

Testing the null of no cointegration when covariates are known to have a unit root*

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

Elena Pesavento
Emory University
1602 Fishburne Dr.
Atlanta, GA 30322

December 18, 2007

Abstract

1 Introduction

In their seminal paper Engle and Granger (1983) suggested tests for the null hypothesis of no cointegration that amounted to testing for a unit root in the estimated least squares cointegrating vector. Since this suggestion, a large number of tests have been considered for testing this null. When there are a large number of tests for a particular hypothesis, it is natural to consider which one should be employed in practice. Such considerations rest in large samples on power considerations and in smaller samples on size control and power. Such large sample power comparisons are thus interesting and have been made in the literature. Understanding which, if any, test is best is a first step, however power comparisons leave open the question of whether or not there exists some as yet unexamined test that is even better. This paper attempts to provide numerical bounds as to the maximal power of tests

*The authors

for the null of no cointegration when the cointegrating vector is not known. With these bounds we can examine both whether or not existing tests are close to the bounds (in which case it makes little sense to try and improve over existing tests on this metric) or in the case they are substantially below the upper bounds provide potential new tests that are closer to the theoretical bounds.

That there is no obviously optimal test follows directly from the result that in special cases of the test no uniformly maximum power test exists across the set of alternatives. For example when the cointegrating vector is known and the covariates of the cointegration vector (in some rotation) are strictly exogenous, the optimal test for the null of a unit root in the (known) cointegrating vector collapses to the usual unit root test for a single variable, a problem that is studied in Elliott, Rothenberg and Stock (1996) and Mueller and Elliott (2003). Since no uniformly most powerful test exists for a submodel, it is no surprise that there does not exist such a test in general. Indeed, the case of an unknown cointegrating vector pulls us further from the situation where standard generalized likelihood ratio tests — that is likelihood ratio tests where the nuisance parameters are replaced by their maximum likelihood estimates — have any known optimality properties. Since under the null hypothesis there is no cointegrating vector, as all linear combinations of the data have unit roots, then we are faced with the problem of a nuisance parameter that is not identified under the null hypothesis. It is well understood in these situations (see Davies (1977,87), Hansen (1996), and Andrews and Ploberger (1984)) that standard tests do not in general have optimality properties.

The following section examines the problem of testing for the null of no cointegration along with the notions of local power. Standard examinations for power are shown to depend on a nuisance parameter of the problem (this parameter describes the zero frequency correlation between variables) that is not estimable from the data. Hence whilst these power results are interesting, they present no clear picture as to which test statistic to employ for a particular situation. We set the problem up based on nuisance parameters that are estimable, motivating the model to be examined.

In section 3 we take up the idea of deriving the optimal tests for the null of no cointegration. For general sets of deterministic terms we show the form of the optimal test against a particular mean reverting cointegrating vector when both the scale of the variables and

the cointegrating vector is known. We assume the scale is known as it will be consistently estimable in large samples. We assume a point alternative since no test will be uniformly most powerful over a range of values. We assume the cointegrating vector is known as a step towards the unknown cointegrating vector problem. With this test, optimal tests against the alternative of cointegration when the cointegrating power is unknown take the form of a weighted average test where the weights are over the possible cointegrating vectors. When one has some information as to the likely values of the cointegrating vector, this should be employed in the test. We take the view here that we have no information, and construct a test based on a uniform distribution for the vectors. The result is a test with an asymptotic distribution that is not dependent under the null on the actual cointegrating vector (so size is controlled) and surprisingly also not dependent on the correlation structure of the shocks driving the model (making the bounds and the tests that can be constructed from these results particularly simple).

The main point of the paper was to compare the bounds with currently popular test statistics. We do this in section 4. We show that existing popular tests do not achieve power bounds for moderate local alternatives, however the difference is not great. The tests that are relatively close to the power bounds are those that rely on the systems information (the tests of Johansen (1988) based on the MLE and the ECM-Wald tests, based on the error correction model). We find that the Engle and Granger (1983) test, which is based on the estimated cointegrating vector does not exploit system information, has power far below the power bound for the experiments examined.

2 Tests of no Cointegration and Local Power

We are interested in testing the null of no cointegration against the alternative of a single cointegrating vector when the cointegrating vector is unknown to the researcher. The researcher observes an $(k + 1)$ -dimensional vector time series $z_t = (x_t', y_t)'$ generated by the model

$$x_t = \phi_x' D_t + u_{1,t} \tag{1}$$

$$y_t = \phi_c' D_t + \beta' x_t + u_{2,t} \tag{2}$$

where D_t are deterministic terms and

$$F(L) \begin{pmatrix} \Delta u_{1,t} \\ (1 - \rho L) u_{2,t} \end{pmatrix} = \varepsilon_t^*, \quad (3)$$

where y_t is univariate, x_t is of dimension $k \times 1$, $F(L) = I_{k+1} - \sum_{j=1}^p F_j L^j$ is a matrix polynomial of finite (known) order p in the lag operator L , and the following assumptions hold.

Condition 1 (*assumptions on residuals*)

(i) $\max_{-p \leq t \leq 0} \left\| (u'_{1,t}, u_{2,t})' \right\| = O_p(1)$, where $\|\cdot\|$ is the Euclidean norm.

(ii) $|F(r)| = 0$ has roots outside the unit circle.

(iii) $E_{t-1}(\varepsilon_t^*) = 0$ (a.s.), $E_{t-1}(\varepsilon_t^* \varepsilon_t^{*'}) = \Sigma^*$ (a.s.), and $\sup_t E \|\varepsilon_t^*\|^{4+\delta} < \infty$ for some $\delta > 0$, where Σ^* is positive definite, and $E_{t-1}(\cdot)$ refers to the expectation conditional on $\{\varepsilon_{t-1}^*, \varepsilon_{t-2}^*, \dots\}$.

We are interested in the problem of testing

$$H_0 : \rho = 1 \quad \text{vs.} \quad H_1 : -1 < \rho < 1.$$

Under the maintained hypothesis, x_t is vector integrated process whose elements are not mutually cointegrated. There is no cointegration between y_t and x_t under the null, whereas y_t and x_t are cointegrated under the alternative because $y_t - \beta' x_t = \phi_c' D_t + u_{2,t}$ mean reverts to its deterministic component when $-1 < \rho < 1$.

Assumptions (i)-(iii) are fairly standard and are the same as (A1)-(A3) of Elliott and Jansson (2003) apart from additional moment restrictions. Assumption 1 ensures that the initial values are asymptotically negligible, Assumption 2 is a stationarity condition, and Assumption 3 implies that $\{\varepsilon_t^*\}$ satisfies a functional central limit theorem (e.g. Phillips and Solo (1992)).

There exist many tests for this null and alternative in the literature. In their seminal paper introducing cointegration, Engle and Granger (1987) suggest tests based on least squares residuals from a regression of y_t on deterministic terms and the x_t variables. Phillips and

Ouliaris (1990) derived the asymptotic distributions for these tests. Banerjee et al. (1986), Kremers et al (1992), Boswijk's (1994), and Zivot (2000) looked at tests for no cointegration in single equation conditional error correction models. Finally, one of Johansen's tests, the maximum eigenvalue test (Johansen and Juselius, 1990) can be used to test this hypothesis.

In the context of the above model, local power functions have been considered. Kremers et al (1992), Zivot (2000), Banerjee et al. (1986) computed the local power of the error corrections test when β is prespecified, while Johansen (2005) derived the local power for rank tests. Using model (1), Pesavento (2004), and Pesavento (2006) derived and compared, the local power functions of all the tests for the null of no cointegration when β is unknown.

One reason for the plethora of tests is that the typical generalized likelihood ratio test — likelihood ratio tests that are evaluated with nuisance parameters replaced by their MLE's — is neither locally asymptotically normal nor is it optimal. The lack of normality means that the usual Wald tradeoffs in terms of directions for power may not be particularly interesting. The lack of optimality arises both because of the curvature of the model as well as the problem that the cointegrating vector is a set of unidentified parameters under the null hypothesis.

Power curves derived from the above model have the nice property that they do not depend on the unknown values for the cointegrating vector. Hence choice between methods and in assessing which method to employ the researcher does not need to consider the effect the particular value the cointegrating vector takes for their problem. However, both power and the relative rankings of the test — and hence the choice of method — do rely on the correlation between $u_{1,t}$ and $u_{2,t}$. This is a problem because in practice this parameter is not only unknown but under the null and local alternative this parameter is not consistently estimable. Hence the researcher can look at the curves for various values for the zero frequency correlation but they will have no idea which of these curves is appropriate for any particular application.

A solution to this could be to take a weighted average of power over various values for the long run correlation, with weights related to the types of models the researcher expects to encounter in practice. Whilst this is a reasonable solution, the path that we take here will be to transform the model into its error correction representation, and consider the zero frequency correlation matrix in that representation fixed. This avoids the problem of

producing results that rely on an unknowable nuisance parameter, since in this rotation we can estimate the correlation consistently.

Ignoring deterministic terms for the sake of exposition, a general error correction model (ECM) representation for the data is

$$\tilde{A}(L) \left[I_{k+1} (1 - L) - \alpha \bar{\beta}' L \right] z_t = \varepsilon_t, \quad (4)$$

where $\bar{\beta} = (1, -\beta)'$ and $\varepsilon_t = P^{-1} \varepsilon_t^*$ where $P = \begin{pmatrix} I_m & 0 \\ -\beta & 1 \end{pmatrix}$. In this rotation, it is possible to consistently estimate the spectral density $\tilde{A}(1)^{-1} \Sigma \tilde{A}(1)^{-1'}$. This means that power comparisons based on these nuisance parameters are directly useful for an implication of which method to employ. In the following section we will develop methods based on fixing this spectral density.

Regardless of the particular rotation of the model chosen, under the null hypothesis both of these equivalent models have the property that they are vector autoregressions in the differenced data $(\Delta x'_t, \Delta y_t)$ and the models do not depend on β . Hence the cointegrating vector is an unidentified nuisance parameter under the null hypothesis. This problem commonly arises when the model is written where unknown parameters appear as products (such as $a' \bar{\beta}$ in the ECM model). In situations where this problem arises it is still possible to use standard statistical optimality results based on the Neyman Pearson lemma, however for maximising power against a weighted average over the unidentified nuisance parameter. Essentially what can be constructed is a test statistic that maximises power against a random alternative where the randomness in the nuisance parameter is exactly that chosen to construct the test statistic.

The weighting, to the extent that it affects power, should thus be chosen to reflect the types of cointegrating vectors that the researcher expects to see. Choosing the weighting would be moot if the power of the tests did not depend on the particular cointegrating vector, however such a dependence arises necessarily given the results mentioned above. There power having fixed Σ^* does not depend on the cointegrating vector β . However since we fix $\Sigma = P^{-1} \Sigma^* P^{-1'}$ where P is a function of β , then it will mean that power will depend on the cointegrating vector β .

So in order to construct tests, we will take the approach of computing tests that are weighted average optimal for a chosen weighting over the possible values for the cointegrating vector. Since power does depend on β then it is likely that the weighting will matter for power performance. Understanding the extent of this dependence is one of the goals of the paper. A number of weightings are possible, although as we show below the simplest test statistic arises with an even weighting across the cointegrating vectors.

An additional dimension for consideration in constructing tests aimed at generating good power is the direction in which power is to be maximised. This is a problematic issue in all but the simplest of testing problems, since gaining power in one direction often comes at the expense of power in an alternative direction. In considering cointegrating models that are local to no cointegration, an issue that arises is the behavior of x_t under the local alternative. Compare the ECM form of the model to the (4) form. The two representations are equivalent when we set $\varepsilon_t = P^{-1}\varepsilon_t^*$, $\tilde{A}(L) = P^{-1}A(L)P$ and we have $\alpha = (0, (\rho - 1))'$. Thus the models are equivalent with the restriction that the cointegrating vector does not enter the first k equations. This restriction is precisely the restriction that x_t are constrained to be $I(1)$ under the null and alternative — this is the known information that the testing procedures developed here are intended to exploit (see Elliott et. al. (2005) for a detailed discussion). Essentially what happens is that if this restriction does not hold then under the local alternative the nearly integrated cointegrating term is part of the shock to the change in x_t , hence the change in x_t would have a local to $I(1)$ component. Here we assume that x_t is $I(1)$ in the sense that the weak limit of $T^{-1/2}x_{[T\cdot]}$ is a Brownian motion under the null and local alternatives of the form $\rho = 1 - T^{-1}\gamma$, where γ is a fixed constant.

>From the perspective of comparing power curves, local power will differ when one includes this assumption or not. We are not aware of any analytical work examining local power without this assumption, although often this assumption is left implicit.

3 Tests for the null of no Cointegration

In this section we derive point optimal tests that maximise weighted average power for various weightings on the unknown cointegrating vector β . The tests that we derive have optimality properties for normal errors in the model where we have no serial correlation ($F(L) = I$) at the point alternative chosen. This family of optimal tests can be used to trace

out an approximate power envelope for tests of the null of no cointegration with the stated weights on potential cointegrating vectors. This enables the numerical examination of how well various tests perform in terms of power relative to what is possible. We then construct tests that have the same asymptotic power properties as these point optimal tests under the usual milder conditions on the error process and dependence structure.

3.1 Derivation of the family of tests.

Consider the model (4) with the deterministic terms included, and having set $F(L) = I$. Define $y = (y_1, \dots, y_T)'$, $x = (x_1, \dots, x_T)'$, $u_i = (u_{i,1}, \dots, u_{i,T})'$ for $i = 1, 2$ and $D = (D_1, \dots, D_T)'$. The model can be written in matrix notation as

$$\tilde{P} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} = \begin{pmatrix} \text{vec}(x) \\ y - x\beta \end{pmatrix} = \begin{pmatrix} (I_k \otimes D)\text{vec}(\alpha_x) \\ D\alpha_c \end{pmatrix} + \begin{pmatrix} \text{vec}(u_1) \\ u_2 \end{pmatrix} = u \quad (5)$$

where

$$\begin{pmatrix} I_k \otimes A(1) & 0 \\ 0 & A(\rho) \end{pmatrix} u = \tilde{A}u = \varepsilon^*. \quad (6)$$

and we define

$$A(r) = \begin{pmatrix} 1 & 0 & & 0 \\ -r & 1 & 0 & \\ 0 & -r & \ddots & \ddots \\ & \ddots & \ddots & \ddots & 0 \\ 0 & & 0 & -r & 1 \end{pmatrix}$$

for any scalar r and $\tilde{P} = P \otimes I_T$. Then $\tilde{P}^{-1}Au = \tilde{P}^{-1}\varepsilon^* = \varepsilon$ where we assume $\varepsilon \sim N(0, \tilde{\Sigma})$ with $\tilde{\Sigma} = \Sigma \otimes I_T$. Here Σ is a general $(k+1) \times (k+1)$ variance covariance matrix (the contemporaneous covariance matrix between the residuals). We will partition this matrix after the k^{th} row and column, writing the elements of the partitioned matrix as Σ_{ij} for blocks $i, j = 1, 2$. Define $\Sigma_{2.1} = \Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12}$. It follows that $u = \tilde{A}^{-1}\tilde{P}\varepsilon$ and hence $u \sim N(0, \tilde{A}^{-1}\tilde{P}\tilde{\Sigma}\tilde{P}'\tilde{A}^{-1})$.

Since we observe $(\text{vec}(x)', y')$ and not u we impose invariance to translations of the form

$$(\text{vec}(x)', y') \mapsto (\text{vec}(x)' + \text{vec}(a_x)'(I_k \otimes D'), y' + a'_c D') \quad (7)$$

for any real conformable a_x, a_c . We have

$$\begin{pmatrix} vec(x) \\ y \end{pmatrix} = \tilde{P}^{-1} \begin{pmatrix} (I_k \otimes D)vec(\phi_x) \\ D\phi_c \end{pmatrix} + \tilde{P}^{-1}u.$$

Define $M = I_T - D'(D'D)^{-1}D'$ and $\tilde{M} = I_{k+1} \otimes M$. Then

$$\begin{aligned} \tilde{M} \begin{pmatrix} vec(x) \\ y \end{pmatrix} &= \tilde{M}\tilde{P}^{-1}u \\ &= \tilde{M}\tilde{P}^{-1}\tilde{A}^{-1}\tilde{P}\varepsilon. \end{aligned}$$

The Likelihood function invariant to the above translations is then proportional to

$$\exp \left[-\frac{1}{2} \begin{pmatrix} vec(x)' & y' \end{pmatrix} \tilde{M} \left(\tilde{M}\tilde{P}^{-1}\tilde{A}^{-1}\tilde{P}\tilde{\Sigma}\tilde{P}'\tilde{A}^{-1}\tilde{P}^{-1'}\tilde{M} \right)^{-1} \tilde{M} \begin{pmatrix} vec(x) \\ y \end{pmatrix} \right].$$

This likelihood function depends on the cointegrating vector β through the dependence on \tilde{P} and the value for ρ through \tilde{A} .

In order to construct a feasible test, an issue that arises is the lack of uniformity of the test statistic with respect to the size of the departure of the cointegrating relation from a unit root. It is well understood by now that in testing for a unit root (see Elliott et. al. (1996)) that no uniformly most powerful test for the null of a unit root exists, even asymptotically. This result directly implies the lack of uniformity here too, since a special case of a known cointegrating vector and $\Sigma_{12}=0$ makes the test based on a known cointegrating vector equivalent to the models in that paper. We follow the suggestion of King (1980) as employed in Elliott et. al. (1996) and construct a point optimal test. This enables both the construction of a feasible test as well as the evaluation of maximal power of a classical test since the tests have optimality properties at the chosen alternative.

Theorem 1 *For the model in (5) and (6) the optimal test for the hypothesis $H_0 : \rho = 0$ versus $H_1 : \rho = \bar{\rho}$ where $\varepsilon_t \sim N(0, \Sigma)$ (with known β, Σ) invariant to translations (7) rejects for large values of the statistic*

$$LR_T(\beta) = \exp \left[-\frac{1}{2} \begin{pmatrix} \Sigma_{2,1}^{-1} ((1 - \bar{\rho})^2 c'_{-1} M c_{-1} + 2(1 - \bar{\rho}) c'_{-1} M A(1) M \tilde{y}) \\ - \begin{pmatrix} vec(x)' & y' \end{pmatrix} \tilde{M} G \tilde{M} \begin{pmatrix} vec(x) \\ y \end{pmatrix} \end{pmatrix} \right] \quad (8)$$

where $\tilde{y} = y - x\Sigma_{11}^{-1}\Sigma_{12}$, $c_{-1} = y_{-1} - x_{-1}\beta$,

$$G = K(\bar{\rho})'\tilde{\Sigma}^{-1}K(\bar{\rho})\tilde{D} \left(\tilde{D}'K(\bar{\rho})'\tilde{\Sigma}^{-1}K(\bar{\rho})\tilde{D} \right)^{-1} \tilde{D}K(\bar{\rho})\tilde{\Sigma}^{-1}K(\bar{\rho}) \\ - K(1)'\tilde{\Sigma}^{-1}K(1)\tilde{D} \left(\tilde{D}'K(1)'\tilde{\Sigma}^{-1}K(1)\tilde{D} \right)^{-1} \tilde{D}K(1)\tilde{\Sigma}^{-1}K(1),$$

$K(r) = \tilde{P}^{-1}\tilde{A}\tilde{P}$ which is a function of r through \tilde{A} and $\tilde{D} = I_{k+1} \otimes D$.

This result gives the optimal test under normality as a function of the known cointegrating vector β and the variance covariance matrix of the residuals.

We now turn to the issue of the unknown cointegrating vector β , a nuisance parameter that arises under the alternative only, by constructing a test that has weighted average power over some weights on β . To keep the difference between the true cointegrating vector, β , and the hypothesized ones let the hypothesized cointegrating vector be b , and define $\tilde{b} = \Sigma_{2,1}^{-1/2}\Sigma_{11}^{1/2}b$. We will average over models for different normalized cointegrating vectors \tilde{b} with some weighting function on this parameter vector. The test statistic is then

$$C_T(v_{\tilde{b}}) = 2 \ln \int LR_T(\tilde{b}) dv_{\tilde{b}}.$$

We have a choice of measures for the integration, each of which will result in tests which have optimal power in the experiment that \tilde{b} has against that distribution. Different choices for measures on $\tilde{\beta}$ will yield different tests with different power properties.

The chief difficulty in constructing closed form statistics in the family $C_T(v_{\tilde{b}})$ is that the statistic of Theorem 1 is a complicated function of the cointegrating vector b , and hence this complicates the integration over this parameter vector. However it turns out that for the case with no deterministic terms then $G = 0$ and the test of Theorem 1 is quadratic in b , hence a normal weighting or a constant weighting (the limit of a normal weighting when the variance of the weights is allowed to become large) provide closed form tests. In the case of a constant in the model — $D = \iota$ where ι is a $T \times 1$ vector of ones — then this remains asymptotically true.

Denote the case of constant weights as $C_T(\tilde{b})$. This is a reasonable choice in the absence of any information as to the likely values for the cointegrating vector. If we integrate over \tilde{b} with a flat prior and use a fixed alternative for the largest root of the dynamics for the cointegrating vector, then ignoring terms that are small asymptotically we can obtain a

closed form solution for the test. Set the fixed alternative $\bar{\rho} = 1 - T^{-1}\bar{g}$ with \bar{g} fixed. The statistic that arises from this weighting is

$$\begin{aligned}
C_T(\bar{g}) &= -\Sigma_{2,1}^{-1} \left(\bar{g}^2 T^{-2} y'_{-1} M y_{-1} + 2\bar{g} T^{-1} y'_{-1} M A(1) M \tilde{y} \right) \\
&\quad + \Sigma_{2,1}^{-1} \left(\bar{g} T^{-2} y'_{-1} M x_{-1} \Sigma_{11}^{-1/2} + T^{-1} \tilde{y}' M A(1) M x_{-1} \Sigma_{11}^{-1/2} \right)' \left(\Sigma_{11}^{-1/2'} T^{-2} x'_{-1} M x_{-1} \Sigma_{11}^{-1/2} \right)^{-1} \\
&\quad * \left(\bar{g} T^{-2} y'_{-1} M x_{-1} \Sigma_{11}^{-1/2} + T^{-1} \tilde{y}' M A(1) M x_{-1} \Sigma_{11}^{-1/2} \right) \\
&\quad - \log \left(\det \left(\bar{g}^2 \Sigma_{11}^{-1/2'} T^{-2} x'_{-1} M x_{-1} \Sigma_{11}^{-1/2} \right) \right)
\end{aligned}$$

where, $\tilde{y} = y_t - x_t \Sigma_{11}^{-1} \Sigma_{12}$, and $M = I$ for the no deterministic case and $M = I_T - \iota(\iota' \iota)^{-1} \iota$ for the case where $D = \iota$. The test statistics $C_T(\bar{g})$ are a family of test statistics (indexed by the choice of the local alternative \bar{g}) that provide tests for the null of no cointegration against the alternative that there exists a cointegrating vector where the cointegrating vector is unknown. The following result establishes the distribution of the test under the null and local alternative for any chosen true cointegrating vector.

The limit distributions for these tests can be derived.

Theorem 2 (*limiting properties of $C_T(\bar{g})$*). *Under Condition 1 then for $\gamma = T(1 - \rho)$ fixed as $T \rightarrow \infty$ then when the true cointegrating vector is $\beta = \Sigma_{2,1}^{1/2} \Sigma_{11}^{-1/2} \tilde{\beta}$ then*

$$\begin{aligned}
C_T(\bar{g}) &\Rightarrow -(\bar{g}^2 - 2\bar{g}\gamma) \int (W_{2\gamma}^d)^2 - (\bar{g}^2 - 2\bar{g}\gamma)(\tilde{\beta} - \tilde{R})' \int W_{1\gamma}^d W_{1\gamma}^{d'} (\tilde{\beta} - \tilde{R}) - \bar{g}^2 \tilde{\beta}' \int W_1^d W_1^{d'} \tilde{\beta} \\
&\quad + 2(\bar{g}^2 - \bar{g}\gamma) (\tilde{\beta} - \tilde{R})' \int W_{1\gamma}^d W_{2\gamma}^d + 2(\bar{g}^2 - \bar{g}\gamma)(\tilde{\beta} - \tilde{R})' \int W_{1\gamma}^d W_1^{d'} \tilde{\beta} \\
&\quad - 2(\bar{g}^2 - \bar{g}\gamma) \tilde{\beta}' \int W_1^d W_{2\gamma}^d - 2\bar{g} \int W_{2\gamma}^d dW_2 + 2\bar{g}(\tilde{\beta} - \tilde{R})' \int W_{1\gamma}^d dW_2 - 2\bar{g} \tilde{\beta}' \int W_1^d dW_2 \\
&\quad + \left(\bar{g} \left(\int W_{2\gamma}^d W_1^{d'} - (\tilde{\beta} - \tilde{R})' \int W_{1\gamma}^d W_1^{d'} + \tilde{\beta}' \int W_1^d W_1^{d'} \right) + \right. \\
&\quad \left. \int dW_2^d W_1^{d'} - \gamma \int W_{2\gamma}^d W_1^{d'} + \gamma(\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_1^{d'} \right) \left(\int W_1^d W_1^{d'} \right)^{-1} \\
&\quad \bullet \left(\bar{g} \left(\int W_{2\gamma} W_1^{d'} - (\tilde{\beta} - \tilde{R})' \int W_{1\gamma}^d W_1^{d'} + \tilde{\beta}' \int W_1^d W_1^{d'} \right) + \right. \\
&\quad \left. \int dW_2 W_1^{d'} - \gamma \int W_{2\gamma}^d W_1^{d'} + \gamma(\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_1^{d'} \right)' \\
&\quad - \log \left(\det \left(\bar{g}^2 \int W_1^d W_1^{d'} \right) \right).
\end{aligned}$$

where $\tilde{R} = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{2,1}^{-1/2}$, $W_1(\lambda)$ and $W_2(\lambda)$ are $k \times 1$ and 1×1 standard Brownian motion processes respectively, $W_{i\gamma}(\lambda) = -\gamma \int_0^\lambda e^{-\gamma(\lambda-r)} W_i(r) dr + W_i(\lambda)$ for $i = 1, 2$, and $W_{i\gamma}^d(\lambda) = W_{i\gamma}(\lambda)$ for the no deterministic case and $W_{i\gamma}^d(\lambda) = W_{i\gamma}(\lambda) - \int W_{i\gamma}(r) dr$ for the case where $D = \iota$.

Under the local alternative, for a fixed (normalized) cointegrating vector $\tilde{\beta}$ the limit distributions depend on the choice of the fixed alternative \bar{g} , and the alternative model values for $\tilde{\beta}$, \tilde{R} and γ . Here $\tilde{R} = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{2.1}^{-1/2} = \delta(1 - \delta'\delta)^{-1/2}$ where $\delta = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}$, which is the vector of correlations between ε_{1t} and ε_{2t} . Since, as noted in the previous section, this parameter is consistently estimable under the null and local alternative (and so will its analog in more general settings), then a researcher can compute this number to access the relevant power function. The dependence of the power function on $\tilde{\beta}$ was also suggested in the previous section.

The null distribution is given when $\gamma = 0$, in which case the distribution simplifies and no longer depends on either \tilde{R} or $\tilde{\beta}$, and thus is independent of all nuisance parameters. The null distribution is given by

$$\begin{aligned}
C_T(\bar{g}) \Rightarrow & -\bar{g}^2 \int (W_2^d)^2 - 2\bar{g} \int W_2^d dW_2 + \bar{g}^2 \int W_2^d W_1^{d'} \left(\int W_1^d W_1^{d'} \right)^{-1} \int W_1^d W_2^d \\
& + \int dW_2 W_1^{d'} \left(\int W_1^d W_1^{d'} \right)^{-1} \int W_1^d dW_2 + 2\bar{g} \int W_2^d W_1^{d'} \left(\int W_1^d W_1^{d'} \right)^{-1} \int W_1^d dW_2 \\
& - \log \left(\det \left(\bar{g}^2 \int W_1^d W_1^{d'} \right) \right).
\end{aligned} \tag{9}$$

The distribution does however depend on the alternative \bar{g} chosen along with the dimension of x_t .

The test $C_T(\bar{g})$ is obtained by integrating the first two terms of (8) across $\tilde{\beta}$. Since the third term is zero for the case where $D = 0$ and asymptotically negligible when $D = \iota$ then this is asymptotically equivalent to the optimal test under the null and local alternative hypotheses.

Theorem 3 (*constant weights on \tilde{b}*) *For the testing problem of Theorem 1 when $\bar{p} = 1 + T^{-1}\bar{g}$, then under Condition 1 then for the cases where $D = 0$ and $D = \iota$ then the family of tests $C_T(\bar{g})$ are asymptotically equivalent to the family of tests $C_T(\tilde{\beta})$ under the null hypothesis and local alternative.*

This result shows that the tests are, at each point of the local alternative \bar{g} , optimal for the testing problem considered here. The tests can be used to trace out maximal power against models where $\tilde{\beta}$ is drawn from the uniform distribution.

Similar results can be obtained using a normal weighting on the distribution for $\tilde{\beta}$, however in this case the null distribution depends on the nuisance parameter \tilde{R} .

3.2 Tests under more general assumptions.

The assumption of normality, necessary for writing the likelihood, affects the form of the statistic. In this section we show that the optimal power under normality and no serial correlation can be achieved with $C_T(\bar{g})$ asymptotically under a much wider range of assumptions on the distribution of the errors (including allowing serial correlation). For a researcher that knows the distribution of the errors when it is not normal could construct a test statistic that is potentially more powerful than that of the previous section. When such a distribution is unknown, potential gains may still exist, as examined by Rothenberg and Stock (1993) and Jansson (2007) in the univariate unit root test problem. However such gains will likely come at the expense of robustness, as shown by Mueller (2007). We will remove the assumption of a known Σ by employing consistent estimates of the scale parameters, which for a wide set of assumptions will not affect asymptotic power of the resulting statistic.

When the errors satisfy condition 1, we replace ε_t in the previous subsection with $v_t = F(L)\varepsilon_t$. We are able to consistently estimate both $\Omega = \lim_{T \rightarrow \infty} E\left[\left(T^{-1/2} \sum_{t=1}^T v_t\right) \left(T^{-1/2} \sum_{t=1}^T v_t\right)'\right]$ and $\Lambda = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=2}^T \sum_{j=1}^{t-1} E[v_j v_t']$ under the null and local alternative from the residuals of the first order vector autoregression using $(x'_t, y_t)'$. A sums of covariances estimator is convenient. Denote the consistent estimator $\hat{\Omega}$.

For the construction of the statistic, we redefine $\tilde{y} = y - x\hat{\Omega}_{11}^{-1}\hat{\Omega}_{12}$, and replace scaling matrices by their zero frequency counterparts. Also, partition both Ω and Λ after the k^{th} row and column and define Λ_{12} and Λ_{22} accordingly (these matrices allow corrections similar to the form of Hansen and Phillips (1989)). The statistic is now.

$$\begin{aligned}
C_T(\bar{g}) = & -\Omega_{2,1}^{-1} \left(\bar{g}^2 T^{-2} y'_{-1} M y_{-1} + 2\bar{g} T^{-1} y'_{-1} M A(1) \tilde{y} - \hat{\Lambda}_{22} \right) \\
& + \Omega_{2,1}^{-1} \left(\bar{g} T^{-2} y'_{-1} M x_{-1} \Sigma_{11}^{-1/2} + T^{-1} \tilde{y}' A(1) M x_{-1} \Omega_{11}^{-1/2} - \hat{\Lambda}'_{12} \right)' \left(\Omega_{11}^{-1/2'} T^{-2} x'_{-1} M x_{-1} \Omega_{11}^{-1/2} \right)^{-1} \\
& * \left(\bar{g} T^{-2} y'_{-1} M x_{-1} \Omega_{11}^{-1/2} + T^{-1} \tilde{y}' A(1) M x_{-1} \Omega_{11}^{-1/2} - \hat{\Lambda}'_{12} \right) \\
& - \log \left(\det \left(\bar{g}^2 \Omega_{11}^{-1/2'} T^{-2} x'_{-1} M x_{-1} \Omega_{11}^{-1/2} \right) \right).
\end{aligned}$$

When Condition 1 holds then under the null and local alternatives this test statistic has the same asymptotic distribution as given in Theorem 1 above. The null distribution depends only on \bar{g} and the dimension of the X variables.

We construct in the next section tests based on various weightings over the cointegrating vector, although the simplest test arises using equal weights. The test can be constructed through the following steps.

- Step 1: Run a vector autoregression of z_t on z_{t-1} including deterministic terms (for example if $D = \iota$ include a constant). From the residuals, construct an estimate of the (scaled by 2π) spectral density at frequency zero, denoted $\hat{\Omega}$, along with a consistent estimate $\hat{\Lambda}$ of Λ . If the data are not serially correlated then $\hat{\Omega}$ is simply the variance covariance matrix of the residuals of the VAR and $\hat{\Lambda} = 0$. Partition the matrices after the m^{th} row and column and let $\hat{\Omega}_{ij}$ and $\hat{\Lambda}_{ij}$ denote the ij^{th} blocks. Construct $\hat{\Omega}_{11}$ and $\hat{\Omega}_{2,1} = \hat{\Omega}_{22} - \hat{\Omega}'_{12}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{12}$. Also construct $\hat{\Lambda}_{22}$ and $\hat{\Lambda}_{12}$.
- Step 2: If $D = \iota$, construct $y_t^d = y_t - T^{-1} \sum_{s=1}^T y_s$ and $x_t^d = x_t - T^{-1} \sum_{s=1}^T x_s$, otherwise $y_t^d = y_t, x_t^d = x_t$.
- Step 3. Construct w_t such $w_1 = y_1^d - x_1^d \hat{\Omega}_{11}^{-1} \hat{\Omega}_{12}$ and $w_t = (y_t^d - y_{t-1}^d) - (x_t^d - x_{t-1}^d)' \hat{\Omega}_{11}^{-1} \hat{\Omega}_{12}$ for $t = 2, \dots, T$.
- Step 4: Construct $\tilde{x}_t^d = \hat{\Omega}_{11}^{-1/2} x_t^d$.
- Step 5: The statistic is then computed as

$$\begin{aligned}
C_T(\bar{g}) &= -\hat{\Omega}_{2,1}^{-1} \left(\bar{g}^2 T^{-2} \sum_{t=2}^T (y_{t-1}^d)^2 + 2\bar{g} T^{-1} \sum_{t=2}^T y_{t-1}^d w_t - \hat{\Lambda}_{22} \right) \\
&\quad + \hat{\Omega}_{2,1}^{-1} \left(\bar{g} T^{-2} \sum_{t=2}^T \tilde{x}_{t-1}^d y_{t-1}^d + T^{-1} \sum_{t=2}^T \tilde{x}_{t-1}^d w_t - \hat{\Lambda}'_{12} \right)' \left(T^{-2} \sum_{t=2}^T \tilde{x}_{t-1}^d \tilde{x}_{t-1}^{d'} \right)^{-1} \\
&\quad * \left(\bar{g} T^{-2} \sum_{t=2}^T \tilde{x}_{t-1}^d y_{t-1}^d + T^{-1} \sum_{t=2}^T \tilde{x}_{t-1}^d w_t - \hat{\Lambda}'_{12} \right) \\
&\quad - \log \left(\det \left(\bar{g}^2 T^{-2} \sum_{t=2}^T \tilde{x}_{t-1}^d \tilde{x}_{t-1}^{d'} \right) \right).
\end{aligned}$$

- Step 6: The test rejects for large values, with critical values given in Table 1.

4 Evaluating the Tests

In this section we evaluate the tests presented and compare against popularly employed tests for this null hypothesis.

4.1 Large Sample Evaluation

The tests were designed to have power against the cointegrating vector β being unknown and drawn from an even weighted distribution on its (infinite) support. We approximate this support by considering the case where β is drawn from a uniform distribution.

More precisely. We set $m = 1$ so that x_t is univariate. Under the null hypothesis the data are drawn as random walks with mean zero normal innovations with variance covariance matrix Σ , which defines \tilde{R} . We set $\Sigma_{11} = \Sigma_{22} = 1$, and vary Σ_{12} . Under the local alternative we draw residuals as under the null hypothesis, however now the data are drawn from a VAR(1) with coefficient matrix

$$\begin{pmatrix} 1 & 0 \\ \beta(\rho - 1) & \rho \end{pmatrix}$$

where $\rho = 1 - \gamma T^{-1}$. We draw β from a $U(-10, 10)$ distribution and we construct power curves against this random alternative. We employ 5,000 Monte Carlo replications and a sample of $T = 1000$ to approximate large sample performance. Results are presented for $D = \iota$ and $\Sigma_{12} = -0.5, -0.25, 0, 0.25$, and 0.5 . Figures 1-5 presents the rejection rates of $C_T(\bar{g})$ for $\bar{g} = 3$ (LR in the graphs) in comparison with Engle and Granger (1987) test (EG), Boswijk's (1994) Wald test, the maximum eigenvalue test (JOH) of Johansen and Juselius (1990), and Banerjee et al. (1986) test on the Error correction equation (ECR). We also plot the rejection rates for $C_T(\bar{g})$ when $\bar{g} = c$ for each c , $C_T(c)$ (MAX in the graphs). This is the power bound or the maximum attainable power at each point in the alternative when β is drawn from an uniform distribution.

While the asymptotic distribution of the tests under the local alternative depend on the value of Σ_{12} is clear from Figure 1-5 that the power curves are very similar. None of the existing test achieve the power bound for alternatives close to the null. $C_T(3)$ is tangent to $C_T(c)$ by definition and close to the null it outperforms all other tests although the gain in power over using a system tests (ECR, Wald or JOH) is not large. The unit root test on the

residuals (EG) has very low power in all four scenarios as it is based on a single equation for y_t and does not use the information contained in the equation for x_t . Even the case of $\Sigma_{12} = 0$ in the rotated model, is equivalent to $\Sigma_{12}^* = -\beta$ in (3) which corresponds to cases in Pesavento (2004) in which Engle and Granger's (1987) test does not perform well and system tests perform very well. In fact, in all four scenarios residual tests for cointegration are biased (see also Pesavento (2004)). For alternative points further away from the null, ECR, Wald, and JOH have power than is higher than $C_T(3)$.

Overall it is clear that all the system tests are very close to each other and not too far from the maximal power. This result suggests that existing tests are indeed close to the power bounds at least in the case in which we have no information on likely values of the cointegration vector.

5 Conclusion

Tests for the null of no cointegration are nonstandard in a number of ways — as with tests for a unit root null there is no uniformly most powerful test even when the cointegrating vector is known. When these nuisance parameters are not known, they do not appear under the null hypothesis. For both these reasons we do not expect that tests based on MLE's or regression approaches to have any optimality properties (at least none follow from the usual arguments). This paper sets out to compute bounds for testing the null hypothesis of no cointegration to evaluate whether or not the tests are close to optimal in a practical sense (that there is not much in the way of gains in large sample power to be had from constructing new tests) or if they fall short.

We characterize optimal tests for this problem. As is typical with a nuisance parameter not known under the null distribution, power depends on the weighting over possible sets of nuisance parameters. We show that for an equal weighting over possible cointegrating vectors, that the resulting family of point optimal tests (which in turn trace out the power bound) do not depend under the null hypothesis on any of the nuisance parameters of the problem (they still however depend on the dimensions of the cointegrating vector, however this is obviously known). Power does depend on a nuisance parameter describing the simultaneity of the model as well as the true cointegrating vector.

Averaging across possible cointegrating vectors in accordance with the flat weightings

over potential vectors, we are able to show that (i) existing systems test such as the Johansen (1988) trace test and the ECM-Wald test are close to the power bounds, however there is some difference in power for intermediate local alternatives, and (ii) that univariate tests such as the Engle and Granger (1983) test do not have good power against this alternative hypothesis. Overall the differences between the power bounds and the leading tests are not so great and so these results suggest the use of the popular statistics in practice. The main caveat to the results rests on the chosen weighting function — a more informative weighting function will result in a different power bound and different power functions for the popularly employed tests, so the differences may well be larger (or smaller) in such a situation. From the results of Elliott et. al. (2005) that a single point weight on the cointegrating vector (i.e. when it is known) results in substantial power gains.

Table 1:Critical values

		1%	5%	10%
$D = 0$	$\bar{g} = 3$	9.04	6.34	5.30
	$\bar{g} = 5$	9.45	6.87	5.83
	$\bar{g} = 10$	10.43	7.81	6.31
$D = \iota$	$\bar{g} = 3$	9.12	6.48	5.29
	$\bar{g} = 5$	9.46	6.63	5.40
	$\bar{g} = 10$	9.74	6.98	5.67

The critical values are computed by approximating the Brownian Motions in (9) by step functions using Gaussian random walk with 1000 steps and 50000 replications for the corresponding \bar{g} .

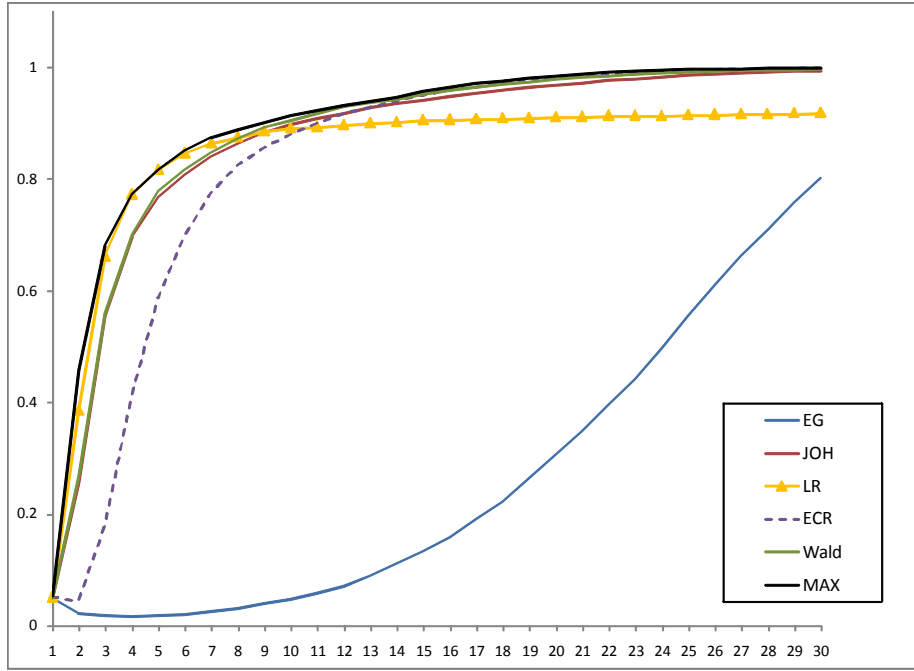


Figure1: Rejection Rates, $\Sigma_{12} = -0.5$, $\bar{g} = 3$, $D = \iota$

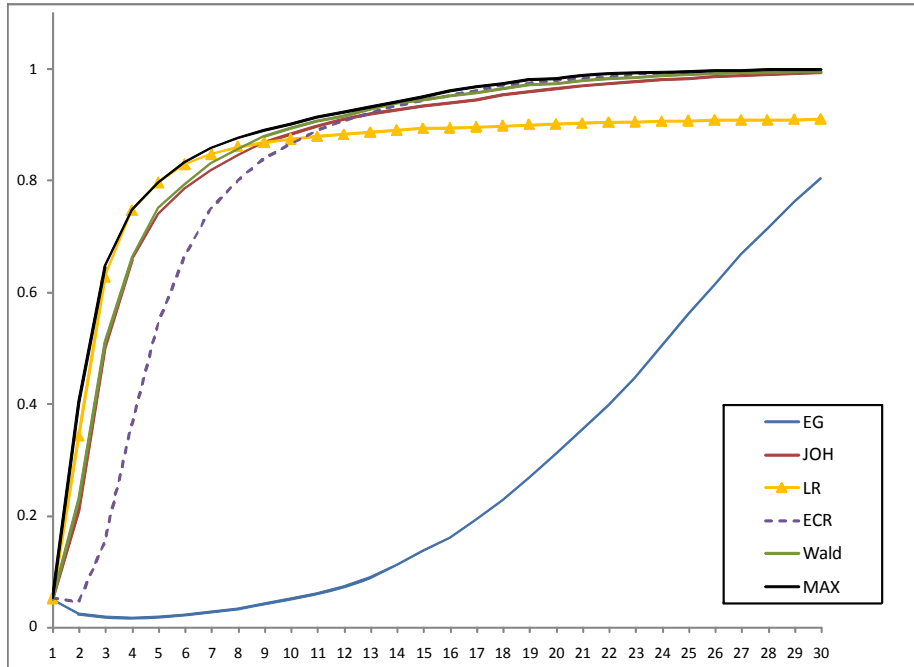


Figure 2: Rejection Rates, $\Sigma_{12} = -0.25$, $\bar{g} = 3$, $D = \iota$

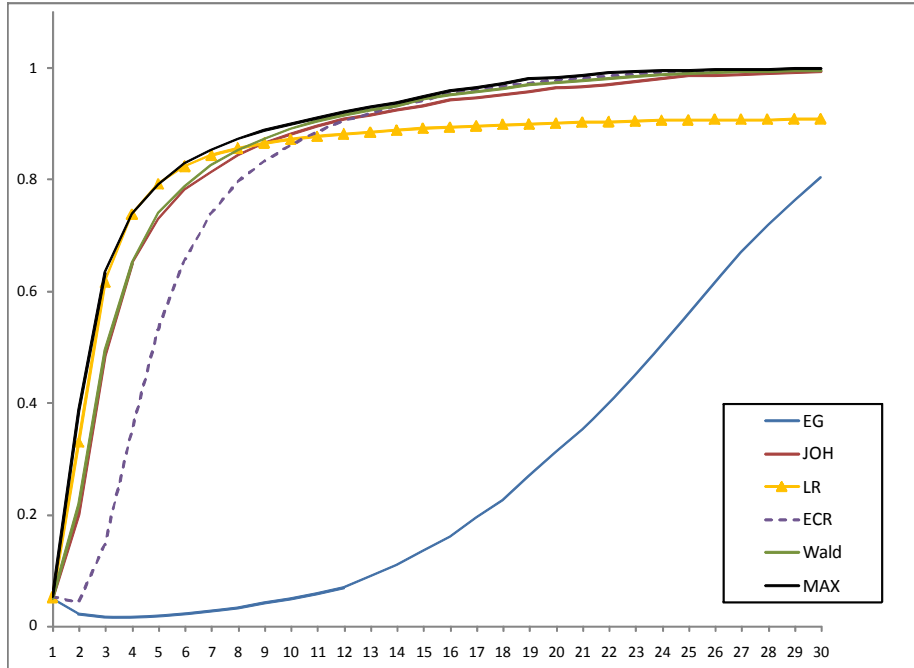


Figure 3: Rejection Rates, $\Sigma_{12} = 0$, $\bar{g} = 3$, $D = \iota$

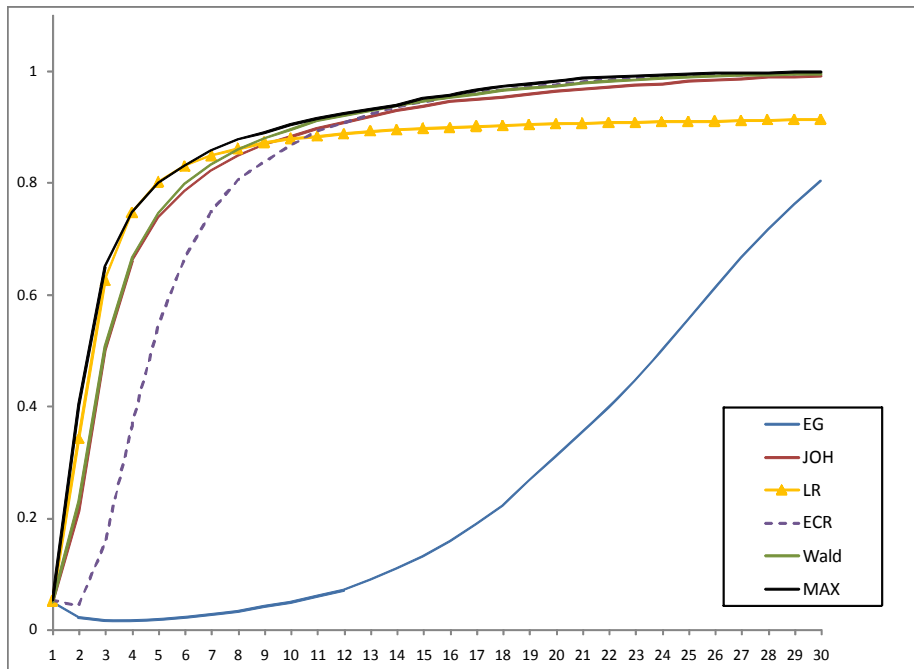


Figure 4: Rejection Rates, $\Sigma_{12} = 0.25$, $\bar{g} = 3$, $D = \iota$

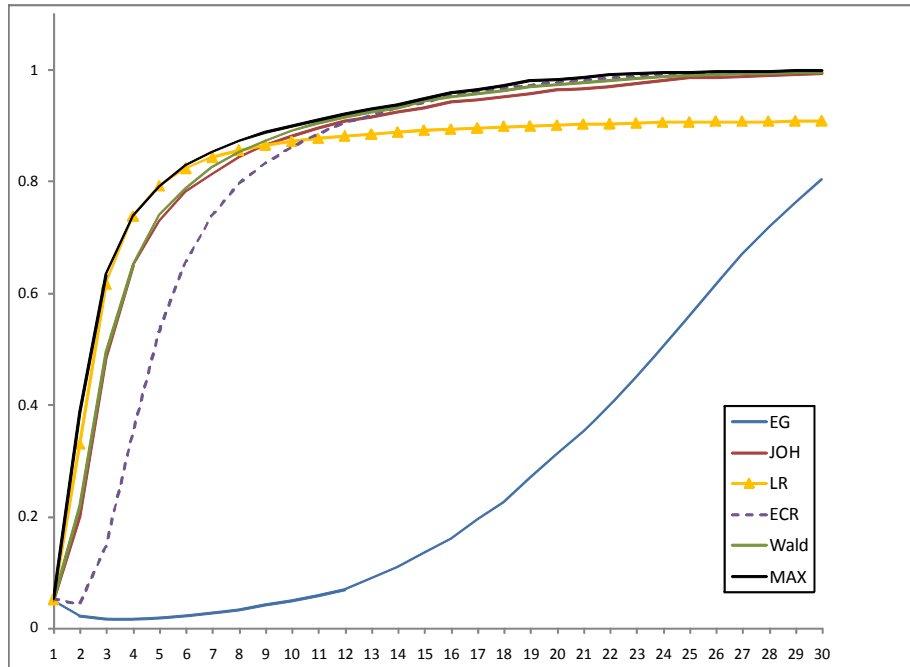


Figure 5: Rejection Rates, $\Sigma_{12} = 0.5$, $\bar{g} = 3$, $D = \iota$

6 References

Andrews, D., and W. Ploberger., 1994, Optimal Tests when the Nuisance Parameter is only present under the Alternative, *Econometrica*, 62, 1383-1414.

Banerjee, A., Dolado, J.J., Hendry, D.F. and G.W. Smith, 1986, Exploring Equilibrium Relationship in Econometrics through Static Models: Some Monte Carlo Evidence, *Oxford Bulletin of Economics and Statistics* 48, 253-277

Boswijk, H. P., 1994, Testing for an Unstable Root in Conditional and Structural Error Correction Models, *Journal of Econometrics* 63, 37-60.

Elliott, G. and M. Jansson, 2003: "Testing for Unit Roots with Stationary Covariates", *Journal of Econometrics*, 115, pp. 75-89.

Elliott, G., M. Jansson, and E. Pesavento, 2005: "Optimal Power for Testing Potential Cointegrating Vectors with Known Parameters for Nonstationarity", *Journal of Business & Economic Statistics*, 23:1, pp.34-48.

Engle, R.F. and C. W. Granger, 1987, Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica* 55, 251-276

Hansen, B.E., 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, 64, 413-430

King, M., 1988, Towards a Theory of Point Optimal Testing, *Econometric Reviews*, 6, 169-218.

Jansson, M., 2007., Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis, manuscript, UC Berkeley.

Johansen, S., 1995, Likelihood Based Inference in Cointegrated autoregressive Models. (Oxford University Press, Oxford).

Johansen, S, 1988, Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. and K. Juselius, 1990, Maximum Likelihood Estimation and Inference on Cointegration with Application to the Demand for Money, *Oxford Bulletin of Economics and Statistics* 52, 169-210.

Mueller, U.K. and G. Elliott., 2003, Tests for Unit Roots and the Initial Observation, *Econometrica*, 71, 1269-1286.

Mueller, U.K., 2007, An Alternative Sense of Asymptotic Efficiency, manuscript, Princeton University.

Pesavento, E., 2004, "An Analytical Evaluation of the power of tests for the absence of cointegration", *Journal of Econometrics*, vol. 122, No. 2, October 2004, pp. 349-384.

Pesavento, E., 2006, "Residuals Based Tests for the Null of No Cointegration: an Analytical Comparison", forthcoming, *Journal of Time Series Analysis*.

Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic Properties of Residuals Based Tests for Cointegration, *Econometrica* 58, 165-193.

Phillips, P.C.B. and V. Solo, 1992, "Asymptotics of Linear Processes", *The Annals Of Statistics* 20, pp. 971-1001.

Rothenberg, T.J. and J.H. Stock., 1993, Inference in a Nearly Integrated Model with Nonnormal Innovations, *Journal of Econometrics*, 80, 269-286.

Zivot, E. (2000): "The Power of Single Equation Tests for Cointegration when the Cointegrating Vector is prespecified", *Econometric Theory*, 16, 407-439

7 Appendix

Proof. (Theorem 1).

Since $u \sim N(0, \tilde{A}^{-1}\tilde{P}\tilde{\Sigma}\tilde{P}'\tilde{A}^{-1})$ and $\tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} = \tilde{M}\tilde{P}^{-1}u$ then the likelihood at $\rho = r$ is proportional to

$$\begin{aligned} & \exp \left[-\frac{1}{2} \begin{pmatrix} \text{vec}(x)' & y' \end{pmatrix} \tilde{M} \left(\tilde{M}\tilde{P}^{-1}\tilde{A}^{-1}\tilde{P}\tilde{\Sigma}\tilde{P}'\tilde{A}^{-1}\tilde{P}^{-1}\tilde{M} \right)^{-1} \tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} \right] \\ & = \exp \left[-\frac{1}{2} \begin{pmatrix} \text{vec}(x)' & y' \end{pmatrix} \tilde{M} \left(\tilde{M}K(r)\tilde{\Sigma}K(r)'\tilde{M} \right)^{-1} \tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} \right] \end{aligned}$$

Now $G(r) = K(r)'\tilde{\Sigma}^{-1}K(r) - K(r)'\tilde{\Sigma}^{-1}K(r)\tilde{D} \left(\tilde{D}'K(r)'\tilde{\Sigma}^{-1}K(r)\tilde{D} \right)^{-1} \tilde{D}K(r)\tilde{\Sigma}^{-1}K(r)$ is a candidate generalised inverse for $\tilde{M}\tilde{P}^{-1}\tilde{A}^{-1}\tilde{P}\tilde{\Sigma}\tilde{P}'\tilde{A}^{-1}\tilde{P}^{-1}\tilde{M}$.

The likelihood can be written as

$$\exp \left[-\frac{1}{2} \begin{pmatrix} \text{vec}(x)' & y' \end{pmatrix} \tilde{M}G(r)\tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} \right].$$

Hence the likelihood ratio test is proportional to

$$\exp \left[-\frac{1}{2} \begin{pmatrix} \text{vec}(x)' & y' \end{pmatrix} \tilde{M}(G(\bar{\rho}) - G(1))\tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} \right].$$

By definition,

$$\begin{aligned} G(\bar{\rho}) - G(1) &= K(\bar{\rho})'\tilde{\Sigma}^{-1}K(\bar{\rho}) - K(1)'\tilde{\Sigma}^{-1}K(1) \\ &\quad - K(\bar{\rho})'\tilde{\Sigma}^{-1}K(\bar{\rho})\tilde{D} \left(\tilde{D}'K(\bar{\rho})'\tilde{\Sigma}^{-1}K(\bar{\rho})\tilde{D} \right)^{-1} \tilde{D}K(\bar{\rho})\tilde{\Sigma}^{-1}K(\bar{\rho}) \\ &\quad + K(1)'\tilde{\Sigma}^{-1}K(1)\tilde{D} \left(\tilde{D}'K(1)'\tilde{\Sigma}^{-1}K(1)\tilde{D} \right)^{-1} \tilde{D}K(1)\tilde{\Sigma}^{-1}K(1). \end{aligned} \tag{10}$$

We can write

$$\begin{aligned}
K(r) &= \tilde{P}^{-1} \tilde{A} \tilde{P} \\
&= \begin{pmatrix} I_k \otimes A(1) & 0 \\ (r-1)(\beta' \otimes B) & A(1) + (1-r)B \end{pmatrix} \\
&= K(0) + (1-r)K_d
\end{aligned}$$

where B is a backward shift matrix

$$B = \begin{pmatrix} 0 & & & 0 \\ 1 & 0 & & \\ 0 & \ddots & & \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and

$$K_d = \begin{pmatrix} 0 & 0 \\ (-\beta' \otimes B) & B \end{pmatrix}.$$

Thus $K(\bar{\rho})' \tilde{\Sigma}^{-1} K(\bar{\rho}) - K(1)' \tilde{\Sigma}^{-1} K(1) = (1-\bar{\rho})^2 K_d' \tilde{\Sigma}^{-1} K_d + (1-\bar{\rho})(K_d' \tilde{\Sigma}^{-1} K(1) + K(1)' \tilde{\Sigma}^{-1} K_d)$.

We have for the first term in (10)

$$\begin{aligned}
& \begin{pmatrix} \text{vec}(x)' & y' \end{pmatrix} \tilde{M} (K(\bar{\rho})' \tilde{\Sigma}^{-1} K(\bar{\rho}) - K(1)' \tilde{\Sigma}^{-1} K(1)) \tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} \\
&= \begin{pmatrix} \text{vec}(Mx)' & y'M \end{pmatrix} ((1-\bar{\rho})^2 K_d' \tilde{\Sigma}^{-1} K_d + (1-\bar{\rho})(K_d' \tilde{\Sigma}^{-1} K(1) + K(1)' \tilde{\Sigma}^{-1} K_d)) \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix}.
\end{aligned}$$

The first of these terms is

$$\begin{aligned}
& (1-\bar{\rho})^2 \begin{pmatrix} \text{vec}(Mx)' & y'M \end{pmatrix} K_d' \tilde{\Sigma}^{-1} K_d \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\
&= \Sigma_{2,1}^{-1} (1-\bar{\rho})^2 \begin{pmatrix} \text{vec}(Mx)' & y'M \end{pmatrix} \left(\begin{pmatrix} -\beta \\ 1 \end{pmatrix} \begin{pmatrix} -\beta' & 1 \end{pmatrix} \otimes BB' \right) \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\
&= \Sigma_{2,1}^{-1} (1-\bar{\rho})^2 (BMy - BMx\beta)' (BMy - BMx\beta) \\
&= \Sigma_{2,1}^{-1} (1-\bar{\rho})^2 c'_{-1} M c_{-1}
\end{aligned}$$

where $c = y - x\beta$ is the error correction term.

For the second term we have

$$\begin{aligned}
& (1 - \bar{\rho}) \begin{pmatrix} \text{vec}(Mx)' & y'M \end{pmatrix} K_0' \tilde{\Sigma}^{-1} K_d \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\
&= (1 - \bar{\rho}) \begin{pmatrix} \text{vec}(Mx)' & y'M \end{pmatrix} \begin{pmatrix} (\Sigma^{12} \otimes A(1)')(-\beta' \otimes B) & \Sigma^{12} \otimes A(1)'B \\ \Sigma^{22} A(1)'(-\beta' \otimes B) & \Sigma^{22} A(1)'B \end{pmatrix} \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\
&= (1 - \bar{\rho}) (-\Sigma^{12'} x' MA(1)' B M x \beta + \Sigma^{22} y' MA(1)' B M y - \Sigma^{22} y' MA(1)' B M x \beta + \Sigma^{12'} x' MA(1)' B M y) \\
&= (1 - \bar{\rho}) (\Sigma^{22} y' MA(1)' M c_{-1} + \Sigma^{12'} x' MA(1)' M c_{-1}) \\
&= (1 - \bar{\rho}) \Sigma_{2,1}^{-1} (y' - \Sigma'_{12} \Sigma_{11}^{-1} x') MA(1)' M c_{-1} \\
&= (1 - \bar{\rho}) \Sigma_{2,1}^{-1} \tilde{y}' MA(1)' M c_{-1}.
\end{aligned}$$

Putting this together yeilds

$$\begin{aligned}
& \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix}' \left(K_1' \tilde{\Sigma}^{-1} K_1 - K_0' \tilde{\Sigma}^{-1} K_0 \right) \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\
&= (1 - \bar{\rho})^2 \Sigma_{2,1}^{-1} c'_{-1} M c_{-1} + 2(1 - \bar{\rho}) \Sigma_{2,1}^{-1} c'_{-1} MA(1) M \tilde{y}.
\end{aligned}$$

The result follows with the remaining terms in $G(\bar{\rho}) - G(0) = G$. ■

To obtain asymptotic distributions, let the disturbances be such that suitably scaled they satisfy an invariance principle, i.e. $T^{-1/2} \xi_{[T\lambda]} = T^{-1/2} \sum_{t=1}^{[T\lambda]} v_t \Rightarrow \Omega^{1/2} W(\lambda)$ uniformly in $\lambda \in [0, 1]$ where \Rightarrow denotes weak convergence, $\Omega = 2\pi S_v(0)$ where $S_v(0)$ is the spectral density of v_t at frequency zero and $W(\lambda)$ is a $k \times 1$ standard Brownian motion. This follows from Condition 1. It follows directly that $T^{-1/2} \sum_{t=1}^{[T\lambda]} P v_t \Rightarrow P \Omega^{1/2} W(\lambda)$ where

$$\begin{aligned}
P \Omega^{1/2} W(\lambda) &= \begin{pmatrix} 1 & 0 \\ -\beta' & 1 \end{pmatrix} \begin{pmatrix} \Omega_{11}^{1/2} & 0 \\ \Omega_{12}' \Omega_{11}^{-1} & \Omega_{2,1}^{1/2} \end{pmatrix} \begin{pmatrix} W_1(\lambda) \\ W_2(\lambda) \end{pmatrix} \\
&= \begin{pmatrix} \Omega_{11}^{1/2} W_1(\lambda) \\ \Omega_{2,1}^{1/2} (W_2(\lambda) - (\tilde{\beta} - \tilde{R}) W_1(\lambda)) \end{pmatrix}.
\end{aligned}$$

where the partition is after the k^{th} row and column. It also follows from Condition 1 that $T^{-1} \sum_{t=2}^T \xi_{t-1} v_t' \Rightarrow \Omega^{1/2} \int W(\lambda) dW(\lambda)' + \underline{\Delta}$.

Let (u_1, u_2) have rows $u_t' = (u_{1t}', u_{2t}')$ and hence we have that $T^{-1/2} u_{1[\lambda T]} \Rightarrow \Omega_{11}^{1/2} W_1(\lambda)$ and $T^{-1/2} u_{2[\lambda T]} \Rightarrow \Omega_{2,1}^{1/2} (W_{2\gamma}(\cdot) - (\tilde{\beta} - \tilde{R}) W_{1\gamma}(\cdot))$ where $W_{i\gamma}(\cdot)$ are Ornstein Uhlenbeck

processes constructed from $W_i(\cdot)$ for $i = 1, 2$. Define $u_{i,-1} = (0, u_{i1}, \dots, u_{iT-1})$. Further define $W_i^j(\lambda) = W_i(\lambda)$ for the no deterministic case and $W^j(\lambda) = W_i(\lambda) - \int W_i(s)ds$ for the case where $D = \iota$. Define $\tilde{v}_{2t} = v_{2t} + (\beta - \Omega_{11}^{-1}\Omega_{12})'v_{1t} = v_{2t} + \Omega_{2.1}^{1/2}(\tilde{\beta} - \tilde{R})'\Omega_{11}^{-1/2}v_{1t}$ so $T^{-1/2} \sum_{t=2}^T \tilde{v}_{2t} \Rightarrow \Omega_{2.1}^{1/2}W_2(\cdot)$.

A number of results used in proving the theorems follow directly. First, $T^{-2}u'_{1,-1}Mu_{1,-1} \Rightarrow \Omega_{11}^{1/2} \int W_1^j(\lambda)W_1^j(\lambda)'\Omega_{11}^{1/2'}$ under both H0 and H1. Also $T^{-2}u'_{2,-1}Mu_{2,-1} \Rightarrow \Omega_{2.1} \int \left(W_{2\gamma}^j(\cdot) - (\tilde{\beta} - \tilde{R})W_{1\gamma}^j(\cdot) \right)$ and $T^{-2}u'_{2,-1}Mu_{1,-1} \Rightarrow \Omega_{2.1}^{1/2} \int \left(W_{2\gamma}^j(\cdot) - (\tilde{\beta} - \tilde{R})W_{1\gamma}^j(\cdot) \right) W_1^j(\cdot)'\Omega_{11}^{1/2'}$.

Also, note that $T^{-1}u'_{1,-1}M\tilde{v}_2 \Rightarrow \Omega_{11}^{1/2} \int W_1^j(\lambda)dW_2^j(\lambda)\Omega_{2.1}^{1/2} + \Lambda_{12}$, and $T^{-1}u'_{2,-1}M\tilde{v}_2 \Rightarrow \Omega_{2.1} \int (W_{1\gamma}^j(\lambda) - (\tilde{\beta} - \tilde{R})W_{1\gamma}^j(\lambda))dW_2^j(\lambda) + \Lambda_{22}$.

Proof. (Theorem 2).

Results will be constructed for the serially correlated case, so results of the theorem are for the special case of $\Omega = \Sigma$ known and $\Lambda = 0$. Rearranging, $y_{-1} = (y_{-1} - x_{-1}\beta + x_{-1}\beta)$, so $T^{-2}\Omega_{2.1}^{-1}y'_{-1}My_{-1} = T^{-2}\Omega_{2.1}^{-1} \left(u'_{2,-1}Mu_{2,-1} + \tilde{\beta}'\Omega_{11}^{1/2'}u'_{1,-1}Mu_{1,-1}\Omega_{11}^{1/2}\tilde{\beta} + 2\Omega_{2.1}^{1/2}u'_{2,-1}Mu_{1,-1}\Omega_{11}^{1/2}\tilde{\beta} \right)$. Limit results follow from the results above for each of these peices. Similarly the terms $T^{-2}\Omega_{11}^{-1/2}u'_{1,-1}Mu_{1,-1}\Omega_{11}^{-1/2'}$ and $T^{-2}\Omega_{2.1}^{-1/2}u'_{2,-1}Mu_{1,-1}\Omega_{11}^{-1/2'}$ follow directly. For the terms involving $A(1)M\tilde{y}$, notice that $A(1)M\tilde{y} = M\tilde{v}_2(1-\rho)Mu_{2,-1}$. Hence we have $T^{-1}\Omega_{2.1}^{-1/2}\Omega_{11}^{-1/2}u'_{1,-1}MA(1)M\tilde{y} = T^{-1}\Omega_{2.1}^{-1/2}\Omega_{11}^{-1/2}u'_{1,-1}M\tilde{v}_2 - \gamma T^{-2}\Omega_{2.1}^{-1/2}\Omega_{11}^{-1/2}u'_{1,-1}Mu_{2,-1}$ where the limit results follow from above. Similarly $T^{-1}\Omega_{2.1}^{-1}u'_{2,-1}MA(1)M\tilde{y} = T^{-1}\Omega_{2.1}^{-1}(u'_{2,-1} + \beta u'_{1,-1})M(\tilde{v}_2 - \gamma T^{-1}u_{2,-1})$ where all terms follow from above.

For the case where $D = \iota$, we still have the third term, which we show here to be $o_p(1)$. First note that the second denominator term $\tilde{D}'K'_0\tilde{\Sigma}^{-1}K_0\tilde{D} = \Sigma^{-1} \otimes \iota'A(1)'A(1)\iota = \Sigma^{-1}$. The first denominator term is

$$\begin{aligned} \tilde{D}'K'_1\tilde{\Sigma}^{-1}K_1\tilde{D} &= \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_0\tilde{D} + (1-\rho)^2\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_d\tilde{D} + (1-\rho)(\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_0\tilde{D} + \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_d\tilde{D}) \\ &= \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_0\tilde{D} + \gamma^2T^{-2}\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_d\tilde{D} + \gamma T^{-1}(\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_0\tilde{D} + \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_d\tilde{D}). \end{aligned}$$

The first term is Σ^{-1} , take the remaining terms independently. First,

$$T^{-2}\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_d\tilde{D} = \Sigma_{2.1}^{-1} \left(\begin{pmatrix} -\beta \\ 1 \end{pmatrix} \begin{pmatrix} -\beta' & 1 \end{pmatrix} \otimes T^{-2}\iota'B'B\iota \right)$$

but $T^{-2}\iota'B'B\iota = T^{-2}(T-1) \rightarrow 0$ and so this term dissapears at rate T . Also

$$T^{-1}\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_0\tilde{D} = \begin{pmatrix} (-\beta\Sigma^{21} \otimes T^{-1}D'B'A(1)D) & \Sigma^{22}(-\beta \otimes T^{-1}D'B'A(1)D) \\ (\Sigma^{21} \otimes T^{-1}D'B'A(1)D) & \Sigma^{22}T^{-1}D'B'A(1)D \end{pmatrix}$$

and $T^{-1}\iota'B'A(1)\iota = T^{-1}0 = 0$. Finally

$$T^{-1}\tilde{D}'K'_0\tilde{\Sigma}^{-1}K_d\tilde{D} = \begin{pmatrix} (-\Sigma^{12}\beta' \otimes T^{-1}D'A(1)'BD) & \Sigma^{12} \otimes T^{-1}D'A(1)'BD \\ \Sigma^{22}(-\beta' \otimes T^{-1}D'A(1)'BD) & \Sigma^{22}T^{-1}D'A(1)'BD \end{pmatrix}$$

and $T^{-1}\iota'A(1)'B\iota T^{-1}\iota'B'A(1)\iota = T^{-1}0 = 0$. Hence we have that for $D = \iota$

$$\begin{aligned} \tilde{D}'K'_1\tilde{\Sigma}^{-1}K_1\tilde{D} &= \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_0\tilde{D} + o(T^{-1}) \\ &= \Sigma^{-1} + o(T^{-1}). \end{aligned}$$

>From this it follows that $K'_1\tilde{\Sigma}^{-1}K_1\tilde{D} \left(\tilde{D}'K'_1\tilde{\Sigma}^{-1}K_1\tilde{D} \right)^{-1} \tilde{D}'K'_1\tilde{\Sigma}^{-1}K_1 = K'_1\tilde{\Sigma}^{-1}K_1\tilde{D}\Sigma\tilde{D}'K'_1\tilde{\Sigma}^{-1}K_1 + o_p(1)$ provided that the numerator terms are at most $0_p(1)$.

For the numerator.

We have

$$\begin{aligned} \tilde{D}'K'_1\tilde{\Sigma}^{-1}K_1 &= \tilde{D}'(K_0 + (1 - \rho)K_d)'\tilde{\Sigma}^{-1}(K_0 + (1 - \rho)K_d) \\ &= \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_0 + (1 - \rho)^2\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_d + (1 - \rho)\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_0 + (1 - \rho)\tilde{D}'K'_0\tilde{\Sigma}^{-1}K_d \\ &= \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_0 + \gamma^2T^{-2}\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_d + \gamma T^{-1} \left(\tilde{D}'K'_d\tilde{\Sigma}^{-1}K_0 + \tilde{D}'K'_0\tilde{\Sigma}^{-1}K_d \right). \end{aligned}$$

Now,

$$\begin{aligned} \tilde{D}'K'_d\tilde{\Sigma}^{-1}K_d \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} &= \Sigma^{22} \begin{pmatrix} (\beta\beta' \otimes \iota'B'B) & (-\beta \otimes \iota'B'B) \\ (-\beta' \otimes \iota'B'B) & \iota'B'B \end{pmatrix} \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\ &= \Sigma^{22} \begin{pmatrix} \beta[\iota'B'BMx\beta - \iota'B'BM_y] \\ -\iota'B'BMx\beta + \iota'B'BM_y \end{pmatrix} \\ &= \Sigma^{22} \begin{pmatrix} -\beta\iota'B'BM_c \\ \iota'B'BM_c \end{pmatrix}. \end{aligned}$$

For the next term

$$\begin{aligned} K'_0\tilde{\Sigma}^{-1}K_d \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} &= \begin{pmatrix} (\Sigma^{12} \otimes A(1)')(-\beta' \otimes B) & \Sigma^{12} \otimes A(1)'B \\ \Sigma^{22}A(1)'(-\beta' \otimes B) & \Sigma^{22}A(1)'B \end{pmatrix} \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\ &= \begin{pmatrix} -(\Sigma^{12} \otimes A(1)')BMx\beta + \Sigma^{12} \otimes A(1)'BM_y \\ -\Sigma^{22}A(1)'BMx\beta + \Sigma^{22}A(1)'BM_y \end{pmatrix} \\ &= \begin{pmatrix} (\Sigma^{12} \otimes A(1)')BM_c \\ \Sigma^{22}A(1)'BM_c \end{pmatrix} \end{aligned}$$

so

$$\tilde{D}'K_0'\tilde{\Sigma}^{-1}K_d \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} = (1 - \rho) \begin{pmatrix} \Sigma^{12} \otimes D'A(1)'BMc \\ \Sigma^{22}D'A(1)'BMc \end{pmatrix}.$$

Finally

$$\begin{aligned} K_d'\tilde{\Sigma}^{-1}K_0 &= \begin{pmatrix} 0 & (-\beta \otimes B') \\ 0 & B' \end{pmatrix} \begin{pmatrix} \Sigma^{11} \otimes A(1) & \Sigma^{12} \otimes A(1) \\ \Sigma^{21} \otimes A(1) & \Sigma^{22} \otimes A(1) \end{pmatrix} \\ &= \begin{pmatrix} (-\beta \otimes B')(\Sigma^{21} \otimes A(1)) & (-\beta \otimes B')(\Sigma^{22} \otimes A(1)) \\ B'(\Sigma^{21} \otimes A(1)) & \Sigma^{22}B'A(1) \end{pmatrix} \end{aligned}$$

so

$$\begin{aligned} K_d'\tilde{\Sigma}^{-1}K_0 \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} &= \begin{pmatrix} (-\beta \otimes B')(\Sigma^{21} \otimes A(1)) & (-\beta \otimes B')(\Sigma^{22} \otimes A(1)) \\ B'(\Sigma^{21} \otimes A(1)) & \Sigma^{22}B'A(1) \end{pmatrix} \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} \\ &= \begin{pmatrix} (-\beta \otimes B')A(1)Mx\Sigma^{12} + \Sigma^{22}(-\beta \otimes B')A(1)My \\ B'A(1)Mx\Sigma^{12} + \Sigma^{22}B'A(1)My \end{pmatrix} \\ &= \Sigma_{2,1}^{-1} \begin{pmatrix} (\beta \otimes B')A(1)Mx\Sigma_{11}^{-1}\Sigma_{12} + (-\beta \otimes B')A(1)My \\ -B'A(1)Mx\Sigma_{11}^{-1}\Sigma_{12} + B'A(1)My \end{pmatrix} \\ &= \Sigma_{2,1}^{-1} \begin{pmatrix} -\beta \otimes B'A(1)M\tilde{y} \\ B'A(1)M\tilde{y} \end{pmatrix}. \end{aligned}$$

Finally

$$\tilde{D}'K_d'\tilde{\Sigma}^{-1}K_0 \begin{pmatrix} \text{vec}(Mx) \\ My \end{pmatrix} = \Sigma_{2,1}^{-1} \begin{pmatrix} -\beta \otimes D'B'A(1)M\tilde{y} \\ D'B'A(1)M\tilde{y} \end{pmatrix}.$$

For a Txx vector z

$$l'B'A(1)Mz = (z_T - z_1)$$

$$l'A(1)BMz = 0$$

$$l'B'BMz = \bar{z} - z_1$$

For the mean case we have that for $T^{-1/2}z_{[T]}$ being $0_p(1)$ then

$$T^{-2}l'B'BMz = T^{-3/2}(T^{-1/2}\bar{z} - T^{-1/2}z_1) \rightarrow^p 0$$

$$T^{-1}l'A(1)'BMz = T^{-1}0 = 0$$

$$T^{-1}l'B'A(1)Mz = T^{-1/2}(T^{-1/2}z_T - T^{-1/2}z_1) \rightarrow^p 0.$$

Then we have

$$\begin{aligned} & \begin{pmatrix} Mx \\ My \end{pmatrix}' \left(K_0' \tilde{\Sigma}^{-1} K_0 \tilde{D} \left(\tilde{D}' K_0' \tilde{\Sigma}^{-1} K_0 \tilde{D} \right)^{-1} \tilde{D} K_0' \tilde{\Sigma}^{-1} K_0 - K_1' \tilde{\Sigma}^{-1} K_1 \tilde{D} \left(\tilde{D}' K_1' \tilde{\Sigma}^{-1} K_1 \tilde{D} \right)^{-1} \tilde{D} K_1' \tilde{\Sigma}^{-1} K_1 \right) \\ &= 0 + o_p(1). \end{aligned}$$

Hence the second term disappears asymptotically in the case where $D = \iota$.

Null distributions follow from setting $\gamma = 0$ so $\int W_{2\gamma} = \int W_2$ etc. ■

Proof. (Theorem 3).

Optimality is for the serially uncorrelated case at the points $\gamma = \bar{g}$ for the uniform distribution over \tilde{b} . This result follows by construction, directly for the case of $D = 0$ and via dominated convergence for the case where $D = \iota$. ■