denotes a generalized inverse of Ω.

To find a suitable Ω^-, write

$$M\Sigma_\phi(\alpha)M = CC'\Sigma_\phi(\alpha)CC' = C\Lambda C',$$

where the $T \times n$ matrix C is the singular value decomposition of M defined in (7). Consequently, note that

$$\begin{aligned}
\Omega C\Lambda^{-1}C'\Omega &= C\Lambda C'C\Lambda C'C\Lambda^{-1}C'C\Lambda C' = C\Lambda C' = \Omega \\
C\Lambda^{-1}C'\Omega C\Lambda^{-1}C &= C\Lambda^{-1}C'C\Lambda C'C\Lambda^{-1}C = C\Lambda^{-1}C' \\
\left(C\Lambda^{-1}C'\Omega\right)' &= (CC')' = M' = M = CC' = \Omega C\Lambda^{-1}C' \\
\left(\Omega C\Lambda^{-1}C'\right)' &= (CC')' = M' = M = CC' = C\Lambda^{-1}C'\Omega,
\end{aligned}$$

so that $C\Lambda^{-1}C'$ satisfies the four conditions which define the unique Moore-Penrose inverse of Ω, i.e. we can use,

$$\Omega^- = \Omega^+ = C\Lambda^{-1}C'.$$

Since the singular value decomposition is unique only up to orthogonal transformation, and since the matrix A is symmetric, it has a spectral decomposition of

$$A = R\Lambda R', \quad \Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}$$

where $R'R = RR' = I_n$ and $R = (r_1, r_2, ..r_n)$ is the matrix of eigenvectors, which then satisfy the equations

$$Ar_i = \lambda_i r_i, \quad i = 1, 2, .., n,$$

that is the λ_i are the eigenvalues of A. Moreover, we can also assume, without loss of generality that in addition to (7), C also satisfies,

$$C'AC = C'\Lambda C \quad \text{and} \quad C'A^{-1}C = C'\Lambda^{-1}C.$$

The log-likelihood based on w is thus given by

$$\begin{aligned} L(\alpha) &= -\frac{1}{2\sigma^2}w'\Omega^- w - \frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{n}\ln\lambda_i \\ &= -\frac{1}{2\sigma^2}w'CA^{-1}C'w - \frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{n}\ln\lambda_i, \\ &= -\frac{1}{2\sigma^2}y'CA^{-1}C'y - \frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{n}\ln\lambda_i, \end{aligned}$$

as required, since $C'w = C'My = C'CC'y = C'y$.

Parts (ii) & (iii) follow from straightforward application of the definitions given in (5) and (4) to an $n \times 1$ Gaussian random vector,

$$z = C'y \sim N(0, \sigma^2 A). \quad \blacksquare$$

Proof of Theorem 1

Since $W = K_\phi^{-1}T_1X$, then immediately W is differentiable with respect to ϕ_j. Now under Assumption 2, 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 $M_W = I_T - W(W'W)^{-1}W'$, and is therefore the unique solution (up to orthogonal transformation), in $R^{T\times n}$, to the equations

$$M_W = CC' \quad \text{and} \quad C'C = I_n. \tag{21}$$

We first show that (21) implies and is implied by

$$W'C = \mathbf{0} \quad \text{and} \quad C'C = I_n. \tag{22}$$

To do this note that

$$M_W = CC' \Longleftrightarrow (I_T - M_W)C = \mathbf{0}, \tag{23}$$

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 (23) as $\left(W^+\right)' W'C = \mathbf{0}$, then since

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

we have

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

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 $h : \mathbb{R}^{T\times k} \times \mathbb{R}^{T\times n} \to \mathbb{R}^{T\times n}$ of C and W by

$$h(C, W) = \begin{pmatrix} W'C \\ C'C - I_{T-k} \end{pmatrix},$$

then following a similar argument to Magnus and Neudecker (1988), Theorem 8.7, h is differentiable on $\mathbb{R}^{T\times k} \times \mathbb{R}^{T\times n}$. Letting the point C_0, W_0 in $\mathbb{R}^{T\times k} \times \mathbb{R}^{T\times n}$ 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 $\mathbb{R}^{T\times k}$, $V(W_0)$ and a unique (up to orthogonal transformation) matrix valued function $C : V(W_0) \to \mathbb{R}^{T\times n}$ 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 $C'C = I_{T-k}$ 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 (22) 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

$$P = (W')^{+}W + CC' = P_W + M_W = I_{T-k},$$

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 (9). ■

Proof of Theorem 2

For part (i), and from Lemma 1, we can write

$$\begin{aligned} SE &= \frac{1}{2}\left[\ln|A| + n\left(1 + \ln\left(2\pi\sigma^2\right)\right)\right], \\ KL &= \frac{1}{2}\left[Tr[A^{-1}] + \ln|A| - n\right], \end{aligned}$$

where A is a function of p. In order to establish differentiability we utilize *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

$$\partial_p A = \left[\partial_p C'\Sigma_{\alpha,\phi}C + C'\Sigma_{\alpha,\phi}\partial_p C\right], \tag{24}$$

so that substitution of (9) into (24), yields

$$\partial_p A = \frac{\partial A}{\partial p} = -C'[D + D']C, \tag{25}$$

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

$$\frac{\partial \ln |A|}{\partial p} = Tr\left[A^{-1}\left(\partial_p A\right)\right], \quad \text{and} \quad \partial_p A^{-1} = -A^{-1}\left(\partial_p A\right)A^{-1}$$

so that

$$\frac{\partial KL}{\partial p} = \frac{1}{2}\left(Tr\left[\left(\partial_p A\right)A^{-1}\left(I - A^{-1}\right)\right]\right) \text{ and } \frac{\partial SE}{\partial p} = \frac{1}{2}\left(Tr\left[\left(\partial_p A\right)A^{-1}\right]\right), \tag{26}$$

substituting (25) into (26) and rearranging proves part (i).

For part (ii) differentiability is established in exactly the same way as in part (i). The required derivatives are

$$\frac{\partial KL}{\partial \phi_j} = \frac{1}{2}\left(Tr\left[\left(\partial_{\phi_j} A\right)A^{-1}\left(I - A^{-1}\right)\right]\right) \text{ and } \frac{\partial KL}{\partial \phi_j} = \frac{1}{2}\left(Tr\left[\left(\partial_{\phi_j} A\right)A^{-1}\right]\right), \tag{27}$$

For this case the derivative of A is

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

however, from the definition of $\Sigma_{\alpha,\phi}$, $\partial_{\phi_j}\Sigma_{\alpha,\phi} = 0$, so that the second term in (28) vanishes. From (9), 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

$$\partial_{\phi_j} A = C'(H + H')C, \tag{29}$$

where $H = K_\phi^{-1}L^{(i)}P_W\Sigma_{\alpha,\phi}$, so that substituting (29) into (27) gives the required derivative.

For part (iii), notice that A and hence A^{-1} are positive definite matrices of rank n. Consequently, following Magnus and Neudecker (1999, p. 152), the k^{th} derivatives (guaranteed to exist by Theorem 1) of the matrix functions A^{-1} and $\ln|A|$, with respect to p, are given by

$$\begin{aligned}\frac{\partial^k A^{-1}}{\partial p^k} &= (-1)^k\, k!\left(A^{-1}(\partial_p A)\right)^k A^{-1}, \\ &\text{and} \\ \frac{\partial^k \ln|A|}{\partial p^k} &= (-1)^{k-1}(k-1)!Tr\left[\left(A^{-1}(\partial_p A)\right)^k\right],\end{aligned}$$

with very similar expressions for the derivatives with respect to the ϕ_j. As a consequence the second derivatives of entropy are given by

$$\begin{aligned}\frac{\partial^2 KL}{\partial p^2} &= \frac{1}{2}Tr\left[\left(A^{-1}\partial_p A\right)^2\left(2A^{-1} - I\right)\right] \\ &\text{and} \\ \frac{\partial^2 SE}{\partial p^2} &= -\frac{1}{2}Tr\left[\left(A^{-1}\partial_p A\right)^2\right],\end{aligned}$$

again with very similar expressions for the derivatives with respect to the ϕ_j.

As an immediate consequence we have that

$$\frac{\partial^2 SE}{\partial p^2} \leq 0$$

and so $SE(\alpha, p, \phi)$ is (Quasi) concave in p and so any solution in p, to either

$$\frac{\partial SE(\alpha, p, \phi)}{\partial p} = 0,$$

is at a maximum. For KL, since, both $Tr[A^{-1}]$ and $Tr[A]$ are minimized at $A = I_n$, then

$$Tr\left[\left(2A^{-1} - I\right)\right] > 0 \Rightarrow \frac{\partial^2 KL}{\partial p^2} \geq 0,$$

and so $KL(\alpha, p, \phi)$ is (Quasi) convex in p and so any solution in p, to

$$\frac{\partial KL(\alpha, p, \phi)}{\partial p} = 0,$$

is at a minimum.

Similarly, given a value of polynomial trending p and for any value of α both $KL(\alpha, p, \phi)$ and $SE(\alpha, p, \phi)$ are respectively convex and concave functions of each ϕ_j, and so solutions to

$$\frac{\partial KL(\alpha, p, \phi)}{\partial \phi_j} = 0 \quad or \quad \frac{\partial SE(\alpha, p, \phi)}{\partial \phi_j} = 0,$$

are also at a minimum and maximum, respectively.

Figures and Tables

Fig.1a: p^*_{KL} for model (14) for
$T = 100$ (—), $T = 50$ (- - -) and $T = 25$ ($\cdots$).

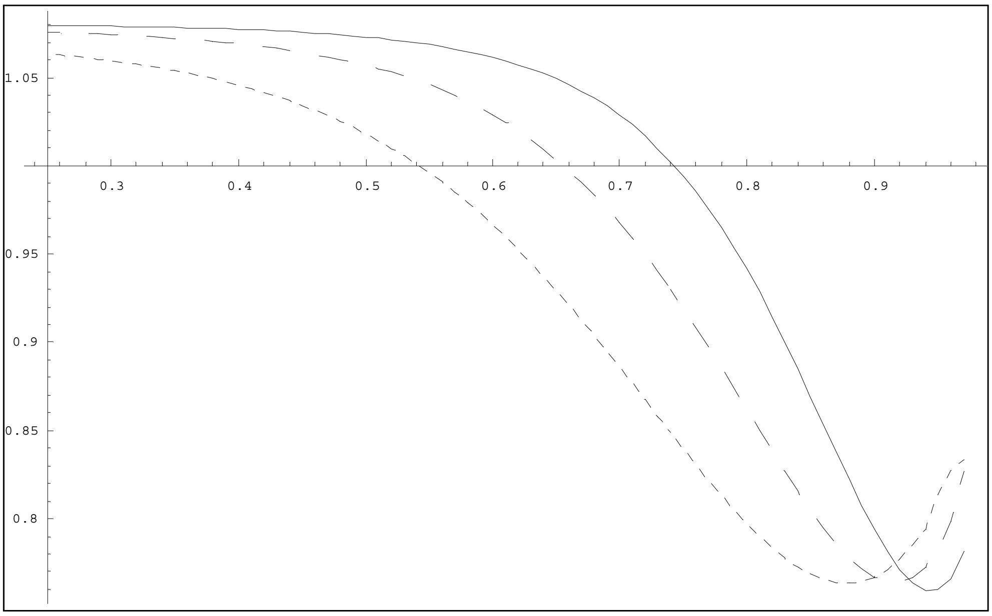

Fig.1b: p^*_{SE} for model (14) for

$T = 100$ (—), $T = 50$ (- - -) and $T = 25$ ($\cdots$).

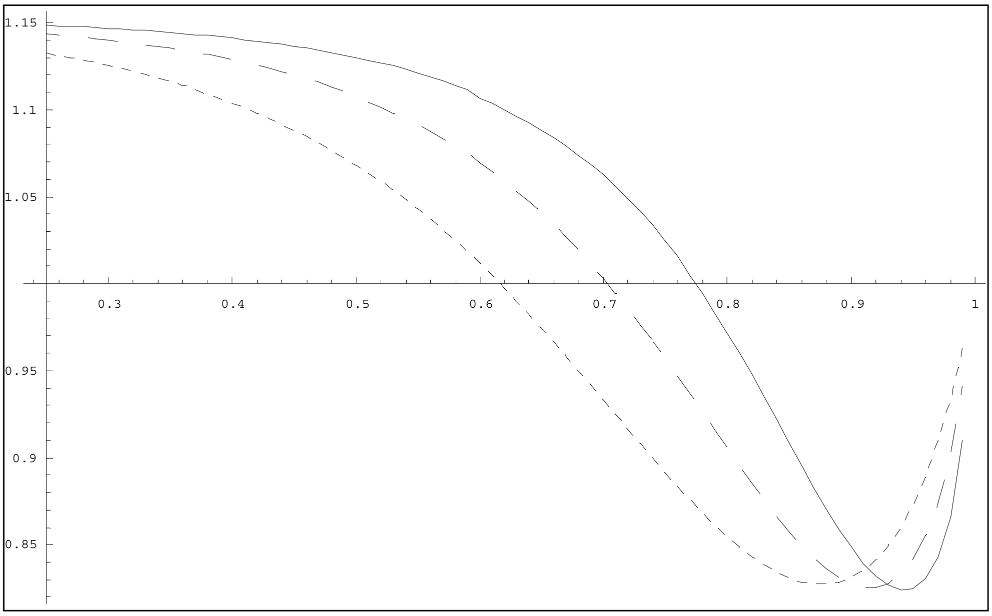

Table 1a: KL in (15) for different trends

ρ / d_t	.975	.850	.925	.900	.875	.850	.825	.800
$\beta_1 + \beta_2 t^2$	.186	.878	2.05	3.58	5.42	7.53	9.91	12.6
$\beta_1 + \beta_2 t$	.057	.406	1.15	2.29	3.80	5.63	7.79	10.3
$\beta_1 + \beta_2 \sqrt{t}$	.071	.423	1.19	2.41	4.06	6.11	8.55	11.4
$\beta_1 + \beta_2 \log(t)$	.344	.987	2.12	3.84	6.15	9.03	12.5	16.4

Table 1b: $SE - 140$ in (15) for different trends

ρ / d_t	.975	.850	.925	.900	.875	.850	.825	.800
$\beta_1 + \beta_2 t^2$	-.271	.161	.500	.790	1.05	1.28	1.49	1.70
$\beta_1 + \beta_2 t$	-.092	.391	.740	1.02	1.27	1.49	1.69	1.87
$\beta_1 + \beta_2 \sqrt{t}$	-.163	.327	.670	.940	1.17	1.37	1.55	1.71
$\beta_1 + \beta_2 \log(t)$	-.496	-.073	.245	.493	.698	.874	1.03	1.17

Table 2a: KL in (16) for different breakpoints τ.

τ \ ρ	.975	.850	.925	.900	.875	.850	.825	.800
0.2	1.11	3.28	5.53	7.68	9.84	12.1	14.6	17.3
0.4	.612	1.51	2.52	3.77	5.31	7.14	9.29	11.7
0.6	.337	.874	1.69	2.85	4.34	6.17	8.31	10.8
0.8	.203	.629	1.40	2.53	4.03	5.86	8.01	10.5

Table 2b: $SE - 140$ in (16) for different breakpoints τ.

τ \ ρ	.975	.850	.925	.900	.875	.850	.825	.800
0.2	-.851	-.678	-.472	-.258	-.047	.155	.349	.530
0.4	-.685	-.374	-.088	.169	.403	.619	.818	1.00
0.6	-.538	-.157	.158	.427	.666	.884	1.09	1.26
0.8	-.429	-.011	.317	.592	.833	1.05	1.26	1.43

Fig.2a: $\bar{\phi}_1^{KL}$ derived for model (17) and
for $T = 100$ (—), $T = 50$ (- - -) and $T = 25$ ($\cdots$).

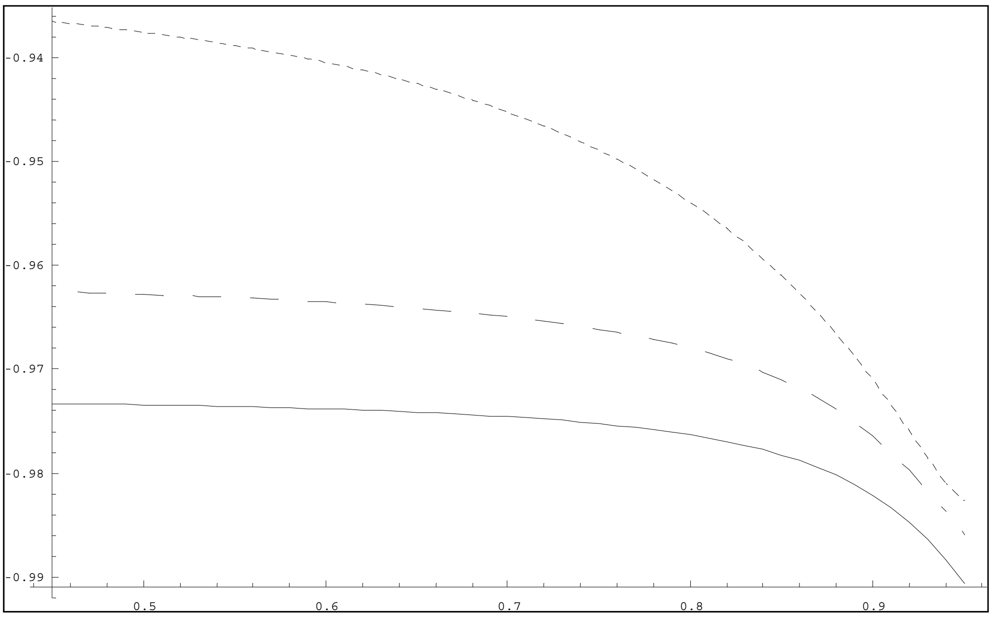

Fig.2b: $\bar{\phi}_1^{SE}$ derived for model (17) and

for $T = 100$ (—), $T = 50$ (- - -) and $T = 25$ (···).

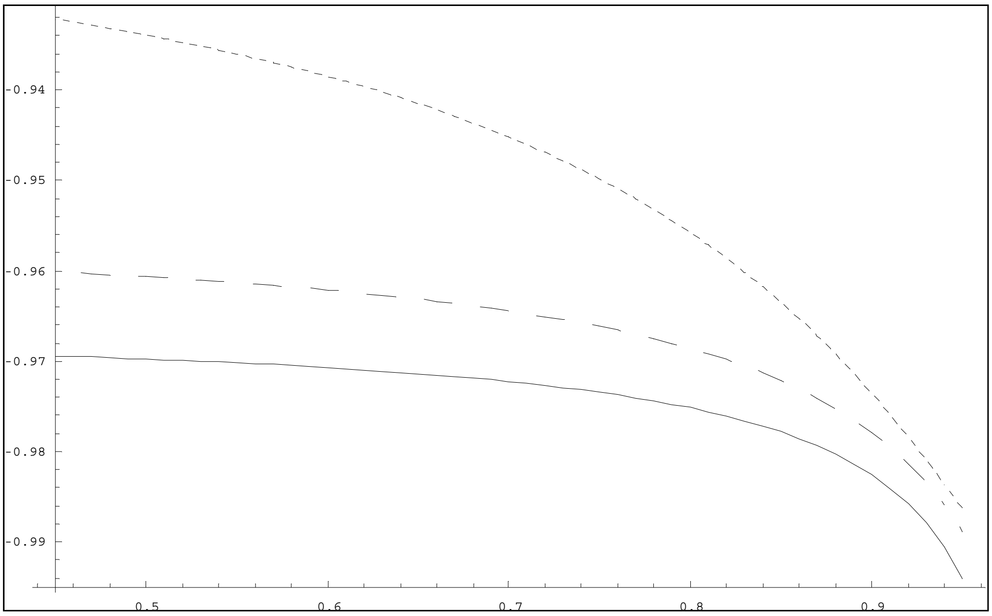

Table 3: Estimated Values for α, KL and SE in (18) and (19)

applied in the Nelson & Plosser data set.

		M_1			M_2		
	T	$\hat{\alpha}_1$	$\widehat{KL}_1$	$\widehat{SE}_1$	$\hat{\alpha}_2$	$\widehat{KL}_2$	$\widehat{SE}_2$
Real GNP	80	1.001	.0031	-111.4	0.991	.0018	-110.1
Nom. GNP	80	1.008	.1036	-79.95	0.990	.0023	-79.46
GNP Per.Cap.	80	1.002	.0085	-109.2	0.993	.0011	-107.9
Bond	89	1.000	.0001	85.82	0.949	.3020	83.95
Nom. Wage	89	1.006	.0809	-122.2	0.989	.0033	-121.1
Real Wage	89	1.000	.0001	-165.8	0.998	.0001	-163.9
Unemp.	99	0.755	52.38	56.51	0.751	15.80	57.98
Employ.	99	1.001	.0011	-135.5	0.998	.0001	-134.5
Money	100	1.003	.0231	-166.2	0.975	.0553	-164.7
S&P500	118	1.009	.3048	-44.41	0.930	1.519	-45.91
Velocity	120	0.959	4.862	-161.4	1.011	.0077	-157.3
I. Prod.	129	1.000	.0000	166.9	0.854	9.622	168.8
CPI	129	1.021	1.942	-236.9	0.983	.0353	-236.3