Power of tests for unit roots in the presence of a linear trend

By Bent Nielsen
Department of Economics, University of Oxford.
bent.nielsen@nuf.ox.ac.uk
5 November 2003

Abstract: Dickey and Fuller (1981) suggested unit root tests for an autoregressive model with a linear trend and a fixed initial value. This model has nuisance parameters so later authors have often worked with a slightly different model with a random initial value in which nuisance parameters can be eliminated by an invariant reduction of the model. This facilitates computation of envelope power functions and comparison of the relative performance of different unit root tests. It is shown here that invariance arguments also can be used when comparing power within the model with fixed initial value. Despite the apparently small difference between the two models the relative performance of unit root tests turns out to be very different.

Keywords: Envelope power function, maximal invariant parameter, maximal invariant statistic, most stringent test, unit root tests.

1 Introduction

Dickey and Fuller (1981) suggested unit root tests for an autoregressive model with a linear trend and a fixed initial value. Although these tests are widely used and have appeal as maximum likelihood ratio tests a wide range of alternative unit root tests have been developed. Many of these tests are formulated for a slightly different model with a random initial value. The change of scene allows nuisance parameters to be eliminated by an invariant reduction of the model. In the reduced model the envelope power function can be computed using the Neyman-Pearson Lemma and the power of different tests can be compared to this and to each other. The problem of comparing power is thereby solved, but it comes at the cost of using invariant reductions without any immediate interpretation from an economic subject matter view. An alternative approach is to leave nuisance parameters in the model and only eliminate them in the power comparison using the notion of maximal invariant parameters. While this idea is implicitly used by Elliot, Rothenberg, and Stock (1996) in their power comparison for a model with random initial value the contribution of this paper is to discuss these concepts more precisely and apply them for power comparison within the model with a fixed initial value.

Four different unit root tests are considered in the power comparisons. The first of these is the log maximum likelihood ratio test statistic for the model with fixed initial value that was formulated as a F-type statistic by Dickey and Fuller (1981). The other three tests were suggested by Bhargava (1986), Ahn (1993), and Elliot, Stock and Rothenberg (1996) and they are constructed for models with a random initial value. A simulation study supports the finding of Elliot, Rothenberg, and Stock (1996) that Dickey-Fuller tests are not particular good in a model with random initial value while it is also shown that the Dickey-Fuller test is actually most stringent in a model with fixed initial value. The practitioner is therefore faced with the issue of choosing the model carefully before choosing a test.

In the event that both the models with fixed and random initial values describe a data series the practitioner may be interested in comparing power across the models. Since the models are not nested statistical theory give no guidance for such a comparison. This issue is therefore addressed by fitting empirical models to a time series of quarterly UK production in order to get comparable alternative hypotheses. With these particular choices of alternatives it is found that the maximal possible power is about the same in the different models.

The above described analysis assumes that the considered models are well-specified. In applications this will have to be checked. For the model with fixed initial value the distributional assumptions are that the innovations are independently, identical normal distributed. These assumptions can be tested consistently, whereas distributional assumptions to the initial value will only concern one particular observation and cannot be tested consistently. For the considered data the assumption to the initial value happen to be strongly rejected. While this of course not implies that a model with random initial value would always be very wrong it does suggest that some care should be made in making such assumptions.

The paper is organised so §2 first reviews the relevant maximal invariant statistics and then introduces the models and associated maximal invariant parameters, while §3 describes the considered tests. The power of the tests are then compared in §4, 5 for models with random and fixed initial value respectively, whereas §6 attempts a comparison across models. §7 concludes and a few mathematical derivations are given in an Appendix.

2 Models

Suppose a linearly trending time series, $w_1, \ldots, w_T$, is observed for which the economic question of interest is whether its growth rate depends on its lagged levels. Some invariance considerations are introduced before actually formalising this as a testing problem. Two slightly different statistical models are then presented. One is to assume that w_1 is fixed leading to a conditional model whereas conveniently chosen distributions for w_1 lead to unconditional models.

2.1 Maximal invariant statistics

When formalising a statistical model the economic context will often indicate that inferences should be invariant to certain aspects of the sample variation. An example would be indifference to measuring log productivity in Pounds or in pence. This is formalised as saying that the statistical analysis should be invariant to location transformations of the type

$$g_c: \qquad w_t \mapsto w_t + b \qquad \text{for all } b \in \mathbf{R}.$$

The analysis should therefore depend on the data vector $W = (w_1, \ldots, w_T)'$ through a vector, called a maximal invariant statistic, that varies in a $(T-1)$-dimensional space determined by the orbits of the sample space under the transformation g_c, see Lehman (1997, p. 284). A convenient choice of maximal invariant statistic is

$$X = (x_1, \ldots, x_T)' \qquad \text{where} \qquad x_t = w_t - w_1,$$

which would indeed take the same value if a constant b were added to each observation w_t.

Groups of transformations can of course be formulated arbitrarily. Given the linear trend in the data translation by a linear trend like

$$g_l: \qquad w_t \mapsto w_t + b + ct \qquad \text{for all } b, c \in \mathbf{R},$$

or a transformation involving scaling as in

$$g_s: \qquad w_t \mapsto a w_t + b + ct \qquad \text{for all } a \neq 0 \text{ and } b, c \in \mathbf{R}$$

come to mind but the list of potential transformations is endless. From a subject matter perspective invariance with respect to g_l or g_s do not have any interpretation as simple as a question of units so it would be somewhat unnatural to impose these when formalising the model. See also Cox and Hinkley (1974, §2.3) for a discussion in relation to the invariance principle. When it comes to power comparison these transformations can be used more naturally and it is useful to note some associated maximal invariants, Y and Z, given by

$$y_t = x_t - \frac{t-1}{T-1} x_T, \qquad z_t = \frac{y_2 y_t}{\sqrt{y_2^2 \sum_{u=1}^T y_u^2}}.$$

2.2 Modelling by conditioning on the initial value

The first statistical model is autoregressive and takes the initial value, w_1, as fixed:

$$\mathsf{M}_{Fixed}: \qquad \begin{cases} \Delta w_t = \alpha w_{t-1} + \beta + \gamma(t-1) + \sigma\varepsilon_t, & \text{for } t = 2, \ldots, T, \\ w_1 \text{ is fixed,} \end{cases} \qquad (1)$$

where the innovations $\varepsilon_2, \ldots, \varepsilon_T$ are independently, standard normally distributed and the parameter $\theta_{Fixed} = (\alpha, \beta, \gamma, \sigma)$ satisfies $\theta_{Fixed} \in \Theta_{Fixed} = \mathbf{R}^3 \times \mathbf{R}_+$. Since w_1 is fixed the model can equivalently be formulated for the maximal invariant X by a simple reparametrisation given by

$$\Delta x_t = \alpha x_{t-1} + \delta + \gamma (t-1) + \sigma \varepsilon_t, \qquad \text{for } t = 2, \ldots, T, \tag{2}$$

where $x_1 = 0$ and the parameters $\alpha, \delta, \gamma, \sigma$ vary in the same parameter space as before, but $\delta = \beta + \alpha w_1$ has a different interpretation from β.

A process within the model M_{Fixed} can be given a trend stationary initial distribution when $|1 + \alpha| < 1$, while it is a random walk with a linear trend when $\alpha = \gamma = 0$. In order to compare these two cases Dickey and Fuller (1981) formulated the unit root hypothesis:

$$\mathsf{H}_{Fixed}: \qquad \alpha = \gamma = 0.$$

This hypothesis is preferred to a hypothesis like $\alpha = 0$ which has less compelling interpretation and generates tests that are typically suffering from lack of similarity with respect to γ, see Nielsen and Rahbek (2000).

The testing problem given by $\mathsf{M}_{Fixed}, \mathsf{H}_{Fixed}$ is invariant to the sample transformations g_c, g_l and g_s. The more general of these, g_s, induces a parameter translation

$$g^*_{s,Fixed}: \qquad (\alpha, \delta, \gamma, \sigma, w_1) \mapsto (\alpha, a\delta - b\alpha + c, a\gamma - c\alpha, a\sigma, aw_1 + b + c), \tag{3}$$

mapping Θ_{Fixed} into Θ_{Fixed} and where w_1 is included as it plays a role similar to a parameter. This has the property

$$\mathsf{P}_\theta (g_s W \in A | w_1) = \mathsf{P}_{g^*_{s,Fixed}\theta}(W \in A | w_1), \tag{4}$$

for all events A so that for each θ in the restricted parameter space $\Theta^\circ_{Fixed} = (0) \times \mathbf{R} \times (0) \times \mathbf{R}_+$ then $g^*_{s,Fixed}\theta$ is also Θ°_{Fixed}, see Lehmann (1997, p.282) and Cox and Hinkley (1974, p.157f).

The transformations g_l, g_s do not offer any scope for eliminating nuisance parameters by invariant reduction of the model. This is because the nuisance parameter δ does not in general have interpretation as either a level or a trend parameter and the marginal distributions of Y and Z will therefore depend on δ. This is perhaps not such a big loss since on the one hand the transformations g_l, g_s as opposed to g_c do not have straight forward economic interpretations and on the other hand the invariance can be exploited when it comes to comparing power functions of different tests.

2.3 Maximal invariant parameters for the model M_{Fixed}

Later on four tests will be discussed that are all functions of the maximal invariant Z under g_s. It will be therefore be useful to describe how the distribution of the maximal invariant statistic Z varies across the parameter space, Θ_{Fixed}. This is done by finding a maximal invariant parameter under the induced transformation g_s^*, see Lehman (1997, p. 292).

By definition the maximal invariant statistic Z takes the same value for a realization W and for any transformation thereof g_sW given by $g_s(w_t) = aw_t + b + ct$. The definition of the induced transformation, (4), shows that for any value of (θ_{Fixed}, w_1) then Z has the same distribution under any probability measure indexed by $g^*_{s,Fixed}(\theta_{Fixed}, w_1)$ as given in (3). This implies that although the distribution of W varies freely in a five dimensional space given by Θ_{Fixed} and the range of w_1 the distribution of Z only varies in a bivariate subspace. Choosing $a = \text{sign}\{\gamma + \alpha(\alpha w_1 + \beta)/(\alpha + 1)\}$, $b = a(\beta - w_1)/(\alpha + 1)$, $c = -a(\alpha w_1 + \beta)/(\alpha + 1)$ shows that a maximal invariant parameter under $g^*_{s,Fixed}$ is given by

$$\theta^*_{Fixed} = (\alpha^*, 0, \gamma^*, 1) \text{ for } \alpha^* \in \mathbf{R},\ \gamma^* \geq 0, \text{ and } w_1 = 0, \tag{5}$$

and where γ^* is given by $\text{abs}(\gamma + \delta\alpha)/\sigma$. It implies that in a power comparison of tests which are only depending on the data through the maximal invariant statistic Z under g_s it suffices to look at the maximal invariant parameter θ^*_{Fixed} and the nuisance parameters of the testing problem can be ignored since both α and γ are parameters of interest.

2.4 Modelling the joint distribution of the time series

The second type of statistical models gives a joint distribution of the time series W including the initial value w_1. Such models are often written in unobserved components form,

$$\mathsf{M}_{Random}: \quad \left\{ \begin{array}{ll} w_t = \tau_c + \tau_l t + v_t, & \\ \text{and} \left\{ \begin{array}{l} \Delta v_t = \sigma\varepsilon_t + \alpha v_{t-1} \\ v_1 = \omega\sigma\varepsilon_1 \end{array} \right. & \begin{array}{l} \text{for } t = 2, \ldots, T, \\ \text{for } t = 1, \end{array} \end{array} \right.$$

where the innovations $\varepsilon_1, \ldots, \varepsilon_T$ are independently, standard normally distributed, the parameter $\theta_{Random} = (\alpha, \tau_c, \tau_l, \sigma)$ takes values in $\Theta_{Random} = \mathbf{R}^3 \times \mathbf{R}_+$ and ω is chosen as some function of α. In this model the unit root hypothesis is formulated as

$$\mathsf{H}_{Random}: \qquad \alpha = 0.$$

In this type of models the conditional distribution of $w_2, \ldots, w_T$ given w_1 is autoregressive as in (1) with

$$\beta = \tau_l - \alpha\tau_c, \qquad \gamma = -\alpha\tau_l, \qquad w_1 = \tau_c + \omega\sigma\varepsilon_1, \tag{6}$$

and thus restricting the parameters $\alpha, \beta, \gamma, \sigma^2$ so $\gamma = 0$ if $\alpha = 0$. The exclusion of parameter values where $\gamma \neq 0$ but $\alpha = 0$ turns out to be important when considering power functions of tests derived in the state space model but used in the conditional model.

Several choices of ω have been suggested in the literature. The simplest is

$$\mathsf{M}_{Random,A}: \qquad \omega^2 = 1, \tag{7}$$

which was used by Ahn (1993) and Elliot, Rothenberg and Stock (1996). A variation thereof is the model by Müller and Elliot (2002) which essentially lets $\omega^2 = k$ for some known constant k. Bhargava (1986) uses

$$\mathsf{M}_{Random,B}: \qquad \omega^2 = \left[\begin{array}{ll} 1 & \text{if } |1+\alpha| \geq 1, \\ \{1-(1+\alpha)^2\}^{-1} & \text{if } |1+\alpha| < 1. \end{array} \right. \tag{8}$$

These choices are different in that specification B generates a stationary distribution for $(1+\alpha)^2 < 1$ but this comes at the price that the likelihood has poles for $(1+\alpha)^2 = 1$.

The models can be written in a vector form as

$$A_{\alpha,\omega}(W - D\tau) = \sigma\varepsilon \qquad \text{or} \qquad W = D\tau + \sigma A_{\alpha,\omega}^{-1}\varepsilon, \tag{9}$$

where $D = (1, t-1)_{t=1,\ldots,T}$ is a $(T \times 2)$ matrix and $\varepsilon = (\varepsilon_t)_{t=1,\ldots,T}$ is a T-vector, while

$$\tau = \begin{pmatrix} \tau_c \\ \tau_l \end{pmatrix}, \qquad A_{\alpha,\omega} = \left\{ \begin{array}{cccc} \omega & & & \\ -(1+\alpha) & 1 & & \\ & \ddots & \ddots & \\ & & -(1+\alpha) & 1 \end{array} \right\}.$$

The testing problem given by $\mathsf{M}_{Random}, \mathsf{H}_{Random}$ for either choice of ω is invariant to the sample transformations g_c, g_l and g_s where g_s induces the translation

$$g^*_{s,Random}: \qquad (\alpha, \tau_c, \tau_l, \sigma) \mapsto (\alpha, a\tau_c + b, a\tau_l + c, a\sigma). \tag{10}$$

For this model it is straight forward to impose an invariant reduction with respect to g_l as done by for instance Bhargava (1986) and Elliot, Rothenberg and Stock (1996). The reduced model is based on the marginal distribution of the maximal invariant Y under g_l. This is easy to find since $Y = D_\perp' W$ for some $(T \times T)$-matrix $D_\perp$ with the property $D_\perp' D = 0$ and hence by (9) it holds that $Y = \sigma D_\perp' A_{\alpha,\omega}^{-1}\varepsilon$. The parameter space is accordingly reduced to $\alpha, \sigma \in \mathbf{R} \times \mathbf{R}_+$, so σ remains a nuisance parameter for the testing problem. This reduction of the model results in a likelihood function that is somewhat complicated and which has apparently not been analysed in the literature.

2.5 Maximal invariant parameters for the model M_{Random}

The testing problem $\mathsf{M}_{Random}, \mathsf{H}_{Random}$ reduced by g_l has a nuisance parameter, σ. As in §2.3 this is dealt with by finding a maximal invariant parameter under g_s^*. An example is

$$\theta^*_{Random} = (\alpha^*, 0, 0, 1) \text{ varying in } \Theta^*_{Random} \text{ given by } \alpha^* \in \mathbf{R}. \tag{11}$$

For a power comparison of tests based on the maximal invariant statistic Z under g_s it therefore suffices to look at a simple hypothesis, $\alpha^* = 0$, for the scalar parameter α^*.

3 Tests for the unit root hypotheses

Four unit root test are considered. These could be applied to either of the testing problems, $\mathsf{M}_{Fixed}, \mathsf{H}_{Fixed}$ and $\mathsf{M}_{Random}, \mathsf{H}_{Random}$. The first test is the maximum likelihood ratio test in the conditional model which was first discussed by Dickey and Fuller (1981). The other three tests were suggested by Ahn (1993), Bhargava (1986), and Elliot, Rothenberg, and Stock (1996). They are intended for the models with random initial conditions, but have ad hoc motivations.

The four tests have the common property that they all depend on the time series W only through the maximal invariant Z under g_s although none of them are actually designed with that in mind. After having described the four test statistics their null distributions can therefore be discussed with relative ease.

3.1 Maximum likelihood ratio test statistic in conditional model

The log maximum likelihood ratio test for the testing problem $(\mathsf{M}_{Fixed}, \mathsf{H}_{Fixed})$ is given by

$$LR = -T \log(1 - \hat{\lambda}^2),$$

where $\hat{\lambda}^2$ is the squared sample multiple correlation of the residuals from regressing Δw_t and $(w_{t-1}, t-1)'$ on a constant. Since multiple correlations are invariant to non-singular linear transformations and since the computation involves regression on a constant the variable w_t can be replaced by z_t. In short, the correlation $\hat{\lambda}^2$ can be written as

$$\hat{\lambda}^2 = \widehat{\mathsf{Corr}}^2 \left\{ \Delta z_t, \left(\begin{array}{c} z_{t-1} \\ t-1 \end{array} \right) \middle| 1 \right\}.$$

Tests based on the statistic LR are equivalent to tests based on the F-type statistic Φ_3 suggested by Dickey and Fuller (1981).

3.2 Ahn's test

Ahn's (1993) test is designed for the testing problem $(\mathsf{M}_{Random,A}, \mathsf{H}_{Random,A})$ where $\omega^2 = 1$, see (7). This was later generalised to a cointegration rank test by Lütkepohl and Saikkonen (2000). It is given in terms of the statistic

$$t^2_{Ahn} = \frac{T\lambda^2_{Ahn}}{1-\lambda^2_{Ahn}}$$

where λ^2_{Ahn} is the squared sample correlation of Δz_t and z_{t-1} satisfying

$$\lambda^2_{Ahn} = \widehat{\mathsf{Corr}}^2\left(\Delta z_t, z_{t-1}\right) = \frac{\sum_{t=2}^T \left(\Delta z_t\right)^2}{4\sum_{t=2}^T z^2_{t-1}},$$

due to the identities $2\sum_{t=2}^T z_{t-1}\Delta z_t = (z_T - z_1)^2 - \sum_{t=2}^T \left(\Delta z_t\right)^2$ and $z_1 = z_T = 0$.

3.3 Bhargava's test

Bhargava's (1986) test is designed for the testing problem $(\mathsf{M}_{Random,B}, \mathsf{H}_{Random,B})$ reduced by g_l. Since ω^2 depends on α two different tests are proposed depending on the direction of the alternative. Here the focus will be on tests against $\alpha < 0$ for which Bhargava proposes the test statistic

$$R_2 = T\frac{\sum_{t=2}^T \left(\Delta z_t\right)^2}{\sum_{t=1}^T \left(z_t - \overline{z}\right)^2} \qquad \text{where} \qquad \overline{z} = \frac{1}{T}\sum_{t=1}^T z_t.$$

Bhargava (1986, Proposition 3) proves that this test asymptotically is locally most powerful, see also Lehman (1997, p. 527). The normalisation with T is actually not in Bhargava's paper, but it indicates the appropriate scaling needed to ensure convergence to a non-degenerate distribution under the null hypothesis.

3.4 Elliot, Rothenberg and Stock's DF-GLS$^{\tau}$ test

The Elliot, Rothenberg and Stock (1996) test is designed for the testing problem $(\mathsf{M}_{Random,B}, \mathsf{H}_{Random,B})$ reduced by g_l. It can be computed as follows. First the variables $W_{\overline{\alpha},1} = A_{\overline{\alpha},1}W$ and $D_{\overline{\alpha},1} = A_{\overline{\alpha},1}D$ are formed with $\overline{\alpha} = -13.5/T$. They chose this value through a simulation study that indicated that an asymptotic envelope power function for 5% level tests would reach 50% at this point. Secondly, $W_{\overline{\alpha},1}$ is regressed on $D_{\overline{\alpha},1}$ giving the estimate $\tilde{\tau}_{\overline{\alpha},1} = (D'_{\overline{\alpha},1}D_{\overline{\alpha},1})^{-1}D'_{\overline{\alpha},1}W_{\overline{\alpha},1}$. Using this estimate the original data series is de-trended giving $W^d = W - D\tilde{\tau}_{\overline{\alpha},1}$. Regressing Δw^d_t

on w^d_{t-1} then gives the test statistic

$$t_{ERS} = \sqrt{T}\frac{\lambda_{ERS}}{\sqrt{1-\lambda^2_{ERS}}}$$

where λ_{ERS} is the sample correlation of Δw^d_t and w^d_{t-1} satisfying

$$\lambda_{ERS} = \widehat{\mathsf{Corr}}\left(\Delta w^d_t, w^d_{t-1}\right) = \widehat{\mathsf{Corr}}\left(\Delta z^d_t, z^d_{t-1}\right).$$

The latter equality holds since the regression $W_{\overline{\alpha},1}$ on $D_{\overline{\alpha},1}$ ensures invariance with respect to g_l and the scale invariance of sample correlations ensures invariance with respect to g_s so W can be replaced by Z.

3.5 Null distribution of considered tests

In each of the three models M_{Fixed}, $\mathsf{M}_{Random,A}$, and $\mathsf{M}_{Random,B}$ the restricted parameter spaces under the respective null hypotheses reduce to single point under g^*_s. In either case the maximal invariant X under g_c is given by

$$x_t = w_t - w_1 = \begin{cases} \sum_{s=2}^t \varepsilon_s & \text{for } t > 1, \\ 0 & \text{for } t = 1, \end{cases}$$

and the maximal invariant Z is a function thereof. As a consequence the maximal invariant statistic Z is pivotal and has exactly the same distribution under each of the restricted models H_{Fixed}, $\mathsf{H}_{Random,A}$, and $\mathsf{H}_{Random,B}$. The four test statistics under consideration are all functions of Z and thus leading to similar tests and their critical values are the same in all of the models. Based on a simulation study using 10^6 repetitions the 5% critical values for the LR, t^2_{Ahn}, R_2, and t_{ERS} tests are found to be 12.511, 7.002, 0.348, and -3.039, respectively, for $T = 100$.

The four statistics all converge to non-degenerate distributions under the null hypotheses. Dickey and Fuller (1981) prove this for LR while Ahn (1993, Lemma 1 and Theorem 1) prove this for t^2_{Ahn} and a similar argument can be made for the statistic R_2. Elliot, Rothenberg and Stock (1996) state the result for t_{ERS}.

4 Power comparison for models with fixed initial value

A power comparison is first done for the testing problem M_{Random},H_{Random}, reduced by the transformation g_l, which has the variance parameter σ as a nuisance parameter. The four tests of interest all depend on the maximal invariant statistic Z under g_s. It therefore suffices to reduce the parameter space by the transformation $g^*_{s,Random}$ with associated maximal invariant parameter θ^*_{Random} varying in a univariate space as discussed in §2.5. For each value of the maximal invariant parameter the power of the four tests can then easily be compared with each other and also with point optimal tests found from the Neyman-Pearson Lemma.

In the following point optimal tests are discussed first. Power comparisons for each of the models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ then follow and finally the results from these two slightly different models are compared.

4.1 Envelope power functions under $g^*_{s,Random}$

As a first step towards using the Neyman-Pearson Lemma the likelihood for the maximal invariant parameter θ^*_{Random} based on the data Y is discussed. As deduced in §2.4 the marginal distribution of Y is given by

$$Y = D'_{\perp} A^{-1}_{\alpha,\omega} \varepsilon \overset{\mathsf{D}}{=} \mathsf{N}_T \left\{ 0, D'_{\perp} \left(A'_{\alpha,\omega} A_{\alpha,\omega} \right)^{-1} D_{\perp} \right\},$$

for any θ^*_{Random}. Since $y_1 = y_T = 0$ the above covariance matrix is singular, but $(y_2, \ldots, y_{T-1})$ will have an invertible covariance matrix. In this case the result of Lehman (1997, Exercise 6.3.5) shows that the likelihood based on Y can be written conveniently in terms of the likelihood based on W as

$$L_Y(\alpha) = c_\alpha \max_\tau L_W(\alpha, \tau),$$

where c_α is a function of α, while $\tau = (\tau_c, \tau_l)'$ and $\sigma = 1$. The likelihood based on W can in turn be written as

$$-2 \log L_W(\alpha, \tau) = d_\alpha + \{A_{\alpha,\omega}(W - D\tau)\}' \{A_{\alpha,\omega}(W - D\tau)\},$$

where d_α is another function of α. Introducing the variables $W_{\alpha,\omega} = A_{\alpha,\omega} W$ and $D_{\alpha,\omega} = A_{\alpha,\omega} D$ the maximum with respect to τ is found by least squares regression. Denoting the resulting residuals by $W^\tau_{\alpha,\omega} = \{I_T - D_{\alpha,\omega}\left(D'_{\alpha,\omega} D_{\alpha,\omega}\right)^{-1} D_{\alpha,\omega}\} W_{\alpha,\omega}$ it follows that

$$-2 \max_\tau \log L_X(\alpha, \tau) = d_\alpha + W^{\tau\prime}_{\alpha,\omega} W^\tau_{\alpha,\omega},$$

which can be computed by a singular value decomposition, see Doornik and O'Brien (2002).

Taking two values of θ^*_{Random} satisfying $\alpha = 0$ and $\alpha \neq 0$ the Neyman-Pearson Lemma shows that a most powerful test is given in terms of the log likelihood statistics

$$Q_{Random,\omega} = -2 \log L_Y(0) + 2 \log L_Y(\alpha).$$

By the above considerations the test statistic can be rewritten as

$$Q_{Random,\omega} = e_\alpha + \tilde{Q}_{Random,\omega} \quad \text{where} \quad \tilde{Q}_{Random,\omega} = W^{\tau\prime}_{0,\omega} W^{\tau}_{0,\omega} - W^{\tau\prime}_{\alpha,\omega} W^{\tau}_{\alpha,\omega},$$

where e_α is some constant depending on α. Tests based on $Q_{Random,\omega}$ and $\tilde{Q}_{Random,\omega}$ are therefore equivalent and their critical values can be found by simulation.

An envelope power function for the testing problem $\mathsf{M}_{Random}, \mathsf{H}_{Random}$ reduced by g_l and with parameter space reduced by g_s^* can now be formed. For each $\alpha \neq 0$ its value is given by the power of the point optimal test found above. It is worth noting that this envelope power function will not be envelope power function in the more general testing problem $\mathsf{M}_{Random}, \mathsf{H}_{Random}$ reduced by g_l although one can get that impression from Elliot, Rothenberg and Stock (1996). This is because the latter problem has a nuisance parameter and a composite null hypothesis so the Neyman-Pearson Lemma cannot be used to construct the envelope power function.

4.2 Comparing the tests under $\mathsf{M}_{Random,A}$

Power comparisons for the testing problem $\mathsf{M}_{Random,A}, \mathsf{H}_{Random,A}$ reduced by g_l and with parameter space reduced by g_s^* have previously been reported by Elliot, Rothenberg and Stock. This will be redone here in the situation where $T = 100$ which matches the sample size in the empirical illustration of §6.1.

Power curves for the four tests of §3 are reported in Figure 1 along with the power envelope. These simulated curves as well as subsequent simulation results are based on 10^6 repetitions with the same set of random numbers for each of the simulated probability measures, and the simulated critical values reported in §3.5. If the critical values had been known this would result in a simulation standard deviation of about 0.07% for powers in the range $5\% - 95\%$. Since the critical values also simulated the overall standard deviation on the reported numbers is a little larger. For a more precise evaluation of the simulation uncertainty, see Paruolo (2001).

The results largely confirm those of Elliot, Stock and Rothenberg (1996). The maximum likelihood ratio test LR from the conditional model M_{Fixed} does not perform well here, so a considerable gain in power is obtained by modelling the initial value. In their work Elliot, Stock and Rothenberg (1996) reported that the power of their test t_{ERS} is visibly indistinguishable from the asymptotic power curve. As seen from Figure 1 this is not the case at this sample length, while a simulation study with $T = 500$ confirmed their large sample result. Having said this the test t_{ERS} is more powerful than the other considered tests for a large part of the parameter space.

As a final remark it is noted that the tests t^2_{Ahn}, R_2, t_{ERS} are designed to be used against negative alternatives, $\alpha^* < 0$. Unlike the LR test these tests are indeed biased when used against positive alternatives.

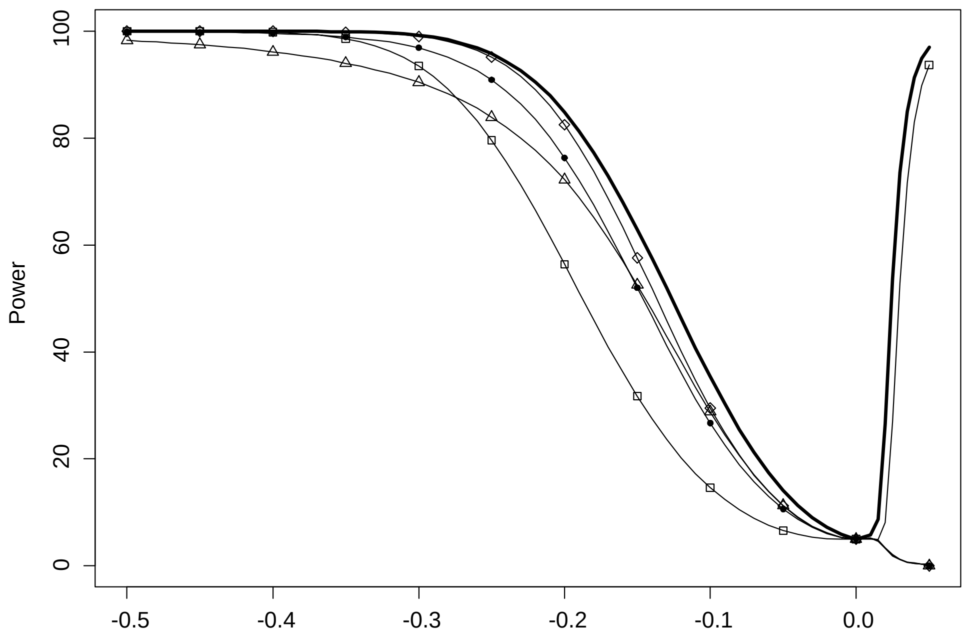

Figure 1: Power curves as a function of α^* for 5% tests for the testing problem $\mathsf{M}_{Random,A}, \mathsf{H}_{Random,A}$ reduced by g_l and with parameter space reduced by g_s^* for $T = 100$. The upper curve is the envelope power function, while the other four curves are: boxes: LR, triangles: t^2_{Ahn}, bullets: Bhargava's R_2, and diamonds: t_{ERS}.

4.3 Comparing the tests under $\mathsf{M}_{Random,B}$

Simulated power curves for the testing problem $\mathsf{M}_{Random,B}, \mathsf{H}_{Random,B}$ reduced by g_l and with parameter space reduced by g_s^* are reported in Figure 2. Only negative alternatives, $\alpha^* < 0$, are considered as the models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ are identical for $\alpha^* > 0$, see (7), (8).

The overall ranking of the different tests is the same as in Figure 1. The tests LM_{Ahn} and t_{ERS} which are based on the model $\mathsf{M}_{Random,A}$ rather than $\mathsf{M}_{Random,B}$ are a bit further away from the envelope power function than seen in Figure 1.

4.4 Comparing the models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$

It is tempting to compare the power functions in Figure 1 and 2 since the null hypotheses are the same. The alternatives are actually different so interpretations should be done with care. A pointwise comparison of the envelope power functions shows a difference of up to 6% in favour of the alternatives of model $\mathsf{M}_{Random,A}$. For the t^2_{Ahn} and t_{ERS} tests the differences are up to 9% and 7%, respectively, in favour of model

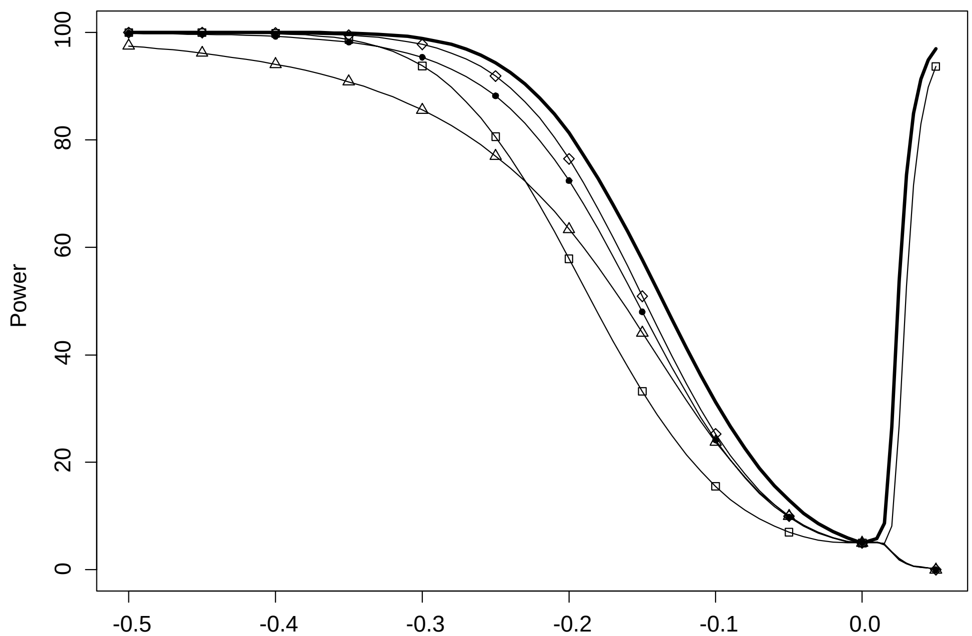

Figure 2: Power curves as a function of α^* for 5% tests for the testing problem $\mathsf{M}_{Random,B}, \mathsf{H}_{Random,B}$ reduced by g_l and with parameter space reduced by g_s^* for $T = 100$. The upper curve is the envelope power function, while the other four curves are: boxes: LR, triangles: t^2_{Ahn}, bullets: Bhargava's R_2, and diamonds: t_{ERS}.

$\mathsf{M}_{Random,A}$. For the test R_1 which is designed for the model $\mathsf{M}_{Random,B}$ this difference is reduced to only 4%. The LR-test is most robust to changing the distributional assumption to the initial value, w_1, by being shifted only by up to 2% and that in favour of the model $\mathsf{M}_{Random,B}$ with the more diffuse assumptions.

5 Power comparison for the conditional model

The testing problem M_{Fixed},H_{Fixed}, is now considered, with parameter space reduced by the transformation $g^*_{s,Fixed}$ with associated maximal invariant parameter θ^*_{Fixed}, see §2.3. An asymptotic analysis shows that the tests t^2_{Ahn}, R_2, and t_{ERS} are biased against the alternative $\alpha = 0$, but $\gamma \neq 0$, that is when the time series is a random walk with a quadratic trend. Since this alternative is not directly of interest a simulations study is used to compare the four tests with each other and with point optimal tests for other points in the alternative.

5.1 Asymptotic theory for the four tests

The following theorem shows that in large samples the test statistics t^2_{Ahn}, R_2, and t_{ERS} converge to zero while LR diverges when $\alpha = 0$, but $\gamma \neq 0$. This implies that the power of the first three tests goes to zero and these tests are asymptotically biased while the LR test is consistent against those alternatives. The proof is given in an Appendix.

Theorem 1 *Consider the model* M_{Fixed} *with* $\alpha = 0$ *but* $\gamma \neq 0$. *Then, for* $T \to \infty$,

$$LR \xrightarrow{\mathsf{P}} \infty, \qquad \textit{while} \qquad t^2_{Ahn} \xrightarrow{\mathsf{P}} 0, \qquad R_2 \xrightarrow{\mathsf{P}} 0, \qquad t_{ERS} \xrightarrow{\mathsf{P}} 0.$$

The parameter values $\alpha = 0, \gamma \neq 0$ in them selves are not so interesting in applications as they correspond to random walks with quadratic trends. In finite samples the asymptotic bias at these points will contaminate the power function in a wider area of the parameter space where $\alpha \neq 0$. This contamination will be studied by simulation methods in the following.

5.2 Envelope power functions under $g^*_{s,Fixed}$

The nuisance parameters of the testing problem $\mathsf{M}_{Fixed}, \mathsf{H}_{Fixed}$ are eliminated when reducing the parameter space to Θ^*_{Fixed} given by the maximal invariant parameter θ^*_{Fixed} under $g^*_{s,Fixed}$ as reported in §2.3. The null hypothesis is then reduced to a simple hypothesis allowing the Neyman-Pearson Lemma to be use in comparing two points θ^*_{Fixed} in the parameter space satisfying $(\alpha^*, \gamma^*) \neq 0$ and $(\alpha^*, \gamma^*) = 0$, respectively. The lemma shows that a most powerful test rejects for small values of the log likelihood ratio statistics

$$\begin{aligned} Q_{Fixed} &= -2 \log L_Y (0, 0) + 2 \log L_Y (\alpha^*, \gamma^*) \\ &= \sum_{t=1}^{T} (\Delta y_t)^2 - \sum_{t=1}^{T} \{\Delta y_t - \alpha^* y_{t-1} - \gamma^* (t-1)\}^2 . \end{aligned}$$

5.3 Comparing the tests under M_{Fixed}

Figure 3 and 4 show level curves of the envelope power function together with level curves of the power functions for the LR-test and the t^2_{Ahn}, respectively, for $T = 100$. They show how the asymptotic results of Theorem 1 contaminate the power function in a wide area of the parameter space in a finite sample. These plots are based on fine a grid with intervals 0.005 in α^* and 0.00025 in γ^* with each point based on simulations with 10^6 repetitions.

From Figure 3 it is seen that both the envelope power function and the power function for the LR are shaped like valleys with minimum at the origin. None of

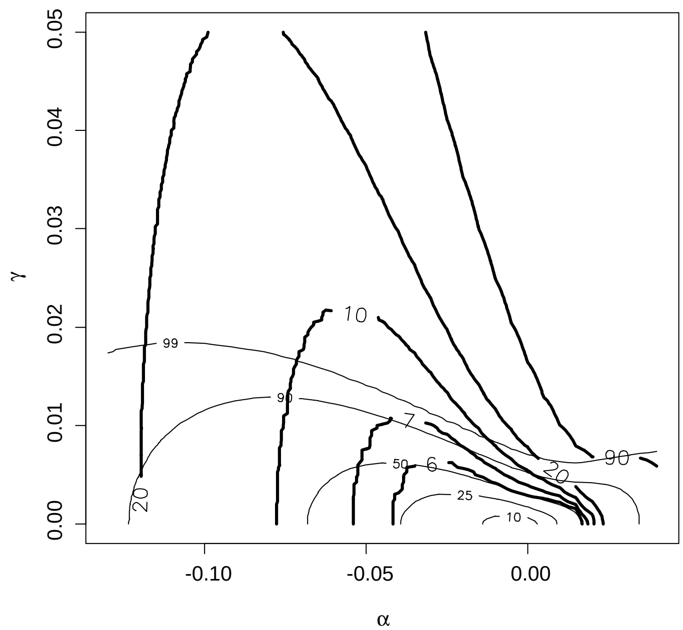

Figure 3: Level curves for 5% level point optimal tests (thin lines) and for LR test (thick lines) in the model M_{Fixed}.

the level curves are elliptic as would have been the case in a standard test situation. The difference between the power of the LR and the envelope power function is more pronounced than seen in §4, in particular for $\alpha^* < 0$. The reason is presumably that here the envelope power function is based on the full model M_{Fixed} rather than a model reduced by g_l.

Figure 4 shows that the power function for the t^2_{Ahn}-test is very different in that it slopes towards 0 when either α^* or γ^* increases. The power surfaces for the R_2, and t_{ERS}-tests are similar in shape and therefore not reported. For the t_{ERS}-test the surfaces is nearly exactly identical whereas the R_2-surface is slightly less steep with a little less power towards the bottom left of the plotted area and slightly more power towards the right of the area.

A first comparison of Figure 3 and 4 shows that the LR test is vastly superior outside the bottom left of the plotted area, which is roughly when γ^* is larger than about $2(\alpha^*)^2$. It is also seen that the power functions arising for varying α^* but $\gamma^* = 0$ is more or less identical to what is seen in the models M_{Random} in Figures 1 and 2.

A more formal comparison of tests with several degrees of freedom can be done in terms of their average power or their stringency as suggested by Wald (1943). Wald showed that for large samples from standard models the maximum likelihood

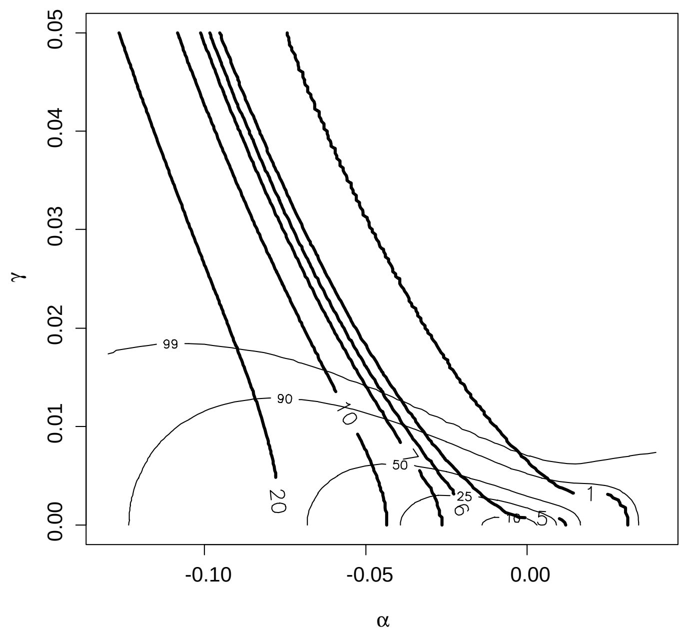

Figure 4: Level curves for 5% level point optimal tests (thin lines) and for Ahn's test (thick lines) in the model M_{Fixed}.

ratio test is optimal according to both criterions, see also Andersen, Borgan, Gill and Keiding (1991, p. 603) and Lehmann (1997, p.525f), for more recent overviews.

The criterion of finding the test with best average power is difficult to use in this situation, in that the average power is defined as the integral of the power function over a surface K chosen as a level curve of the envelope power function and integrated with respect to a certain weight function. This weight function is defined through an asymptotically local argument so each arc of K has same weight as that of the projection on to circle with uniform distribution. This integral would be rather hard to compute analytically as it involves locally asymptotically Brownian functionals as described by Jeganathan (1995).

It is more feasible to compare the tests in terms of their stringency. The idea is to find the maximal shortcoming of each test which is the maximal power loss over the surface K so the test with smallest maximal shortcoming is most stringent. Looking at the 99% level curve of the envelope power function, say, the maximal shortcoming of the LR test is about 91%. For the t^2_{Ahn} test the maximal shortcoming is more than 98.9%. The maximal shortcoming reduces to 98% if only the alternatives $\alpha^* < 0$ are considered, so in any case the LR test is substantially more stringent than the t^2_{Ahn} test. The same result applies for other level curves of the envelope power function and also for the other considered tests.

Regardless of the choice of method for comparison there is some variation in the results depending on which area of the parameter space is considered. In §6 an empirical model is therefore used to highlight a parameter value that may be of particular interest in applications and power comparisons are done for this value.

6 Comparing tests across models

Apart from a brief discussion in §4.4 the power comparisons in §4, 5 have been done within specific models which is a situation where statistical theory gives guidelines. It is an entirely different issue to compare across models since the different families of probability measures are not nested. Still it is of considerable interest to compare across models.

A power comparison only makes sense if one is indifferent to the different interpretations of the different tests and models. Interpreting tests by the strong repeated sampling principle, see Cox and Hinkley (1974, p. 45), the idea of testing is to compare the observed sample with hypothetical samples from the same model. In the present situation this relates to the invariant reduction by g_l in the model M_{Random} and to the way the initial value is modelled. In model M_{Fixed} the initial value is in all the hypothetical samples fixed at the observed initial value, whereas in the models M_{Random} the initial value is random so that it starts sometimes below and sometimes above the trend line. Whether there is an indifference to these interpretations must depend on the specific situation.

To get comparable alternatives a UK production series is considered and each of the models M_{Fixed}, $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ is estimated. In the first instance the models are assumed to be well-specified to facilitate model comparison. This is then followed by a discussion of the validity of the model assumptions.

6.1 Models for UK production

Figure 5(a) shows a time series, w_t, of quarterly log real total final expenditure for the UK for the period 1963:1-1989:3. A detailed econometric analysis of this series in conjunction with other macroeconomic series can be found in Doornik, Hendry and Nielsen (1998), who also list a series of previous papers analysing this data set. To match the above discussion the initial value $w_1 = 11.038$ is subtracted from the series rendering a times series $x_t = w_t - w_1$ with $x_1 = 0$ and $T = 107$.

The conditional model M_{Fixed} is estimated by least squares regression giving

$$\Delta x_t = -0.081 x_{t-1} + 0.0128 + 0.00051\,(t-1) + 0.0139\hat{\varepsilon}_t. \qquad (12)$$

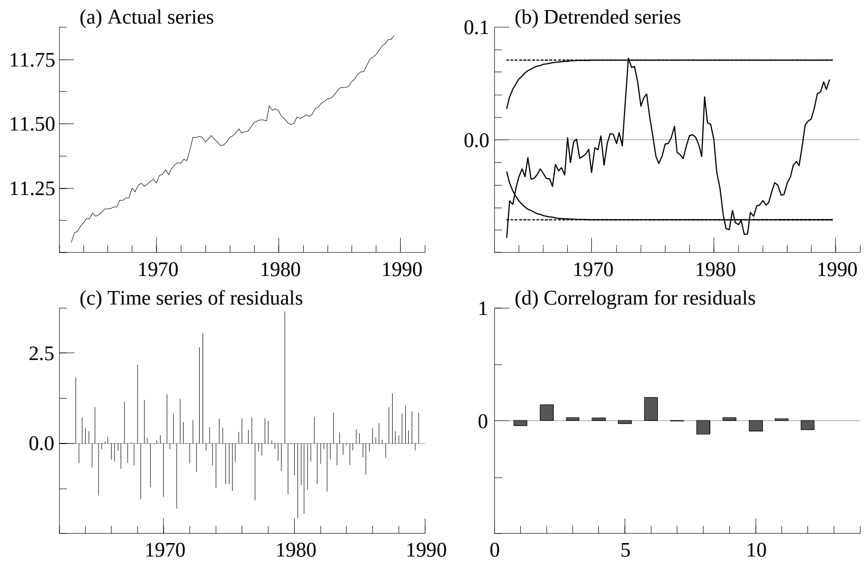

Figure 5: (a) the actual series w_t. (b) the detrended series $\tilde{v}_t$ along with confidence bands derived from models $\mathsf{M}_{Random,A}$ (solid) and $\mathsf{M}_{Random,B}$ (dashed). (c) & (d) time series and correlogram of scaled residuals from fitted model.

The unconditional model $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ are estimated through a two stage procedure, as outlined by Lütkepohl and Saikkonen (2000). In the first step the regression (12) is carried out. Using the definitions in (6) this leads to estimates

$$\tilde{\tau}_l = -\hat{\gamma}/\hat{\alpha} = 0.0063, \qquad \tilde{\tau}_c = (\tilde{\tau}_l - \hat{\delta})/\hat{\alpha} + w_1 = 0.080 + 11.038,$$

and a de-trended variable $\tilde{v}_t = w_t - \tilde{\tau}_c - \tilde{\tau}_l t$ can be computed. A second regression then gives the result

$$\Delta\tilde{v}_t = -0.081\tilde{v}_{t-1} + 0.0139\tilde{\varepsilon}_t. \tag{13}$$

6.2 Comparing power

The null hypotheses of the testing problems $(\mathsf{M}_{Fixed}, \mathsf{H}_{Fixed})$, $(\mathsf{M}_{Random,A}, \mathsf{H}_{Random,A})$, and $(\mathsf{M}_{Random,B}, \mathsf{H}_{Random,B})$ are comparable in that they all generate random walks $w_1 + \sum_{s=2}^{t} \varepsilon_t$ although with varying properties of the initial value. Since the empirical models (12) and (13) are estimated from the same data they are also considered to be comparable. The corresponding maximal invariant parameters defined in (5) and (11) are estimated by

$$\hat{\theta}^*_{Fixed} = (-0.081, 0.038), \qquad \hat{\theta}^*_{Random} = -0.081,$$

model	alternative	LR	t^2_{Ahn}	R_2	t_{ERS}
$\mathsf{M}_{Random,A}$	$\hat{\theta}^*_{Random}$	11.7	23.1	21.4	23.5
$\mathsf{M}_{Random,B}$	$\hat{\theta}^*_{Random}$	12.5	19.2	19.2	20.1
M_{Fixed}	$\hat{\theta}^*_{Fixed}$	16.7	5.3	10.3	7.0

Table 1: Simulated power of 5% level tests for alternatives given by the estimated maximal invariant parameters for the UK production data. Based on $T = 107$ and 10^6 repetitions and critical values of 12.506, 6.989, 0.326, -3.025 respectively.

which define points in the alternatives of the three testing problems. Table 1 shows the power at these points.

There are several observations to be made from Table 1. The tests designed for the models M_{Random} perform well in those models but poorly in model M_{Fixed} and in particular the test t^2_{Ahn} has virtually no power in the model M_{Fixed}. The reason for the poor performance in the model M_{Fixed} is that the estimated trend parameter is so large in relation to the autoregressive parameter that the asymptotic bias described in Theorem 1 is influential. The LR-test is a maximum likelihood ratio test in the model M_{Fixed} and performs well in that model but relatively poorly in M_{Random}.

Comparing the results for the two models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ it is seen that the LR-test is most robust to the different assumptions to the initial value as discussed in §4.4. Of the remaining tests the R_2-test is most robust. This is probably because it is designed for the model $\mathsf{M}_{Random,B}$ while the t^2_{Ahn} and t_{ERS}-tests are designed for $\mathsf{M}_{Random,A}$.

Table 1 indicates that for this particular choice of parameters there is some variation in the maximal power that can be achieved in the three models with the model $\mathsf{M}_{Random,A}$ having the most powerful tests. These results will of course depend a lot on the considered parameters with results that would be less favourable to $\mathsf{M}_{Random,A}$ if α^* were a little smaller or γ^* were a little larger.

6.3 Testing the model assumptions

For any econometric analysis it is important to establish to what extent the model assumptions are met. There is an intriguing difference between the kind of misspecification tests that can be performed for the models since the model M_{Fixed} can be tested with consistent tests while the assumptions to the initial value in the models M_{Random} cannot be tested consistently. While the consistency of course not helps much in a finite sample it does indicate that in principle more powerful conclusions could be reached by awaiting the arrival of future observations.

In order to use a model like M_{Fixed} the distributional assumptions to $w_2, \ldots, w_T$ given w_1, or, equivalently, to $\varepsilon_2, \ldots, \varepsilon_T$ must be checked. This issue is addressed more generally by Andreou and Spanos (2003) and their discussants. The assumptions can be tested informally by graphing for instance the residuals and their correlogram as done in Figure 5(c,d) or formally through statistical tests that are consistent. In this way a normality test shows that non-normality is just significant at a 1% level. This signals that it is potentially not straight forward to draw inference from the estimated model (12). In their analysis of this data series Hendry and Mizon (1993) addressed this issue by introducing a dummy variable taking the value one in 1972:4, 1973:1, 1979:2 matching fiscal expansions. While it is not the point of this paper to consider such modifications it is the case that practitioners have means to find different and possibly more appropriate descriptions of the data.

When using the model M_{Random} it is necessary to test the assumptions to the conditional distribution of $w_2, \ldots, w_T$ given w_1 which can be done as above as well as testing the distributional assumptions to w_1. Figure 5(b) shows the de-trended series $\tilde{v}_t$ together with pointwise confidence bands for the marginal distribution of v_t based on the variances $\sigma^2\{1-(1+\alpha)^{2t}\}/\{1-(1+\alpha)^2\}$ and $\sigma^2/\{1-(1+\alpha)^2\}$, respectively, in the models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$. It is seen that even for the model $\mathsf{M}_{Random,B}$ there is a very small probability of observing an $\tilde{v}_1$ that fits worse with the distributional assumption to v_1. Such a test concerning the distribution of a single observation will inevitably be inconsistent, so apart for a more precise estimate of the confidence band no power can be gained by looking at a longer time series. The reason that the models seem to perform so badly is possibly that the growth is above average in the first two years of this time series, see Figure 5(a). Better fitting models would probably be found by discarding the first eight observations and thereby loosing information which would not be palatable in many applications.

7 Conclusions

The power functions of four unit root tests defined by the LR, t^2_{Ahn}, R_2 and t_{ERS} statistics have been compared with each other and envelope power functions within three different models. The notion of maximal invariant parameters has been used more explicitly than before to simplify the problem of comparing power by essentially reducing the null hypotheses to simple hypotheses which in turn facilitates use of the Neyman-Pearson Lemma.

It has been demonstrated that the power of the four unit root tests is specific to the model that generates the data. Out of the considered four tests the test suggested by Elliot, Rothenberg and Stock (1996) seems most powerful within the models M_{Random} with random initial value while the LR-test is most stringent in the model with fixed initial value M_{Fixed} where it is a maximum likelihood ratio test. The LR-test appears to be most robust to variations in the specification of the model in that the power varies least from model to model, whereas the tests based on t^2_{Ahn}, R_2 and t_{ERS} have more power in the model M_{Random} and are biased in the model M_{Fixed}.

In empirical work the model M_{Fixed} seems easier to use than $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$. While there potentially is a power gain by working with the latter models rather than the former, this is based on two assumptions that need to be motivated in each application. First, the models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ involve an additional distributional assumption to the initial observation w_1. In cases like the example discussed in §6 where this assumption is not met any power gain from working with $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ rather than M_{Fixed} will be illusive. Secondly, the tests based on R_2 and t_{ERS} as well as the power results for the models $\mathsf{M}_{Random,A}$ and $\mathsf{M}_{Random,B}$ are based on an invariant reduction of the models by the transformation g_l. For each application this reduction would have to motivated in terms of the subject matter, which is not simple when modelling a variable like the logarithm of production as in the example of §6. When choosing the model M_{Fixed} the inferences concerning unit roots are therefore less distracted by issues related to the initial value and to invariant reductions.

8 Acknowledgments

The numerical results were generated using Ox, see Doornik (1999) and PcGive, see Doornik and Hendry (2001), while the power curves were drawn using R, see Ihaka and Gentleman (1996). Discussions with D.R. Cox, G. Hillier, S. Johansen and N. Shephard and research assistance from T. Kurita are gratefully acknowledged.

9 Appendix: Proof of Theorem 1

Consider the model M_{Fixed} with $\alpha = 0$, $\gamma \neq 0$. Since the four tests are scale invariant it suffices to assume $\sigma^2 = 1$ and to work with Y instead of Z. Then

$$\Delta w_t = \beta + \gamma(t-1) + \varepsilon_t, \qquad w_t = w_1 + \beta\,(t-1) + \frac{\gamma}{2} t\,(t-1) + \sum_{u=2}^{t} \varepsilon_u.$$

The maximal invariant Y then satisfies

$$y_t = \frac{\gamma}{2}\,(t-1)\,(t-T) + \mathrm{O}_{\mathsf{P}}(T^{1/2}), \qquad \Delta z_t = \frac{\gamma}{2}\,(2t - T - 2) + \mathrm{O}_{\mathsf{P}}(1).$$

It holds that $2\sum_{t=2}^{T} y_{t-1}\Delta y_t = (\sum_{t=2}^{T}\Delta y_t)^2 - \sum_{t=2}^{T}(\Delta y_t)^2$, while, for $T \to \infty$,

$$T^{-3}\sum_{t=2}^{T}(\Delta y_t)^2 = \frac{\gamma^2}{12} + \mathrm{o}_{\mathsf{P}}(1), \qquad T^{-5}\sum_{t=2}^{T} y_{t-1}^2 = \frac{\gamma^2}{120} + \mathrm{o}_{\mathsf{P}}(1),$$

$$\sum_{t=2}^{T}\Delta y_t = 0, \qquad T^{-3}\sum_{t=2}^{T} y_{t-1} = \frac{-\gamma}{12} + \mathrm{O}_{\mathsf{P}}(T^{-1}),$$

$$T^{-3}\sum_{t=2}^{T}(\Delta y_t)\,(t-1) = \frac{\gamma}{12} + \mathrm{o}_{\mathsf{P}}(1), \qquad T^{-4}\sum_{t=2}^{T} y_{t-1}\,(t-1) = \frac{-\gamma}{24} + \mathrm{O}_{\mathsf{P}}(T^{-1}).$$

It then follows that

$$\widehat{\mathsf{Corr}}^2\left\{\Delta z_t, \begin{pmatrix} z_{t-1} \\ t-1 \end{pmatrix} \middle|\, 1\right\} \overset{\mathsf{P}}{\to} 1, \qquad T^2\lambda_{Ahn}^2 \overset{\mathsf{P}}{\to} \frac{5}{2}, \qquad T^2 R_2 \overset{\mathsf{P}}{\to} 60,$$

and therefore LR diverges, while LM_{Ahn} and R_2 converge to zero.

It is left to consider the test given by t_{ERS}. Since

$$D'_{\overline{\alpha},1}D_{\overline{\alpha},1} = \begin{pmatrix} T & T^2/2 \\ T^2/2 & T^3/3 \end{pmatrix}\left\{1 + \mathrm{O}(T^{-1})\right\},$$

$$D'_{\overline{\alpha},1}Y_{\overline{\alpha},1} = \frac{-\gamma}{12}\begin{pmatrix} T^3 \\ T^4/2 \end{pmatrix}\left\{1 + \mathrm{O}_{\mathsf{P}}(T^{-1})\right\},$$

the de-trending estimator satisfies

$$(D'_{\overline{\alpha},1}D_{\overline{\alpha},1})^{-1}(D'_{\overline{\alpha},1}Y_{\overline{\alpha},1}) = \left[-\frac{\gamma}{12}T^2\left\{1 + \mathrm{o}_{\mathsf{P}}(1)\right\}, \mathrm{O}_{\mathsf{P}}(1)\right]'$$

and therefore it holds, uniformly in t that

$$y_t^d = y_t + \gamma T^2/12 + t(1), \qquad \Delta y_t^d = \Delta y_t + \mathrm{O}_{\mathsf{P}}(1).$$

Applying the above results for the sums of y it is seen that

$$\sum_{t=2}^{T}\Delta y_t^d = \mathrm{O}_{\mathsf{P}}(T), \qquad T^{-3}\sum_{t=2}^{T}\left(\Delta y_t^d\right)^2 \overset{\mathsf{P}}{\to} \frac{\gamma^2}{12}, \qquad T^{-5}\sum_{t=2}^{T}\left(y_{t-1}^d\right)^2 \overset{\mathsf{P}}{\to} \frac{\gamma^2}{720},$$

and thus $T^2\lambda_{ERS}^2$ converges to 15 in probability, implying that t_{ERS} converges to 0.

10 References

Ahn, S.K. (1993) Some tests for unit roots in autoregressive-integrated-moving average models with deterministic trends, *Biometrika*, 80, 855-868.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993) *Statistical models based on counting processes.* New York: Springer.

Andreou, E. and Spanos, A. (2003) Statistical adequacy and the testing of trend versus difference stationarity, *Econometric Reviews*, 22, 217-237.

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

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

Dickey, D.A. and Fuller, W.A. (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, 1057-1072.

Doornik, J.A. (1999) *Object-oriented matrix programming using Ox*, 3rd ed. London: Timberlake Consultants Press.

Doornik, J.A. and Hendry, D.F. (2001) *Empirical econometric modelling using PcGive 10, vol. 1.* London: Timberlake Consultants Press.

Doornik, J.A., Hendry, D.F. and Nielsen, B. (1998) Inference in cointegrating models: UK M1 revisited. *Journal of Economic Surveys*, 12, 533-572.

Doornik, J.A. and O'Brien, R.J. (2002) Numerically stable cointegration analysis. *Computational Statistics & Data Analysis*, 41, 185-193.

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

Hendry, D.F., and Mizon, G.E. (1993) Evaluating dynamic econometric models by encompassing the VAR, in *Models, methods, and applications of econometrics*, ed. by P.C.B. Phillips. Oxford: Blackwells.

Ihaka, R. and Gentleman, R. (1996) R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, 5, 299-314.

Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models, *Econometric Theory*, 11, 818-887.

Lehman, E.L. (1997) *Testing statistical hypotheses.* 2nd edition. New York: Springer.

Lütkepohl, H., and Saikkonen, P. (2000) Testing for the cointegrating rank of a VAR process with a time trend, *Journal of Econometrics*, 95, 177-198.

Müller, U.K., and Elliot, G. (2003) Tests for unit roots and the initial condition, *Econometrica*, 71, 1269-1286.

Nielsen, B. and Rahbek, A. (2000) Similarity issues in cointegration analysis, *Oxford Bulletin of Economics and Statistics, 62*, 5-22.

Paruolo, P. (2001) On Monte Carlo estimation of relative power, *Econometrics Journal*, 5, 65-75.

Wald, A. (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical Society*, 54, 426-482.