Title

Confidence Intervals for Autoregressive Coefficients Near One

Authors
Elliott, Graham
STOCK, JAMES H

Publication Date
2000-07-01

2000-19

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

CONFIDENCE INTERVALS FOR AUTOREGRESSIVE COEFFICIENTS NEAR ONE

BY

GRAHAM ELLIOTT

AND

JAMES H. STOCK

DISCUSSION PAPER 2000-19
JULY 2000

Confidence Intervals for Autoregressive Coefficients Near One

Graham Elliott
Department of Economics
University of California, San Diego
9500 Gilman Drive, LA JOLLA, CA, 92093-0508

and

James H. Stock
Kennedy School of Government
Harvard University
CAMBRIDGE, MA, 02138

First Draft: October 1997
This Draft: June 2000

Abstract: Often we are interested in the largest root of an autoregressive process. Available methods rely on inverting t-tests to obtain confidence intervals. However, for large autoregressive roots, t-tests do not approximate asymptotically uniformly most powerful tests and do not have optimality properties when inverted for confidence intervals. We exploit the relationship between the power of tests and accuracy of confidence intervals, and suggest methods which are asymptotically more accurate than available interval construction methods. One interval, based on inverting the P_T or Q_T statistic, has good asymptotic accuracy and is easy to compute.

JEL classification: C32.

Keywords: Unit root, confidence intervals, Point Optimal Tests.

Corresponding Author: Graham Elliott.

We thank R. Engle, C.Granger, J.Hamilton, H.White, two anonymous referees and seminar participants of Camp Econometrics 1998, NBER/NSF times series meetings, the University of NSW and University of Virginia for comments. Elena Pesavento provided excellent research assistance. This research was supported in part by NSF grants SBR-9409629, SABR-9730489 and SBR-9720675. All errors are ours.

I. Introduction

The value of the largest autoregressive root of a univariate time series can be of interest for various reasons. The largest root can be of direct economic interest. For example, Sargent (1998) argues that learning about the largest root in inflation (specifically, learning that the coefficients on lagged inflation in a Phillips Curve sum to one) led to the willingness of Governors of the U.S. Federal Reserve Systems to initiate the disinflationary recessions of the early 1980's. Alternatively, if primary interest is in multivariate regression models, inference in such circumstances typically depends on the value of the largest autoregressive root of the regressor(s), see for example Cavanagh et. al. (1995). In both circumstances, researchers might be interested not just in whether or not the largest autoregressive root is one or not, but in the more general question of what the root is and in the construction of confidence intervals for that root.

Various methods now exist for the construction of such confidence sets. Stock (1991) proposed constructing confidence sets by inverting augmented Dickey Fuller (1979) (ADF) t-tests, and showed the asymptotic validity of this procedure using the local to unity asymptotic framework of Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987) and Phillips (1987). Andrews (1993) proposed finite sample methods for the construction of confidence intervals in the Gaussian AR(1) model with $\rho \in (-1,1]$. Hansen (1999) proposes a bootstrap procedure that is asymptotically valid both for roots in a neighborhood of one and for ρ bounded away from one. These papers focus on tests constructed by inverting t ratios. This choice is, however, one of convenience. In many economic settings, standard t ratios produce tests that are asymptotically uniformly most powerful invariant (with Gaussian errors) (UMPI), and inverting these tests results in asymptotically uniformly most accurate invariant confidence sets. But in the unit root problem this is not the case: as shown by Elliott et. al (1996), there does not exist a UMP or UMPI test of the unit root hypothesis. Moreover, the ADF unit root tests has power which is far from the Gaussian power envelope, and far less than alternative (feasible) point optimal tests.

This paper proposes new asymptotic methods for the construction of confidence sets for the largest autoregressive root, ρ. These methods build on the theory of asymptotically optimal tests in Gaussian autoregressions and extend these results to the realm of confidence intervals for ρ when ρ is local to unity. The specific approach here is to consider confidence intervals constructed as the acceptance region of a sequence of tests, where each test in the sequence is constructed as the point optimal test of a particular null against a particular alternative. The corresponding confidence set is the set of values that are not rejected by this sequence of tests.

The key motivation for this approach is the link between the power of tests and the accuracy of confidence intervals: the hope is that the asymptotic point optimality of the constituent tests will induce good accuracy for the resulting confidence set. These 'sequence tests' are computationally intensive, so in addition we consider approaches based on approximate optimality results in the unit root case that are computationally simpler and thus more appealing as practical methods.

The paper is set out as follows. The next section details the main model of the data we are concerned with, and discusses the theory for classical confidence intervals and the relationship between such confidence intervals and power of tests. The third section discusses a number of methods for the construction of confidence intervals based on powerful tests. The fourth section considers an alternate assumption on the initialization of the data when roots are less than one. In section five we examine exact methods for the calculation of critical values required to construct the intervals. We then evaluate the methods in large and small samples in the sixth section. A final section concludes. All proofs are contained in an appendix.

II. The Model and the Theory of Confidence Intervals.

The time series y_t is assumed to have the representation

$$y_t = d_t + u_t, \qquad t = 1,\ldots T. \tag{1}$$

where

$$u_t = \rho u_{t-1} + \nu_t, \quad t = 2,\ldots T \tag{2}$$

and $\{d_t = \beta' z_t\}$ are deterministic components with $z_t=1$ or $z_t= [1,t]'$ (which we will call the trend), $\nu_t=C(L)\varepsilon_t$ is a zero mean stationary process with finite autocovariances $\gamma(k) = E\nu_t\nu_{t-k}$ and $0 < \omega^2 = \sum_{k=-\infty}^{\infty}\gamma(k) < \infty$ such that $T^{-1/2}\sum_{i=1}^{[Ts]}\nu_i \Rightarrow \omega W(s)$ where W(s) is a standard Brownian Motion, [.] denotes the greatest lessor integer function and $\Rightarrow$ denotes weak convergence.

Intervals will be derived and examined under two different assumptions on the initial condition u_1. These are

A1. u_1 has mean zero and variance $\gamma(0)$.

A2. u_1 has mean zero and if $\rho \geq 1$ has variance $\gamma(0)$, otherwise u_1 has variance $\gamma(0)/(1-\rho^2)$.

We model the largest coefficient as being local to unity, i.e. ρ =1+c/T (see Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987)). Under the latter of these two sets of assumptions, we have the following limit result for u_t;

$$\frac{1}{w\sqrt{T}} u_{[Ts]} \Rightarrow W_c(s) \equiv e^{cs} W_c(0) + e^{cs} \int_0^s e^{-cr} dW(r) \qquad (3)$$

where $W_c(0)$ ~N(0,$(-2c)^{-1}$) is independent of $W_c(s)$, s>0 if c<0 and is zero otherwise. From Lemma 2 of Elliott (1999) this distributional result is continuous in c at c=0. In the fixed initial case we have the limit result in (3) with $W_c(0)$=0.

For the time series representation above and tests for the null of a unit root against alternatives less than one, the results of Elliott et al (1996) and Elliott (1999) show that no uniformly most powerful test exists, even asymptotically. Thus there is no uniformly most accurate confidence interval for c=T(ρ -1) asymptotically. Despite the lack of existence of a uniformly most accurate confidence interval for this model, the general theory underlying the construction of confidence intervals (see Lehmann (1994), section 5.6) can be used to construct confidence intervals for ρ that have useful accuracy properties. Further, the theory can also be employed to evaluate confidence intervals constructed from the inversions of available tests, such as proposed by Stock (1991).

Suppose that one is interested in constructing a confidence interval for the (one dimensional) parameter θ . A 100(1-α)% confidence set $S_\alpha(y)$ where y is the data is a set-valued function of the data that has the property that $P_q[q \in S_a(y)] \geq 1 - a$ for all values of θ. Such a confidence set can be derived from the following sequence of tests. Suppose we have a sequence of tests of asymptotic size α for the hypothesis H_0: θ = θ* vs H_a: θ ≠ θ* for all θ* ∈ Θ where Θ is the parameter space (the sequence is one test for each θ*, in practice the test could be the same at each θ). We conduct these tests for all possible θ* ∈ Θ, and define $S_\alpha(y)$ as the set of θ* that we cannot reject. Such sets have the desired property that in the limit then
$P_q[q \in S_a(y)] \geq 1 - a$ for any true value for θ.

The probability that any θ* is included in the confidence set, denoted $P_{q'}[q^* \in S_a(y)]$ is the probability that we fail to reject that θ = θ* at the true value for θ, say θ'. The ability of such a set to exclude θ* thus depends on the power of the test of H_0: θ =θ* against the alternative H_a: θ = θ'. The accuracy of the confidence interval at any particular θ* is assessed by examining the power of the test of the null θ = θ* against other parameter values.

When there is a uniformly most powerful test, a uniformly most accurate confidence interval can be constructed by inverting this test. But this is not the case here, we therefore consider tests that differ depending on θ*, and on whether the test is against the alternative θ > θ* or θ < θ*.

III. The Tests and their Inversion to Intervals : Fixed Case.

1. Sequence Tests.

Returning to the model above, this means that in choosing the sequence of tests to invert we want tests that have high power at any particular value for ρ against alternatives in each direction. The results of Elliott et al. (1996) show that for the case of ρ=1 and normal errors, no uniformly most powerful test exists. This result will hold for all ρ in the set we are considering, making the choice of tests to invert at each value for ρ somewhat arbitrary.

Consider the following tests. For each ρ we construct the likelihood ratio test statistics

H_0: ρ = ρ* vs H_a : $r = \bar{r} < r^*$ (lower tail test)

and

H_0: ρ = ρ* vs H_a : $r = \bar{r} > r^*$ (upper tail test)

Choosing values for the point alternative $\bar{r}$ for the upper and lower tail tests chooses the particular tests within the family of point optimal tests.

Assuming no serial correlation and $v_t \sim N(0,\sigma^2)$ the log likelihood for the data is of the form

$$L(r, s^2, b) = A - \frac{(y_1 - b'z_1)^2}{2s^2} - \frac{1}{2s^2}\sum_{t=2}^{T}\left[(1 - rL)(y_t - b'z_t)\right]^2 \qquad (4)$$

The likelihood ratio tests invariant to the trend components are of the form

$$LR = -2\left\{L\left[\bar{r}, s^2, b(\bar{r})\right] - L\left[r^*, s^2, b(r^*)\right]\right\} \qquad (5)$$

where $b(r)$ are the GLS estimates of the deterministic trend when we set $r = \bar{r}, r^*$ (these reject for small values as the statistics are the negative of the usual transformed LR test).

Proposition 1.

Under the assumptions of the model in (1) and (2), A1 holding with $v_t \sim N(0,s^2)$ and serially uncorrelated the Neyman Pearson tests of $H_0 : r = r^*$ vs $H_a: r = \bar{r}$ when $r^* = 1 + c^*/T$ and $\bar{r} = 1 + \bar{c}/T$, invariant to d_t and s^2 have power functions of the form

$$\Gamma(c, c^*, \bar{c}) = \Pr\left[P(c, c^*, \bar{c}) < b(c^*, \bar{c})\right] \qquad (6)$$

where $b(c^*, \bar{c})$ *is a critical value,*

$$P(c, c^*, \bar{c}) = (\bar{c}^2 - c^{*2})\int_0^1 W_c(s)^2 ds + x_c^{d}{}' \Lambda^d (c^*, \bar{c}) x_c^d \text{ and}$$

a)Demeaned case(d=m)

$$x_c^m{}' = \left[W_c(1), \int W_c(t)dt\right] \text{ and } \Lambda^m(c^*, \bar{c}) = \begin{bmatrix} c^* - \bar{c} & 0 \\ 0 & 0 \end{bmatrix}.$$

b) Detrended case(d=t)

$$x_c^t{}' = \left[W_c(1), \int W_c(t)dt, \int tW_c(t)dt\right] \text{ and } \Lambda^t(c^*, \bar{c}) = K^t(\bar{c}) - K^t(c^*),$$

$$K^t(c) = \begin{bmatrix} (1-c)(1-l) & 0 & -c^2 l \\ 0 & 0 & 0 \\ -c^2 l & 0 & -3c^2(1-l) \end{bmatrix}$$

where $l = \dfrac{1-c}{1 + \frac{c^2}{3} - c}$.

(all proofs are given in the appendix).

The above results give the power curves for a family of tests for each null ρ* (the family being one for each $\bar{r}$). When ρ*=1 and $\bar{r}$ <1 this is a special case of the results in Elliott et al (1996). The results hold also for other values for ρ*. The critical values b(c*, $\bar{c}$) depend on both the null and the alternative for the test, as well as the extent of the detrending. It can also be seen from the above results that the most powerful test tends to an asymptotic chi-squared distribution when $\bar{c}^2 \to c^{*2}$ (for the locally best test). Further, as the best test against the alternative $\bar{c}$ depends on $\bar{c}$, we have no uniformly most powerful tests at any of the relevant null hypotheses. We can use these results to compute power envelopes of tests for each null hypothesis, and examine the power properties of the tests we employ at different null values.

To make this operational for the more general assumptions on the residuals presented in the previous section, we suggest inverting the following test statistic

$$P_T(c^*,\bar{c}) = \frac{1}{\hat{w}^2}\left[\sum_{t=1}^{T}(\bar{u}_t)^2 - \frac{\bar{r}}{r^*}\sum_{t=1}^{T}(u_t^*)^2\right] \quad (7)$$

where $\bar{u}_t$ and u_t^* are elements of the vectors $\bar{u}$ and u^* with $\bar{u} = \bar{y} - \bar{b}'\bar{z}$ and $u^* = y^* - b^{*\prime} z^*$ where $\bar{b}, b^*$ are the GLS estimates of the trend terms under the alternative and null respectively,

$\bar{z} = [z_1, (1-\bar{r}L)z_2, \ldots, (1-\bar{r}L)z_T]$,

$\bar{y} = [y_1, (1-\bar{r}L)y_2, \ldots, (1-\bar{r}L)y_T]$,

z^*, y^* are defined similarly with ρ* and c* replacing $\bar{r}$ and $\bar{c}$ respectively, and $\hat{w}^2$ is a consistent estimator of w^2 (a specific estimator $\hat{w}^2$ is discussed in Section V).

Proposition 2.

When the data are generated according to (1) and (2) with $E[u_1]=0$ and $Var(u_1)=g(0)$, the statistic $P_T(c^*, \bar{c})$ has the asymptotic distribution $P(c,c^*,\bar{c})$ as defined in Proposition 1.

The results of Proposition 2 show that the statistics have the same distribution as the most powerful tests against a point alternative from the previous section. Thus they achieve the asymptotic Gaussian power envelope for the null hypothesis of c=c* against the alternative c = $\bar{c}$. The tests are one sided (the side depends on whether $\bar{c}$ is greater or less than c*) and reject the null for sufficiently small or negative values. We can define the acceptance regions of the tests with size α_i as

$$A(c^*,\bar{c};a_i) = \{y : P_T(c^*,\bar{c}) > p_{c^*,\bar{c},a_i}\}$$

where $p_{c^*,\bar{c},a_i}$ is the lower 100(α_i)% percentile of the null distribution. We can construct a two sided test for the null hypothesis of c=c* by combining an upper and lower tail test. For simplicity, rather than searching over various alternatives (and hence tests) for each c* we will instead consider testing against alternatives that are a fixed distance from c* for all c*. So for upper tail tests we set $\bar{c} = c^* + \bar{c}_u$ and for lower tail tests $\bar{c} = c^* + \bar{c}_l$ where $\bar{c}_l$ <0. The size of the test will be $\alpha \leq \alpha_l + \alpha_u$ (the sum of the size for the lower and upper tail tests respectively). We can define now the acceptance region at each c* as

$$A(c^*, \bar{c}_u, \bar{c}_l; a) = \{y : P_T(c^*, c^* + \bar{c}_u) > p_{c^*, c^* + \bar{c}_u, a_u}\} \cap \{y : P_T(c^*, c^* + \bar{c}_l) > p_{c^*, c^* + \bar{c}_l, a_l}\}$$

We then define the sequence test confidence interval with coverage 100(1-α)% $S_a^s(y, \bar{c}_u, \bar{c}_l)$ as the set of c* that are in the acceptance regions $A(c^*, \bar{c}_u, \bar{c}_l; a)$. This definition admits the possibility of disjoint sets, however for practical reasons we rule out this case in numerical results reported below.

This confidence set has the parameters $\bar{c}_u, \bar{c}_l, a_l, a_u$ and the accuracy of the confidence set depends on these parameters. Our choice for $\bar{c}_u$ and $\bar{c}_l$ is guided by the results of Elliott et. al. (1996). For c*=0 they found $\bar{c}_l = -7$ in the case of a mean and $\bar{c}_l = -13.5$ in the case of a constant and time trend provide good power for all lower tail alternatives. By numerical experiments we select $\bar{c}_u = 2$ for the demeaned case and $\bar{c}_u = 5$ for the detrended case, α_l=3% and α_u= 2% (so $\alpha \leq$5%)[1].

2. Inverting a Single Test

A computationally simpler alternative to the method described above is to simply invert a test for a unit root, as undertaken by Stock (1991) for the augmented Dickey and Fuller (1979) statistic and the (modified) Sargan and Bhargava (1983) statistic. In this method all is the same as the above except that a single statistic is used for all of the values for c.

For any test for a unit root, T(y), we can determine the asymptotic distribution as a function of c. We can then use the percentiles of these distributions to construct acceptance regions for any particular size test and null hypothesis (c*) in our set. These acceptance regions have the form

---

[1] The power curve is much steeper for alternatives that are greater than the null than alternatives less than the null which suggests choosing $\alpha_u < \alpha_l$.

$$A(c^*;a) = \{y : q_{c^*,a_1} < T(y) < q_{c^*,1-a_u}\}$$

where q_{c^*,a_l} and $q_{c^*,1-a_u}$ are the 100(α_l) and 100(1-α_u) percentiles of the limit distribution for T(y) when c* is true. The confidence set then collects, for any data set y, all the values for c* where y is in the acceptance region given above, i.e. we then define the test inversion confidence interval with coverage 100(1-α)% $S(y,a)$ as the set of c* that are in the acceptance regions $A(c^* : a)$ defined above.

We construct such tests for various unit root tests which have been found to have good power properties. These include the P_T (0, $\bar{c}$) and the Sargan and Bhargava (1983) statistic as modified for general residuals in Stock (1991), which we denote as MSB.

The first of these statistics is simply the sequence test where c*=0 and $\bar{c}$ equal to the values in Elliott et al (1996). Thus the test statistic to invert is

$$P_T(0,\bar{c}) = \frac{1}{\hat{w}^2}\left[\sum_{t=1}^{T}(\bar{u}_t)^2 - \bar{r}\sum_{t=1}^{T}(u_t^*)^2\right]$$

The limit distribution of this statistic for any c is

$$P_T(0,\bar{c}) \Rightarrow \bar{c}^2 \int W_c(t)^2 dt + x_c^d{}' \Lambda^d(0,\bar{c}) x_c^d \equiv P(c,0,\bar{c})\,.$$

The MSB statistic is constructed as MSB = $\frac{1}{\hat{w}^2 T^2}\sum_{t=1}^{T}\left(y_t^{d,SB}\right)^2$ where

$y_t^{m,SB} = y_t - \frac{1}{T}\sum y_t$ and

$y_t^{t,SB} = y_t - \frac{1}{T}\sum y_t + \left[(T+1)/2(T-1)\right](y_T - y_1) - t(y_T - y_1)/(T-1)$.

The MSB statistic also has an asymptotic representation of a form similar to the above tests, namely we can write the asymptotic distribution as

$$MSB^d \Rightarrow \int W_c(t)^2 dt + x_c^d{}' \Lambda_{SB}^d(c) x_c^d$$

where

$$\Lambda_{SB}^m = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

and

$$\Lambda_{SB}^{t} = \begin{bmatrix} \frac{1}{12} & \frac{1}{2} & -1 \\ \frac{1}{2} & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

IV. The Tests and their Inversion to Intervals : Unconditional Case.

The results of the previous section show dependence on the assumption on the initial condition in two ways. Firstly, the optimal tests to invert (sequence tests) depend on this assumption through the specification of the likelihood (this is shown for the case where c*=0 in Elliott (1999)). Second, the limiting distribution in equation (3) and hence critical values for each of the tests when $\rho<1$ differ depending on the assumption made, hence for all of the confidence intervals different intervals are appropriate depending on the assumption made. In this subsection, we derive limit results for the sequence confidence interval that would be appropriate under the assumption that the initial condition comes from its stationary distribution when $\rho<1$. We also consider the inversion of the 'near' optimal test for a unit root in this case and again the Sargan and Bhargava statistic under this different distribution.

1. Sequence Tests.

In this model the log likelihood for the data when there is no serial correlation and $v_t \sim N(0,\sigma^2)$ is of the form

$$L(r,s^2,b) = \begin{cases} A - \dfrac{(1-r^2)(y_1 - b'z_1)^2}{2s^2} - \dfrac{1}{2s^2}\sum_{t=2}^{T}\left[(1-rL)(y_t - b'z_t)\right]^2 & r<1 \\ A - \dfrac{(y_1 - b'z_1)^2}{2s^2} - \dfrac{1}{2s^2}\sum_{t=2}^{T}\left[(1-rL)(y_t - b'z_t)\right]^2 & r \geq 1 \end{cases} \quad (8)$$

where the likelihood changes discretely at $\rho=1$.

The likelihood ratio tests invariant to the trend components are again of the form

$$LR = -2\left\{L\left[\bar{r}, s^2, b(\bar{r})\right] - L\left[r^*, s^2, b(r^*)\right]\right\} \quad (9)$$

where $b(r)$ are the GLS estimates of the deterministic trend given ρ for this alternate model.

Proposition 3.

Under the assumptions of the model in (1) and (2), $v_t \sim N(0,s^2)$ and is not serially correlated, u_0 drawn from its unconditional distribution when $r<1$ the Neyman Pearson tests of $H_0: r = r^*$ vs $H_a: r = \overline{r}$ when $r^*=1+c^*/T$ and $\overline{r}=1+\overline{c}/T$, invariant to d_t and s^2 have asymptotic power functions of the form

$$\Gamma_s(c,c^*,\overline{c}) = \Pr\left[Q(c,c^*,\overline{c}) < b(c^*,\overline{c})\right] \qquad (10)$$

where $Q(c,c^*,\overline{c})$ depends on the signs of c, c and $\overline{c}$ and is defined by*

$$Q(c,c^*,\overline{c}) = (\overline{c}^2 - c^{*2})\int_0^1 W_c(s)^2\,ds + x_{s,c}^{d}{}'\Lambda_s^d(c^*,\overline{c})x_{s,c}^d \text{ where}$$

a)Demeaned case

For $r<1$ $x_{s,c}^{m}{}' = \left[W_c(0), W_c(1), \int W_c(t)dt\right]$, $\Lambda_s^m(c^*,\overline{c}) = K_s^m(\overline{c}) - K_s^m(c^*)$ and

$$K_s^m(c) = \frac{c}{c-2}\begin{bmatrix}(1-c) & -1 & c\\ -1 & (1-c) & c\\ c & c & -c^2\end{bmatrix} \text{ for } c<0 \text{ and}$$

$$K_s^m(c) = \begin{bmatrix}(c^2-c) & c & -c^2\\ c & -c & 0\\ -c^2 & 0 & 0\end{bmatrix} \text{ for } c \geq 0.$$

For $r \geq 1$ $x_{s,c}^m = x_c^m$ *and* $\Lambda_s^m(c^*,\overline{c}) = \underline{K}_s^m(\overline{c}) - \underline{K}_s^m(c^*)$ *where*

$\underline{K}_s^m(c)$ *is* $K_s^m(c)$ *as above with the first row and column deleted.*

b) Detrended case

for $r<1$ $x_{s,c}^{t}{}' = \left[W_c(0), W_c(1), \int W_c(t)dt, \int tW_c(t)dt\right]$ and $\Lambda_s^t(c^*,\overline{c}) = K_s^t(\overline{c}) - K_s^t(c^*)$ where

$$K_s^t(c)=\begin{bmatrix} (1-f)(1-c)-c^2fk & f\left(c^2(\tfrac{1}{2}-k)-c\right)-(1-f) & fc^2(ck-\tfrac{1}{2}) & fc^2(1-\tfrac{c}{2}) \\ f\left(c^2(\tfrac{1}{2}-k)-c\right)-(1-f) & (1-f)(1-c)-c^2fk & fc^2\left((\tfrac{1}{2}+ck)-\tfrac{c}{2}\right) & -fc^2(1-\tfrac{c}{2}) \\ -\quad fc^2(ck-\tfrac{1}{2}) & fc^2\left((\tfrac{1}{2}+ck)-\tfrac{c}{2}\right) & -fc^4k & \frac{fc^4}{2} \\ fc^2(1-\tfrac{c}{2}) & -fc^2(1-\tfrac{c}{2}) & \frac{fc^4}{2} & -fc^4 \end{bmatrix}$$

for $c<0$ *where* $f=1/(1+\frac{c^2}{12}-\frac{c}{2})$, $k=(1+\frac{c^2}{3}-c)/(c^2-2c)$

$$K_s^t(c)=\begin{bmatrix} \frac{c^2}{4}+\frac{1}{4}(1-c)(1-l) & \frac{1}{2}(1-c)(1-l) & -c^2 & \frac{3c^2}{2}-\frac{3}{2}(1-c)(1-l) \\ \frac{1}{2}(1-c)(1-l) & (1-c)(1-l) & 0 & -3(1-c)(1-l) \\ -c^2 & 0 & 0 & 0 \\ \frac{3c^2}{2}-\frac{3}{2}(1-c)(1-l) & -3(1-c)(1-l) & 0 & 9(1-c)(1-l)-3c^2 \end{bmatrix} \text{for } c\geq 0.$$

For $r\geq 1$ $x_{s,c}^t=x_c^t$ *and* $\Lambda_s^t(c*,\bar{c})=\underline{K}_s^t(\bar{c})-\underline{K}_s^t(c*)$ *where* $\underline{K}_s^t(c)$ *is* $K_s^t(c)$ *as above with the first row and column deleted.*

It is clear that the sequence tests are different from those in the fixed case, so the confidence intervals constructed from them will also differ. Further, as the best test against the alternative $\bar{c}$ depends on $\bar{c}$, we again have no uniformly most powerful tests at any of the relevant null hypotheses.

To make this operational for the more general assumptions on the residuals we suggest the following test statistic for each side of the hypothesis

$$Q_T(c*,\bar{c})=\frac{1}{\hat{w}^2}\left[\sum_{t=1}^{T}(\bar{u}_t)^2-\frac{\bar{r}}{r*}\sum_{t=1}^{T}(u_t*)^2\right] \tag{11}$$

where $\bar{u}_t$ and u_t* are elements of the vectors $\bar{u}$ and $u*$ with $\bar{u}=\bar{y}-\bar{b}'\bar{z}$ and $u*=y*-b*'z*$ where $\bar{b}, b*$ are the GLS estimates of the trend terms under the alternative and null respectively,

$$\bar{z}=\begin{cases}\left[(1-\bar{r}^2)^{1/2}z_1,(1-\bar{r}L)z_2,\ldots,(1-\bar{r}L)z_T\right] & \bar{c}<0 \\ \left[z_1,(1-\bar{r}L)z_2,\ldots,(1-\bar{r}L)z_T\right] & \bar{c}\geq 0\end{cases},$$

$$\bar{y} = \begin{cases} [(1-\bar{r}^2)^{1/2} y_1, (1-\bar{r}L) y_2, \ldots, (1-\bar{r}L) y_T] & \bar{c} < 0 \\ [y_1, (1-\bar{r}L) y_2, \ldots, (1-\bar{r}L) y_T] & \bar{c} \geq 0 \end{cases}$$

and z^*, y^* are defined similarly with ρ^* and c* replacing $\bar{r}$ and $\bar{c}$ respectively.

Proposition 4.

When the data are generated according to (1) and (2) with the u_1 drawn from its unconditional distribution when r <1 the statistic $Q_T(c^*, \bar{c})$ has the asymptotic distribution $Q(c, c^*, \bar{c})$ as defined in Proposition 3.

We can construct sequence confidence intervals using the same method as the previous section. For the results in this paper we have set $\bar{c}_l = -10$ as this was found to provide good power for all alternatives for the null c*=0 and $\bar{c}_u = 2$ for the demeaned case and $\bar{c}_u = 5$ for the detrended case. We again choose α_l=3% and α_u= 2% .

2. Inverting a Single Test

For this alternative assumption on the initial condition we construct such tests for various unit root tests. These include the $Q_T(0, \bar{c})$ (constructed in the same way as the $P_T(0, \bar{c})$) and the Sargan and Bhargava statistic.

The first of these statistics is simply the sequence test where c*=0 and $\bar{c}$ = -10. Thus the test statistic to invert is

$$Q_T(0, \bar{c}) = \frac{1}{\hat{w}^2}\left[\sum_{t=1}^{T} (\bar{u}_t)^2 - \bar{r}\sum_{t=1}^{T} (u_t^*)^2\right]$$

The limit distributions for this statistic as a function of c is thus a special case of the above results and are given in Elliott (1999).

The MSB statistic is constructed as in the previous section, but under the assumptions of this section its asymptotic representation is

$$MSB^d \Rightarrow \int W_c(t)^2 dt + x_{s,c}^{d}{}' \Lambda_{s,SB}^d(c) x_{s,c}^d$$

where

$$\Lambda^{m}_{s,SB} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

and

$$\Lambda^{t}_{s,SB} = \begin{bmatrix} \frac{1}{12} & -\frac{1}{12} & -\frac{1}{2} & 1 \\ -\frac{1}{12} & \frac{1}{12} & \frac{1}{2} & -1 \\ -\frac{1}{2} & \frac{1}{2} & -1 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}$$

V. Computing critical values.

The construction of confidence intervals will require the computation of asymptotic percentage points of the above distributions. This can be accomplished either by the use of Monte Carlo simulation or through the computation of the characteristic functions of the above statistics and then by use of the methods of Imhof (1961) to compute percentage points from these characteristic functions. The Monte Carlo approach, while conceptually straightforward, is computationally very intensive in this application because of the large number of tests, indexed by the null hypothesis: moreover it is prone to numerical imprecision. We therefore adopt the latter approach here. This requires deriving the limiting characteristic functions for the various test statistics.

All of the above statistics have limit distributions with the similar form

$$T(y) \Rightarrow a\int W_c(t)^2\,dt + x'\Lambda x$$

where x is a kx1 vector of functionals of Brownian motion which have a joint normal distribution and {a,Λ} are constants depending on the particular test.

Proposition 5.. The characteristic functions for the limit distribution of T(y) are

$$f(t) = \exp\left\{\frac{-c-b}{2}\right\}\left|I_k - 2\Sigma\Lambda^*\right|^{-1/2}$$

where $b = \sqrt{c^2 - 2ita}$ *and*

$$\Lambda^* = it\Lambda + \begin{bmatrix} -\frac{(c+b)}{2} & 0 & 0 & 0 \\ 0 & \frac{(c+b)}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

for the case when $W_c(0)$ is non zero and d=t. For d=m we have k=3 and ignore the last row and column and for $W_c(0)=0$ we ignore the first row and column. The elements of S can be computed using stochastic calculus and are given in Appendix 2.

The proof of the Theorem follows Tanaka (1996) directly and is omitted. The characteristic functions make it fairly straightforward to compute p-values and percentages of the distributions. Again following Tanaka (1996), percentiles of the distributions are given by

$$F(x) = \frac{1}{2} - \frac{1}{p}\int_0^\infty \frac{2}{t}\,\mathrm{Im}\left[e^{-it^2x} f(t^2)\right] dt$$

where Im[.] takes the imaginary part of the argument. This is the Imhof (1961) formula (see Tanaka (1996), equation (6.13)). A change of variables is used to ensure that the function integrated is zero at t=0. Simpson's rule is used here to evaluate the integral numerically.

VI. Evaluating the methods

This section considers large and small sample properties of the methods for constructing confidence intervals on ρ discussed above under each of the assumptions on the initial condition. As large sample properties of the confidence sets map directly to asymptotic power properties of the tests, these are just various views of the asymptotic power of the tests presented above for different null and alternative values for ρ. The small sample results given are from directly computing the confidence intervals as discussed above.

1. Large Sample Results.

As there is a direct relationship between power and exclusion of false values of ρ, we examine the power of the tests at various null hypotheses. Figures 1a to 1f show asymptotic power of two sided tests in the conditional case (where the lower tail test has size 0.03 and the upper 0.02) for the sequence test, the $P_T(0, \bar{c})$ test, and the modified Sargan and Bhargava test in the d = μ and τ under the assumptions of section 3. The power curves are given for three values for c* ; c*=0, -5 and -10.

A number of features are noteworthy. First, in each case we see that the sequence test has slightly better power than the other two tests. Second, the $P_T(0, \bar{c})$ test is very similar to the sequence test for c* near zero; this is partially by construction as the tests are the same at c*=0 in the lower tail. Third, the MSB test has the lowest power curve in each case and when the model is close to a unit root this power loss is more severe. Fourth, even though the optimality properties for the $P_T(0, \bar{c})$ statistic and the SB statistic are for the null of ρ =1 we see that the tests based on these statistics at other nulls have reasonable power. There is not so great a reason to consider using a sequence of statistics rather than inverting $P_T(0, \bar{c})$, at least asymptotically. All of the results are true for both the demeaned and detrended models, although the differences are greater in the demeaned case.

These results suggest that each of these statistics are likely to invert to similarly short intervals, with $P_T(0, \bar{c})$ likely to be preferred over MSB. Note that we do not consider the augmented Dickey and Fuller (1979) statistic as it has power significantly below that of the $P_T(0,-10)$ statistic (Elliott et al (1996)) at c*=0 in the lower tail.

Similar figures are given for the unconditional case in Figures 2a through 2f. Here the $Q_T(0,-10)$ test is examined along with the Sargan and Bhargava test and the sequence test that is appropriate for this different assumption. Essentially the results are similar to the fixed initial condition case except that the differences are not as large in the demeaned case.

2. Small Sample Results

The coverage rates for the confidence intervals can be examined in Monte Carlo experiments. There is a large body of Monte Carlo results on the size performance of tests for a unit root in a wide variety of models for the error terms. The results are not greatly encouraging - size performance is usually poor when the extent of serial correlation in ν_t is unknown, and is also poor for models in which ν_t follows a moving average process with a large root (which to some extent cancels the unit root in the model). This size performance carries across to coverage performance when the true value for ρ is 1.

Here we examine in a number of models the size performance at various true values for ρ. Each of the statistics requires estimation of the nuisance parameter $\hat{w}^2$. We use here autoregressive estimates of this parameter from estimating the regression

$$\Delta y_t^d = a_0 y_{t-1}^d + \sum_{i=1}^{m} a_i \Delta y_{t-i}^d + e_t \qquad (12)$$

where the data is first detrended according to the model (i.e. if a constant or a constant and time trend is in the model). Two methods of detrending are employed; OLS detrending (indicated by having (ols) after the name of the statistic) and GLS detrending under the local alternative[2] (indicated by having (loc) after the name of the statistic). The estimate for $\hat{w}^2$ is then given by $\hat{w}^2 = \frac{\hat{s}^2}{[1-\hat{a}(1)]^2}$ where $\hat{s}^2 = T^{-1}\sum \hat{e}_t^2$, $\hat{a}(1) = \sum_{i=1}^{m} \hat{a}_i$ and $\hat{e}_t$ and $\hat{a}_i$ are least squares estimates from the above regression. To make this operational we need to select the generally unknown lag length m. Here, m is selected by the MAIC criterion (recommended in Ng and Perron (1998)) with a maximum of four lags.

Results when there is no serial correlation and u_0=0 are compiled in Table 1, where we examine the performance of the constructed confidence intervals when T=100 for three true values for ρ, these are ρ = 1, 0.95 and 0.9. For each of these models we examine one minus the coverage rate for a variety of hypothesized roots ρ*. In the case of ρ=1, this means that the column marked ρ*=1 gives the probability of false exclusion and the columns marked ρ* =0.95 and 0.9 give the probability of correctly excluding these values from the confidence interval.

A number of features are noteworthy. First, coverage rates are somewhat different than the nominal coverage rates. The undercoverage arises from the upper tail rejections. Second, the ordering of the ability of the confidence intervals to rule out false values is as implied by the asymptotic power results. The Sequence test does indeed do the best job in most cases in this regard, followed closely by the inverted $P_T(0,\bar{c})$ statistic, with the MSB statistic worst. Given the extra computation effort required by the sequence tests, then it may be considered just as useful to use the $P_T(0,\bar{c})$ statistic for inversion for a confidence interval.

Table 2 examines the probability of false exclusions when there are various forms of serial correlation in the model. Both MA(1) and AR(1) models for the errors are considered. These results are similar to those in the unit root testing literature (see Stock (1994) for a detailed

2 This follows Ng and Perron (1998).

examination of this case). Size distortions (undercoverage) arise in most cases but especially when there is a negative MA coefficient. These distortions are more severe when the true root is less than one, regardless of the model.

Table 3 reports results for models with no serial correlation in the unconditional case, examining the sequence test for this case, the $Q_T(0,-10)$ statistic and the MSB statistic (using the correct limit result for this model). The results are essentially similar to those above, with less of a difference between the sequence method and simply inverting $Q_T(0,-10)$. Also, there is a smaller difference between the performance of the intervals constructed from inverting the MSB statistic and the other confidence intervals. Results are not reported for the serial correlation case as they are similar to those in the conditional case.

In addition to coverage and exclusion, we are interested in interval length. Figures 3a to 3d show histograms of the interval lengths for the sequence test, $P_T(0,-\bar{c})$ test, MSB, and ADF (From Stock (1991)) test from 1000 random samples where the model is as per Table 1 with ρ_0=0.95, however the lag length in estimating nuisance parameters is known to be zero (in each case a constant only is estimated). The ADF test, whilst having some short intervals, has more often than the other tests many long intervals. The MSB test has more longer intervals than the $P_T(0,-\bar{c})$ test, which is fairly similar to the Sequence test. For other models (values for c, unconditional initial condition) similar results apply so the figures are not included. In practice the differences between the interval estimators in the unconditional case are much smaller.

VI. Conclusion

Most accurate confidence intervals are constructed from the inversion of a sequence of most powerful tests. In the case of large roots in autoregressive models, for any root (modeled as local to unity) no UMP test exists against even one sided alternatives, ruling out the possibility of inverting a sequence of UMP tests. We instead choose to invert a sequence of point optimal tests, which although only most powerful at a single point have quite reasonable properties at other alternatives. By computing the characteristic functions of this family of tests we are able to provide methods for making computation of the confidence intervals feasible without relying on a large number of Monte Carlo estimates of critical values.

We find that the confidence intervals constructed from a sequence of point optimal tests have quite similar power properties to inverting near optimal tests for a unit root. Given that the latter confidence intervals are simpler to compute in practice, these are the suggested method.

We also find that this asymptotic property holds up reasonably well for smaller samples, so inversion of the $P_T(0,\bar{c})$ and $Q_T(0,\bar{c})$ statistics work well in terms of constructing short confidence intervals for the largest autoregressive root. The main small sample problem that arises is undercoverage when there is serial correlation of an unknown form, this is directly analogous to the problem of overrejection in the unit root testing literature. Confidence intervals constructed from inverting these point optimal tests were found to have better properties than that for the ADF or Sargan Bhargava type tests.

Because no asymptotically most accurate confidence interval exists in this problem, the work is not definitive. For example, the parameters $\bar{c}_u, \bar{c}_l, a_l, a_u$ were fixed here after preliminary investigation for computational reasons; they could however vary with c*. This may yield more accurate intervals. A quite different approach not considered here is to focus on Euclidean length of confidence intervals. In these and other dimensions, interesting work remains.

References:

Andrews, D.W.K, 1993, Exactly Median Unbiased Estimation of First Order Autoregressive/Unit Root Models, Econometrica, 61, pp 139-166.

Andrews, D.W.K and Y-C Chen, 1992, Approximately Median Unbiased Estimation of Autoregressive Models with Applications to U.S. Macroeconomic and Financial Time Series, Cowles Foundation Discussion paper #1026, Yale University.

Bobkoski, M.J., 1983, Hypothesis Testing in Nonstationary Time Series, unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin.

Cavanagh, C., 1985, Roots Local to Unity, manuscript, Department of Economics, Harvard University.

Cavanagh, C.L., G. Elliott and J.H.Stock, 1995, Inference in Models with Nearly Nonstationary Regressors, Econometric Theory, v11, 1131-47.

Chan, N.H. and C.Z.Wei, 1987, Asymptotic Inference for Nearly Nonstationary AR(1) Processes, Annals of Statistics, 15, pp1050-63.

Dickey, D.A. and W.A.Fuller, 1979, Distribution of Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74, pp427-431.

Elliott, G., 1999, Efficient Tests for a Unit Root when the Initial Observation is Drawn From its Unconditional Distribution, International Economic Review, 40, pp767-783..

Elliott, G., T.J.Rothenberg and J.H.Stock, 1992, Efficient Tests for an Autoregressive Unit Root, NBER Technical Working Paper #130.

Elliott, G., T.J.Rothenberg and J.H.Stock, 1996, Efficient Tests for an Autoregressive Unit Root, Econometrica, 64, pp813-36.

Hansen, B.E., 1999, Bootstrapping the Autoregressive Model", Review of Economics and Statistics, forthcoming.

Imhof, J.P., 1961, Computing the distribution of quadratic forms in normal variables, Biometrika, 48, 419-26.

Lehmann, E.L., 1994, Testing Statistical Hypotheses, Second edition, (Chapman and Hall:New York).

Ng, S. and P. Perron, 1998, Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, mimeo, Boston College, February.

Phillips, P.C.B., 1987, Towards a Unified Asymptotic Theory for Autoregression, Biometrika, 74, 535-47.

Sargan, J.D and A. Bhargava, 1983, Testing residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk, Econometrica, 51, 153-74.

Sargent, T.J., 1998, The Conquest of American Inflation, (Harvard University Press, Cambridge, MA).

Stock, J.H., 1991, Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series, Journal of Monetary Economics, 28, 435-459.

Stock, J.H., 1994, Unit Roots, Structural Breaks, and Trends, in D.McFadden and R.F.Engle eds. Handbook of Econometrics V4, (North Holland: Amsterdam), pp2739-841.

Tanaka, K., 1996, Time Series Analysis: Nonstationary and Noninvertible Distribution Theory, (Wiley : New York).

Proofs of Propositions.

Lemma 1 (Limit results for data detrended under local to unit root detrending)

When the data is generated according to (1) and (2) with ρ=1+c/T and detrending is under $\tilde{r} = 1 + \frac{\tilde{c}}{T}$ then

a) Demeaned Case.

When ρ<1,

1. For the unconditional model with $\tilde{r}$ <1

$$T^{-1/2}\left[b(\tilde{r}) - b\right] \Rightarrow w\begin{bmatrix} \frac{-1}{\tilde{c}-2} & \frac{-1}{\tilde{c}-2} & \frac{\tilde{c}}{\tilde{c}-2} \end{bmatrix} x^m_{s,c} \text{ and}$$

$$T^{-1/2}\left[y_t - b(\tilde{r})' z_t\right] \Rightarrow w\left[W_c(s) + \begin{bmatrix} \frac{1}{\tilde{c}-2} & \frac{1}{\tilde{c}-2} & \frac{-\tilde{c}}{\tilde{c}-2} \end{bmatrix} x^m_{s,c}\right].$$

2. For $\tilde{r}$ ≥1

$$T^{1/2}\left[b(\tilde{r}) - b - u_1\right] \Rightarrow w\left[(\tilde{c} - \tilde{c}^2) \quad -\tilde{c} \quad \tilde{c}^2\right] x^m_{s,c} \text{ and}$$

$$T^{-1/2}\left[y_t - b(\tilde{r})' z_t\right] \Rightarrow w\left[W_c(s) - W_c(0)\right].$$

b) Detrended case.

1. For $\tilde{r}$ <1

$$\Psi^{-1}\left[b(\tilde{r}) - b\right] \Rightarrow wf\begin{bmatrix} -ck + \frac{1}{2} & -ck - \frac{1-c}{2} & c^2 k & -\frac{c^2}{2} \\ \frac{c}{2} - 1 & 1 - \frac{c}{2} & -\frac{c^2}{2} & c^2 \end{bmatrix} x^t_{s,c} \text{ and}$$

$$T^{-1/2}\left[y_t - b(\tilde{r})' z_t\right] \Rightarrow w\left[W_c(s) - f(1 \quad s)\begin{bmatrix} -ck + \frac{1}{2} & -ck - \frac{1-c}{2} & c^2 k & -\frac{c^2}{2} \\ \frac{c}{2} - 1 & 1 - \frac{c}{2} & -\frac{c^2}{2} & c^2 \end{bmatrix} x^t_{s,c}\right]$$

where $\Psi = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{-1/2} \end{bmatrix}$.

2. For $\tilde{r} \geq 1$

$$\Psi^{-1}[b(\tilde{r})-b] \Rightarrow w\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{l-3}{2} & l & 0 & 3(1-l) \end{bmatrix} x^t_{s,c}$$

$$T^{-1/2}[y_t - b(\tilde{r})' z_t] \Rightarrow w\left\{ W_c(s) - (1 \quad s)\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{l-3}{2} & l & 0 & 3(1-l) \end{bmatrix} x^t_{s,c} \right\}$$

For the fixed initial condition case the results are as above in case 2 for all $\tilde{r}$ with 0 in place of $W_c(0)$ in $x^d_{s,c}$. For the unconditional initial condition case when ρ≥1, the results are as above with 0 in place of $W_c(0)$.

Proof:

For both the demeaned and detrended results where $\tilde{r}<1$, the results are derived in Elliott (1999) Lemma 3. The GLS estimators for the case where $\tilde{r} \geq 1$ are those of Elliott et. al. (1996), where results for $W_c(0)=0$ are shown. When $u_1 \neq 0$, the results follow directly from the expressions (A.2) and (A.3) of Elliott et. al. (1992) and application of the functional central limit theorem and continuous mapping theorem.

Proof of Proposition 1.

The LR test in (5) can be rewritten as the difference between two LR tests with the null of a unit root

$$LR = -2\left\{ L\left[\bar{r}, s^2, b(\bar{r})\right] - L\left[1, s^2, b(1)\right]\right\} + 2\left\{ L\left[r^*, s^2, b(r^*)\right] - L\left[1, s^2, b(1)\right]\right\}.$$

Each of these tests for a unit root are of the form

$$-2\left\{ L\left[\tilde{r}, s^2, b(\tilde{r})\right] - L\left[1, s^2, b(1)\right]\right\}.$$

As noted after equation (6) of Elliott et. al. (1996) these are equivalent to test statistics of the form

$$\frac{T(\tilde{s}^2 - \hat{s}^2)}{s^2}$$

where $T\tilde{s}^2 = \sum_{t=1}^{T}\tilde{u}_t^2$ where $\tilde{u}_t$ is defined after equation (7) with $\tilde{r}$ in place of $\bar{r}$ in the text and $T\hat{s}^2 = \sum_{t=1}^{T}\hat{u}_t^2$ is similarly defined using 1 in place of $\tilde{r}$.

The statistics $\frac{T(\tilde{s}^2 - \hat{s}^2)}{s^2}$ were examined in Elliott et al (1996). With some rearrangement we have the results

(a) Demeaned Case

$$T(\tilde{s}^2 - \hat{s}^2) = (1-\tilde{r})^2\sum_{t=2}^{T}(y_{t-1}^d)^2 + (1-\tilde{r})\left[(y_T^d)^2\right] + B(\tilde{r}) + o_p(1)$$

(b) Detrended Case

$$T(\tilde{s}^2 - \hat{s}^2) = (1-\tilde{r})^2\sum_{t=2}^{T}(y_{t-1}^d)^2 + \left[T^{-1} + (1-\tilde{r})\right]\left[(y_T^d)^2\right] + B(\tilde{r}) + o_p(1)$$

where $y_t^d = y_t - b(\tilde{r})'z_t$ and (from lemma 1) have limit results that depend on the value for $\tilde{r}$. From the results of Lemma 1 and the continuous mapping theorem we have the limit results (ignoring $b(\tilde{r})$ which is a constant depending on $\tilde{r}$ and can be subsumed into the critical value)

$$\frac{T(\tilde{s}^2 - \hat{s}^2)}{w^2} \Rightarrow \tilde{c}^2\int_0^1 W_c(t)^2 dt + x_c^d{}'K^d(\tilde{c})x_c^d$$

where $K(\tilde{c})$ depends on the (demeaned or detrended) case considered, $K^m(c) = \begin{bmatrix} -c & 0 \\ 0 & 0 \end{bmatrix}$ and K^τ is given in Proposition 1 . The results are shown for the statistic with ν_t satisfying the more general assumptions after equation (2) rather than that of the special case in the proposition. In the special case of the proposition $w^2 = s^2$. The limit results for the general LR tests thus derive from noting that the limit results are the difference between the limit results replacing $\tilde{c}$ with $\bar{c}$ and c* respectively.

Proof of Proposition 2.

The method of proof follows the intuition of the first proposition - rewriting the statistic into those previously analyzed in Elliott et. al. (1996).

$$
\begin{aligned}
P_T(c^*,\bar{c}) &= \frac{1}{\hat{w}^2}\left[\sum_{t=1}^{T}(\bar{u}_t)^2 - \frac{\bar{r}}{r^*}\sum_{t=1}^{T}(u_t^*)^2\right] \\
&= \frac{1}{\hat{w}^2}\left[\left\{\sum_{t=1}^{T}(\bar{u}_t)^2 - \bar{r}\sum_{t=1}^{T}(\hat{u}_t)^2\right\} - \frac{\bar{r}}{r^*}\left\{\sum_{t=1}^{T}(u_t^*)^2 - r^*\sum_{t=1}^{T}(\hat{u}_t)^2\right\}\right] \\
&= P_T(0,\bar{c}) - \frac{\bar{r}}{r^*}P_T(0,c^*)
\end{aligned}
$$

The statistics $P_T(0,\bar{c})$ were analyzed in Elliott et. al. (1996) . The limit distribution for $P_T(0,\bar{c})$ differs from that of $\frac{T(s^2-\hat{s}^2)}{w^2}$ by the constant B($\bar{r}$) which can be subsumed into the critical value. As the weight for the second part of the statistic $\frac{\bar{r}}{r^*}$ converges to one asymptotically the statistics have the same rejection regions as the likelihood ratio statistics in Proposition 1.

Proposition 3.

We can exploit the same rewriting of the LR statistic here as in the proof of proposition 1, and then employ the results of Elliott (1999) along with lemma 1.

Proposition 4.

The proof of this follows directly that of Proposition 2 with $Q_T(c^*,\bar{c})$ replacing $P_T(c^*,\bar{c})$ and employing the results of Elliott (1999).

Appendix 2 :

The elements of Σ for the most general (k=4, where x= x_s^t) case are

$$\Sigma_{11} = -\frac{1}{2c},\ \Sigma_{12} = -\frac{e^{-b}}{2c},\ \Sigma_{13} = -\frac{(1-e^{-b})}{2cb},\ \ \Sigma_{22} = -\frac{e^{-2b}}{2c} + \frac{(1-e^{-2b})}{2b},$$

$$\Sigma_{23} = \frac{1}{b^2}\left[\frac{1}{2} + \left(-\frac{b}{2c} - 1\right)e^{-b} + \left(\frac{1}{2} + \frac{b}{2c}\right)e^{-2b}\right],$$

$$\Sigma_{33} = \frac{1}{b^2}\left[1 - \frac{1}{2c} - \frac{3}{2b} + \left(\frac{2}{b} + \frac{1}{c}\right)e^{-b} + \left(-\frac{1}{2c} - \frac{1}{2b}\right)e^{-2b}\right],$$

$$\Sigma_{14} = \frac{1}{2cb^2}\left[(b+1)e^{-b} - 1\right],\ \Sigma_{24} = \frac{1}{2cb^3}\left[c(b-1) - be^{-b} + (c+b)(b+1)e^{-2b}\right],$$

$$\Sigma_{34} = \frac{1}{2b^3}\left[b - 1 + e^{-b}\right] + \frac{1}{2}\left(-\frac{1}{cb^3} - \frac{1}{b^4}\right)\left[(1-e^{-b})(1-(b+1)e^{-b}\right],$$

$$\Sigma_{44} = \frac{1}{b^4}\left(-\frac{1}{2c} - \frac{1}{2b}\right)\left[(b+1)e^{-b} - 1\right]^2 + \frac{1}{b^5}\left[\frac{b^3}{3} + 1 - \frac{b^2}{2} - (b+1)e^{-b}\right]$$

This is the relevent variance covariance matrix for the detrended case where the initial condition is drawn from its stationary (unconditional) distribution and $\rho < 1$. In this case where d=μ the relevent variance covariance matrix omits the fourth row and column.

For the other cases (fixed initial condition or unconditional case where $\rho \geq 1$) we have Σ being in the most general case a 3x3 matrix as $W_c(0)=0$. The elements of this symmetric matrix are

$$\Sigma_{22} = \frac{(1-e^{-2b})}{2b},\ \Sigma_{23} = \frac{1}{b^2}\left[\frac{1}{2} - e^{-b} + \frac{1}{2}e^{-2b}\right],\ \Sigma_{24} = \frac{1}{2b^3}\left[(b-1) + (b+1)e^{-2b}\right],$$

$$\Sigma_{33} = \frac{1}{b^3}\left[b - \frac{3}{2} + 2e^{-b} - \frac{1}{2}e^{-2b}\right],$$

$$\Sigma_{34} = \frac{1}{2b^4}\left[b^2 - b - 1 + 2(b+1)e^{-b} - (b+1)e^{-2b}\right],$$

$$\Sigma_{44} = \frac{1}{2b^5}\left[1 + \frac{2}{3}b^3 - b^2 - (b+1)^2 e^{-2b}\right].$$

For the demeaned case we have the same as above with the last row and column deleted.

TABLE 1: Small Sample results. Zero initial condition.

	$r = 1$			$r = 0.95$			$r = 0.9$		
	$r^* = 1$	$r^* = 0.95$	$r^* = 0.9$	$r^* = 1$	$r^* = 0.95$	$r^* = 0.9$	$r^* = 1$	$r^* = 0.95$	$r^* = 0.9$
Demeaned									
Sequence	0.052	0.540	0.770	0.206	0.038	0.211	0.579	0.111	0.039
$P_T(ols)$	0.034	0.513	0.817	0.206	0.029	0.252	0.579	0.097	0.038
$P_T(loc)$	0.031	0.508	0.817	0.199	0.027	0.254	0.568	0.092	0.039
$MSB(ols)$	0.036	0.375	0.695	0.091	0.035	0.233	0.268	0.044	0.049
$MSB(loc)$	0.033	0.367	0.690	0.088	0.034	0.236	0.260	0.039	0.048
Detrended									
Sequence	0.047	0.182	0.478	0.054	0.046	0.227	0.146	0.060	0.051
$P_T(ols)$	0.031	0.120	0.420	0.049	0.028	0.175	0.146	0.058	0.038
$P_T(loc)$	0.026	0.110	0.417	0.043	0.024	0.174	0.132	0.048	0.034
$MSB(ols)$	0.035	0.118	0.369	0.033	0.036	0.173	0.092	0.040	0.046
$MSB(loc)$	0.030	0.109	0.362	0.029	0.034	0.174	0.083	0.033	0.044

Note: The pseudo data are drawn from the model in (1) and (2) processes with N(0,1) errors, ρ given in the first column heading and u_0=0. Nominal coverage is set to 95%. Tests are implemented allowing for deterministics as appropriate for the test (see text). Results are proportion of rejections based on 20000 Monte Carlo replications with sample sizes of 100 observations.

TABLE 2: Small Sample results with serial dependence: One minus Coverage Rate,Conditional case.

	Demeaned			Detrended		
	$r_0 = 1$	$r_0 = 0.95$	$r_0 = 0.9$	$r_0 = 1$	$r_0 = 0.95$	$r_0 = 0.9$
π=0.3,θ=0.						
Sequence	0.071	0.090	0.107	0.093	0.102	0.122
$P_T(ols)$	0.057	0.073	0.095	0.080	0.084	0.106
$P_T(loc)$	0.046	0.063	0.084	0.060	0.064	0.086
$MSB(ols)$	0.058	0.068	0.097	0.071	0.076	0.100
$MSB(loc)$	0.038	0.052	0.083	0.050	0.058	0.082
π=-0.3,θ=0.						
Sequence	0.077	0.095	0.125	0.110	0.121	0.170
$P_T(ols)$	0.056	0.074	0.121	0.098	0.093	0.132
$P_T(loc)$	0.042	0.067	0.118	0.071	0.074	0.123
$MSB(ols)$	0.073	0.091	0.154	0.091	0.098	0.143
$MSB(loc)$	0.048	0.078	0.149	0.066	0.083	0.137
π=0,θ=0.3						
Sequence	0.076	0.087	0.101	0.106	0.119	0.142
$P_T(ols)$	0.055	0.068	0.094	0.092	0.090	0.111
$P_T(loc)$	0.040	0.058	0.089	0.067	0.069	0.098
$MSB(ols)$	0.074	0.080	0.122	0.091	0.095	0.119
$MSB(loc)$	0.047	0.065	0.111	0.063	0.077	0.109
π=0,θ=-0.3						
Sequence	0.075	0.108	0.134	0.102	0.126	0.153
$P_T(ols)$	0.064	0.089	0.121	0.090	0.108	0.139
$P_T(loc)$	0.053	0.075	0.104	0.069	0.079	0.110
$MSB(ols)$	0.063	0.084	0.117	0.080	0.098	0.128
$MSB(loc)$	0.045	0.063	0.096	0.057	0.076	0.102

Note: The data is generated according to equations (1) and (2) of the text with $(1+pL)n_t = (1+qL)e_t$ and $e_t \sim N(0,1)$. Otherwise the pseudo data is generated as in Table 1. In each case the spectral density at frequency zero is estimated using the method described in section VI.2. Reported is one minus the coverage rate, i.e. the probability of false exclusion.

TABLE 3: Small Sample results. Unconditional case.

	$r = 1$			$r = 0.95$			$r = 0.9$		
	$r^* = 1$	$r^* = 0.95$	$r^* = 0.9$	$r^* = 1$	$r^* = 0.95$	$r^* = 0.9$	$r^* = 1$	$r^* = 0.95$	$r^* = 0.9$
Demeaned									
Sequence	0.042	0.378	0.717	0.089	0.043	0.305	0.284	0.059	0.060
$Q_T(ols)$	0.037	0.358	0.683	0.089	0.039	0.253	0.284	0.066	0.047
$Q_T(loc)$	0.033	0.354	0.684	0.086	0.037	0.254	0.276	0.062	0.046
$MSB(ols)$	0.036	0.328	0.664	0.070	0.036	0.259	0.226	0.051	0.048
$MSB(loc)$	0.032	0.325	0.666	0.067	0.035	0.263	0.219	0.047	0.048
Detrended									
Sequence	0.042	0.140	0.418	0.040	0.045	0.225	0.094	0.040	0.063
$Q_T(ols)$	0.035	0.120	0.370	0.035	0.038	0.184	0.093	0.043	0.049
$Q_T(loc)$	0.029	0.114	0.370	0.030	0.033	0.184	0.083	0.035	0.045
$MSB(ols)$	0.032	0.104	0.318	0.026	0.038	0.159	0.067	0.040	0.048
$MSB(loc)$	0.027	0.099	0.316	0.022	0.035	0.159	0.059	0.035	0.045

Note: As per Table 1 with u_0 drawn from $N(0, \sigma^2/(1-\rho))$ distribution.

Figure 1a.

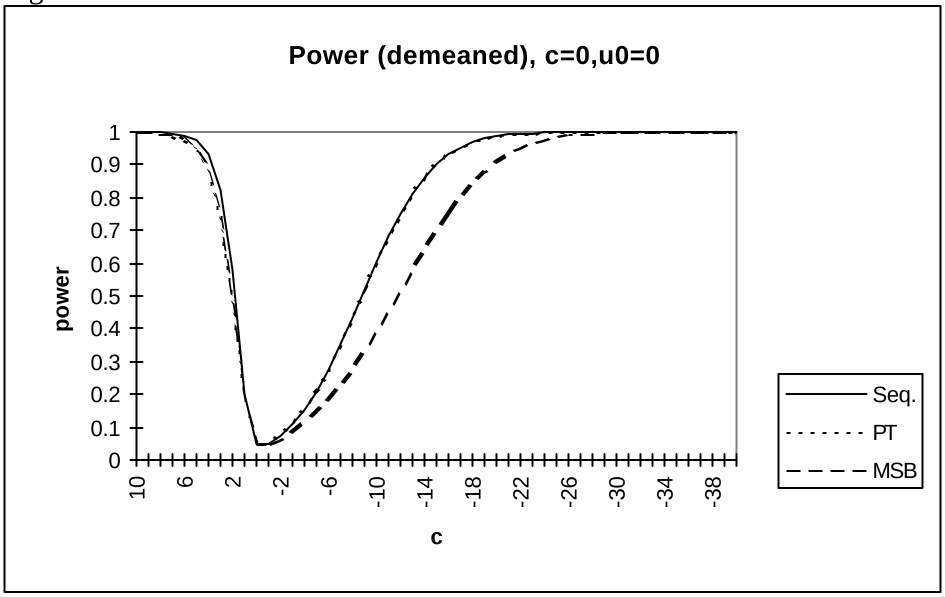


Figure 1b.

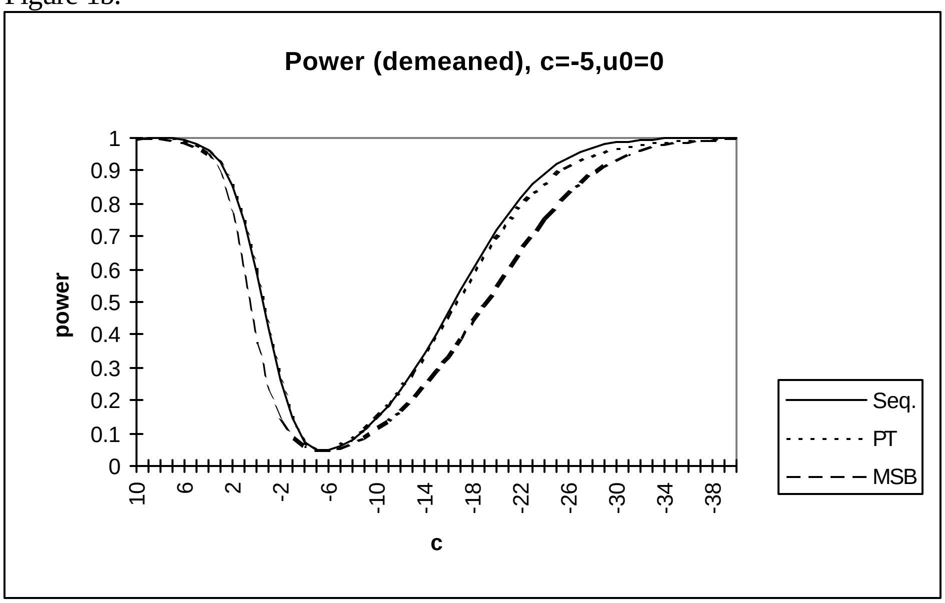


Figure 1c:

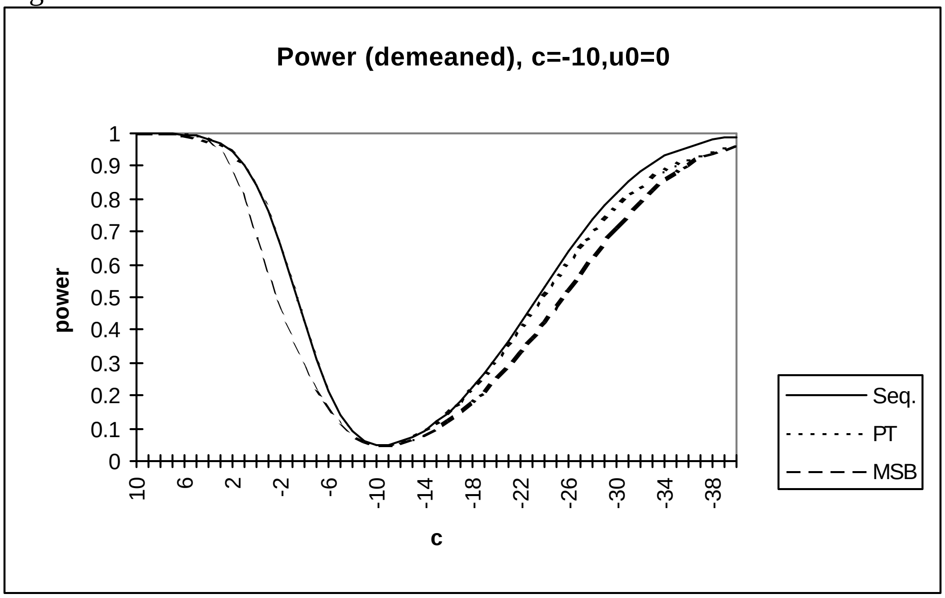


Figure 1d.

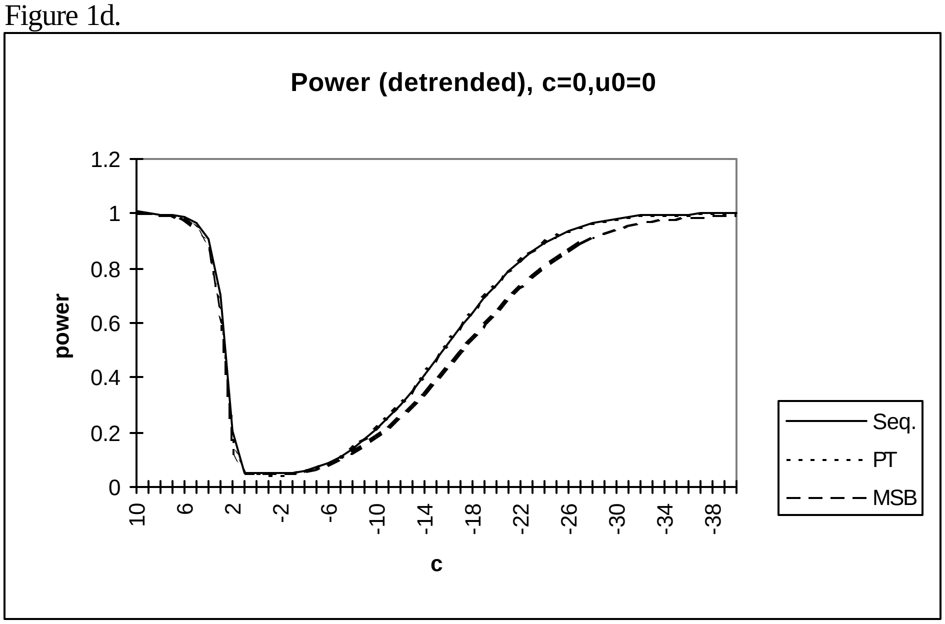


Figure 1e:

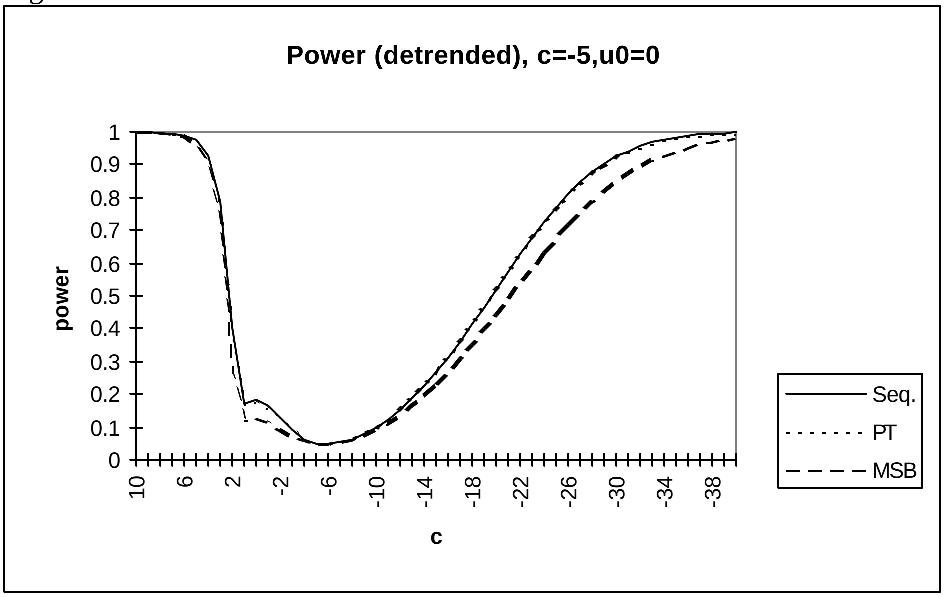


Figure 1f:

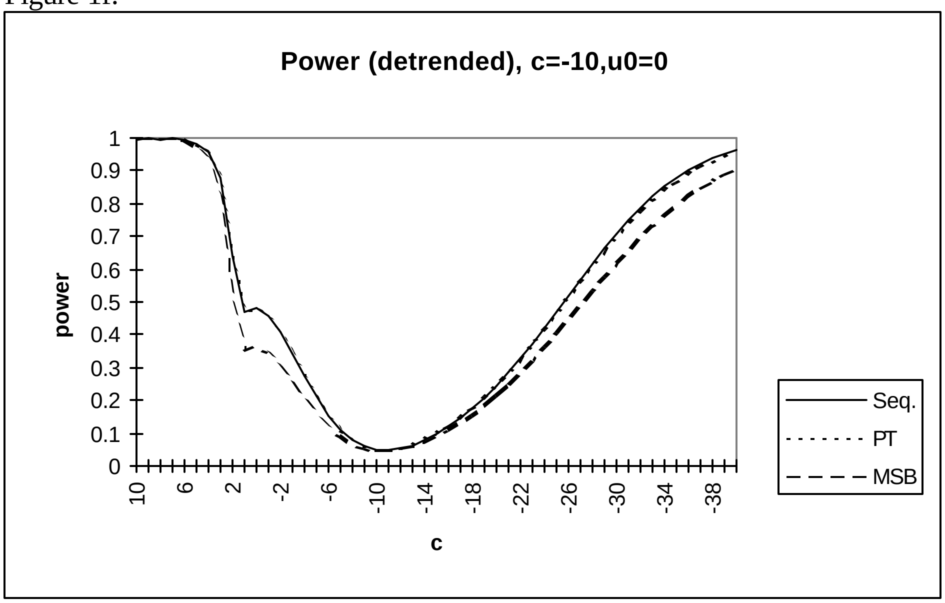


Figure 2a.

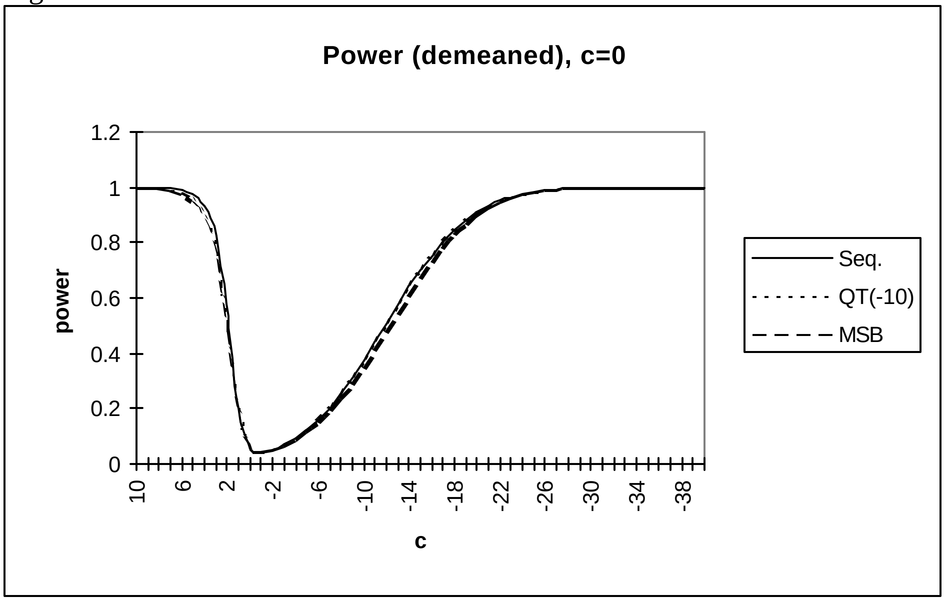


Figure 2b.

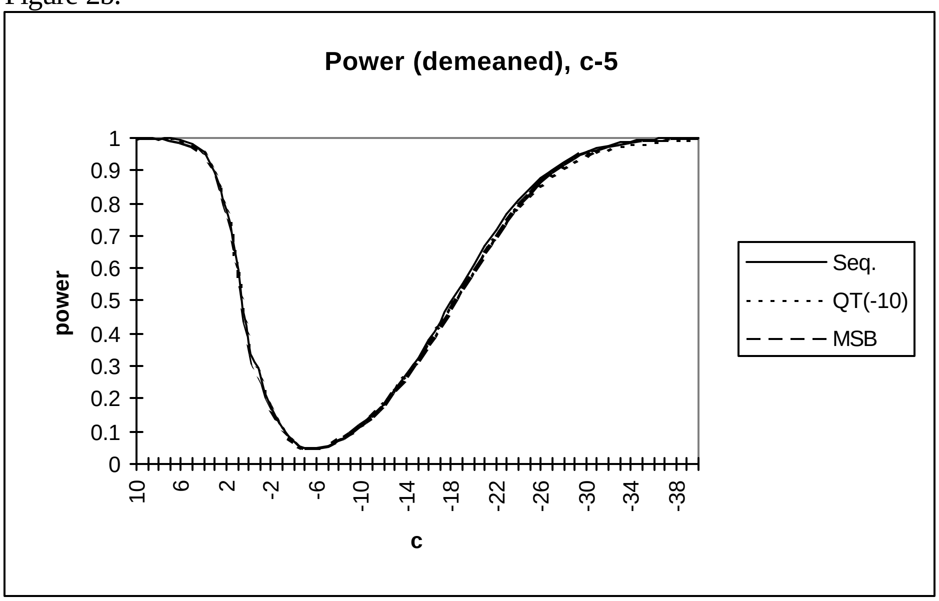


Figure 2c.

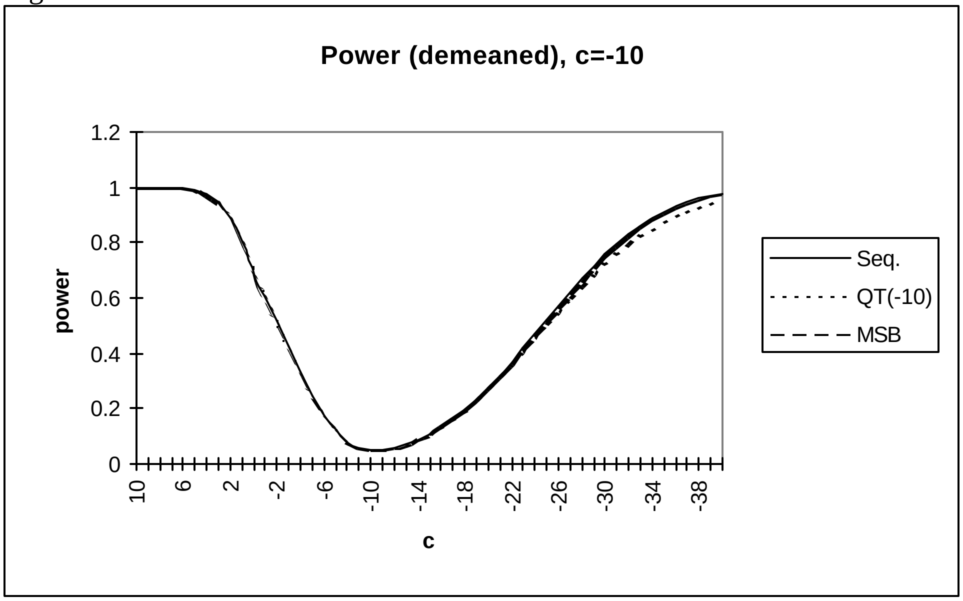


Figure 2d.

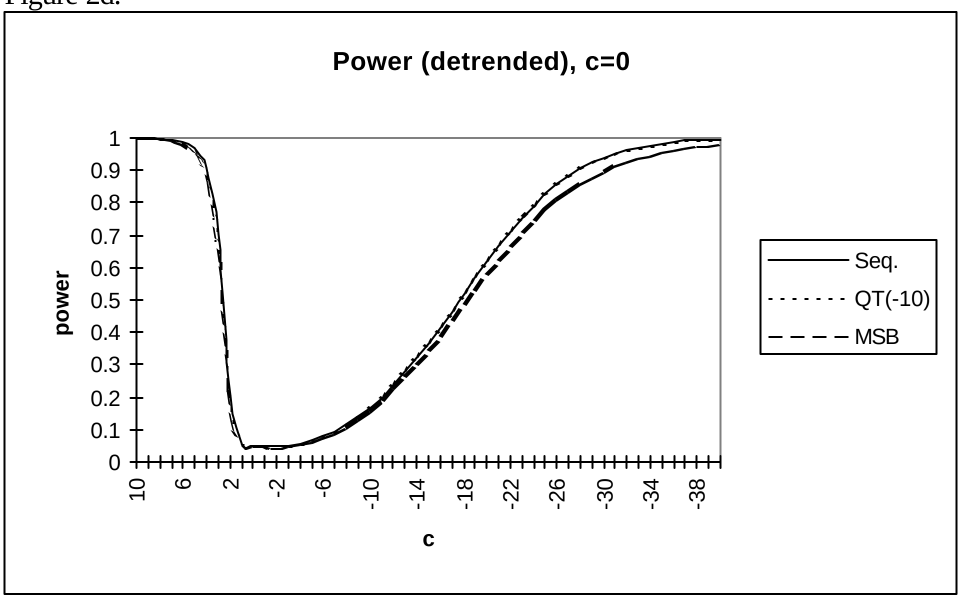


Figure 2e.

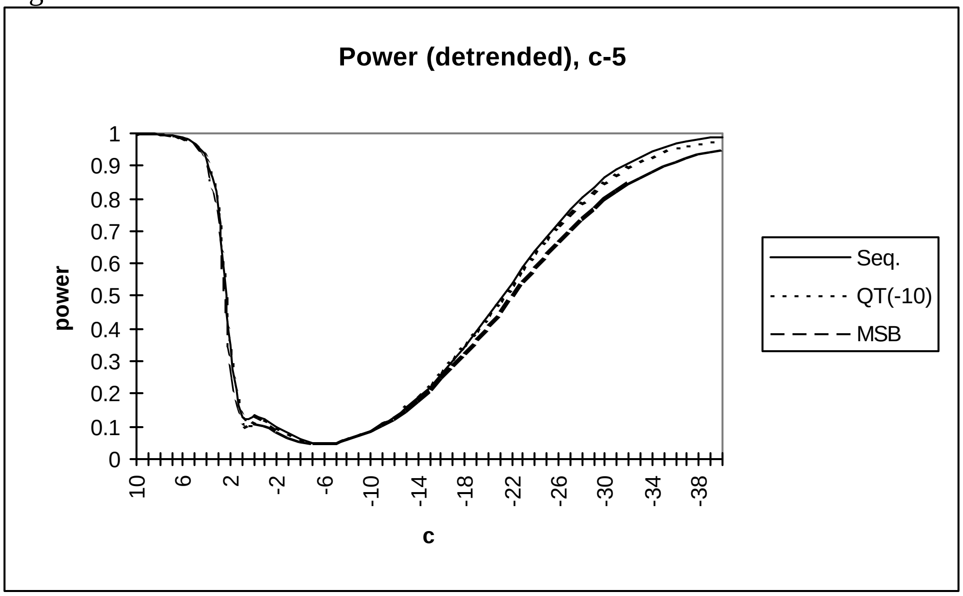


Figure 2f.

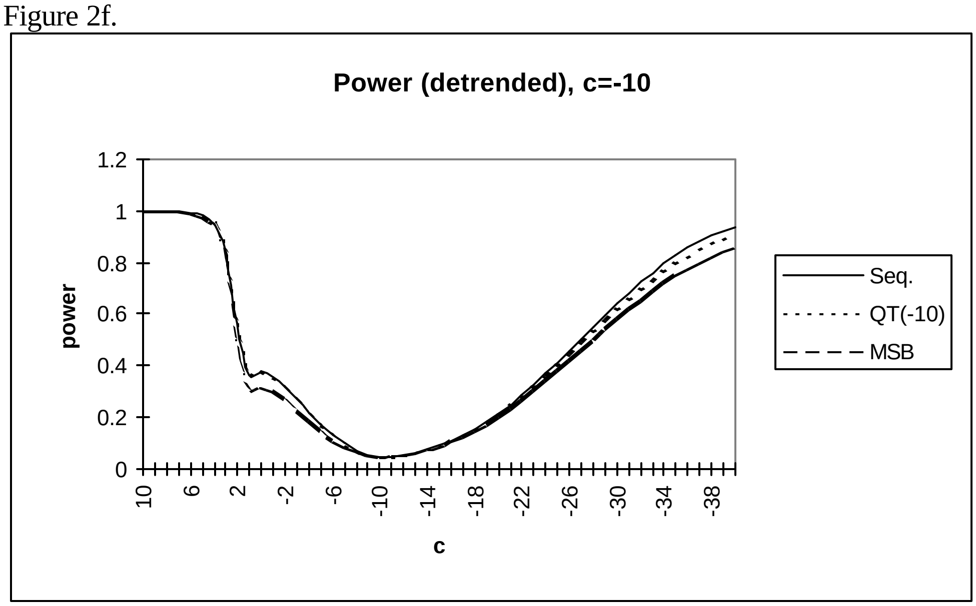
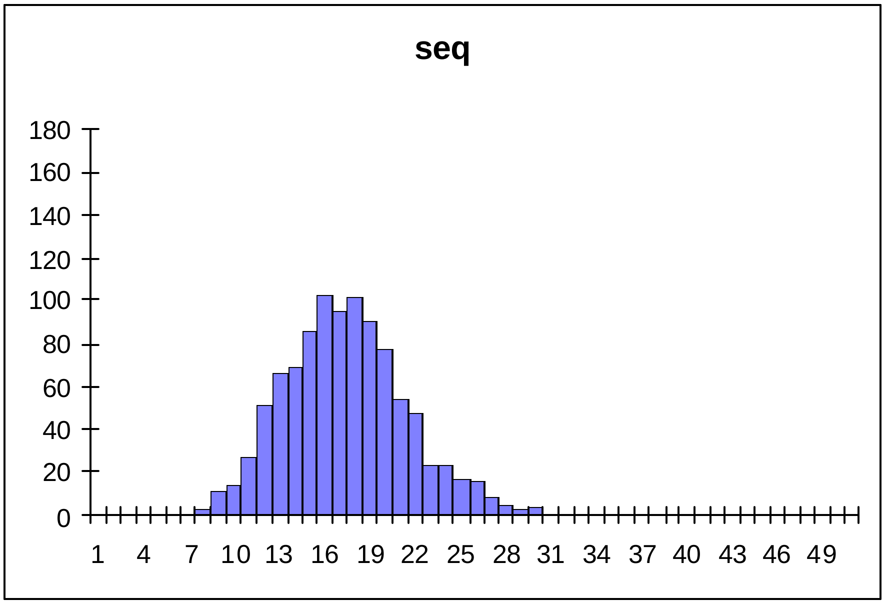
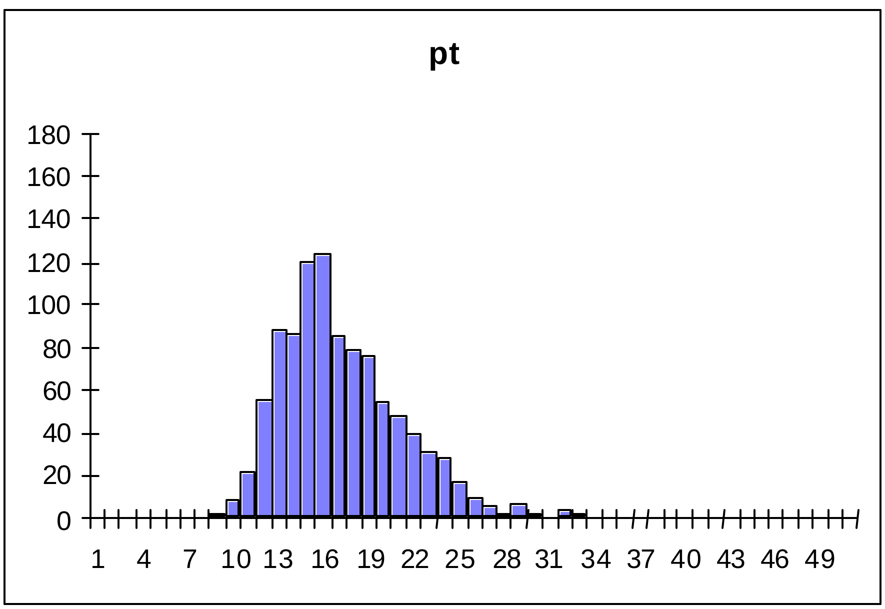


Figure 3a: Sequence Interval Length.



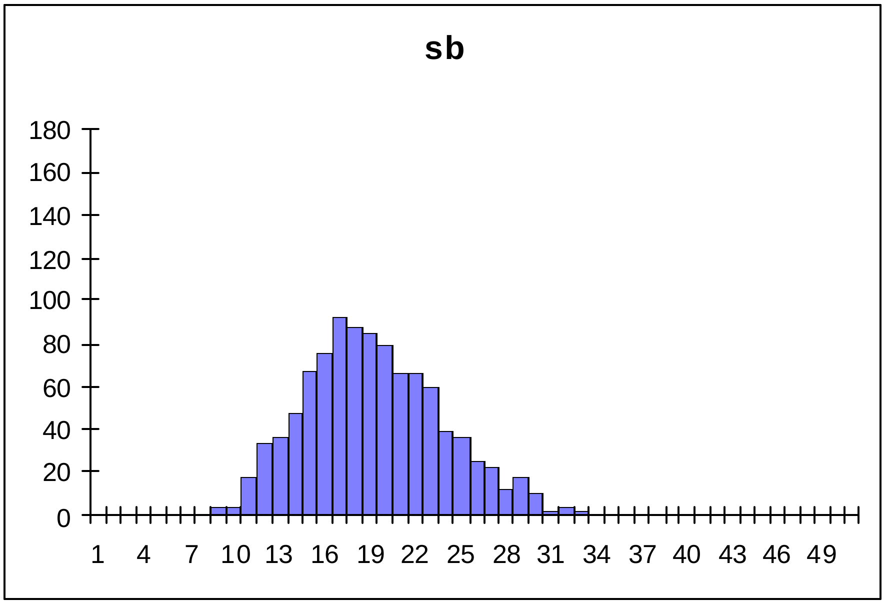
Notes: Vertical axis gives # occurrences in the Monte Carlo, the horizontal axis gives the lengths of the intervals in terms of $c = 100(\rho - 1)$ where $\rho = 0.95$.

Fig 3b: P_T interval length



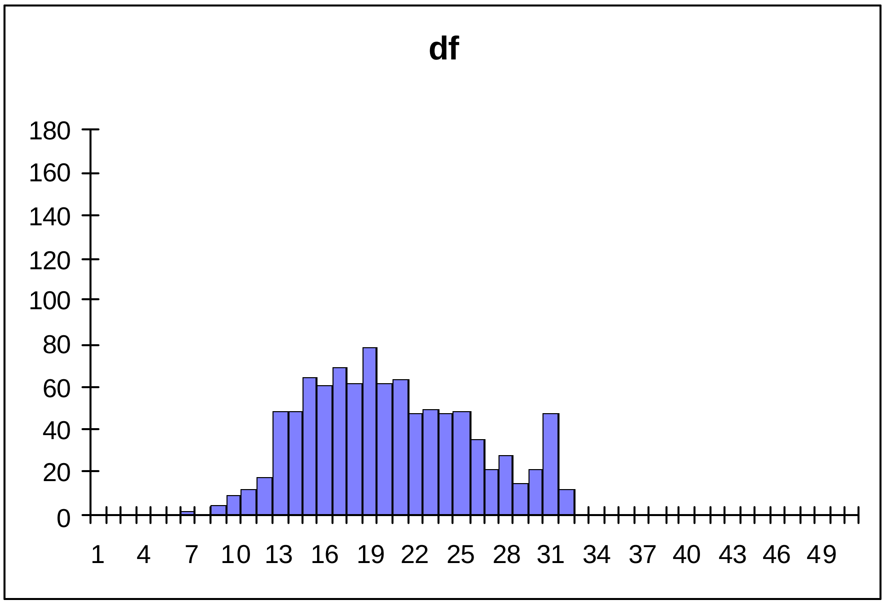
Notes: As per Figure 3a.

Fig 3c: MSB figure Length



Notes: As per Figure 3a.

Fig 3d: Dickey Fuller Length



Notes: As per figure 3a.