*Econometric Theory*, **25**, 2009, 1829–1850. Printed in the United States of America.
doi:10.1017/S026646660999034X

# TESTING THE NULL OF NO COINTEGRATION WHEN COVARIATES ARE KNOWN TO HAVE A UNIT ROOT

GRAHAM ELLIOTT
*University of California, San Diego*

ELENA PESAVENTO
*Emory University*

A number of tests have been suggested for the test of the null of no cointegration. Under this null, correlations are spurious in the sense of Granger and Newbold (1974) and Phillips (1986). We examine a set of models local to the null of no cointegration and derive tests with optimality properties in order to examine the efficiency of available tests. We find that, for a sufficiently tight weighting over potential cointegrating vectors, commonly employed full system tests have power that can, in some situations, be quite far from the power bounds for the models examined.

## 1. INTRODUCTION

Given that many economic time series display trending, it is important to distinguish spurious correlation (as in Granger and Newbold, 1974; Phillips, 1986) from meaningful economic relationships. When trending is due to unit roots in the data, correlations between the levels of the data are cointegrating relationships. In their seminal paper, Engle and Granger (1987) proposed tests for the null hypothesis of no cointegration that amounted to testing for a unit root in the estimated least squares cointegrating residuals. Their contribution triggered a wave of alternative tests for this null hypothesis. Whenever there is a variety of tests for a given hypothesis, it is natural to ask which one should be employed in practice. This questions rests, in large samples, on power considerations and, in smaller samples, on size control and power. Understanding which of the tests, if any, is best, is only a first step as relative power comparisons leave open the question of whether or not there exists a test, yet to be derived, that has even better power properties.

The authors would like to thank participants at the Conference in Honour of Paul Newbold, as well as the editor and referees for comments. All errors remain ours. Address correspondence to Elena Pesavento, Emory University, 1602 Fishburne Dr., Atlanta, GA 30322; e-mail: epesave@emory.edu.

© 2009 Cambridge University Press 0266-4666/09 $15.00 

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

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

The lack of an obvious optimal test follows directly from the result that no uniformly maximum power test exists across the set of alternatives even in special cases of the test. 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 such a test does not exist in general. Indeed, the case of an unknown cointegrating vector pulls us farther away 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. In fact, since under the null hypothesis there is no cointegrating vector, as all linear combinations of the data have unit roots, we face the additional problem of a nuisance parameter that is not identified under the null hypothesis. It is well understood that in these situations (see Davies, 1977, 1987; Hansen, 1996; and Andrews and Ploberger, 1994), standard tests do not in general have optimality properties.

Section 2 examines the problem of testing for the null of no cointegration. Previous examination of the power of these tests (Pesavento, 2004) have shown dependence of local power curves on a nuisance parameter (which describes the zero frequency correlation between variables) that is not estimable from the data. So while we can see which test should be chosen for any particular value of the nuisance parameter, for any given data set, applied researchers do not know the value of this nuisance parameter and hence, which power curve is relevant for their problem. In this paper, the model is rotated so that local power results are based on nuisance parameters that are estimable with sample data and thus informative for applied researchers.

In Section 3, we take up the idea of deriving the optimal tests for the null of no cointegration. 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 are known. From a testing perspective, an unknown cointegrating vector is a nuisance parameter that is identified only under the alternative hypothesis, so optimal tests against the alternative of cointegration take the form of a weighted average test where the weights are over the possible cointegrating vectors. We employ a normal distribution for the weighting of the cointegrating vector in order to provide analytic results. Power functions for optimal tests will depend on the weighting chosen.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

The primary point of the paper is to compare the bounds derived here with currently popular test statistics. This is undertaken in Section 4. We show that existing popular tests do not achieve power bounds for moderate local alternatives; however, the difference depends on the weights and the amount of correlation in the data. Existing system tests (the tests of Johansen (1988), based on the maximum likelihood estimation (MLE), and the ECM-Wald tests, based on the error correction model) have power close to the envelope for more distant alternatives; however, some gains are available for alternatives where the tests have intermediate (closer to 50%) power. We find that the Engle and Granger (1987) test, which is based on the estimated cointegrating vector and 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 a $(k+1)$-dimensional vector time series $z_t = \left(x_t', y_t\right)'$ 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

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

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 Assumptions 1–3 (assumptions on residuals) hold.

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

**Assumption 2.** $|F(r)| = 0$ has roots outside the unit circle.

**Assumption 3.** $\mathrm{E}_{t-1}(\varepsilon_t^*) = 0$ (a.s.), $\mathrm{E}_{t-1}\left(\varepsilon_t^* \varepsilon_t^{*\prime}\right) = \Sigma^*$ (a.s.), and $\sup_t \mathrm{E}\left\|\varepsilon_t^*\right\|^{4+\delta} < \infty$ for some $\delta > 0$, where $\Sigma^*$ is positive definite and $\mathrm{E}_{t-1}(\cdot)$ refers to the expectation conditional on $\left\{\varepsilon_{t-1}^*, \varepsilon_{t-2}^*, \ldots\right\}$.

We are interested in the problem of testing $H_0 : \rho = 1$ versus $H_1 : -1 < \rho < 1$.

Assumptions 1–3 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

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

a stationarity condition, and Assumption 3 is stronger than necessary in imposing conditional heteroskedasticity and implies that $\{\varepsilon_t^*\}$ satisfies a functional central limit theorem (e.g., Phillips and Solo, 1992). Under the maintained hypothesis, $x_t$ is a vector integrated process whose elements are not mutually cointegrated, which follows from (3) and Assumption 2. 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 $-1 < \rho < 1$.

There exist many tests for this null and alternative in the literature. In their seminal paper introducing cointegration, Engle and Granger (1987) suggested 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, Dolado, Hendry, and Smith (1986), Kremers, Ericsson, and Dolado (1992), Boswijk (1994), and Zivot (2000) looked at tests for no cointegration in single equation conditional error correction models; and, finally, the maximum eigenvalue test of Johansen and Juselius (1990) can be used to test this hypothesis.

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 MLEs—is neither locally asymptotically normal nor does it have known optimality properties. The lack of normality means that the usual Wald trade-offs in terms of directions for power may not be particularly interesting. The lack of optimality arises both because of the curvature of the model and because the cointegrating vector is a set of unidentified parameters under the null hypothesis. However, local power curves can be constructed and compared in order to consider which test statistic to use in practice.[1]

Power curves derived from the above model have the nice property that they do not depend on the unknown values for the cointegrating vector, so the researcher does not need to know the particular value of the cointegrating vector when assessing which method to use for a particular empirical application. However, both power and the relative ranking of the test—and hence the choice of method—do rely on the zero frequency correlation between $\Delta u_{1,t}$ and $(1 - \rho L)u_{2,t}$. This is a problem because, in practice, this parameter is not only unknown, but also, under the null and local alternative, there is no obvious way to obtain a good estimate of this parameter. If we consider the case where there are no deterministic terms, we would want to estimate the zero frequency correlation between $\Delta x_t$ and $(1 - \rho L)(y_t - x_t\beta)$, however, when cointegration does not hold ($\rho = 1$), this is simply a rotation of $\Delta x_t$ and $\Delta y_t$, and any identification of this zero-frequency correlation depends on the unknown $\beta$. Under the local alternative, the problem persists because $\rho$ is local to one. Since this correlation strongly affects the ranking of the power curves, the choice between the tests depends on the value of this unknown nuisance parameter, not estimable under the null, and the researcher is left with no idea as to which of these curves is appropriate for any specific application.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

A solution 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. While this is a reasonable solution, the path that we take here is to transform the model into its error correction representation and to consider the zero-frequency correlation matrix in that representation as 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. It does, however, mean that the test will require a weighting over the cointegrating vector, which we find preferable, since economic theory is more likely to be informative for the cointegrating vector than for the zero-frequency correlation.

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

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

where (partitioning after $k$ rows and/or columns) $\alpha=[\alpha_x\ \ \alpha_c]'$, $\bar{\beta}=(1,-\beta')'$, $\Sigma=\mathrm{E}_{t-1}\left(\varepsilon_t\varepsilon_t'\right)=P^{-1}\Sigma^*P^{-1\prime}$, and $\varepsilon_t=P^{-1}\varepsilon_t^*$, where $P=\begin{pmatrix} I_k & 0 \\ -\beta' & 1\end{pmatrix}$. In this rotation, it is possible to consistently estimate the spectral density $\tilde{F}(1)^{-1}\Sigma\tilde{F}(1)^{-1\prime}$ without knowledge of the cointegrating vector. In the following section we will develop methods based on fixing this spectral density.

Regardless of the particular rotation of the model chosen, (3) or (4), 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 $\beta$. Hence, the cointegrating vector is an unidentified nuisance parameter under the null hypothesis. This problem commonly arises in models where unknown parameters appear as products (such as $a\bar{\beta}'$ in the ECM model). In these situations, it is still possible to use standard statistical optimality results based on the Neyman Pearson lemma to maximize power against a weighted average over the unindentified nuisance parameter and obtain tests that will in general depend (and do so here) on the weighting function chosen. Essentially, a test statistic can be constructed that maximizes power against a random alternative where the randomness in the nuisance parameter is exactly that chosen to construct the test statistic.

Since power depends on $\beta$ in the error correction representation, the weighting will matter for power performance. The weighting should thus be chosen to reflect the types of cointegrating vectors that the researcher expects to see. A number of weightings are possible, although, as we show below, a simple test statistic arises with normal or constant weighting across possible cointegrating vectors.

An additional dimension for consideration in constructing tests aimed at generating good power is the direction in which power is to be maximized. 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.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

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 (4) to the (1)–(3) forms. The two representations are equivalent when we set $\varepsilon_t = P^{-1}\varepsilon_t^*$, $\tilde{F}(L) = P^{-1}F(L)P$ and we restrict $\alpha$ so that $\alpha = (0, (\rho - 1))'$. This restriction implies 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, Jansson, and Pesavento, 2005, for a detailed discussion). Essentially, if this restriction does not hold, under the local alternative the nearly integrated cointegrating term is part of the shock to the change in $x_t$, and the change in $x_t$ will 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 $\gamma$ is a fixed constant. Since the error correction term does not enter in (4), under the null hypothesis of no cointegration, this is a restriction only on the model under the alternative and hence is a choice over the direction of power of the tests considered in this paper.

## 3. DERIVATION OF A FAMILY OF TESTS FOR THE NULL OF NO COINTEGRATION

In this section, we derive point optimal tests that maximize weighted average power for various weightings on the unknown cointegrating vector $\beta$. 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 the perfomance, in terms of power, of various currently available tests relatively to the maximal attainable power.

Consider the models (1)–(3) with the deterministic terms included, and having set $F(L) = I$. Define $y = (y_1, \ldots, y_T)', x = (x_1', \ldots, x_T')'$, $u_i = (u_{i,1}', \ldots, u_{i,T}')'$ for $i = 1, 2$, and $D = (D_1', \ldots, D_T')'$. The model can be written in matrix notation as

$$\tilde{P}\begin{pmatrix} \mathrm{vec}(x) \\ y \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(x) \\ y - x\beta \end{pmatrix} = \begin{pmatrix} (I_k \otimes D)\mathrm{vec}(\phi_x) \\ D\phi_c \end{pmatrix} + \begin{pmatrix} \mathrm{vec}(u_1) \\ u_2 \end{pmatrix} = u, \tag{5}$$

where

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

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

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

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

$$(\text{vec}(x)', y') \mapsto (\text{vec}(x)' + \text{vec}(\tilde{\phi}_x)'(I_k \otimes D'), y' + \tilde{\phi}_c' D') \tag{7}$$

for any real conformable $\tilde{\phi}_x, \tilde{\phi}_c$. We have

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

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

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

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

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

where $()^-$ is the generalized inverse. This likelihood function depends on the cointegrating vector $\beta$ through the dependence on $\tilde{P}$ and the value for $\rho$ through $\tilde{A}$.

In order to construct a feasible test, an issue that arises is the dependence of the testing problem on the size of the departure of the largest root in the cointegrating relation from a unit root, i.e., the value of $\rho$. It is well understood by now that in testing for a unit root (see Elliott et al., 1996), no uniformly most powerful test for the null of a unit root exists, even asymptotically. This result directly implies

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

a lack of existence of a uniformly most powerful test for this problem as well, 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 (1988) as employed in Elliott et al. (1996) and construct a family of point optimal tests. This enables the evaluation of maximal power since the tests have optimality properties at the chosen alternative. First, consider the testing problem when $\beta$ is known.

THEOREM 1 (Known β). *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* $\beta, \Sigma$*) invariant to translations, (7) rejects for large values of the statistic*

$$LR_T(\beta) = \exp\left[-\frac{1}{2}\begin{pmatrix} \Sigma_{2.1}^{-1}\left((1-\bar{\rho})^2 c'_{-1} M c_{-1} + 2(1-\bar{\rho}) c'_{-1} M A(1) M \tilde{y}\right) \\ -\begin{pmatrix} \text{vec}(x)' & y' \end{pmatrix} \tilde{M} G \tilde{M} \begin{pmatrix} \text{vec}(x) \\ y \end{pmatrix} \end{pmatrix}\right], \tag{8}$$

*where* $\tilde{y} = y - x\Sigma_{11}^{-1}\Sigma_{12}, c_{-1} = y_{-1} - x_{-1}\beta,\ y_{-1} = (0, y_1, \ldots, y_{T-1}),\ x_{-1} = (0, x'_1, \ldots, x'_{T-1})$, *and*

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

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

*where* $K(r) = \tilde{P}^{-1}\tilde{A}\tilde{P}$, *which is a function of* $r = (\bar{\rho}, 1)$ *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 $\beta$ and the variance covariance matrix of the residuals.

We now turn to the issue of the unknown cointegrating vector $\beta$, a nuisance parameter that arises only under the alternative, by constructing a test that has weighted average power over chosen weights on $\beta$. To keep the difference between the true cointegrating vector, $\beta$, and the hypothesized ones, let the hypothesized cointegrating vector be $b$. Further, define $\tilde{\beta} = \Sigma_{2.1}^{-1/2}\Sigma_{11}^{1/2\prime}\beta$ and $\tilde{b} = \Sigma_{2.1}^{-1/2}\Sigma_{11}^{1/2\prime}b$. 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

$$C_T(v_{\tilde{b}}, \bar{\rho}) = 2\ln\int LR_T(\tilde{b})\, dv_{\tilde{b}}. \tag{9}$$

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

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

The chief difficulty in constructing closed-form statistics in the family $C_T(v_{\tilde{b}}, \bar{\rho})$ is that the statistic of Theorem 1 is a complicated function of the cointegrating vector $b$, and hence the integration over this parameter vector is very complex. However, for the case with no deterministic terms ($G = 0$), 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) provides closed-form tests. In the remainder of the paper, we restrict results to those we can derive analytically, i.e., without deterministic terms. While we recognize that this case is limited in its applicability, this is the first optimality result in the literature for the case of unknown cointegration vector, and we hope this will provide a first step in the direction of deriving a family of optimal tests for more general cases.

For the remainder of the paper, we choose normally distributed weights $b \sim N(b_0, h^{-1}\Sigma_{2.1}(T^{-2}x'_{-1}x_{-1})^{-1})$. Here the choice of the spread of the weights is driven by a chosen parameter $h$, scaled by the Hessian (evaluated at the data without taking expectations). Set the alternative $\bar{\rho} = 1 - T^{-1}\bar{g}$ with $\bar{g}$ fixed and consider the statistic

$$
\begin{aligned}
C_T(\bar{g}, \tilde{b}_0, h) = &-\Sigma_{2.1}^{-1}\left(\bar{g}^2 T^{-2} y'_{-1}y_{-1} + 2\bar{g}T^{-1}y'_{-1}A(1)\tilde{y}\right) \\
&+ \Sigma_{2.1}^{-1}\left(\bar{g}^2 T^{-2} y'_{-1}x_{-1} + \bar{g}T^{-1}\tilde{y}'A(1)'x_{-1}\right) \\
&\quad \times \left((\bar{g}^2 + h)T^{-2}x'_{-1}x_{-1}\right)^{-1} \\
&* \left(\bar{g}^2 T^{-2} y'_{-1}x_{-1} + \bar{g}T^{-1}\tilde{y}'A(1)'x_{-1}\right)' \\
&- \bar{g}^2\left(\frac{h}{\bar{g}^2 + h}\right) T^{-2}\tilde{b}'_0 \Sigma_{11}^{-1/2} x'_{-1}x_{-1}\Sigma_{11}^{-1/2\prime}\tilde{b}_0 \\
&+ 2\Sigma_{2.1}^{-1/2}\left(\frac{h}{\bar{g}^2 + h}\right)\left(\bar{g}^2 T^{-2} y'_{-1}x_{-1}\Sigma_{11}^{-1/2\prime}\tilde{b}_0\right. \\
&\qquad \left. + \bar{g}T^{-1}\tilde{y}'A(1)'x_{-1}\Sigma_{11}^{-1/2\prime}\tilde{b}_0\right), \qquad \textbf{(10)}
\end{aligned}
$$

where, $\tilde{y} = y - x\Sigma_{11}^{-1}\Sigma_{12}$. The weights have been rescaled so $\tilde{b}_0 = \Sigma_{2.1}^{-1/2}\Sigma_{11}^{1/2\prime}b_0$. $C_T(\bar{g}, \tilde{b}_0, h)$ is a natural family of statistics to consider, as it is asymptotically equivalent to the statistic that arises from integrating (9) with the normal weights given above (i.e., completing the square). This is of course just one possible choice. We could also have employed constant weights to approximate the situation where we have no information on the cointegrating case, although due to the nonlinearity of the model this choice is not completely uninformative.[2] It can of course be approximated by choosing a very small $h$. When $h \to \infty$, (10) is equivalent to the family of tests for known $\beta$ of Elliott et al. (2005).

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

The family of test statistics $C_T(\bar{g}, \tilde{b}_0, h)$ (indexed by the choice of the local alternative $\bar{g}$ as well as the choices for $\tilde{b}_0$ and $h$) provides tests for the null of no cointegration against the alternative that a cointegrating vector exists where the cointegrating vector is unknown. With $\gamma = T(1-\rho)$ fixed, the power of a sequence of tests setting $\bar{g} = \gamma$ forms an envelope of highest possible power given the weighting $v_{\tilde{b}}$. The following result establishes the distribution of the test under the null and local alternative for any true cointegrating vector.

THEOREM 2 (Limiting properties of $C_T(\bar{g}, \tilde{b}_0, h)$. *For the model in equations (1)–(3) with $F(L) = I$ where Assumptions 1–3 hold, for $\gamma = T(1-\rho)$ fixed as $T \to \infty$, when the true cointegrating vector is $\beta = \Sigma_{2.1}^{1/2} \Sigma_{11}^{-1/2\prime} \tilde{\beta}$, then*

$$
\begin{aligned}
C_T(\bar{g}) \Rightarrow & -(\bar{g}^2 - 2\bar{g}\gamma) \int \left(W_{2\gamma}\right)^2 - (\bar{g}^2 - 2\bar{g}\gamma)(\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_{1\gamma}' (\tilde{\beta} - \tilde{R}) \\
& - \bar{g}^2 \tilde{\beta}' \int W_1 W_1' \tilde{\beta} \\
& + 2\left(\bar{g}^2 - \bar{g}\gamma\right)(\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_{2\gamma} + 2(\bar{g}^2 - \bar{g}\gamma)(\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_1' \tilde{\beta} \\
& - 2(\bar{g}^2 - \bar{g}\gamma)\tilde{\beta}' \int W_1 W_{2\gamma} - 2\bar{g} \int W_{2\gamma} dW_2 + 2\bar{g}(\tilde{\beta} - \tilde{R})' \int W_{1\gamma} dW_2 \\
& - 2\bar{g}\tilde{\beta}' \int W_1 dW_2 \\
& + \begin{pmatrix} (\bar{g}^2 - \bar{g}\gamma)\left(\int W_{2\gamma} W_1' - (\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_1'\right) \\ + \bar{g}^2 \tilde{\beta}' \int W_1 W_1' + \bar{g} \int dW_2 W_1' \end{pmatrix} \left((\bar{g}^2 + h) \int W_1 W_1'\right)^{-1} \\
& * \begin{pmatrix} (\bar{g}^2 - \bar{g}\gamma)\left(\int W_{2\gamma} W_1' - (\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_1'\right) \\ + \bar{g}^2 \tilde{\beta}' \int W_1 W_1' + \bar{g} \int dW_2 W_1' \end{pmatrix}' \\
& - \bar{g}^2 \left(\frac{h}{\bar{g}^2 + h}\right) \tilde{b}_0' \int W_1 W_1' \tilde{b}_0 \\
& + 2\left(\frac{h}{\bar{g}^2 + h}\right) \begin{pmatrix} (\bar{g}^2 - \bar{g}\gamma)\left(\int W_{2\gamma} W_1' \tilde{b}_0 - (\tilde{\beta} - \tilde{R})' \int W_{1\gamma} W_1' \tilde{b}_0\right) \\ + \bar{g}^2 \tilde{\beta}' \int W_1 W_1' \tilde{b}_0 + \bar{g} \int dW_2 W_1' \tilde{b}_0 \end{pmatrix},
\end{aligned}
$$

*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 $1 \times 1$ standard Brownian motion processes, respectively, and $W_{i\gamma}(\lambda) = -\gamma \int_0^\lambda e^{-\gamma(\lambda - r)} W_i(r)\, dr + W_i(\lambda)$ for $i = 1, 2$.*

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

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

alternative model values for $\tilde{\beta}, \tilde{R}$, and $\gamma$. 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 $\varepsilon_{1t}$ and $\varepsilon_{2t}$. Since, as noted in the previous section, this parameter is consistently estimable under the null and local alternative (as is its analog in more general settings), then a researcher can compute this number to access the relevant power function.

The null distribution is given when $\gamma = 0$ and simplifies to

$$
\begin{aligned}
C_T(\bar{g}) \Rightarrow\ & -\bar{g}^2 \int (W_2)^2 - \bar{g}^2 \tilde{R}' \int W_1 W_1' \tilde{R} - 2\bar{g}^2 \tilde{R}' \int W_1 W_2 \\
& - 2\bar{g} \int W_2\, dW_2 - 2\bar{g}\tilde{R}' \int W_1\, dW_2 \\
& + \left(\bar{g}^2 \left(\int W_2 W_1' + \tilde{R}' \int W_1 W_1'\right) + \bar{g} \int dW_2 W_1'\right)\left((\bar{g}^2 + h)\int W_1 W_1'\right)^{-1} \\
& * \left(\bar{g}^2 \left(\int W_2 W_1' + \tilde{R}' \int W_1 W_1'\right) + \bar{g} \int dW_2 W_1'\right) \\
& - \bar{g}^2 \left(\frac{h}{\bar{g}^2 + h}\right) \tilde{b}_0' \int W_1 W_1' \tilde{b}_0 \\
& + 2\left(\frac{h}{\bar{g}^2 + h}\right)\left(\bar{g}^2 \left(\int W_2 W_1' \tilde{b}_0 + \tilde{R}' \int W_1 W_1' \tilde{b}_0\right) + \bar{g} \int dW_2 W_1' \tilde{b}_0\right).
\end{aligned}
\tag{11}
$$

The distribution depends on the weights chosen along with the alternative $\bar{g}$, and the coefficient describing the correlation structure, $\tilde{R}$, along with the dimension of $x_t$. As $h \to 0$, the last two terms drop and the limit distribution can be shown to be independent of $\tilde{R}$. So for very wide weights over the cointegrating vector, there will be little dependence of the null distribution on the correlation structure. For tighter sets of weights, this is not true.

Because the statistic $C_T(\bar{g}, \tilde{b}_0, h)$ is obtained by integrating the first two terms of (8) across $\tilde{\beta}$ and the third term is zero for the case where $D = 0$, then a test based on $C_T(\bar{g}, \tilde{b}_0, h)$ is asymptotically equivalent to the optimal test under the null and local alternative hypotheses. Theorem 3 formalizes this result.

THEOREM 3 (Constant weights on $\tilde{b}$). *For the testing problem of Theorem 1, when $\bar{\rho} = 1 + T^{-1}\bar{g}$, then under Assumptions 1–3, for the cases where $D = 0$, the family of tests $C_T(\bar{g})$ are asymptotically equivalent to the family of tests (9) under the null hypothesis and the local alternative where $v_{\tilde{b}}$ is set to 1 for all real $\tilde{b}$.*

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 $\tilde{b}$ is drawn from the distribution used to construct the weighting function.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

The assumption of a density for the innovations of the model, 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). A researcher who 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).

## 4. EVALUATING THE TESTS

In Section 4 we numerically evaluate the upper power bounds derived in Section 3 and compare them with popularly employed tests for this null hypothesis. Figure 1 plots the rejections rates for $C_T(\bar{g})$ when $\bar{g} = \gamma$, for various $\delta$ when we choose $h = 0.5$. This is the power bound or the maximum attainable power at each point in the alternative when $\beta$ is drawn from a normal distribution. The curves are produced using normal step functions to approximate the asymptotic distributions derived in Theorem 2, where $\widetilde{R} = \delta\left(1-\delta^2\right)^{1/2}$, and $\widetilde{\beta}$ is drawn from a normal distribution with mean $\beta_0 = 1$ and variance $h^{-1}\left(\int W_1 W_1'\right)^{-1}$, the limit of $h^{-1}\Sigma_{2.1}(T^{-2}x_{-1}'x_{-1})^{-1}$. Figure 1 shows that the power bounds depend on $\delta$, and that the effect of $\delta$ is not symmetric. The power is greater for $\delta$ closer to $-1$, and falls as $\delta$ nears $+1$.

The remaining figures are designed to examine how close the power of commonly applied tests is compared to the maximal power envelopes. Results are from univariate ($k = 1$) models. Under the null hypothesis, the data are drawn as random walks with mean zero normal innovations with variance covariance matrix $\Sigma$, which defines $\tilde{R}$. We set $\Sigma_{11} = \Sigma_{22} = 1$, and vary $\Sigma_{12} = \delta$. Under the local alternative, we draw residuals as under the null; however, now the data are drawn from a VAR(1) with coefficient matrix

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

where $\rho = 1 - \gamma T^{-1}$. We draw $\beta$ from $N(\beta_0, h^{-1}\Sigma_{2.1}(T^{-2}x_{-1}'x_{-1})^{-1})$ with $\beta_0 = 1$, and we construct power curves against this random alternative. We employ 10,000 Monte Carlo replications and a sample of $T = 1{,}000$ to approximate large sample performance. Results are presented for various values of $\delta$, and $h$. Figures 2–7 present the rejection rates of the Engle and Granger (1987) test (EG), Boswijk's (1994) Wald test (WALD), the maximum eigenvalue test of Johansen and Juselius (1990) (JOH), and the Banerjee et al. (1986) test on the error correction equation (ECR). We also plot the power bound $C_T(\gamma)$ (MAX in the graphs).

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

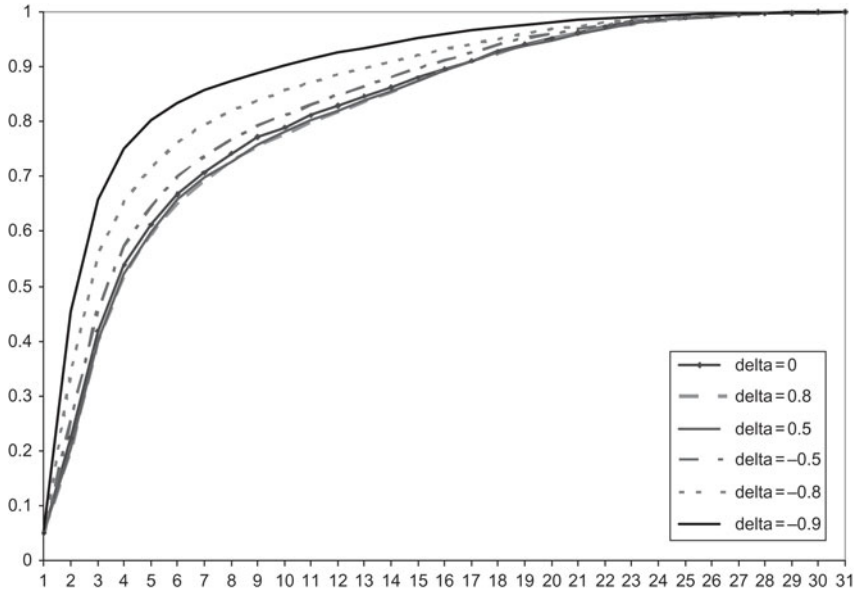



**FIGURE 1.** Power bounds, $h = 0.5$, $\beta_0 = 1$.

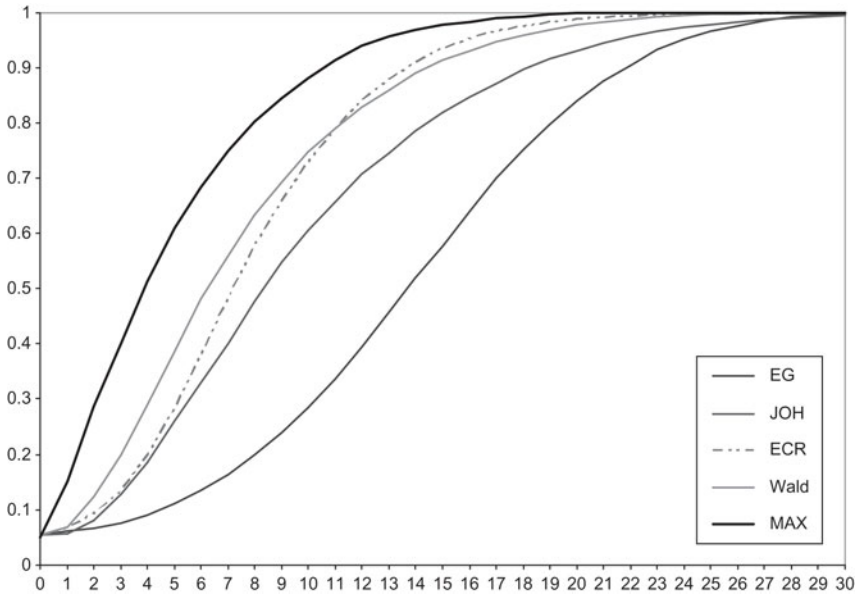



**FIGURE 2.** Power curves, $\delta = 0$, $h = 10$, $\beta_0 = 1$.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

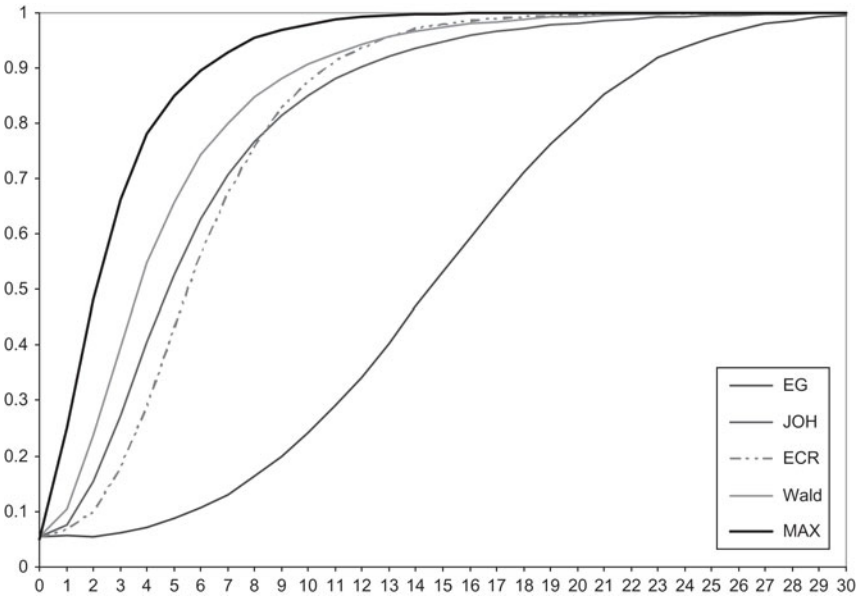



**FIGURE 3.** Power curves, $\delta = -0.5$, $h = 10$, $\beta_0 = 1$.

None of the existing tests achieves the power bound for alternatives close to the null. How far the tests are from the power bounds depends on $\delta$ and $h$. Figures show that the current employed test can be far from the power bounds for alternatives where the tests have intermediate (closer to 50%) power, especially for $\delta$ zero or positive. The magnitude of the power curves against local alternatives is affected by the tightness of the weight distribution. Comparing Figure 5 with Figure 6 and Figure 4 with Figure 7, we see that for large $h$, i.e., when the cointegrating vectors are more tightly distributed around a single value and hence close to the "known" cointegrating vector case, the power curves of usual tests are lower, while the power bounds are higher: The distance between the local power of the currently used tests and the maximal power gets larger as the uncertainty about the cointegration vector diminishes. At the limit in which $h \to \infty$, we are back to the situation in which $\beta$ is known and the difference between the test and the power envelope is large, as shown in Elliott et al. (2005).

Overall, it is clear that the power of the EG test is far from maximal. As for the system tests, while they are not too far from the maximal power in cases in which we have no information on likely values of the cointegration vector and more distant alternatives, the maximal attainable power can be higher as soon as we introduce information on $\beta$ in the form of a tighter prior distribution, and therefore the difference can be larger.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

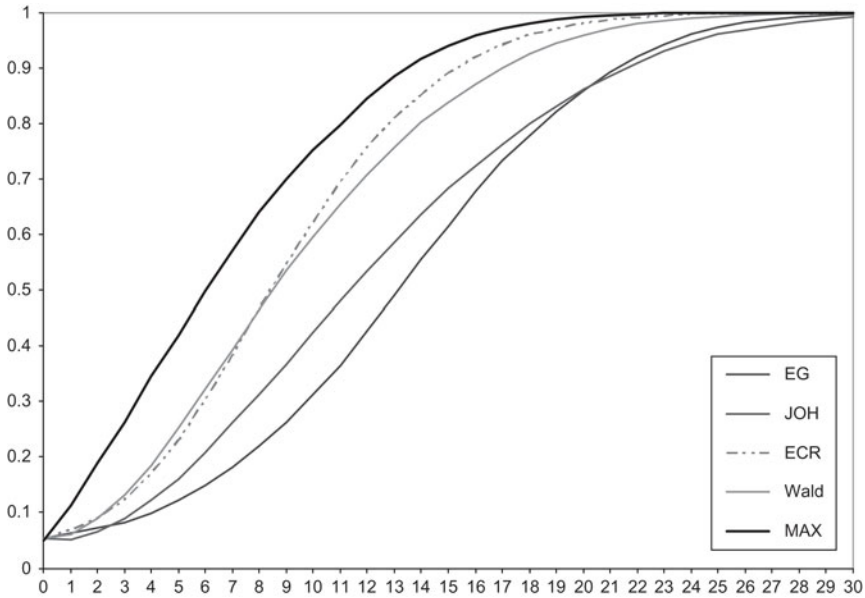



**FIGURE 4.** Power curves, $\delta = 0.5$, $h = 10$, $\beta_0 = 1$.

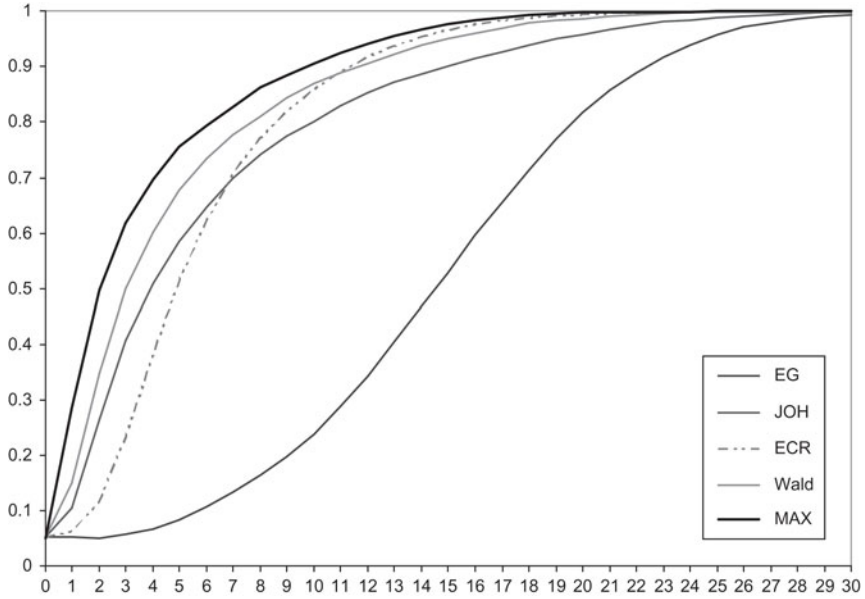



**FIGURE 5.** Power curves, $\delta = -0.5$, $h = 1$, $\beta_0 = 1$.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

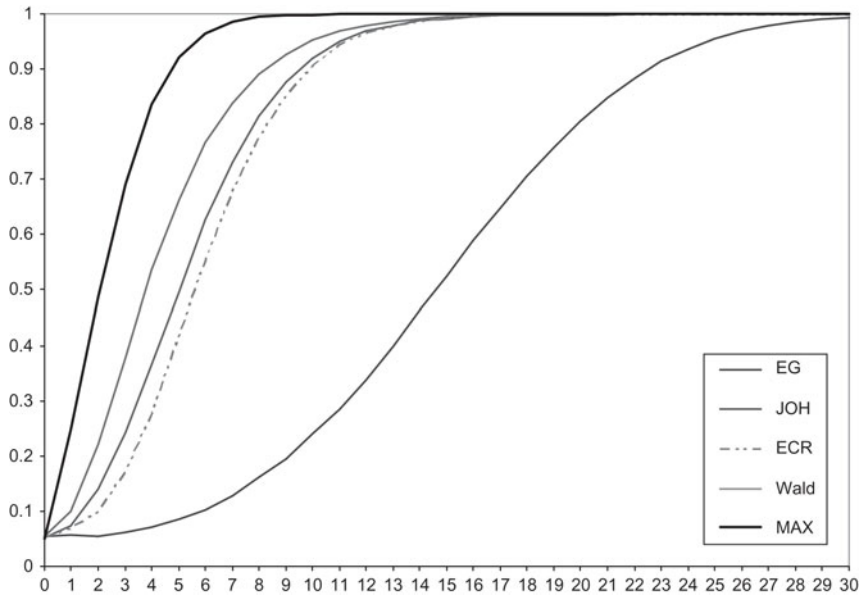



**FIGURE 6.** Power curves, $\delta = -0.5$, $h = 100$, $\beta_0 = 1$.

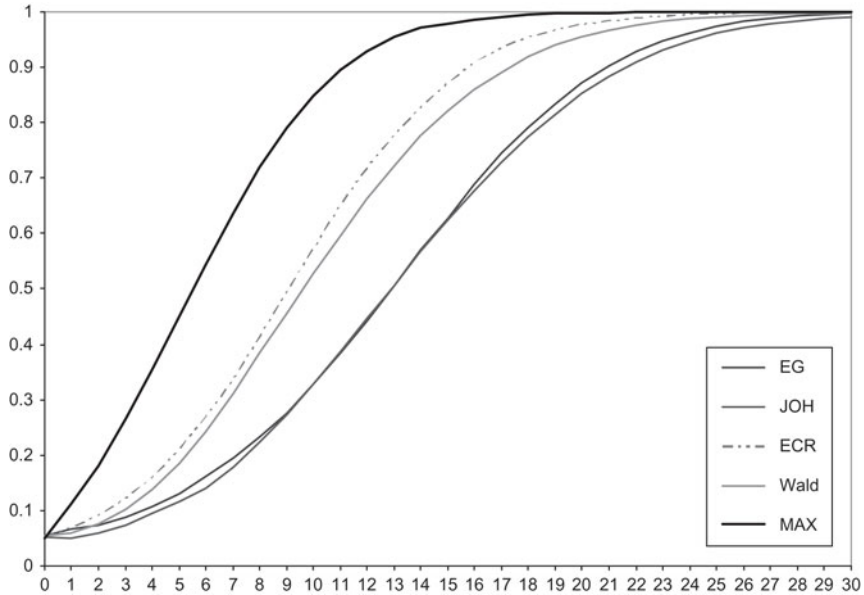



**FIGURE 7.** Power curves, $\delta = 0.5$, $h = 100$, $\beta_0 = 1$.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

## 5. CONCLUSION

Tests for the null of no cointegration are nonstandard in a number of ways. First, as with tests for a unit root null, there is no uniformly most powerful test for this hypothesis (in fact, this is the case even when the cointegrating vector is known). Second, the unknown cointegrating vector does not appear under the null hypothesis, adding the additional issue of unidentified parameters under the null. For both of these reasons we do not expect tests based on MLEs or regression approaches to have any optimality properties (at least none follow from the usual arguments). This paper sets out to compute upper bounds for the power curves of tests for this hypothesis to evaluate whether tests commonly used in the literature are close to optimal in a practical sense. We first characterize optimal tests in this context. As is typical with a nuisance parameter not known under the null distribution, the power depends on the weighting over possible sets of nuisance parameters — here, weightings over possible cointegrating vectors. We provide analytic solutions for normal weightings (including the usual limiting case of a flat prior) for the no deterministic case, and we show that the family of optimal tests depends asymptotically on the correlation structure of the model, summarized by a vector with dimension equal to that of the cointegrating vector.

Averaging across possible cointegrating vectors in accordance with the normal weightings over potential vectors, we are able to show that, for a sufficiently tight weighting, commonly employed full system tests, such as the Johansen (1988) trace test and the ECM-Wald test, have power close to the maximal power for distant alternatives; however, some gains are available for alternatives where the tests have intermediate (closer to 50%) power. The Engle and Granger (1987) test, instead, has power far below the power bound for the experiments examined.

### NOTES

1. See Kremers et al. (1992), Zivot (2000), Banerjee et al. (1986), Johansen (1995), and Pesavento (2004, 2007).

2. Since in the ECM form the cointegrating vector appears as a product, the model is nonlinear in the parameters.

### REFERENCES

Andrews, D. & W. Ploberger (1994) Optimal tests when the nuisance parameter is only present under the alternative. *Econometrica* 62, 1383–1414.

Banerjee, A., J.J. Dolado, D.F. Hendry, & 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.

Davies, R.B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

Davies, R.B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

Elliott, G. & M. Jansson (2003) Testing for unit roots with stationary covariates. *Journal of Econometrics* 115, 75–89.
Elliott, G., M. Jansson, & E. Pesavento (2005) Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. *Journal of Business & Economic Statistics* 23:1, 34–48.
Elliott, G., T.J. Rothenberg, & J. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
Engle, R.F. & C.W. Granger (1987) Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251–276.
Granger, C.W. & P. Newbold (1974) Spurious regressions in econometrics. *Journal of Econometrics* 2, 111–120.
Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64(1), 413–430.
Jansson, M. (2007) Semiparmetric Power Envelopes for Tests of the Unit Root Hypothesis. Mimeo, UC Berkeley.
Johansen, S. (1988) Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control* 12, 231–254.
Johansen, S. (1995) *Likelihood Based Inference in Cointegrated Autoregressive Models*. Oxford University Press.
Johansen, S. & 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.
King, M. (1988) Towards a theory of point optimal testing. *Econometric Reviews* 6, 169–218.
Kremers, J.M., N.R. Ericsson, & J.J. Dolado (1992) The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* 54, 325–348.
Mueller, U.K. (2007) An Alternative Sense of Asymptotic Efficiency. Mimeo, Princeton University.
Mueller, U.K. & G. Elliott (2003) Tests for unit roots and the initial observation. *Econometrica* 71, 1269–1286.
Pesavento, E. (2004) An analytical evaluation of the power of tests for the absence of cointegration. *Journal of Econometrics* 122(2), 349–384.
Pesavento, E. (2007) Residuals based tests for the null of no cointegration: An analytical comparison. *Journal of Time Series Analysis* 28(1), 111–137.
Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33(3), 311–340.
Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residuals based tests for cointegration. *Econometrica* 58, 165–193.
Phillips, P.C.B. & V. Solo (1992) Asymptotics of linear processes. *Annals Of Statistics* 20, 971–1001.
Rothenberg, T.J. & 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.

# APPENDIX

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

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

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

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

where $K(r) = \tilde{P}^{-1}\tilde{A}\tilde{P}$ and $\tilde{D} = I_{k+1} \otimes D$.

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 generalized inverse for $\tilde{M}\tilde{P}^{-1}\tilde{A}^{-1}\tilde{P}\tilde{\Sigma}\tilde{P}'\tilde{A}^{-1'}\tilde{P}^{-1'}\tilde{M}$, where $\tilde{A}$ is evaluated at $r$ instead of $\rho$.

The likelihood can be written as

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

Hence, the likelihood ratio test is proportional to

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

By definition,

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

We can write (for $r = \bar{\rho}, 1$)

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

where $B$ is a backward shift matrix

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

and

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

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

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

We have, for the first term in (A.1),

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

The first of these terms is

$$\begin{aligned}
&(1-\bar{\rho})^2\begin{pmatrix}\mathrm{vec}(Mx)' & y'M\end{pmatrix}K_d'\tilde{\Sigma}^{-1}K_d\begin{pmatrix}\mathrm{vec}(Mx)\\ My\end{pmatrix}\\
&\quad=\Sigma_{2.1}^{-1}(1-\bar{\rho})^2\begin{pmatrix}\mathrm{vec}(Mx)' & y'M\end{pmatrix}\left(\begin{pmatrix}-\beta\\ 1\end{pmatrix}\begin{pmatrix}-\beta' & 1\end{pmatrix}\otimes BB'\right)\begin{pmatrix}\mathrm{vec}(Mx)\\ My\end{pmatrix}\\
&\quad=\Sigma_{2.1}^{-1}(1-\bar{\rho})^2(BMy-BMx\beta)'(BMy-BMx\beta)\\
&\quad=\Sigma_{2.1}^{-1}(1-\bar{\rho})^2c_{-1}'Mc_{-1},
\end{aligned}$$

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

Partition the matrix $\Sigma^{-1}$ into blocks $\Sigma^{ij}$ for $i,j=1,2$ where the partition is after the $k^{\text{th}}$ row and column. For the second and third terms, we have

$$\begin{aligned}
&(1-\bar{\rho})\begin{pmatrix}\mathrm{vec}(Mx)' & y'M\end{pmatrix}K(1)'\tilde{\Sigma}^{-1}K_d\begin{pmatrix}\mathrm{vec}(Mx)\\ My\end{pmatrix}\\
&\quad=(1-\bar{\rho})\begin{pmatrix}\mathrm{vec}(Mx)' & y'M\end{pmatrix}\begin{pmatrix}\left(\Sigma^{12}\otimes A(1)'\right)(-\beta'\otimes B) & \Sigma^{12}\otimes A(1)'B\\ \Sigma^{22}A(1)'(-\beta'\otimes B) & \Sigma^{22}A(1)'B\end{pmatrix}\begin{pmatrix}\mathrm{vec}(Mx)\\ My\end{pmatrix}\\
&\quad=(1-\bar{\rho})\Big(-\Sigma^{12\prime}x'MA(1)'BMx\beta+\Sigma^{22}y'MA(1)'BMy-\Sigma^{22}y'MA(1)'BMx\beta\\
&\qquad\qquad+\Sigma^{12\prime}x'MA(1)'BMy\Big)\\
&\quad=(1-\bar{\rho})\left(\Sigma^{22}y'MA(1)'Mc_{-1}+\Sigma^{12\prime}x'MA(1)'Mc_{-1}\right)\\
&\quad=(1-\bar{\rho})\Sigma_{2.1}^{-1}\left(y'-\Sigma_{12}'\Sigma_{11}^{-1}x'\right)MA(1)'Mc_{-1}\\
&\quad=(1-\bar{\rho})\Sigma_{2.1}^{-1}\tilde{y}'MA(1)'Mc_{-1}.
\end{aligned}$$

Putting this together yields

$$\begin{aligned}
&\begin{pmatrix}\mathrm{vec}(Mx)\\ My\end{pmatrix}'\left(K(\bar{\rho})'\tilde{\Sigma}^{-1}K(\bar{\rho})-K(1)'\tilde{\Sigma}^{-1}K(1)\right)\begin{pmatrix}\mathrm{vec}(Mx)\\ My\end{pmatrix}\\
&\quad=(1-\bar{\rho})^2\Sigma_{2.1}^{-1}c_{-1}'Mc_{-1}+2(1-\bar{\rho})\Sigma_{2.1}^{-1}c_{-1}'MA(1)M\tilde{y}.
\end{aligned}$$

The result follows with the remaining terms in $G$. ■

Consider the general case in which serial correlation is present. When the errors satisfy Assumptions 1–3, we replace $\varepsilon_t$ in the previous subsection with $\upsilon_t=\tilde{F}(L)^{-1}\varepsilon_t$. We are

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

able to consistently estimate both $\Omega = \lim_{T\to\infty} \mathrm{E}\left[\left(T^{-1/2}\sum_{t=1}^{T} v_t\right)\left(T^{-1/2}\sum_{t=1}^{T} v_t\right)'\right]$ and $\Lambda = \lim_{T\to\infty} T^{-1}\sum_{t-2}^{T}\sum_{j-1}^{t-1}\mathrm{E}[v_j v_t']$. From Assumptions 1–3, suitably scaled partial sums of the innovations 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 $kx1$ standard Brownian motion. It follows directly that $T^{-1/2}\sum_{t=1}^{[T\lambda]} Pv_t = T^{-1/2}\sum_{t=1}^{[T\lambda]} v_t^* \Rightarrow P\Omega^{1/2}W(\lambda)$ where

$$P\Omega^{1/2}W(\lambda) = \begin{pmatrix} I_k & 0 \\ -\beta' & 1 \end{pmatrix}\begin{pmatrix} \Omega_{11}^{1/2} & 0 \\ \Omega_{12}'\Omega_{11}^{-1/2\prime} & \Omega_{2.1}^{1/2} \end{pmatrix}\begin{pmatrix} W_1(\lambda) \\ W_2(\lambda) \end{pmatrix}$$

$$= \begin{pmatrix} \Omega_{11}^{1/2}W_1(\lambda) \\ \Omega_{2.1}^{1/2}\left(W_2(\lambda) - (\tilde{\beta}-\tilde{R})'W_1(\lambda)\right) \end{pmatrix},$$

and the partition is after the $k^{\text{th}}$ row and column and $\tilde{R} = \Omega_{11}^{-1/2}\Omega_{12}\Omega_{2.1}^{1/2}$. It also follows from Assumptions 1–3 that $T^{-1}\sum_{t=2}^{T}\xi_{t-1}v_t' \Rightarrow \Omega^{1/2}\int W(\lambda)\,dW(\lambda)' + \Lambda$.

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 $T^{-1/2}u_{2[T\lambda]} \Rightarrow \Omega_{2.1}^{1/2}\left(W_{2\gamma}(.) - (\tilde{\beta}-\tilde{R})'W_{1\gamma}(.)\right)$, where $W_{i\gamma}(.)$ are Ornstein Uhlenbeck processes constructed from $W_i(.)$ for $i = 1, 2$. Define $u_{i,-1} = (0, u_{i1}, \ldots, u_{iT-1})$. 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 $T^{-1/2}\sum_{t=2}^{[T.]}\tilde{v}_{2t} \Rightarrow \Omega_{2.1}^{1/2}W_2(.)$.

A number of results used in proving the theorems follow directly. First, $T^{-2}u_{1,-1}'u_{1,-1} \Rightarrow \Omega_{11}^{1/2}\int W_1(\lambda)W_1(\lambda)'\Omega_{11}^{1/2\prime}d\lambda$ under both $H_0$ and $H_1$.

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

**Proof of 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 $T^{-2}\Omega_{2.1}^{-1}y_{-1}'y_{-1} = T^{-2}\Omega_{2.1}^{-1}\left(u_{2,-1}'u_{2,-1} + \tilde{\beta}'\Omega_{2.1}\Omega_{11}^{-1/2\prime}u_{1,-1}'u_{1,-1}\Omega_{11}^{-1/2}\tilde{\beta} + 2\Omega_{2.1}^{1/2}u_{2,-1}'u_{1,-1}\Omega_{11}^{-1/2}\tilde{\beta}\right)$, where $\tilde{\beta} = \Omega_{2.1}^{1/2}\Omega_{11}\beta$. Limit results follow from the results above for each of these pieces. Similarly, the terms $T^{-2}\Omega_{11}^{-1/2}u_{1,-1}'u_{1,-1}\Omega_{11}^{-1/2\prime}$ and $T^{-2}\Omega_{2.1}^{-1/2}u_{2,-1}'u_{1,-1}\Omega_{11}^{-1/2\prime}$ follow directly. For the terms involving $A(1)\tilde{y}$, notice that $\tilde{y} = (u_2 + \Omega_{2.1}^{1/2}u_1\Omega_{11}^{-1/2\prime}(\tilde{\beta}-\tilde{R})) = u_2^*$. For $i = 1, 2$, $T^{-1}u_{i,-1}'A(1)u_2^* = T^{-1}u_{i,-1}'A(1)\left(u_2 + \Omega_{2.1}^{1/2}u_1\Omega_{11}^{-1/2\prime}(\tilde{\beta}-\tilde{R})\right) = T^{-1}u_{i,-1}'A(\rho)u_2 - \gamma T^{-2}u_{i,-1}'u_{2,-1} + \Omega_{2.1}^{1/2}T^{-1}u_{i,-1}'A(1)u_1\Omega_{11}^{-1/2\prime}(\tilde{\beta}-\tilde{R})$.

Limit results for $\gamma T^{-2}u_{i,-1}'u_{2,-1}$ follow from the results above.

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press

Since $T^{-1}u'_{i,-1}A(\rho)u_2 + \Omega_{2.1}^{1/2}T^{-1}u'_{i,-1}A(1)u_1\Omega_{11}^{-1/2\prime}(\tilde{\beta}-\tilde{R}) = T^{-1}u'_{i,-1}\tilde{v}_{2t}$, we have that, for $i=1$,

$$T^{-1}u'_{1,-1}A(\rho)u_2 + \Omega_{2.1}^{1/2}T^{-1}u'_{1,-1}A(1)u_1\Omega_{11}^{-1/2\prime}(\tilde{\beta}-\tilde{R})$$
$$\Rightarrow \Omega_{2.1}^{1/2}\Omega_{11}^{1/2}\int W_1(\lambda)\,dW_2(\lambda) + \Lambda_{12},$$

and, for $i=2$,

$$T^{-1}u'_{2,-1}A(\rho)u_2 + \Omega_{2.1}^{1/2}T^{-1}u'_{2,-1}A(1)u_1\Omega_{11}^{-1/2\prime}(\tilde{\beta}-\tilde{R})$$
$$\Rightarrow \Omega_{2.1}\int (W_{2\gamma} - (\tilde{\beta}-\tilde{R})'W_{1\gamma})\,dW_2(\lambda) + \Lambda_{22}.$$

Putting these terms together and setting $\Omega=\Sigma$ and $\Lambda=0$ yields the stated result. The null distribution follows from setting $\gamma=0$, so $\int W_{2\gamma} = \int W_2$, etc. ■

**Proof of 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. ■

https://doi.org/10.1017/S026646660999034X Published online by Cambridge University Press