DÉPARTEMENT DE SCIENCE ÉCONOMIQUE
DEPARTMENT OF ECONOMICS

CAHIERS DE RECHERCHE / WORKING PAPERS

 # 0004E

Residual Based Tests for Cointegration with GLS Detrended Data

by
Pierre Perron and Gabriel Rodriguez

ISSN: 0225-3860

uOttawa

Faculté des sciences sociales
Faculty of Social Sciences

CP 450 SUCC. A
OTTAWA (ONTARIO)
CANADA K1N 6N5

P.O. BOX 450 STN. A
OTTAWA, ONTARIO
CANADA K1N 6N5

Residual Based Tests for Cointegration with GLS Detrended Data*

Pierre Perron
Boston University

Gabriel Rodríguez
University of Ottawa

First Draft: October 8, 1999; This version: June 17, 2002

Abstract

We propose residual based tests for cointegration using local GLS detrending (Elliott, Rothemberg and Stock, 1996, ERS) applied separately to each variable of the system. We consider two cases, one where only a constant is included and one where a constant and a time trend are included. We derive the asymptotic distribution of a feasible point optimal test which allows us to derive the power envelope. The quasi-differencing parameter $\overline{c}$ is selected such as the asymptotic power of the feasible point optimal test is 50.0%. Different $\overline{c}$ exists because the limiting distribution of the feasible point optimal test depends of the number of right-hand regressors and the specification of the deterministic components. The limiting distributions of various residuals based tests are derived for a general quasi-differencing parameter $\overline{c}$ and critical values are tabulated for values of $\overline{c} = 0$ irrespective of the nature of the deterministic components and the values suggested by the power envelope. Simulations show that using GLS detrending allows tests with higher power and that using $\overline{c}$ selected by the power envelope, as the quasi-differencing parameter, according to the two cases analyzed, is preferable. An explanation for this feature is provided.

Keywords: Unit root, M-tests, ADF test, quasi-differencing, hypothesis testing.

JEL Classification: C2, C3, C5

*This paper is drawn from chapter 2 of Gabriel Rodríguez's PhD Dissertation at the Université de Montréal (Rodríguez, 1999). Perron acknowledges financial support form the Social Sciences and Humanities Research Council of Canada (SSHRC), the Natural Sciences and Engineering Council of Canada (NSERC), the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR) and the National Science Foundation under Grant SES-0078492. Address for correspondence: Pierre Perron, Department of Economics, Boston University, 270 Bay State Road, Boston, 02215, USA (e-mail: perron@bu.edu).

1 Introduction

Even though they are applicable only under some specific conditions, residuals based tests for cointegration, developed by Phillips and Ouliaris (1990), have been quite popular in applied work mostly because of their computational simplicity. The statistics introduced are designed to test the null hypothesis of no cointegration in a single equation setting assuming that the variables introduced as regressors are not cointegrated. These tests also have some appeal because they follow quite intuitively from the basic definition of cointegration as laid out in Engle and Granger (1987). If the system of variables is cointegrated, then there exists a linear combination (given by the cointegrating vector) that is stationary. In this case, the residuals from a simple static regression are stationary and, as shown by Stock (1987), this regression estimated by OLS will provide a consistent estimate of the cointegrating vector. In the absence of cointegration, the residuals from the static regression are nonstationary for any choice of the parameter vector and we have what has been labelled, following Granger and Newbold (1974) and later Phillips (1986), a spurious regression. Hence, an obvious testing strategy is to test the null hypothesis of no cointegration using some unit root test on the estimated residuals from the simple static regression.

Of course, there are many alternative approaches available, some applicable under less restrictive conditions; for example, the system based tests of Johansen (1991) and Stock and Watson (1988). The reader is referred to one of the many available surveys; for example, Watson (1994), Perron and Campbell (1992), Banerjee, Dolado, Galbraith and Hendry (1993) and Lütkepohl (1999).

In an important paper, Elliott, Rothenberg and Stock (1996, hereafter ERS), following the work of Dufour and King (1991), show that several unit root tests constructed using GLS or quasi-differenced data [1] have asymptotic power functions close to the Gaussian local asymptotic power envelope. Hence, they enjoy some optimal properties over tests constructed using OLS detrended data and the simulations in ERS showed substantial power gains in finite samples. If such a detrending device is beneficial for unit root tests, it is natural to think that it would also be for cointegration tests.

Our aim, accordingly, is to analyze residual based tests for cointegration when they are constructed using GLS detrended or quasi-differenced data. We consider the standard ADF and the Z_α and Z_t tests analyzed in Phillips and Ouliaris (1990) as well as the class of modified unit root tests analyzed in Stock (1999), Perron and Ng (1996) and Ng and Perron (2000). We derive their asymptotic distribution assuming a general quasi-differencing parameter $\bar{c}$ and tabulate critical values for two choices: a) $\bar{c} = 0$ irrespective of the nature of the deterministic components; b) the quasi-differencing parameter selected by

[1] There is some argument to the fact that the use of the terminology "GLS detrending" is not appropriate given that the procedure does not consider a full GLS transformation (since it only considers the leading root modelled as local to unity). An alternative terminology is that of "quasi-differenced" data. We still believe, for reasons that will become clear later in the text, that the use of "GLS detrending" is meaningful since it is this feature of the procedure that is of importance, even if constructed in a partial fashion. We shall use both terminology interchangeably.

the power envelope. Our simulation results about power reveal that important power gains can indeed be achieved by using GLS detrended data, especially if the quasi-difference parameter is set as suggested by the power envelope. An explanation for this result is presented. Our work is related to that of Lütkepohl and Saikkonen (1999), Saikkonen and Lütkepohl (1997, 1998) and Xiao and Phillips (1999) who considered the use of GLS or quasi-differenced data when testing for cointegration in a multivariate setting, i.e. extending the tests proposed by Johansen (1991).

This paper is organized as follows. Section 2 presents the data-generating process considered and offers preliminary results concerning the limits of the estimates of the coefficients of the trend function and the estimates from the static regression with GLS detrended data. Section 3 presents the tests considered and derives their asymptotic distribution. Tabulated asymptotic and finite sample critical values are presented in Section 4 where we further discuss the role of the quasi-difference transformation. Section 5 assesses the size and power of the tests in a simple bivariate setting. In this section we also present results from size and power simulations for the Error Correction Tests (hereafter ECM) to compare with the performance of our tests. Section 6 offers brief concluding remarks and an appendix contains technical derivations.

2 The Model and Preliminary Results

We consider the following Data Generating Process:

$$\begin{aligned} z_t &= d_t + u_t \\ u_t &= Au_{t-1} + v_t \end{aligned} \quad (1)$$

where $z_t = (y_t, x_t')'$ and $u_t = (u_{1t}, u_{2t}')'$ are n-vectors, x_t and u_{2t} are m-vectors $(n = m + 1)$, y_t is a scalar, A is a diagonal matrix with typical element given by $\alpha = 1 + c/T$. Then u_t is a near-integrated process with c as a non-centrality parameter, and $\{v_t\}$ is an unobserved stationary mean-zero process with finite variance and spectral density matrix $f_{vv}(\lambda)$. We use the assumption that $u_0 = 0$ throughout, though the results hold for the weaker requirement that $E(u_0^2) < \infty$. The noise function is $v_t = \Phi(L)\tau_t = \sum_{i=0}^{\infty} \Phi_i \tau_{t-i}$ with $\sum_{i=0}^{\infty} i \det |\Phi_i| < \infty$ $(\Phi_0 = I_n)$ and $\tau_t \sim i.i.d.(0, \sigma_\tau^2)$ [2]. Under these conditions, the partial sum process constructed from $\{v_t\}$ satisfies a multivariate invariance principle. More specifically, for $r \in [0, 1]$, and as $T \to \infty$, we have (where here, and throughout this paper, $\Rightarrow$ denotes weak convergence in distribution),

$$X_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} v_t \Rightarrow B(r),$$

[2] This assumption is made for simplicity of exposition. All results holds under more general conditions, in particular allowing τ_t to be a martingale difference sequence.

where $B(r)$ is a n-vector Brownian motion with covariance matrix given by

$$\begin{aligned} &= \lim_{T\to\infty} T^{-1}E\{[\sum_{t=1}^{T} v_t][\sum_{t=1}^{T} v_t']\} \\ &= 2\pi f_{vv}(0) \\ &= 2\pi\sigma_\tau^2\Phi(1)\Phi(1)'. \end{aligned}$$

We define the following partition of   and $B(r)$:

$$\begin{aligned} &= \begin{bmatrix} w_{11} & w_{21}' \\ w_{21} & {}_{22} \end{bmatrix}, \\ B(r) &= \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix}, \end{aligned}$$

where w_{11} and $B_1(r)$ are scalars, w_{21} and $B_2(r)$ are vectors of dimension m and ${}_{22}$ is a positive definite matrix of dimension m by m. We also define the block triangular decomposition $= L'L$, with

$$L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix},$$

where

$$\begin{aligned} l_{11} &= (w_{11} - w_{21}' \; {}_{22}^{-1} w_{21})^{1/2}, \\ l_{21} &= {}_{22}^{-1/2} w_{21}, \\ L_{22} &= {}_{22}^{-1/2}. \end{aligned}$$

Note that $B(r) = L'W(r)$, where $W(r)$ is an n-dimensional Wiener process with identity covariance matrix. Given that ${}_{22}$ is a positive definite matrix, the null hypothesis of no cointegration implies that $l_{11} \neq 0$ and, accordingly that   is of full rank. This assumption is maintained throughout.

In (1), the deterministic component d_t is specified as $d_t = \psi' m_t$, where m_t is a set of k deterministic terms and ψ a k by n matrix of coefficients. In this paper, we only consider d_t to be a polynomial in time, i.e. $d_t = \sum_{i=0}^{p} \psi_i t^i$, with special focus on $p = 0, 1$.

2.1 Estimates of the Trend Function under GLS Detrending

We start with the case where the set of deterministic components present in the data generating process is the same as the set of deterministic terms used to detrend the data. We discuss in Section 3.3, the case where the data are trending ($p = 1$) and only a constant is included in the regression. To apply the GLS detrending procedure suggested by ERS, we start by defining the transformed data $z_t^{\bar\alpha}$ and $m_t^{\bar\alpha}$ as:

$$\begin{aligned} z_t^{\bar\alpha} &= (1 - \bar\alpha L) z_t, \\ m_t^{\bar\alpha} &= (1 - \bar\alpha L) m_t, \end{aligned}$$

for $t = 2, ..., T$, and $z_1^{\bar{\alpha}} = z_1$, $m_1^{\bar{\alpha}} = m_1$. Let $Z^{\bar{\alpha}} = [z_1^{\bar{\alpha}}, ..., z_T^{\bar{\alpha}}]'$, $M^{\bar{\alpha}} = [m_1^{\bar{\alpha}}, ..., m_T^{\bar{\alpha}}]'$, then

$$\hat{\psi} = (M^{\bar{\alpha}\prime} M^{\bar{\alpha}})^{-1} M^{\bar{\alpha}\prime} Z^{\bar{\alpha}}. \tag{2}$$

This implies that each series is detrended separately by an OLS regression on quasi-transformed data with differencing parameter $\bar{\alpha}$. As a preliminary result, we consider the limiting distribution of the estimates $\hat{\psi}$ obtained from local to unity GLS detrending. The following lemma gives the asymptotic properties of the estimates of the coefficients of the trend function in this case.

Lemma 1 *Suppose that z_t is generated by (1) with $d_t = \psi' m_t$ and that each variable in the $n-$vector z_t is detrended separately. Let $\hat{\psi}$ be the GLS estimates, defined by (2), of the coefficients of the trend function obtained using $\bar{\alpha} = 1 + \bar{c}/T$.*

a. *If $p = 0$, with $m_t = 1$ for all t, then: $\Upsilon_T(\hat{\psi} - \psi)' \Rightarrow \mathbf{0}_m$, where $\Upsilon_T = diag(T^{-1/2}, ..., T^{-1/2})$, a n by n matrix and $\mathbf{0}_m$ denotes an m by 1 vector of zeros.*

b. *If $p = 1$, with $m_t' = (1, t)$ for all t, then:*

$$\Upsilon_T vec[\hat{\psi} - \psi] \Rightarrow \begin{bmatrix} \mathbf{0}_m \\ \Lambda(\bar{c}) B_c(1) + 3(I_n - \Lambda(\bar{c})) \int_0^1 s B_c(r) ds \end{bmatrix} \equiv \begin{bmatrix} \mathbf{0}_m \\ D(\bar{c}) \end{bmatrix}$$

where $\Upsilon_T = [diag(T^{-1/2}, ..., T^{-1/2}), diag(T^{1/2}, ..., T^{1/2})]$, a $2n$ by $2n$ matrix, $\Lambda(\bar{c})$ is a n by n diagonal matrix with typical elements given by $\lambda = (1 - \bar{c})/(1 - \bar{c} + \bar{c}^2/3)$, $B_c(r)$ is defined by the stochastic differential equation $dB_c(r) = CB_c(r)dr + dB(r)$ with $B_c(0) = 0$, and the vec operator stacks the rows of a matrix into a column vector.

2.2 Limit Distributions for the Static Regression

We now consider residuals based tests for cointegration in the spirit of Phillips and Ouliaris (1990) but using GLS detrended variables defined by $(y_t^a, x_t^{a\prime})' = (y_t, x_t')' - \hat{\psi}' m_t$ where $\hat{\psi}$ is defined by (2). The relevant regression estimated by OLS is

$$y_t^a = \gamma' x_t^a + e_t. \tag{3}$$

The following Theorem gives the limiting behavior of the estimate $\hat{\gamma}$ under the null hypothesis of no cointegration.

Theorem 1 *Suppose that z_t is generated by (1) with $d_t = \psi' m_t$. Let y_t^a and x_t^a be GLS detrended variables with non-centrality parameter $\bar{\alpha} = 1 + \bar{c}/T$. Let $\hat{g}' = (1, -\hat{\gamma}')$ be the OLS estimates, from (3), of the cointegrating vector. Let $B_c(r)$ be as defined in Lemma 1 with the partition $B_c(r) = [B_{1c}(r), B_{2c}(r)]'$ where $B_{1c}(r)$ is a scalar and $B_{2c}(r)$ a vector of dimension m.*

1. *If $p = 0$, with $m_t = 1$ for all t, then:*

$$\hat{g}' \equiv (1, -\widehat{\gamma}') \Rightarrow \eta' \equiv (1, -g_{21}' G_{22}^{-1})$$

where $g_{21} = \int_0^1 B_{2c}(r) B_{1c}(r) dr$ *and* $G_{22} = \int_0^1 B_{2c}(r) B_{2c}'(r) dr$.

2. *If $p = 1$, with $m_t' = (1, t)$ for all t, then:*

$$\hat{g}' \equiv (1, -\widehat{\gamma}') \Rightarrow \overline{\eta}' \equiv (1, -\overline{g}_{21}' \overline{G}_{22}^{-1})$$

where $\overline{g}_{21} = \int_0^1 \overline{B}_{2c}(r) \overline{B}_{1c}(r) dr$, $\overline{G}_{22} = \int_0^1 \overline{B}_{2c}(r) \overline{B}_{2c}'(r) dr$ *and* $\overline{B}_c(r) = B_c(r) - rD(\bar{c})'$.

3 The Tests and their Asymptotic Distributions

3.1 The Tests

All tests considered are based on $\hat{e}_t$, the residual obtained from the static cointegration regression (3), i.e.,

$$\hat{e}_t = y_t^a - \hat{\gamma}' x_t^a. \tag{4}$$

The class of Z tests, analyzed by Phillips (1987) and Phillips and Perron (1988) in the context of testing for a unit root, can be applied to test the null hypothesis of no-cointegration as showed by Phillips and Ouliaris (1990). In the present context, with GLS detrended data, these are defined by

$$Z_\alpha^{GLS} = T(\hat{\alpha} - 1) - (s^2 - s_u^2)/(2T^{-2} \sum_{t=1}^{T} \hat{e}_{t-1}^2), \tag{5}$$

$$Z_t^{GLS} = \frac{s_u}{s} t_{\hat{\alpha}} - (s^2 - s_u^2)/(4s^2 T^{-2} \sum_{t=1}^{T} \hat{e}_{t-1}^2)^{1/2}, \tag{6}$$

where $\hat{\alpha}$ is the OLS estimate in the following regression:

$$\hat{e}_t = \hat{\alpha} \hat{e}_{t-1} + \hat{\omega}_t,$$

$t_{\hat{\alpha}}$ is the corresponding t-statistic for testing $\alpha = 1$, $s_u^2 = T^{-1} \sum_{t=1}^{T} \hat{\omega}_t^2$ and s^2 is described below.

The M-tests, originally proposed by Stock (1999), and further analyzed by Perron and Ng (1996) and Ng and Perron (2000), exploit the feature that a series converges with different rates of normalization under the null and the alternative hypotheses. These were shown to have far less size distortions than the Z tests in the presence of important negative serial correlation in the first-differences of the data. Constructed using the residuals from the cointegrating regression, they are defined by:

$$MZ_\alpha^{GLS} = \left(T^{-1}\hat{e}_T^2 - s^2\right)\left(2T^{-2}\sum_{t=1}^{T}\hat{e}_t^2\right)^{-1}, \tag{7}$$

$$MSB^{GLS} = \left(T^{-2}\sum_{t=1}^{T}\hat{e}_t^2/s^2\right)^{1/2}, \tag{8}$$

$$MZ_t^{GLS} = \left(T^{-1}\hat{e}_T^2 - s^2\right)\left(4s^2T^{-2}\sum_{t=1}^{T}\hat{e}_t^2\right)^{-1/2}. \tag{9}$$

The term s^2 is an autoregressive estimate of (2π times) the spectral density at frequency zero of v_t, defined as:

$$s^2 = s_{\eta k}^2 / \left(1 - \hat{b}(1)\right)^2, \tag{10}$$

where $s_{\eta k}^2 = T^{-1}\sum_{t=k+1}^{T}\hat{\eta}_{tk}^2$, $\hat{b}(1) = \sum_{j=1}^{k}\hat{b}_j,$ with $\hat{b}_j$ and $\{\hat{\eta}_{tk}\}$ obtained from the autoregression [3]:

$$\Delta\hat{e}_t = b_0\hat{e}_{t-1} + \sum_{j=1}^{k} b_j\Delta\hat{e}_{t-j} + \eta_{tk}. \tag{11}$$

The first statistic is a modified version of the Z_α test, the second statistic is a modified version of Bhargava's (1986) R_1 statistic which builds upon the work of Sargan and Bhargava (1983), and the third statistic is a modified version of the Z_t test. As Perron and Ng (1996) showed, the MSB and Z_α tests are approximately related by $Z_t \approx MSB \cdot Z_\alpha$. This relation suggests the MZ_t^{GLS} test defined by (9) since it satisfies the relation

$$MZ_t^{GLS} = MSB^{GLS} \cdot MZ_\alpha^{GLS}.$$

Another test of interest is the so-called ADF test (Dickey and Fuller, 1979, Said and Dickey, 1984) which is the t-statistic for testing $b_0 = 0$ in regression (11). We denote this test by ADF^{GLS}.

3.2 The Asymptotic Distributions of the Tests

The following theorem gives the asymptotic distribution of the residual based tests for cointegration under the null hypothesis of no-cointegration.

Theorem 2 *Suppose that z_t is generated by (1) with $d_t = \psi' m_t$. Then, as $T \to \infty$.*

[3] The advantages of using this autoregressive-based spectral density estimator over the more traditional kernel-based methods are discussed in Perron and Ng (1998).

a. *If $p = 0$, with $m_t = 1$ for all t:*

$$
\begin{aligned}
Z_\alpha^{GLS},\ MZ_\alpha^{GLS} &\Rightarrow \frac{0.5\ K_1(c,m)}{K_2(c,m)}, \\
MSB^{GLS} &\Rightarrow \left(\frac{K_2(c,m)}{\kappa'\kappa}\right)^{1/2}, \\
ADF^{GLS},\ Z_t^{GLS},\ MZ_t^{GLS} &\Rightarrow \frac{0.5\ K_1(c,m)}{(K_2(c,m)\ \kappa'\kappa)^{1/2}},
\end{aligned} \tag{12}
$$

where all these limiting distributions can be denoted by $H_\mu^{J^{GLS}}(c,m)$ for $J = Z_\alpha^{GLS}$, MZ_α^{GLS}, ADF^{GLS}, Z_t^{GLS} and MZ_t^{GLS} and where

$$
\begin{aligned}
K_1(c,m) &= Q_c(1)^2 - k'k, \\
K_2(c,m) &= \int_0^1 Q_c(r)^2 dr, \\
Q_c(r) &= W_{1c}(r) - (\int_0^1 W_{1c}(s)W_{2c}'(s)ds)(\int_0^1 W_{2c}(s)W_{2c}'(s)ds)^{-1}W_{2c}(r), \\
k' &= (1, -f_{21}'F_{22}^{-1}),
\end{aligned}
$$

with $f_{21} = \int_0^1 W_{2c}(r)W_{1c}(r)dr$, $F_{22} = \int_0^1 W_{2c}(r)W_{2c}'(r)dr$ and $W_c(r) = [W_{1c}(r), W_{2c}(r)]'$ is a multivariate Ornstein-Uhlenbeck process, i.e. the solution to the stochastic differential equation $dW_c(r) = CW_c(r) + dW(r)$ with $W_c(0) = 0$, where $C = diag(c, ..., c)$ a n by n diagonal matrix and $W(r)$ is a n vector of independent Wiener processes.

b. *If $p = 1$, with $m_t' = (1, t)$ for all t:*

$$
\begin{aligned}
Z_\alpha^{GLS},\ MZ_\alpha^{GLS} &\Rightarrow \frac{0.5\ \overline{K}_1(c,\bar{c},m)}{\overline{K}_2(c,\bar{c},m)}, \\
MSB^{GLS} &\Rightarrow \left(\frac{\overline{K}_2(c,\bar{c},m)}{\overline{\kappa}'\overline{\kappa}}\right)^{1/2}, \\
ADF^{GLS},\ Z_t^{GLS},\ MZ_t^{GLS} &\Rightarrow \frac{0.5\ \overline{K}_1(c,\bar{c},m)}{(\overline{K}_2(c,\bar{c},m)\ \overline{\kappa}'\overline{\kappa})^{1/2}},
\end{aligned} \tag{13}
$$

where all these limiting distributions can be denoted by $H_\tau^{J^{GLS}}(c,m)$ for $J = Z_\alpha^{GLS}$, MZ_α^{GLS}, ADF^{GLS}, Z_t^{GLS} and MZ_t^{GLS} and where

$$
\begin{aligned}
\overline{K}_1(c,\bar{c},m) &= \overline{Q}_c(1)^2 - \overline{k}'\overline{k}, \\
\overline{K}_2(c,\bar{c},m) &= \int_0^1 \overline{Q}_c(r)^2 dr, \\
\overline{Q}_c(r) &= \overline{W}_{1c}(r) - (\int_0^1 \overline{W}_{1c}(s)\overline{W}_{2c}'(s)ds)(\int_0^1 \overline{W}_{2c}(s)\overline{W}_{2c}'(s)ds)^{-1}\overline{W}_{2c}(r), \\
\overline{W}_c &= W_c(r) - rE(\bar{c}),
\end{aligned}
$$

$$E(\bar{c}, m) = \Lambda(\bar{c})W_c(1) + 3(I_n - \Lambda(\bar{c}))\int_0^1 sW_c(s)ds,$$
$$\overline{k}' = (1, -\overline{f}'_{21}\overline{F}_{22}^{-1}).$$

where $\overline{f}_{21} = \int_0^1 \overline{W}_{2c}(r)\overline{W}_{1c}(r)dr$ *and* $\overline{F}_{22} = \int_0^1 \overline{W}_{2c}(r)\overline{W}'_{2c}(r)dr$.

Note that when $c = 0$ the limiting distributions for the tests Z_α^{GLS}-MZ_α^{GLS}, Z_t^{GLS}-MZ_t^{GLS} and ADF^{GLS} shown in (12) for the case $p = 0$ are those derived by Phillips and Ouliaris (1990) for the Z_α, Z_t and ADF tests when no deterministic component is present. This is a simple consequence of the use of local GLS detrending because the intercept is bounded in probability.

3.3 The case with $p = 1$ and only a constant is used to detrend the data

When dealing with trending date, it is often still the case that a researcher may want to include only an intercept in the set of deterministic components. This is the case when the hypothesis of interest is that of deterministic cointegration in the terminology of Campbell and Perron (1991). This refers to the case where it is supposed that the cointegrating vector annihilates both the stochastic and deterministic nonstationarity in the system. With least-squares detrending, this case was analyzed by Hansen (1992). Using a proof similar to his we prove the following result in the appendix. Note that we concentrate on the case with $c = 0$, i.e., pure unit root processes.

Theorem 3 *Suppose that z_t is generated by (1) with $d_t = \mu + \beta t$ and β non-zero. Let $\beta = (\beta_y, \beta_x)$ where β_y is a scalar and β_x an m by 1 vector of rank 1. Consider the case where only a constant is included in the set of deterministic components used to detrend the data, i.e., $m_t = 1$. The limiting distributions of the tests are given by:*

$$Z_\alpha^{GLS},\ MZ_\alpha^{GLS} \Rightarrow \frac{0.5\ (Q^*(1)^2 - \kappa^{*\prime}\kappa^*)}{\int_0^1 Q^*(r)^2 dr},$$
$$MSB^{GLS} \Rightarrow \left(\int_0^1 Q^*(r)^2 dr/\kappa^{*\prime}\kappa^*\right)^{1/2},$$
$$ADF^{GLS},\ Z_t^{GLS},\ MZ_t^{GLS} \Rightarrow \frac{0.5\ (Q^*(1)^2 - \kappa^{*\prime}\kappa^*)}{\left(\kappa^{*\prime}\kappa^* \int_0^1 Q^*(r)^2 dr\right)^{1/2}},$$

where all these limiting distributions can be denoted by $H_{\mu\tau}^{J^{GLS}}(c, m)$ for $J = Z_\alpha^{GLS}$, MZ_α^{GLS}, ADF^{GLS}, Z_t^{GLS} and MZ_t^{GLS} and where

$$Q^*(r) = W_1(r) - (\int_0^1 W_1(r)J_R'(r))(\int_0^1 J_R(r)J_R'(r))^{-1}J_R(r)$$

with $W_1(r)$ a scalar unit Wiener process, and

$$J_R(r) = \begin{pmatrix} W_{m-1}(r) \\ r \end{pmatrix}.$$

Also

$$\kappa^* = (1, -(\int_0^1 W_1(r)J_R'(r))(\int_0^1 J_R(r)J_R'(r))^{-1})'.$$

Note that the limit distributions are identical to those that would obtain if $m-1$ (instead of m) $I(1)$ regressors were included and the data was detrended by least-squares using a time trend only without an intercept. When $m = 1$, the limit distributions correspond to those of unit root tests where the variable is again detrended using a time trend only without an intercept.

This result is similar in spirit to that obtained by Hansen (1992). The limit distribution has, however, not been tabulated. We shall therefore provide critical values in the next section.

4 Feasible Point Optimal Test and the Power Envelope

While a uniformly most powerful test is not attainable, following the univariate suggestion of Dufour and King (1991) and ERS, it is possible to define a point optimal test against the alternative $\alpha = \overline{\alpha}$. If v_t is *i.i.d.*, this is provided by the likelihood ratio statistic, which simplifies, under normality, to $L = S(\overline{\alpha}) - S(1)$, where $S(\overline{\alpha})$ and $S(1)$ are the sums of squared errors from a GLS regression with $\alpha = \overline{\alpha}$ and $\alpha = 1$, respectively. Varying the value of $\overline{\alpha}$, gives a family of point optimal tests and the Gaussian power envelope for testing $\alpha = 1$. In our contexte, the testing hypothesis is performed on the residuals from (3). We use a similar specification to the univariate formula suggested by ERS allowing for correlation in v_t. Recalling that $\overline{\alpha} = 1 + \overline{c}T^{-1}$, let $S(\overline{\alpha}) = \sum_{t=1}^T (\widehat{e}_t^{\overline{\alpha}})^2 = \sum_{t=1}^T (\Delta\widehat{e}_t - \overline{c}T^{-1}\widehat{e}_{t-1})^2$ and $S(1) = \sum_{t=1}^T (\widehat{e}_t^{(1)})^2 = \sum_{t=1}^T (\Delta\widehat{e}_t)^2$, then the feasible point optimal test is defined as

$$P_{1T}^{GLS} = \frac{S(\overline{\alpha}) - \overline{\alpha}S(1)}{s^2} \tag{14}$$

where $\widehat{e}_t$ are the residuals from the equation (3). The following theorem specifies the limiting distribution of this feasible point optimal test.

Theorem 4 *Suppose that z_t is generated by (1) with $d_t = \psi' m_t$. Then, as $T \to \infty$, the limiting distribution of the feasible point optimal test P_{1T}^{GLS} is given by:*

$$P_{1T}^{GLS} \Rightarrow \frac{-\overline{c}Q(1)^2 + \overline{c}^2 \int_0^1 Q(r)^2 dr}{\overline{\kappa}'\overline{\kappa}} \equiv H_\mu^{P_{1T}^{GLS}}(c, \overline{c}, m), \qquad p = 0 \quad (15)$$

$$P_{1T}^{GLS} \Rightarrow \frac{-\overline{c}\overline{Q}(1)^2 + \overline{c}^2 \int_0^1 \overline{Q}(r)^2 dr}{\overline{\kappa}'\overline{\kappa}} \equiv H_\tau^{P_{1T}^{GLS}}(c, \overline{c}, m), \qquad p = 1, \quad (16)$$

where the notation is the same as before.

The asymptotic expressions (15) and (16) for the P_{1T}^{GLS} test allow us to define the asymptotic power envelope for the model where there is only a constante and for the model where there is also a time trend, respectively. It is given by $\pi(c) = \Pr[H^{P_{1T}^{GLS}}(c, c, m) < b^{P_{1T}^{GLS}}(c, m)]$, where $b^{P_{1T}^{GLS}}(c, m)$ is such that $\Pr[H^{P_{1T}^{GLS}}(0, c, m) < b^{P_{T}^{GLS}}(c, m)] = \upsilon$, with υ the size of the test.

Note that $Q(.)$ and $\overline{Q}(.)$ are functions of the dimension of $W_2(.)$ and consequently, a different power envelope exists for different number of regressors. Because this dependence of dimensionality, a different "optimal" non-centrality parameter $\overline{c}$ will exist for each model and for each number of regressors. We follow the univariate recommendation of ERS to choose the value of $\overline{c}$ such the asymptotic power of the feasible point optimal test is 50.0%, i.e., $\overline{c}$ such that $\Pr[H^{P_{1T}^{GLS}}(\overline{c}, \overline{c}, m) < b^{P_{T}^{GLS}}(\overline{c}, m)] = 0.5$. Using simulations, we found that $\overline{c} = -12.75, -17.0, -21.5, -24.75 - 28.5$ for the model where $m_t = \{1\}$ and for $m = 1, 2, 3, 4, 5$ regressors. In the case of the model $m_t = \{1, t\}$ and same number of regressors, we found $\overline{c} = -18.25, -22.50, -27.0, -31.0, -35.5$, respectively. We use these values in the rest of the paper.

5 Asymptotic Critical Values and Asymptotic Power Functions

As seen in Theorems 2 and 4 (specially in the case with a linear time trend $(p = 1)$), the limiting distributions of the test statistics depend on both noncentrality parameters c and $\bar{c}$ and also the dimension of $W_2(r)$. Since the concept of cointegration usually involves variables with a unit root and that residual based tests are constructed from a regression with $I(1)$ variables (most often as a result of some unit root pre-test), it is natural to consider tabulating critical values setting $c = 0$. We maintain this specification when computing asymptotic critical values.

It has been argued that the GLS or quasi-difference device introduced by ERS allows unit root tests with better power because the coefficients of the trend function are more precisely estimated (see Canjels and Watson, 1997, and Phillips and Lee, 1996 for evidence on the latter assertion). Following this argument, the proper choice would be to set the non-centrality parameter $\bar{c} = 0$, since all series are effectively pure $I(1)$ processes. As will be verified by the simulations reported below, this argument is, however, incorrect. What permits an increase in power is not a more precise estimate of the trend function but rather the fact that by applying a GLS transformation we get more precise estimates of the coefficients in the static regression under the alternative hypothesis of cointegration. To see this, consider the triangular representation of the data generating process (Phillips, 1991):

$$\begin{aligned} y_t &= \mu' m_t + \gamma' x_t + \nu_t, \\ x_t &= \delta + x_{t-1} + u_{2t}, \end{aligned} \tag{17}$$

where

$$\nu_t = \bar{\alpha}\nu_{t-1} + \varepsilon_t,$$

with $\bar{\alpha} = 1 + \bar{c}/T$ and, for simplicity, $\varepsilon_t \sim i.i.d.\ (0, \sigma_\varepsilon^2)$. Hence, in this setup, y_t and x_t are cointegrated for any finite samples but no cointegrated in the limit as $T \to \infty$. Hence, this is a case of a nearly non-cointegrated process similar to the nearly integrated framework used by ERS. Ignoring the first observation, we can apply a GLS transformation and write (17) as

$$y_t^{\bar{\alpha}} = \mu' m_t^{\bar{\alpha}} + \gamma' x_t^{\bar{\alpha}} + \varepsilon_t \tag{18}$$

and the GLS estimates of μ and γ are obtained by applying OLS to the equation (18) which is equivalent to applying OLS to the regression

$$y_t^a = \gamma' x_t^a + \varepsilon_t, \tag{19}$$

where y_t^a and x_t^a are the GLS detrended data defined in Section 2.2, i.e. $(y_t^a, x_t^{a\prime})' = (y_t, x_t')' - \hat{\psi}' m_t$ where $\hat{\psi}$ is defined by (2). Accordingly, the method to construct the test statistic is one that provides, via the GLS transform, better estimates of the parameters of the cointegrating regression, μ and γ, and not better estimates of the coefficients of the trend function of the variables of the system, e.g. δ for the regressors x_t. This becomes even more transparent in the case where the data are trending and only a constant is included in the set of deterministic components to detrend the data. In the latter case, there is not even an attempt to estimate the deterministic components.

To assess the extent to which the quality of the inference is dependent on that particular choice, we shall consider the size and power of the tests for the values found using power envelope as well as for $\overline{c} = 0$ irrespective of the nature of the deterministic components (which is the optimal choice if the goal is to get better estimates of the coefficients of the trend function of the variables).

We simulate directly the asymptotic distributions using $1,000$ steps to approximate the Wiener process $W(r)$ as the partial sums of a vector of n independent $i.i.d.\ N(0,1)$ random variables. In all cases, $10,000$ replications are used and we present critical values for $m = 1, 2, 3, 4$, and 5, where m is the number of right-hand side variables in the cointegrating regression. To assess the adequacy of the asymptotic distribution, we also present critical values for the finite sample distributions of selected statistics for a sample of size $T = 100$ with data generated by independent random walks with zero initial condition and $i.i.d.\ N(0,1)$ errors. Here, lag lenght k is set to unity to construct s^2.

The results are presented in Table 1 for the case where $p = 0$ and $\overline{c} = 0$; in Table 2 for $p = 0$ and $\overline{c}$ as chosen by the power envelope; in Table 3 for $p = 1$ and $\overline{c} = 0$, and in Table 4 for $p = 1$ and $\overline{c}$ as chosen by the power envelope. In general, the asymptotic distributions provide good approximations to the finite sample distributions. Note that when the model includes a time trend and GLS detrending is performed at $\overline{c} = 0$ the asymptotic critical values are similar to those derived by Phillips and Ouliaris (1990) for the case where only an intercept is included and removed by OLS, even though the theoretical limit distributions are different.

For completeness, Table 5 presents the asymptotic critical values in the case with $p = 1$ and only a constant is used to detrend the data. These were obtained using 20,000 replications of the limiting representation in Theorem 3. Two results of interest stand out. First, unlike the case where both a constant and a time trend are included, the distribution is little affected by the value of $\overline{c}$. Indeed, the critical values are basically the same with $\overline{c} = 0$ and $\overline{c}$ selected by the power envelope. This accords with our theoretical result which showed the asymptotic distribution to be invariant to $\bar{c}$. Second, comparing Table 5 and the asymptotic critical values in Table 1 or 2 (for the case with non-trending data and only a constant included), reveals that the asymptotic distribution are very similar. Hence, little loss in accuracy would be incurred from using the critical values for the case with non-trending data even if the data are trending.

The asymptotic power functions of the tests are defined by $\pi_{J^{GLS}}(c, \overline{c}) = \Pr[H^{J^{GLS}}(c, \bar{c}, m) < b^{J^{GLS}}(\bar{c}, m)]$ for $J = Z_\alpha, Z_t, MZ_\alpha$, MSB, MZ_t, ADF and with $H^i(c, \bar{c})$ defined in Theorems 2 and 4. The constants $b^{J^{GLS}}(\bar{c}, m)$ is such that $\Pr[H^{J^{GLS}}(0, \bar{c}, m) < b^{J^{GLS}}(\bar{c}, m)] = \upsilon$, the size of the tests. The asymptotic power functions are shown in Figures 1 and 2 for the two cases considered and where the solid line denotes the power envelope. All tests have have asymptotic power functions very close to the power envelope. Naturally, the power decreases when the number of right-hand regressors increase and when more deterministic components are included.

6 Size and Power of the Tests in Finite Samples

In this section, we evaluate the size and power properties of the tests considered in section 3. For comparison, we also include tests based on OLS detrended series, namely the Z_α, Z_t and ADF tests of Phillips and Ouliaris (1990) and the MZ_α test proposed by Stock (1999) (we do not report results for the MZ_t and MSB as they are very similar to those of MZ_α). This will allow us to assess the advantages of using GLS detrended series to construct the tests. Also, we only report results for MZ_α^{GLS} as they are similar to those for MZ_t^{GLS} and MSB^{GLS}. At the end, we report results of size and power for the single-equation based ECM tests.

6.1 The Monte Carlo Design

The data-generating process considered is given by:

$$\begin{aligned} y_t - \beta x_t &= u_{1t}, \\ a_1 y_t - a_2 x_t &= u_{2t}, \\ u_{1t} &= \alpha u_{1t-1} + e_{1t}, \\ u_{2t} &= u_{2t-1} + e_{2t}, \\ \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} &= i.i.d. \quad N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \eta\sigma \\ \eta\sigma & \sigma^2 \end{pmatrix} \right]. \end{aligned} \tag{20}$$

This data-generating process is similar to that used by Banerjee et al. (1986), Gonzalo (1994) and Haug (1996). Here, y_t and x_t are scalars. For each Monte-Carlo experiment, we generate $2,000$ series of length T and we start at $u_{20} = 0$ and $u_{10} = 0$. We use the *rndns* function in Gauss-Unix with a $seed = 123456$ in order to generate the pseudo normal variates e_{1t} and e_{2t}.

The values of the various parameters considered are similar to those used in Gonzalo (1994) and Haug (1996), among others. The parameter β is assumed unity and a_2 is assumed minus unity. For the parameter a_1, two cases are considered. The first sets $a_1 = 0$, which is equivalent to assuming that the variable x_t is exogenous with respect to the parameter β. We also consider $a_1 = 1$, in which case x_t is endogenous. The parameter η measures the presence of contemporaneous correlation across the errors. Here, we consider three possible values (-0.5, 0.0 and 0.5). Notice that the case where $\eta = 0$ in conjunction with $a_1 = 0$ implies strong exogeneity. The parameter σ is known as the signal to noise ratio and it measures how big is the random walk component in the series. We consider three possible values for this parameter (0.25, 1.0 and 4.0). Two sample sizes are used, $T = 100$ and 200.

Note that we consider only the case with *i.i.d.* errors and, hence, the effects induced by the presence of correlation are not analyzed. Our principal goal is to appreciate gains in power from tests constructed using GLS detrended data compared to the case where OLS detrended series are used[4]. Finally, the BIC is used to select lag length k when estimating the autoregression (11) to construct s^2 and the ADF tests.

The size and power are evaluated using asymptotic critical values. For the Z_α, MZ_α, Z_t and ADF tests, we use those from Phillips and Ouliaris (1990). For the GLS-based tests, we use the asymptotic critical values tabulated in Tables 1 to 4 for the case where $m = 1$.

6.2 Size of the Tests

Size is evaluated at $\alpha = 1$. The results are reported in Tables 6 and 7 for the case where the model includes only a constant and in Tables 8 and 9 for the case where both a constant and a time trend are included. Tables 6 and 8 consider the case where x_t is endogenous and Tables 7 and 9 the case where x_t is weakly exogenous. Given that we use $2,000$ replications an approximate 95% confidence interval for the p-value .05 is $(.04, .06)$. What transpires from the results is that the exact sizes of all tests, except MZ_α, are within this confidence interval (or very close to it) in all cases considered. The tests MZ_α have exact sizes slightly below the nominal 5% size when $T = 100$. These distortions are reduced when $T = 200$ for the case $p = 0$ but remain somewhat, though to a lesser extent, when $p = 1$. The conservative nature of the M tests was also noted by Perron and Ng (1996) in the case of unit root tests. It is also clear that some higher

[4] We also conducted experiments for MA(1) and AR(1) process in the errors. The relative ranking and the general conclusions remain the same unless these is strong correlation in which case the class of M tests suffer much less from size distortions. The results are available upon request.

exact sizes (around 7-8%) are obtained when $\overline{c}$ selected by the power envelope is used in the simulations. However, these small distortions are reduced when $T = 200$.

6.3 Power of the Tests

Power is evaluated at $\alpha = 0.85$ when $T = 100$ and at $\alpha = 0.90$ when $T = 200$. The results are reported in Tables 10 and 11 for the case where the model includes only a constant and in Tables 12 and 13 for the case where both a constant and a time trend are included. Tables 10 and 12 consider the case where x_t is endogenous and Tables 11 and 13 the case where x_t is weakly exogenous.

The qualitative results are consistent across the various configurations considered. The most striking and important feature is the fact that substantial power gains can be achieved using GLS detrended data instead of OLS detrended data. These gains are more important when the quasi-differencing parameter is set to $\bar{c} = -12.75$ (if $p = 0$) or $\overline{c} = -18.25$ (if $p = 1$), particularly when $\sigma \geqslant 1$. For example, consider the ADF test in the case $p = 0$, $\sigma = 4$, $a_1 = 1$ and $\eta = 0.0$ with $T = 100$ (whose results are presented in the second to last column of Table 10). The power of the test using the standard test of Phillips and Ouliaris (1990) based on OLS detrended data is .449. With GLS detrended data this power increases to .617 when using a quasi-differencing parameter $\overline{c} = 0$. When using a the non-centrality parameter $\bar{c} = -12.75$ (as suggested by ERS for unit root tests), the power is further increased to .764. This increase in power holds in all configurations considered.

Concerning the relative ranking of the various tests considered, the most powerful are the ADF^{GLS} and the Z_t^{GLS} tests followed by the Z_α^{GLS}. The test MZ_α^{GLS} appears significantly less powerful, no doubt due to the fact that its exact size is conservative. Note, however, that in the presence of a large negative moving-average component, the MZ_α^{GLS} test (as well as the MZ_t^{GLS} and MSB^{GLS} tests) has the least size distortions.

Xiao and Phillips (1999), in the context of tests within a full system, obtained similar simulation results, namely that power was higher with the non-centrality parameter $\bar{c}$ set at -7.0 ($p = 0$) or -13.5 ($p = 1$) compared to using $\bar{c} = 0$. Their explanation was that the estimates were numerically less reliable when $\bar{c} = 0$ which caused a loss of power in finite samples. We believe, instead, that the simulations show support for our interpretation outlined in Section 4, that in the context of unit root or cointegration tests, the use of GLS detrended data does not arise from a concern to get better estimates of the coefficients of the trend functions of the variables in the system but rather arises as a consequence of trying to obtain better estimates of the cointegrating regression when the errors are modelled as nearly-integrated.

6.4 A Comparison with ECM Tests

Another kind of cointegration tests based on estimates of a single equation are the so called ECM tests; see Banerjee, Dolado and Mestre (1998), Kremers, Ericsson and Dolado (1995), Boswijk (1994) and Zivot (2000) among many others. These tests are also easy to apply empirically although critical values are not available in the same way. For a reference about the citical values and new tabulated critical values, see Ericsson and MacKinnon (1999).

Essentially these literature has shown that ECM tests have more power to OLS residual-based tests for cointegration. Using the same DGP as before, we calculated size and power of three ECM tests. The first is the t-statistic testing the coefficient associated to the error mechanism. The second test was proposed by Bowsjik (1994) and consists in an F-statistic about the significance of the coefficient of the error mechanism and when deterministic components (a constant and/or also a time trend) are unrestricted under the null hypothesis. The third tests (Bowsjik, 1994) is similar as before but the deterministic components are restricted to enter only in the error mechanism. Asymptotic critical values for the t-statistic were taken from Ericsson and MacKinnon (1999) and for the other two statistics, the source is Bowsjik (1994). Same number of replications were used and no lags were used because we are considering i.i.d. disturbances. The goal is only to compare, in equal circunstances, size and power properties of ECM tests against GLS residual-based tests for cointegration.

The results for size and power of the ECM tests are presented in Tables 14-17 following a similar structure in the presentation as in the Tables for GLS residual-based tests for cointegration. Overall, we observe that our tests have similar or higher power compared to those values obtained from ECM tests. This is true when there is included only a constant in the regression and when there is also a time trend included. Some more power is observed from ECM tests when $a_1 = 0$, which represents the case where right-hand regressors are exogenous; consequently this escenario is not really an important advantage. In another hand, our tests have less power when $\eta = -0.50$ but they have higher power when η is zero or positive.

7 Conclusions

We have considered residuals-based tests for cointegration using GLS detrended data and provided asymptotic critical values for use in applied work. Our simulation results show that substantial gains in power can be obtained relative to the OLS residuals-based tests of Phillips and Ouliaris (1990) especially when using a quasi-differencing parameter selected using the power envelope and according to the nature of the deterministic components included in the regression. Given that our tests are basically as easy to construct, they provide a useful alternative for empirical applications. Comparison of size and power from ECM tests are favorabe to our tests.

8 Appendix

Throughout, we use the following lemmas which are by now standard (see Lemma 3.1 of Phillips (1988) and Lemma 2.2 of Phillips and Ouliaris (1990)).

Lemma A.1: Let $\{u_t\}$ be a near-integrated series generated by (1), then: a) $T^{-1/2}u_{[Tr]} \Rightarrow B_c(r)$; b) $T^{-3/2}\sum_{t=1}^{T} u_t \Rightarrow \int_0^1 B_c(r)dr$; c) $T^{-2}\sum_{t=1}^{T} u_t u_t' \Rightarrow \int_0^1 B_c(r)B_c'(r)dr$; d) $T^{-1}\sum_{t=1}^{T} u_{t-1}v_t$ $\Rightarrow \{\int_0^1 B_c(r)dB_c(r)' + \quad _1\}$ with $\quad _1 = \sum_{k=1}^{\infty} E(v_0 v_k')$.

Lemma A.2: Using the notation defined in Section 2, we have: a) $B_c(r) \equiv L'W_c(r)$; b) $L\eta \equiv l_{11}k$, $\quad \eta' \quad \eta \equiv w_{11.2}k'k$; c) $\eta' B_c(r) \equiv l_{11}Q_c(r)$; d) $\eta' \int_0^1 B_c(r)dB_c'(r)\eta \equiv w_{11.2}\int_0^1 Q_c(r)dQ_c(r)$; e) $\eta' \int_0^1 B_c(r)B_c'(r)dr\eta \equiv w_{11.2}\int_0^1 Q_c(r)^2 dr$; with $w_{11.2} = w_{11} - w_{21}' \quad _{22}^{-1} w_{21} = l_{11}^2$.

Proof of Lemma 1: In both cases, the proof is a multivariate extension of arguments given in ERS (1996) for the univariate case and is, hence, omitted.

Proof of Theorem 1: Consider first the case where $m_t = \{1\}$. Let $z_t^a = (y_t^a, x_t^{a\prime})'$ be a n-vector, $M_t' = (u_t', m_t')$ be a $n{+}1$-vector and let $\Upsilon_T = diag(I_n, T^{-1/2})$ be a diagonal matrix of dimension $n+1$. We can write:

$$T^{-2}\sum_{t=1}^{T} z_t^a z_t^{a\prime} = \begin{bmatrix} I_n \\ -(\hat{\psi} - \psi) \end{bmatrix}' \Upsilon_T \begin{bmatrix} T^{-2}\Upsilon_T^{-1}\sum_{t=1}^{T} M_t M_t' \Upsilon_T^{-1} \end{bmatrix} \Upsilon_T \begin{bmatrix} I_n \\ -(\hat{\psi} - \psi) \end{bmatrix} \tag{A.1}$$

Using the notation $\psi_0 = (\psi_{0,y}, \psi_{0,x})$ where $\psi_{0,y}$ corresponds to the mean of the dependent variable and $\psi_{0,x}$ is an m-vector of means associated with the regressors x_t (a similar partition is used for $\hat{\psi}_0$). The first element in (A.1) can be written as

$$\begin{aligned} \begin{bmatrix} I_n & -(\hat{\psi} - \psi)' \end{bmatrix} \Upsilon_T &= \begin{bmatrix} 1 & \mathbf{0}_m' & -(\hat{\psi}_{0,y} - \psi_{0,y}) \\ \mathbf{0}_m & I_m & -(\hat{\psi}_{0,x} - \psi_{0,x}) \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_m' & 0 \\ \mathbf{0}_m & I_m & 0 \\ 0 & \mathbf{0}_m' & T^{-1/2} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \mathbf{0}_m' & -T^{-1/2}(\hat{\psi}_{0,y} - \psi_{0,y}) \\ \mathbf{0}_m & I_m & -T^{-1/2}(\hat{\psi}_{0,x} - \psi_{0,x}) \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 1 & \mathbf{0}_m' & 0 \\ \mathbf{0}_m & I_m & \mathbf{0}_m \end{bmatrix} \end{aligned}$$

The limit of the second element is:

$$\left[T^{-2}\Upsilon_T^{-1}\sum_{t=1}^{T} M_t M_t' \Upsilon_T^{-1}\right] = T^{-2}\begin{bmatrix} 1 & \mathbf{0}_m' & 0 \\ \mathbf{0}_m & I_m & \mathbf{0}_m \\ 0 & \mathbf{0}_m' & T^{1/2} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^{T} u_{1t}u_{1t} & \sum_{t=1}^{T} u_{1t}u_{2t}' & \sum_{t=1}^{T} u_{1t} \\ \sum_{t=1}^{T} u_{2t}u_{1t} & \sum_{t=1}^{T} u_{2t}u_{2t}' & \sum_{t=1}^{T} u_{2t} \\ \sum_{t=1}^{T} u_{1t} & \sum_{t=1}^{T} u_{2t}' & T \end{bmatrix}$$

$$\begin{bmatrix} 1 & \mathbf{0}_m' & 0 \\ \mathbf{0}_m & I_m & \mathbf{0}_m \\ 0 & \mathbf{0}_m' & T^{1/2} \end{bmatrix}$$

$$= \begin{bmatrix} T^{-2}\sum_{t=1}^T u_{1t}u_{1t} & T^{-2}\sum_{t=1}^T u_{1t}u_{2t}' & T^{-3/2}\sum_{t=1}^T u_{1t} \\ T^{-2}\sum_{t=1}^T u_{2t}u_{1t} & T^{-2}\sum_{t=1}^T u_{2t}u_{2t}' & \sum_{t=1}^T u_{2t} \\ T^{-3/2}\sum_{t=1}^T u_{1t} & T^{-3/2}\sum_{t=1}^T u_{2t}' & 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \int_0^1 B_{1c}(r)B_{1c}(r)dr & \int_0^1 B_{1c}(r)B_{2c}'(r)dr & \int_0^1 B_{1c}(r)dr \\ \int_0^1 B_{2c}(r)B_{1c}(r)dr & \int_0^1 B_{2c}(r)B_{2c}'(r)dr & \int_0^1 B_{2c}(r)dr \\ \int_0^1 B_{1c}(r)dr & \int_0^1 B_{2c}'(r)dr & 1 \end{bmatrix}$$

By symmetry, the third element converges to the transpose of the limit of the first and we have

$$T^{-2}\sum_{t=1}^T z_t^a z_t^{a\prime} \Rightarrow \begin{bmatrix} \int_0^1 B_{1c}(r)B_{1c}(r)dr & \int_0^1 B_{1c}(r)B_{2c}'(r)dr \\ \int_0^1 B_{2c}(r)B_{1c}(r)dr & \int_0^1 B_{2c}(r)B_{2c}'(r)dr \end{bmatrix} \quad \text{(A.2)}$$

$$\equiv \begin{bmatrix} g_{11} & g_{21}' \\ g_{21} & G_{22} \end{bmatrix} \equiv G$$

The limit of $\hat{g}$ follows directly. Now, consider the case where $m_t = \{1, t\}$. Let $\Upsilon_T = diag(I_n, T^{-1/2}, T^{1/2})$ be a diagonal matrix and using a similar partition for $(\hat{\psi} - \psi)$.

$$\begin{bmatrix} I_n & -(\hat{\psi}-\psi)' \end{bmatrix} \Upsilon_T$$

$$= \begin{bmatrix} 1 & \mathbf{0}_m' & -(\hat{\psi}_{0,y}-\psi_{0,y}) & -(\hat{\psi}_{1,y}-\psi_{1,y}) \\ \mathbf{0}_m & I_m & -(\hat{\psi}_{0,x}-\psi_{0,x}) & -(\hat{\psi}_{1,x}-\psi_{1,x}) \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_m' & 0 & 0 \\ \mathbf{0}_m & I_m & \mathbf{0}_m & \mathbf{0}_m \\ 0 & \mathbf{0}_m' & T^{-1/2} & 0 \\ 0 & \mathbf{0}_m' & 0 & T^{1/2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \mathbf{0}_m' & -T^{-1/2}(\hat{\psi}_{0,y}-\psi_{0,y}) & -T^{-1/2}(\hat{\psi}_{1,y}-\psi_{1,y}) \\ \mathbf{0}_m & I_m & -T^{-1/2}(\hat{\psi}_{0,x}-\psi_{0,x}) & -T^{-1/2}(\hat{\psi}_{1,x}-\psi_{1,x}) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & \mathbf{0}_m' & 0 & -D_y \\ \mathbf{0}_m & I_m & \mathbf{0}_m & -D_x \end{bmatrix}$$

where $D_y(\bar{c}) = \lambda B_{1c}(1) + 3(1-\lambda)\int_0^1 sB_{1c}(r)dr$, $D_x(\bar{c}) = \Lambda_x(\bar{c})B_{2c}(1) + 3(I_m - \Lambda_x(\bar{c}))\int_0^1 sB_{2c}(r)dr$ and $\Lambda_x(\bar{c}) = diag(\lambda, ..., \lambda)$ a m by m diagonal matrix. Also,

$$\left[T^{-2}\Upsilon_T^{-1}\sum_{t=1}^T M_t M_t' \Upsilon_T^{-1} \right]$$

$$= \begin{bmatrix} T^{-2}\sum_{t=1}^T u_{1t}u_{1t} & T^{-2}\sum_{t=1}^T u_{1t}u_{2t}' & T^{-3/2}\sum_{t=1}^T u_{1t} & T^{-5/2}\sum_{t=1}^T tu_{1t} \\ T^{-2}\sum_{t=1}^T u_{2t}u_{1t} & T^{-2}\sum_{t=1}^T u_{2t}u_{2t}' & T^{-3/2}\sum_{t=1}^T u_{2t} & T^{-5/2}\sum_{t=1}^T tu_{2t} \\ T^{-3/2}\sum_{t=1}^T u_{1t} & T^{-3/2}\sum_{t=1}^T u_{2t}' & 1 & T^{-2}\sum_{t=1}^T t \\ T^{-5/2}\sum_{t=1}^T tu_{1t} & T^{-5/2}\sum_{t=1}^T tu_{2t}' & T^{-2}\sum_{t=1}^T t & T^{-3}\sum_{t=1}^T t^2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \int_0^1 B_{1c}(r)B_{1c}(r)dr & \int_0^1 B_{1c}(r)B_{2c}'(r)dr & \int_0^1 B_{1c}(r)dr & \int_0^1 rB_{1c}(r)dr \\ \int_0^1 B_{2c}(r)B_{1c}(r)dr & \int_0^1 B_{2c}(r)B_{2c}'(r)dr & \int_0^1 B_{2c}(r)dr & \int_0^1 rB_{2c}(r)dr \\ \int_0^1 B_{1c}(r)dr & \int_0^1 B_{2c}'(r)dr & 1 & \int_0^1 rdr \\ \int_0^1 rB_{1c}(r)dr & \int_0^1 rB_{2c}'(r)dr & \int_0^1 rdr & \int_0^1 r^2dr \end{bmatrix}$$

Using these results, we have

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} z_t^a z_t^{a\prime} \quad &\Rightarrow \begin{bmatrix} \int_0^1 \overline{B}_{1c}(r)\overline{B}_{1c}(r)dr & \int_0^1 \overline{B}_{1c}(r)\overline{B}_{2c}'(r)dr \\ \int_0^1 \overline{B}_{2c}(r)\overline{B}_{1c}(r)dr & \int_0^1 \overline{B}_{2c}(r)\overline{B}_{2c}'(r)dr \end{bmatrix} \qquad \text{(A.3)} \\ &\equiv \begin{bmatrix} \overline{g}_{11} & \overline{g}_{21}' \\ \overline{g}_{21} & \overline{G}_{22} \end{bmatrix} \equiv \overline{G} \end{aligned}$$

where $\overline{B}_c = B_c(r) - rD(\bar{c})$, and the final expression for the limit of $\hat{g}$ follows directly.

Proof of Theorem 2: We present derivations only for the M^{GLS} tests; the results for the other tests follow using similar arguments and from the asymptotic equivalence with one of these three tests. Consider first the case where $m_t = \{1\}$. We have the following results:

a. $2T^{-2}\sum_{t=1}^{T} \hat{e}_t^2 = 2T^{-2}\sum_{t=1}^{T}\{y_t^a - \hat{\gamma}' x_t^a\}^2 = 2\hat{g}'\{T^{-2}\sum_{t=1}^{T} z_t^a z_t^{\prime a}\}\hat{g} \Rightarrow 2\eta' G\eta$.

From expression (A.2), we have $G = \int_0^1 B_c(r)B_c(r)'dr$, and, from Lemma A.2, $B_c(r) = L'W_c(r)$ and $L\eta = l_{11}k$. Furthermore, using the results $k'W_c(r) = Q_c(r)$ and $w_{11.2} = w_{11} - w_{21}' \;{}_{22}^{-1} w_{21} = l_{11}^2$, we have:

$$\begin{aligned} 2\eta' G\eta &= 2\int_0^1 [\eta' B_c(r)][\eta' B_c(r)]' dr \\ &= 2\int_0^1 [\eta' L' W_c(r)][\eta' L' W_c(r)]' dr \\ &= 2l_{11}^2 \int_0^1 [k' W_c(r)][k' W_c(r)]' dr \\ &= 2l_{11}^2 \int_0^1 Q_c(r)^2 dr \\ &= 2w_{11.2}\int_0^1 Q_c(r)^2 dr. \end{aligned}$$

b. $T^{-1}\hat{e}_T^2 = T^{-1}\{y_T^a - \hat{\gamma}' x_T^a\}^2 = \hat{g}'\{T^{-1} z_T^a z_T^{\prime a}\}\hat{g} \Rightarrow \eta' B_c(1)B_c(1)'\eta$.

Using similar arguments as in (a), we have:

$$\begin{aligned} [\eta' B_c(1)][\eta' B_c(1)]' &= [\eta' L' W_c(1)][\eta' L' W_c(1)]' \\ &= l_{11}^2 [k' W_c(1)][k' W_c(1)]' \\ &= l_{11}^2 Q_c(1)^2 \\ &= w_{11.2} Q_c(1)^2. \end{aligned}$$

c. $s^2 \Rightarrow \eta' \quad \eta$.

We follow the argument of Phillips and Ouliaris (1991). s^2 is a zero frequency spectral density estimate based on $\Delta\hat{e}_t = \hat{g}'\Delta z_t^a = \hat{g}'v_t$. Since $\hat{g} \Rightarrow \eta$, we can condition on the value of η. Now, as v_t, ηv_t is also a linear process satisfying the conditions of Berk (1974). Hence, s^2 is a consistent estimate of (2π times) the spectral density function at frequency zero of ηv_t given by $\eta' \quad \eta$. Using Lemma A.2, $\eta' \quad \eta = w_{11.2}\kappa'\kappa$, where $k' = (1, -f_{21}'F_{22}^{-1})$ and f_{21} and F_{22} are elements of the n by n matrix F defined as

$$\begin{aligned} F &= \begin{bmatrix} f_{11} & f_{21}' \\ f_{21} & F_{22} \end{bmatrix} \\ &= \begin{bmatrix} \int_0^1 W_{1c}(r)W_{1c}(r)dr & \int_0^1 W_{1c}(r)W_{2c}'(r)dr \\ \int_0^1 W_{2c}(r)W_{1c}(r)dr & \int_0^1 W_{2c}(r)W_{2c}'(r)dr \end{bmatrix} \end{aligned} \tag{A.4}$$

Using these results in expression (7), the proof is complete for the MZ_α^{GLS} test. For the MSB^{GLS} test, it is sufficient to use (a) and (c) in expression (8). Given the equivalence $MZ_t^{GLS} = MZ_\alpha^{GLS} * MSB^{GLS}$, the proof for the MZ_t^{GLS} test follows.

Consider now the case where $m_t = \{1, t\}$. We have following results:

a. $2T^{-2}\sum_{t=1}^T \hat{e}_t^2 = 2T^{-2}\sum_{t=1}^T\{y_t^a - \hat{\gamma}' x_t^a\}^2 = 2\hat{g}'\{T^{-2}\sum_{t=1}^T z_t^a z_t'^a\}\hat{g} \Rightarrow 2\overline{\eta}'\bar{G}\overline{\eta}$

From expression (A.3), we have $\overline{G} = \int_0^1 \overline{B}_c(r)\overline{B}_c(r)'dr$, and, from Lemma A.2, $\overline{B}_c(r) = L'\bar{W}_c(r)$ and $L\bar{\eta} = l_{11}\bar{k}$. Furthermore, using the results $\bar{k}'\bar{W}_c(r) = \bar{Q}_c(r)$ and $w_{11.2} = w_{11} - w_{21}' \quad {}_{22}^{-1} w_{21} = l_{11}^2$, we have:

$$\begin{aligned} 2\overline{\eta}'\overline{G}\overline{\eta} &= 2\int_0^1 [\overline{\eta}'\overline{B}_c(r)][\overline{\eta}'\overline{B}_c(r)]'dr \\ &= 2\int_0^1 [\overline{\eta}'L'\overline{W}_c(r)][\overline{\eta}'L'\overline{W}_c(r)]'dr \\ &= 2l_{11}^2\int_0^1 [\overline{k}'\overline{W}_c(r)][\overline{k}'\overline{W}_c(r)]'dr \\ &= 2l_{11}^2\int_0^1 \bar{Q}_c(r)^2 dr \\ &= 2w_{11.2}\int_0^1 \bar{Q}_c(r)^2 dr \end{aligned}$$

b. $T^{-1}\hat{e}_T^2 = T^{-1}\{y_T^a - \hat{\gamma}' x_T^a\}^2 = \hat{g}'\{T^{-1}z_T^a z_T'^a\}\hat{g} \Rightarrow \overline{\eta}'\overline{B}_c(1)\overline{B}_c(1)'\overline{\eta}$.

Using similar arguments as in (a), we have:

$$\begin{aligned} [\overline{\eta}'\overline{B}_c(1)][\overline{\eta}'\overline{B}_c(1)]' &= [\overline{\eta}'L'\overline{W}_c(1)][\overline{\eta}'L'\overline{W}_c(1)]' \\ &= l_{11}^2[\overline{k}'\overline{W}_c(1)][\overline{k}'\overline{W}_c(1)]' \\ &= l_{11}^2\bar{Q}_c(1)^2 \\ &= w_{11.2}\bar{Q}_c(1)^2 \end{aligned}$$

c. $s^2 \Rightarrow \bar{\eta}' \quad \bar{\eta}$, using arguments similar to the case with $m_t = \{1\}$.

Using Lemma A.2, $\overline{\eta}' \quad \overline{\eta} = w_{11.2}\overline{\kappa}'\overline{\kappa}$, where $\overline{k}' = (1, -\overline{f}'_{21}\overline{F}_{22}^{-1})$ and $\overline{f}_{21}$ and $\overline{F}_{22}$ are elements from a n by n matrix $\bar{F}$ given by (A.4) with $\bar{W}_{1c}(r)$ and $\bar{W}'_{2c}(r)$ instead of $W_{1c}(r)$ and $W'_{2c}(r)$. Using these results in expression (7), the proof is complete for the MZ_{α}^{GLS} test. For the MSB^{GLS} test, it is sufficient to use (a) and (c) in expression (8). Given the equivalence $MZ_t^{GLS} = MZ_{\alpha}^{GLS} * MSB^{GLS}$, the proof for the MZ_t^{GLS} test follows.

Proof of Theorem 3. The proof is similar to that in Hansen (1992). First from (1), $x_t = \mu_x + \beta_x t + u_{2t}$ and we can deduce that $x_t^a = x_t - \hat{\psi}_x$, with $\hat{\psi}_x \Rightarrow \beta_x(1 - \overline{c} + \overline{c}^2/2) + u_{21}$. Note that a similar result still for y_t. Consider the following m by $(m-1)$ matrix β_x^* which spans the null space of β_x and let

$$C = [C_1, C_2] = [\beta_x^*(\beta_x^{*\prime} \quad _{22}\beta_x^*)^{-1/2}, \beta_x(\beta_x'\beta_x)^{-1}].$$

Given that $(\beta_x^{*\prime} \quad _{22}\beta_x^*)$ is positive definite (because $\quad _{22}$ is positive definite) then C is well-defined. We have

$$C'x_t^a = \begin{pmatrix} C_1'x_t^a \\ C_2'x_t^a \end{pmatrix} = \begin{pmatrix} C_1'u_{2t} + O_p(1) \\ C_2'u_{2t} + t + O_p(1) \end{pmatrix}.$$

Defining the weight matrix $D_T = diag(I_m T^{1/2}, T)$, we have

$$\begin{aligned} D_T^{-1}C'x_{[Tr]}^a &= \begin{pmatrix} T^{-1/2}C_1'u_{2[Tr]} + o_p(1) \\ T^{-1}C_2'u_{2[Tr]} + T^{-1}[Tr] + o_p(1) \end{pmatrix} \\ &\Rightarrow \begin{pmatrix} W_{m-1}(r) \\ r \end{pmatrix} \\ &\equiv J_R \end{aligned}$$

where, following Hansen (1992),

$$W_{m-1}(r) = C_1'B_2(r) = (\beta_x^{*\prime} \quad _{22}\beta_x^*)^{-1/2}\beta_x^{*\prime}B_2(r) \equiv BM(I_{m-1})$$

Now, define the following n by n matrix

$$F = \begin{bmatrix} 1 & \mathbf{0}_m' \\ \mathbf{0}_m & C' \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_m' \\ \mathbf{0}_{m-1} & C_1' \\ 0 & C_2' \end{bmatrix}.$$

We first consider the limit distribution of the coefficients in the cointegrating equation given by

$$y_t^a = \widehat{\gamma}'x_t^a + \widehat{e}_t.$$

We have

$$\widehat{\gamma} = (x_t^a x_t^{a\prime})^{-1}(x_t^a y_t^a).$$

The following results are obtained in exactly the same way as in Hansen (1992). The limiting distribution of the estimates in the cointegrating equation is

$$T^{-1/2}D_T F^{-1}\widehat{g} \Rightarrow \begin{bmatrix} 1 \\ -(\int_0^1 J_R(r)J_R'(r))^{-1}(\int_0^1 J_R(r)B_1(r)) \end{bmatrix} \equiv \eta$$

where

$$B_1(r) = \overline{\omega}^{1/2}W_1(r) + \omega_{21}'C_1 W_{m-1}(r) \tag{A.5}$$

with $\overline{\omega} = \omega_{11} - \omega_{21}'C_1C_1'\omega_{21}$. To obtain the limiting distribution of the statistic MZ_α^{GLS}, we need the limiting distribution of $T^{-2}\sum_{t=1}^T \widehat{e}_t^2$, $T^{-1}\widehat{e}_T^2$ and s^2. From (A.18) of Hansen (1992),

$$2T^{-2}\sum_{t=1}^{T}\widehat{e}_t^2 \Rightarrow 2\bar{\omega}\int_0^1 Q^*(r)^2 dr$$

It can be shown using the same arguments that

$$T^{-1}\widehat{e}_T^2 \Rightarrow \bar{\omega}Q^*(1)^2.$$

Consider now s^2 which is a zero frequency spectral density estimate based on $\Delta\widehat{e}_t = \widehat{g}'\Delta z_t^a = \widehat{g}'(\beta + v_t) = (T^{-1/2}D_T F^{-1}\widehat{g})'T^{1/2}D_T^{-1}F(\beta + v_t)$. Since, we have $T^{-1/2}D_T F^{-1}\widehat{g} \Rightarrow \eta$, we can condition on the value of η. Now, as v_t, $\eta T^{1/2}D_T^{-1}Fv_t$ is also a linear process satisfying the conditions of Berk (1974). Hence, s^2 is a consistent estimate of (2π times) the spectral density function at frequency zero of $\eta T^{1/2}D_T^{-1}Fv_t$ given by $\lim_{T\to\infty}\eta' T^{1/2}D_T^{-1}F \quad F'T^{1/2}D_T^{-1}\eta$. We have:

$$\begin{aligned}
F \quad F' &= \begin{bmatrix} 1 & \mathbf{0}_m' \\ \mathbf{0}_{m-1} & C_1' \\ 0 & C_2' \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & {}_{22} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{m-1}' & 0 \\ \mathbf{0}_m & C_1 & C_2 \end{bmatrix} \\
&= \begin{bmatrix} \omega_{11} & \omega_{21}' \\ C_1'\omega_{21} & C_1' \quad {}_{22} \\ C_2'\omega_{21} & C_2' \quad {}_{22} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{m-1}' & 0 \\ \mathbf{0}_m & C_1 & C_2 \end{bmatrix} \\
&= \begin{bmatrix} \omega_{11} & \omega_{21}'C_1 & \omega_{21}'C_2 \\ C_1'\omega_{21} & C_1' \quad {}_{22}C_1 & C_1' \quad {}_{22}C_2 \\ C_2'\omega_{21} & C_2' \quad {}_{22}C_1 & C_2' \quad {}_{22}C_2 \end{bmatrix}
\end{aligned}$$

Note that $C_2' \quad {}_{22}C_2 = I_{m-1}$ and $C_1' \quad {}_{22}C_2 = \mathbf{0}_{m-1}$. Hence,

$$\begin{aligned}
\lim_{T\to\infty} T^{1/2}D_T^{-1}F \quad F'T^{1/2}D_T^{-1} &= \begin{bmatrix} \omega_{11} & \omega_{21}'C_1 & 0 \\ C_1'\omega_{21} & I_{m-1} & \mathbf{0}_{m-1} \\ 0 & \mathbf{0}_{m-1}' & 0 \end{bmatrix} \\
&\equiv \quad {}_0
\end{aligned}$$

and

$$s^2 \Rightarrow \eta' \quad {}_0\eta.$$

Now consider the matrix L_0 such that $\;_0 = L_0' L_0$ defined by

$$L_0 = \begin{bmatrix} \overline{\omega}^{1/2} & \mathbf{0}'_{m-1} & 0 \\ C_1'\omega_{21} & I_{m-1} & \mathbf{0}_{m-1} \\ 0 & \mathbf{0}'_{m-1} & 0 \end{bmatrix}.$$

We then have

$$\begin{aligned} L_0\eta &= \begin{bmatrix} \overline{\omega}^{1/2} & \mathbf{0}'_{m-1} & 0 \\ C_1'\omega_{21} & I_{m-1} & \mathbf{0}_{m-1} \\ 0 & \mathbf{0}'_{m-1} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -(\int_0^1 J_R(r)J_R'(r))^{-1}(\int_0^1 J_R(r)B_1(r)) \end{bmatrix} \\ &= \begin{bmatrix} \overline{\omega}^{1/2} \\ C_1'\omega_{21} - \begin{bmatrix} I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \left[(\int_0^1 J_R(r)J_R'(r))^{-1}(\int_0^1 J_R(r)B_1(r))\right] \\ 0 \end{bmatrix}. \end{aligned}$$

Using the representation (A.5) for $B_1(r)$, the second element is

$$C_1'\omega_{21} - \begin{bmatrix} I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \left\{ (\int_0^1 J_R(r)J_R'(r))^{-1}(\int_0^1 J_R(r)\left[\overline{\omega}^{1/2}W_1(r) + \omega_{21}'C_1W_{m-1}(r)\right]\right\}.$$

Finally, notice that $\begin{bmatrix} I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} J_R(r) = W_{m-1}(r)$. Using this expression the term $C_1'\omega_{21}$ disappears and

$$\begin{aligned} L_0\eta &= \begin{bmatrix} \overline{\omega}^{1/2} \\ -\overline{\omega}^{1/2} \begin{bmatrix} I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \left[(\int_0^1 J_R(r)J_R'(r))^{-1}(\int_0^1 J_R(r)W_1(r))\right] \\ 0 \end{bmatrix} \\ &= \overline{\omega}^{1/2} \begin{bmatrix} 1 \\ -\begin{bmatrix} I_{m-1} & \mathbf{0}_{m-1} \end{bmatrix} \left[(\int_0^1 J_R(r)J_R'(r))^{-1}(\int_0^1 J_R(r)W_1(r))\right] \\ 0 \end{bmatrix} \\ &= \overline{\omega}^{1/2}\kappa^* \end{aligned}$$

This shows that $s^2 \Rightarrow \overline{\omega}\kappa^{*\prime}\kappa^*$. Using these results, the proof is complete for the MZ_α^{GLS} and MSB^{GLS} tests. Given the equivalence $MZ_t^{GLS} = MZ_\alpha^{GLS} * MSB^{GLS}$, the proof for the MZ_t^{GLS} test follows.

Proof of Theorem 4. We know that we have cointegration when these residuals have not a uni root. In the converse case, we have an spuriouos regression. Following same way as in the univariate case (ERS, 1996) we propose the following feasible optimal point test P_{1T}^{GLS}:

$$s^2 P_{1T}^{GLS} = S(\overline{\alpha}) - \overline{\alpha}S(1) \tag{A.6}$$

where $S(\overline{\alpha})$ are the GLS detrended residuals under the alternative hypotehsis of stationarity and $S(1)$ are same residuals under the null hypothesis of a unit root. Hence, using GLS transformed data:

$$\begin{aligned} S(\overline{\alpha}) &= \sum(\widehat{e}_t^{\overline{\alpha}})^2 = \sum(\Delta\widehat{e}_t - \overline{c}T^{-1}\widehat{e}_{t-1})^2 && \text{(A.7)} \\ S(1) &= \sum(\widehat{e}_t^{(1)})^2 = \sum(\Delta\widehat{e}_t)^2 && \text{(A.8)} \end{aligned}$$

Then, we replace these terms in equation (A.6)

$$s^2 P_{1T}^{GLS} = \sum(\Delta\widehat{e}_t - \overline{c}T^{-1}\widehat{e}_{t-1})^2 - (1+\overline{c}T^{-1})\sum(\Delta\widehat{e}_t)^2 \tag{A.9}$$

We also know that $-2\overline{c}T^{-1}\Delta\widehat{e}_t\widehat{e}_{t-1} = -\overline{c}\widehat{e}_T^2 + \overline{c}T^{-1}\Delta\widehat{e}_t^2$. Using this we have

$$\begin{aligned} s^2 P_{1T}^{GLS} &= -\overline{c}T^{-1}\widehat{e}_T^2 + \overline{c}T^{-1}\sum\Delta\widehat{e}_t^2 + \overline{c}^2T^{-2}\sum\widehat{e}_{t-1}^2 - \overline{c}T^{-1}\sum\Delta\widehat{e}_t^2 \\ &= -\overline{c}T^{-1}\widehat{e}_T^2 + \overline{c}^2T^{-2}\sum\widehat{e}_{t-1}^2 \end{aligned} \tag{A.10}$$

Using arguments or results of convergencefound in Theorem 3, we have that

$$s^2 P_{1T}^{GLS} \Rightarrow \omega_{11.2}\{-\overline{c}\overline{Q}(1)^2 + \overline{c}^2\int\overline{Q}(r)^2dr\} \tag{A.11}$$

In another hand, we know that $s^2 \Rightarrow \overline{\eta}'$ $\overline{\eta} = \omega_{11.2}\overline{\kappa}'\overline{\kappa}$. With these results we have

$$P_{1T}^{GLS} \Rightarrow \frac{-\overline{c}\overline{Q}(1)^2 + \overline{c}^2\int\overline{Q}(r)^2dr}{\overline{\kappa}'\overline{\kappa}} \tag{A.12}$$

which completes the proof.

References

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

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

[3] Banerjee, A., J. J. Dolado and R. Mestre (1998), "Error-Corrcetion Mechanism Tests for Cointegration in a Single-equation Framework," *Journal of Time Series Analysis* **19 (3)**, 267-283.

[4] Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics* **2**, 489-502.

[5] Bhargava, A. (1986), "On the Theory of Testing for Unit Root in Observed Time Series," *Review of Economic Studies* **53**, 369-384.

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

[7] Campbell, J. Y. and P. Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," *NBER Macroeconomics Annual*, Vol. 6, O.J. Blanchard and S. Fisher (eds.), 141-201.

[8] Canjels, E. and M. W. Watson (1997), "Estimating Deterministic Trends in the Presence of Serialy Correlated Errors," *Review of Economics and Statistics* **79**, 184-200.

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

[10] Dufour, J.-M. and M. King (1991), "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary Errors," *Journal of Econometrics* **47**, 115-143.

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

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

[13] Ericsson, N. R. and J. G. MacKinnon (1999), "Distribution of Error Correction Tests for Cointegration", International Finance Discussion Paper No. 655, Board of Governors of the Federal Reserve System.

[14] Gonzalo, J. (1994), "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships," *Journal of Econometrics* **60**, 203-233.

[15] Granger, C. W. J. and P. Newbold (1974), "Spurious Regression in Econometrics," *Journal of Econometrics* **2**, 111-120.

[16] Hansen, B. E. (1992), "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends," *Journal of Econometrics* **53**, 87-121.

[17] Haug, A. A. (1996), "Tests for Cointegration. A Monte Carlo Comparison," *Journal of Econometrics* **71**, 89-115.

[18] Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* **59**, 1551-1580.

[19] Kremers, J. J., N. R. Ericsson and J. J. Dolado (1992), "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statisitics* **54 (3)**, 325-348.

[20] Lütkepohl, H. (1999), "Vector Autoregression," Discussion Paper 4, SFB 373, Humboldt-Universität zu Berlin.

[21] Lütkepohl, H. and P. Saikkonen (1999), "Testing for the Cointegrating Rank of a VAR Process with a Time Trend," *Journal of Econometrics*, forthcoming.

[22] Ng, S. and P. Perron (2000), "Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power," forthcoming in *Econometrica*.

[23] Perron, P. and J. Y. Campbell (1992), "Racines Unitaires en Macroéconomie: Le Cas Multidimensionnel," *Annales D'Économie et de Statistique* **27**, 1-50.

[24] Perron, P. and S. Ng (1996), "Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties," *Review of Economic Studies* **63**, 435-463.

[25] Perron, P. and S. Ng (1998), "An Autoregressive Spectral Density Estimator at Frequency Zero for Nonstationarity Tests," *Econometric Theory* **14**, 560-603.

[26] Phillips, P. C. B. (1986), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics* **33**, 311-340.

[27] Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root," *Econometrica* **55**, 277-302.

[28] Phillips, P. C. B. (1988), "Regression Theory for Near-Integrated Time Series," *Econometrica* **56**, 1021-1043.

[29] Phillips, P. C. B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica* **59**, 283-306.

[30] Phillips, P. C. B. and C. C. Lee (1996), "Efficiency Gains from Quasi-Differencing under Nonstationarity," in P. M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series: Essays in Memory of E. J. Hannan*, Springer-Verlag: New York.

[31] Phillips, P. C. B. and S. Ouliaris (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica* **58**, 165-193.

[32] Phillips, P. C. B. and P. Perron (1988), "Testing for Unit Roots in Time Series Regression," *Biometrika* **75**, 335-346.

[33] Rodríguez, G. (1999), *Unit Root, Additive Outliers and Cointegration Analysis with Macroeconomic Applications*, Unpublished PhD Dissertation, Université de Montréal.

[34] Saikkonen, P. and H. Lütkepohl (1997), "Trend Adjustment Prior to Testing for the Cointegrating Rank of a VAR Process," Discussion Paper 84, SFB 373, Humboldt-Universität zu Berlin.

[35] Saikkonen, P. and H. Lütkepohl (1998), "Testing for the Cointegrating Rank of a VAR process with an Intercept," Discussion Paper 51, SFB 373, Humboldt-Universität zu Berlin.

[36] Sargan, J. D. and A. Bhargava (1983), "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica* **51**, 153-174.

[37] Stock, J. H. (1987), "Asymptotic Properties of Least-Squares Estimators of Cointegrating Vectors," *Econometrica* **55**, 1035-1056.

[38] Stock, J. H. (1999): "A Class of Tests for Integration and Cointegration," in Engle, R.F. and H. White (eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, 137-167.

[39] Stock, J. H. and M. W. Watson (1988), "Testing for Common Trends," *Journal of the American Statistical Association* **83**, 1097-1107.

[40] Watson, M. W. (1994), "Vector Autoregressions and Cointegration,"in *Handbook of Econometrics*, vol. IV, Engle, R. F. and D. L. McFadden (eds.), chapter 47.

[41] Xiao, Z. and P. C. B. Phillips (1999), "Efficient Detrending in Cointegrating Regression," *Econometric Theory* **15**, 519-548.

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