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

Elena Pesavento
Emory University

May 2005
Preliminary, do not cite without permission

Abstract. Numerous tests for integration and cointegration have been proposed in the literature. Since Elliott, Rothenberg 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. Similarly to Hansen (1995), Elliott and Jansson (2003), Zivot (2000), and Elliott, Jansson and Pesavento (2003) the proposed test achieves higher power by using additional information contained in covariates correlated with the variable we are testing. The test is constructed by applying Hansen’s test to variables that are detrended under the alternative in regression augmented with leads and lags of the stationary covariates. Using local to unity parametrization, this paper analytically compute the asymptotic distribution of the test under the null and the local alternative. I show 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 that are indistinguishable from the power curves of optimal tests.

Keywords: Unit Root Test, GLS detrending.

JEL Classification: C32.

I thank Michael Jansson as the idea of this paper came after a conversation with him.

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, they compute the power envelope for point-optimal tests of a unit root in an univariate model, and they derive a family of feasible tests (\(P_T\) thereafter) that have asymptotic power close to the power envelope. In fact the asymptotic power of the ERS test is never 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 \(ADF\) \(t\)-test where the variable have been GLS detrended before running the augmented Dickey Fuller 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. In fact, practitioners use the \(ADF - GLS\) more often that the \(P_T\) test, presumably because it is more intuitive to understand the origin of the increase in power while being simple to compute.

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 include constants and/or time trends. Both papers illustrate the significant increase in the asymptotic power envelope in multivariate models that include 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 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 \(P_T\) when there is no information in the stationary covariates, i.e. the

\(^1\)Exceptions are Rossi (2004), Rossi (2005), Pesavento Rossi (2004), and Jansson and Moreira (2005), where inference in robust to the presence of not of exact unit roots.
correlation between the stationary covariate and the variable we are testing is zero. Both the CADF and EJ tests have power than is higher that 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 outperforms univariate tests, but there are also significant differences between them. The differences are similar to the differences between ADF and \( P_T \) in univariate models. This is to be expected as ADF and \( P_T \) are special cases of CADF and EJ when no stationary covariates are included. 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 assumption on the deterministic terms and then estimating an augmented regression with lags and leads of the stationary covariates. To keep with the Hansen’s notation, I will call this test CADF − GLS. Similarly to the ADF − GLS test, the proposed test is intuitive and it is easy to compute. Section 2 analytically computes the asymptotic distribution of the test under the null and the local alternative. Using Monte Carlo simulations I show 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 that are indistinguishable from the power curves of optimal tests. Although the general model of Section 2 does not allow for cointegration\(^2\), 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 XX, study the behavior of CADF − GLS tests when used to test for the null of no cointegration.

Section XX 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 triangular model

\[
\begin{align*}
x_t &= \mu_x + \tau_x t + u_{x,t} \\
y_t &= \mu_y + \tau_y t + u_{y,t}
\end{align*}
\]  

\(^2\)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.
and
\[ \Phi(L) \begin{bmatrix} u_{x,t} \\ (1 - \rho L) u_{y,t} \end{bmatrix} = \varepsilon_t, \]  
(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.

**Assumption 1:**
\[ \max_{-k \leq t \leq 0} \mathbb{E} \left| u_{x,t}, u_{y,t} \right|^2 = O_p(1), \]
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 \( \sup_t E \| \varepsilon_t \|^{2+\delta} < \infty \) for some \( \delta > 0 \), where \( \Sigma \) is positive definite, \( E_t^{-1} (\cdot) \) refers to the expectation conditional on \( \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\} \).

I am interested in the problem of testing for the presence of a unit root in \( y_t \):

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

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)). This is the same problem analyzed by Elliott and Jansson (2003). Following their notation, define \( u_t(\rho) = (1 - \rho L) u_{y,t} = \Phi^{-1}(L) \varepsilon_t \). Assumption 1-3 mean that

\[ T^{-1/2} \mathbb{P} \left[ \sum_{t=1}^T u_t(\rho) \right] \Rightarrow \Omega^{1/2} W(\cdot) \]

where \( \Omega = \Phi^{-1}(1) \Sigma \Phi^{-1}(1) \) is 2\( \pi \) times the spectral density at frequency zero of \( u_t(\rho) \). Partition \( \Omega \) 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 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 \( y_t \) is determinable from the \( m \)-vector valued \( x_t \) by linear time invariant operations. \( R^2 \) lies between zero and one, and represents the contribution of the stationary variable as 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. The only exception is when there is only one cointegration vector that is known a-priori as in Elliott, Jansson and Pesavento (2005). This case is analyzed in section XX.

As in Elliott and Jansson (2003) 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.

Both Hansen (1995) and Elliott and Jansson (2003) show that in the triangular model (1) - (3), no uniformly most powerful test exists. When \( R^2 \) is different than 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 tests (EJ) that are feasible with data and that are close to the power envelope constructed from a point optimal family of tests.

Model (1 - 3) is slightly more general than Hansen (1995) and Elliott and Jansson (2003) as it allows for a short run dynamics of unknown and possibly infinite order. Define \( \Gamma (k) = E [u_t (\rho) u_{t+k} (\rho)] \) the autocovariance function of \( u_t (\rho) \). Implicit in A1-A3 is the summability condition \( \lim_{j \to \infty} \| \Gamma (k) \| < \infty \) and the condition that the spectral density of \( u_t (\rho) \), \( f_{u(\rho)u(\rho)} (\lambda) \) is bounded away from zero (check this). It is well known that when those two conditions hold (Saikonnen, 1991) we can write

\[
\begin{align*}
    u_{y,t} (\rho) &= \sum_{j=-\infty}^{\infty} \mathbf{e}_{x,j} u_{x,t-j} (\rho) + \eta_t
\end{align*}
\]  

(4)

where the summability condition \( \lim_{j \to \infty} \| \mathbf{e}_{x,j} \| < \infty \) holds and \( \eta_t \) is a serially correlated stationary process such that \( E u'_{x,t} \eta_{t+k} = 0 \) for any \( k = 0, \pm 1, \pm 2, \ldots \).
spectral density of $\eta_t$ is $f_{\eta t}(\lambda) = f_{u y}(\rho) u_y(\lambda) - f_{u y}(\rho) u_x(\lambda) f_{u x u y}(\lambda)^{-1} f_{u x u y}(\lambda)$ so

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

Substituting (4) into (1) we have that

$$\Delta y_t = d_t + (\rho - 1) y_{t-1} + \sum_{j=-\infty}^{\infty} \epsilon'_{x,j} x_{t-j} + \eta_t$$

where

$$d_t = (1 - \rho) \mu_y + (1 - \rho) \tau_y t - \epsilon'_x (1)' \mu_x - \epsilon'_x (1)' \tau_x t - \tau_x \sum_{j=-\infty}^{\infty} \epsilon'_{x,j}$$

with $\epsilon_x(1) = \sum_{j=-\infty}^{\infty} \epsilon'_{x,j}$. Since the sequence $\{\epsilon_{x,j}\}$ is absolute summable $\epsilon_{x,j} \approx 0$ for $|j| > k$ for $k$ large enough and we can approximate (5) with

$$\Delta y_t = d_t + \alpha y_{t-1} + \sum_{j=-k}^{k} \epsilon'_{x,j} x_{t-j} + \eta_{tk}$$

where $\alpha = (\rho - 1)$, and $\eta_{tk} = \eta_t + \sum_{|j| > k} \epsilon'_{x,j} x_{t-j}$. The deterministic term $d_t$ is such that

**Case 1**: $d_t = 0$

**Case 2**: $d_t = (1 - \rho) \mu_y$

**Case 3**: $d_t = (1 - \rho) \mu_y - \epsilon'_x (1)' \mu_x$

**Case 4**: $d_t = (1 - \rho) \mu_y + (1 - \rho) \tau_y t - \epsilon'_x (1)' \mu_x$

**Case 5**: $d_t = (1 - \rho) \mu_y + (1 - \rho) \tau_y t - \epsilon'_x (1)' \mu_x - \epsilon'_x (1)' \tau_x t - \tau_x \sum_{j=-\infty}^{\infty} \epsilon'_{x,j}$

Notice that $\eta_{tk}$ in (6) is uncorrelated at all leads and lags with $x_t$ but it is serially correlated and the asymptotic distribution of tests on $\alpha$ in (6) will depend on nuisance parameters. Modified version of the tests can still be constructed as in Phillips and Perron (1988) by using non parametric estimates of the nuisance parameters. Alternatively, the regression can be augmented with lags of $\Delta y_t$ to obtain errors that are white noise as in Hansen (1995). Hansen’s CADF test is then based on the $t$-statistics on an augmented regression in which lagged, contemporaneous and future values of the stationary covariate are included:
The intuition behind this approach is that the correlation between $y_t$ and $x_t$ can help in reducing the error variance thus resulting in more precisely regression parameters estimates. Hansen (1995) shows that the asymptotic distribution for the t-statistics on $\varphi$ is different that the distribution of the ADF test and that a significant increase in the asymptotic power for local alternative can be obtained with the inclusion of the covariate and shows, for the case with no mean, that the asymptotic power is close to the power envelope. Hansen (1995) considers cases equivalent to Case1-Case 4 in this paper, where the regression is estimated with no mean for Case 1, with a mean for Case 2 and Case 3, and with mean and trend for Case 4. As in Elliott and Jansson (2003) I also consider the general case in which there are no restrictions on the deterministic terms.

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

The condition in Assumption 4 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 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, Lutkephol and Saikonnen, 1999). Since I am only interested in the t-ratio statistics, Assumption 4 is necessary and sufficient to prove the asymptotic distribution of the tests.

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.

**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 $T \to \infty$:

$$
\hat{\varphi} \Rightarrow t_R \frac{c}{\sqrt{d_2}} + i_R \frac{j_{xy}}{j_{xyc}} \frac{c_{-1} i_R d_{xyc} dW_2}{c_{-1/2}}
$$
where \( J_{xy}(r) \) is a Ornstein-Uhlenbeck process such that
\[
J_{xy}(r) = W_{xy}(r) + c \int_0^1 e^{(\lambda-s)cW_{xy}(s)} ds,
\]
\[
W_{xy}(r) = q \frac{R^2}{1-R^2} W_x(r) + W_y(r),
\]
\( W_x(r) \) and \( W_y(r) \) are independent standard Brownian Motions, and \( J^d_{xy}(r) = J_{xy}(r) \) if no deterministic terms are included in the regression, \( J^d_{xy}(r) = J_{xy}(r) - \int sJ_{xy}(s) ds \) if a mean is included in the regression, and \( J^d_{xy}(r) = J_{xy}(r) - (4-6r) \int sJ_{xy}(s) ds - (12r-6) \int s^2J_{xy}(s) ds \) if 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 (7) are uncorrelated with \( x_{t-k} \), which holds in well-specified dynamic regressions.\(^3\)

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). Elliott Rothenberg and Stock (1996) first applied this methodology to construct point optimal tests in unit root testing. The additional complication in the context of model (1) – (3) is that the likelihood ratio test is computed using the entire system. Elliott and Jansson (2003)\(^4\) compute the power envelope for the family of point optimal tests for each possible case of the deterministic terms (Case 1 to Case 5) under some simplifying assumptions, and they construct feasible general tests that are asymptotically equivalent to the power envelope and that are valid under more general assumptions. In both Hansen (1995) and Elliott and Jansson (2003), the asymptotic power under a local alternative \( \rho = 1 + (c/T) \) depend on the parameter \( R^2 \); as \( R^2 \) increases, there is a larger gain in using the information contained in the stationary covariate over an univariate test. For this reason both the \( CADF \) and \( EJ \) tests have power that is equal to the power of \( ADF \) and \( P_t \) tests respectively when \( R^2 \) is zero. When \( R^2 \) is different than zero, the gain from using a multivariate test over an univariate test gets larger as \( R^2 \) increases.

Although both \( CADF \) and \( EJ \) have power that is larger than univariate tests, Elliott and Jansson (2003) show that 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

\(^3\)Note that \( R^2 \) in this paper corresponds to \( 1 - \rho^2 \) in Hansen(1995)’s notation.

\(^4\)Hansen (1995) also computes the power envelope for the less general case in which there are not deterministic terms present.
CADF and EJ are equivalent respectively to the asymptotic distributions of ADF and PT.

In the context of univariate tests, Elliott, Rothenberg and Stock (1996) show that, although PT has higher power that ADF, the ADF test has slightly 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 mean (ADF-GLS). Although this test does not have the same optimality justification of the PT test, simulations show that it has similar power properties while having slightly better size properties. As the ADF-GLS is easier to implement, while having similar properties, often practitioner prefer to use the ADF-GLS over PT tests.

This paper proposes a test that it 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.

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 $T \to \infty$:

\[
\hat{\theta} \Rightarrow \frac{c}{\sqrt{J_{xyc}^2}} + \frac{iR J_{xyc}^d \epsilon_{-1} iR J_{xyc} dW_2}{\sqrt{J_{xyc}^d}}
\]

where $J_{xyc}(r)$ and $W_{xy}(r)$ are as defined in Theorem 1, and $J_{xyc}^d(r) = J_{xyc}(r)$ if no deterministic terms are included in the regression, and if $R$ mean is included in the regression, while $J_{xyc}^d(r) = J_{xyc}(r) - \lambda J_{xyc}(1) + (1 - \lambda)3 \int s J_{xyc}(s) ds$ r where $\lambda = \frac{1-\epsilon}{1-\epsilon^2}$ if a mean and trend are included in the regression.

3. Size and Power Comparison

3.1. Power Comparison. All three tests studied in this paper have 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 a unit root can be sensitive to the choice of the estimation method. To study the small sample behavior of the proposed test, we simulate equation with scalar $x_t$:

\[
\begin{align*}
x_t &= u_{x,t}, \\
\Delta y_t &= (\rho - 1) y_{t-1} + u_{y,t}
\end{align*}
\]
To study the small sample power of the tests we simply allow the error process to be \( \text{i.i.d. } \mathcal{N}(0, \Sigma) \), so that \( \Omega = \Sigma \) where we choose \( \Omega = \frac{1}{R} \begin{bmatrix} R \\ 1 \end{bmatrix} \) for \( R^2 = 0, 0.3, 0.5, 0.7 \). To be able to make a meaningful comparison or the rejection rates under the alternative Table 5, reports the size adjusted power\(^5\).

**TABLES 5- ABOUT HERE**

For Case 1 no deterministic terms are included in the regression, for Case 2-Case3, the regressions are estimated with a mean, and for Case 4-5 the regressions include mean and trend. In each case \( R^2 \) is estimated as suggested by Elliott and Jansson (2003) and Hansen (1995)\(^6\). Table 5 presents the rejection rates over 10,000 replications of samples of length 100. As Table 5 shows the all three tests have power that is increasing with \( R^2 \). The relative ranking of the tests in finite samples is consistent with the asymptotic results. As documented in Elliott and Jansson (2003), \( EJ \) can have significant higher power than \( CADF \) while the power of \( CADF - GLS \) is higher that the power of \( CADF \) and it is very similar to the power of \( EJ \).

### 3.2. Size Comparison.

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

\[
A = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_1 \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_1 \end{bmatrix},
\]

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

\[
\Omega = (I_2 - A)^{-1} (I_2 + \Theta) \Sigma (I_2 + \Theta)' (I_2 - A)^{-1} = \begin{bmatrix} \mu & R \\ R & 1 \end{bmatrix}, \quad R \in [0, 1].
\]

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. The same lag length chosen by BIC is used for all three tests. For case 2-Case3, the regressions are estimated with a mean. For Case 4-5 the model is estimated with mean and trend. The sample size is \( T = 100 \) and 10,000 replications are used.

**TABLES 1-4 ABOUT HERE**

\(^5\)The small sample power is computed from the tests. It would also be interesting to simulated the asymptotic power directly from the asymptotic distribution and see if the two are different. This comparison will be available in a future version of the paper.

\(^6\)In the former test \( R^2 \) is estimated parametrically while in the latter it is estimated non-parametrically.
Tables XX and XX 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 $\Theta$ 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 $\Theta$ 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 the serial correlation present in the data.\footnote{Ng and Perron (2001) provide a method for better lag selection in the univariate case.}
### Table 1: Small Sample Size, Deterministic Case 2.

<table>
<thead>
<tr>
<th>( A )</th>
<th>( a_1 )</th>
<th>0</th>
<th>0.2</th>
<th>0.8</th>
<th>0.2</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \Theta )</td>
<td>( \theta_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.2</td>
<td>0.8</td>
<td>-0.5</td>
<td>-0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( R^2 )</th>
<th>0.059</th>
<th>0.059</th>
<th>0.076</th>
<th>0.121</th>
<th>0.102</th>
<th>0.165</th>
<th>0.346</th>
<th>0.158</th>
</tr>
</thead>
<tbody>
<tr>
<td>( EJ )</td>
<td>( R^2 = 0.3 )</td>
<td>0.065</td>
<td>0.067</td>
<td>0.098</td>
<td>0.104</td>
<td>0.099</td>
<td>0.128</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.5 )</td>
<td>0.060</td>
<td>0.071</td>
<td>0.119</td>
<td>0.097</td>
<td>0.094</td>
<td>0.164</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.7 )</td>
<td>0.066</td>
<td>0.083</td>
<td>0.153</td>
<td>0.102</td>
<td>0.099</td>
<td>0.260</td>
<td>0.159</td>
</tr>
<tr>
<td>( CADF )</td>
<td>( R^2 = 0 )</td>
<td>0.051</td>
<td>0.045</td>
<td>0.076</td>
<td>0.103</td>
<td>0.075</td>
<td>0.056</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.3 )</td>
<td>0.055</td>
<td>0.056</td>
<td>0.082</td>
<td>0.073</td>
<td>0.070</td>
<td>0.061</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.5 )</td>
<td>0.057</td>
<td>0.064</td>
<td>0.086</td>
<td>0.056</td>
<td>0.062</td>
<td>0.067</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.7 )</td>
<td>0.053</td>
<td>0.065</td>
<td>0.089</td>
<td>0.047</td>
<td>0.055</td>
<td>0.069</td>
<td>0.056</td>
</tr>
<tr>
<td>( CADF-GLS )</td>
<td>( R^2 = 0 )</td>
<td>0.072</td>
<td>0.058</td>
<td>0.089</td>
<td>0.086</td>
<td>0.111</td>
<td>0.070</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.3 )</td>
<td>0.078</td>
<td>0.072</td>
<td>0.093</td>
<td>0.075</td>
<td>0.099</td>
<td>0.076</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.5 )</td>
<td>0.072</td>
<td>0.071</td>
<td>0.090</td>
<td>0.075</td>
<td>0.078</td>
<td>0.073</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>( R^2 = 0.7 )</td>
<td>0.066</td>
<td>0.072</td>
<td>0.087</td>
<td>0.076</td>
<td>0.067</td>
<td>0.071</td>
<td>0.066</td>
</tr>
</tbody>
</table>

Lags by BIC with a maximum of 8 (should use MAIC for \( CADF-GLS \)), \( T = 100 \), \( NMC = 10,000 \).
Table 2: Small Sample Size, Deterministic Case 3.

<table>
<thead>
<tr>
<th></th>
<th>( a_1 )</th>
<th>0</th>
<th>0.2</th>
<th>0.2</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( a_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \Theta )</td>
<td>( \theta_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.2</td>
<td>0.8</td>
<td>-0.5</td>
<td>-0.8</td>
</tr>
<tr>
<td></td>
<td>( \theta_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
R^2 = 0 \\
R^2 = 0.3 \\
R^2 = 0.5 \\
R^2 = 0.7
\]

\[
EJ \\
R^2 = 0 \\
R^2 = 0.3 \\
R^2 = 0.5 \\
R^2 = 0.7
\]

\[
CADF \\
R^2 = 0 \\
R^2 = 0.3 \\
R^2 = 0.5 \\
R^2 = 0.7
\]

\[
CADF-GLS \\
R^2 = 0.3 \\
R^2 = 0.5 \\
R^2 = 0.7
\]
Table 3: Small Sample Size, Deterministic Case 4.

<table>
<thead>
<tr>
<th>A</th>
<th>$a_1$</th>
<th>0.2</th>
<th>0.8</th>
<th>0.2</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_2$</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>$\theta_1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.2</td>
<td>0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$EJ$</th>
<th>$R^2 = 0$</th>
<th>$R^2 = 0.3$</th>
<th>$R^2 = 0.5$</th>
<th>$R^2 = 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CADF$</td>
<td>$R^2 = 0$</td>
<td>$R^2 = 0.3$</td>
<td>$R^2 = 0.5$</td>
<td>$R^2 = 0.7$</td>
</tr>
<tr>
<td>$CADF-GLS$</td>
<td>$R^2 = 0$</td>
<td>$R^2 = 0.3$</td>
<td>$R^2 = 0.5$</td>
<td>$R^2 = 0.7$</td>
</tr>
</tbody>
</table>
Table 4: Small Sample Size, Deterministic Case 5.

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>0</th>
<th>0.2</th>
<th>0.8</th>
<th>0.2</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>$\theta_1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.2</td>
<td>0.8</td>
<td>-0.5</td>
<td>-0.8</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$R^2 = 0$

$EJ$

$R^2 = 0$

$R^2 = 0.3$

$R^2 = 0.5$

$R^2 = 0.7$

$CADF$

$R^2 = 0$

$R^2 = 0.3$

$R^2 = 0.5$

$R^2 = 0.7$

$CADF-GLS$

$R^2 = 0$

$R^2 = 0.3$

$R^2 = 0.5$

$R^2 = 0.7$
Table 5: Size adjusted, Small Sample Power.

<table>
<thead>
<tr>
<th></th>
<th>Case 2</th>
<th></th>
<th></th>
<th></th>
<th>Case 3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_{c}$-2</td>
<td>-5</td>
<td>-10</td>
<td>-15</td>
<td>$R^2_{c}$-2</td>
<td>-5</td>
<td>-10</td>
<td>-15</td>
</tr>
<tr>
<td>$EJ$</td>
<td>0</td>
<td>0.119</td>
<td>0.292</td>
<td>0.654</td>
<td>0.873</td>
<td>0.148</td>
<td>0.438</td>
<td>0.847</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.208</td>
<td>0.653</td>
<td>0.965</td>
<td>0.996</td>
<td>0.270</td>
<td>0.858</td>
<td>0.998</td>
</tr>
<tr>
<td>$CADF$</td>
<td>0</td>
<td>0.065</td>
<td>0.115</td>
<td>0.310</td>
<td>0.606</td>
<td>0.090</td>
<td>0.202</td>
<td>0.553</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.126</td>
<td>0.339</td>
<td>0.782</td>
<td>0.961</td>
<td>0.208</td>
<td>0.604</td>
<td>0.948</td>
</tr>
<tr>
<td>$CADF$-$GLS$</td>
<td>0</td>
<td>0.119</td>
<td>0.286</td>
<td>0.664</td>
<td>0.902</td>
<td>0.151</td>
<td>0.409</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.237</td>
<td>0.601</td>
<td>0.911</td>
<td>0.980</td>
<td>0.381</td>
<td>0.796</td>
<td>0.974</td>
</tr>
</tbody>
</table>

Table reports the size adjusted power for each test. The rejection rate for $c=0$ is 5% in all cases and it is not reported. $T=100$, NMC=10,000
4. References


5. Appendix

Notation used: $\|\cdot\|$ 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) $\omega^{-1/2} T^{-1/2} u_y[T_] \Rightarrow J^d_{xyc} (\cdot)$

$J^d_{xyc} (r)$ is a Ornstein-Uhlenbeck process such that

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

$W_{xy} (r) = \frac{q}{1 - R^2} W_x (r) + W_y (r)$, $W_x (r)$ and $W_y (r)$ are independent standard Brownian Motions, and $J^d_{xyc} (r) = J^d_{xy} R (r)$ if no deterministic terms are included in the regression, $J^d_{xyc} (r) = J^d_{xyc} (r) - \int_{\mathbb{R}} R (s) ds$ if a mean is included in the regression, and $J^d_{xyc} (r) = J^d_{xyc} (r) - (4 - 6r) J^d_{xyc} (s) ds - (12r - 6) s J^d_{xyc} (s) ds$ if a mean and trend are included in the regression.

Proof. (i) Assumption A1-A3 imply $T^{-1/2} \sum_{t=1}^{T_0} u_t (\rho) \Rightarrow \Omega^{1/2} W (\cdot)$ where $\Omega^{1/2} = \begin{bmatrix} \Omega_{xx}^{1/2} & \omega_{y,x} \Omega_{xx}^{1/2} \\ \omega_{y,x} \Omega_{xx}^{1/2} & \omega_{y,y} \end{bmatrix}$, $\omega_{y,x} = \omega_{yy} - \omega_{yx} \Omega_{xx}^{-1} \omega_{xy}$ and $W' = \begin{bmatrix} W_x' \\ W_y' \end{bmatrix}$. Using this notation $T^{-1/2} \sum_{t=1}^{T_0} u_t (\rho) \Rightarrow \omega_{y,x} \Omega_{xx}^{-1/2} W_x' + \omega_{y,y} W_y$. Define $\delta = \omega_{y,x} \omega_{yx} \Omega_{xx}^{-1}$ so that $\delta^2 = \frac{R^2}{1 - R^2}$; then from Phillips (1987 a,b) and the multivariate Functional Central Limit Theorem, $\omega_{y,x} \Omega_{xx}^{-1/2} u_y[T_0] \Rightarrow \delta^2 W_x' + W_y = \frac{R^2 - \Omega_{xx}}{1 - R^2} W_x + W_y$ where $W_x$ is an univariate standard Brownian Motion independent of $W_y$.

(ii) and (iii) Follows directly from the Chan and Wei (1988) and Phillips (1987 a).

Proof also in Pesavento (2004)

Recall equation (5) $\Delta y_t = d_t + (\rho - 1) y_{t-1} + \sum_{j=-\infty}^{+\infty} \theta_{x,j} x_{t-j} + \eta_t$. $\eta_t$ is uncorrelated at all lags and lags with $x_t$ but is serially correlated. Assume for example that $\psi (L) \eta_t = \xi_t$ where $\xi_t$ is white noise and The regression (5) can be augmented with lags of $\Delta y_t$ to obtain errors that are white noise 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 included:

$$\Delta y_t = \delta \xi_t + \varphi y_{t-1} + \sum_{j=-\infty}^{+\infty} \pi_{x,j} x_{t-j} + \sum_{j=1}^{+\infty} \pi_{y,j} \Delta y_{t-j} + \xi_t$$

where $\varphi = \psi (1) (\rho - 1)$. The deterministic is such that there is no mean in Case 1, a mean in Case 2 and 3 and a mean and trend in Cases 4 and 5. Given
the absolute summability condition we can approximate the regression with a finite number of lags $k$:

$$\Delta y_t = \mathbf{0} + \varphi y_{t-1} + \mathbf{p}_j \pi'_{x,j} x_{t-j} + \mathbf{p}_j \pi'_{y,j} \Delta y_{t-j} + \xi_{tk} \tag{9}$$

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

To prove Theorem 1 let's first prove some auxiliary results. For easiness of notation I will consider the case in which there are no deterministic terms. The extension to the other cases is straightforward, as they are applications of the same approach to demeaned or demeaned and detrended variables.

Following the same methodology of Sims, Stock and Watson (1990) rewrite (9) as

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

where $\Pi' = \mathbf{\varphi} \pi' = \mathbf{\varphi} \pi'_{x,-k} \cdots \pi'_{x,k} \pi'_{y,k} \cdots \pi'_{y,k}, w_{tk} = y_{t-1} X' \Pi', X' = x_{t+k} \cdots x_{t} \cdots x_{t-k} \Delta y_{t-1} \cdots \Delta y_{t-k}$. 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 $\|B\|^2 \leq \sum_{i,j} b_{ij}$ and that $\|B\|$ is dominated by the largest modulus of the eigenvalues of $B$.

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

$$\Upsilon = \text{diag} (T - 2k) (T - 2k)^{1/2} I_m \cdots (T - 2k)^{1/2} I_m (T - 2k)^{1/2} \cdots (T - 2k)^{1/2}$$

and $\hat{R} = \Upsilon^{-1} \mathbf{p}^T \mathbf{w}_{tk} \mathbf{w}_{tk}' \Upsilon^{-1}$. We are interested in the difference between $\hat{R}$ and $R = \text{diag} (T - 2k)^{-2} \mathbf{p}^T \mathbf{y}_{tk} \mathbf{y}_{tk}' \Upsilon^{-1}$ with $\Upsilon^{-1} \mathbf{G}_X = E [XX']$.

Lemma 2. $\hat{R} - R^o = O_p i k^2 / T^\xi$

Proof. Denote $Q = [q_{ij}] = \hat{R} - R$. By definition $q_{11} = 0$. When $i > 1$ and $j > 1$ Dickey and Fuller (1984) show that $(T - 2k) E q_{ij}^2 \leq C$ for some $C$, where $0 < C < \infty$ and it is independent of $i, j$, and $T$. Since $Q$ has dimensions $m (2k + 1) + 1 + k$,

$$E \|Q\|^2 \leq \frac{m (2k + 1) + 1 + k}{T - 2k}$$

so if $k^2 / T \rightarrow 0$, $\|Q\|$ converges in probability to zero. \hfill \blacksquare

Lemma 3. $R^{-1} = O_p (1)$
**Proof.** Since $R^{-1}$ is black diagonal, $R^{-1}$ is bounded by the sum of the norms of the diagonal blocks. Under Lemma 1, and if $k/T \to 0$, $(T - 2k)^{-1} \mathbb{P}_{t=k+1}^{T-k} y_{t-1}^2 \Rightarrow \omega_{y,x}$ while the lower right corner of $R^{-1}$ is $\Gamma_{X}^{-1}$ which is bounded since all the elements of $X$ are stationary. ■

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

**Proof.** The proof follows directly from Dickey and Fuller (1984). ■

Denote $e_t = \overset{\circ}{\mu}_p \sum_{j \leq k} \pi_{x,j} \hat{e}_{t-j} + \overset{\circ}{\pi}_y j \overset{\circ}{\pi}_{y,j} y_{t-j}$ so that $e_{tk} = e_t + e_t$. Note that $E\|e_t\|^2 \leq C$.

**Lemma 5.** $\overset{\circ}{\Upsilon}^{-1} \mathbb{P}_{t=k+1}^{T-k} w_{tk} e_t = O_p \left( k^{1/2} \right)$

**Proof.** $\overset{\circ}{\Upsilon}^{-1} \mathbb{P}_{t=k+1}^{T-k} w_{tk} e_t^2 = \overset{\circ}{\Upsilon}^{-1} (T - 2k)^{-1} \mathbb{P}_{t=k+1}^{T-k} y_{t-1}^2 e_t^2 + \overset{\circ}{\Upsilon}^{-1} (T - 2k)^{-1} \mathbb{P}_{t=k+1}^{T-k} X e_t^2$.

$E\|e_t\|^2 \leq C tr(\Gamma_X) \sum_{j \leq k} \pi_{x,j} + \overset{\circ}{\pi}_y j \overset{\circ}{\pi}_{y,j}$ under Assumption 1, $\pi_{x,j}$ and $\overset{\circ}{\pi}_y j \overset{\circ}{\pi}_{y,j}$ are bounded, $E\|e_t\|^2 \leq O_p \left( k^{1/2} \right)$.

**Lemma 6.** $\overset{\circ}{\Upsilon}^{-1} \mathbb{P}_{t=k+1}^{T-k} w_{tk} e_t = O_p \left( k^{1/2} \right)$

**Proof.** $\overset{\circ}{\Upsilon}^{-1} \mathbb{P}_{t=k+1}^{T-k} w_{tk} e_t^2 = \overset{\circ}{\Upsilon}^{-1} (T - 2k)^{-1} \mathbb{P}_{t=k+1}^{T-k} y_{t-1}^2 e_t^2 + \overset{\circ}{\Upsilon}^{-1} (T - 2k)^{-1} \mathbb{P}_{t=k+1}^{T-k} X e_t^2$.

Under Lemma 1 and given that $k/T \to 0$. Additionally, because all the elements of $X$ are...
stationary and uncorrelated at all leads and lags with \( \varepsilon_t, E \xi_0 (T - 2k)^{-1/2} \operatorname{P} \frac{T-k}{t=k+1} X \xi_0^2 \), and 
\[
(T - 2k)^{-1} \operatorname{P} \frac{T-k}{t=k+1} E \| X \|^2 E \| \xi_t \|^2 = (C (2k + 1) tr (\Gamma_x) + k tr (\Gamma_\Delta y)) \sigma^2_x = O_p (k). 
\]

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

**Proof.** [Theorem 1] \( \Sigma = \frac{R^{-1} \Sigma^{-1}}{\Sigma} \frac{T-k}{t=k+1} w_{tk} \xi_{tk} = \frac{T-k}{t=k+1} \frac{y_{tk}^2}{\Sigma} + \frac{R^{-1} \Sigma^{-1}}{\Sigma} \frac{T-k}{t=k+1} w_{tk} \xi_{tk} = \frac{R^{-1} \Sigma^{-1}}{\Sigma} \frac{T-k}{t=k+1} w_{tk} \xi_{tk} = \frac{R^{-1} \Sigma^{-1}}{\Sigma} \frac{T-k}{t=k+1} \xi_{tk} + O_p (1). \) Because \( R^{-1} \) is black diagonal, 
\[
(T - 2k) (\hat{\varphi} - \varphi) = (T - 2k) \frac{1}{\Sigma} \frac{T-k}{t=k+1} y_{tk} \frac{(T - 2k)^{-1}}{\Sigma} \frac{T-k}{t=k+1} y_{tk} - \frac{1}{\Sigma} \frac{T-k}{t=k+1} y_{tk} \xi_{tk} + O_p (1).
\] See also Ng and Perron (1995) p. 278. Under Assumption 1, by Chan and Wei (1988), Phillips (1987) and Lemma 1, it is easy to show that if \( k/T \to 0 \), \( (T - 2k)^{-1} \frac{T-k}{t=k+1} y_{tk} \xi_{tk} \to \omega_{y,x} \psi (1) \) and \( \omega_{y,x} \frac{J_{yyc}^2 dW_2}{\xi_{tk}} \). Define \( s_{tk}^2 \) as \( \frac{R^{-1} \Sigma^{-1}}{\Sigma} \frac{T-k}{t=k+1} \xi_{tk}^2 \), 
\[
\hat{s}_{tk} \text{ converges in probability to the variance of } \xi_{tk} \text{ which is } \omega_{y,x}^{1/2} \psi (1) \text{ and } (T - 2k) SE (\hat{\varphi}) \to \psi (1) \frac{J_{yyc}^2 dW_2}{\xi_{tk}^2}.
\]

If the deterministic terms are not zero, the same proof can be applied to demeaned variables for Cases 2 and 3 and to demeaned and detrended variables for Cases 4 and 5.

**Proof.** [Theorem 2] The proof of Theorem 2 follows directly from the proof of Theorem 1, with different Brownian Motions deriving from the different detrending procedure.