EUI WORKING PAPERS
ECO No. 2006/18

Near-Optimal Unit Root Tests with Stationary Covariates with Better Finite Sample Size

ELENA PESAVENTO

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Near-Optimal Unit Root Tests with Stationary Covariates with Better Finite Sample Size

ELENA PESAVENTO

EUI Working Paper **ECO** No. 2006/18)

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50016 San Domenico di Fiesole (FI)
Italy

http://www.iue.it/
http://cadmus.iue.it/dspace/index.jsp

Near-Optimal Unit Root Tests with Stationary Covariates with Better Finite Sample Size.

Elena Pesavento
Emory University

April, 2006

Abstract. Numerous tests for integration and cointegration have been proposed in the literature. Since Elliott, Rothemberg and Stock (1996) the search for tests with better power has moved in the direction of finding tests with some optimality properties both in univariate and multivariate models. Although the optimal tests constructed so far have asymptotic power that is indistinguishable from the power envelope, it is well known that they can have severe size distortions in finite samples. This paper proposes a simple and powerful test that can be used to test for unit root or for no cointegration when the cointegration vector is known. Although this test is not optimal in the sense of Elliott and Jansson (2003), it has better finite sample size properties while having asymptotic power curves that are indistinguishable from the power curves of optimal tests. Similarly to Hansen (1995), Elliott and Jansson (2003), Zivot (2000), and Elliott, Jansson and Pesavento (2005) the proposed test achieves higher power by using additional information contained in covariates correlated with the variable being tested. The test is constructed by applying Hansen's test to variables that are detrended under the alternative in a regression augmented with leads and lags of the stationary covariates. Using local to unity parametrization, the asymptotic distribution of the test under the null and the local alternative is analytically computed.

Keywords: Unit Root Test, GLS detrending.

JEL Classification: C32.

I thank Michael Jansson, Alex Maynard and Barbara Rossi. Part of this research was done as a Jean Monnet Fellow in the Economics Department at the European University Institute. Presentation of this paper was aided by a faculty travel grant from ICIS at Emory University.

Corresponding author: Elena Pesavento, Department of Economics, Emory University, Emory GA30322, USA. Phone: (404) 712 9297. E-mail: epesave@emory.edu.

1. Introduction

Since the work of Fuller (1976) and Dickey and Fuller (1979) a large number of tests have been developed for the hypothesis that a variable is integrated of order one against the hypothesis that it is integrated of order zero. Motivating this considerable body of literature is the knowledge that a root equal to one can have a significant impact on the analysis of the long- and short-run dynamics of economic variables. Unit root testing is therefore considered an important step in economic modeling[1].

The seminal paper of Elliott, Rothenberg and Stock (1996, ERS thereafter) marked the point at which to stop the search for unit root tests with better power in an *univariate setting*. They show that no uniformly most powerful test for this problem exits, compute the power envelope for point-optimal tests of a unit root in an univariate model, and they derive a family of feasible tests (P_T) that have asymptotic power close to the power envelope. In fact the asymptotic power of the P_T test is never much below the envelope and it is tangent to the power envelope at one point. In this sense, the ERS tests are approximately most powerful.

One feature of the ERS approach is that the variables are detrended under the alternative (or GLS detrended). ERS also propose a version of the Augmented Dickey Fuller (ADF) t-test where the variable have been GLS detrended before estimating the regression ($ADF-GLS$ test). Although the $ADF-GLS$ does not have the same optimality justifications of the P_T test, it performs similarly in term of power while having better size properties, and it is easier to compute. For this reason, practitioners use the $ADF-GLS$ more often that the P_T test.

The search for tests with better power is now moving in the directions of multivariate models. Hansen (1995) shows that additional information contained in stationary covariates that are correlated with the variables of interest can be exploited to obtain tests that have higher power than univariate tests. Hansen (1995) computes the power envelope for unit root tests in the presence of stationary covariates in a model with no deterministic terms, while Elliott and Jansson (2003) generalize the results to the case in which the model includes a constant and/or time trends. Both papers illustrate the significant increase in the asymptotic power envelope in multivariate models achieved by including stationary covariates. To implement a feasible test, Hansen (1995) proposes covariate augmented Dickey-Fuller ($CADF$) tests computed as $t-$tests in a ADF regression augmented by leads and lags of the stationary covariates. Elliott and Jansson (2003) construct a family of point-optimal tests (EJ thereafter), similar in spirit to the P_T tests, that are feasible and that attain the power envelope at a point. Both Hansen (1995) and Elliott and Jansson (2003) tests are generalization of the ADF and P_T tests and, in fact, they have the same asymptotical distribution of ADF and P_T respectively when there is no information in the stationary covariates, i.e. the correlation between the stationary covariate and

[1] Exceptions are Rossi (2005), Rossi (2006), Pesavento and Rossi (2006), and Jansson and Moreira (2006), where inference in robust to the presence of exact unit roots.

the variable being tested is zero. Both the $CADF$ and EJ tests have power than is higher than the power of ADF and P_T when the correlation is different than zero with gains that get larger as the correlation increases. Not only both the $CADF$ and EJ tests outperform univariate tests, but there are also significant differences between them. As expected, given that ADF and P_T are special cases of $CADF$ and EJ when no stationary covariates are included, the differences are similar to the differences between ADF and P_T in univariate models. Elliott and Jansson (2003) show that EJ can significantly outperform $CADF$ in term of power although it can be slightly worse in term of size distortions.

The goal of this paper is to propose a generalization of the $CADF$ test that is similar to the GLS generalization of the ADF test, and that apply to a model with stationary covariates. The test is constructed by applying GLS detrending to each variable according to the assumptions on the deterministic terms, and then estimating an augmented regression with lags and leads of the stationary covariates. To keep with Hansen's notation, the test is called $CADF-GLS$. Similarly to the $ADF-GLS$ test, the proposed test is intuitive and it is easy to compute. Section 2 describes the model while Section 3 analytically computes the asymptotic distribution of the test under the null and local alternative hypotheses. Section 4 shows that, although this test is not optimal in the sense of Elliott and Jansson (2003), it has better finite sample size properties while having asymptotic power curves close to the power envelope. Although the general model of Section 2 does not allow for cointegration, Elliott, Jansson and Pesavento (2005) show that the problem of testing for the null of no cointegration in cases in which there is only one cointegration vector that is known a-priori is isomorphic to the unit root testing problem studied in Elliott and Jansson (2003). Section 5 briefly discusses the known cointegration case. The more general case of unknown cointegration vectors should be modeled accordingly and it is left for future research. Section 6 discusses the relevance of the assumption of stationarity of the covariate and Section 7 concludes.

2. Model: no cointegration

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

$$x_t = \mu_x + \tau_x t + u_{x,t} \tag{1}$$

$$y_t = \mu_y + \tau_y t + u_{y,t} \tag{2}$$

and

$$\Phi(L) \begin{pmatrix} u_{x,t} \\ (1-\rho L) u_{y,t} \end{pmatrix} = \varepsilon_t, \tag{3}$$

where y_t is univariate, x_t is of dimension $m \times 1$, $\Phi(L)$ is a matrix polynomial of possible infinite order in the lag operator L with first element equal to the identity matrix. I am interested in the problem of testing for the presence of a unit root in y_t:

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

Following Elliott and Jansson (2003), define $u_t(\rho) = \left[u'_{x,t}, u_{y,t}(\rho)\right]' = \left[u'_{x,t}, (1-\rho L) u_{y,t}\right]' = \Phi(L)^{-1} \varepsilon_t$ and $\Gamma(k) = E\left[u_t(\rho) u_{t+k}(\rho)\right]$ the autocovariance function of $u_t(\rho)$. The following will be assumed throughout the paper:

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

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

Assumption 3: $E_{t-1}(\varepsilon_t) = 0$ (a.s.), $E_{t-1}(\varepsilon_t \varepsilon'_t) = \Sigma$ (a.s.), and $\sup_t E \|\varepsilon_t\|^{4+\delta} < \infty$ for some $\delta > 0$, where Σ is positive definite, $E_{t-1}(\cdot)$ refers to the expectation conditional on $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

Assumption 4: The covariance function of $u_t(\rho)$ is absolute summable such that $\sum_{j=-\infty}^{+\infty} \|\Gamma(k)\| < \infty$ and $\sum_{j=-\infty}^{+\infty} j \ \ \|\Gamma(k)\| < \infty$.

Assumptions 1-3 are fairly standard and are similar to (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)).

Assumption 1-3 imply that

$$T^{-1/2} \sum_{t-1}^{[T.]} u_t(\rho) \Rightarrow \Omega^{1/2} W(\cdot)$$

where $\Omega = \Phi(1)^{-1} \Sigma \Phi(1)'^{-1}$ is 2π times the spectral density at frequency zero of $u_t(\rho)$ such that the spectral density of $u_t(\rho)$, $f_{u(\rho)u(\rho)}(\lambda)$ is bounded away from zero. Partition Ω and $\Phi(L)$ conformably to z_t as

$$\Omega = \begin{bmatrix} \Omega_{xx} & \omega_{xy} \\ \omega_{yx} & \omega_{yy} \end{bmatrix}$$

and

$$\Phi(L) = \begin{bmatrix} \Phi_{xx}(L) & \Phi_{xy}(L) \\ \Phi_{yx}(L) & \Phi_{yy}(L) \end{bmatrix}$$

and define $R^2 = \delta'\delta$ where $\delta = \Omega_{xx}^{-1/2} \omega_{xy} \omega_{yy}^{-1/2}$ is a vector containing the bivariate zero frequency correlations between the shocks to x_t and the quasi-difference of the shocks to y_t. R^2 is the multiple coherence of $(1-\rho L) y_t$ with x_t at frequency zero (Brillinger (2001), p. 296) and it measures the extent to which the quasi-difference of y_t is determinable from the $m-$vector valued x_t by linear time invariant operations.

R^2 lies between zero and it is zero when there is no long run correlation between x_t and the quasi-difference of y_t. As in Elliott and Jansson (2003), R^2 is assumed to be strictly less than one, thus ruling out the possibility that under the null, the partial sum of x_t cointegrates with y_t. The case in which a cointegration vector is present should be modeled to take account of cointegration and it is outside the scope of this paper unless an unique cointegration vector is known a-priori as in Elliott, Jansson and Pesavento (2005). This case is discussed in section 5.

I will consider five cases for the deterministic part of the model:

Case 1: $\mu_x = \mu_y = 0$ and $\tau_x = \tau_y = 0$.

Case 2: $\mu_x = 0$ and $\tau_x = \tau_y = 0$.

Case 3: $\tau_x = \tau_y = 0$.

Case 4: $\tau_x = 0$.

Case 5: no restrictions.

These cases represent a fairly general set of models that are relevant in empirical applications.

Hansen (1995) and Elliott and Jansson (2003) show that when R^2 is different from zero, the stationary covariate x_t contains information that can be exploited to obtain unit root tests that have power higher than standard univariate tests. Hansen (1995) suggests a covariate augmented Dickey-Fuller test ($CADF$) while Elliott and Jansson (2003) constructs a family of feasible tests (EJ) that are close to the power envelope.

Model (1) $-$ (3) is slightly more general than Hansen (1995) and Elliott and Jansson (2003) as it allows for short run dynamics of unknown and possibly infinite order.[2] Under assumption (A1)-(A4) we can write (Saikonnen, 1991):

$$u_{y,t}(\rho) = \sum_{j=-\infty}^{+\infty} \widetilde{\pi}'_{x,j} u_{x,t-j} + \eta_t \tag{4}$$

where the summability condition $\sum_{j=-\infty}^{+\infty} \|\widetilde{\pi}_{x,j}\| < \infty$ holds and η_t is a serially correlated stationary process such that $E\left(u'_{x,t}\eta_{t+k}\right) = 0$ for any $k = 0, \pm 1, \pm 2, ...$ The spectral density of η_t is $f_{\eta\eta}(\lambda) = f_{u_y(\rho)u_y(\rho)}(\lambda) - f_{u_y(\rho)u_x}(\lambda) f_{u_x u_x}(\lambda)^{-1} f_{u_x u_y(\rho)}(\lambda)$ so

$$2\pi f_{\eta\eta}(0) = \omega_{y.x} = \omega_{yy} - \omega_{yx}\Omega_{xx}^{-1}\omega_{xy}$$

Denote the detrended variables with superscript d so, $x_t^d = x_t - \hat{\mu}_x - \hat{\tau}_x t$, where $\hat{\mu}_x$ and $\hat{\tau}_x$ are OLS estimates of mean and trend of the stationary variable x_t and

[2] Details on how to choose a finite lag length in practice will be discussed later.

$y_t^d = y_t - \tilde{\mu}_y - \tilde{\tau}_y t$ where $\tilde{\mu}_y$ and $\tilde{\tau}_y$ are estimates of mean and trend of y_t either by OLS or by GLS. We can write $y_t^d = u_{y,t} + \left(\mu_y - \tilde{\mu}_y\right) + \left(\tau_y - \tilde{\tau}_y\right) t$ so that $\Delta y_t^d = (\rho - 1)\, y_{t-1}^d + (1-\rho)\left(\mu_y - \tilde{\mu}_y\right) + (1-\rho L)\left(\tau_y - \tilde{\tau}_y\right) t\ + u_{y,t}(\rho)$. Using (4) and the fact that we can write $u_{x,t} = x_t^d + \left(\hat{\mu}_{x-}\mu_x\right) + \left(\hat{\tau}_x - \tau_x\right) t$ we have that

$$\Delta y_t^d = \alpha y_{t-1}^d + \sum_{j=-\infty}^{+\infty} \widetilde{\pi}'_{x,j} x_{t-j}^d + \bar{\eta}_t \tag{5}$$

where $\alpha = (\rho - 1)$ and

$$\bar{\eta}_t = \widetilde{\pi}_x(1)'\left(\hat{\mu}_{x-}\mu_x\right) + \sum_{j=-\infty}^{+\infty} \widetilde{\pi}'_{x,j}\left(\hat{\tau}_x - \tau_x\right)(t-j) \tag{6}$$

$$+ (\rho - 1)\left(\tilde{\mu}_y - \mu_y\right) - (1-\rho L)\left(\tilde{\tau}_y - \tau_y\right) t + \eta_t \tag{7}$$

with $\widetilde{\pi}_x(1) = \sum_{j=-\infty}^{+\infty} \widetilde{\pi}'_{x,j} = \omega_{yx}\Omega_{xx}^{-1}$. η_t in is uncorrelated at all leads and lags with x_t^d but it is serially correlated and the asymptotic distribution of tests on α in (5) will depend on nuisance parameters. Modified version of the tests can be constructed as in Phillips and Perron (1988) by using non parametric estimates of the nuisance parameters or by augmenting the regression with lags of Δy_t^d to obtain errors that are white noises as in Hansen (1995). The test suggested by Hansen (1995) is then based on the t-statistics on a augmented regression in which lagged, contemporaneous and future values of the stationary covariate are included[3]:

$$\Delta y_t^d = \varphi y_{t-1}^d + \textstyle\sum_{j=-\infty}^{\infty} \pi'_{x,j} x_{t-j}^d + \sum_{j=1}^{\infty} \pi_{y,j} \Delta y_{t-j}^d + \xi_t \tag{8}$$

where $\varphi = \psi(1)(\rho - 1)$. Since the sequence $\{\widetilde{\pi}_{x,j}\}$ is absolute summable $\widetilde{\pi}_{x,j} \approx 0$ for $|j| > k$ for k large enough and we can in practice approximate (8) with a finite number of lags:

$$\Delta y_t^d = \varphi y_{t-1}^d + \textstyle\sum_{j=-k}^{k} \pi'_{x,j} x_{t-j}^d + \sum_{j=1}^{k} \pi_{y,j} \Delta y_{t-j}^d + \xi_{tk} \tag{9}$$

with $\xi_{tk} = \xi_t + \sum_{|j|>k} \pi'_{x,j} x_{t-j}^d + \sum_{j>k} \pi_{y,j} \Delta y_{t-j}^d$. The intuition behind this approach is that the correlation between y_t^d and x_t^d can help in reducing the error

[3] Derivation can be found in the Appendix.

variance thus resulting in more precise regression parameter estimates. The asymptotic distribution for the t-statistics on φ is different from the distribution of the ADF test and a significant increase in the asymptotic power for local alternatives can be obtained with the inclusion of the covariate.

Model $(1)-(3)$ allows for autoregressive processes of infinite order and a condition on the expansion rate of the truncation lag k is necessary. The following condition is assumed throughout the paper:

Assumption 5: $T^{-1/3}k \to 0$ and $k \to \infty$ as $T \to \infty$.

The condition in Assumption 5 specifies an upper bound for the rate at which the value k is allowed to tend to infinity with the sample size. Ng and Perron (1995) show that conventional model selection criteria like AIC and BIC yield $k = O_p(\log T)$, which satisfies Assumption 5.[4]

To implement the test in practice we recommend the following steps:

1. Construct the quasi-differenced y_t as $y_t(\bar{\rho}) = (1-\bar{\rho}L)\,y_t$ for $t>1$ and $y_1(\bar{\rho}) = y_1$ where $\bar{\rho} = 1 + (\bar{c}/T)$ and $\bar{c} = -7$ for Case 1-3 and $\bar{c} = -13.5$ for Cases 4-5[5]. The GLS or quasi-difference detrended $y_t^d(\bar{\rho})$ is computed as $y_t(\bar{\rho}) - d_t(\bar{\rho})'\,\hat{\mu}(\bar{\rho})$ where $\hat{\mu}(\bar{\rho})$ is the OLS estimator from regressing $y_t(\bar{\rho})$ on $d_t(\bar{\rho})$. The choice of $d_t(\bar{\rho})$ will depend of the determinist case chosen: For Case 1 $d_t(\overline{\rho})' = 0$, for Cases 2 and 3 $d_t(\overline{\rho})' = (1-\bar{\rho})$ when $t>1$ and $d_1(\bar{\rho}) = 1$, and for Cases 4 and 5 $d_t(\overline{\rho})' = \begin{bmatrix} (1-\bar{\rho}) & (1-\bar{\rho}L)\,t \end{bmatrix}$ $t>1$ and $d_1(\bar{\rho})' = \begin{bmatrix} 1 & 1 \end{bmatrix}$.
2. Detrend the stationary covariates x_t. Given the assumption that x_t is stationary there is no reason to use GLS detrending so x_t^d is OLS demeaned for Case 3 and 4 and OLS demeaned and detrended for Case 5.
3. Estimate $\hat{R}^2 = \hat{\omega}_{yx}\hat{\Omega}_{xx}^{-1}\hat{\omega}_{xy}\omega_{yy}^{-1}$ where $\hat{\Omega}$ is estimated non parametrically as in Hansen (1995) as the LR variance covariance matrix of $\hat{\xi}_{tk}$ and $\hat{\xi}_{tk} + \sum_{j=-k}^{k} \hat{\pi}'_{x,j} x_{t-j}^d$.

[4] Often, a second condition is also assumed to impose a lower bound on k. The lower bound condition is only necessary to obtain consistency of the parameters on the stationary variables, and it is sufficient but not necessary to prove the limiting distribution of the relevant test statistics (Ng and Perron, 1995, Lutkepohl and Saikonnen, 1999). Because I am only interested in the t-ratio statistics, Assumption 5 is necessary and sufficient to prove the asymptotic distribution of the tests.

[5] Although the choice of $\bar{c}$ is not irrelevant, a complete discussion on the optimal choice of $\bar{c}$ is outside the scope of this paper. To make a resonable comparison with existing tests I will then use the same values for $\bar{c}$ that were originally suggested by Elliott and Jansson (2003), that is $\bar{c} = -7$ for Case 1-3 and $\bar{c} = -13.5$ for Cases 4 and 5.

4. Estimate the augmented regression (9) without any deterministic terms with the lag length chosen by some criteria satisfying assumption 4 as BIC. The $CADF-GLS$ test is obtained using the t-statistics on φ and the critical values in Table 1 for the corresponding estimated $\hat{R}^2$.

One key assumption for the validity of all three tests compared in this paper is the stationarity of the covariate x_t, Assumption 2. Violations of this assumption will invalidate the results. Of course this problem is only relevant in situations in which x_t is persistent and it is difficult to detect if it is stationary. To avoid this problem, Hansen (1995) recommends taking first differences before including highly serially correlated variables in the augmented regressions. He shows by simulations that when, x_t is persistent but not a unit root, over differencing x_t and therefore including Δx_t as the stationary covariate results in only mild power losses[6]. In empirical applications then it is recommended to include stationary covariates in level when the researcher is sure they are stationary, and in first differences when they are persistent even if not exactly a unit root.

3. Asymptotic Power Functions

Since all tests are consistent, they all have power equal to one asymptotically and the asymptotic power for fixed alternatives cannot be used to rank the tests. As it is standard in the literature, I will compute the asymptotic distribution of the test under a sequence of local alternatives of the type $\rho = 1 + \frac{c}{T}$ where c is a constant less than zero. When c is equal to zero, the errors are integrated of order one. For negative c and fixed T, the variables in equation (2) are stationary. The local to unity asymptotics is used to obtain an approximation to the distribution of consistent tests that mimics their behavior in finite sample and allows a meaningful comparison of their power properties. By using this parameterization and the results of Phillips (1988), I evaluate and compare the power of the tests for integration presented in the previous section.

Theorem 1 generalizes the results proved by Hansen (1995) in the context of model $(1)-(3)$ for more general cases for the deterministic terms and for an infinite order polynomial, which is approximated by a finite lag length k chosen by a data dependent criteria.

[6] The intuition is that when x_t is highly persistent we can modeled as local to unity so that

$$\Delta x_t = \frac{g}{T} x_{t-1} + u_{xt}$$

and the first difference of x_t is equal to the stationary variable u_{xt} plus one extra term $\frac{g}{T} x_{t-1}$ that disappear when $T \to \infty$.

Theorem 1 [OLS Detrending]**.** *When the model is generated according to* (1)−(3) *,with* $T(\rho-1)=c$*, and Assumption 1 to 4 are valid, then, as* $T\to\infty$*:*

$$\hat{t}_{\hat{\varphi}}^{ADF}\Rightarrow c\left(\int J_{xyc}^{d2}\right)^{1/2}+\left(\int J_{xyc}^{d2}\right)^{-1/2}\left(\int J_{xyc}^{d}dW_2\right)$$

where $J_{xyc}(r)$ *is a Ornstein-Uhlenbeck process such that*

$$J_{xyc}(r)=W_{xy}(r)+c\int_0^1 e^{(\lambda-s)c}W_{xy}(s)\,ds,$$

$W_{xy}(r)=\sqrt{\frac{R^2}{1-R^2}}W_x(r)+W_y(r)$*,* $W_x(r)$ *and* $W_y(r)$ *are independent standard Brownian Motions, and*

1. *(Case 1)* $J_{xyc}^{d}(r)=J_{xyc}(r)$ *and no deterministic terms are included in the regression,*
2. *(Case 2-Case 3)* $J_{xyc}^{d}(r)=J_{xyc}(r)-\int J_{xyc}(s)\,ds$ *and a constant is included in the regression,*
3. *(Case 4-Case 5) and* $J_{xyc}^{d}(r)=J_{xyc}(r)-(4-6r)\int J_{xyc}(s)\,ds-(12r-6)\int sJ_{xyc}(s)\,ds$ *and a mean and trend are included in the regression.*

The asymptotic distribution of the test is the same as Hansen (1995) in his special case in which the errors terms in equation (9) are uncorrelated with x_{t-k}, which holds in well-specified dynamic regressions. As expected the local power depends only on R^2, the long run correlation between the shocks to x_t and the quasi-difference of the shocks to y_t.[7]

Elliott and Jansson (2003) follow the general methods of King (1980, 1988) and examine Neyman-Pearson type of tests in the context of model $(1)-(3)$ to compute the power envelope for the family of point optimal tests for each possible case of the deterministic terms (Case 1 to Case 5)[8], and to construct feasible general tests that are asymptotically equivalent to the power envelope. As it is intuitive, the more the covariate is correlated with the quasi-difference of y_t the higher the power of the test: the asymptotic distribution of the tests depends on the parameter R^2 both under the null and the local alternative $\rho=1+(c/T)$. As R^2 increases, there is a larger gain in using the information contained in the stationary covariate over an univariate test and the power increases.

---

[7] Note that R^2 in this paper corresponds to $1-\rho^2$ in Hansen(1995)'s notation.

[8] Hansen (1995) also computes the power envelope for the less general case in which there are not deterministic terms present.

Although both $CADF$ and EJ have power that is larger than univariate tests, the gain in term of power from using an optimal test over the standard t-test in Hansen (1995) can be quite large. In some cases (depending on the deterministic case considered) the power of EJ can be up to 2-3 times larger than the power of $CADF$. The difference between EJ and $CADF$ is similar to the difference between ADF and P_T in the univariate case. In fact, when R^2 is zero and there is no gain in using the stationary covariate, the asymptotic distributions of $CADF$ and EJ are equivalent respectively to the asymptotic distributions of ADF and P_T.

In the context of univariate tests, Elliott, Rothenberg and Stock (1996) show that, although P_T has higher power than ADF, the ADF test has smaller size distortions. Interestingly, Elliott, Rothenberg and Stock (1996) propose an alternative test that is computed by first detrending the variable under a local alternative and then applying the ADF with no deterministic terms (*ADF-GLS*). Although this test does not have the same optimality justification of the P_T test, simulations show that it has almost identical power properties while having slightly better size properties. As the *ADF-GLS* is easier to implement while having similar power and better size properties, in most cases practitioners prefer to use the *ADF-GLS* over P_T tests.

The test proposed in this paper is similar in spirit to the *ADF-GLS* test, in the context of the multivariate model $(1)-(3)$. The test is derived by applying Hansen's $CADF$ test to variables that have been previously detrended under the alternative as described in the previous section. The following theorem characterize the asymptotic distribution of the proposed test.

Theorem 2 [GLS Detrending]**.** *When the model is generated according to* $(1)-(3)$ *,with* $T(\rho-1)=c$, *and Assumption 1 to 4 are valid, then, as* $T\rightarrow\infty$*:*

$$\hat{t}_{\hat{\varphi}}^{GLS} \Rightarrow c\left(\int J_{xyc}^{d2}\right)^{1/2}+\left(\int J_{xyc}^{d2}\right)^{-1/2}\left(\Lambda_c(r)+\int J_{xyc}^{d}dW_2\right)$$

where $J_{xyc}(r)$ *and* $W_{xy}(r)$ *are as defined in Theorem 1, and*

1. *(Case 1- Case2)* $\Lambda_c(r)=0$ *and* $J_{xyc}^d(r)=J_{xyc}(r)$.

2. *(Case 3)* $\Lambda_c(r)=\int J_{xyc}^d\sqrt{\frac{R^2}{1-R^2}}W_x(1)$ *and* $J_{xyc}^d(r)=J_{xyc}(r)$.

3. *(Case 4)* $\Lambda_c(r)=\int J_{xyc}^d\sqrt{\frac{R^2}{1-R^2}}W_x(1)+V_c\cdot\left[c\int rJ_{xyc}^d-\int J_{xyc}^d\right]$, $V_c=\lambda J_{xyc}(1)+(1-\lambda)3\int sJ_{xyc}(s)\,ds$, *with* $\lambda=\frac{1-\bar{c}}{1-\bar{c}+\bar{c}^2/3}$ *and* $J_{xyc}^d(r)=J_{xyc}(r)-rV_c$.

4. *(Case 5)* $\Lambda_c(r)=\int J_{xyc}^d\sqrt{\frac{R^2}{1-R^2}}\left[-2W_x(1)+6\int W\right]+\int rJ_{xyc}^d\sqrt{\frac{R^2}{1-R^2}}\left[-6W_x(1)+12\int W\right]+V_c\cdot\left[c\int rJ_{xyc}^d-\int J_{xyc}^d\right]$, *where* V_c *and* $J_{xyc}^d(r)$ *are as in Case 4.*

As in Theorem 1, the local power of $CADF - GLS$ depends on the point in the alternatice space c and on the single nuisance parameter R^2 determining the usefulness of the stationary covariates. As in Hansen (1995) and Elliott and Jansson (2003), R^2 also affects the distribution under the null so we need a different critical value for each R^2. Table 1 report the critical values for the $CADF - GLS$. As in Elliott Rothemberg and Stock (1996) for Case 2, the estimated mean is stochastically bounded and the asymptotic distribution is independent of on the values of $\bar{c}$ used in the detrending and identical to the case with no mean (i.e identical to the standard case in Hansen's (1995) Table 1).

TABLE 1- ABOUT HERE

Figure 1 compares the power functions of the $CADF-GLS$, $CADF$ and EJ tests for the deterministic case 2.[9]. The power functions are computed as the probability that the tests are less than some critical value. Given the expression for the limit distribution of all the tests, the asymptotic local power can be approximated by simulating the distributions presented in in Theorem 1 and 2 and in Elliott and Jansson (2003). Each Brownian Motion's piece in the asymptotic distribution is approximated by step functions using Gaussian random walk with $T = 1000$ observations. $10,000$ replications are used to find the critical values and the rejection probabilities for each c and R^2. Since the local power for all the tests depends solely on one nuisance parameter, the power functions of the tests are compared for different values of R^2. Notice that while the asymptotic distributions in Figure 1 do not depend on the particular estimator used to estimate other nuisance parameters, the finite size sample properties of tests can be sensitive to the choice of estimation method for the nuisance parameters. In this respect, the local asymptotic curves presented in this section should be interpreted as approximations to the finite sample size-adjusted power curves of the corresponding tests.

FIGURE 1- ABOUT HERE

As Figure 1 shows the asymptotic power function of the proposed test is very close to the asymptotic power function of EJ test. For high values of R^2, and point in the alternative close to the null, the rejection probability of $CADF - GLS$ can be even slightly higher although still below the power envelope (not reported). As R^2 increases the power curves of all three tests shift to the right as more information can be exploited to increase power. Although the asymptotic power is independent of nuisance parameters other than R^2, this is not the case in finite samples. Comparison of the rejection rates in small samples not reported[10] confirm that the asymptotic ranking of the tests in Figure 1 is a good approximation for the small sample behavior of the tests.

[9] Results for other cases are similar and available upon request.

[10] The results are available from the author upon request.

4. Size Comparison in Small Samples

To compare the tests in term of size distortions more dynamic in the error terms is allowed. The error process $u_t = (u_{y,t}, u_{x,t})'$ is generated by the VARMA(1,1) model $(I_2 - AL)\, u_t = (I_2 + \Theta L)\, \varepsilon_t$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 u_t satisfies

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

To replicate what empirical practitioners face, the true number of lags is assumed unknown and it is estimated using the $MAIC$ by Ng and Perron (2001) on a univariate regression on the GLS detrended y_t. The maximum number of lags allowed is 8. To have a meaningful comparison the same number of lags is used for all three tests.[11] For Case 2 and 3, the regressions are estimated with a mean. For Case 4 and 5 the model is estimated with mean and trend. The sample size is $T = 100$ and $10,000$ replications are used.

TABLES 2-5 ABOUT HERE

Tables 2-5 compare the small sample size of Elliott and Jansson's (2003) test, Hansen's (1995) $CADF$ test, and $CADF$-GLS test for various values of Θ and A. To compute the critical values in each case we interpolate the critical values and estimate R^2as 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 $CADF$ tests emphasizing a trade-off between size and power. This is not surprising as this is the same type of difference found between the P_T and ADF tests in the univariate case and 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 trend (Cases 4 and 5). The proposed $CADF - GLS$ test, while having power similar to EJ test as we saw, has better size properties. The improvement in size is more evident when a deterministic trend is present in y_t, as in the univariate case, and for large values of R^2. With a large autoregressive root for example, size of EJ can be between 13% and 30% (Case 3, $a_1 = 0.8$) while in that case the size of $CADF - GLS$ is around 7%. When Θ is nonzero both tests present size distortions that are severe in the presence of a large negative moving average root (as is the case for unit root tests), emphasizing the need of proper modeling of

[11] The choice of the number of lags of course will affect the performance of the test. Another option would be to estimate the number of lags by BIC or AIC in a VAR in first difference (under the null). Simulations not reported show that this method estimates very conservative number of lags and delivers large size distortions.

the serial correlation present in the data.[12] Although the size is still not equal to the nominal values of 5%, the gains in term of better size properties from using the proposed tests is quite remarkable in the presence of MA roots. For large R^2 and $\theta_1 = -0.5$ and -0.8 the size of $CADF - GLS$ can be less than half the size of EJ.

5. Model: Cointegration

The case of one known cointegration vector can be modeled in the same framework of Section 2. Consider the problem of testing for a unit root in a cointegrating vector when the cointegrating vector is known and the variables are known to be $I(1)$. The model for this problem is:

$$x_t = \mu_x + \tau_x t + u_{x,t} \tag{10}$$
$$y_t = \mu_y + \tau_y t + \gamma' x_t + u_{y,t} \tag{11}$$

and

$$\Phi(L)\begin{pmatrix} \Delta u_{x,t} \\ (1-\rho L)\, u_{y,t} \end{pmatrix} = \varepsilon_t, \tag{12}$$

where as before where y_t is univariate, x_t is of dimension $m \times 1$, $\Phi(L) = I_{m+1} - \sum_{j=1}^{k} \Phi_j L^j$ is a matrix polynomial in the lag operator L. The hypothesis of interest is again $H_0 : \rho = 1$ vs. $H_1 : -1 < \rho < 1$. Under the null hypothesis and the assumptions of section 2, 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 under the alternative. The value of γ, the parameter that characterizes the potentially cointegrating relation between y_t and x_t is assumed to be known to the researcher. The model is similar to the unit root case analyzed in Section 2, except that now x_t has a unit root and it is the first difference of x_t that helps in explaining the variability in the quasi-difference of the errors in the cointegration regression (11). The relevant cases for the deterministics in this case are:

Case 1: $\mu_y = 0$, $\tau_x = 0$, $\tau_y = 0$.

Case 2: $\tau_x = 0$, $\tau_y = 0$.

Case 3: $\tau_y = 0$.

[12] In fact, size properties of all three tests are very sensitive to the choice of the lag length. There is a large literature on the choice of the order of AR and VAR models, A complete analysis of the sensitivities of the tests to the choice of the lag length would be interesting but outside the scope of this paper.

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 and fourth cases have x_t with a unit root and drift with a constant (Case 3) or constant and trend (Case 4) in the cointegrating vector. The case in which Δx_t has a drift and time trend which would corresponds to Case 5 in section 2 seems unlikely in practice and so it is not considered. .

Elliott, Jansson and Pesavento (2005) show that the 'known a priori' information (that is that x_t has a unit root under both the null and alternative hypothesis) imposes restrictions on the error correction coefficient in a VECM that renders the representation $(10)-(11)$ equivalent to a VECM with one cointegration vector. Apart from deterministic terms, the problem of testing for no cointegrating vector is therefore isomorphic to the unit root testing problem Section 2 and a point optimal test for this hypothesis is again the test proposed by Elliott and Jansson (2003). Hansen's (1995) $CADF$ test can also be applied and it is in fact equivalent to a t-test on φ, the error correction term, in the error correction regression:

$$\Delta y_t^d = \varphi \left(y_t^d - \gamma' x_t^d \right) + \textstyle\sum_{j=-k}^{k} \pi'_{x,j} x_{t-j}^d + \sum_{j=1}^{k} \pi_{y,j} \Delta y_{t-j}^d + \xi_{tk} \qquad (13)$$

With the exception of small differences in the treatment of the first observation for x_t in the GLS detrending,[13] the $CADF-GLS$ test proposed in this paper can then be used to test for no cointegration when the cointegration vector is known.

6. Conclusion

This paper proposes a simple and powerful unit root test that has power close to the most power unit root tests currently available, while having better size properties. Similarly to recent literature on unit root testing, the proposed test achieves higher power by using additional information contained in covariates correlated with the variable being tested. The test is constructed by applying Hansen's (1995) tests to variables that are detrended under the alternative in a regression augmented with leads and lags of the stationary covariates. The proposed tests is easy to compute, has power higher than Hansen's (1995) test and close to the power of Elliott and Jansson's (2003) optimal test and to the power envelope, and displays better size properties than Elliott and Jansson (2003).

[13] See Elliott, Jansson and Pesavento (2005).

Table 1: Asymptotic 5% Critical Values for the CADF-GLS t-statistics.

R^2	Case1, 2	Case 3	Case 4	Case 5
0	-1.948	-1.948	-2.836	-2.835
0.1	-1.939	-1.909	-2.786	-2.780
0.2	-1.929	-1.866	-2.738	-2.730
0.3	-1.918	-1.812	-2.688	-2.664
0.4	-1.905	-1.760	-2.628	-2.586
0.5	-1.881	-1.707	-2.568	-2.497
0.6	-1.864	-1.647	-2.498	-2.401
0.7	-1.839	-1.579	-2.418	-2.286
0.8	-1.818	-1.497	-2.343	-2.152
0.9	-1.773	-1.405	-2.315	-2.017

Critical values are calculated from 60000 replications of samples of size 1000 drawn with i.i.d. Gaussian innovations.

Table 2: Small Sample Size, Deterministic Case 2.

A	a_1	0	0.2	0.8	0.2	0	0	0	0	0.2
	a_2	0	0	0	0.5	0	0	0	0	0
Θ	θ_1	0	0	0	0	-0.2	0.8	-0.5	-0.8	-0.5
	θ_2	0	0	0	0	0	0	0	0	0
	$R^2 = 0$	0.059	0.029	0.070	0.125	0.139	0.081	0.252	0.639	0.248
EJ	$R^2 = 0.3$	0.064	0.042	0.088	0.103	0.122	0.091	0.209	0.564	0.204
	$R^2 = 0.5$	0.058	0.042	0.099	0.094	0.113	0.102	0.197	0.544	0.197
	$R^2 = 0.7$	0.062	0.046	0.119	0.098	0.115	0.137	0.210	0.561	0.209
	$R^2 = 0$	0.052	0.034	0.077	0.105	0.119	0.055	0.182	0.607	0.222
$CADF$	$R^2 = 0.3$	0.056	0.051	0.082	0.075	0.094	0.061	0.131	0.491	0.157
	$R^2 = 0.5$	0.059	0.064	0.085	0.059	0.071	0.066	0.093	0.383	0.108
	$R^2 = 0.7$	0.054	0.069	0.087	0.049	0.053	0.067	0.063	0.234	0.066
	$R^2 = 0$	0.072	0.033	0.088	0.088	0.166	0.070	0.238	0.589	0.280
$CADF$-GLS	$R^2 = 0.3$	0.079	0.059	0.094	0.078	0.128	0.078	0.175	0.481	0.204
	$R^2 = 0.5$	0.071	0.067	0.091	0.076	0.096	0.075	0.127	0.371	0.147
	$R^2 = 0.7$	0.065	0.074	0.085	0.077	0.068	0.070	0.078	0.229	0.087
$Ave\ \ Lags$		0.4	1.4	2.6	2.8	1.6	5.6	3.7	6.2	2.9

Lags chosen by BIC with a maximum of 8 (could use MAIC for $CADF - GLS$), $T = 100,\ NMC = 10,000$.

Table 3: Small Sample Size, Deterministic Case 3.

A	a_1	0	0.2	0.8	0.2	0	0	0	0	0.2
	a_2	0	0	0	0.5	0	0	0	0	0
Θ	θ_1	0	0	0	0	-0.2	0.8	-0.5	-0.8	-0.5
	θ_2	0	0	0	0	0	0	0	0	0
	$R^2 = 0$	0.055	0.042	0.127	0.082	0.083	0.104	0.094	0.181	0.099
EJ	$R^2 = 0.3$	0.055	0.052	0.161	0.060	0.075	0.128	0.084	0.161	0.090
	$R^2 = 0.5$	0.052	0.059	0.209	0.067	0.071	0.153	0.085	0.159	0.087
	$R^2 = 0.7$	0.054	0.065	0.292	0.112	0.077	0.206	0.093	0.182	0.094
	$R^2 = 0$	0.046	0.038	0.071	0.127	0.057	0.066	0.052	0.096	0.063
$CADF$	$R^2 = 0.3$	0.051	0.057	0.077	0.083	0.055	0.071	0.051	0.080	0.059
	$R^2 = 0.5$	0.057	0.066	0.087	0.056	0.050	0.071	0.051	0.065	0.056
	$R^2 = 0.7$	0.054	0.070	0.091	0.042	0.049	0.070	0.052	0.051	0.052
	$R^2 = 0$	0.063	0.040	0.067	0.095	0.083	0.067	0.072	0.127	0.085
$CADF$-GLS	$R^2 = 0.3$	0.062	0.055	0.072	0.046	0.072	0.071	0.067	0.110	0.076
	$R^2 = 0.5$	0.063	0.062	0.072	0.035	0.064	0.071	0.063	0.098	0.069
	$R^2 = 0.7$	0.059	0.060	0.069	0.041	0.058	0.070	0.060	0.093	0.060
$Ave\ \ Lags$		0.4	1.4	2.6	2.8	1.6	5.6	3.7	3.7	2.9

Lags chosen by BIC with a maximum of 8 (could use MAIC for $CADF - GLS), T = 100,\ NMC = 10,000.$

Table 4: Small Sample Size, Deterministic Case 4.

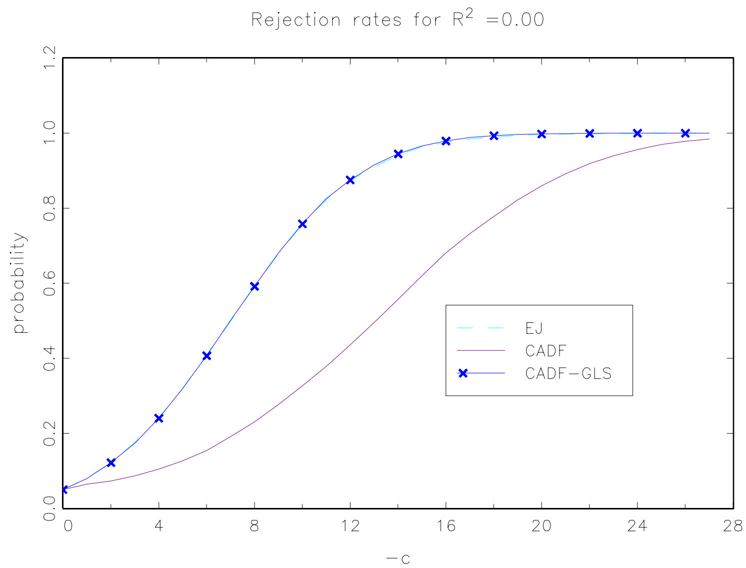
A	a_1	0	0.2	0.8	0.2	0	0	0	0	0.2
	a_2	0	0	0	0.5	0	0	0	0	0
Θ	θ_1	0	0	0	0	-0.2	0.8	-0.5	-0.8	-0.5
	θ_2	0	0	0	0	0	0	0	0	0
	$R^2 = 0$	0.040	0.018	0.061	0.057	0.069	0.066	0.079	0.196	0.085
EJ	$R^2 = 0.3$	0.045	0.037	0.178	0.044	0.068	0.143	0.086	0.187	0.089
	$R^2 = 0.5$	0.045	0.051	0.314	0.058	0.076	0.247	0.110	0.220	0.106
	$R^2 = 0.7$	0.050	0.092	0.516	0.158	0.111	0.452	0.177	0.338	0.165
	$R^2 = 0$	0.044	0.024	0.078	0.137	0.064	0.077	0.060	0.145	0.075
$CADF$	$R^2 = 0.3$	0.048	0.041	0.088	0.065	0.054	0.084	0.055	0.119	0.061
	$R^2 = 0.5$	0.049	0.053	0.098	0.040	0.054	0.094	0.056	0.101	0.061
	$R^2 = 0.7$	0.052	0.064	0.104	0.034	0.054	0.094	0.060	0.082	0.059
	$R^2 = 0$	0.058	0.023	0.074	0.106	0.081	0.068	0.063	0.135	0.087
$CADF$-GLS	$R^2 = 0.3$	0.054	0.041	0.077	0.042	0.064	0.065	0.055	0.100	0.069
	$R^2 = 0.5$	0.054	0.051	0.077	0.023	0.056	0.066	0.049	0.083	0.059
	$R^2 = 0.7$	0.051	0.060	0.076	0.024	0.049	0.063	0.048	0.069	0.051
Ave Lags		0.4	1.0	2.6	2.1	1.9	5.2	4.0	6.2	3.2

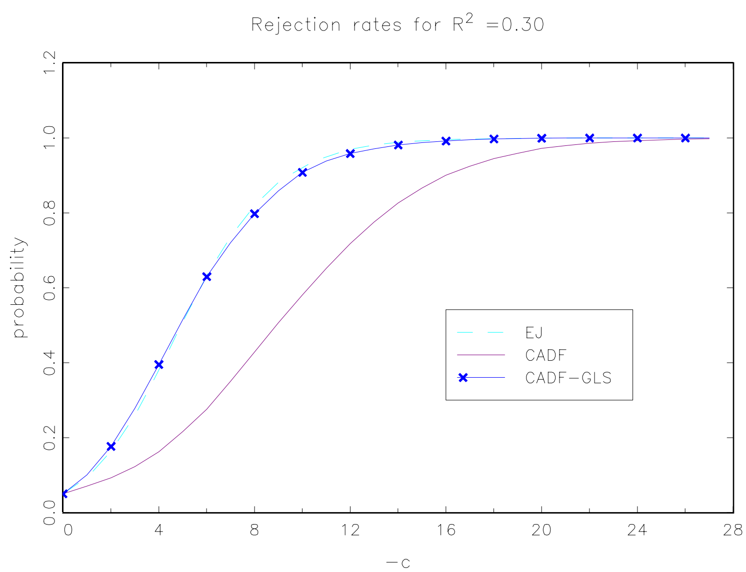
Lags chosen by BIC with a maximum of 8 (could use MAIC for $CADF - GLS$), $T = 100,\ NMC = 10,000$.

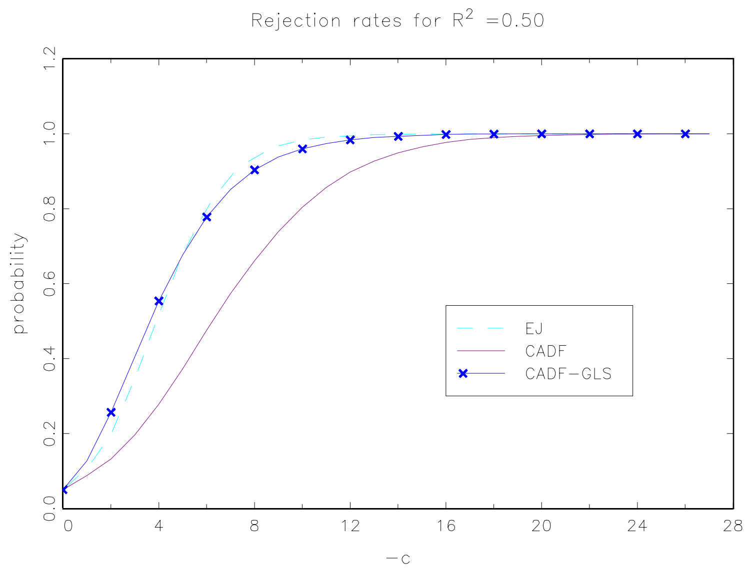
Table 5: Small Sample Size, Deterministic Case 5.

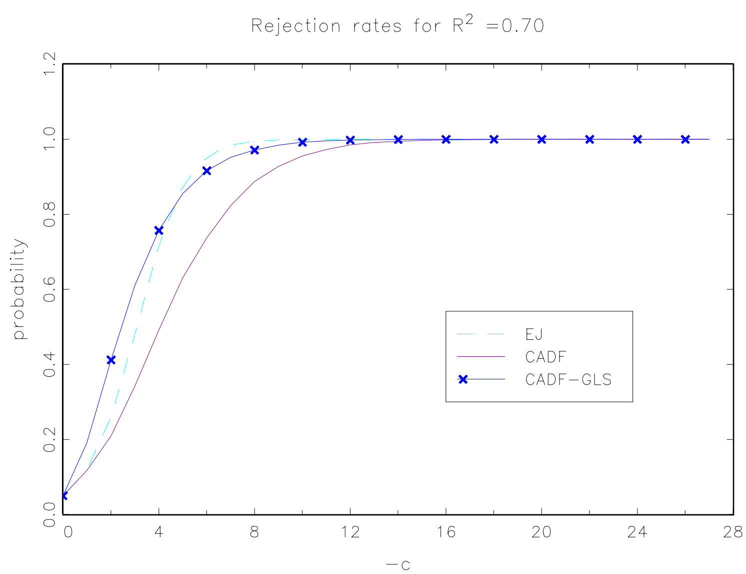
A	a_1	0	0.2	0.8	0.2	0	0	0	0	0.2
	a_2	0	0	0	0.5	0	0	0	0	0
Θ	θ_1	0	0	0	0	-0.2	0.8	-0.5	-0.8	-0.5
	θ_2	0	0	0	0	0	0	0	0	0
	$R^2 = 0$	0.039	0.018	0.066	0.057	0.067	0.067	0.077	0.197	0.084
EJ	$R^2 = 0.3$	0.043	0.035	0.192	0.044	0.066	0.152	0.086	0.197	0.090
	$R^2 = 0.5$	0.048	0.052	0.344	0.065	0.080	0.269	0.114	0.249	0.110
	$R^2 = 0.7$	0.059	0.102	0.566	0.187	0.126	0.501	0.210	0.414	0.191
	$R^2 = 0$	0.044	0.022	0.067	0.127	0.061	0.059	0.052	0.128	0.069
$CADF$	$R^2 = 0.3$	0.047	0.037	0.067	0.060	0.053	0.065	0.051	0.106	0.058
	$R^2 = 0.5$	0.047	0.048	0.074	0.035	0.052	0.076	0.051	0.096	0.055
	$R^2 = 0.7$	0.048	0.056	0.070	0.025	0.048	0.079	0.055	0.093	0.055
	$R^2 = 0$	0.044	0.024	0.070	0.099	0.080	0.065	0.062	0.135	0.086
$CADF$-GLS	$R^2 = 0.3$	0.047	0.035	0.071	0.039	0.059	0.061	0.051	0.107	0.066
	$R^2 = 0.5$	0.047	0.041	0.068	0.019	0.050	0.059	0.048	0.095	0.056
	$R^2 = 0.7$	0.048	0.046	0.062	0.020	0.042	0.059	0.047	0.091	0.049
Ave Lags		0.4	1.0	2.6	2.1	1.8	5.2	4.0	6.2	3.2

Figure 1: Asymptotic Power, Deterministic Case 2






Rejection rates for R^2 =0.50
probability
−c
EJ
CADF
CADF−GLS
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
4
8
12
16
20
24
28


Rejection rates for R^2 =0.70
probability
−c
EJ
CADF
CADF−GLS
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
4
8
12
16
20
24
28

7. References

Berk, K.N., 1974: "Consistent Autoregressive Spectral Estimates", *Annals of Statistics* 2, pp. 486-502.

Brillinger, D.R., 2001: "Time Series: Data Analysis and Theory".

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

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

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

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

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

Jansson, M. and M. Moreira, 2006: "Optimal Inference in Regression Models with Nearly Integrated Regressors", *Econometrica,* forthcoming..

Lutkephol, H. and P. Saikkonen, 1999: "Order Selection in Testing for the Cointegrating Rank of VAR Process", in: R.F. Engle and H. White, eds., Cointegration, Causality and Forecasting, (Oxford University Press, Oxford), pp. 168-199.

Ng, S, and P. Perron, 1995: "Unit Root Tests in ARMA Models with Data-Dependent Methods for Selection of the Truncation Lag", *Journal of the American Statistical Association*, 90, pp. 268-281.

Ng, S, and P. Perron, 2001: "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", *Econometrica*, 69:6, pp. 1519-1554.

Pesavento, E. and B.Rossi (2006): "Small Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons", *Journal of Applied Econometrics,* forthcoming.

Phillips, P.C.B., 1987a: "Toward and Unified Asymptotic Theory for Autoregression", *Biometrica*, 74, pp. 535-547.

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

Rossi B. 2005: "Confidence intervals for half-life deviations from Purchasing Power Parity", *Journal of Business & Economic Statistics,* 23:4, pp. 432-442.

Rossi B. 2006: "Expectations Hypotheses Tests at Long Horizons", *mimeo.*

Said, E.D. and D.A. Dickey, 1984: "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order", *Biometrika* 71, pp. 599-607.

Saikonnen, P, 1991: "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory* 7, pp. 1-21.

Sims, C.A., Stock J.H. and M.W. Watson, 1990: "Inference in Linear Time Series Models with some Unit Roots", *Econometrica* 58, 113-144.

Wooldridge, J., 1994, "Estimation and Inference for Dependent Processes", in D. Mc Fadden and R.F. Engle, eds., Handbook of Econometrics, vol.4, (North-Holland, Amsterdam), pp. 2639-2738.

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

8. Appendix

Notation used: $\|\bullet\|$ is the standard Euclidean norm, $\Rightarrow$ denotes weak convergence.

Lemma 1. *When the model is generated according to* $(1)-(3)$ *,with* $T(\rho-1)=c$*, then, as* $T\to\infty$:

$$(i)\ \omega_{y.x}^{-1/2}T^{-1/2}y_{[T\cdot]}^{d}\Longrightarrow J_{xyc}^{d}(\cdot)$$

$J_{xyc}(r)$ *is a Ornstein-Uhlenbeck process such that*

$$J_{xyc}(r)=W_{xy}(r)+c\int_0^1 e^{(\lambda-s)c}W_{xy}(s)\,ds$$

$W_{xy}(r)=\sqrt{\frac{R^2}{1-R^2}}W_x(r)+W_y(r)$*,* $W_x(r)$ *and* $W_y(r)$ *are independent standard Brownian Motions, and*

- *Under OLS detrending* $J_{xyc}^{d}(r)=J_{xyc}(r)$ *for Case 1,* $J_{xyc}^{d}(r)=J_{xyc}(r)-\int J_{xyc}(s)\,ds$ *for Case 2 and 3, and* $J_{xyc}^{d}(r)=J_{xyc}(r)-(4-6r)\int J_{xyc}(s)\,ds-(12r-6)\int sJ_{xyc}(s)\,ds$ *for Case 4 and 5.*
- *Under GLS detrending (i)* $J_{xyc}^{d}(r)=J_{xyc}(r)$ *for Case 1-3, and* $J_{xyc}^{d}(r)=J_{xyc}(r)-\left[\lambda J_{xyc}(1)+(1-\lambda)3\int sJ_{xyc}(s)\,ds\right]r$ *with* $\lambda=\frac{1-\bar{c}}{1-\bar{c}+\bar{c}^2/3}$ *.for Case 4 and 5.*

Proof. [Lemma 1] Assumption A1-A3 imply $T^{-1/2}\sum_{t=1}^{[T\cdot]}u_t(\rho)\Rightarrow\Omega^{1/2}W(\cdot)$ where $\Omega^{1/2}=\begin{bmatrix}\Omega_{xx}^{1/2} & 0\\ \omega_{yx}\Omega_{xx}^{-1/2} & \omega_{y.x}^{1/2}\end{bmatrix}$, $\omega_{y.x}^{1/2}=\omega_{yy}-\omega_{yx}\Omega_{xx}^{-1}\omega_{xy}$ and $W'=\begin{bmatrix}\widetilde{W}_x' & W_y\end{bmatrix}'$. Define $\overline{\delta}'=\omega_{y.x}^{-1/2}\omega_{yx}\Omega_{xx}^{-1/2}$ so that $\overline{\delta}'\overline{\delta}=\frac{R^2}{1-R^2}$; then from Phillips (1987 a,b) and the multivariate Functional Central Limit Theorem, $\omega_{y.x}^{-1/2}T^{-1/2}\sum_{t=1}^{[T\cdot]}u_{y[T\cdot]}(\rho)\Rightarrow\omega_{y.x}^{-1/2}\omega_{yx}\Omega_{xx}^{-1/2}\widetilde{W}_x'+W_y=\overline{\delta}'\widetilde{W}_x'+W_y=\sqrt{\frac{R^2}{1-R^2}}W_x+W_y$ where W_x is an univariate standard Brownian Motion independent of W_y. By the Continuous Mapping Theorem we have that $\omega_{y.x}^{-1/2}T^{-1/2}u_{y[T\cdot]}\Rightarrow J_{xyc}(r)$. The proof of the Lemma follows directly from simple calculations and the Continous Mapping Theorem. ■

The test suggested by Hansen (1995) is then based on the t-statistics on a augmented regression (8) in which lagged, contemporaneous and future values of the stationary covariate included. Recall equation (5)

$$\Delta y_t^d=\alpha y_{t-1}^d+\sum_{j=-\infty}^{+\infty}\widetilde{\pi}_{x,j}'x_{t-j}^d+\overline{\eta}_t \tag{14}$$

If we assume for example that $\psi(L)\eta_t = \tilde{\xi}_t$ where $\tilde{\xi}_t$ is white noise it easy to obtain the covariate augmented regression:

$$\Delta y_t^d = \varphi y_{t-1}^d + \sum_{j=-\infty}^{\infty} \pi'_{x,j} x_{t-j}^d + \sum_{j=1}^{\infty} \pi_{y,j} \Delta y_{t-j}^d + \xi_t \qquad (15)$$

where $\varphi = \psi(1)(\rho - 1)$ and

$$\begin{aligned} \xi_t &= \psi(1)\widetilde{\pi}_x(1)'\left(\hat{\mu}_{x-}\mu_x\right) + \psi(L)\sum_{j=-\infty}^{+\infty} \widetilde{\pi}'_{x,j}(\hat{\tau}_x - \tau_x)(t-j) + \\ &\qquad +\psi(1)(\rho-1)\left(\tilde{\mu}_y - \mu_y\right) - \psi(L)(1-\rho L)(\tilde{\tau}_y - \tau_y)t + \tilde{\xi}_t \end{aligned}$$

Given the absolute summability condition we can approximate the regression with a finite number of lags k :

$$\Delta y_t^d = \varphi y_{t-1}^d + \sum_{j=-k}^{k} \pi'_{x,j} x_{t-j}^d + \sum_{j=1}^{k} \pi_{y,j} \Delta y_{t-j}^d + \xi_{tk} \qquad (16)$$

where $\xi_{tk} = \xi_t + \sum_{|j|>k} \pi'_{x,j} x_{t-j}^d + \sum_{j>k} \pi'_{y,j} \Delta y_{t-j}^d$ and

To prove Theorem 1 let's first prove some auxiliary results. Following the same methodology of Sims, Stock and Watson (1990) rewrite (16) as

$$\Delta y_t^d = \Pi' w_{tk} + \xi_{tk}$$

where $\Pi' = \begin{bmatrix} \varphi & \pi' \end{bmatrix} = \begin{bmatrix} \varphi & \pi'_{x,-k} & \cdots & \pi'_{x,k} & \pi_{y,1} & \cdots & \pi_{y,k} \end{bmatrix}$, $w'_{tk} = \begin{bmatrix} y_{t-1}^d & X' \end{bmatrix}$, $X' = \begin{bmatrix} x_{t+k}^{d\prime} & \cdots & x_t^{d\prime} & \cdots & x_{t-k}^{d\prime} & \Delta y_{t-1}^d & \cdots & \Delta y_{t-k}^d \end{bmatrix}$. The proof follows closely Berk (1974), Said and Sickey (1984) and Saikonnen (1991). As Berk (1974) I use the standard Euclidean norm $\|z\| = (z'z)^{1/2}$ of a column vector z to define a matrix norm $\|B\|$ such that $\|B\| = \sup\{\|Bz\| : z < 1\}$. Notice that $\|B\|^2 \leq \sum_{ij} b_{ij}$ and that $\|B\|$ is dominated by the largest modulus of the eigenvalues of B.

Let Υ denote the diagonal matrix of dimensions $m(2k+1) + (k+1)$:

$$\Upsilon = diag\begin{bmatrix} (T-2k) & (T-2k)^{1/2} I_m & \cdots & (T-2k)^{1/2} I_m & (T-2k)^{1/2} & \cdots & (T-2k)^{1/2} \end{bmatrix}$$

and $\hat{R} = \Upsilon^{-1}\left(\sum_{t=k+1}^{T-k} w_{tk} w'_{tk}\right)\Upsilon^{-1}$. We are interested in the difference between $\hat{R}$ and $R = diag\begin{bmatrix} (T-2k)^{-2}\sum_{t=k+1}^{T-k} y_{t-1}^{d2} & \Gamma_X \end{bmatrix}$ with $\Gamma_X = E[XX']$.

Lemma 2. $\left\|\hat{R} - R\right\| = O_p\left(k^2/T\right)$

Proof. [Lemma 2] Denote $Q = [q_{ij}] = \hat{R} - R$. By definition $q_{11} = 0$. When $i > 1$ and $j > 1$ Dickey and Fuller (1984) show that $(T-2k)\, E\left(q_{ij}^2\right) \leq C$ for some C, where $0 < C < \infty$ and it is indipendent of i, j and T. Since Q has dimensions $m\,(2k+1) + 1 + k$, $E\left(\|Q\|^2\right) \leq \frac{m(2k+1)+1+k}{T-2k}$ so if $k^2/T \to 0$, $\|Q\|$ converges in probability ot zero. ■

Lemma 3. $\left\|R^{-1}\right\| = O_p\,(1)$

Proof. [Lemma 3] Since R^{-1} is black diagonal, $\left\|R^{-1}\right\|$ is bounded by the sum of the norms of the diagonal blocks. Under Lemma 1, and if $k/T \to 0$, $(T-2k)^{-2} \sum_{t=k+1}^{T-k} y_{t-1}^{d2} \Rightarrow \omega_{y.x} \int J_{xyc}^{d2}$ while the lower right corner of R^{-1} is Γ_X^{-1} which is bounded since all the elements of X are stationary. ■

Lemma 4. $\left\|\hat{R}^{-1} - R^{-1}\right\| = O_p\left(k/T^{1/2}\right)$

Proof. [Lemma 4] The proof follows directly from Dickey and Fuller (1984). ■

Denote $e_t = \sum_{|j|>k} \pi'_{x,j} x_{t-j}^d + \sum_{j>k} \pi'_{y,j} \Delta y_{t-j}^d$ so that $\varepsilon_{tk} = e_t + \varepsilon_t$. Note that $E\left\|e_t\right\|^2 \leq C\left(\sum_{|j|>k} \left\|\pi'_{x,j}\right\|^2 + \sum_{j>k} \left\|\pi'_{y,j}\right\|^2\right)$

Lemma 5. $\left\|\Upsilon^{-1} \sum_{t=k+1}^{T-k} w_{tk} e_t\right\| = O_p\left(k^{1/2}\right)$

Proof. [Lemma 5] $E\left\|\Upsilon^{-1} \sum_{t=k+1}^{T-k} w_{tk} e_t\right\|^2 = E\left\|(T-2k)^{-1} \sum_{t=k+1}^{T-k} y_{t-1}^d e_t\right\|^2 + E\left\|(T-2k)^{-1/2} \sum_{t=k+1}^{T-k} X e_t\right\|^2$.

$E\left\|(T-2k)^{-1} \sum_{t=k+1}^{T-k} y_{t-1}^d e_t\right\|^2 = E\left[(T-2k)^{-1} \sum_{t=k+1}^{T-k} u_{y,t-1} e_t\right]^2$. Under Lemma 1, if $k/T \to 0$, $(T-2k)^{-1} \sum_{t=k+1}^{T-k} u_{y,t-1} e_t = O_p\,(1)$ and $E\left\|(T-2k)^{-1} \sum_{t=k+1}^{T-k} y_{t-1}^d e_t\right\|$ is $O_p\,(1)$. Additionally, $E\left\|(T-2k)^{-1/2} \sum_{t=k+1}^{T-k} X e_t\right\|^2 \leq (T-2k)^{-1} \sum_{t=k+1}^{T-k} E\left\|X e_t\right\|^2 \leq (T-2k)^{-1} \sum_{t=k+1}^{T-k} E\left\|X\right\|^2 E\left\|e_t\right\|^2 \leq$

$Ctr\left(\Gamma_X\right)\left(\sum_{|j|>k} \left\|\pi'_{x,j}\right\|^2 + \sum_{j>k} \left\|\pi_{y,j}\right\|^2\right) \leq$

$\leq \left(C\left(2k+1\right)tr\left(\Gamma_x\right)+ktr\left(\Gamma_{\Delta y}\right)\right)\left(\sum_{|j|>k}\left\|\pi'_{x,j}\right\|^2+\sum_{j>k}\left\|\pi_{y,j}\right\|^2\right)$. Under Assumption 1, $\sum_{|j|>k}\left\|\pi'_{x,j}\right\|^2$ and $\sum_{j>k}\left\|\pi'_{y,j}\right\|^2$ are bounded, $E\left\|(T-2k)^{-1/2}\sum_{t=k+1}^{T-k}Xe_t\right\|^2$ is $O_p\left(k\right)$, $E\left\|\Upsilon^{-1}\sum_{t=k+1}^{T-k}w_{tk}e_t\right\|^2=0_p\left(k\right)$ and $\left\|\Upsilon^{-1}\sum_{t=k+1}^{T-k}w_{tk}e_t\right\|=O_p\left(k^{1/2}\right)$ ■

Lemma 6. $\left\|\Upsilon^{-1}\sum_{t=k+1}^{T-k}w_{tk}\xi_t\right\|=O_p\left(k^{1/2}\right)$

Proof. [Lemma 6] $E\left\|\Upsilon^{-1}\sum_{t=k+1}^{T-k}w_{tk}\xi_t\right\|^2=E\left\|(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\xi_t\right\|^2+E\left\|(T-2k)^{-1/2}\sum_{t=k+1}^{T-k}X\xi_t\right\|^2$.

$E\left\|(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\xi_t\right\|^2=E\left[(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\xi_t\right]^2$ which is $O_p\left(1\right)$ under Lemma 1 if $k/T\rightarrow 0$. Additionally, because all the elements of X are stationary and uncorrelated at all leads and lags with ξ_t, $E\left\|(T-2k)^{-1/2}\sum_{t=k+1}^{T-k}X\xi_t\right\|^2\leq (T-2k)^{-1}\sum_{t=k+1}^{T-k}E\left\|X\right\|^2E\left\|\xi_t\right\|^2=\left(C\left(2k+1\right)tr\left(\Gamma_x\right)+ktr\left(\Gamma_{\Delta y}\right)\right)\sigma_\xi^2=O_p\left(k\right)$ and $\left\|\Upsilon^{-1}\sum_{t=k+1}^{T-k}w_{tk}\xi_t\right\|=O_p\left(k^{1/2}\right)$ by Markov inequality. ■

Before stating and proving the next Lemma, recall, using the notation of model (1) the standard result for OLS estimated mean and trend $\hat{\mu}_x$ and $\hat{\tau}_x$:

$$\begin{bmatrix} T^{1/2}\left(\hat{\mu}_x-\mu_x\right) \\ T^{3/2}\left(\hat{\tau}_x-\tau_x\right)\end{bmatrix}\Rightarrow\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3\end{bmatrix}^{-1}\begin{bmatrix}\Omega_{xx}^{1/2}\tilde{W}_x\left(1\right) \\ \Omega_{xx}^{1/2}\left[\tilde{W}_x\left(1\right)-\int\tilde{W}_x\right]\end{bmatrix}$$

Additionally from Elliott, Rothemberg and Stock (1996), using the result of Lemma 1, we have that $T^{1/2}\left(\tilde{\mu}_y-\mu_y-y_1\right)\Rightarrow\omega_{y.x}^{1/2}\left[\bar{c}^2\int J_{xyc}-\bar{c}J_{xyc}\left(1\right)\right]$ and $T^{1/2}\left(\tilde{\tau}_y-\tau_y\right)\Rightarrow \omega_{y.x}^{1/2}\left[\lambda J_{xyc}\left(1\right)+\left(1-\lambda\right)3\int rJ_{xyc}\right]=\omega_{y.x}^{1/2}V_c$ where $\lambda=\frac{1-\bar{c}}{1-\bar{c}+\bar{c}^2/3}$.

Lemma 7 [OLS Detrending]**.** *When y_t is detrended using OLS, $(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\xi_t\Rightarrow \omega_{y.x}\psi\left(1\right)\int J_{xyc}^d dW_2$ with $J_{xyc}^d\left(r\right)=J_{xyc}\left(r\right)$ for Case 1, $J_{xyc}^d\left(r\right)=J_{xyc}\left(r\right)-\int J_{xyc}\left(s\right)ds$ for Case 2 and 3, and $J_{xyc}^d\left(r\right)=J_{xyc}\left(r\right)-\left(4-6r\right)\int J_{xyc}\left(s\right)ds-\left(12r-6\right)\int sJ_{xyc}\left(s\right)ds$ for Case 4 and 5*

Proof. [Lemma 7] Using the definition of ξ_t we have that

$$\begin{aligned}(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\xi_t &= (T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\tilde{\xi}_t \\ &+(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\psi\left(1\right)\widetilde{\pi}_x\left(1\right)'\left(\hat{\mu}_x-\mu_x\right) \\ &+\psi\left(L\right)(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\sum_{j=-\infty}^{+\infty}\widetilde{\pi}'_{x,j}\left(\hat{\tau}_x-\tau_x\right)\left(t-j\right) \\ &+\psi\left(1\right)\left(\rho-1\right)\left(\tilde{\mu}_y-\mu_y\right)(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d \\ &-(T-2k)^{-1}\sum_{t=k+1}^{T-k}y_{t-1}^d\psi\left(L\right)\left(1-\rho L\right)\left(\tilde{\tau}_y-\tau_y\right)t \quad (17)\end{aligned}$$

When y_t is detrended by OLS, $T^{1/2}(\hat{\mu}_x - \mu_x)$ and $T^{3/2}(\hat{\tau}_x - \tau_x)$ are $O_p(1)$ and under the local alternative $\rho - 1 = c/T$, the last two term in (17) converge in probability to zero. Under Assumption 1, by Chan and Wei (1988), Phillips (1987) and Lemma 1, it is easy to show that if $k/T \to 0$, $(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d \tilde{\xi}_t \Rightarrow \omega_{y.x}\psi(1)\int J_{xyc}^d dW_2$ as $\tilde{\xi}_t = \psi(L)\eta_t$ and 2π times the spectral density at frequency zero of η_t is $\omega_{y.x} = \omega_{yy} - \omega_{yx}\Omega_{xx}^{-1}\omega_{xy}$. For the other terms we can use the convergence results for the estimated deterministic terms of x_t, and as in Lemma 1 the fact that $\bar{\delta}'\widetilde{W}_x' = \sqrt{\frac{R^2}{1-R^2}}W_x$ to derive $(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d \psi(1)\widetilde{\pi}_x(1)'\left(\hat{\mu}_{x-}\mu_x\right) \Rightarrow \omega_{y.x}\psi(1)\int J_{xyc}^d \omega_{y.x}^{-1/2}\omega_{yx}\Omega_{xx}^{-1/2}\left\{4\tilde{W}_x(1) - 6\left[\tilde{W}_x(1) - \int \tilde{W}\right]\right\} =$

$= \omega_{y.x}\psi(1)\int J_{xyc}^d\sqrt{\frac{R^2}{1-R^2}}\left[-2W_x(1) + 6\int W_x\right]$ and

$\psi(L)(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d \sum_{j=-\infty}^{+\infty}\widetilde{\pi}_{x,j}'(\hat{\tau}_x - \tau_x)(t-j)$

$\Rightarrow \omega_{y.x}\psi(1)\int rJ_{xyc}^d\sqrt{\frac{R^2}{1-R^2}}\left[-6W_x(1) + 12\int W_x\right]$. Given the definitions of J_{xyc}^d for OLS detrended variables we have that $\int J_{xyc}^d$ and $\int rJ_{xyc}^d$ are both zero in all casese. ■

Now we have all the results necessary to prove Theorem 1.

Proof. [Theorem 1] $\Upsilon\left(\hat{\Pi} - \Pi\right) = \hat{R}^{-1}\Upsilon^{-1}\sum_{t=k+1}^{T-k} w_{tk}\xi_{tk} =$

$= \left(\hat{R}^{-1} - R^{-1}\right)\Upsilon^{-1}\sum_{t=k+1}^{T-k} w_{tk}\xi_{tk} + R^{-1}\Upsilon^{-1}\sum_{t=k+1}^{T-k} w_{tk}\xi_{tk} =$

$= \left(\hat{R}^{-1} - R^{-1}\right)\Upsilon^{-1}\sum_{t=k+1}^{T-k} w_{tk}\xi_t - \left(\hat{R}^{-1} - R^{-1}\right)\Upsilon^{-1}\sum_{t=k+1}^{T-k} w_{tk}e_t +$

$+ R^{-1}\Upsilon^{-1}\sum_{t=k+1}^{T-k} w_{tk}\xi_{tk} =$

$= E_1 + E_2 + E_3$. By Lemma 4, 5,6 and if $k^3/T \to 0$, both $\|E_1\|$ and $\|E_2\|$ are of order $o_p(1)$. Because R^{-1} is black diagonal

$(T-2k)(\hat{\varphi} - \varphi) = \left((T-2k)^{-2}\sum_{t=k+1}^{T-k} y_{t-1}^{d2}\right)^{-1}\left((T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d\xi_{tk}\right) + o_p(1)$ with $(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d\xi_{tk} = (T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d\xi_t + o_p(1)$. See also Ng and Perron (1995) p. 278 where, from Lemma 7, $(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d\xi_t \Rightarrow \omega_{y.x}\psi(1)\int J_{xyc}^d dW_2$. Recalling that $(T-2k)^{-2}\sum_{t=k+1}^{T-k} y_{t-1}^{d2} \Rightarrow \omega_{y.x}\int J_{xyc}^{d2}$, $(T-2k)(\hat{\varphi} - \varphi) \Rightarrow \psi(1)\left(\int J_{xyc}^{d2}\right)^{-1}\left(\int J_{xyc}^d dW_2\right)$. Define $\hat{s}_{\xi tk}^2 = (T-2k)^{-1}\sum_{t=k+1}^{T-k}\hat{\xi}_{tk}^2$, $\hat{s}_{\xi tk}$ converges in probability to the standard deviation of ξ_{tk} which is $\omega_{y.x}^{1/2}\psi(1)$ and $(T-2k)SE(\hat{\varphi}) \Rightarrow \psi(1)\left(\int J_{xyc}^{d2}\right)^{-1/2}$. Because $\hat{t}_{\hat{\varphi}} = \frac{(T-2k)\varphi}{(T-2k)SE(\hat{\varphi})} + \frac{(T-2k)(\hat{\varphi}-\varphi)}{(T-2k)SE(\hat{\varphi})} = \frac{(T-2k)(\rho-1)\psi(1)}{(T-2k)SE(\hat{\varphi})} + \frac{(T-2k)(\hat{\varphi}-\varphi)}{(T-2k)SE(\hat{\varphi})} = \frac{(T-2k)(c/T)\psi(1)}{(T-2k)SE(\hat{\varphi})} + \frac{(T-2k)(\hat{\varphi}-\varphi)}{(T-2k)SE(\hat{\varphi})}$,

then $t_{\hat{\varphi}} \Rightarrow \frac{c}{\left(\int J_{xyc}^{d2}\right)^{-1/2}} + \frac{\left(\int J_{xyc}^{d2}\right)^{-1}\left(\int J_{xyc}^d dW_2\right)}{\left(\int J_{xyc}^{d2}\right)^{-1/2}}$ as long as $\frac{k}{T} \to 0$. ■

When y_t is quasi-differenced detrended or GLS detrended the results are very similar with the exception that a few terms in (17) do not go to zero as can be seen in the following Lemma.

Lemma 8 [GLS Detrending]**.** *When y_t is detrended using GLS* , $(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d \xi_t \Rightarrow \omega_{y.x}\psi(1)\left[\int J_{xyc}^d dW_2 + \Lambda_c(r)\right]$ *where (i)* $\Lambda_c(r) = 0$ *for Case 1 and 2, (ii)* $\Lambda_c(r) = \left(\int J_{xyc}^d\right)\sqrt{\frac{R^2}{1-R^2}}\left[-2W_x(1) + 6\int W_x\right]$ *for Case 3, (iii)* $\Lambda_c(r) = \left(\int J_{xyc}^d\right)\sqrt{\frac{R^2}{1-R^2}}\left[-2W_x(1) + 6\int W_x\right] + V_c\left[c\int sJ_{xyc}^d - \int J_{xyc}^d\right]$ *for Case 4 and (iv)* $\Lambda_c(r) = \left(\int J_{xyc}^d\right)\sqrt{\frac{R^2}{1-R^2}}\left[-2W_x(1) + 6\int W_x\right] + \left(\int sJ_{xyc}^d\right)\sqrt{\frac{R^2}{1-R^2}}\left[-6W_x(1) + 12\int W_x\right] + V_c\left[c\int sJ_{xyc}^d - \int J_{xyc}^d\right]$ *for Case 5 with* $V_c = \lambda J_{xyc}(1) + (1-\lambda)3\int rJ_{xyc}$*with* $J_{xyc}^d = J_{xyc} - rV_c$

Proof. [Lemma 8] When y_t is GLS detrended, the last term in (17) does not converge to zero as now $\tilde{\tau}_y$ converges at rate $T^{1/2}$ and not $T^{3/2}$ so that

$$(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d \psi(L)(1-\rho L)(\tilde{\tau}_y - \tau_y)t =$$

$$= \tfrac{c}{T}\psi(L)(\tilde{\tau}_y - \tau_y)(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d t - \psi(L)(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d(\tilde{\tau}_y - \tau_y)$$

$$-\tfrac{c}{T}\psi(L)(T-2k)^{-1}\sum_{t=k+1}^{T-k} y_{t-1}^d(\tilde{\tau}_y - \tau_y) \Rightarrow c\psi(1)\omega_{y.x}V_c\int rJ_{xyc}^d - (1)\omega_{y.x}V_c\int J_{xyc}^d =$$

$\psi(1)\omega_{y.x}V_c\left[c\int rJ_{xyc}^d - \int J_{xyc}^d\right]$. Additionally, $\int J_{xyc}^d$ and $\int rJ_{xyc}^d$ are now not zero. . Plugging in the definition of J_{xyc}^d for each case gives the results. ■

Proof. [Theorem 2] Proof of Theorem 2 follows exactly the proof of Theorem 1 using the results of Lemma 8. ■