Title
Testing for Unit Roots with Stationary Covariances

Authors
Elliott, Graham
Jansson, Michael

Publication Date
2000-05-01

2000-06

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

TESTING FOR UNIT ROOTS WITH STATIONARY COVARIATES

BY

GRAHAM ELLIOTT

AND

MICHAEL JANSSON

DISCUSSION PAPER 2000-06
MAY 2000

Testing for Unit Roots with Stationary Covariates*

Graham Elliott
University of California, San Diego

Michael Jansson
University of Aarhus

First Draft: September 9, 1999
This Draft: May 1, 2000

Abstract.

We derive the family of tests for a unit root with maximal power against a point alternative when an arbitrary number of stationary covariates are modeled with the potentially integrated series. We show that very large power gains are available when such covariates are available. We then derive tests which are simple to construct (involving the running of vector autoregressions) and achieve at a point the power envelopes derived under very general conditions. These tests have excellent properties in small samples. We also show that these are obvious and internally consistent tests to run when identifying structural VAR's using long run restrictions.

JEL Code: C3.

Corresponding Author: Michael Jansson
Department of Economics
Building 350
University of Aarhus
DK - 8000 Aarhus C
Denmark

* The authors would like to thank conference participants at the September 1999 CNMLE conference at Svinkløv for comments.

1. Introduction.

Due to the effects of the assumption of a unit root in a variable on both the econometric method used and the economic interpretation of the model examined, it is quite common to pre-test the data for unit roots. This is typically done by either (or both) testing variables one by one for unit roots or by examining cointegrating rank using Johansen (1988) tests or their asymptotic equivalent.

In testing variables one by one, commonly the t-test method of Dickey and Fuller (1979) is employed. This method is asympotically optimal when the data is stationary and is a natural statistic to consider. However in the unit root case there are many other tests available have greater power. Elliott et. al (1996) showed that there is no uniformly most powerful test for this problem and derived tests that were approximately most powerful in the sense that they have asymptotic power close to the envelope of most powerful tests for this problem.

This paper considers a model where there is one series that potentially has a unit root, and that this series potentially covaries with some available stationary variables. In a model similar to the one examined here, Hansen (1995) demonstrated that in a model with no deterministic terms that no uniformly most powerful test for a unit root in the presence of stationary covariates exists and that power gains are to be had from using these covariates. He suggested covariate augmented Dickey Fuller (CADF) tests and showed that these tests had greater power than tests that ignored these covariates[1].

This paper extends the results in Hansen (1995) in a number of ways. First, we show that the point optimal tests implicit in the power envelope derived in Hansen (1995) and computed when all nuisance parameters are known are feasible when these parameters are not known. We also extend the results by deriving the power envelope in the more empirically relevant cases of where constants and/or time trends are also included in the regression. We propose tests that are feasible to construct with data and attain the power envelope at a point. These tests have good power at other points as well. We then show that these are natural tests to report in justifying the unit root assumption in the popular method of identifying structural vector autoregressions from long run restrictions (as suggested by Blanchard and Quah (1989)).

The paper is set up as follows. In the next section the model is introduced, and the power bounds for the problem are established. In the third section, tests which feasibly attain these power bounds at a point are derived and discussed. Section four examines the tests empirically using Monte Carlo methods. A fifth section discusses the tests as they relate to identifying structural VAR’s from long run restrictions. The final section concludes. All proofs are contained in an appendix.

2. Model and Power Envelopes.

Consider the model

$$z_t = b_0 + b_1 t + u_t \qquad (1)$$

and

$$A(L)\begin{pmatrix}(1-rL)u_{y,t}\\ u_{x,t}\end{pmatrix} = e_t \qquad (2)$$

where $z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$, x_t is an mx1 vector, y_t is 1x1, $b_0 = \begin{pmatrix} b_{y0} \\ b_{x0} \end{pmatrix}$, $b_1 = \begin{pmatrix} b_{y1} \\ b_{x1} \end{pmatrix}$, $u_t = \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix}$ and A(L) is a finite polynomial of order k in the lag operator L. For the constructed test statistics we will assume that

A1. |A(z)|=0 has roots outside the unit circle.

A2. e_t is a martingale difference sequence satisfying a multivariate invariance principle, i.e. $T^{-1/2}\sum_{s=1}^{[T\bullet]} e_t \Rightarrow \Sigma^{1/2}\left[W_1(\cdot)\ \ V(\cdot)'\right]'$, where W_1(.) is a univariate standard Brownian Motion on C[0,1], V(.) is and mx1 standard Brownian Motion, Σ is positive definite and ⇒ denotes weak convergence.

A3. $u_0 = 0_p(1)$.

Define $u_t(r) = \left[(1-rL)u_{y,t}\ \ u_{xt}'\right]'$ with spectral density at frequency zero (scaled by 2π) Ω, so we have $\Omega = A(1)^{-1}\Sigma A(1)^{-1\prime}$ where we can partition this after the first column and row so that

$$\Omega = \begin{bmatrix} w_{yy} & w_{yx} \\ w_{yx}' & \Omega_{xx} \end{bmatrix}$$

(we partition Σ similarly). We will further define R^2 = δ'δ where $d = \Omega_{xx}^{-1/2} w_{yx}' w_{yy}^{-1/2}$ is an mx1 vector of correlations between the x's and the quasi difference of y at frequency zero. The R^2 value will represent the contribution of the stationary variables as it is zero when these variables are not correlated in the long run with the shocks to $(1-\rho L)y_t$ at the zero frequency and one if there is perfect correlation.

[1] There is also a discussion of this work in Caporale and Pittis (1999).

In this paper we consider five cases indexed by superscript i (i=1,2,3,4,5) for the deterministic part of the model (where parameters are free unless otherwise stated)

Case 1: $b_{y0} = b_{y1} = 0$ and $b_{x0} = b_{x1} = 0$.

Case 2: $b_{y1} = 0$ and $b_{x0} = b_{x1} = 0$.

Case 3. $b_{y1} = 0$ and $b_{x1} = 0$.

Case 4: $b_{x1} = 0$.

Case 5: No restrictions.

Each of these cases can be characterized by the restriction $\left(I_{2(m+1)} - S_i\right)b = 0$ where β = [β_0' β_1']', S_i is a 2(m+1)x2(m+1) matrix where S_1=0, $S_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $S_3 = \begin{pmatrix} I_{m+1} & 0 \\ 0 & 0 \end{pmatrix}$, $S_4 = \begin{pmatrix} I_{m+2} & 0 \\ 0 & 0 \end{pmatrix}$ and S_5 is the identity matrix.

This represents a fairly general set of models in which we have a VAR in the model of x and the quasi difference of y. We wish to test that the parameter ρ is equal to one (y_t has a unit root) against alternatives that this root is less than one. Following the general methods of King (1980, 1988) we will examine Neyman Pearson tests for this hypothesis. Following the application of these methods to testing for unit roots in Elliott, Rothenberg and Stock (1996), Elliott (1999)) we will examine Neyman Pearson tests for this hypothesis under simplifying assumptions, and then in the following section we will derive general tests that are asymptotically equivalent to these optimal tests.

With the assumption that A(L)=I (so that Ω=Σ) and assuming the e_t are normally distributed and u_{y0}=0 we will examine tests against the local alternative that c = $\bar{c}$ <0 where $r = 1 + c/T$ and $\bar{r} = 1 + \bar{c}/T$ with c, $\bar{c}$ fixed (we will suppress the dependence of ρ on T in the notation).

The likelihood ratio test statistic for the hypothesis is given by

$$\Lambda^i(1, \bar{r}) = \sum_{t=1}^{T} \hat{u}_t^i(\bar{r})'\Sigma^{-1}\hat{u}_t^i(\bar{r}) - \sum_{t=1}^{T} \hat{u}_t^i(1)'\Sigma^{-1}\hat{u}_t^i(1)$$

where we have for r = $\bar{r}$,1 that

$$\hat{u}_t^i(r) = z_t(r) - d_t\,(r)'\,\hat{b}^{\,i}$$

where $z_t\,(r) = \begin{bmatrix} (1-rL)\,y_t \\ x_t \end{bmatrix}$ for t>1 and $z_1\,(r) = \begin{bmatrix} y_1 \\ x_1 \end{bmatrix}$,

$d_t(r)' = \begin{bmatrix} 1-r & 0 & (1-rL)t & 0 \\ 0 & I_m & 0 & I_m t \end{bmatrix}$ for t>1, $d_1(r)' = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & I_m & 0 & I_m \end{bmatrix}$, and

$$\hat{b}^{\,i} = \left[S_i \left(\sum_{t=1}^{T} d_t(r)\Sigma^{-1} d_t(r)' \right) S_i \right]^{-} \left[S_i \sum_{t=1}^{T} d_t(r)\Sigma^{-1} z_t(r) \right]$$

where D^- is the Moore Penrose inverse of D.

The test has rejection regions of the form $\{y_t, x_t\ : \Lambda^i(1, \bar{r}) - \bar{c} < b\}$ where b is a critical value.

Case 1: No Deterministics.

The model above is similar to that of Hansen (1995) when there are no deterministic terms (S_1 =0) in the model. In this case we have $\hat{u}_t(r) = u_t(r)$ and

Theorem 1.

For the model in (1) and (2) with A(L)=I, e_t independent N(0,S) random variables and A3 holding then with $r = 1 + c/T$ *and* $\bar{r} = 1 + \bar{c}/T$ *with c,* $\bar{c}$ *fixed as T ® ¥ then the most powerful test of H_0: c=0 vs. H_a: c=* $\bar{c}$ *<0 has asymptotic power function*

$$P(c, \bar{c}, R^2) = \Pr\left[y^{\,1}(c, \bar{c}, R^2) < b(\bar{c}, R^2) \right]$$

where

$$y^{\,1}(c, \bar{c}, R^2) = (\bar{c}^{\,2} - 2c\bar{c}) \int W_{1c}(l)^2 dl - 2\bar{c} \int W_{1c}(l)\,dW_1(l)$$
$$+ (\bar{c}^{\,2} - 2c\bar{c}) \left(\frac{R^2}{1-R^2} \right) \int W_{1c}(l)^2 dl + 2\bar{c} \frac{R}{\sqrt{1-R^2}} \int W_{1c}(l)\,dW_2(l)$$

and $b(\bar{c}, R^2)$ *is a constant depending on* $\bar{c}$ *and R^2.*

This is apart from a scale factor the same as that reported in Hansen (1995)[2]. A number of features are noteworthy. Firstly, the dependence of the test on $\bar{c}$ indicates that no uniformly most powerful test is available for this problem, power depends on the choice of the alternative. Second, the test is the sum of nonstandard functionals of Brownian motions and a mixed normal term. Third, the test depends on the parameter R^2, which summarizes the extent to which the covariates are correlated with the correctly differenced y_t at the zero frequency. A value of R^2=0 indicates the case where δ_i=0 for all i, the case where none of the x variables are correlated with the y variable at the zero frequency (so the second line of the limit expression is zero). In this case the result in Theorem 1 is equivalent to Theorem 1 of Elliott et. al (1996), thus the most powerful tests coincide asymptotically with tests which do not use the information in the covariates.

Figure 1a examines the power envelopes[3] derived in Theorem 1 (these replicate the results of Hansen (1995)). As can be seen, the power envelope when R^2=0 is the lower bound power - this is the relevant envelope if no covariate information is employed (as derived in Elliott et. al. (1996)) and is equivalent to the case where no useful covariate information is available. When R^2 is greater than zero, the power attainable increases above this lower bound. This indicates that using covariates has the potential to greatly increase the power of tests for a unit root, as indicated by Hansen (1995). The closer is R^2 to one, the more powerful the optimal test[4].

Cases 2-5: Constant and/or Time Trends Included.

The more interesting cases practically are those where β is not fully known.

Theorem 2.

For the models in (1) and (2) with A(L)=I, e_t independent N(0,S) random variables, A_3 holding with $r = 1 + c/T$ *and* $\bar{r} = 1 + \bar{c}/T$ *with c,* $\bar{c}$ *fixed as T ® ¥ then the most powerful test of H_0: c=0 vs. H_a: c=* $\bar{c}$ *<0 invariant to deterministic terms have asymptotic power functions*

$$P(c,\bar{c},R^2) = \Pr\left[y^{i}(c,\bar{c},R^2) < b^{i}(\bar{c},R^2) \right]$$

where

[2] We also have a notational difference in that our R^2 is defined in Hansen (1995) as 1-R^2. We changed the notation to accord with the usual use of R^2.
[3] The power envelope is the power of a test with alternative $\bar{c}$ =c for each c (thus is a different test at each alternative, and is the envelope of power functions of the point optimal tests).
[4] The asymptotic results are not appropriate at R^2=1, which is readily seen from the limit expression which would not be finite at this point.

Case 2: $y^{2}(c,\bar{c},R^{2}) = y^{1}(c,\bar{c},R^{2})$

Case 3:

$$y^{3}(c,\bar{c},R^{2}) = (\bar{c}^{2}-2c\bar{c})\int W_{1c}(l)^{2}\,dl - 2\bar{c}\int W_{1c}(l)dW_{1}(l)$$
$$+(\bar{c}^{2}-2c\bar{c})\left(\frac{R^{2}}{1-R^{2}}\right)\int W_{1c}^{m}(l)^{2}\,dl + 2\bar{c}\frac{R}{\sqrt{1-R^{2}}}\int W_{1c}^{m}(l)dW_{2}(l)$$

Case 4:

$$y^{4}(c,\bar{c},R^{2}) = (\bar{c}^{2}-2c\bar{c})\int W_{1c}(l)^{2}\,dl - 2\bar{c}\int W_{1c}(l)dW_{1}(l) + W_{1c}(1)^{2}$$
$$-\frac{1}{h}\begin{pmatrix}(1-\bar{c})W_{1c}(1)+\bar{c}^{2}\int lW_{1c} + \left(\frac{R^{2}}{1-R^{2}}\right)\left[\frac{\bar{c}^{2}}{2}(c-\bar{c})\int W_{1c} - \bar{c}(c-\bar{c}))\int lW_{1c}\right] + \\ \sqrt{\frac{R^{2}}{1-R^{2}}}\left[\bar{c}\int l\,dW_{2}(l) - \frac{\bar{c}}{2}\int dW_{2}(l)\right]\end{pmatrix}^{2}$$
$$+(\bar{c}^{2}-2c\bar{c})\left(\frac{R^{2}}{1-R^{2}}\right)\int W_{1c}^{m}(l)^{2}\,dl + 2\bar{c}\frac{R}{\sqrt{1-R^{2}}}\int W_{1c}^{m}(l)dW_{2}(l)$$

Case 5:

$$y^{5}(c,\bar{c},R^{2}) = (\bar{c}^{2}-2c\bar{c})\int W_{1c}(l)^{2}\,dl - 2\bar{c}\int W_{1c}(l)dW_{1}(l)$$
$$+W_{1c}(1)^{2} - \left((1-\bar{c})W_{1c}(1)+\bar{c}^{2}\int l\,W_{1c}(l)dl\right)^{2} / \left(1+\frac{\bar{c}^{2}}{3}-\bar{c}\right)$$
$$+(\bar{c}^{2}-2c\bar{c})\left(\frac{R^{2}}{1-R^{2}}\right)\int W_{1c}^{t}(l)^{2}\,dl + 2\bar{c}\frac{R}{\sqrt{1-R^{2}}}\int W_{1c}^{t}(l)dW_{2}(l)$$

and $h = \left(1+\frac{\bar{c}^{2}}{3}-\bar{c}\right)+\frac{\bar{c}^{2}}{12}\frac{R^{2}}{1-R^{2}}$ *where* $b^{i}(\bar{c},R^{2})$ *are constants depending on* $\bar{c}$ *and* R^2 *as well as the case i.*

Figures 1b through to 1d asymptotically approximate the power envelopes for cases 3 through 5 respectively (case 2 is equivalent to case 1). When $R^2=0$, the stationary covariates do not help in the testing procedure and the power of the invariant tests are equivalent to those derived in Elliott et. al. (1996). This means that in case 3 there is no loss of power asymptotically when β_{y0} is unknown, and in cases 4 and 5 there is a loss of power compared to case 1 where the deterministic terms are known (Cases 4 and 5 have identical power functions when R^2=0, and correspond to the case in Elliott et. al. (1996) of the inclusion of a constant and time trend).

When R^2 is nonzero, power functions are affected by not knowing the deterministic part of the model. We also have that the optimal test depends on R^2, the extent to which the stationary covariates are correlated with $(1-\rho L)y_t$ in the long run. Comparing Figures 1a to 1b we see the effect of not knowing the constant terms. This effect is relatively small, for example when R^2=0.5 and c=-5 the power envelope in the constants known case is 70% whilst when the constants are unknown this power is 62%. However both these powers are substantially above that of the case where no covariates are employed, where the power envelope attains a power of 32%.

As in the case where there are no covariates, the effect on the power envelopes for the case where the trend terms (coefficients on time trends) are not known is quite large. In the case mentioned above, where R^2=0.5 and c= -5 the maximal power in case 4 is 33%, far below the 62% when only coefficients on the constants are known. When the coefficient on the trend in the x_t regressions is known, this power rises to 36%. Notice though that the maximal power in this case even when constants and coefficients on the time trend are estimated is (just) above that for the case where stationary covariates are ignored and the coefficient on the time trend is known. In general the power losses from not knowing the coefficient on the trends in the x_t regressions is small, between zero (when R^2 is small) and 6% or so (when R^2 is large).

There is clearly the potential for much to be gained in terms of power from exploiting stationary covariates in constructing tests for a unit root. The construction of tests that achieve these gains is addressed in the next section.

3. Feasible Tests.

In this section we will derive families of tests that asymptotically attain the power bounds derived above at pre-specified points.

The model is as in equations (1) and (2) with assumptions A.1, A.2 and A.3. As in the previous model we consider four cases for the deterministic component of the model. For each case define

$$\tilde{u}_t^i(r) = z_t(r) - d_t(r)'\tilde{b}^i(r)$$

where

$$\tilde{b}^{i} = \left[S_i \left(\sum_{t=1}^{T} d_t(r) \tilde{\Omega}^{-1} d_t(r)' \right) S_i \right]^{-} \left[S_i \sum_{t=1}^{T} d_t(r) \tilde{\Omega}^{-1} z_t(r) \right] \tag{3}$$

and $\tilde{\Omega}^{-1}$ is a consistent estimate of Ω^{-1} under the null.

Run VAR's (for r= $\bar{r}$,1)

$$\tilde{A}(L,r)\tilde{u}_t^{i}(r) = \tilde{e}_t(r)$$

and construct the estimated variance covariance matrices

$$\tilde{\Sigma}(r) = T^{-1} \sum_{t=k+1}^{T} \tilde{e}_t(r)\tilde{e}_t(r)'$$

then the proposed test is

$$\tilde{\Lambda}^{i}(1,\bar{r}) = T\left(tr\left[\tilde{\Sigma}(1)^{-1}\tilde{\Sigma}(\bar{r})\right] - (m+\bar{r})\right)$$

This test will have asymptotic power that achieves the power bound at $\bar{c}$ under the assumptions

Theorem 3.

For the model in (1) and (2) with b=0 with assumptions A1, A2 and A3 holding then as T ® ¥

$$\tilde{\Lambda}^{i}(1,\bar{r}) \Rightarrow y^{i}(c,\bar{c},R^2) - \bar{c} \text{ .}$$

Thus the critical values for the test depend on the alternative chosen ($\bar{c}$) and R^2 . The feasible test in the case of β=0 asymptotically achieves the highest power possible at $\bar{c}$. We have chosen here to let $\bar{c}$ = -7 for cases 1 and 2 and $\bar{c}$ =-13.5 for cases 3 and 4 (which follows the choice of Elliott et. al . (1996), which was shown in this case of R^2=0 to be a choice that ensures maximal power at power 50%). In principle and practice we could choose different values for $\bar{c}$ depending on R^2, however as R^2 rises above zero lack of power is becoming less problematic so it seems reasonable to us to choose $\bar{c}$ for the worst case scenario.

Asymptotic critical values for the test for selected values of R^2 are given in Table 1. In practice we still require knowledge of the value for R^2. This can be estimated consistently from the data without knowledge of r . The method we suggest is the following

a) estimate ρ from a regression of y_t on y_{t-1}, deterministics and lags of changes in y_t.

b) run the VAR $A(L, \hat{r})z_t(\hat{r}) = \det + e_t(\hat{r})$

(choose deterministics appropriate to the case in each of these steps).

c) estimate $\hat{\Omega} = \hat{A}(1, \hat{r})^{-1}\hat{\Sigma}\hat{A}(1, \hat{r})^{-1}{}'$ where $\hat{\Sigma} = T^{-1}\sum_{t=k+1}^{T}\hat{e}_t(\hat{r})\hat{e}_t(\hat{r})'$

d) estimate $\hat{R}^2 = \hat{w}_{yx}\hat{\Omega}_{xx}^{-1}\hat{w}_{yx}'/\hat{w}_{yy}$.

We then propose using the critical value for the estimated $\hat{R}^2$. The estimate of $\hat{\Omega}$ can be used for constructing the local estimates of the deterministic part in equation (3). This is valid asymptotically as this is a consistent estimator. For values of R^2 between the ones given in Table 1, interpolation can be used to estimate the critical value.

4. Evaluation of the Tests.

4.1. Large Sample Evaluation.

Figures 2a to 2d examine the power of the feasible test for each of the four different cases (specification of the deterministic part of the model). The figures give the results for R^2 = 0.1, 0.5 and 0.8. Accompanying the power curves are the power envelopes for the case given. In figures 2a and 2b it is seen clearly that very little power is lost by using a point optimal test. The feasible point optimal test has power that lies almost on top of the power envelope. This is very similar to the results of Elliott et. al. (1996), where for the case of R^2=0 this was found to be true. A similar result is true also when R^2 = 0.5. Here, the difference between the power envelope and the asymptotic power of the feasible test is small for alternatives further from the null, but a little larger for alternatives close to the null. For R^2 large this is even more apparent. Overall, even though allowing the choice of $\bar{c}$ to depend on R^2 may allow us to further minimize the difference between the power curve and power envelope, we do not pursue this here.

In figures 2c and 2d, where time trends are included in the y regression (cases 4 and 5), there is some difference between the power attainable by the point optimal tests and the power envelope (where in both these cases $\bar{c}$ =-13.5). As in Elliott et. al. (1996) when R^2 is close to zero this is not apparent, but becomes more apparent as R^2 gets large. The difference comes are relatively close alternatives. To the extent that very large values for R^2 are probably not too relevant empirically, this may not be too much of a problem. The suggestion from these graphs appears to be that the most useful choice of $\bar{c}$ in practice may depend on R^2. We also examined the power curves for the case where $\bar{c}$ =-7 to perhaps improve the closeness of the power curves to the envelopes for these near alternatives. When this alternative is chosen this indeed happens, however the tradeoff is that the power curves for R^2 small are not as close to the envelope for more distant alternatives. We recommend choosing $\bar{c}$ =-13.5 as power is more of a concern when R^2 is small.

4.2. Small Sample Evaluation.

We will examine various special case models in samples of 100 observations. Along with the above tests, we report results for the commonly applied test of Dickey and Fuller (1979) and also the P_T test of Elliott et. al (1996) as well as the Hansen (1995) CADF test.

Tables 2 through 4 report results of simulations of the model in (1) and (2) for each of the cases (models for the deterministics) respectively where A(L)=I (and this is known), e_t is normally distributed with variances equal to 1 and covariance equal to the value of δ reported in the Table. Results are reported for various values of δ. Size is given in the row corresponding to r =1 and (empirical) power against the indicated alternatives in the following rows. When there are no deterministic terms in the model the DF and P_T single equation tests do similarly well (see Elliott et. al. (1996) for a discussion of this). In the test proposed here, when R^2=0 power and size are comparable to the univariate tests indicating that even in small samples little may be lost by including extraneous information and doing the system test. As δ increases (R^2 increases), size remains well controlled whilst power rises considerably. Consider the case of the true r being equal to 0.96, the P_T test has power around 23% whilst if R^2 =0.25 the system test has power equal to 34%, roughly a 50% gain.

When a constant is included, the P_T statistic gains in power over the Dickey and Fuller (1979) t test are very large. Again, when R^2=0 the test proposed here has similar size and power to the P_T statistic indicating that little is lost adding extraneous stationary covariates. In general, size is less well controlled, especially for R^2 close to one (where the asymptotic theory would no longer be relevant, however it would not be expected that such models would be appropriate for real world data) . There is some evidence of power losses from not knowing the constant term. At a value of r = 0.96 the power when the constant is known (or zero) power is 49% compared to the unknown constant power of 45% when δ=0.7 (R^2=0.49). Even so, power for the test with the constant unknown is quite high in many cases, and is far beyond that achievable when covariates are not employed.

Similar results are found for the partially detrended (case 4) and detrended (case 5) models. In both of these cases we have power when using covariates to be substantially greater than when relevant covariates are ignored (for example, in case 4 when ρ =0.9, power of the test proposed here when δ=0.5 (R^2=0.25) is 20% for the Dickey and Fuller test and is 49% for the test with covariates employed. Overall, there is some loss of power from including the time trend in the x_t equations, which can be seen from comparing tables 4 and 5. In the case of r = 0.96 and δ=0.5 the power drops form 52% in case 3 to 49% in case 4. As indicated by the asymptotic results presented above, these losses are fairly small but not insignificant.

The effect of estimating R^2 in the computation of the test is examined in tables 6 and 7 (for cases 3 and 5 respectively). Here the results when R^2 is estimated are repeated from Tables 3 and 5 on the right hand side panels, whilst the same results using the critical value chosen using the true R^2 are given in the left hand panels. There is very little difference, even in a sample of 100 observations. Most of the differences in size and power are at the third decimal place. It is only for case 5 when R^2 is a little larger that there is much of an effect, but the effect is minor (in these cases there is a small power loss from estimating R^2).

Tables 8 and 9 compared the CADF test of Hansen (1995) with the feasible test derived here (again for the leading cases 3 and 5 respectively). The CADF test augments the usual Dickey and Fuller (1979) test with lags, leads and the contemporaneous values of x_t. In this table, with no serial correlation, this amounts to including x_t as a regressor in the ADF regression and then constructing the t-test of the unit root hypothesis as normal. As shown in Hansen (1995) this test also depends on R^2. In the comparison we use the same value of R^2 to compute critical values for each of the tests. In the first column of the CADF results, where R^2=0, we have essentially the same results as the Dickey and Fuller (1979) test in Tables 3 and 5 that ignores the covariates. This should be the case, the included x_t variable in the ADF regression has a population coefficient of zero in this case. Likewise, the first column of the $\hat{\Lambda}(1,\overline{r})$ test matches with the P_T test for the reasons we have described. This gives an insight into the difference in the two approaches, the difference between the CADF and $\hat{\Lambda}(1,\overline{r})$ is similar to the difference between the Dickey and Fuller (1979) approach and the Elliott et. al. (1996) approach. When R^2>0, we see that the $\hat{\Lambda}(1,\overline{r})$ test outperforms the CADF test in terms of power, although is slightly worse in size performance. The increases in power can be quite large. In the case 3 when δ=0.3 (R^2= 0.91) the power of the $\hat{\Lambda}(1,\overline{r})$ test is two to three times that of the CADF test for alternatives closer than 0.88. For case 5 the effects are not as dramatic, but still power gains of 50% or so are available from using the covariates test proposed here over the CADF test.

5. Unit Root Tests and Long Run Structural VAR Estimation.

Blanchard and Quah (1989) derive a method for identifying structural VAR's from restrictions placed on the spectral density of the data at frequency zero when there are known unit roots in the system. Consider the bivariate version of the model considered in this paper when we impose that the root ρ is equal to unity,

$$A(L)\begin{bmatrix}\Delta y_t \\ x_t\end{bmatrix} = e_t \,.$$

Inverting the lag polynomial gives us

$$\begin{bmatrix}\Delta y_t \\ x_t\end{bmatrix} = C(L)e_t$$

where $C(L)=A(L)^{-1}$. This model is not identified in the usual sense as we can write for any invertible K matrix $C(L)e_t = C(L)KK^{-1}e_t = D(L)h_t$. Since there exist an infinity of choices of K the model is not uniquely identified. In this bivariate system we require a single restriction so that the rotation K is unique for the model to be identified (this would be the order condition).

In such systems, y_t is permanently affected by shock(s) since it is an integrated process. On economic grounds, it may be interesting to identify the model such that only one of the structural shocks has a permanent effect on y_t. In Blanchard and Quah (1989) this argument meant that demand shocks could not have a permanent effect. In King et. al (1991) cointegration was used to imply a smaller number of permanent shocks than total shocks. In such cases it is possible to identify the model as the cumulated sum of the structural impulse responses, D(1), will be triangular as only one of the shocks has a long run effect on y_t.

For the model above, the identification scheme would set $d_{12}(1)=0$ where this is the (1,2) component of D(1). Since[5] the spectral density of the data at frequency zero (scaled by 2π) is $\Omega = D(1)D(1)'$ this amounts to taking the choleski decomposition of the estimated matrix $\hat{\Omega}$. Such a restriction is only interesting and useful in identification when the off diagonals for Ω are indeed nonzero, this is the case when $R^2>0$ also.

The crux of this approach to identification clearly is that y_t indeed does have a unit root. If instead there were no permanent effects then we would interpret D(1) differently and would have no reason to make this matrix triangular. So in practice a useful hypothesis test to report in undertaking this method would be a test

[5] We are using the usual identification from this literature so $E[h_t h_t'] = I$.

for a unit root in y_t. Further, when the imposed restriction is indeed informative, then $R^2>0$ and hence we are exactly in the cases where the tests of this paper yield power gains over univariate testing. Typically, such tests for a unit root to provide evidence of the validity of this restriction are undertaken using Dickey Fuller (1979) tests (see Gali (1999) for example), which neither use the full information in the model nor are they the most powerful univariate tests. The tests derived in this paper provide a natural test of the basic identification assumption of the Blanchard and Quah identification scheme.

By way of illustration we apply the tests derived here and other common tests to the Blanchard-Quah dataset. The data is quarterly data on income and unemployment for the US from 1950:2 to 1987:4, where unemployment is the stationary variable x_t and income is the y_t variable. Table 10 applies the various tests to this data - the univariate tests are the frequently applied augmented Dickey and Fuller (1979) (DF) test, the DF-GLS test of Elliott et. al. (1996) and the test statistic derived here. We include constants and time trends in both unemployment and income[6] so the tests are from case 5. Results are presented for lags from 1 to 8. Except for very short lag lengths (which are most likely too short and hence the tests are not correctly sized), the DF test does not reject - it is not close to the 5% critical value. The DF-GLS test similarly does not come close to rejecting. The $\tilde{\Lambda}^5(1,\bar{r})$ test rejects at 7 lags, although is close for a few other lag lengths. Overall we would probably still conclude that it fails to reject, although we would be worried if the seven lag model is relevant (Blanchard and Quah used eight lags).

6. Conclusion.

Typically in economics correlation between the variables is the rule rather than the exception. Often these are implied by theory. Either way, this information can be extremely valuable in testing assumptions that are ancillary to the modeling process. This appears to be especially true in the case of testing for a unit root. Hansen (1995) showed this with tests he developed based around the statistic of Dickey and Fuller (1979). In a related paper Horvath and Watson (1995) showed that power gains are available when there are known cointegrating relationships (which are then stationary variables). We have shown here that even greater gains are possible. The statistics are simple to implement and yield extremely large gains in power when the covariates are relevant.

---

[6] Blanchard and Quah included a time trend in unemployment on the grounds that it was increasing over the sample. They had the equivalent of a time trend with a break for the oil shocks in income. We do not include a 'known' break such as this, however not including the break if it were truly there (tests which search for such a break typically fail to reject the hypothesis of no break) biases us away from rejecting the unit root.

The statistics we generate, useful in many areas, are directly applicable to testing the unit root assumption in the identification of structural VAR's from long run restrictions. These restrictions do not make sense unless there is a process with a unit root in the model, yet typically very low power tests are used to examine this assumption. The tests derived here will have much better power at detecting the mistaken use of this procedure.

References.

Blanchard, O. and D. Quah (1989), The Dynamic effect of Aggregate Demand and Aggregate Supply Disturbances" *American Economic Review*, 79, pp655-73.

Caporale, G.M. and N. Pittis (1999), "Unit Root Testing Using Covariates: Some Theory and Evidence", *Oxford Bulletin of Economic Studies*, 61, pp583-95.

Chan, N.H. and C.Z.Wei (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes", *Annals of Statistics*, 16, pp 367-401.

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., J.H. Stock and T.J. Rothenberg (1996). "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, pp813-836.

Horvath, M.T.K and M.W. Watson (1995), "Testing for Cointegration when Some of the Cointegrating Vectors are Known", *Econometric Theory*, 5, pp 952-983.

Gali, J. (1999), "Technology, Employment and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?", *American Economic Review*, 89, 249-271.

Hamilton, J.D. (1994) Time Series Analysis, Princeton University Press, Princeton, N.J.

Hansen, B. E. (1995), "Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power", *Econometric Theory*, 11 (5) 1148-1172.

Johansen, S. (1989), "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*, 12, 231-54.

King, M.L. (1980), "Robust Tests for Spherical Symmetry and their Application to Least Squares Regression", *Annals of Statistics*, 8, 1265-1271.

King, M.L. (1988), "Towards a Theory of Point Optimal Testing", *Econometric Reviews*, 6, 169,218.

King, R.G, Plosser, C.I., Stock, J.H. and Watson, M.W. (1991), "Stochastic Trends and Economic Fluctuations", *American Economic Review*, 81, 819-40.

Park, J.Y and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory*, 4, pp468-97.

Phillips, P.C.B. (1987), "Towards a Unified Asymptotic Theory for Autoregression", *Biometrica*, 74, 535-547.

Appendix.

Lemma 1. Distribution results.

Under the Assumptions of the model in (1) and (2) with A1, A2 and A3 we have

a) $T^{-1/2}u_{y[T\cdot]} \Rightarrow w_{yy}^{1/2}W_{1c}(\cdot)$

b) $\frac{1}{Tw_{yy}^{1/2}}\sum_{t=2}^{T}u_{yt-1}\left(\Sigma^{-1/2}e_t(r)\right)' \Rightarrow \int W_{1c}(l)d\left[W_1(l)\ \ V(l)'\right]$

where $\bar{d}'V(l) = \sqrt{\frac{R^2}{1-R^2}}W_2(l)$, $\bar{d}' = w_{yy}^{-1/2}w_{yx}\Omega_{x.y}^{-1/2}$, $\Omega_{x.y} = \Omega_{xx} - w_{yx}'w_{yx}w_{yy}^{-1}$,

$W(l) = \begin{bmatrix} W_1(l) \\ W_2(l) \end{bmatrix}$ *are univariate independent standard Brownian Motions on C[0,1] and*

$W_{1c}(l) = c\int_0^l e^{c(l-s)}W_1(s)ds + W_1(l)$.

Proof: (a) follows as $u_{y,t} = ru_{y,t-1} + v_t$ where $v_t = s_1A(L)^{-1}e_t(r)$. The partial sum $T^{-1/2}\sum_1^{[T.]}v_s \Rightarrow s_1\Omega^{1/2}\begin{pmatrix} W_1(.) \\ V(.) \end{pmatrix} = w_{yy}^{1/2}W_1(.)$ where s_1 = [1 0] is an 1xm+1 vector with partition after the first column. The result then follows setting ρ =1+c/T from Phillips (1987). Part (b) follows from Chan and Wei (1988), Park and Phillips (1988). The relationship between V(λ) and W_2(λ) follows from the relation $\bar{d}'\bar{d} = \frac{R^2}{1-R^2}$.

Proof of Theorems 1 and 2.

The proof for Theorem 1 is a special case of that for Theorem 2 where terms relating to the deterministics are zero, so we proceed in the general case. Throughout we use r for results general for ρ , $\bar{r}$ and 1.

First, define $\hat{u}_t^i(r) = z_t(r) - d_t(r)\hat{b}^i(r) = e_t(r) - d_t(r)'\left(\hat{b}^i(r) - b\right)$, and $e_t(r) = A(L)u_t(r)$.

From the algebra of GLS

$$\sum_{t=1}^{T}\hat{u}_t^i(r)'\Sigma^{-1}\hat{u}_t^i(r) = \sum_{t=1}^{T}e_t(r)'\Sigma^{-1}e_t(r) - \left(S_iN_T(r)\right)'\left(S_iD_T(r)S_i\right)^{-}\left(S_iN_T(r)\right)$$

where

$$N_T(r) = \Psi_T^{-1}\left(\sum_{t=1}^{T} d_t(r)\Sigma^{-1}e_t(r)\right)$$
$$D_T(r) = \Psi_T^{-1}\left(\sum_{t=1}^{T} d_t(r)\Sigma^{-1}d_t(r)'\right)\Psi_T^{-1}{}'$$

and

$$\Psi_T = \begin{pmatrix} w_{yy}^{-1/2} & 0 & 0 & 0 \\ 0 & T^{1/2}\Omega_{x.y}^{-1/2}{}' & 0 & 0 \\ 0 & 0 & T^{1/2}w_{yy}^{-1/2} & 0 \\ 0 & 0 & 0 & T^{3/2}\Omega_{x.y}^{-1/2}{}' \end{pmatrix}$$

Thus,

$$\Lambda^i(1,\bar{r}) = \sum_{t=1}^{T} e_t(\bar{r})'\Sigma^{-1}e_t(\bar{r}) - \sum_{t=1}^{T} e_t(1)'\Sigma^{-1}e_t(1)$$
$$+\left(S_i N_T(1)\right)'\left(S_i D_T(1) S_i\right)^{-}\left(S_i N_T(1)\right) - \left(S_i N_T(\bar{r})\right)'\left(S_i D_T(\bar{r}) S_i\right)^{-}\left(S_i N_T(\bar{r})\right) \qquad \text{(A1)}$$

Notice that for t>1

$$\Sigma^{-1/2}e_t(r) = e_t + (r - r)\Sigma^{-1/2}s_1'u_{y,t-1} \qquad \text{(A2)}$$

(and is ε_1 for t=1) where $e_t = \Sigma^{1/2}e_t$. Using the results $s_1\Sigma^{-1}s_1' = (1+\bar{d}'\bar{d})w_{yy}^{-1}$ and $s_1\Sigma^{-1/2}{}' = w_{yy}^{-1/2}[1 \quad -\bar{d}']'$ then in case 1 where S_i=0 we have

$$\Lambda^1(1,\bar{r}) = \sum_{t=1}^{T} e_t(\bar{r})'\Sigma^{-1}e_t(\bar{r}) - \sum_{t=1}^{T} e_t(1)'\Sigma^{-1}e_t(1)$$
$$= (\bar{c}^2 - 2c\bar{c})(1+\bar{d}'\bar{d})w_{yy}^{-1}\frac{1}{T^2}\sum_{t=1}^{T} u_{y,t-1}^2$$
$$-2\bar{c}\frac{1}{T}\sum_{t=1}^{T}\left[u_{y,t-1}w_{yy}^{-1/2}[1 \quad -\bar{d}']e_t(r)\right]$$

From the limit results in lemma 1

$$\Lambda^1(1,\bar{r}) \Rightarrow (\bar{c}^2 - 2c\bar{c})\left(\frac{R^2}{1-R^2}\right)\int W_{1c}(l)^2\,dl - 2\bar{c}\left[\int W_{1c}(l)dW_1(l) - \frac{R}{\sqrt{1-R^2}}\int W_{1c}(l)dW_2(l)\right]$$

as stated in Theorem 1.

For the other cases, extra terms arise from the final two terms in equation (A1). Defining c_r=T(r-1) we have

$$\lim_{T\to\infty}\left\|\left(\Psi_T^{-1}d_1(r)\Sigma^{-1/2}{}'\right)-\begin{pmatrix}1 & -\bar{d}'\\ 0 & 0\\ 0 & 0\\ 0 & 0\end{pmatrix}\right\|=0$$

and

$$\lim_{T\to\infty}\left(\sup_{2/T\le s\le 1}\left\|\left(T^{1/2}\Psi_T^{-1}d_{[Ts]}(r)\Sigma^{-1/2}{}'\right)-\begin{pmatrix}0 & 0\\ 0 & I_m\\ 1-c_r s & -(1-c_r s)\bar{d}\\ 0 & sI_m\end{pmatrix}\right\|\right)=0$$

Using these two results and the continuous mapping theorem $\left(S_i D_T(r)S_i\right)^- \to \left(S_i D(c_r,\bar{d})S_i\right)^-$ where

$$D(c_r,\bar{d})=\begin{pmatrix}1+\bar{d}'\bar{d} & 0 & 0 & 0\\ 0 & I_m & -\left(1-\frac{c_r}{2}\right)\bar{d} & \frac{1}{2}I_m\\ 0 & -\left(1-\frac{c_r}{2}\right)\bar{d}' & (1+\frac{c_r}{3}-c_r)(1+\bar{d}'\bar{d}) & -\left(\frac{1}{2}-\frac{c_r}{3}\right)\bar{d}'\\ 0 & \frac{1}{2}I_m & -\left(\frac{1}{2}-\frac{c_r}{3}\right)\bar{d} & \frac{1}{3}I_m\end{pmatrix}$$

Using the continuous mapping theorem, equation (A2) and results from lemma 1 we have $N_T(r)\Rightarrow N(c,c_r,\bar{d})$ where

$$N(c,c_r,\bar{d})=\begin{pmatrix}e_{y,1}-\bar{d}'e_{x,1}\\ V(1)-(c-c_r)\bar{d}\int W_{1c}(s)ds\\ \int(1-c_r s)d\left[W_1(s)-\bar{d}'V(s)\right]+(c-c_r)(1+\bar{d}'\bar{d})\int(1-c_r s)W_{1c}(s)ds\\ \int sdV(s)-(c-c_r)\bar{d}\int W_{1c}(s)ds\end{pmatrix}$$

(all integrals are zero to one). Applying these results to (A1) yields

$$\begin{aligned}\Lambda^i(1,\bar{r})\Rightarrow y^1(c,\bar{c},R^2)&+\left(S_iN(c,0,\bar{d})\right)'\left(S_iD(0,\bar{d})S_i\right)^-\left(S_iN(c,0,\bar{d})\right)\\ &-\left(S_iN(c,\bar{c},\bar{d})\right)'\left(S_iD(\bar{c},\bar{d})S_i\right)^-\left(S_iN(c,\bar{c},\bar{d})\right)\end{aligned}$$

The individual results follow by using the relevant S_i and rearranging.

In case 2, we have

$$\left(S_iN(c,c_r,\bar{d})\right)'\left(S_iD(c_r,\bar{d})S_i\right)^-\left(S_iN(c,c_r,\bar{d})\right)=(1+\bar{d}'\bar{d})^{-1}(e_{y,1}-\bar{d}'e_{x,1})^2$$

thus the terms offset giving the result in the Theorem.

In case 3, we have

$$\left(S_3 D(c_r,\bar{d})S_3\right)^- = \begin{pmatrix} (1+\bar{d}'\bar{d})^{-1} & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so

$$\left(S_i N(c,c_r,\bar{d})\right)'\left(S_i D(c_r,\bar{d})S_i\right)^-\left(S_i N(c,c_r,\bar{d})\right) = (1+\bar{d}'\bar{d})^{-1}(e_{y,1}-\bar{d}'e_{x,1})^2 + V(1)'V(1)$$
$$+(c-c_r)^2\bar{d}'\bar{d}\left(\int W_{1c}\right)^2 - 2(c-c_r)\bar{d}'V(1)\int W_{1c}$$

Plugging in 0 and $\bar{c}$ for c_r and taking the difference yields the result.

Case 4.

Here $\left(S_4 D(c_r,\bar{d})S_4\right)^- = \left(S_3 D(c_r,\bar{d})S_3\right)^- + \frac{1}{h(r)}\begin{pmatrix} 0 \\ (1-\frac{c_r}{2})\bar{d} \\ 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 \\ (1-\frac{c_r}{2})\bar{d} \\ 1 \\ 0 \end{pmatrix}'$

Where $h(r) = 1+\frac{c_r^2}{3}-c_r+\frac{c_r^2}{12}\frac{R^2}{1-R^2}$. The result follows after some rearrangement.

Case 5.

Here $\left(S_5 D(c_r,\bar{d})S_5\right)^- = \begin{pmatrix} (1+\bar{d}'\bar{d})^{-1} & 0 & 0 & 0 \\ 0 & I_m & 0 & \frac{1}{2}I_m \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2}I_m & 0 & \frac{1}{3}I_m \end{pmatrix} + \frac{1}{a(r)}\begin{pmatrix} 0 \\ \bar{d} \\ 1 \\ -c_r\bar{d} \end{pmatrix}\begin{pmatrix} 0 \\ \bar{d} \\ 1 \\ -c_r\bar{d} \end{pmatrix}'$

Where $a(r) = 1+\frac{c_r^2}{3}-c_r$.

We have

$$N(c,c_r,\bar{d})'\begin{pmatrix}0 & 0 & 0 & 0\\ 0 & 4I_m & 0 & -6I_m\\ 0 & 0 & 0 & 0\\ 0 & -6I_m & 0 & 12I_m\end{pmatrix}N(c,c_r,\bar{d}) = \begin{pmatrix}\int dV(s)\\ \int sdV(s)\end{pmatrix}'\begin{pmatrix}4I_m & -6I_m\\ -6I_m & 12I_m\end{pmatrix}\begin{pmatrix}\int dV(s)\\ \int sdV(s)\end{pmatrix}$$

$$+(c-c_r)^2\bar{d}'\bar{d}\begin{pmatrix}\int W_{1c}\\ \int sW_{1c})\end{pmatrix}'\begin{pmatrix}4 & -6\\ -6 & 12\end{pmatrix}\begin{pmatrix}\int W_{1c}\\ \int sW_{1c})\end{pmatrix}$$

$$-2(c-c_r)\begin{pmatrix}\int W_{1c}\\ \int sW_{1c})\end{pmatrix}'\begin{pmatrix}4 & -6\\ -6 & 12\end{pmatrix}\begin{pmatrix}\int d\bar{d}'V(s)\\ \int sd\bar{d}'V(s)\end{pmatrix}$$

and also

$$\begin{pmatrix}0\\ \bar{d}\\ 1\\ -c_r\bar{d}\end{pmatrix}'N(c,c_r,\bar{d}) = (1-c_r)W_{1c}(1) + c_r^2\int sW_{1c}$$

The result follows from straightforward algebra.

Proof of Theorem 3.

First, note that

$$\tilde{\Lambda}^i(1,\bar{r}) = T\left(tr\left[\tilde{\Sigma}(1)^{-1}\left(\tilde{\Sigma}(\bar{r})-\tilde{\Sigma}(1)\right)\right]\right)-\bar{c}$$

so we need to show that $T\left(tr\left[\tilde{\Sigma}(1)^{-1}\left(\tilde{\Sigma}(\bar{r})-\tilde{\Sigma}(1)\right)\right]\right)\Rightarrow y^{\,i}(c,\bar{c},R^2)$. To show this we will show

(a) $\sum_{t=k+1}^{T}\tilde{e}_t^{\,i}(\bar{r})\tilde{e}_t^{\,i}(\bar{r})' - \sum_{t=k+1}^{T}\tilde{e}_t^{\,i}(1)\tilde{e}_t^{\,i}(1)' = \sum_{t=k+1}^{T}\hat{e}_t^{\,i}(\bar{r})\hat{e}_t^{\,i}(\bar{r})' - \sum_{t=k+1}^{T}\hat{e}_t^{\,i}(1)\hat{e}_t^{\,i}(1)' + o_p(1)$

where $\hat{e}_t^i(r) = A(L)\tilde{u}_t^{\,i}(r)$.

(b) $\sum_{t=k+1}^{T}\hat{e}_t^i(r)'\Sigma^{-1}\hat{e}_t^i(r) - \sum_{t=k+1}^{T}e_t(r)'\Sigma^{-1}e_t(r) \Rightarrow -\left(S_iN(c,c_r,\bar{d})\right)'\left(S_iD(c_r,\bar{d})S_i\right)^{-}\left(S_iN(c,c_r,\bar{d})\right)$

(c) $\sum_{t=k+1}^{T}e_t(\bar{r})'\Sigma^{-1}e_t(\bar{r}) - \sum_{t=k+1}^{T}e_t(1)'\Sigma^{-1}e_t(1) \Rightarrow y^{\,1}(c,c_r,R^2)$

We take part (b) first.

We have

$$\hat{e}_t^i(r) = A(L)\left[z_t(r) - d_t(r)'\tilde{b}^i(r)\right]$$
$$= e_t(r) - A(L)d_t(r)'S_i\left(S_i\sum d_t(r)\tilde{\Omega}^{-1}d_t(r)'S_i\right)^{-}\left(S_i\sum d_t(r)\tilde{\Omega}^{-1}u_t(r)\right)$$

so

$$\sum_{t=k+1}^{T}\hat{e}_t^i(r)'\Sigma^{-1}\hat{e}_t^i(r) = \sum_{t=k+1}^{T}e_t(r)'\Sigma^{-1}e_t(r)$$
$$+\left(S_iN_T(r)\right)'\left(S_iD_T(r)S_i\right)^{-}\left(S_i\Psi_T^{-1}\sum\left[A(L)d_t(r)'\right]'\Sigma^{-1}\left[A(L)d_t(r)'\right]\Psi_T^{-1}{}'S_i\right)\left(S_iD_T(r)S_i\right)^{-}\left(S_iN_T(r)\right)$$
$$-2\left(S_iN_T(r)\right)'\left(S_iD_T(r)S_i\right)^{-}\left(S_i\Psi_T^{-1}\sum\left[A(L)d_t(r)'\right]'\Sigma^{-1}e_t(r)\right)+o_p(1)$$

where $N_T(r)$ is defined as before replacing $e_t(r)$ is replaced by $u_t(r)$ and Σ is replaced by Ω and similarly for $D_T(r)$ (these are the generalizations to $A(L)\neq I$) and the $o_p(1)$ term arises from replacing the estimated Ω with its true value.

Using the Beveridge Nelson decomposition $A(L)=A(1)+A^*(L)(1-L)$ we have

$$A(L)d_t(r)'\Psi_T^{-1} = A(1)d_t(r)'\Psi_T^{-1}{}'+A^*(1)\Delta d_t(r)\Psi_T^{-1}$$
$$= A(1)d_t(r)'\Psi_T^{-1}{}'+0(T^{-3/2})$$

so

$$S_i\Psi_T^{-1}\sum\left[A(L)d_t(r)'\right]'\Sigma^{-1}\left[A(L)d_t(r)'\right]\Psi_T^{-1}{}'S_i = S_iD_T(r)S_i + o(1)$$

and also

$$\Psi_T^{-1}\sum d_t(r)A(1)'\Sigma^{-1}e_t(r) = \Psi_T^{-1}\sum d_t(r)A(1)'\Sigma^{-1}A(L)u_t(r)$$
$$= \Psi_T^{-1}\sum d_t(r)\Omega^{-1}u_t(r) + \Psi_T^{-1}\sum d_t(r)A(1)'\Sigma^{-1}A^*(L)\Delta u_t(r)$$
$$= \Psi_T^{-1}\sum d_t(r)\Omega^{-1}u_t(r) + o_p(1)$$

This gives the result

$$\sum_{t=k+1}^{T}\hat{e}_t^i(r)'\Sigma^{-1}\hat{e}_t^i(r) = \sum_{t=k+1}^{T}e_t(r)'\Sigma^{-1}e_t(r) - \left(S_iN_T(r)\right)'\left(S_iD_T(r)S_i\right)^{-}\left(S_iN_T(r)\right)+o_p(1)$$

Finally, following steps analogous to those in the proof of Theorem 2 we have that

$$\left(S_iN_T(r)\right)'\left(S_iD_T(r)S_i\right)^{-}\left(S_iN_T(r)\right)\Rightarrow\left(S_iN(c,c_r,\bar{d})\right)'\left(S_iD(c_r,\bar{d})S_i\right)^{-}\left(S_iN(c,c_r,\bar{d})\right).$$

Part (c) follows from noting that

$$\Sigma^{-1/2}e_t(r) = e_t + (r - r)\Sigma^{-1/2}A(L)s_1'u_{y,t-1}$$

so using the Beveridge Nelson decomposition and results above

$$\sum e_t(r)'\Sigma^{-1}e_t(r) = \sum e_t'e_t + (r-r)^2 s_1\Omega^{-1}s_1'\sum u_{y,t-1}^2 + 2(r-r)\sum u_{y,t-1}s_1\Omega^{-1/2}e_t$$

Thus

$$\sum_{t=1}^{T} e_t(\bar{r})'\Sigma^{-1}e_t(\bar{r}) - \sum_{t=1}^{T} e_t(1)'\Sigma^{-1}e_t(1) = (\bar{c}^2 - 2c\bar{c})(1+\bar{d}'\bar{d})w_{yy}^{-1}\frac{1}{T^2}\sum_{t=1}^{T} u_{y,t-1}^2$$
$$-2\bar{c}\frac{1}{T}\sum_{t=1}^{T}\left[u_{y,t-1}w_{yy}^{-1/2}[1\ -\bar{d}']e_t(r)\right]$$

Applying the convergence results in lemma 1 completes the result.

Finally, it remains only to show part (a), that estimating the VAR coefficients assuming the largest root for y_t is r does not matter asymptotically.

We have that

$$\begin{aligned}\tilde{e}_t^i(r) &= \tilde{A}(L,r)\tilde{u}_t^i(r)\\ &= \hat{e}_t^i(r) - \left(\sum U_{t-1}(r)\hat{e}_{t-1}(r)'\right)\left(\sum U_{t-1}(r)U_{t-1}(r)'\right)^{-1}U_{t-1}(r)\end{aligned}$$

where $U_{t-1}(r) = \left[\tilde{u}_{t-1}^i(r)'\ \ \tilde{u}_{t-1}^i(r)'\ \cdots\ \cdots\ \tilde{u}_{t-k}^i(r)'\right]'$

(i.e. the regressors in the VAR to be run). Note that

$$U_{t-1}(r) = \begin{bmatrix}\tilde{u}_{t-1}^i(r)\\ \vdots\\ \vdots\\ \vdots\\ \tilde{u}_{t-k}^i(r)\end{bmatrix} = \begin{bmatrix}\tilde{u}_{t-1}^i(r)\\ \vdots\\ \vdots\\ \vdots\\ \tilde{u}_{t-1}^i(r)\end{bmatrix} + \begin{bmatrix}\begin{pmatrix}(r-r)\tilde{y}_{t-2}^i\\ 0\end{pmatrix}\\ \vdots\\ \vdots\\ \vdots\\ \begin{pmatrix}(r-r)\tilde{y}_{t-k-1}^i\\ 0\end{pmatrix}\end{bmatrix} = U_{t-1}(r) + (r-r)V_y$$

where $\tilde{y}_t^i = y_t - s_1 d_t'\tilde{b}(r)$ (i.e. y_t detrended under the hypothesis that $\rho = r$).

Now,

$$\sum_{t=k+1}^{T}\tilde{e}_t^i(r)\tilde{e}_t^i(r)' = \sum_{t=k+1}^{T}\hat{e}_t^i(r)\hat{e}_t^i(r)'$$
$$-\left(T^{-1/2}\sum U_{t-1}(r)\hat{e}_t^i(r)\right)'\left(T^{-1}\sum U_{t-1}(r)U_{t-1}(r)'\right)^{-1}\left(T^{-1/2}\sum U_{t-1}(r)\hat{e}_t^i(r)\right)$$

and

$$\left(T^{-1}\sum U_{t-1}(r)U_{t-1}(r)'\right)=\left(T^{-1}\sum U_{t-1}(r)U_{t-1}(r)'\right)+T^{2}(r-r)^{2}T^{-3}\sum V_{y}V_{y}'$$
$$+2T(r-r)T^{-2}\sum V_{y}U_{t-1}(r)$$

The second of these terms is $o_p(1)$ as typical terms involve $T^{-3}\sum \tilde{y}_{t-i}^{2}$. These converge to zero as $T^{-1/2}\tilde{y}_t^{i}$ is $0_p(1)$. This follows as

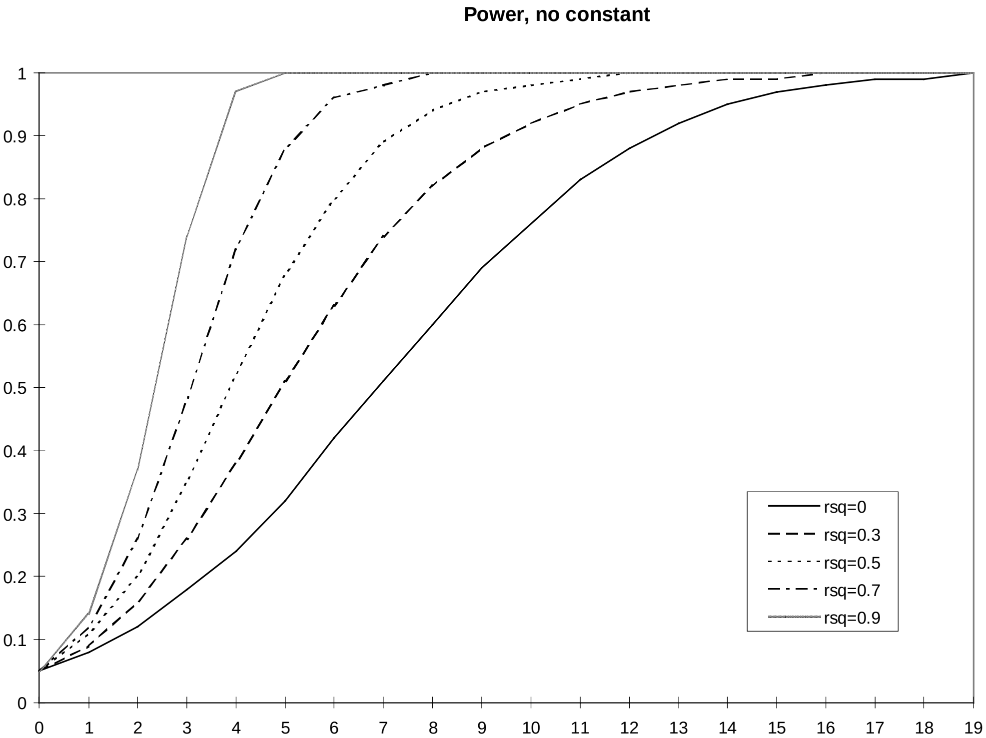
$$\begin{aligned} T^{-1/2}\tilde{y}_t^{i} &= T^{-1/2}u_{y,t-1} - s_1 T^{-1/2} d_t\left(\tilde{b}^{i} - b^{i}\right) \\ &= T^{-1/2}u_{y,t-1} - s_1 T^{-1/2} d_t \Psi_T^{-1}\left(S_i D_T(r) S_i\right)^{-}\left(S_i N_T(r)\right) \\ &= T^{-1/2}u_{y,t-1} - w_{yy}^{-1/2}(T^{-1}t)s_3\left(S_i D_T(r) S_i\right)^{-}\left(S_i N_T(r)\right) + o_p(1) \end{aligned}$$

where s_3 is (2m+2)x1 with the (m+2) element one and is zero everywhere else. Similar results follow for the cross product terms. So we have

$$\sum_{t=k+1}^{T}\tilde{e}_t^{i}(\bar{r})\tilde{e}_t^{i}(\bar{r})' - \sum_{t=k+1}^{T}\tilde{e}_t^{i}(1)\tilde{e}_t^{i}(1)' = \sum_{t=k+1}^{T}\hat{e}_t^{i}(\bar{r})\hat{e}_t^{i}(\bar{r})' - \sum_{t=k+1}^{T}\hat{e}_t^{i}(1)\hat{e}_t^{i}(1)'$$
$$-\left(T^{-1/2}\sum U_{t-1}(r)\hat{e}_t^{i}(r)\right)'\left(T^{-1}\sum U_{t-1}(r)U_{t-1}(r)'\right)^{-1}\left(T^{-1/2}\sum U_{t-1}(r)\hat{e}_t^{i}(r)\right)$$
$$+\left(T^{-1/2}\sum U_{t-1}(r)\hat{e}_t^{i}(r)\right)'\left(T^{-1}\sum U_{t-1}(r)U_{t-1}(r)'\right)^{-1}\left(T^{-1/2}\sum U_{t-1}(r)\hat{e}_t^{i}(r)\right)+o_p(1)$$

and the third and fourth terms cancel obtaining the result in (a).

Figure 1a:



Notes: For all figures the x axis is -c

Figure 1b:

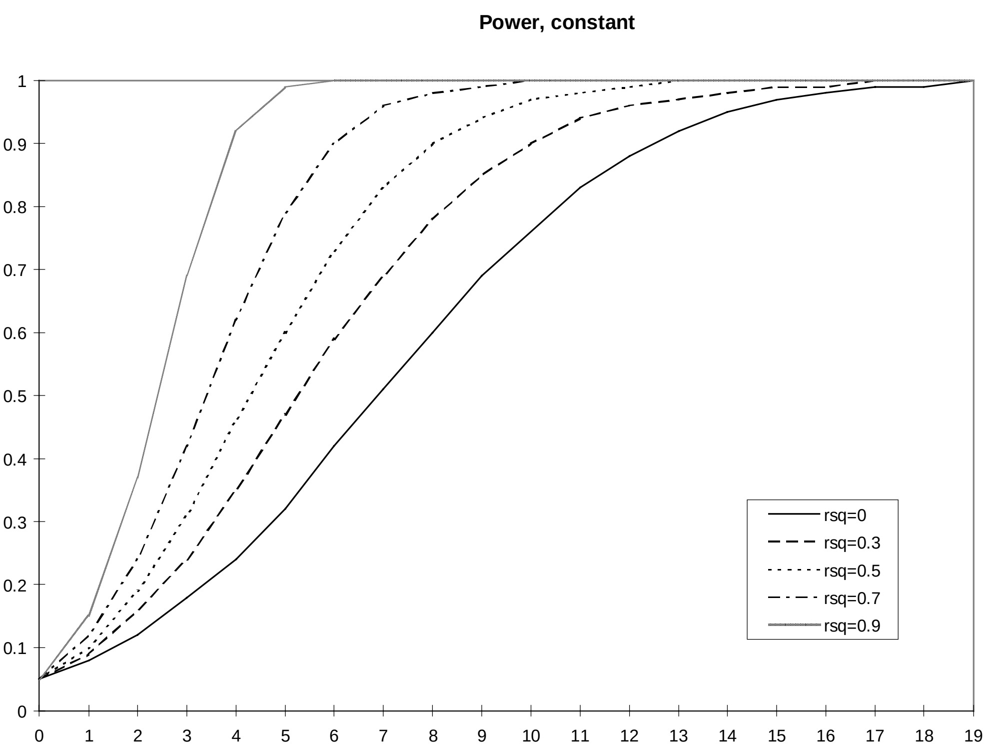


Figure 1c:

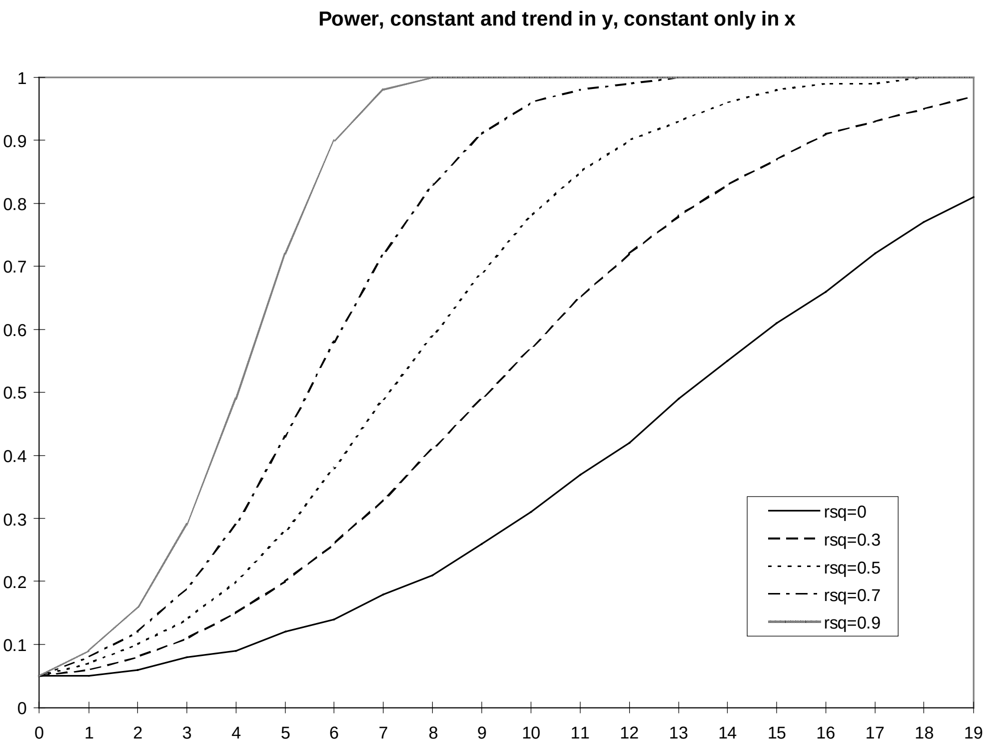


Figure 1d:

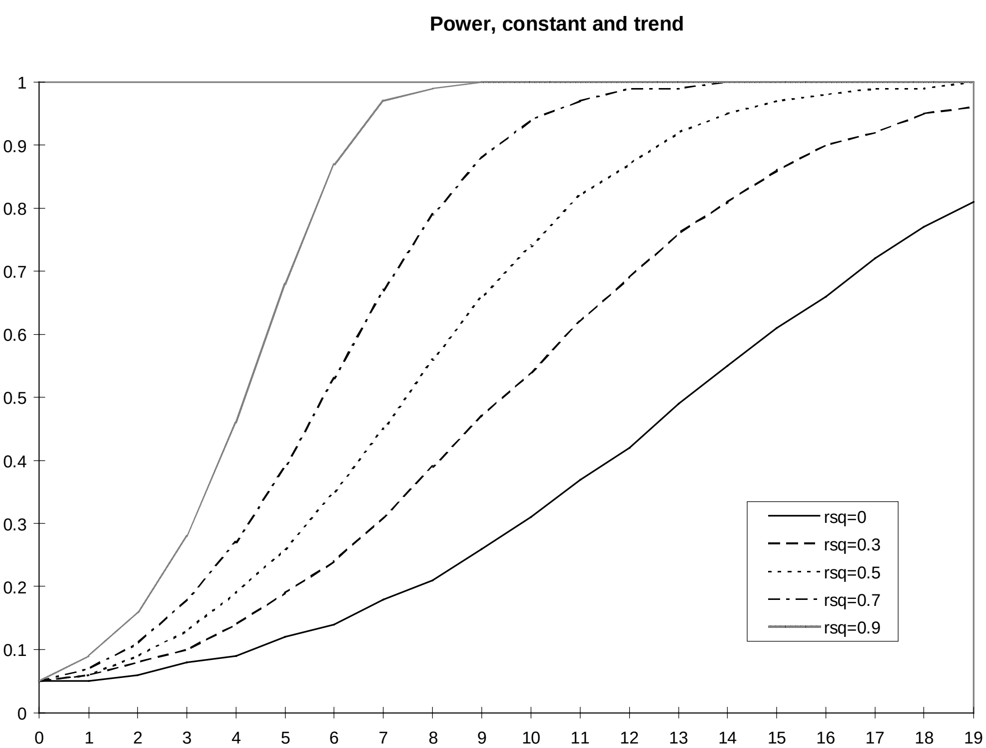


Figure 2a:

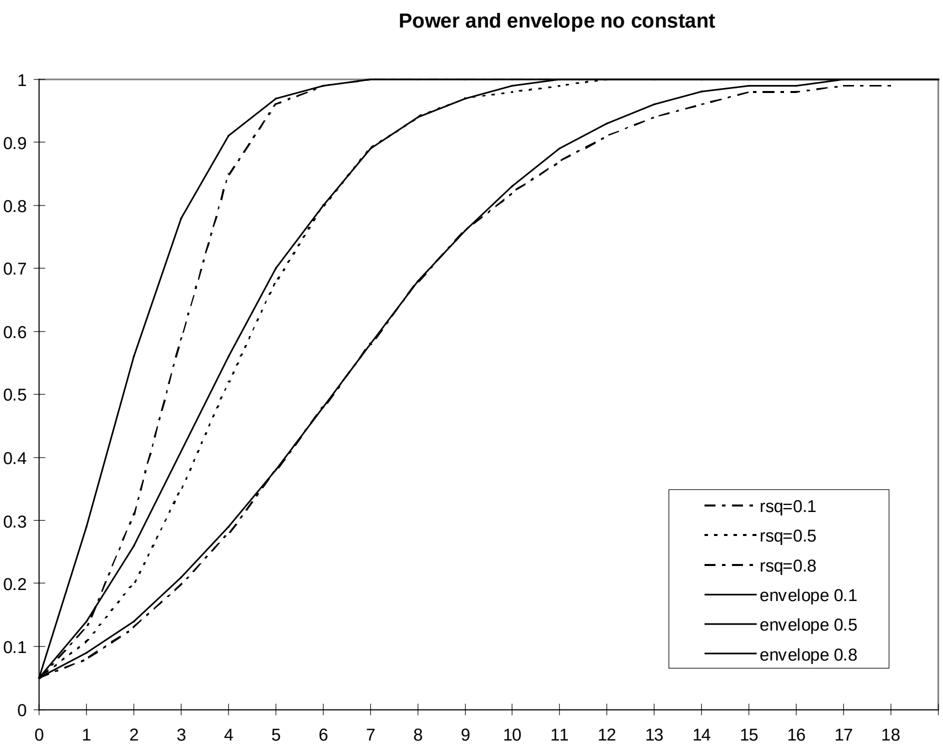


Figure 2b:

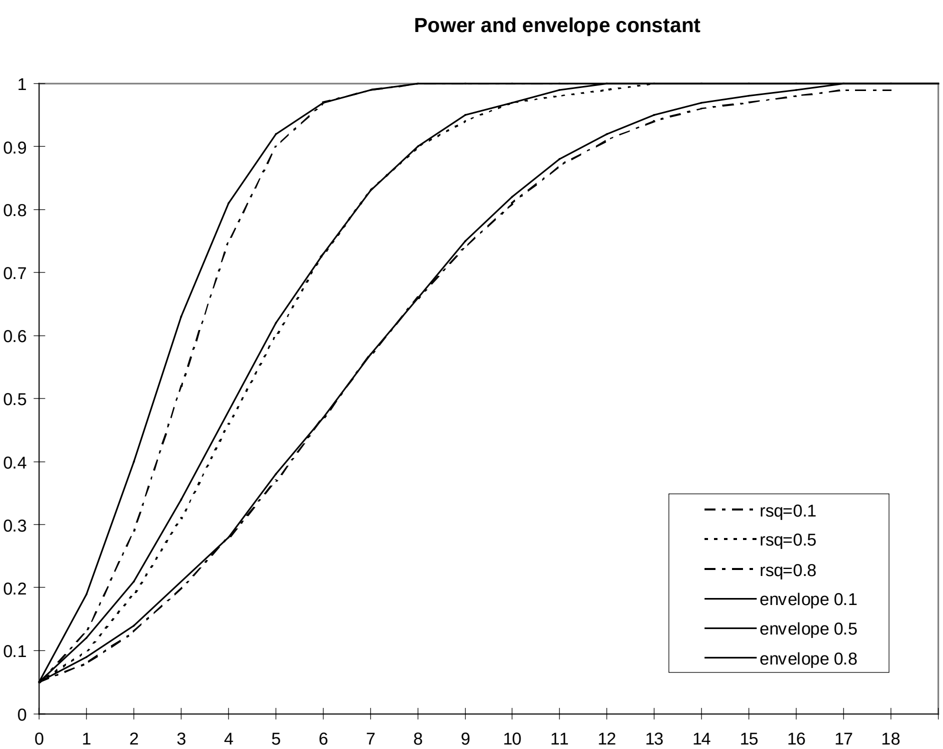


Figure 2c:

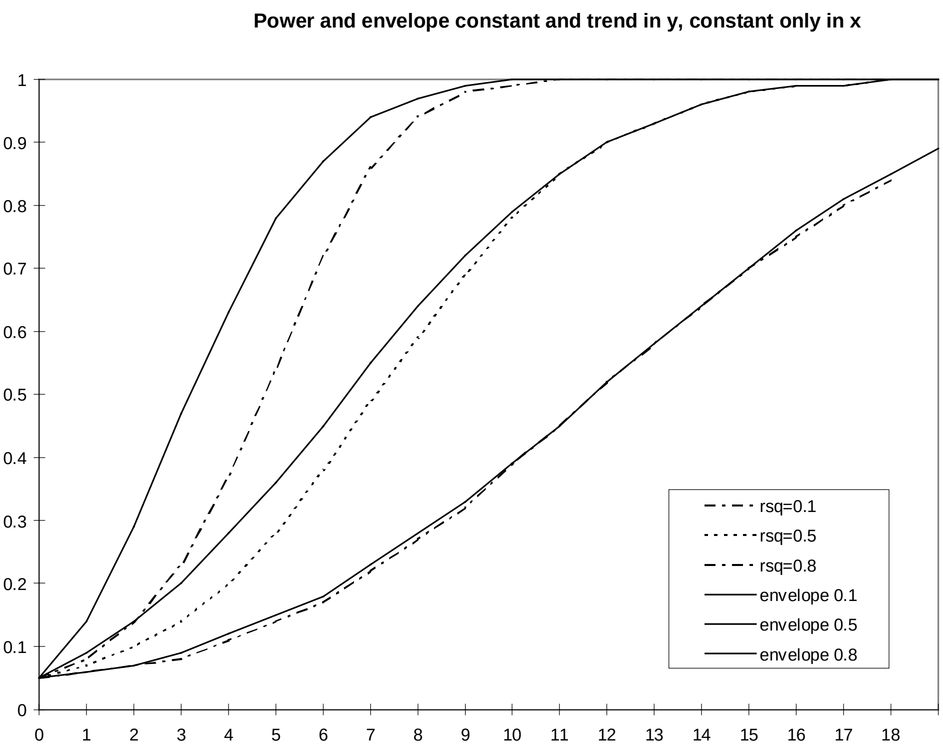


Figure 2d:

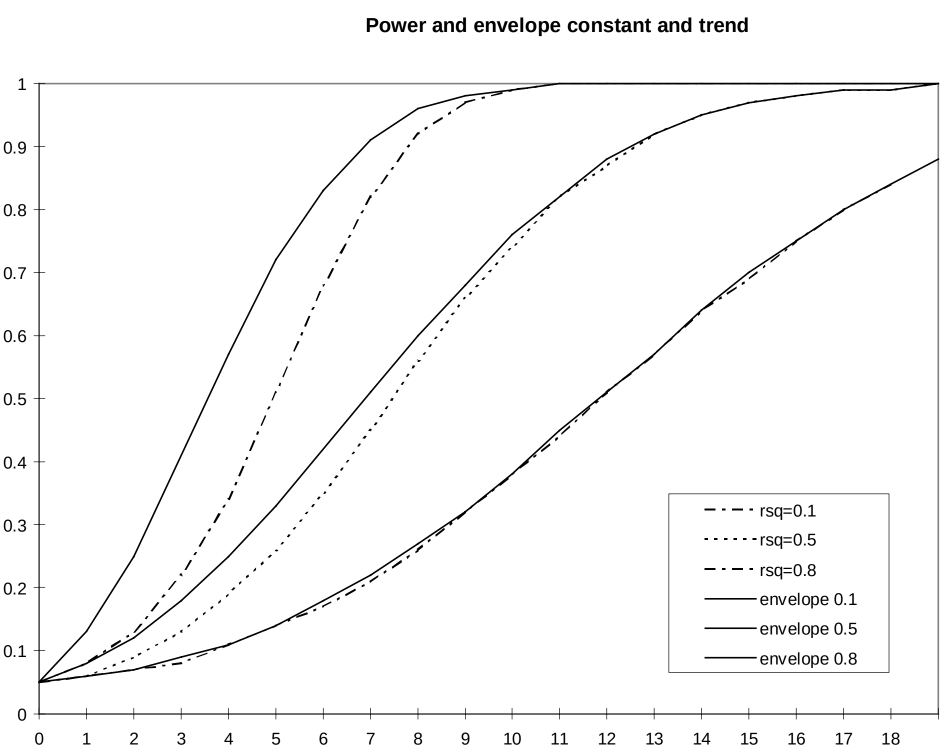


Table 1: Asymptotic Critical Values (Distribution in Theorem 3)

R^2	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Cases 1,2	3.34	3.41	3.54	3.76	4.15	4.79	5.88	7.84	12.12	25.69
Case 3	3.34	3.41	3.54	3.70	3.96	4.41	5.12	6.37	9.17	17.99
Case 4	5.70	5.79	5.98	6.38	6.99	7.97	9.63	12.6	19.03	39.62
Case 5	5.70	5.77	6.00	6.40	7.07	8.15	10.00	13.36	20.35	41.87

Notes: Critical values were computed using 1500 steps as approximations to the Brownian Motion terms in the limit theorem representations and 60000 replications. The critical values reported are for tests of size 5% with $\bar{c}$ =-7 for cases 1, 2 and 3 and $\bar{c}$ = -13.5 for cases 4 and 5.

Table 2: Small Sample results - No Deterministics (case 1)

	DF	PT	$\tilde{\Lambda}^1(1,\bar{r})$				
δ=	0	0	0	0.3	0.5	0.7	0.9
R^2 =	0	0	0	0.09	0.25	0.49	0.81
r							
1	0.05	0.048	0.051	0.049	0.05	0.05	0.044
0.98	0.117	0.113	0.119	0.132	0.153	0.195	0.306
0.96	0.237	0.229	0.239	0.276	0.342	0.493	0.848
0.94	0.407	0.396	0.407	0.463	0.576	0.782	0.992
0.92	0.594	0.581	0.59	0.655	0.774	0.926	0.999
0.9	0.758	0.744	0.748	0.807	0.896	0.977	1
0.88	0.878	0.865	0.867	0.905	0.954	0.993	1
0.86	0.947	0.939	0.936	0.957	0.981	0.998	1

Notes: Based on 20000 replications of the model with T=100, normal errors as discussed in the text. The system test is implemented with R^2 estimated.

Table 3: Small Sample results - Constant Included (case 3)

	DF	PT	$\tilde{\Lambda}^3(1,\bar{r})$				
δ=	0	0	0	0.3	0.5	0.7	0.9
R^2 =	0	0	0	0.09	0.25	0.49	0.81
r							
1	0.054	0.059	0.064	0.061	0.06	0.054	0.039
0.98	0.075	0.138	0.145	0.154	0.167	0.192	0.254
0.96	0.105	0.273	0.285	0.308	0.355	0.445	0.716
0.94	0.159	0.453	0.466	0.499	0.572	0.709	0.946
0.92	0.235	0.64	0.648	0.685	0.759	0.875	0.991
0.9	0.332	0.795	0.797	0.825	0.879	0.951	0.998
0.88	0.448	0.899	0.897	0.914	0.943	0.981	1
0.86	0.573	0.956	0.951	0.959	0.974	0.992	1

Notes: As per Table 2 with a constant included.

Table 4: Small Sample results - Constant Included in both, Time in Y regression (case 4)

	DF	PT	$\tilde{\Lambda}^4(1,\bar{r})$				
$\delta =$	0	0	0	0.3	0.5	0.7	0.9
$R^2 =$	0	0	0	0.09	0.25	0.49	0.81
r							
1	0.057	0.039	0.053	0.054	0.054	0.049	0.024
0.98	0.062	0.049	0.065	0.071	0.082	0.096	0.094
0.96	0.078	0.076	0.099	0.115	0.142	0.192	0.305
0.94	0.106	0.119	0.152	0.179	0.239	0.356	0.663
0.92	0.147	0.184	0.227	0.274	0.368	0.559	0.906
0.9	0.204	0.27	0.325	0.389	0.518	0.744	0.981
0.88	0.277	0.377	0.442	0.519	0.663	0.868	0.995
0.86	0.365	0.503	0.564	0.646	0.783	0.937	0.999

Notes: As per Table 2 with a constant included in both regressions and a time trend in the y_t regression (for the $\tilde{\Lambda}^4(1,\bar{r})$ statistic) and a constant and time trend included in the univariate unit root tests.

Table 5: Small Sample results - Constant and Time Included (case 5)

	DF	PT	$\tilde{\Lambda}^5(1,\bar{r})$				
$\delta =$	0	0	0	0.3	0.5	0.7	0.9
$R^2 =$	0	0	0	0.09	0.25	0.49	0.81
r							
1	0.057	0.039	0.053	0.053	0.051	0.044	0.021
0.98	0.062	0.049	0.065	0.069	0.076	0.085	0.08
0.96	0.078	0.076	0.099	0.111	0.131	0.172	0.262
0.94	0.106	0.119	0.152	0.173	0.223	0.32	0.599
0.92	0.147	0.184	0.226	0.267	0.345	0.511	0.871
0.9	0.204	0.27	0.325	0.379	0.488	0.699	0.971
0.88	0.277	0.377	0.441	0.507	0.634	0.834	0.993
0.86	0.365	0.503	0.564	0.635	0.758	0.919	0.998

Notes: As per Table 2 with a constant and time trend included.

Table 6: Effect of estimating R^2 on test using $\tilde{\Lambda}^3(1, \bar{r})$

	R² known					Estimated R²				
δ =	0	0.3	0.5	0.7	0.9	0	0.3	0.5	0.7	0.9
R^2 =	0	0.09	0.25	0.49	0.81	0	0.09	0.25	0.49	0.81
r										
1	0.063	0.06	0.061	0.056	0.053	0.064	0.061	0.06	0.054	0.039
0.98	0.144	0.152	0.167	0.193	0.29	0.145	0.154	0.167	0.192	0.254
0.96	0.283	0.305	0.356	0.45	0.758	0.285	0.308	0.355	0.445	0.716
0.94	0.465	0.497	0.573	0.716	0.967	0.466	0.499	0.572	0.709	0.946
0.92	0.647	0.684	0.761	0.882	0.997	0.648	0.685	0.759	0.875	0.991
0.9	0.796	0.824	0.881	0.956	1	0.797	0.825	0.879	0.951	0.998
0.88	0.896	0.913	0.944	0.984	1	0.897	0.914	0.943	0.981	1
0.86	0.951	0.958	0.975	0.994	1	0.951	0.959	0.974	0.992	1

Notes: As per Table 3.

Table 7: Effect of estimating R^2 on test using $\tilde{\Lambda}^5(1, \bar{r})$

	R² known					Estimated R²				
δ =	0	0.3	0.5	0.7	0.9	0	0.3	0.5	0.7	0.9
R^2 =	0	0.09	0.25	0.49	0.81	0	0.09	0.25	0.49	0.81
r										
1	0.053	0.052	0.052	0.048	0.05	0.053	0.053	0.051	0.044	0.021
0.98	0.065	0.068	0.076	0.087	0.131	0.065	0.069	0.076	0.085	0.08
0.96	0.099	0.109	0.131	0.176	0.342	0.099	0.111	0.131	0.172	0.262
0.94	0.152	0.172	0.221	0.327	0.686	0.152	0.173	0.223	0.32	0.599
0.92	0.225	0.265	0.345	0.522	0.923	0.226	0.267	0.345	0.511	0.871
0.9	0.323	0.377	0.489	0.714	0.989	0.325	0.379	0.488	0.699	0.971
0.88	0.44	0.504	0.639	0.853	0.999	0.441	0.507	0.634	0.834	0.993
0.86	0.562	0.633	0.764	0.93	1	0.564	0.635	0.758	0.919	0.998

Notes: As per Table 5.

Table 8: CADF and $\tilde{\Lambda}^3(1, \bar{r})$

	CADF					$\tilde{\Lambda}^3(1, \bar{r})$				
δ =	0	0.3	0.5	0.7	0.9	0	0.3	0.5	0.7	0.9
R^2 =	0	0.09	0.25	0.49	0.81	0	0.09	0.25	0.49	0.81
r										
1	0.053	0.055	0.056	0.054	0.051	0.064	0.061	0.06	0.054	0.039
0.98	0.075	0.082	0.098	0.135	0.321	0.145	0.154	0.167	0.192	0.254
0.96	0.107	0.123	0.162	0.272	0.675	0.285	0.308	0.355	0.445	0.716
0.94	0.16	0.188	0.262	0.456	0.885	0.466	0.499	0.572	0.709	0.946
0.92	0.234	0.285	0.396	0.639	0.965	0.648	0.685	0.759	0.875	0.991
0.9	0.332	0.4	0.542	0.79	0.991	0.797	0.825	0.879	0.951	0.998
0.88	0.444	0.527	0.682	0.889	0.998	0.897	0.914	0.943	0.981	1
0.86	0.566	0.654	0.798	0.947	0.999	0.951	0.959	0.974	0.992	1

Notes: As per table 3. The CADF refers to the test procedure in Hansen (1995). In each case the same R^2 estimate is used to determine the critical value.

Table 9: CADF and $\tilde{\Lambda}^5(1, \bar{r})$

	CADF					$\tilde{\Lambda}^5(1, \bar{r})$				
δ =	0	0.3	0.5	0.7	0.9	0	0.3	0.5	0.7	0.9
R^2 =	0	0.09	0.25	0.49	0.81	0	0.09	0.25	0.49	0.81
r										
1	0.057	0.058	0.057	0.053	0.046	0.053	0.053	0.051	0.044	0.021
0.98	0.061	0.067	0.079	0.106	0.219	0.065	0.069	0.076	0.085	0.08
0.96	0.079	0.093	0.121	0.197	0.525	0.099	0.111	0.131	0.172	0.262
0.94	0.105	0.131	0.182	0.327	0.78	0.152	0.173	0.223	0.32	0.599
0.92	0.147	0.186	0.268	0.479	0.916	0.226	0.267	0.345	0.511	0.871
0.9	0.203	0.257	0.375	0.635	0.973	0.325	0.379	0.488	0.699	0.971
0.88	0.276	0.345	0.495	0.766	0.992	0.441	0.507	0.634	0.834	0.993
0.86	0.363	0.451	0.613	0.861	0.998	0.564	0.635	0.758	0.919	0.998

Notes: As per table 5. The CADF refers to the test procedure in Hansen (1995). In each case the same R^2 estimate is used to determine the critical value.

Table 10: Blanchard-Quah Model

#lags	DF	DF-GLS	$\tilde{\Lambda}^5(1, \bar{r})$	R^2	Critical Value
1	-3.06	-1.58	16.21	0.38	6.88
2	-3.80	-1.85	23.27	0.46	7.54
3	-2.87	-1.52	18.16	0.68	12.05
4	-2.59	-1.46	15.70	0.69	12.25
5	-2.34	-1.45	18.21	0.72	13.55
6	-2.34	-1.57	16.73	0.65	11.04
7	-2.32	-1.48	19.08	0.80	19.19
8	-1.78	-1.37	17.93	0.76	16.56

Notes: The Column labelled DF gives the Augmented Dickey Fuller statistic when a constant and time trend are included in the regression for the indicated lag length (the asymptotic critical value is -3.41). The column labeled DF-GLS is the Elliott. et. al. (1996) augmented Dickey Fuller statistic with GLS detrending (the critical value is -2.89). The critical values for the $\tilde{\Lambda}^5(1, \bar{r})$ statistic are in the final column (dependent on the estimated R^2 given in the fifth column).