UC San Diego
Recent Work

Title
Optimal Power for Testing Potential Cointegrating Vectors with Known

Permalink
https://escholarship.org/uc/item/2bv7n071

Authors
Elliott, Graham
Jansson, Michael
Pesavento, Elena

Publication Date
2004-06-01

Optimal Power for Testing Potential Cointegrating Vectors with Known Parameters for Nonstationarity[*]

GRAHAM ELLIOTT

UNIVERSITY OF CALIFORNIA, SAN DIEGO

9500 GILMAN DRIVE,

LA JOLLA, CA, 92093-0508

MICHAEL JANSSON   ELENA PESAVENTO

UNIVERSITY OF CALIFORNIA, BERKELEY   EMORY UNIVERSITY

BERKELEY, CA, 94720-3880   ATLANTA, GA, 30322

June 2004

ABSTRACT. Theory often specifies a particular cointegrating vector amongst integrated variables and it is often required that one test for a unit root in the known cointegrating vector. Although it is common to simply employ a univariate test for a unit root for this test, it is known that this does not take into account all available information. We show here that in such testing situations a family of tests with optimality properties exists. We use this to characterize the extent of the loss in power from using popular methods, as well as to derive a test that works well in practice. We also characterize the extent of the losses of not imposing the cointegrating vector in the testing procedure. We apply various tests to the hypothesis that price forecasts from the Livingston data survey are cointegrated with prices, and find that although most tests fail to reject the presence of a unit root in forecast errors the tests presented here strongly reject this (implausible) hypothesis.

Keywords: Cointegration, optimal tests, unit roots.

---

1. INTRODUCTION

This paper examines tests for cointegration when the researcher knows the cointegrating vector a-priori and when the 'X' variables in the cointegrating regression are known to be integrated of order one (I(1)). In particular, we characterize a family of optimal tests for the null hypothesis of no cointegration when there is one cointegrating vector. This enables us to examine the loss in power from using either suboptimal methods (such as univariate unit root tests on the cointegrating vector) and also losses that arise from testing cointegration with estimated cointegrating vectors.

There are a number of practical reasons for interest in this family of tests. First, in many applications a potential cointegrating vector is specified by economic theory (see Zivot (2000) for a list of examples), and researchers are confident or willing to assume that variables are I(1). The test of interest then becomes testing that the implied cointegrating vector has a unit root (which would falsify the theory). The empirical strategy commonly followed is simply to construct the potential cointegrating vector and employ a univariate test for a unit root. However, this method avoids using useful information in the original multivariate model that could lead to more powerful tests (see Zivot (2000)). While there are tests available that do exploit this extra information in the problem (e.g. those in Horvath and Watson (1995), Johansen (1988, 1995), Kremers, Ericsson and Dolado (1992), and Zivot (2000)), these tests do not use this information optimally. The class of tests suggested below, identical apart from the treatment of deterministic terms to those in Elliott and Jansson (2003) for testing for unit roots with stationary covariates, do have optimality properties.

Second, the optimal family we derive allows the power bound of such tests to be derived. This is interesting in the sense that it gives an objective for examining the loss of power in estimating rather than specifying the cointegrating vector. A quantitative understanding of this loss and how it varies with nuisance parameters of the model is

1

important for understanding differences in empirical results. If one researcher specifies the parameters of the cointegrating vector and rejects, while another estimates the vector and fails to reject, we are more certain this is likely to be due to loss of power when there are large losses in power from estimating the cointegrating vector. If the power losses were small, then we would probably conclude that the imposed parameters are in error. By deriving the results analytically we are able to say what types of models (or more concretely what values for a certain nuisance parameter) are likely to be related to large or small power losses in estimating the cointegrating vector. For many values of the nuisance parameter (which is consistently estimable and is produced as a by-product of the test proposed herein) the differences in power is large.

The next section presents our model and relates it to error correction models (ECM). In the third section we consider tests for cointegration when the cointegrating vector is known. We discuss a number of approaches that have been used in the literature and present the methods of Elliott and Jansson (2003) in the context of this problem. Section four presents numerical results to show the asymptotic and small sample performances of the Elliott and Jansson (2003) test relative to others in the literature. An empirical application relating to the cointegration of forecasts and their outcomes is described in the fifth section.

2. THE MODEL AND ASSUMPTIONS

We consider the case where a researcher observes an $(m + 1)$-dimensional vector time series $z_t = (y_t, x'_t)'$ generated by the triangular model

$$y_t \;=\; \mu_y + \tau_y t + \gamma' x_t + u_{y,t} \qquad\qquad (1)$$

$$x_t \;=\; \mu_x + \tau_x t + u_{x,t} \qquad\qquad (2)$$

and

$$A\left(L\right)\begin{pmatrix} \left(1 - \rho L\right) u_{y,t} \\[2mm] \Delta u_{x,t} \end{pmatrix} = \varepsilon_t, \qquad\qquad (3)$$

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

Assumption 1: $\max_{-k \leq t \leq 0} \left\| \left(u_{y,t}, u'_{x,t}\right)' \right\| = O_p\left(1\right)$, where $\|\cdot\|$ is the Euclidean norm.

Assumption 2: $\left|A\left(r\right)\right| = 0$ has roots outside the unit circle.

Assumption 3: $E_{t-1}\left(\varepsilon_t\right) = 0$ (a.s.), $E_{t-1}\left(\varepsilon_t \varepsilon'_t\right) = \Sigma$ (a.s.), and $\sup_t E \left\|\varepsilon_t\right\|^{2+\delta} < \infty$ for some $\delta > 0$, where Σ is positive definite, and $E_{t-1}\left(\cdot\right)$ refers to the expectation conditional on $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

We are interested in the problem of testing

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

Under the maintained hypothesis, x_t is a 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 $y_t - \gamma' x_t = \mu_y + +\tau_y t + u_{y,t}$ mean reverts to its deterministic component when $-1 < \rho < 1$. We assume that the researcher knows the value of γ, the parameter that characterizes the potentially cointegrating relation between y_t and x_t. This assumption is plausible in many empirical applications, including the one discussed in Section five of this paper. Various assumptions on μ_y, μ_x, τ_y, and τ_x will be entertained.

Assumptions 1-3 are fairly standard and are the same as (A1)-(A3) of Elliott and Jansson (2003). 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 are a number of different VAR-type representations of the model in $(1)-(3)$. Ignoring deterministic terms for the sake of exposition, we now present three such representations, each of which sheds light on the properties of our model and the precise restrictions of the formulation of the problem above. The restrictions are precisely those embodied in the idea that x_t is known to be $I(1)$ under both the null and alternative hypotheses.

A general error correction model (ECM) representation for the data is

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

where $\beta = (1, -\gamma')'$. Comparing this to the (1) form we note that the two representations are equivalent when $\alpha = ((\rho - 1), 0)'$ where α is $(m+1) \times 1$ and also

$$\tilde{A}(L) = A(L)\begin{pmatrix} 1 & -\gamma' \\ 0 & I_m \end{pmatrix}.$$

4

These are precisely the restrictions on the full system that impose the assumed knowledge over the system we are exploiting and also the correct normalization on the root of interest when testing for a unit root in the known cointegrating vector .

First, since $\tilde{A}(L)$ is a rotation of $A(L)$ then the roots of each lag polynomial are equivalent and hence Assumption 2 rules out the possibility of other unit roots in the system. Second, the normalization of the first element of α to $(1-\rho)$ merely ensures that the root that we are testing is correctly scaled under the alternative.

The important restriction is setting elements 2 through $m+1$ of α to zero, which is the typical assumption in the triangular form. 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. 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 $\rho = 1 + T^{-1}c$, where c is a fixed constant.

To see this, we show that if z_t is generated by (4) and Assumptions 1-3 hold, then x_t is I(1) under local alternatives only if $\alpha = (\rho-1)(1,0')'$. Now suppose z_t is generated by (4) and let $\alpha = (\rho-1)(1,\alpha_x')'$. Then

$$
\Delta \begin{pmatrix} \beta' z_t \\ x_t \end{pmatrix} = (\rho-1) \begin{pmatrix} 1 - \gamma' \alpha_x \\ \alpha_x \end{pmatrix} \beta' z_{t-1} + \begin{pmatrix} v_{y,t} \\ v_{x,t} \end{pmatrix}, \tag{5}
$$

where $v_t = (v_{y,t}, v_{x,t}')' = A(L)^{-1}\varepsilon_t$. Using the fact that $T^{-1/2}\sum_{t=1}^{[T\cdot]} v_t \Rightarrow B(\cdot)$, where $B = (B_y, B_x')'$ is a Brownian motion with covariance matrix $\Omega = A(1)^{-1}\Sigma A(1)^{-1\prime}$ and $\Rightarrow$ denotes weak convergence, it is easy to show that $T^{-1/2}\sum_{t=1}^{[T\cdot]} (\beta' z_t) \Rightarrow B_y^c(\cdot)$ and hence

$$T^{-1/2} x_{[T\cdot]} \quad = \quad c\alpha_x T^{-1} \sum_{t=1}^{[T\cdot]} T^{-1/2} \left(\beta' z_{t-1}\right) + T^{-1/2} \sum_{t=1}^{[T\cdot]} v_{x,t}$$

$$\Rightarrow \quad c\alpha_x \int_0^{\cdot} B_y^c(s)\, ds + B_x(\cdot)$$

under local alternatives $\rho = 1 + T^{-1}c$, where $B_y^c(r) = \int_0^r \exp\left(c\left(r - s\right)\right) dB_y(s)$. The process on the last line is a Brownian motion if and only if $c\alpha_x = 0$. As a consequence, x_t is I(1) under local alternatives if and only if $\alpha_x = 0$, as claimed. It follows from the foregoing discussion that if α_x was nonzero in (5), then we would have that x_t is I(1) under the null hypothesis while under the alternative hypothesis we would suddenly have a small but asymptotically non-negligible persistent component in Δx_t. This would be an artificial difference between the null and alternative models that is unlikely to map into any real life problem.

A second form of the model is the error correction model (ECM) representation of the model

$$\Delta z_t = \alpha^* \beta' z_{t-1} + A^*(L) \Delta z_{t-1} + \varepsilon_t, \tag{6}$$

where $A^*(L)$ is a lag polynomial of order k. Under the restrictions we are imposing

$$\alpha^* = \tilde{A}(1)\alpha = (\rho - 1) \begin{pmatrix} A_{11}(1) \\ A_{21}(1) \end{pmatrix} \tag{7}$$

where we have partitioned $A(L)$ after the first row and column. From this formulation we are able to see that x_t is weakly exogenous for γ if and only if $A(1)$ is block upper

6

triangular when partitioned after the first row and column (i.e. elements 2 through $m+1$ of the first column of $A(1)$ are equal to zero). Thus the 'directional' restriction we place on α and the restriction on the ECM implicit in the triangular representation are distinct from the assumption of weak exogeneity. They only become equivalent when there is no serial correlation (i.e. when $A(L) = I_{m+1}$). We do not impose weak exogeneity in general.

Finally, our model can be written as

$$A(L) \begin{pmatrix} (1 - \rho L) Y_t \\ X_t \end{pmatrix} = \varepsilon_t, \tag{8}$$

where $Y_t = y_t - \gamma' x_t$ and $X_t = \Delta x_t$. In other words, the model can be represented as a VAR model of the form examined in Elliott and Jansson (2003). Apart from deterministic terms, the testing problem studied here is therefore isomorphic to the unit root testing problem studied in Elliott and Jansson (2003). In the next section, we utilize the results of that paper to construct powerful tests for the testing problem under consideration here.

3. Testing Potential Cointegrating Vectors with Known Parameters for Nonstationarity

3.1. Existing Methods. There are a number of tests derived for the null hypothesis considered in this paper. An initial approach (for the PPP hypothesis, see for example Cheung and Lai (2000); for income convergence, see Greasley and Oxley (1997) and, in a multivariate setting, Bernard and Durlauf (1995)) was to realize that with γ known one could simply undertake a univariate test for a unit root in $y_t - \gamma' x_t$. Any univariate test for a unit root in $y_t - \gamma' x_t$ is indeed a feasible and consistent test, however this amounts to examining (1) ignoring information in the

7

remaining m equations in the model. As is well understood in the stationary context, correlations between the error terms in such a system can be exploited to improve estimation properties and the power of hypothesis tests. For the testing problem under consideration here, we have that under both the null and alternative hypotheses the remaining m equations can be fully exploited to improve the power of the unit root test. Specifically, there is extra exploitable information available in the 'known' stationarity of Δx_t. (That such stationary variables can be utilized to improve power is evident from the results of Hansen (1995) and Elliott and Jansson (2003)). The key correlation that will describe the availability of power gains is the long-run ('zero frequency') correlation between $\Delta u_{y,t}$ and $\Delta u_{x,t}$. In the case where this correlation is zero, an optimal univariate unit root test is optimal for this problem. Outside of this special case many tests have better power properties. There is still a small sample issue — univariate tests require fewer estimated parameters. This is analyzed in small sample simulations in section 4.4.

For a testing problem analogous to ours, Zivot (2000) employs the covariate augmented Dickey Fuller test of Hansen (1995) using Δx_t as a covariate. Hansen's (1995) approach extends the Dickey and Fuller (1979) approach to testing for a unit root by exploiting the information in stationary covariates. These tests deliver large improvements in power over univariate unit root tests, but they do not make optimal use of all available information. The analysis of Zivot (2000) proceeds under the assumption that x_t is weakly exogenous for the cointegrating parameters γ under the alternative. In addition, Zivot (2000, p. 415) assumes (as do we) that x_t is I(1). As discussed in the previous section, these two assumptions are different in general even though they are equivalent in the leading special case when there is no serial correlation (i.e. when $A(L) = I_{m+1}$). The theoretical analysis of Zivot (2000) therefore proceeds under assumptions that are identical to ours in the absence of serial correlation and strictly stronger than ours in the presence of serial correlation. Similarly, all local

asymptotic power curves in Zivot (2000) are computed under assumptions that are strictly stronger than ours in the presence of serial correlation.

One could also use a trace test to test for the number of cointegrating vectors when the cointegrating vectors are prespecified. Horvath and Watson (1995) compute the asymptotic distribution of the test under the null and the local alternative. The trace test does exploit the correlation between the errors to increase power, but does not do so optimally for the model we consider. We will show numerically below that the power of Horvath and Watson's (1995) test is always below the power of Elliott and Jansson's (2003) test when it is known that the covariates are I(1). All of the tests we consider are able to distinguish alternatives other than the one we focus on in this paper, however for none of these alternatives (say α_x nonzero) are optimality results available for any of the tests. This lack of any optimality result means that we have no theoretical prediction as to which test is the best test for those models. This implication is brought out in simulation results that show that the rankings between the statistics change for various models when the assumption that x_t is I(1) under the alternative is relaxed.

3.2. Optimal Tests. The development of optimality theory for the testing problem considered here is complicated by the nonexistence of a uniformly most powerful (UMP) test. Nonexistence of a UMP test is most easily seen in the special case where $\gamma = 0$ and x_t is strictly exogenous. In that case, our testing problem is simply that of testing if y_t has a unit root and it follows from the results of Elliott, Rothenberg and Stock (1996) that no UMP invariant (to the deterministic terms) test exists. By implication, a likelihood ratio test statistic constructed in the standard way will not give tests that are asymptotically optimal. That the more general model is more complicated than this special case will not override this lack of optimality on the part of the likelihood ratio test. Elliott and Jansson (2003) therefore return to first

principles in order to construct tests that will enjoy optimality properties. Apart from the test suggested here, none of the tests discussed in the previous section are in the family of optimal tests discussed here except in special cases (i.e. particular values for the nuisance parameters). This subsection describes the derivation of the Elliott and Jansson (2003) test, the functional form of which is presented in the next subsection.

Given values for the parameters of the model other than ρ (i.e. $A(L)$, μ_y, μ_x, τ_y, and τ_x) and a distributional assumption of the form $\varepsilon_t \sim i.i.d.\ F(\cdot)$ (for some known cdf $F(\cdot)$), it follows from the Neyman-Pearson lemma that a UMP test of

$$H_0 : \rho = 1 \qquad \text{vs.} \qquad H_{\bar{\rho}} : \rho = \bar{\rho}$$

exists and can be constructed as the likelihood ratio test between the simple hypotheses H_0 and $H_{\bar{\rho}}$. Even in this special case the statistical curvature of the model implies that the functional form of the optimal test statistic depends on the simple alternative $H_{\bar{\rho}}$ chosen. To address this problem, one could follow Cox and Hinkley (1974, Section 4.6) and construct a test which maximizes power for a given size against a weighted average of possible alternatives. (The existence of such a test follows from the Neyman-Pearson lemma.) Such a test, with test statistic denoted $\psi\left(\{z_t\}, G | A(L), \mu_y, \mu_x, \tau_y, \tau_x, F\right)$, would then maximize (among all tests with the same size) the weighted power function

$$\int \mathrm{Pr}_\rho\left(\psi \text{ rejects } | A(L), \mu_y, \mu_x, \tau_y, \tau_x, F\right) dG(\rho),$$

where $G(\cdot)$ is the chosen weighting function and the subscript on Pr denotes the distribution with respect to which the probability is evaluated. An obvious shortcoming

10

of this approach is that all nuisance parameters of the model are assumed to be known, as is the joint distribution of $\{\varepsilon_t\}$. Nevertheless, a variant of the approach is applicable. Specifically, we can use various elimination arguments to remove the unknown nuisance features from the problem and then construct a test which maximizes a weighted average power function for the transformed problem.

To eliminate the unknown nuisance features from the model, we first make the testing problem invariant to the joint distribution of $\{\varepsilon_t\}$ by making a 'least favorable' distributional assumption. Specifically, we assume that $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0, \Sigma)$. This distributional assumption is least favorable in the sense that the power envelope developed under that assumption is attainable even under the more general Assumption 3 of the previous section. In particular, the limiting distributional properties of the test statistic described below are invariant to distributional assumptions although the optimality properties of the associated test are with respect to the normality assumption. With other distributional assumptions further gains in power may be available (Rothenberg and Stock (1997)) Next, we remove the nuisance parameters Σ, $A(L)$, and μ_x from the problem. Under Assumption 2 and the assumption that $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0, \Sigma)$, the parameters Σ and $A(L)$ are consistently estimable and (due to asymptotic block diagonality of the information matrix) the asymptotic power of a test that uses consistent estimators of Σ and $A(L)$ is the same as if the true values were employed. Under Assumption 1, the parameter μ_x does not affect the analysis because it is differenced out by the known unit root in x_t. For these reasons, we can and will assume that the nuisance parameters Σ, $A(L)$, and μ_x are known.

Finally, we follow the general unit root literature by considering $\phi = \left(\mu_y, \tau_x', \tau_y\right)'$ to partially known and using invariance restrictions to remove the unknown elements of ϕ. We will consider the following four combinations of restrictions.

Case 1: (no deterministics) $\mu_y = 0$, $\tau_x = 0$, $\tau_y = 0$.

Case 2: (constants, no trend) $\tau_x = 0$, $\tau_y = 0$.

Case 3: (constants, no trend in $\beta' z_t$) $\tau_y = 0$.

Case 4: No restrictions.

The first of these cases corresponds to a model with no deterministic terms. The second has no drift or trend in Δx_t but a constant in the cointegrating vector, and the third case has x_t with a unit root and drift with a constant in the cointegrating vector. The no restrictions case adds a time trend to the cointegrating vector. Cases 1-4 correspond to cases 1-4 considered in Elliott and Jansson (2003). Elliott and Jansson (2003) also consider the case corresponding to a model where Δx_t has a drift and time trend. That case seems unlikely in the present problem so we will ignore it (the extension is straightforward).

Having reduced the testing problem to a scalar parameter problem (involving only the parameter of interest, ρ), it remains to specify the weighting function $G(\cdot)$. We follow the suggestions of King (1988) and consider a point optimal test. The weighting functions associated with point optimal tests are of the form $G(\rho) = 1\,(\rho \geq \bar{\rho})$, where $1\,(\cdot)$ is the indicator function and $\bar{\rho}$ is a prespecified number. In other words, the weighting function places all weight on the single point $\rho = \bar{\rho}$. By construction, a test derived using this weighting function has maximal power against the simple alternative $H_{\bar{\rho}} : \rho = \bar{\rho}$. As it turns out, it is possible to choose $\bar{\rho}$ in such a way that the corresponding point optimal test delivers 'nearly' optimal tests against alternatives other than the specific alternative $H_{\bar{\rho}} : \rho = \bar{\rho}$ (against which optimal power is achieved by construction). Studying a testing problem equivalent (in the appropriate sense) to the one considered here, Elliott and Jansson (2003) found that the point optimal tests with $\bar{\rho} = 1 + \bar{c}/T$, where $\bar{c} = -7$ for all but the no restrictions model and $\bar{c} = -13.5$ for the model with a trend in the cointegrating vector, are 'nearly' optimal against a

12

wide range of alternatives (in the sense of having 'nearly' the same local asymptotic power as the point optimal test designed for that particular alternative).These choices accord with alternatives where local power is 50% when x_t is strictly exogenous. The functional form of the point optimal test statistics will be given below.

The shape of the local asymptotic power function of our tests is determined by an important nuisance parameter. This nuisance parameter, denoted R^2, measures of the usefulness of the stationary covariates Δx_t and is given by $R^2 = \omega'_{xy}\Omega^{-1}_{xx}\omega_{xy}/\omega_{yy}$, where

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega'_{xy} \\ \omega_{xy} & \Omega_{xx} \end{pmatrix} = A(1)^{-1}\Sigma A(1)^{-1\prime}$$

and we have partitioned Ω after the first row and column. Being a squared correlation coefficient, R^2 lies between zero and one. When $R^2 = 0$, there is no useful information in the stationary covariates, and so univariate tests on the cointegrating vector are not ignoring exploitable information. The 'common factors' restriction discussed in Kremers et. al. (1992) provides an example of a model where $R^2 = 0$.. As R^2 gets larger, the potential power gained from exploiting the extra information in the stationary covariates gets larger.

3.3. Our Method. Our proposed test statistic for cases $i = 1, ..., 4$, denoted $\tilde{\Lambda}^i(1, \bar{\rho})$, can be constructed by following the five step procedure described below. The test based on $\tilde{\Lambda}^i(1, \bar{\rho})$ is a point optimal (invariant) test against a fixed alternative $\rho = \bar{\rho}$.

For $r \in \{1, \bar{\rho}\}$, let

$$Z_1(r) = \begin{pmatrix} y_1 - \gamma' x_1 \\ 0 \end{pmatrix},$$

$$Z_t(r) = \begin{pmatrix} (1 - rL)(y_t - \gamma' x_t) \\ \Delta x_t \end{pmatrix}, \qquad t = 2, \ldots, T,$$

and

$$d_1(r) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & I_m & 0 \end{pmatrix}',$$

$$d_t(r) = \begin{pmatrix} 1 - r & 0 & (1 - rL)t \\ 0 & I_m & 0 \end{pmatrix}', \qquad t = 2, \ldots, T.$$

Step 1: Impose the null $(\rho = 1)$ and estimate the VAR

$$\hat{A}(L) z_t(1) = \widehat{\det}_t + \hat{e}_t, \qquad t = k + 2, \ldots, T,$$

where the deterministic terms are as according to the case under consideration and we have dropped the first observation. (We drop this observation only in this step.). From this regression, we obtain consistent estimates of the nuisance parameters Ω and R^2 :

$$\hat{\Omega} = \begin{pmatrix} \hat{\omega}_{yy} & \hat{\omega}'_{xy} \\ \hat{\omega}_{xy} & \hat{\Omega}_{xx} \end{pmatrix} = \hat{A}(1)^{-1} \hat{\Sigma} \hat{A}(1)^{-1\prime}$$

and $\hat{R}^2 = \hat{\omega}'_{xy} \hat{\Omega}_{xx}^{-1} \hat{\omega}_{xy} / \hat{\omega}_{yy}$, where $\hat{\Sigma} = T^{-1} \sum_{t=k+2}^{T} \hat{e}_t \hat{e}'_t$.

14

Step 2: Estimate the coefficients ϕ on the deterministic terms under the null and alternative hypotheses. Each case $i = 1, ..., 4$ imposes a restriction of the form $(I_{m+2} - S_i) \phi = 0$, where S_i is a $(m+2) \times (m+2)$ matrix and $S_1 = 0$, $S_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $S_3 = \begin{pmatrix} I_{m+1} & 0 \\ 0 & 0 \end{pmatrix}$, and $S_4 = I_{m+2}$. For $r \in \{1, \bar{\rho}\}$, the formula for the case i estimate is

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

where $[\cdot]^{-}$ is the Moore Penrose inverse of the argument.

Step 3: Construct the detrended series under the null and alternative hypotheses. For $r \in \{1, \bar{\rho}\}$, let

$$\tilde{u}_t^i(r) = z_t(r) - d_t(r)' \hat{\phi}^i(r), \qquad t = 1, \ldots, T.$$

Step 4: For $r \in \{1, \bar{\rho}\}$, run the VAR

$$\tilde{A}(L) \tilde{u}_t^i(r) = \tilde{e}_t^i(r), \qquad t = k+1, \ldots, T,$$

and construct the estimated variance covariance matrices

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

Step 5: The test statistic is constructed as

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

15

The test rejects for small values of $\tilde{\Lambda}^i(1, \bar{\rho})$. As noted above we suggest using the alternatives $\bar{\rho}_i = 1 + \bar{c}_i/T$ where $\bar{c}_1 = \bar{c}_2 = \bar{c}_3 = -7$ and $\bar{c}_4 = -13.5$. Critical values for these tests are contained in Elliott and Jansson (2003) and reprinted in Table 1. The critical values are valid under Assumptions 1-3.

$$\boxed{\text{TABLE 1 ABOUT HERE}}$$

The equivalence between our testing problem and that of Elliott and Jansson (2003) enables us to use the results of that paper, reinterpreted in the setting of our model, to show the upper power bound for tests for cointegration when the cointegrating vector is known and the x_t variables are known to be I(1). This power bound is of practical value. First, it gives an objective standard to compare how efficiently tests use the data. Second, it allows us to study the size of the loss in power when the cointegrating vector is not known. Both of these examinations will be undertaken in the following section.

4. GAINS OVER ALTERNATIVE METHODS AND COMPARISON OF LOSSES FROM NOT KNOWING THE COINTEGRATING VECTORS

The results derived above enable a number of useful asymptotic comparisons to be made. First, there are a number of methods available to a researcher in testing for the possibility that a pre-specified cointegrating vector is (under the null) not a cointegrating vector. We can directly examine the relative powers of these methods in relation to the envelope of possible powers for tests. (The nonexistence of a UMP test for the hypothesis means that there is no test that has power identical to the power envelope, however some or all of the tests may well have power close to this envelope with the implication that they are 'nearly' efficient tests.) We examine power for various values for R^2. A second comparison that can be made is the examination of the loss involved in estimating cointegrating vectors in testing the hypothesis that the cointegrating vector does not exist. Pesavento (2003) examines the local asymptotic

16

power properties of a number of methods that do not require the cointegrating vector to be known. However, little is known regarding the extent of the loss involved in knowing or not knowing the cointegrating vector. All results are for tests with asymptotic size equal to 5%.

4.1. Relative Powers of Tests when the Cointegrating Vector is Known.

Figures 1 through 4 examine the case where the cointegrating vector of interest is known, and we are testing for a unit root in the cointegrating vector (null of no cointegration). As noted in Section 2, tests available for testing this null include univariate unit root test methods, represented here by the ADF test of Dickey and Fuller (1979) (note that the Z_τ test of Phillips (1987) and Phillips and Perron (1988) has the same local asymptotic power) and the P_T test of Elliott et al. (1996). They also include methods that exploit information in the covariates, which in addition to the test presented above include Hansen's (1995) CADF test and Horvath and Watson (1995) Wald test. Because power is influenced by the assumptions on the deterministics we present results for each of the four cases for the deterministics (Figures 1 through 4 are cases 1 through 4 respectively). The power also depends on R^2, the squared zero frequency correlation between the shocks driving the potentially cointegrating relation and the 'X' variables, respectively. We present three sets of results for each case. Figures 1a through 1c are for the model with no deterministic terms with $R^2 = 0, 0.3$ and 0.5 respectively (and similarly for each of the other models of the deterministic component).

FIGURES 1-4 ABOUT HERE

When $R^2 = 0$ there is no gain in using the system methods over the univariate unit root methods as there is no exploitable information in the extra equations. In this

17

case the Elliott and Jansson (2003) test is equivalent to the P_T test and the CADF test has equivalent power to the ADF test. This is clear from Figures 1a, 2a, 3a, and 4a, where the power curves lie on top of each other for these pairs of tests. When there are no deterministic terms, it has previously been shown (Stock (1994), Elliott et. al. (1996)) that there is very little distinction between the power envelope, the P_T test and the ADF test. This is evident in Figure 1a, which shows all of the tests as having virtually identical power curves to the power envelope. When there are deterministic terms (the remaining cases), these papers show that the P_T test remains close to the power envelope whilst the ADF test has lower power. This is also clear from the results in Figures 2a, 3a, and 4a. As the equivalence between the test presented here and the P_T test holds (as does the equivalence between CADF and ADF), the test presented herein has similarly better power than the CADF test.

In all cases when $R^2 > 0$, the multivariate tests have extra information to exploit. In parts b and c of Figures 1 through 4 we see that the power of the test presented herein is greater than that of the P_T test (and the power of the CADF test is greater than that of the ADF test). As was to be expected, these differences are increasing as R^2 gets larger. The differences are smaller when there is a trend in the potentially cointegrating relation (case 4) than when the specification restricts such trends to be absent (cases 1-3). There is a trade-off between using the most efficient univariate method (the P_T statistic) and using the system information inefficiently (the CADF statistic and the Horvath and Watson (1995) Wald test). In Figure 2b we see that the P_T test has power in excess of the CADF test, whereas the ranking is reversed in Figure 2c, where the system information is stronger. In the model with a trend in $\beta' z_t$ the reversal of the ranking is already apparent when $R^2 = 0.3$, implying that the relative value of the system information is larger in that case.

In all of the models, the power functions for the Elliott and Jansson (2003) tests are quite close to the power envelopes. In this sense they are 'nearly' efficient tests.

18

For the choices of the point alternatives suggested above, there is some distinction between the power curve and the power envelope in the model where the cointegrating vector has a trend and R^2 is very large (not shown in the figures). However, this appears to be an unlikely model in practice. The power is adversely affected by less information on the deterministic terms (this is a common result in the unit root testing literature). We can see this clearly by holding R^2 constant and looking across the figures. Comparing the constants only model to the model where there is a trend in the cointegrating vector when $R^2 = 0.3$, we have that the test achieves power at 50% for c around 4.9 when there are constants only, whereas with trends this requires a c around 8.8, a more distant alternative. This difference essentially means that in the model with a trend in the cointegrating vector we require about 80% more observations to achieve power at 50% against the same alternative value for ρ.

Also in all models, the test presented herein has higher power than the CADF test for the null hypothesis. Again using the comparisons at power equal to 50%, we have that in the model with constants only and $R^2 = 0.3$ we would require 80% more observations for the same power at the same value for ρ. As R^2 rises, this distinction lessens. When $R^2 = 0.5$, the extra number of observations is around 60%. The distinctions are smaller when trends are possibly present. When $R^2 = 0.3$ we would require only 28% more observations when there is a trend in $\beta' z_t$, whilst if $R^2 = 0.5$ this falls to 15%. For these alternatives the Horvath and Watson test tends to have lower power, although as noted this test was not designed directly for this particular set of alternatives.

Overall, large power increases are available through employing system tests over univariate tests except in the special case of R^2 very small. Since this nuisance parameter is simply estimated, it seems that one could simply evaluate for a particular study the likely power gains using the system tests from the graphs presented.

4.2. Power Losses When the Cointegrating Vector is Unknown. When the parameters of the cointegrating vector of interest are not specified, they are typically estimated as part of the testing procedure. Methods to do this include the Engle and Granger (1987) two step method of estimating the cointegrating vector and then testing the residuals for a unit root using the ADF test, and the Zivot (2000) or Boswijk (1994) tests in the error correction models. One could also simply use a rank test, testing the null hypothesis of $m + 1$ unit roots versus m unit roots. These tests include the Johansen (1988, 1995) and Johansen and Juselius (1990) methods, and the Harbo et al. (1998) rank test in partial systems. The Zivot (2000) test is equivalent to Banerjee, Dolado, Hendry, and Smith's (1986) and Banerjee, Dolado, Galbraith, and Hendry's (1993) t-test in a conditional error correction model with unknown cointegration vector (ECR thereafter). Additionally, for the case examined in this paper in which the right-hand variables are not mutually cointegrated and there is at most one cointegration vector, the Harbo, Johansen, Nielsen, and Rahbek (1998) rank test is equivalent to Boswijk's (1994) Wald test. The rank tests are not derived under the assumption that x_t is I(1), implying that they spread power (in an arbitrary and random way) amongst the alternatives we examine here as well as alternatives in the direction of one (or more) of the x_t variables being stationary. Although the rank tests do not make optimal use of the information about x_t, these tests will of course still be consistent against the alternatives considered in this paper. We will see numerically that the lack of imposing the information on x_t comes at a relatively high cost.

Pesavento (2003) gives a detailed account of the abovementioned methods and computes power functions for tests of the null hypothesis that there is no cointegrating vector. The powers of the tests are found to depend asymptotically on the specification of the deterministic terms and R^2, just as in the known cointegrating vector case. Pesavento (2003) finds that the error correction methods outperform the

other methods for all models, and the ranking between the Engle and Granger (1987) method and the Johansen (1988, 1995) methods depends on the value for R^2, the first method (a univariate method) being useful when there is little extra information in the remaining equations of the system (i.e. when R^2 is small) and the Johansen full system method being better when the amount of extra exploitable information is substantial.

The absolute loss from not knowing the cointegrating vector can be assessed by examining the difference between the power envelope when the cointegrating vector is known versus the power functions for these tests. The quantification of this gap is useful for researchers in examining results where one estimates the cointegrating vector even though theory specifies the coefficients of the vector (a failure to reject may be due to a large decrease in power) and also provide guidance for testing in practice when one has a vector and does not know if they should specify it for the test. In this case, if the power losses are small, then it would be prudent not to specify the coefficients of the cointegrating vector but instead estimate it.

<div align="center">

FIGURES 5-7 ABOUT HERE

</div>

Figures 5 through 7 show the results for these power functions for values for $R^2 = 0$, 0.3, and 0.5, respectively. Each figure has two panels, the first for the model with constants only and the second for the model with a trend in the cointegrating vector. The first point to note is that in all cases, the gap between the power envelope when the coefficients of the cointegrating vector are known and the best test is very large. This means that there is a large loss in power from estimating the cointegrating vector. Comparing the first and second panels for each of these figures, we see that additional deterministic terms (trend versus constant) results in a smaller gap. This is most apparent when $R^2 = 0$ and lessens as R^2 gets larger. In case 2, when $R^2 = 0.3$

(Figure 6, panel a) at $c = -5$ we have power of 53% for the power envelope and just 13% for the best test examined that does not have the coefficients of the cointegrating vector known (the ECR test). In case 2 with $R^2 = 0.3$, we have that at $c = -5$ the envelope is 53% but power of the ECR test is 13% as noted above. For the model with trends the envelope is 24% and power of the ECR test is 8%. The difference falls from 40% to 16%. In general for the model with trends the power curves tend to be closer together than in the model with constants only, compressing all of the differences.

As R^2 rises, the gap between the power envelope and the power of the best test falls. This is true for both cases. In the constants only model when $R^2 = 0$ we have at $c = -10$ we have 79% power for the envelope and 27% power for the ECR test. When R^2 increases to 0.5, we have power for the envelope of 98% and 54% for the ECR test. The difference falls from 52% to 44%.

4.3.   Power of the Test When α_x is Different From Zero in Equation (5).

Although the assumption that x_t is I(1) under both the null and local alternative is a reasonable assumption in the context of cointegration, we can also examine the sensitivity of the proposed tests to different values of α_x in equation (5). Recall that for these models x_t is I(1) under the null but under the alternative has a small additional local to I(2) component. We simulate equation (5) with scalar x_t and $\beta = (1, -1)'$:

$$
\begin{aligned}
\Delta \left(y_t - x_t \right) &= \left(\rho - 1 \right) \left(y_{t-1} - x_{t-1} \right) + v_{y,t}, \\
\Delta x_t &= \left(\rho - 1 \right) \alpha_x \left(y_{t-1} - x_{t-1} \right) + v_{x,t}.
\end{aligned}
$$

The error process $v_t = (v_{y,t}, v_{x,t})'$ is generated by the VAR(1) model

$$v_t = Av_{t-1} + \varepsilon_t, \qquad A = \begin{pmatrix} 0.3 & 0.2 \\ 0.1 & 0.2 \end{pmatrix},$$

where $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0, \Sigma)$ and Σ is chosen in such a way that $\Omega = (I_2 - A)^{-1} \Sigma (I_2 - A)^{-1\prime}$, the long-run variance covariance matrix of v_t, is given by

$$\Omega = \begin{pmatrix} 1 & R \\ R & 1 \end{pmatrix},$$

where $R \in [0,1)$ is the positive square root of R^2, the nuisance parameter that determines asymptotic power (when $\alpha_x = 0$).

The system can be written as

$$\Delta(y_t - x_t) = (\rho - 1)(1 - R\alpha_x)(y_{t-1} - x_{t-1}) + R\Delta x_t + \sqrt{1 - R^2}\eta_{y,t}, \qquad (9)$$

$$\Delta x_t = (\rho - 1)\alpha_x(y_{t-1} - x_{t-1}) + \eta_{x,t}, \qquad (10)$$

where $\eta_t = (\eta_{y,t}, \eta_{x,t})' = \Omega^{-1/2}v_t$ has long-run variance covariance matrix I_2. In this example, if $\rho < 1$ but $R\alpha_x > 1$, the error correction term, y_{t-1}, is not mean reverting and tests based on equation (9) will be inconsistent (see also Zivot (2000, page 429)). For this reason, only values for α_x such that $R\alpha_x < 1$ will be considered.

The number of lags (one) is assumed known (for the Hansen (1995) and Zivot (2000) tests one lead of Δx_t is also included). The regressions are estimated for the models with no deterministic terms, with constants only, and with no restrictions.

We do not report results for case 3 as they are similar to the included results. The sample size is $T = 1,500$ and $10,000$ replications are used.

<div style="text-align: center; border: 1px solid black; display: inline-block;">TABLES 2-4 ABOUT HERE</div>

Tables 2 though 4 report the rejection rates for various values of α_x. The rejection rates for $c = 0$ do not vary with α_x and are not reported as they equal 0.05. When $\alpha_x = 0$, the power of Elliott and Jansson (2003) test is higher than the power of any other tests, including the Horvath and Watson (1995) test. In the simulated DGP, A is not lower triangular, so $\alpha_x = 0$ does not coincide with weak exogeneity. The Elliott and Jansson (2003) test exploits the information contained in the covariates in an optimal way and therefore rejects the null hypothesis with higher probability than Horvath and Watson's (1995) test.

When α_x is positive, the root of y_t in equation (9) is larger (in our simulations ω_{xy} is positive so $R\alpha_x > 0$), so all the tests based on equation (9) will have smaller power, as will the test proposed herein. For positive but small values of α_x, Elliott and Jansson's (2003) test still performs relatively well compared to other tests. Only when there are no deterministic terms and R^2 is zero, the proposed test rejects a false null with probability smaller than Hansen's (1995) CADF test. As α_x increases, Horvath and Watson's (1995) trace test outperforms the other tests in most cases. When the deterministic terms include a constant, but not a trend, Elliott and Jansson (2003) test has power similar to Horvath and Watson (1995) test in a neighborhood of the null.

The P_T test of Elliott et al. (1996) does not use the information in the covariates and the power for $\alpha_x = 0$ is lower than Elliott and Jansson's (2003) test when R^2 is different than zero. Given that the P_T test is based on the single equation, it is not sensitive to α_x and it rejects with higher probability than Elliott and Jansson's (2003)

24

test for large positive values of α_x when R^2 is positive. Finally, the test proposed by Zivot (2000) rejects the null with lower probability than Elliott and Jansson's (2003) test for any value of α_x. This is not surprising given that Zivot's test does not fully utilize the information that the cointegrating vector is known.

When α_x is negative, the coefficient for the error correction term in the conditional equation is further away from zero, and all the tests rejects the null of no cointegration more often with Elliott and Jansson (2003) test having the highest power as soon as R^2 departs from zero.

4.4. Small Sample Comparisons. The results of the previous sections show that the Elliott and Jansson (2003) family of tests has optimality properties when applied in the context of model $(1) - (3)$ and has asymptotic power that depends on the nuisance parameter R^2. Although the particular estimator used to estimate the nuisance parameters does not affect the asymptotic distributions under the local alternatives, the finite sample properties of tests for no cointegration can be sensitive to the choice of the estimation method. To study the small sample behavior of the proposed test, we simulate equation (5) with scalar x_t, $\alpha_x = 0$, and $\beta = (1, -1)'$:

$$
\begin{aligned}
\Delta\left(y_t - x_t\right) &= (\rho - 1)\left(y_{t-1} - x_{t-1}\right) + v_{y,t}, \\
\Delta x_t &= v_{x,t}.
\end{aligned}
$$

The error process $v_t = (v_{y,t}, v_{x,t})'$ is generated by the VARMA(1,1) model

$$
(I_2 - AL)\, v_t = (I_2 + \Theta L)\, \varepsilon_t,
$$

where

$$A = \begin{pmatrix} a_1 & a_2 \\ a_2 & a_1 \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_1 \end{pmatrix},$$

and $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0, \Sigma)$, where Σ is chosen in such a way that the long-run variance covariance matrix of v_t satisfies

$$\Omega = (I_2 - A)^{-1} (I_2 + \Theta) \Sigma (I_2 + \Theta)' (I_2 - A)^{-1\prime} = \begin{pmatrix} 1 & R \\ R & 1 \end{pmatrix}, \qquad R \in [0, 1).$$

The number of lags and leads is estimated by BIC on a VAR on the first differences (under the null) with a maximum of 8 lags. For case 2, the regressions are estimated with a mean. For the model with a trend in the cointegrating vector, the regressions are estimated with a mean and trend (results for other cases were similar). The sample size is $T = 100$ and 10,000 replications are used.

$$\boxed{\text{TABLES 5-6 ABOUT HERE}}$$

Tables 5 and 6 compare the small sample size of Elliott and Jansson's (2003) test and Hansen's (1995) CADF test for various values of Θ and A. To compute the critical values in each case we estimate the value of R^2 as suggested by Elliott and Jansson (2003) and Hansen (1995). Overall the Elliott and Jansson (2003) test is worse in term of size performance than the CADF test. This is the same type of difference found between the P_T and DF tests in the univariate case, so is not surprising given that these methods are extensions of the two univariate tests respectively. The difference between the two tests is more evident for large values of R^2 and for the case with no trend. When Θ is nonzero both tests present size distortions that are severe in the

26

presence of a large negative moving average root (as is the case for unit root tests), emphasizing the need of proper modeling of the serial correlation present in the data..

5. Cointegration Between Forecasts and Outcomes

There are a number of situations where if there is a cointegrating vector we have theory that suggests the form of the cointegrating vector. In the purchasing power parity literature, the typical assumption is that logs of the nominal exchange rate, home and foreign prices all have unit roots and the real exchange rate does not. The real exchange rate is constructed from the $I(1)$ variables with the cointegrating vector $(1, 1, -1)'$. In examining interest rates, term structure theories often imply a cointegrating structure of $(1, -1)'$ between interest rates of different maturities (however one might find it difficult to believe that the log interest rate is unbounded, and hence is unlikely to have a unit root).

Another example is forecasts and outcomes of the variable of interest. Since many variables that macroeconomists would like to forecast have trending behavior, often taken to be unit root behavior, some researchers have examined whether or not the forecasts made in practice are indeed cointegrated with the variable being forecast. The expected cointegrating vector is $(1, -1)'$, implying that the forecast error is stationary. This has been undertaken for exchange rates (Liu and Maddala (1992)) and macroeconomic data (Aggarwal, Mohanty and Song (1995)). In the context of macroeconomic forecasts, Cheung and Chinn (1999) also relax the cointegrating vector assumption.

The requirement that forecasts be cointegrated with outcomes is a very weak requirement. Note that the forecasters information set includes the current value of the outcome variable. Since the current value of the outcome variable is trivially cointegrated with the future outcome variable to be forecast (they differ by the change, which is stationary) then the forecaster has a simple observable forecast that satisfies

the requirement that the forecast and outcome variable be cointegrated. We can also imagine what happens under the null hypothesis of no cointegration. Under the null, forecast errors are I(1) and hence become arbitrarily far from zero with probability one. It is hard to imagine that a forecaster would stick with such a method when the forecast becomes further from the current value of the outcome than typical changes in the outcome variable would suggest are plausible.

<div align="center">

FIGURE 8 ABOUT HERE

</div>

This being said, of course it is useful that tests reject the hypothesis of no cointegration and quite indicative of power problems if they do not. Here we employ forecasts of the price level from the Livingston data set over the period 1971 to 2000. The survey recipients forecast the consumer price index six months ahead. Figure 8 shows the forecast errors. As the variables are indexes a value of 1 is a 1% difference relative to the base 1982-1984. Forecast errors at different times have been quite large, especially around the times of the oil shocks in the 1970's. They have been smaller and more often negative over the last two decades - this was a period of falling inflation rates which appears to have induced the error on average of overestimating prices. There is no indication from the data that the errors are getting larger in variance over time, although there are long swings in the forecast errors that may lead lower power tests into failing to reject the hypothesis that the forecast error does not have a unit root.

<div align="center">

TABLE 7 ABOUT HERE

</div>

Indeed, the Dickey and Fuller (1979) test is unable to reject a unit root in the forecast error, even at the 10% level. This is shown in Table 7, which provides the results of the test in column 1. We have allowed under the alternative for a nonzero mean (i.e.

the constant included case of the test). Commonly employed multivariate tests do a little better. The Horvath and Watson (1995) test fails to reject at the 10% level, while the CADF test rejects at 5% but not the 1% level. Thus, even though the null hypothesis is an extremely weak requirement of the data, the forecasts fail the test in most cases. However, the problem could be one of power rather than extremely poor forecasting. This is further backed up by the Elliott et. al. (1996) tests, which reject at the 5% level but not at the 1% level. The p-value for the P_T test is 0.02.

$$\boxed{\text{TABLE 8 ABOUT HERE}}$$

The first two columns of Table 8 presents results for the Elliott and Jansson (2003) test under the assumption that the change in the forecasts is on the right hand side of the cointegrating regression. Our X variable is chosen to be price expectations. Results are similar when prices are chosen as the X variable, as reported in Table 8. We are examining case 3 from the paper, i.e. the statistic is invariant to a mean in the change in forecasts (so this variable has a drift, prices rise over time suggesting a positive drift) and a mean in the quasi difference of the cointegrating vector under the alternative. For there to be gains over univariate tests, the R^2 value should be different from 0. Here we estimate R^2 to be 0.19, suggesting there are gains from using this multivariate approach. Comparing the statistic developed here with its critical value we are able to reject not only at the 5% level but at the 1% level. Comparing this results with those for the previous tests, we see results that we may have expected from the asymptotic theory. Standard unit root tests have low power (and we do not reject with DF). We can improve power through using additional information such as using CADF or through more efficient use of the data through P_T. In these cases we do reject at the 5% level. Finally, using the additional information and using all information efficiently, where we expect to have the best power, we reject at not only

the 5% level but also at the 1% level. We are able to reject that the cointegrating vector has a unit root and conclude that the forecast errors are indeed mean reverting, a result not available with current multivariate tests and less assured from the higher power univariate tests.

As a robustness check, we also tested the data for a unit root allowing for a break at unknown time. The forecast errors in Figure 8 appear to have a shift in the level around 1980 to 1983 that could lower the probability of rejection of conventional tests. To test the data for a unit root with break, we use Perron and Vogelsang (1992) test. Denote $DU_t = 1$ if $t > T_b$ and 0 otherwise, where T_b is the break date. Following Perron and Vogelsang (1992) we first remove the deterministic part of the series for a given break T_b by estimating the regression

$$y_t = \mu + \delta DU_t + \widetilde{y}_t.$$

The unit root test is then computed as the t-test for $\alpha = 1$ in the regression

$$\widetilde{y}_t = \sum_{i=0}^{k} \omega_i D(TB)_{t-1} + \alpha \widetilde{y}_{t-1} + \sum_{i=1}^{k} c_i \Delta \widetilde{y}_{t-1} + e_t$$

where $D(TB)_t = 1$ if $t = T_b = 1$ and 0 otherwise.

<div align="center">

TABLE 9 ABOUT HERE

</div>

Panel A in Table 9 corresponds to the case in which the break is estimated as the date that minimizes the t-statistics $t_{\hat{\sigma}}$ in the unit root test. The number of lags is chosen for a given break such that the coefficient on the last included lag of the first differences of the data is significant at 10% level (for details, see Perron and Vogelsang

(1992)). Panel B corresponds to the case in which the break is chosen to minimize the t-statistics testing $\delta = 0$ in the first regression.

As the table shows, standard methods reject for some cases but not everywhere. When the break is chosen to minimize the t-statistic in the unit root test, the unit root with break test rejects at 10% level. When the break is chosen as the date that minimizes the t-statistic in the regression for the deterministic, we cannot reject the unit root hypothesis. Overall it appears that if there is a break, it is small. This is all the more reason to employ tests that use the data as efficiently as possible.

6. Conclusion

This paper examines the idea of testing for a unit root in a cointegrating vector when the cointegrating vector is known and the variables are known to be $I(1)$. Early studies simply performed unit root tests on the cointegrating vector, however this approach omits information that can be very useful in improving power of the test for a unit root. The restrictions placed on the multivariate model for this 'known a priori' information renders the testing problem equivalent to that in Elliott and Jansson (2003), and so those tests are employed here. Whilst there exists no uniformly most powerful test for the problem, the point optimal tests derived in that paper and appropriate here are amongst the asymptotically admissible class (as they are asymptotically equivalent to the optimal test under normality at a point in the alternative) and were shown to perform well in general.

The method is quite simple, requiring the running of a vector autoregression to estimate nuisance parameters, detrending the data (under both the null and the alternative) and then running two vector autoregressions, one on the data detrended under the null and another based on the data detrended under the alternative. The statistic is then constructed from the variance covariance matrices of the residuals of these vector autoregressions.

We then applied the method to examine the cointegration of forecasts of the price level with the actual price levels. The idea that forecasts and their outcomes are cointegrated with cointegrating vector $(1, -1)'$ (so forecast errors are stationary) is a very weak property. It is difficult to see that such a property could be violated by any serious forecaster. The data we examine here is six month ahead forecasts from the Livingston data for prices from 1971-2000. However, most simple univariate tests and some of the more sophisticated multivariate tests currently available to test the proposition do not reject the null that the forecast errors have a unit root. The tests derived here are able to reject this hypothesis with a great degree of certainty.

7. APPENDIX: NOTES ON THE DATA

The current CPI is the non seasonally adjusted CPI for all Urban Consumers from the Bureau of Labor and Statistics (code CUUR0000SA0) corresponding to the month being forecasted. All the current values for the CPI are in 1982-1984 base year

The forecasts CPI data are the six month price forecasts from the Livingston Tables at the Philadelphia Fed from June 1971 to December 2001. The survey is conducted twice a year (early June and early December) to obtain the six month ahead forecasts from a number of respondents. The number of respondents varies for each survey so each forecast in our sample is computed as the average of the forecasts from all the respondents from each survey. The data in the Livingston Tables are available in 1967 base up to December 1987 and 1982-1984 base thereafter. Given that there are not overlapping forecasts at both base years, we transformed all the forecasts to a 1982-1984 base as follows. We first computed the average of the actual values for the 1982-1984 base CPI for the year 1967. We then used this value to multiply all the forecasts prior to 1987 to transform the forecasts to a 1982-1984 base.

At the time of the survey the respondents were also given current figures on which

to base their forecasts. The surveys are sent out early in the month so the available information to the respondents for the June and December survey are, respectively, April and October. For this reason, although traditionally the forecasts are denoted as 6 month ahead forecasts, they are truly 7 month ahead forecasts. Carlson (1977) presents a detailed description of the issues related with the price forecasts from the Livingston Survey.

8.   REFERENCES

Aggarwal, R., S. Mohanty, and F. Song (1995): "Are Survey forecasts of Macroeconomic Variables Rational?", *Journal of Business*, 68, 99-119.

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

Banerjee, A., J. Dolado, J.W. Galbraith, and D.F. Hendry (1993): *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Bernard, A., and S. Durlauf (1995): "Convergence in International Output", *Journal of Applied Econometrics*, 10, 97-108.

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

Carlson, J.A. (1977): "A Study of Price Forecasts", *Annals of Economic and Social Measurement*, 6, 27-56.

Cheung, Y-W., and M.D. Chinn (1999): "Are Macroeconomic Forecasts Informative? Cointegration Evidence from the ASA-NBER Surveys", NBER discussion

paper 6926.

Cheung, Y-W., and K. Lai (2000): "On Cross Country Differences in the persistence of Real Exchange Rates", *Journal of International Economics*, 50, 375-97.

Cox, D.R., and D.V. Hinkley (1974): *Theoretical Statistics.* New York: Chapman & Hall.

Dickey, D.A., and W.A. Fuller (1979): "Distribution of Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association,* 74, 427-431.

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

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 813-836.

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

Greasley, D., and L. Oxley (1997): "Time-series Based Tests of the Convergence Hypothesis: Some Positive Results", *Economics Letters*, 56, 143-147.

Hansen, B.E. (1995): "Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power", *Econometric Theory*, 11, 1148-1171.

Harbo, I. S. Johansen, B. Nielsen, and A. Rahbek (1998): "Asymptotic Inference on Cointegrating Rank in Partial Systems", *Journal of Business and Economic Statistic*, 16, 388-399.

Horvath, M.T.K., and M.W. Watson (1995): "Testing for Cointegration When Some of the Cointegrating Vectors are Prespecified", *Econometric Theory*, 11, 984-1014.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

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

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.

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

Kremers, J.J.M., N.R. Ericsson, and J. Dolado (1992): "The Power of Cointegration Tests", *Oxford Bulletin of Economics and Statistics*, 54, 325-348.

Liu, T., and G.S. Maddala, (1992): "Rationality of Survey data and tests for Market Efficiency in the Foreign Exchange Markets", *Journal of International Money and Finance*, 11, 366-81.

Perron P., and T.J. Vogelsang (1992): "Non stationarity and Level Shifts With an Application to Purchasing Power Parity", *Journal of Business and Economic Statistics*, 10, 301-320.

Pesavento, E. (2003): "An Analytical Evaluation of the Power of Tests for the Absence of Cointegration", *Journal of Econometrics*, forthcoming.

Phillips, P.C.B. (1987): "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.

Phillips, P.C.B, and P. Perron (1988): "Testing for a Unit Root in Time Series Regression", *Biometrika*, 75, 335-346.

Phillips, P.C.B., and V. Solo (1992): "Asymptotic for Linear Processes", *The Annals of Statistics*, 20, 971-1001.

Rothenberg, T.J., and J.H. Stock (1997): "Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations", *Journal of Econometrics*, 80, 269-286.

Stock, J.H. (1994): "Unit Roots, Structural Breaks and Trends", in *Handbook of Econometrics, Volume IV*, ed. by R.F. Engle and D.L. McFadden. New York: North Holland, 2739-2841.

Zivot, E. (2000): "The Power of Single Equation Tests for Cointegration when the Cointegrating Vector is prespecified", *Econometric Theory*, 16, 407-439.
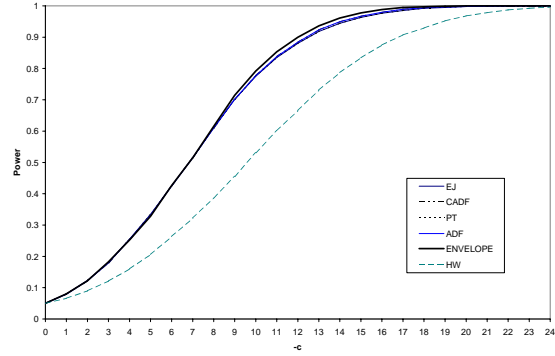
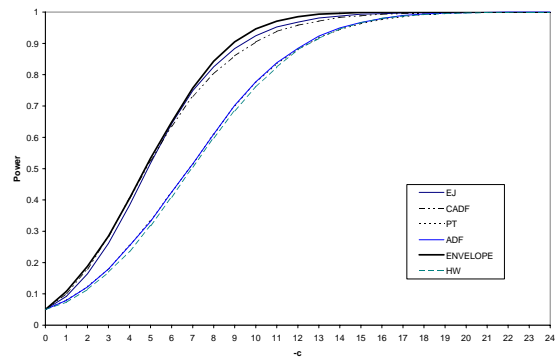Figure 1a: Asymptotic Power, $R^2 = 0$, No deterministic terms.

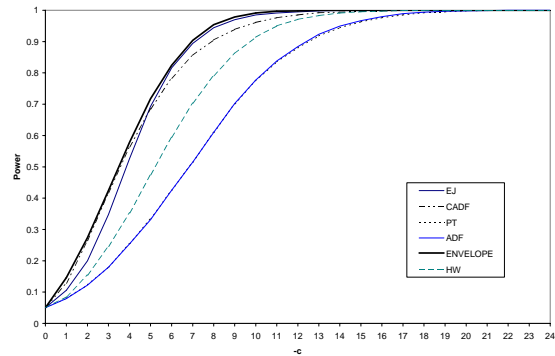Figure 1b: Asymptotic Power, $R^2 = 0.3$, No deterministic terms.

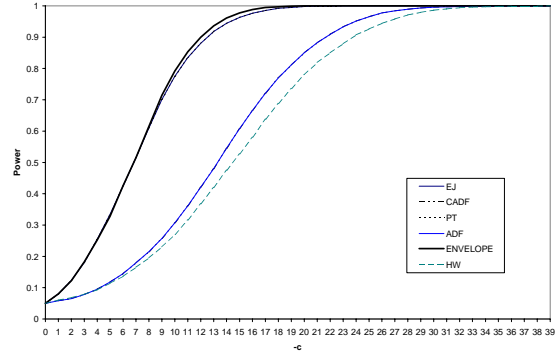Figure 1c: Asymptotic Power, $R^2 = 0.5$, No deterministic terms.

Figure 2a: Asymptotic Power, $R^2 = 0$, Constants, no Trend.

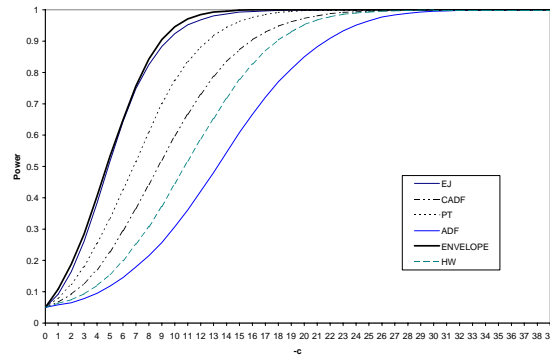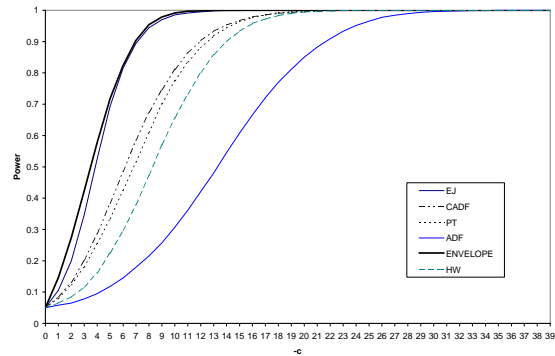Figure 2b: Asymptotic Power, $R^2 = 0.3$, Constants, no Trend.

Figure 2c: Asymptotic Power, $R^2 = 0.5$, Constants, no Trend.

Figure 3a: Asymptotic Power, $R^2 = 0$, Constants, no Trend in the Cointegrating Vector.
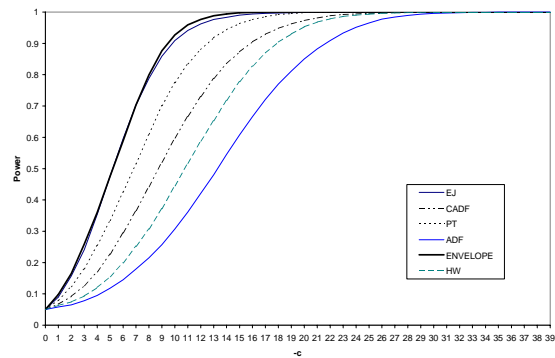
Figure 3b: Asymptotic Power, $R^2 = 0.3$, Constants, no Trend in the Cointegrating Vector.
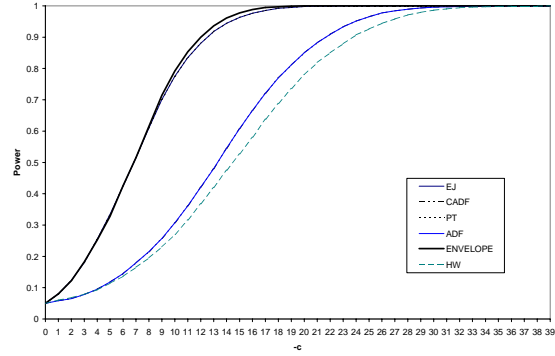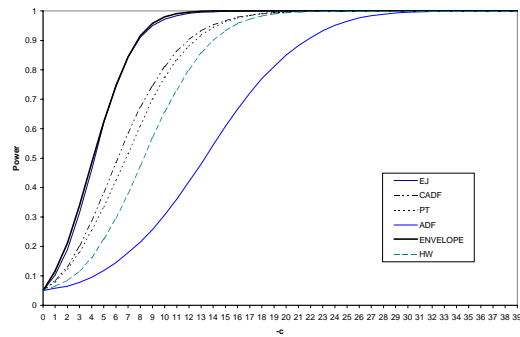

Figure 3c: Asymptotic Power, $R^2 = 0.5$, Constants, no Trend in the Cointegrating Vector.
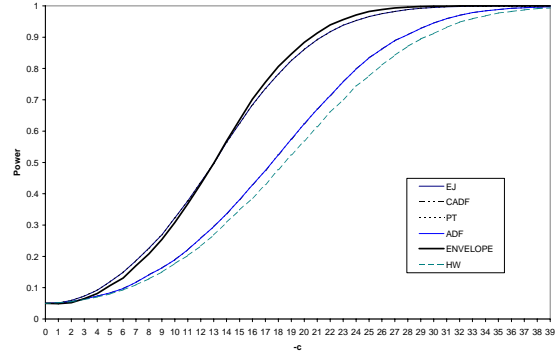
Figure 4a: Asymptotic Power, $R^2 = 0$, No Restrictions in the Deterministic Terms.
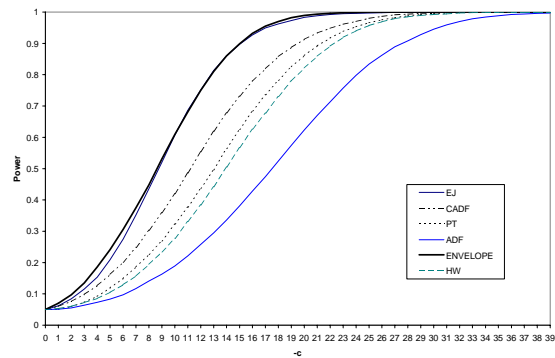
Figure 4b: Asymptotic Power, $R^2 = 0.3$, No Restrictions in the Deterministic Terms.
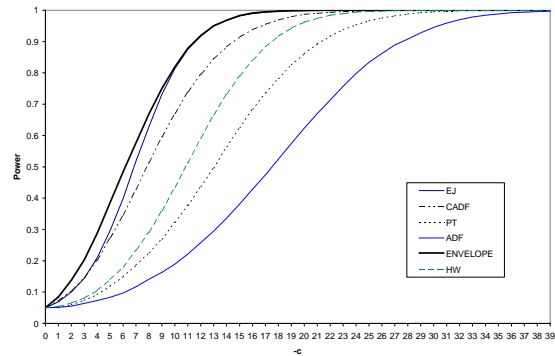
Figure 4c: Asymptotic Power, $R^2 = 0.5$, No Restrictions in the Deterministic Terms.
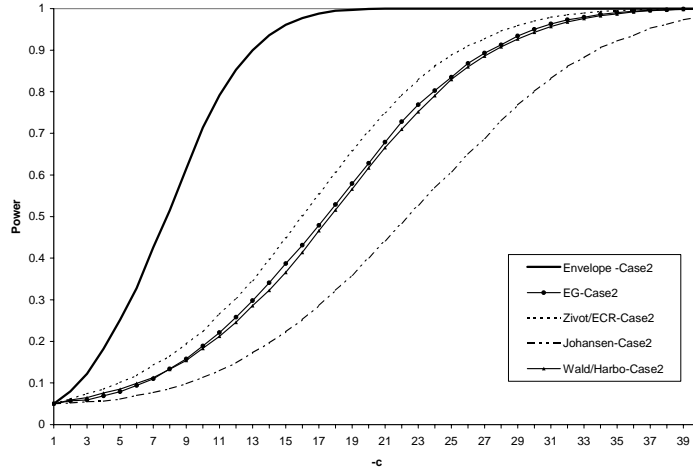
Figure 5a: Asymptotic Power Known and Unknown Cointegration Vector, $R^2 = 0$, Constants, no Trend.
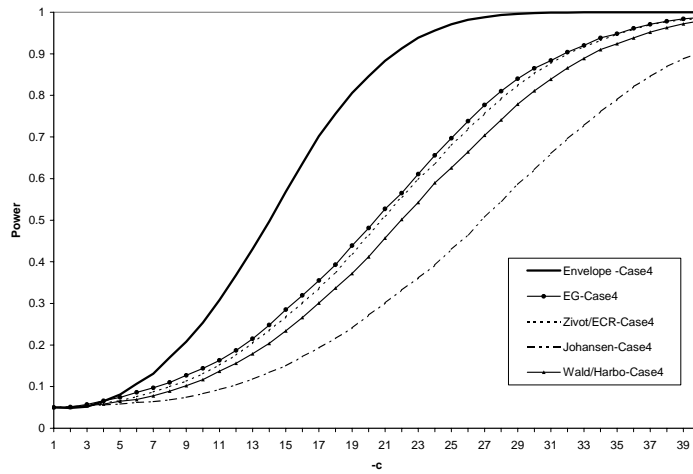
Figure 5b: Asymptotic Power Known and Unknown Cointegration Vector, $R^2 = 0$, No Restrictions in the Deterministic Terms.

41
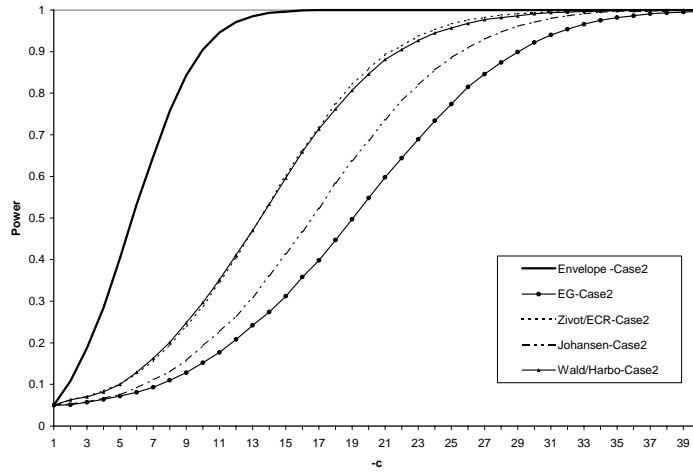
Figure 6a: Asymptotic Power Known and Unknown Cointegration Vector, $R^2 = 0.3$, Constants, no Trend.
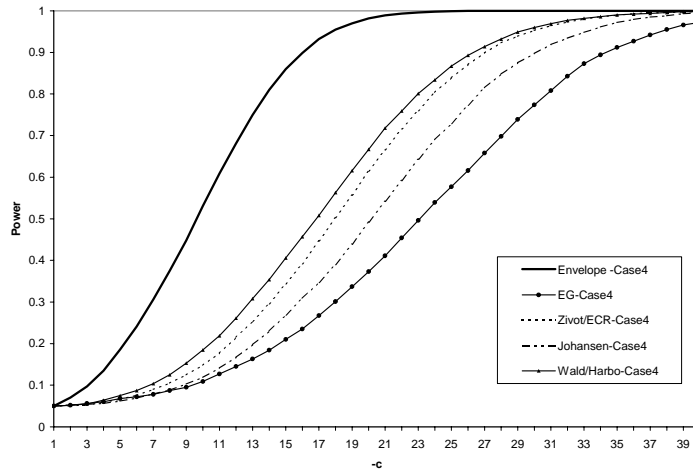
Figure 6b: Asymptotic Power Known and Unknown Cointegration Vector, $R^2 = 0.3$, No Restrictions in the Deterministic Terms.
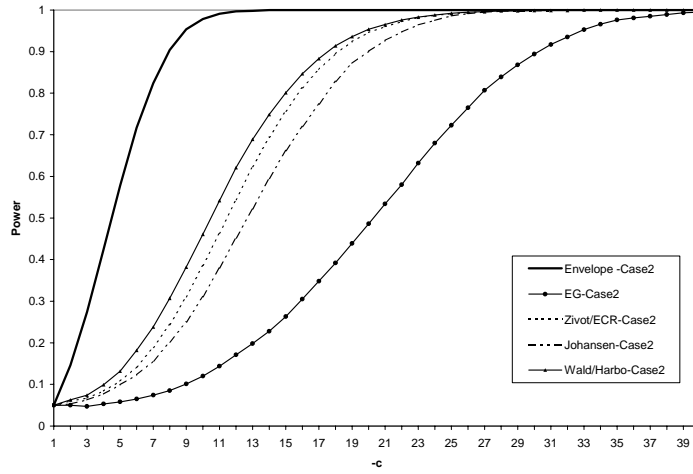
Figure 7a: Asymptotic Power Known and Unknown Cointegration Vector, $R^2 = 0.5$,Constants, no Trend.
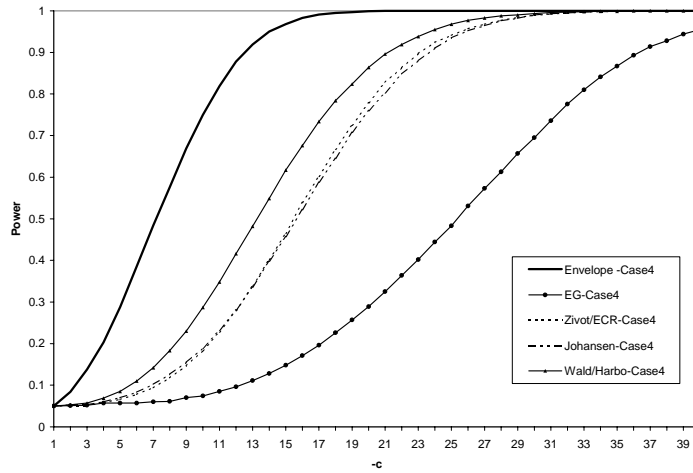
Figure 7b: Asymptotic Power Known and Unknown Cointegration Vector, $R^2 = 0.5$, No Restrictions in the Deterministic Terms.

43
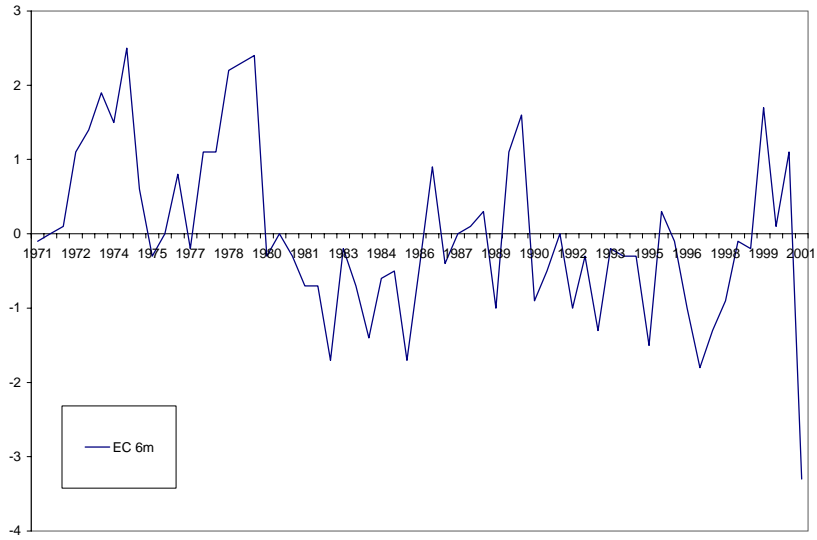
Figure 8: Forecasts errors for 6 months ahead forecasts.

Table 1: Critical Values

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

Notes: These are reprinted from Elliott and Jansson (2003).

Table 2: Rejection Rates when $\alpha_x \neq 0$ in equation (5), No deterministic terms.

α_x	$R^2\backslash c$	EJ -5	EJ -10	HW -5	HW -10	CADF -5	CADF -10	Zivot -5	Zivot -10	P_T -5	P_T -10
-1	0	0.33	0.76	0.43	0.89	0.28	0.63	0.14	0.32	0.30	0.73
	0.3	0.93	1.00	0.82	1.00	0.93	1.00	0.50	0.98	0.29	0.72
	0.5	0.99	1.00	0.94	1.00	0.99	1.00	0.76	1.00	0.29	0.72
-0.5	0	0.33	0.76	0.25	0.64	0.30	0.69	0.16	0.38	0.30	0.73
	0.3	0.77	0.99	0.56	0.95	0.78	0.99	0.32	0.86	0.29	0.72
	0.5	0.93	1.00	0.77	0.99	0.92	1.00	0.49	0.96	0.29	0.72
0	0	0.33	0.76	0.18	0.49	0.32	0.75	0.18	0.45	0.30	0.73
	0.3	0.51	0.92	0.29	0.72	0.50	0.90	0.21	0.58	0.29	0.72
	0.5	0.68	0.98	0.43	0.88	0.67	0.96	0.25	0.72	0.29	0.72
0.5	0	0.33	0.76	0.25	0.64	0.35	0.81	0.19	0.48	0.30	0.73
	0.3	0.29	0.61	0.18	0.49	0.26	0.57	0.14	0.31	0.29	0.72
	0.5	0.36	0.73	0.20	0.54	0.31	0.61	0.13	0.30	0.29	0.72
1	0	0.33	0.76	0.44	0.89	0.39	0.86	0.19	0.50	0.30	0.73
	0.3	0.15	0.26	0.26	0.66	0.13	0.25	0.08	0.14	0.29	0.72
	0.5	0.16	0.27	0.22	0.59	0.11	0.17	0.06	0.09	0.29	0.72
1.3	0	0.33	0.76	0.59	0.97	0.41	0.88	0.19	0.51	0.30	0.73
	0.3	0.10	0.14	0.39	0.85	0.09	0.14	0.06	0.08	0.29	0.72
	0.5	0.10	0.12	0.36	0.82	0.05	0.06	0.04	0.04	0.29	0.72
1.4	0	0.33	0.76	0.64	0.98	0.42	0.89	0.19	0.52	0.30	0.73
	0.3	0.09	0.11	0.44	0.89	0.08	0.12	0.05	0.06	0.29	0.72
	0.5	0.08	0.09	0.42	0.88	0.04	0.04	0.03	0.03	0.29	0.72

Table 3: Rejection Rates when $\alpha_x \neq 0$ in equation (5), Constants, no trend.

α_x	$R^2\backslash c$	EJ -5	EJ -10	HW -5	HW -10	CADF -5	CADF -10	Zivot -5	Zivot -10	P_T -5	P_T -10
	0	0.34	0.76	0.19	0.59	0.11	0.24	0.10	0.19	0.30	0.73
-1	0.3	0.94	1.00	0.48	0.96	0.61	0.98	0.28	0.85	0.30	0.73
	0.5	0.99	1.00	0.72	0.99	0.89	1.00	0.50	0.98	0.29	0.72
	0	0.34	0.76	0.12	0.31	0.11	0.27	0.10	0.22	0.30	0.73
-0.5	0.3	0.78	1.00	0.25	0.74	0.38	0.88	0.17	0.58	0.30	0.73
	0.5	0.93	1.00	0.43	0.93	0.66	0.98	0.26	0.82	0.29	0.72
	0	0.33	0.76	0.10	0.23	0.12	0.31	0.11	0.24	0.30	0.73
0	0.3	0.53	0.92	0.13	0.38	0.21	0.56	0.11	0.32	0.30	0.73
	0.5	0.69	0.98	0.19	0.59	0.35	0.79	0.13	0.44	0.29	0.72
	0	0.33	0.76	0.12	0.32	0.13	0.35	0.12	0.27	0.30	0.73
0.5	0.3	0.30	0.61	0.09	0.24	0.11	0.25	0.09	0.18	0.30	0.73
	0.5	0.38	0.74	0.10	0.26	0.14	0.32	0.08	0.17	0.29	0.72
	0	0.34	0.76	0.19	0.59	0.14	0.40	0.13	0.31	0.30	0.73
1	0.3	0.16	0.27	0.11	0.34	0.05	0.09	0.07	0.10	0.30	0.73
	0.5	0.17	0.29	0.11	0.29	0.05	0.07	0.06	0.07	0.29	0.72
	0	0.34	0.76	0.27	0.77	0.15	0.43	0.13	0.32	0.30	0.73
1.3	0.3	0.11	0.15	0.17	0.53	0.04	0.05	0.05	0.06	0.30	0.73
	0.5	0.10	0.13	0.15	0.49	0.03	0.02	0.04	0.03	0.29	0.72
	0	0.34	0.76	0.31	0.82	0.15	0.44	0.14	0.33	0.30	0.73
1.4	0.3	0.09	0.12	0.19	0.61	0.03	0.04	0.05	0.05	0.30	0.73
	0.5	0.08	0.10	0.18	0.58	0.02	0.02	0.03	0.02	0.29	0.72

Table 4: Rejection Rates when $\alpha_x \neq 0$ in equation (5), No restrictions.

α_x	$R^2 \backslash c$	EJ		HW		CADF		Zivot		P_T	
		-5	-10	-5	-10	-5	-10	-5	-10	-5	-10
	0	0.12	0.32	0.13	0.41	0.08	0.15	0.08	0.13	0.10	0.27
-1	0.3	0.51	0.98	0.31	0.87	0.40	0.93	0.13	0.56	0.09	0.26
	0.5	0.75	1.00	0.51	0.98	0.74	0.99	0.21	0.85	0.09	0.26
	0	0.12	0.31	0.09	0.22	0.08	0.17	0.08	0.15	0.10	0.27
-0.5	0.3	0.33	0.87	0.16	0.56	0.24	0.69	0.09	0.32	0.09	0.26
	0.5	0.50	0.98	0.26	0.81	0.48	0.94	0.12	0.52	0.09	0.26
	0	0.11	0.31	0.08	0.16	0.09	0.19	0.08	0.16	0.10	0.27
0	0.3	0.20	0.56	0.10	0.26	0.13	0.36	0.08	0.18	0.09	0.26
	0.5	0.28	0.77	0.13	0.40	0.23	0.62	0.08	0.22	0.09	0.26
	0	0.11	0.31	0.09	0.22	0.09	0.21	0.09	0.18	0.10	0.27
0.5	0.3	0.11	0.26	0.08	0.16	0.07	0.15	0.07	0.12	0.09	0.26
	0.5	0.13	0.35	0.08	0.18	0.10	0.21	0.06	0.11	0.09	0.26
	0	0.12	0.31	0.13	0.41	0.10	0.24	0.09	0.21	0.10	0.27
1	0.3	0.06	0.10	0.09	0.23	0.04	0.05	0.06	0.09	0.09	0.26
	0.5	0.06	0.10	0.08	0.20	0.04	0.04	0.06	0.07	0.09	0.26
	0	0.12	0.32	0.17	0.60	0.10	0.26	0.10	0.22	0.10	0.27
1.3	0.3	0.04	0.05	0.12	0.36	0.03	0.03	0.06	0.06	0.09	0.26
	0.5	0.04	0.04	0.11	0.33	0.02	0.01	0.05	0.04	0.09	0.26
	0	0.12	0.32	0.19	0.65	0.11	0.27	0.10	0.23	0.10	0.27
1.4	0.3	0.04	0.04	0.13	0.42	0.02	0.02	0.05	0.05	0.09	0.26
	0.5	0.03	0.03	0.12	0.40	0.02	0.01	0.04	0.03	0.09	0.26

Table 5: Small Sample Size, Constants, no trend.

A		Θ		EJ			CADF		
a_1	a_2	θ_1	θ_2	$R^2=0$	$R^2=0.3$	$R^2=0.5$	$R^2=0$	$R^2=0.3$	$R^2=0.5$
0	0	0	0	0.059	0.070	0.074	0.051	0.055	0.058
0.2	0	0	0	0.060	0.077	0.089	0.047	0.058	0.062
0.8	0	0	0	0.066	0.083	0.103	0.075	0.085	0.086
0.2	0.5	0	0	0.089	0.078	0.076	0.104	0.069	0.059
0	0	-0.2	0	0.104	0.112	0.113	0.077	0.068	0.061
0	0	0.8	0	0.106	0.152	0.207	0.057	0.063	0.067
0	0	-0.5	0	0.167	0.169	0.189	0.102	0.082	0.068
0	0	-0.8	0	0.352	0.341	0.372	0.256	0.197	0.144
0.2	0	-0.5	0	0.160	0.166	0.181	0.104	0.085	0.070

Table 6: Small Sample Size, No restrictions.

A		Θ		EJ			CADF		
a_1	a_2	θ_1	θ_2	$R^2=0$	$R^2=0.3$	$R^2=0.5$	$R^2=0$	$R^2=0.3$	$R^2=0.5$
0	0	0	0	0.048	0.054	0.057	0.055	0.051	0.052
0.2	0	0	0	0.037	0.052	0.062	0.039	0.051	0.059
0.8	0	0	0	0.043	0.088	0.143	0.083	0.091	0.100
0.2	0.5	0	0	0.061	0.057	0.060	0.142	0.082	0.063
0	0	-0.2	0	0.098	0.104	0.123	0.091	0.073	0.064
0	0	0.8	0	0.057	0.122	0.207	0.059	0.067	0.075
0	0	-0.5	0	0.173	0.187	0.219	0.131	0.097	0.082
0	0	-0.8	0	0.371	0.324	0.350	0.299	0.232	0.176
0.2	0	-0.5	0	0.161	0.148	0.153	0.134	0.100	0.080

Table 7: Cointegration tests

	ADF	HW	CADF	P_T	DF-GLS
Forecast Errors	-2.72	7.43	-2.48*	2.38*	-2.72*
	(2)	(2)	(2)		(2)

Notes: The number of lags was chosen by MAIC and is reported in parenthesis. The star indicates significance at 5%.

Table 8: Cointegration tests stationary covariate.

	EJ, Case 3, Δp_t^e cov	$\hat{R}^2$	EJ, Case 3, Δp_t cov	$\hat{R}^2$
Forecast Errors	0.60**		0.60**	
	(2)	0.19	(1)	0.10
p-value	0.003		0.001	

Notes: The first two columns give results for tests where $X_t = p_t^e$, the remaining two columns give results for $X_t = p_t$. Two stars indicates significance at 1%.

Table 9: Unit Root Test with Breaks

	Min t-stat	Lags	Estimated Break
Panel A			
6 months	-4.37†	0	1994:II
Panel B			
6 months	-0.10	4	1979:II

Note: Panel A corresponds to the case in which the break is estimated as the date that minimizes the t-statistics in the unit root test, Panel B corresponds to the case in which the break is chosen to minimize the t-statistics testing in the break regression. (See Perron and Vogelsang (1992)). The small sample 5% and 10% critical value that take into account the lag selection procedure are -4.67 and -4.33 for Panel A and -3.68 and -3.35 for Panel B. * Denotes rejection at 5% while † denotes rejection at 10%.