2016-02-01

Residuals-based tests for cointegration with generalized least-squares detrended data

Pierre Perron, Gabriel Rodriguez. 2016. "Residuals-based tests for cointegration with generalized least-squares detrended data." ECONOMETRICS JOURNAL, Volume 19, Issue 1, pp. 84 - 111 (28).

https://hdl.handle.net/2144/26277

Residuals-based Tests for Cointegration with GLS Detrended Data*

Pierre Perron† Gabriel Rodríguez‡

Boston University Pontificia Universidad Católica del Perú

Revised: October 19, 2015

Abstract

We provide GLS–detrended versions of single-equation static regression or residuals-based tests for testing whether or not non-stationary time series are cointegrated. Our approach is to consider nearly optimal tests for unit roots and apply them in the cointegration context. We derive the local asymptotic power functions of all tests considered for a triangular DGP imposing a directional restriction such that the regressors are pure integrated processes. Our GLS versions of the tests do indeed provide substantial power improvements over their OLS counterparts. Simulations show that the gains in power are important and stable across various configurations.

JEL Classification Number: C22, C32, C52.

Keywords: Cointegration, Residuals-Based Unit Root Tests, ECR Tests, OLS and GLS Detrended Data, Hypothesis Testing.

*A preliminary version was presented at the European Society Econometric Meeting in Málaga, Spain, August 27-31, 2012. Rodríguez acknowledges financial support from the International Cooperation Office and the Department of Economics of the Pontificia Universidad Católica del Perú for a Visitor Scholarship to the Department of Economics at Boston University (January-April 2013). We also thank three referees and the co-editor Michael Jansson for useful comments. The comments of one referee were especially constructive and helpful.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

‡Address for Correspondence: Gabriel Rodríguez, Departamento de Economia, Pontificia Universidad Católica del Perú, Av. Universitaria 1801, Lima 32, Perú (gabriel.rodriguez@pucp.edu.pe).

1 Introduction

Even though they are applicable only under some specific conditions, single-equation static regression or residuals-based tests for cointegration, proposed by Engle and Granger (1987) and developed by Phillips and Ouliaris (1990), have been quite popular in applied work mostly because of their computational simplicity. The statistics 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 Ordinary Least Squares (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.

Another type of single-equation cointegration tests is based on estimates of a conditional error correction model (ECM); see, e.g., Kremers, Ericsson and Dolado (1992), Boswijk (1994), Banerjee, Dolado and Mestre (1998), and Zivot (2000). Pesavento (2004) analyzes the local asymptotic power function of various single-equation cointegration tests. Many alternative approaches are available, some applicable under less restrictive conditions; e.g., the system based tests of Johansen (1991) and Stock and Watson (1988).

Elliott, Rothenberg and Stock (1996, henceforth ERS), following Dufour and King (1991), showed that several unit root tests constructed using Generalized Least Squares (GLS) or quasi-differenced data[1] have asymptotic power functions close to the Gaussian local asymptotic power envelope and offer substantial power gains over tests constructed using OLS detrended data. It is natural to think that such a detrending device would also be beneficial for cointegration tests. Our aim, accordingly, is to analyze residuals-based tests for coin-

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

tegration when they are constructed using GLS detrended or quasi-differenced data. We consider the standard ADF and the $Z_{\hat{\rho}}$ 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 (2001).

Unfortunately, in the context of testing for cointegration, the problem of finding the Gaussian local asymptotic power envelope in the general case with deterministic components, as done by ERS for unit root tests, is intractable (at least to us) given the fact that the cointegrating vector is not identified under the null hypothesis. Elliott and Pesavento (2009) derived a class of tests that have some optimality property by specifying some weight function over the possible values of the cointegrating vector. However, this was done only for the case with no deterministic components, limiting its practical usefulness. It nevertheless seems sensible to argue that tests that perform best in the unit root case would also allow power gains in the cointegration context. This is the approach we follow in this paper. Hence, we consider nearly optimal tests for unit roots and apply them in the cointegration context. We cannot claim that any of the tests achieves the best power possible. Given the approach taken, we also cannot claim that the non-centrality parameters $\bar{c}$ suggested correspond exactly to the values that would yield a 50% local asymptotic power for the tests. While we agree that it would be of theoretical interest to be able to remedy these caveats, we believe that, for all practical purposes, it would make little differences. Our GLS versions of the tests do indeed provide substantial improvements over their OLS counterparts. The end-products do provide improvements over currently available methods and are therefore useful for empirical applications. Our approach is similar to that used by Boswijk, Jansson and Nielsen (2015) who considered improved full-system cointegration tests in VAR models by using a likelihood ratio test (based on the full likelihood) which was shown to have optimality properties in the unit root context (as in ERS), as show in Jansson and Nielsen (2012).

We derive the asymptotic distributions of all tests for a Data Generating Process (DGP) with a so-called "directional restriction" which imposes an exact unit root on the regressors. This is the same DGP adopted by, among others, Elliott, Jansson and Pesavento (2005) to analyze optimal tests with a known cointegrating vector, Pesavento (2004, 2007) to analyze the local asymptotic power functions of various tests for cointegration, and Zivot (2000) to analyze the power of single equation ECM tests for cointegration when the cointegrating vector is known. It permits non-weakly exogenous regressors. With such a DGP, the local asymptotic power functions depend on a nuisance parameter, R^2, which represents the long-run contribution of the m regressors in the cointegrating relation and the dependent variable

y_t, as well as m. Also, as will be shown in the simulations, the power functions of the residuals-based tests are not much influenced by nuisance parameters other than R^2 and m, even though they are not invariant to them. Using this DGP will allow obtaining the relevant local asymptotic power functions of the tests. We then calibrate the relevant non-centrality parameter to quasi-difference the data from a residuals-based test that corresponds to a nearly optimal unit root test. Even with this less than optimal approach to select the non-centrality parameter, all tests proposed will have much improved power over their OLS-based versions in all DGPs considered, unless the initial condition is very large in which case the OLS-based ADF test has higher power. The power functions of the ECM-based tests are very sensitive to various nuisance parameters, with power that can often be near the size. Hence, these will not be considered in this paper.

Our work is related to that of Lütkepohl and Saikkonen (2000), Saikkonen and Lütkepohl (2000a,b) 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). It is also related to Pesavento (2007) who analyzed the local asymptotic power functions of similar tests we suggested in an earlier version of this paper. She showed that our approach did lead to marked improvements in power. This paper offers improved tests with non-centrality parameters to quasi-difference the data that are better grounded in theory compared to what was available in the previous version.

This paper is organized as follows. Section 2 presents the data-generating process considered. Section 3 presents a motivation for the use of GLS detrended data and outlines our strategy for the construction of the tests and the selection of the non-centrality parameter to quasi-difference the data. Section 4 describes the various tests considered. Section 5 is concerned with the local asymptotic power functions of the tests, including the limit null distributions that allow tabulating critical values. Section 6 describes issues related to the implementation of the tests, in particular the evaluation of the relevant non-centrality parameter to quasi-difference the data. It also provides asymptotic critical values and a comparison of the local asymptotic power functions of the tests, both with OLS and GLS detrended data. Section 7 presents simulation results about the size and power of the tests in finite samples, with Section 7.1 concentrating on the effect of the initial condition. Section 8 offers brief concluding remarks and recommendations for practical implementations. An Appendix contains technical derivations.

2 The Data Generating Process

We consider the following Data Generating Process (DGP):

$$\mathbf{x}_t = \mathbf{d}_{xt} + \mathbf{u}_{xt}; \qquad \mathbf{u}_{xt} = \mathbf{u}_{xt-1} + \mathbf{v}_{1t} \tag{1}$$

$$y_t = d_{yt} + \boldsymbol{\beta}'\mathbf{x}_t + u_t; \qquad u_t = \rho u_{t-1} + v_{2t},$$

for $t = 1, ..., T$, where $\mathbf{x}_t$ is a $m \times 1$ vector, y_t is a scalar, $\mathbf{d}_{xt}$ and d_{yt} are the deterministic components: $\mathbf{d}_{xt} = \sum_{i=0}^{p_x} \boldsymbol{\psi}_{ix} t^i \equiv \boldsymbol{\psi}_x' \mathbf{m}_t^x$, $d_{yt} = \sum_{i=0}^{p_y} \psi_{iy} t^i \equiv \boldsymbol{\psi}_y' \mathbf{m}_t^y$, where $\mathbf{m}_t^x = (1, t, ..., t^{p_x})'$ and $\mathbf{m}_t^y = (1, t, ..., t^{p_y})'$. The vector $\mathbf{v}_t = (\mathbf{v}_{1t}', v_{2t})'$ contains (potentially) serially correlated errors with $\mathbf{v}_t = \boldsymbol{\Phi}(L)\boldsymbol{\epsilon}_t = \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{\epsilon}_{t-i}$ and $\sum_{i=0}^{\infty} i \det |\boldsymbol{\Phi}_i| < \infty$ ($\boldsymbol{\Phi}_0 = \mathbf{I}_n$) where $\boldsymbol{\epsilon}_t$ is a martingale difference sequence with respect to the sigma-field $\mathcal{F}_t = \sigma\text{-}field\ \{..., \boldsymbol{\epsilon}_{t-2}, \boldsymbol{\epsilon}_{t-1}\}$ with $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' | \mathcal{F}_{t-1}) = \boldsymbol{\Sigma}$. The total number of variables in the system is $n = m + 1$.

Assumption 1 $\{\boldsymbol{\epsilon}_t\}$ *satisfies a functional central limit theorem, that is* $T^{-1/2} \sum_{t=1}^{[Tr]} \boldsymbol{\epsilon}_t \Rightarrow \boldsymbol{\Sigma}^{1/2} \mathbf{W}(r)$ *where* $\mathbf{W}(r)$ *is a standard* $n \times 1$ vector of independent Wiener processes, $\Rightarrow$ *denotes weak convergence and* $\boldsymbol{\Sigma}$ *is the variance-covariance matrix of* $\boldsymbol{\epsilon}_t$.

Remark 1 *Note that we impose that* u_0 *and* u_{x0} *are* $O_p(1)$ *random variables. This is not innocuous, especially with regards to* u_0*, as the results are qualitatively different if this initial value is "large" in a well-defined sense. We shall assess the influence of the initial condition via simulations in Section 7.1.*

As a matter of notation, let $\mathbf{W} = (\mathbf{W}_1', W_2)'$ where $\mathbf{W}_1$ is an m-vector and W_2 is a scalar. Also, Assumption 1 along with the conditions on $\boldsymbol{\Phi}(L)$ imply that

$$T^{-1/2} \sum_{t=1}^{[Tr]} \mathbf{v}_t \Rightarrow \mathbf{B}(\mathbf{r}) = \boldsymbol{\Omega}^{1/2} \mathbf{W}(\mathbf{r}) \tag{2}$$

where $\boldsymbol{\Omega}$ is the spectral density at frequency zero of $\mathbf{v}_t$ scaled by 2π, that is,

$$\boldsymbol{\Omega} = \lim_{T \to \infty} T^{-1} E\{[\textstyle\sum_{t=1}^{T} \mathbf{v}_t][\sum_{t=1}^{T} \mathbf{v}_t']\} = 2\pi \mathbf{f}_{vv}(0) = \boldsymbol{\Phi}(1)\boldsymbol{\Sigma}\boldsymbol{\Phi}(1)'.$$

We define the following partition of the $n \times n$ matrix $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \omega_{22} \end{bmatrix},$$

and define $R^2 = \boldsymbol{\delta}'\boldsymbol{\delta}$, where $\boldsymbol{\delta} = \boldsymbol{\Omega}_{11}^{-1/2} \boldsymbol{\omega}_{12} \omega_{22}^{-1/2}$. The coefficient R^2 lies between zero and one and represents the long-run contribution of the right-hand side variables in the second equation of (1). It is zero when the variables $\mathbf{x}_t$ are not correlated in the long run with u_t. Furthermore, under the assumption that $\mathbf{x}_t$ are not individually cointegrated, $\mathbf{\Omega}_{11}$ is non singular, an assumption that we maintain throughout.

The null hypothesis is $\rho = 1$. Under the alternative hypothesis $|\rho| < 1$, the linear combination $y_t - d_{yt} - \boldsymbol{\beta}'\mathbf{x}_t$ is stationary, and y_t and $\mathbf{x}_t$ are cointegrated. Under the null hypothesis, the variables are not cointegrated and there are n unit roots in the system.

Some comments about the specifications of the deterministic components are in order. First, we restrict the analysis to $p_x = 0$ (non-trending data) and $p_x = 1$ (trending data) as these are of most interest in practice. When $p_x = 0$, we set $p_y = 0$ (only a constant in the cointegrating regression). When $p_x = 1$, we consider both cases with $p_y = 0$ and $p_y = 1$. In the terminology of Campbell and Perron (1991) and Perron and Campbell (1992, 1993), setting $p_y = 0$ amounts to testing for "deterministic cointegration", i.e., the cointegrating vector eliminates both the deterministic and stochastic non-stationarity. This is the case of most interest in practice. On the other hand, setting $p_y = 1$ amounts to testing for "stochastic cointegration" meaning that the cointegrating vector eliminates the non-stationarity due to the stochastic components only so that the cointegrating relationship is allowed to be a trend-stationary process. This case is less commonly encountered in practice but has nevertheless been useful in a variety of applications. Note finally that the cases with ($p_x = 0$, $p_y = 1$) and ($p_x = 1$, $p_y = 1$) are asymptotically the same since when a time trend is included as a regressor the tests are invariant to the value of the trend function of the regressors. Since the case ($p_x = 0$, $p_y = 1$) corresponds to one with an irrelevant trend regressor included, we shall not discuss it further. Note that when adopting the specification $p_x = 1$ and $p_y = 0$, we shall impose that $\boldsymbol{\psi}_{1x} \neq 0$. Hence, the tests will not be asymptotically similar since the limit distributions will be different whether $\boldsymbol{\psi}_{1x} \neq 0$ or $\boldsymbol{\psi}_{1x} = 0$. However, it will turn out that the critical values in the case with $\boldsymbol{\psi}_{1x} \neq 0$ are slightly more negative than the critical values that apply when specifying $p_x = p_y = 0$. Hence, though the tests with $p_x = 1$ and $p_y = 0$ will not be similar, they will have the correct asymptotic size (at least pointwise in the value of $\boldsymbol{\psi}_{1x}$) and be only slightly conservative when $\boldsymbol{\psi}_{1x} = 0$. One could have a similar test using $p_x = p_y = 1$ even if $\psi_{1y} = 0$. This would, however, imply a substantial loss in power. Hence, we shall continue recommending using the case $p_x = 1$ and $p_y = 0$ when $\boldsymbol{\psi}_{1x} \neq 0$ and $\psi_{1y} = 0$, i.e., trending regressors and testing for deterministic cointegration.

The DGP adopted is the same as in Elliott, Jansson and Pesavento (2005) to analyze optimal tests with a known cointegrating vector, Pesavento (2004, 2007) to analyze the local asymptotic power functions of various tests for cointegration, and Zivot (2000) to analyze the power of single equation ECM tests for cointegration when the cointegrating vector is known. It imposes what Elliott, Jansson and Pesavento (2005) refer to as a "directional restriction" so that the regressors $\mathbf{x}_t$ are imposed to have an exact unit root under both the null and alternative hypotheses, i.e, in both cases the limit of $T^{-1/2}\mathbf{x}_{[Tr]}$ is a Brownian motion. In general, endogensous regressors are allowed. The system contains weakly exogenous regressors when $\mathbf{\Phi}(1)$ is block lower triangular. Ideally, one would like to adopt a DGP that is the most general possible, not imposing the "directional restriction". The problem is that in this general case, the local asymptotic power functions of the tests are difficult to obtain and very complex functions of many nuisance parameters (e.g., Zivot, 2000, p. 427). On the other hand, with the "directional restriction", the local asymptotic power functions depend only on R^2 and m. Under the null hypothesis, the stated limit distributions will depend only upon m and will be the same with or without the "directional restriction" as shown in an earlier version of this paper. Accordingly, the strategy we adopt is to work with the DGP (1). This will allow obtaining the local asymptotic power functions from which we can select the relevant non-centrality parameter to quasi-difference the data. Even with this less than optimal approach, the GLS residuals-based tests will have much improved power over their OLS-based versions.

3 Motivation

Consider a special case of the DGP given by:

$$y_t = \boldsymbol{\psi}_y' \mathbf{m}_t^y + \boldsymbol{\beta}' \mathbf{x}_t + u_t, \tag{3}$$

where $u_t = \overline{\rho} u_{t-1} + \epsilon_t$, with $\overline{\rho} = 1 + \bar{c}/T$ and, for simplicity, $\epsilon_t \sim i.i.d.\ (0, \sigma_\epsilon^2)$. In this setup, y_t and $\mathbf{x}_t$ are cointegrated for any finite samples but not 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. Let $\mathbf{z}_t = (\mathbf{x}_t', y_t)'$ and $\mathbf{z}_t^{\overline{\rho}} = (1 - \overline{\rho} L)\mathbf{z}_t$, $\mathbf{m}_t^{y\overline{\rho}} = (1 - \overline{\rho} L)\mathbf{m}_t^y$. Ignoring the first observation, we can apply a GLS transformation and write (3) as

$$y_t^{\overline{\rho}} = \boldsymbol{\psi}_y' \mathbf{m}_t^{y\overline{\rho}} + \boldsymbol{\beta}' \mathbf{x}_t^{\overline{\rho}} + \epsilon_t. \tag{4}$$

Concentrating out $\boldsymbol{\psi}_y$, we can write (4) as

$$y_t^{\overline{\rho}} - \hat{\boldsymbol{\psi}}_y' \mathbf{m}_t^y = \boldsymbol{\beta}' (\mathbf{x}_t^{\overline{\rho}} - \hat{\boldsymbol{\psi}}_x^{*\prime} \mathbf{m}_t^y) + \epsilon_t \tag{5}$$

where $(\hat{\boldsymbol{\psi}}_y, \hat{\boldsymbol{\psi}}_x^*) = (\mathbf{M}_y^{\overline{\rho}\prime} \mathbf{M}_y^{\overline{\rho}})^{-1} \mathbf{M}_y^{\overline{\rho}\prime} \mathbf{Z}^{\overline{\rho}}$, with $\mathbf{Z}^{\overline{\rho}} = [\mathbf{z}_1^{\overline{\rho}}, ..., \mathbf{z}_T^{\overline{\rho}}]'$ and $\mathbf{M}_y^{\overline{\rho}} = [\mathbf{m}_1^{*y\overline{\rho}}, ..., \mathbf{m}_T^{*y\overline{\rho}}]'$. Let the GLS detrended data be defined by $y_t^d = y_t - \hat{\boldsymbol{\psi}}_y' \mathbf{m}_t^y$ and $\mathbf{x}_t^d = \mathbf{x}_t - \hat{\boldsymbol{\psi}}_x^{*\prime} \mathbf{m}_t^y$, then (5) can be written as

$$y_t^d = \boldsymbol{\beta}'\mathbf{x}_t^d + u_t^d \tag{6}$$

where

$$u_t^d = u_t - (\hat{\boldsymbol{\psi}}_y - \boldsymbol{\psi}_y)'\mathbf{m}_t^y + \boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_x^{*\prime}\mathbf{m}_t^y. \tag{7}$$

This implies that each series is "detrended" separately by an OLS regression on quasi-differenced data with differencing parameter $\overline{\rho}$. Accordingly, the method to construct the test statistic is one that provides, via the GLS transformation, a better method to concentrate out the parameter $\boldsymbol{\psi}_y$ of the cointegrating regression, and not better estimates of the coefficients of the trend function of the variables of the system, e.g., $\boldsymbol{\psi}_x$ for the regressors $\mathbf{x}_t$. For instance, when $p_y = 0$, the quasi-differenced operation only "quasi-demean" whether the data are trending or not. Xiao and Phillips (1999), in the context of tests within a full system with unit root processes, obtained simulation results showing that power was higher with the non-centrality parameter $\bar{c}$ set at -7.0 ($p_x = p_y = 0$) or -13.5 ($p_x = p_y = 1$) 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 that in the context of 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.

Following ERS, the next step would be to derive the Gaussian local power envelope of the likelihood ratio test for testing $H_0 : \bar{c} = 0$ versus the family of point alternatives $H_1 : \bar{c}$ some fixed negative value. Unfortunately, this problem is intractable (at least to us) given the fact that the cointegrating vector is not identified under the null hypothesis. It nevertheless seems sensible to argue that tests that perform best in the unit root case would also allow power gains in the cointegration context. This is the approach we follow. To motivate further, consider the case with a known cointegrating vector and $R^2 = 0$, then the tests of ERS and those of Ng and Perron (2001) have power functions nearly identical to the local Gaussian power envelope. As shown by Elliott et al. (2005), this is no longer the case when $R^2 > 0$ with a known cointegrating vector, though it make sense to argue that power gains are nevertheless possible even when $R^2 > 0$ and also in the case of an unknown cointegrating vector even though the tests cannot be claimed to be optimal. Hence, the approach taken is to consider nearly optimal tests for unit roots and apply them in the cointegration context. The tests are then based on the OLS residuals $\hat{u}_t$ using GLS detrended data,

$$\hat{u}_t = y_t^d - \hat{\boldsymbol{\beta}}'\mathbf{x}_t^d. \tag{8}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimate obtained from (6). This will lead to considerable power gains over existing residuals-based tests for cointegration (e.g., Phillips and Ouliaris, 1990).

4 The Tests

The residuals-based tests considered are constructed from $\hat{u}_t$, the residuals obtained from the static cointegration regression (8) with GLS detrended variables. Consider first the MP_T^{GLS} proposed by Ng and Perron (2001). In our case, it is defined by (for a sample of size $T+1$):

$$MP_{T,\mu}^{GLS} = \frac{\bar{c}^2 T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2 - \bar{c} T^{-1} \hat{u}_T^2}{s^2}, \quad MP_{T,\tau}^{GLS} = \frac{\bar{c}^2 T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2 + (1-\bar{c}) T^{-1} \hat{u}_T^2}{s^2},$$

for the demeaned (μ) and detrended (τ) cases, respectively. Another class of tests analyzed is the class of Z tests of Phillips and Perron (1988) in the context of testing for a unit root. These statistics can be applied to test the null hypothesis of no-cointegration as showed by Phillips and Ouliaris (1990). These are defined by

$$Z_{\hat{\rho}}^{GLS} = T(\hat{\rho}-1) - \frac{(s^2 - s_u^2)}{(2T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2)}, \quad Z_{t_{\hat{\rho}}}^{GLS} = \frac{s_u}{s} t_{\hat{\rho}} - \frac{(s^2 - s_u^2)}{(4s^2 T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2)^{1/2}},$$

where $\hat{\rho}$ is the OLS estimate in the regression $\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + \hat{\omega}_t$, and $t_{\hat{\rho}}$ is the corresponding t-statistic for testing $\rho = 1$, $s_u^2 = T^{-1} \sum_{t=1}^{T} \hat{\omega}_t^2$ and s^2 is described below. The class of M-tests, originally proposed by Stock (1999), and further analyzed by Perron and Ng (1996) and Ng and Perron (2001), 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. They are defined by:

$$MZ_{\hat{\rho}}^{GLS} = \frac{T^{-1}\hat{u}_T^2 - s^2}{2T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2}, \; MSB^{GLS} = [\frac{T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2}{s^2}]^{1/2}, \; MZ_{t_{\hat{\rho}}}^{GLS} = \frac{T^{-1}\hat{u}_T^2 - s^2}{[4s^2 T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^2]^{1/2}}.$$

The statistics are modified versions of the $Z_{\hat{\rho}}$ test, Bhargava's (1986) R_1 statistic, and the $Z_{t_{\hat{\rho}}}$ test. The term s^2 is an autoregressive estimate of (2π times) the spectral density at frequency zero of u_t^d, defined by $s^2 = s_{\eta k}^2 / [1 - \hat{b}(1)]^2$, 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[2]:

$$\Delta \hat{u}_t = \rho_0 \hat{u}_{t-1} + \sum_{j=1}^{k} b_j \Delta \hat{u}_{t-j} + \eta_{tk}. \tag{9}$$

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 $\rho_0 = 0$ in regression (9), denoted by ADF^{GLS}.

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

5 Asymptotic Distributions

In order to derive the limit distributions of the tests under the null hypothesis of no cointegration and their local asymptotic power functions, the DGP is assumed to be given by (1) with the local to unity specification $\rho = 1 + c/T$. We consider in turn the limit distribution of the estimate of the trend function under GLS detrending, the limit distribution of the estimate of the cointegrating vector and finally the limit distributions of the various tests [3].

5.1 Preliminaries

We start with a weak convergence result about the limit of $u_{[Tr]}$.

Lemma 1 *$T^{-1/2}u_{[Tr]} \Rightarrow \omega_{2.1}^{1/2} J_{2.1c}(r)$, where $\omega_{2.1} = \omega_{22} - \boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}$, $J_{2.1c}(r) = W_{2.1}(r) + c\int_0^r e^{(r-s)c} W_{2.1}(s)ds$ and $W_{2.1}(r)$ is a scalar Wiener process such that*

$$W_{2.1}(r) = \left[R^2/(1-R^2)\right]^{1/2} \overline{W}_1(r) + W_2(r)$$

with $\overline{W}_1(r) = m^{-1/2}\sum_{i=1}^m W_i(r)$.

The following Lemma gives the asymptotic properties of the estimates of the coefficients of the trend function using the GLS approach.

Lemma 2 *Let $\mathbf{z}_t = [\mathbf{x}_t', y_t]'$ be generated by (1). Each variable in the vector $\mathbf{z}_t$ is detrended separately using $\bar{\rho} = 1 + \bar{c}/T$. Then, 1) If $p_y = 0$, with $p_x = 0$ or $p_x = 1$: $T^{-1/2}vec(\hat{\boldsymbol{\psi}}_x^*) \Rightarrow \mathbf{0}_m$, an $m \times 1$ vector of zeros; also $T^{-1/2}(\hat{\psi}_y - \psi_y) \Rightarrow 0$. 2) If $p_y = 1$:*

$$\boldsymbol{\Upsilon}_T vec[\hat{\boldsymbol{\psi}}_x^* - \boldsymbol{\psi}_x] \Rightarrow \begin{bmatrix} \mathbf{0}_m \\ \lambda \mathbf{B}_1(\mathbf{1}) + 3(1-\lambda)\int_0^1 r\mathbf{B}_1(\mathbf{r}) \end{bmatrix} \equiv \begin{bmatrix} \mathbf{0}_m \\ \mathbf{H}_x \end{bmatrix}$$

where $\boldsymbol{\Upsilon}_T = [diag(T^{-1/2}, ..., T^{-1/2}), diag(T^{1/2}, ..., T^{1/2})]$, a $2m \times 2m$ matrix, $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$. Also, with $\boldsymbol{\Upsilon}_T = diag[T^{-1/2}, T^{1/2}]$, a diagonal 2×2 matrix,

$$\boldsymbol{\Upsilon}_T vec[\hat{\boldsymbol{\psi}}_y - \boldsymbol{\psi}_y] \Rightarrow \begin{bmatrix} 0 \\ \{\lambda J_{2.1c}(1) + 3(1-\lambda)\int_0^1 r J_{2.1c}(r)\} + \boldsymbol{\beta}'\mathbf{H}_x \end{bmatrix} \equiv \begin{bmatrix} 0 \\ H_y + \boldsymbol{\beta}'\mathbf{H}_x \end{bmatrix}$$

[3] Many of our proofs are similar to those of Pesavento (2004) who use OLS detrended data. However, her proofs are only valid with $m = 1$. The case with multiple regressors adds complications, which we address.

Lemma 2 will be useful in proving subsequent results. In particular, $T^{-1/2}vec(\hat{\boldsymbol{\psi}}_x^*) \Rightarrow \mathbf{0}_m$ and $T^{-1/2}(\hat{\psi}_y - \psi_y) \Rightarrow 0$ are the rates needed to have the limit distributions not influenced by the constant terms. The following Lemma provides the limit distribution of the errors from the regression involving GLS-detrended variables, i.e., u_t^d defined by (7).

Lemma 3 $T^{-1/2}u_{[Tr]}^d \Rightarrow \omega_{2.1}J_{2.1c}^d(r)$, *where; a) if* $p_y = 0$*:* $J_{2.1c}^d(r) = J_{2.1c}(r)$*; b) if* $p_y = 1$*:* $J_{2.1c}^d = J_{2.1c} - rH_y$*, as defined in Lemmas 1 and 2.*

The following Theorem gives the limit of the estimate $\hat{\boldsymbol{\beta}}$ obtained from applying OLS to (6) with GLS-detrended variables.

Theorem 1 *Suppose that* $\mathbf{z}_t = [\mathbf{x}_t', y_t]'$ *is generated by (1) with* $\rho = 1 + c/T$*. Let* y_t^d *and* $\mathbf{x}_t^d$ *be GLS detrended variables with non-centrality parameter* $\overline{\rho} = 1 + \overline{c}/T$*. Let* $\hat{\boldsymbol{\beta}}$ *be the OLS estimates of the cointegrating vector defined by (8). Then:*

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2}[\textstyle\int_0^1 \mathbf{W}_1^d\mathbf{W}_1^{d\prime}]^{-1}[\int_0^1 \mathbf{W}_1^d J_{2.1c}^d] \equiv \omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2}\boldsymbol{\kappa}_c^d,$$

where 1) If $p_x = 0$*,* $p_y = 0$*:* $\mathbf{W}_1^d = \mathbf{W}_1$ *and* $J_{2.1c}^d = J_{2.1c}$*; 2) If* $p_x = 1$*,* $p_y = 1$*:* $\mathbf{W}_1^d = \mathbf{W}_1 - [\lambda\mathbf{W}_1(1) + 3(1-\lambda)\int_0^1 s\mathbf{W}_1(s)ds]r$*,* $J_{2.1c}^d = J_{2.1c} - [\lambda J_{2.1c}(1) + 3(1-\lambda)\int_0^1 sJ_{2.1c}(s)ds]r$ *and* $\lambda = (1-\overline{c})/(1-\overline{c}+\overline{c}^2/3)$*; 3) If* $p_x = 1$*,* $p_y = 0$ *and* $\boldsymbol{\psi}_{1x} \neq 0$*:*

$$T^{-1/2}\mathbf{D}_T\mathbf{C}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \omega_{2.1}^{1/2}[\textstyle\int \mathbf{W}_{1R}^d\mathbf{W}_{1R}^{d\prime}]^{-1}[\int \mathbf{W}_{1R}^d J_{2.1c}] \equiv \omega_{2.1}^{1/2}\widetilde{\boldsymbol{\kappa}}_c^d$$

where $\mathbf{C} = [\ \mathbf{C}_1 \quad \mathbf{C}_2\ ] = [\ \boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\boldsymbol{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2} \quad \boldsymbol{\psi}_{1x}(\boldsymbol{\psi}_{1x}'\boldsymbol{\psi}_{1x})^{-1}\ ]$*, with* $\boldsymbol{\psi}_{1x}^{\perp}$ *a* $m \times (m-1)$ *matrix which spans the null space of* $\boldsymbol{\psi}_{1x}$*,* $\mathbf{W}_{1R}^d = [\mathbf{W}_{1(m-1)}', r]'$ *and* $\mathbf{D}_T = diag(T^{1/2}\mathbf{I}_{m-1}, T)$*.*

5.2 The Asymptotic Distributions of the Tests

We now consider the limit distributions of the tests. Before, we state a required Assumption. **Assumption 2**. *For regression (9),* $T^{-1/3}k \to 0$ *and* $k \to \infty$ as $T \to \infty$.

Theorem 2 *Let the data be generated by (1) with* $\rho = 1 + c/T$*, Assumptions 1-2 holding and the residuals* $\hat{u}_t$ *obtained from (8) with* $\overline{\rho} = 1 + \overline{c}/T$*. Then, as* $T \to \infty$*:*

$$MP_{T,\mu}^{GLS} \Rightarrow \frac{\overline{c}^2\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right] - \overline{c}\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(\mathbf{1})\boldsymbol{\eta}_c^d\right]}{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d\right]}, \quad MP_{T,\tau}^{GLS} \Rightarrow \frac{\overline{c}^2\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right] + (1-\overline{c})\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(\mathbf{1})\boldsymbol{\eta}_c^d\right]}{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d\right]},$$

$$Z_{t_{\hat{\rho}}}^{GLS}, ADF^{GLS} \Rightarrow c\frac{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right]^{1/2}}{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d\right]^{1/2}} + \frac{\left[\boldsymbol{\eta}_c^{d\prime}\int \mathbf{W}_c^d\mathbf{d}\widetilde{\mathbf{W}}\boldsymbol{\eta}_c^d\right]^{1/2}}{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right]^{1/2}\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d\right]^{1/2}},$$

$$Z_{\widehat{\rho}}^{GLS} \Rightarrow c+\frac{\left[\boldsymbol{\eta}_c^{d\prime}\int \mathbf{W}_c^d \mathbf{d}\widetilde{\mathbf{W}}\boldsymbol{\eta}_c^d\right]^{1/2}}{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right]^{1/2}}, \quad MSB^{GLS} \Rightarrow \frac{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right]^{1/2}}{\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d\right]^{1/2}},$$

$$MZ_{\widehat{\rho}}^{GLS} \Rightarrow \frac{\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(\mathbf{1})\boldsymbol{\eta}_c^d - \boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d}{2\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d}, \quad MZ_{t_{\widehat{\rho}}}^{GLS} \Rightarrow \frac{\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(\mathbf{1})\boldsymbol{\eta}_c^d - \boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d}{2\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d\right]^{1/2}\left[\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d\right]^{1/2}},$$

where $\boldsymbol{\eta}_c^{d\prime} = [-(\int \mathbf{W}_1^d J_{2.1c}^d)(\int \mathbf{W}_1^d \mathbf{W}_1^{d\prime})^{-1}, 1]$, $\mathbf{W}_c^d = \left[\mathbf{W}_1^{d\prime}, J_{2.1c}^d\right]$, $\mathbf{A}_c^d = \int_0^1 \mathbf{W}_c^d \mathbf{W}_c^{d\prime}$, $\widetilde{\mathbf{W}}' = [\mathbf{W}_1', W_{2.1}]'$, *and, with* $\overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}} = R^2/(1-R^2)$,

$$\mathbf{D} = \begin{bmatrix} \mathbf{I} & \overline{\boldsymbol{\delta}} \\ \overline{\boldsymbol{\delta}}' & 1+\overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}} \end{bmatrix}, \quad \mathbf{A}_c^d(\mathbf{1}) = \begin{bmatrix} \mathbf{W}_1^d(\mathbf{1})\mathbf{W}_1^d(\mathbf{1})' & \mathbf{W}_1^d(\mathbf{1})J_{2.1c}^{d\prime}(1) \\ J_{2.1c}^d(1)\mathbf{W}_1^d(\mathbf{1})' & (J_{2.1c}^d(1))^2 \end{bmatrix},$$

with $\mathbf{W}_1^d$ *and* $J_{2.1c}^d$ *as defined in Theorem 1 for* $p_x = 0$, $p_y = 0$ *and* $p_x = 1$, $p_y = 1$ *and when* $p_x = 1$, $p_y = 0$*:* $\mathbf{W}_1^d = \mathbf{W}_{1R}^d = [\mathbf{W}_{1(m-1)}', r]'$ *and* $J_{2.1c}^d = J_{2.1c}$.

6 Implementation of the Procedures and Power Results

The theoretical results obtained are used to address the following issues: 1) the choice of the relevant non-centrality parameter $\bar{c}$; 2) given a choice for $\bar{c}$, the derivation of the critical values; 3) an analysis of the local asymptotic power functions of the tests. The results are obtained from 10,000 replications using 1,000 steps to approximate the Wiener processes.

6.1 The Choice of $\bar{c}$ and the Asymptotic Critical Values

Of importance for the implementation of the tests is the choice of $\bar{c}$. We performed extensive simulations about the local asymptotic power functions of the tests for all cases with a wide range of values for R^2 and up to 5 right-hand-side regressors. In each case, we computed the non-centrality parameter $\bar{c}$ that corresponds to a local asymptotic power of 50%. In order to provide the non-centrality parameters used to construct all tests, we calibrated the results using the MP_T^{GLS} test for the three cases analyzed for the deterministic components. This is motivated in part by the fact that the MP_T^{GLS} test has a local asymptotic power function closest to the Gaussian local power envelope in the case of unit root tests. In any event, the differences in power are very minor across the residuals-based tests and, hence, similar results would apply had we settled on any other test in this class.

What transpired from these results are the following features. First, the local asymptotic power functions of the residuals-based tests are very similar implying an "optimal $\bar{c}$" that is nearly the same for a given number of $I(1)$ regressors m. Second, variations in R^2 do affect the local power function and the corresponding "optimal $\bar{c}$" but not as much as variations in m. In order not to overburden practitioners with endless tables of critical values, we opted to select a single representative value for R^2 and for each m compute the associated "optimal $\bar{c}$". Since extreme values of R^2 near 0 or 1 are unlikely in practice, it made sense to select a value in the mid-range. We settled upon using $R^2 = 0.4$ which appears to be a sensible value often considered in the literature; see Pesavento (2007) for a similar argument. The "optimal $\bar{c}$" does not vary much for small deviations, e.g., values of R^2 between .25 and .55. So it indeed appears to be a sensible choice. Since the power functions vary considerably with m, we report a separate "optimal $\bar{c}$" for $m = 1, ..., 5$, in Table 1.

As can be seen in Table 1, the "optimal $\bar{c}$" are very different from the unit root case. They also are larger in absolute value (more negative) with more deterministic components and with more regressors. Using these values of $\bar{c}$, we simulated the limit distributions of the tests from Theorem 2, under the null hypothesis that $c = 0$. Notice that under the null hypothesis none of the statistics depends on R^2 or any other nuisance parameters asymptotically. Tables 2, 3 and 4 present the asymptotic critical values for the three sets of possibilities for the deterministic components ($p_x = 0$, $p_y = 0$; $p_x = 1$, $p_y = 1$; and $p_x = 1$, $p_y = 0$). We present the critical values for tests with nominal sizes ranging from 1% to 20%. Note that the limit distributions of the GLS-based tests for the case ($p_x = 0$, $p_y = 0$) are the same as their OLS-based counterparts when no deterministic component is present in the data and the regression, e.g., Phillips and Ouliaris (1990), Table Ia for $Z_{\widehat{\rho}}^{GLS}$ and $MZ_{\widehat{\rho}}^{GLS}$ and Table IIa for $Z_{t_{\widehat{\rho}}}^{GLS}$, $MZ_{t_{\widehat{\rho}}}^{GLS}$, ADF^{GLS}. The limit distributions for $MP_{T,\mu}^{GLS}$ and MSB^{GLS} are, however, new. We present the critical values for all tests for completeness and ease of reference.

6.2 Asymptotic Power Functions

This section presents the local asymptotic power functions of the statistics proposed and compare them to those of tests constructed with OLS detrended data. We present results only for the case $p_y = p_x = 0$, that is, the demeaned case. The results for the other cases are qualitatively similar with power being lower overall when more deterministic components are included. We also consider results for $R^2 = 0, .2, .4$ and .8 and up to five right hand-side variables. For the OLS version of the tests, we used results in Pesavento (2004) and for the test MP_T we used unreported limit distributions that we derived.

Figures 1 shows the asymptotic power functions of some tests based on OLS and GLS detrending: MP_T^{OLS} and MP_T^{GLS} in Figure 1.a, ADF^{OLS} and ADF^{GLS} in Figure 1.b. It is clear that the power of a test that uses GLS-detrended data is higher than its OLS-based counterpart for all cases, although two features are important. First, the advantages of GLS-based tests are clear when $|c| \geq 10$. Second, for small R^2 (0.0, 0.2 in Figure 1) the advantages of GLS-based over OLS-based tests are small when there are 5 regressors. Also, as expected, power decreases when R^2 increases and when m increases. The increase in power when using GLS instead of OLS detrended data can be substantial. It also applies to all tests considered. We do not present detailed results to that effect as they are similar. What transpires is that the local asymptotic power functions are essentially the same unless the number of regressors is very large, in which case the ADF^{GLS} is slightly more powerful.

7 Size and Power of the Tests in Finite Samples

We now evaluate the size and power of the tests in finite samples. The data-generating process is a bivariate system ($m = 1$) given by $y_t = \beta x_t + u_t$ and $u_t = \rho u_{t-1} + v_{2t}$, with $x_t = x_{t-1} + v_{1t}$ (and $x_0 = 0$) where $\mathbf{v}_t = \mathbf{A}\mathbf{v}_{t-1} + \boldsymbol{\epsilon}_t$, $\boldsymbol{\epsilon}_t \sim i.i.d.\ N(0, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ chosen such that $\boldsymbol{\Omega} = (\mathbf{I}_2 - \mathbf{A})^{-1}\boldsymbol{\Sigma}(\mathbf{I}_2 - \mathbf{A})^{-1\prime}$, the long-run variance-covariance matrix of $\mathbf{v}_t$, is given by:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix} \text{ with } A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \text{ as a matter of notation.}$$

We set $\beta = 1$, $R^2 = 0.0, 0.4, 0.8$, $c = 0, -10, -20$, $T = 200$, 500 and use 5,000 replications. We present results only for the demeaned case ($p_x = p_y = 0$) since they are qualitatively similar for the other cases. We consider six configurations for the matrix $\mathbf{A}$:

$$\begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.0 \\ 0.2 & 0.3 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.0 \\ 0.2 & 0.7 \end{bmatrix}, \begin{bmatrix} 0.0 & 0.3 \\ 0.0 & 0.0 \end{bmatrix}, \begin{bmatrix} 0.0 & 0.7 \\ 0.0 & 0.0 \end{bmatrix}, \begin{bmatrix} 0.0 & 0.9 \\ 0.0 & 0.0 \end{bmatrix}$$

When the upper-right element of $\mathbf{A}$ is 0, x_t is weakly exogenous, as in the first three cases. The first case is the base case with no serial correlation. The next two cases imply weakly exogenous regressors and vary the extent of the serial correlation in v_{2t}, while keeping the serial correlation in v_{1t} fixed at 0.5. Cases 2 and 3 also introduce feedback from v_{1t} to v_{2t}. The last three cases pertain to regressors x_t that are not weakly exogenous. We present results for the tests MP_T, $Z_{\hat{\rho}}$, ADF using both OLS and GLS detrended data. They are presented in Table 5 for $T = 200$ and $T = 500$. The lag length was selected using the BIC with $k_{min} = 0$ and $k_{max} = (4 \times (T/100))^{1/4}$, which for $T = 200$ and $T = 500$ imply $k_{max} = 5$ and 6, respectively. The critical values for the OLS residuals-based tests are from Phillips and Ouliaris (1990) for the ADF and $Z_{\hat{\rho}}$ tests and those for the test MP_T were obtained from an unreported limit distribution that we derived. The nominal size of the tests is set at 5%. We start with the case with $u_0 = 0$ and consider large non-zero value afterwards.

Consider first the base case with $\mathbf{A}$ a matrix of $\mathbf{0}$, so that there is no serial correlation. Note first that the exact size of the tests is close to the nominal 5% level. The test MP_T is slightly conservative when $T = 200$ but the distortions disappear when $T = 500$. Consider now, more importantly, the power of the tests. The first feature to note is that for all tests and all parameter configurations, the GLS versions are more powerful than their OLS counterparts. The difference in power can indeed be quite substantial. For example, for the ADF test with $T = 200$ and $R^2 = 0$, the power of the OLS version is .243 and that of the GLS version is .462 when $c = -10$; when $c = -20$ the corresponding figures are .702 and .927. Hence, there are clear gains in using our GLS-based tests.

Consider now cases 2-3 involving serial correlation in v_{2t}. The same qualitative features remain albeit with a slight decrease in power. To gain some insight into the cause of these results, consider the case with $R^2 = 0$ and note that $(\rho - 1)(1 - a_{22})$ is one minus the sum of the autoregressive coefficients in the autoregressive representation of u_t, one may expect a reduction in power. This is not the case because with the long-run variance Ω held fixed throughout all experiments, σ_{22}, the variance of ϵ_{2t}, is proportional to $(1 - a_{22})^2$. Hence, the variance of v_{2t} is fixed. Accordingly, the increase in the sum of the autoregressive coefficients is compensated by a reduction in the variance of the noise so that the power of the residuals-based test is unchanged. Things are more complex when R^2 is non-zero but the insight remains the same.

Consider now the cases with regressors that are not weakly exogenous. Again, the power of the residuals-based tests are little influenced by the value of a_{12}. We conducted extensive additional simulations with a variety of other configurations. What transpires from the results is that the local asymptotic power functions provide a reliable guide about the relative power of the tests.

7.1 The effect of a large initial value u_0

In the context of testing for a unit root, Müller and Elliott (2003) have shown that the relative power of GLS-based versus OLS-based tests can be reversed when the initial condition is large, with the OLS-based tests having higher power. To address this issue in our cointegration setup, we use simulations with non-zero initial values. The specification adopted for

u_0 follows Harvey et al. (2009), namely

$$u_0 = \alpha/(1-\rho^2)^{1/2} \text{ where } \alpha \sim N(\mu_\alpha I(\sigma_\alpha^2 = 0), \sigma_\alpha^2),$$

with $I(\cdot)$ the indicator function and $\rho = 1 + c/T$. When $\sigma_\alpha^2 = 0$, the initial value is a non-random quantity so that $u_0 = \mu_\alpha/(1-\rho^2)^{1/2}$, referred to as the "fixed case". Here, the parameter μ_α dictates the magnitude of the initial condition. When $\sigma_\alpha^2 \neq 0$, the initial value is a random variable such that $u_0 = N(0, \sigma_\alpha^2/(1-\rho^2))$, referred to as the "random case" for which σ_α^2 dictates the magnitude of the initial condition. For each of the fixed and random cases, we set $c = -10$, -20 and $T = 200, 500$. As previously, we consider the tests MP_T, $Z_{\hat{\rho}}$, ADF using both OLS and GLS detrended data. We only considered the case $A = 0$, as the results are qualitatively similar for the other cases. The results are presented in Figure 2 for the "fixed case" and Figure 3 for the "random case".

The results indicate the following features. First, the power functions are very similar for the fixed and random cases. Second, the power functions of all tests are (nearly) monotonically decreasing as the magnitude of the initial condition increases except for the OLS-based ADF test. The latter has a power function that increases with the magnitude of the initial condition. In all cases, the cross-over point at which the OLS-based ADF becomes more powerful than a GLS-based test is (roughly) a value of μ_α in the fixed case or σ_α^2 in the random case of 1.5. Whether this implies a realistic value for the initial condition is uncertain (and readers can differ about this issue). The results imply that the OLS-based ADF can have higher power than the GLS-based tests only for very large initial values, i.e., when the system starts very far from equlibrium. Indeed, note that when $\sigma_\alpha^2 = 1.5$ for the random case, the initial value is a draw from a normal with mean 0 and variance which takes the following values: 37.5 when $T = 500$ and $c = -10$, 19.4 when $T = 500$ and $c = -20$, 15.5 when $T = 200$ and $c = -10$, 7.9 when $T = 200$ and $c = -20$. In the fixed case, the corresponding values for u_0 are 7.7, 5.4, 4.8 and 3.4. These large values are induced by the fact that the the initial condition is modelled to follow a draw from the unconditional distribution, which is unbounded as both T increases and ρ approaches 1. Under the null hypothesis it is not well defined. If one thinks such large values are practically relevant, then it would be feasible to extend the analysis of Harvey et al. (2009) to use both a GLS-based test and the OLS-based ADF test to have a hybrid procedure that has the correct size for all values of the initial condition. This is outside the scope of this paper. Nevertheless, as noted by Harvey et al. (2009), a simple union of rejection procedure (rejecting if either a GLS-based test or the OLS-based ADF test rejects) is likely to have mild liberal size distortions and could be used.

8 Conclusions

We analyzed residuals-based tests for cointegration with GLS-detrended data and derived their local asymptotic power functions using a DGP with a "directional restriction" assumption from which we calibrated parameters $\bar{c}$ to quasi-difference the data. The asymptotic distributions depend of the number of right-hand side variables, the type of deterministic components in the cointegration equation and present in the data, and R^2 which measures the long-run correlation between $\mathbf{x}_t$ and y_t. The theoretical results showed that important power gains can be achieved using GLS detrended data. Simulations have shown that these gains are indeed important in finite samples and robust across a wide variety of DGPs, unless the initial condition is very large, in which case the OLS-based ADF test is more powerful.

Appendix

Proof of Lemma 1: Assumption 1 along with the conditions on $\mathbf{\Phi}(L)$ implies that

$$T^{-1/2}\sum_{t=1}^{[Tr]}\mathbf{v}_t \Rightarrow \mathbf{B}(r) = \mathbf{\Omega}^{1/2}\mathbf{W}(r)$$

for $r \in [0,1]$, where

$$\mathbf{\Omega}^{1/2} = \begin{bmatrix} \mathbf{\Omega}_{11}^{1/2} & 0 \\ \boldsymbol{\omega}_{21}\mathbf{\Omega}_{11}^{1/2} & \omega_{2.1}^{1/2} \end{bmatrix},$$

$\omega_{2.1} = \omega_{22} - \boldsymbol{\omega}_{21}\mathbf{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}$, and $\mathbf{W} = [\mathbf{W}_1', W_2]'$. Using this notation, we have $T^{-1/2}\sum_{t=1}^{[Tr]}\mathbf{v}_{1t} \Rightarrow \mathbf{\Omega}_{11}^{1/2}\mathbf{W}_1(r)$ and $T^{-1/2}\sum_{t=1}^{[Tr]} v_{2t} \Rightarrow \boldsymbol{\omega}_{21}\mathbf{\Omega}_{11}^{-1/2}\mathbf{W}_1 + \omega_{2.1}^{1/2}W_2$. Define $\overline{\boldsymbol{\delta}}' = \omega_{2.1}^{-1/2}\boldsymbol{\omega}_{21}\mathbf{\Omega}_{11}^{-1/2}$ so that $\overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}} = R^2/(1-R^2)$. Therefore $\boldsymbol{\omega}_{21}\mathbf{\Omega}_{11}^{-1/2}\mathbf{W}_1 = \omega_{2.1}^{1/2}\overline{\boldsymbol{\delta}}'\mathbf{W}_1$. Using Lemma 5.6 of Park and Phillips (1988), $T^{-1/2}\sum_{t=1}^{[Tr]} v_{2t}$ may be written in such way that it depends on only one nuisance parameter, namely R^2. Because $\overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}} = R^2/(1-R^2)$, $W_{2.1} = [R^2/(1-R^2)]^{1/2}\overline{W}_1(r) + W_2(r)$ where $\overline{W}_1(r) = m^{-1/2}\sum_{i=1}^{m} W_i(r)$. The result follows since u_t is a near-integrated process in terms of the errors v_{2t}.

Proof of Lemma 2: Consider first the cases $p_y = p_x = 0$ or $p_y = p_x = 1$. The results for $\hat{\boldsymbol{\psi}}_x^*$ follow from arguments as in ERS using the fact that the noise component of $\mathbf{x}_t$ is a vector of integrated processes and the limit result (2). For $\hat{\boldsymbol{\psi}}_y$, note that $y_t = d_{yt} + \boldsymbol{\beta}'\mathbf{x}_t + u_t$ and

$$(\mathbf{M}_y^{\bar{\rho}\prime}\mathbf{M}_y^{\bar{\rho}})^{-1}\mathbf{M}_y^{\bar{\rho}\prime}\mathbf{Y}^{\bar{\rho}} = (\mathbf{M}_y^{\bar{\rho}\prime}\mathbf{M}_y^{\bar{\rho}})^{-1}\mathbf{M}_y^{\bar{\rho}\prime}(\mathbf{U}^{*\bar{\rho}} + \mathbf{X}^{\bar{\rho}}\boldsymbol{\beta})$$

where $\mathbf{Y}^{\bar{\rho}} = [y_1^{\bar{\rho}}, ..., y_T^{\bar{\rho}}]'$, $\mathbf{U}^{*\bar{\rho}} = [(d_{y1}+u_1)^{\bar{\rho}}, ..., (d_{yT}+u_T)^{\bar{\rho}}]'$ and $\mathbf{X}^{\bar{\rho}} = [\mathbf{x}_1^{\bar{\rho}}, ..., \mathbf{x}_T^{\bar{\rho}}]'$. The results follows using arguments as in ERS and Lemma 1 for the first term, and the results for $\hat{\boldsymbol{\psi}}_x^*$ for the second term. It remains to consider the case $p_y = 0$ and $p_x = 1$. Without loss of generality suppose $\mathbf{x}_t$ is a scalar ($m = 1$), given by

$$x_t = \psi_{0x} + \psi_{1x}t + u_{xt} = m_{1t}^x + m_{2t}^x + u_{xt}.$$

Then,

$$\begin{aligned} \hat{\psi}_{0x}^* &= (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}x^{\bar{\rho}}) \\ &= (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}m_1^{y\bar{\rho}\prime}\{\psi_{0x}m_1^{x\bar{\rho}} + \psi_{1x}m_2^{x\bar{\rho}} + u_x^{\bar{\rho}}\} \\ (\hat{\psi}_{0x}^* - \psi_{0x}) &= \psi_{1x}(m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}m_2^{x\bar{\rho}}) + (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}u_x^{\bar{\rho}}). \end{aligned}$$

From straightforward derivations, we have: $(m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}m_2^{x\bar{\rho}}) \Rightarrow (1 - \bar{c} + \bar{c}^2/2)$ and $(m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}u_x^{\bar{\rho}}) \Rightarrow u_{x1}$. Therefore, $(\hat{\psi}_{0x}^* - \psi_{0x}) \Rightarrow \psi_{1x}(1 - \bar{c} + \bar{c}^2/2) + u_{x1}$. Hence, $T^{-1/2}\hat{\psi}_{0x}^* \Rightarrow 0$. Now,

$$\begin{aligned} \hat{\psi}_{0y} &= (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}y_t^{\bar{\rho}}) \\ &= (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}m_1^{y\bar{\rho}\prime}\{\psi_{0y}m_1^{y\bar{\rho}} + \beta x^{\bar{\rho}} + u^{\bar{\rho}}\} \\ (\hat{\psi}_{0y} - \psi_{0y}) &= (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}\beta x^{\bar{\rho}}) + (m_1^{y\bar{\rho}\prime}m_1^{y\bar{\rho}})^{-1}(m_1^{y\bar{\rho}\prime}u^{\bar{\rho}}). \end{aligned}$$

From straightforward derivations, we have: $(m_1^{y\bar{\rho}\prime} m_1^{y\bar{\rho}}) \Rightarrow 1$, $(m_1^{y\bar{\rho}\prime} \beta x^{\bar{\rho}}) \Rightarrow u_{x1}$ and $(m_1^{y\bar{\rho}\prime} u^{\bar{\rho}}) \Rightarrow u_1$. Hence, $(\hat{\psi}_{0y} - \psi_{0y}) \Rightarrow \beta u_{x1} + u_1$ and $T^{-1/2}(\hat{\psi}_{0y} - \psi_{0y}) \Rightarrow 0$.

Proof of Lemma 3: We first consider the limit of $\mathbf{x}_t^d$. When $p_y = p_x = 0$, we have:

$$\begin{aligned} \mathbf{x}_t^d &= \mathbf{x}_t - \hat{\boldsymbol{\psi}}_{0x}^* = \mathbf{u}_{xt} - (\hat{\boldsymbol{\psi}}_{0x}^* - \boldsymbol{\psi}_{0x}) \\ T^{-1/2}\mathbf{x}_t^d &= T^{-1/2}\mathbf{u}_{xt} - T^{-1/2}(\hat{\boldsymbol{\psi}}_{0x}^* - \boldsymbol{\psi}_{0x}) \end{aligned}$$

Using (2) and Lemma 2:

$$T^{-1/2}\mathbf{x}_{[Tr]}^d \Rightarrow \mathbf{B_1}(\mathbf{r}) \equiv \boldsymbol{\Omega}_{11}^{1/2}\mathbf{W_1}(\mathbf{r}). \tag{A.1}$$

When $p_x = p_y = 1$, we have:

$$\begin{aligned} \mathbf{x}_t^d &= \mathbf{x}_t - \hat{\boldsymbol{\psi}}_{0x}^* - \hat{\boldsymbol{\psi}}_{1x}^* t = \mathbf{u}_{xt} - (\hat{\boldsymbol{\psi}}_{0x}^* - \boldsymbol{\psi}_{0x}) - (\hat{\boldsymbol{\psi}}_{1x}^* - \boldsymbol{\psi}_{1x})t \\ T^{-1/2}\mathbf{x}_t^d &= T^{-1/2}\mathbf{u}_{xt} - T^{-1/2}(\hat{\boldsymbol{\psi}}_{0x}^* - \boldsymbol{\psi}_{0x}) - T^{1/2}(\hat{\boldsymbol{\psi}}_{1x}^* - \boldsymbol{\psi}_{1x})(t/T). \end{aligned}$$

Using Lemma 2: $T^{-1/2}\mathbf{x}_{[Tr]}^d \Rightarrow \mathbf{B_1}(\mathbf{r}) - r\mathbf{H}_x \equiv \mathbf{B_1^d}(\mathbf{r}) = \boldsymbol{\Omega}_{11}^{1/2}\mathbf{W_1^d}(\mathbf{r})$. When $p_x = 1$ and $p_y = 0$,

$$\mathbf{x}_t^d = \mathbf{x}_t - \hat{\boldsymbol{\psi}}_{0x}^* = \mathbf{u}_{xt} + \boldsymbol{\psi}_{1x}t - (\hat{\boldsymbol{\psi}}_{0x}^* - \boldsymbol{\psi}_{0x})$$

and $T^{-1/2}\mathbf{x}_{[Tr]}^d \Rightarrow \mathbf{B_1}(\mathbf{r}) - r\boldsymbol{\psi}_{1x}$. We now consider the limit of u_t^d. From (7), we have:

$$u_t^d = u_t - (\hat{\boldsymbol{\psi}}_y - \boldsymbol{\psi}_y)'\mathbf{m}_t^y + \boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_x^{*\prime}\mathbf{m}_{xt}^y. \tag{A.2}$$

Consider first the case with $p_y = 0$ and $p_x = 0$ or $p_x = 1$. Then:

$$\begin{aligned} u_t^d &= u_t - (\hat{\psi}_{0y} - \psi_{0y}) + \boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_{0x}^{*\prime} \\ T^{-1/2}u_t^d &= T^{-1/2}u_t - T^{-1/2}(\hat{\psi}_{0y} - \psi_{0y}) + T^{-1/2}\boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_{0x}^{*\prime}. \end{aligned}$$

Using Lemmas 1 and 2: $T^{-1/2}u_{[Tr]}^d \Rightarrow \omega_{2.1}^{1/2}J_{2.1c}(r)$. Consider now the case with $p_y = 1$, then (A.2) is

$$u_t^d = u_t - (\hat{\psi}_{0y} - \psi_{0y}) - (\hat{\psi}_{1y} - \psi_{1y})t + \boldsymbol{\beta}'[\hat{\boldsymbol{\psi}}_{0x}^{*\prime} + \hat{\boldsymbol{\psi}}_{1x}^{*\prime}t]$$

and

$$\begin{aligned} T^{-1/2}u_t^d &= T^{-1/2}u_t - T^{-1/2}(\hat{\psi}_{0y} - \psi_{0y}) - T^{1/2}(\hat{\psi}_{1y} - \psi_{1y})(t/T) \\ &\quad + T^{-1/2}\boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_{0x}^{*\prime} + T^{1/2}\boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_{1x}^{*\prime}(t/T). \end{aligned}$$

Since the estimates of the residuals and the test statistics are invariant to the value $\boldsymbol{\psi}_{1x}$, we can without loss of generality set $\boldsymbol{\psi}_{1x} = \mathbf{0}$. Then, using Lemmas 1 and 2, we have:

$$T^{-1/2}u_{[Tr]}^d \Rightarrow B_{2c}(r) - r(H_y - \boldsymbol{\beta}'\mathbf{H}_x) + r\boldsymbol{\beta}'\mathbf{H}_x = B_{2c}(r) - rH_y = \omega_{2.1}^{1/2}J_{2.1c}^d(r).$$

Lemma A.1. When the model is generated according to (1) with $T(\rho-1)=c$, then, as $T\to\infty$, we have: 1) $T^{-2}\sum \mathbf{x}_t^d\mathbf{x}_t^{d\prime} \Rightarrow \mathbf{\Omega}_{11}^{1/2}\int_0^1 \mathbf{W}_1^d\mathbf{W}_1^{d\prime}\mathbf{\Omega}_{11}^{1/2}$; 2) $T^{-2}\sum \mathbf{x}_t^d u_t^d \Rightarrow \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{1/2}\int_0^1 \mathbf{W}_1^d J_{2.1c}^d$; 3) $T^{-2}\sum (u_t^d)^2 \Rightarrow \omega_{2.1}\int_0^1 (J_{2.1c}^d)^2$; 4) $T^{-1}\sum u_{t-1}^d\boldsymbol{\epsilon}_t' \Rightarrow \mathbf{\Omega}_{11}^{1/2}\int_0^1 J_{2.1c}^d d\mathbf{W}'\mathbf{\Sigma}^{1/2}$; 5) $T^{-1}\sum \mathbf{x}_{t-1}^d\boldsymbol{\epsilon}_t' \Rightarrow \mathbf{\Omega}_{11}^{1/2}\int_0^1 \mathbf{W}_1^d d\mathbf{W}'\mathbf{\Sigma}^{1/2}$; 6) $T^{-1}(u_T^d)^2 \Rightarrow \omega_{2.1}J_{2.1c}^d(1)^2$; 7) $T^{-1}\mathbf{x}_T^d\mathbf{x}_T^{d\prime} \Rightarrow \mathbf{\Omega}_{11}^{1/2}\mathbf{W}_1^d(1)\mathbf{W}_1^d(1)'\mathbf{\Omega}_{11}^{1/2}$.

Proof of Lemma A.1. The results follow straightfowardly from Lemmas 1-3, the continuous mapping theorem (CMT) and results from Chang and Wei (1988) and Phillips (1987), with appropriate modifications for the fact that the variables are quasi-differenced.

Proof of Theorem 1. The cointegration vector $\hat{\boldsymbol{\beta}}$ is estimated by regressing y_t^d on $\mathbf{x}_t^d$. In the first case ($p_x=0$, $p_y=0$), both variables are demeaned and for the second case ($p_x=1$, $p_y=1$) both variables are linearly detrended. In general, for both cases, $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = [T^{-2}\sum_{t=1}^T \mathbf{x}_t^d\mathbf{x}_t^{d\prime}]^{-1}[T^{-2}\sum_{t=1}^T \mathbf{x}_t^d u_t^d]$. Using Lemma A.1 and the CMT, we have

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \Rightarrow \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}\left[\int_0^1 \mathbf{W}_1^d\mathbf{W}_1^{d\prime}\right]^{-1}\left[\int_0^1 \mathbf{W}_1^d J_{2.1c}^d\right] \equiv \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}\boldsymbol{\kappa}_c^d$$

with all terms as defined in the text. In the case where $p_x=1$, $p_y=0$ we follow Hansen (1992). First, note that $\mathbf{x}_t = \boldsymbol{\psi}_{0x}+\boldsymbol{\psi}_{1x}t+\mathbf{u}_{xt}$ and we can deduce that $\mathbf{x}_t^d = \mathbf{x}_t - \hat{\boldsymbol{\psi}}_{0x}^*$, with $\hat{\boldsymbol{\psi}}_{0x}^* \Rightarrow \boldsymbol{\psi}_{1x}(1-\bar{c}+\bar{c}^2/2)+\mathbf{u}_{x1}$. Note that a similar result holds for y_t. Let $\boldsymbol{\psi}_{1x}^{\perp}$ be a $m\times(m-1)$ matrix which spans the null space of $\boldsymbol{\psi}_{1x}$ and let

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\mathbf{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2} & \boldsymbol{\psi}_{1x}(\boldsymbol{\psi}_{1x}'\boldsymbol{\psi}_{1x})^{-1} \end{bmatrix}.$$

Given that $(\boldsymbol{\psi}_{1x}^{\perp\prime}\mathbf{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})$ is positive definite (because $\mathbf{\Omega}_{11}$ is positive definite), $\mathbf{C}$ is well-defined. Therefore, we have:

$$\begin{aligned}
\mathbf{C}'\mathbf{x}_t^d &= \begin{bmatrix} \mathbf{C}_1'\mathbf{x}_t^d \\ \mathbf{C}_2'\mathbf{x}_t^d \end{bmatrix} = \begin{bmatrix} (\boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\mathbf{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2})'\mathbf{x}_t^d \\ (\boldsymbol{\psi}_{1x}(\boldsymbol{\psi}_{1x}'\boldsymbol{\psi}_{1x})^{-1})'\mathbf{x}_t^d \end{bmatrix} \\
&= \begin{bmatrix} (\boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\mathbf{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2})'\{\boldsymbol{\psi}_{0x}+\boldsymbol{\psi}_{1x}t+\mathbf{u}_{xt}-\hat{\boldsymbol{\psi}}_{0x}^*\} \\ (\boldsymbol{\psi}_{1x}(\boldsymbol{\psi}_{1x}'\boldsymbol{\psi}_{1x})^{-1})'\{\boldsymbol{\psi}_{0x}+\boldsymbol{\psi}_{1x}t+\mathbf{u}_{xt}-\hat{\boldsymbol{\psi}}_{0x}^*\} \end{bmatrix} \\
&= \begin{bmatrix} (\boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\mathbf{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2})'\{(\boldsymbol{\psi}_{0x}-\hat{\boldsymbol{\psi}}_{0x}^*)+\boldsymbol{\psi}_{1x}t+\mathbf{u}_{xt}\} \\ (\boldsymbol{\psi}_{1x}(\boldsymbol{\psi}_{1x}'\boldsymbol{\psi}_{1x})^{-1})'\{(\boldsymbol{\psi}_{0x}-\hat{\boldsymbol{\psi}}_{0x}^*)+\boldsymbol{\psi}_{1x}t+\mathbf{u}_{xt}\} \end{bmatrix}.
\end{aligned}$$

Defining the weight matrix $\mathbf{D}_T = diag(\mathbf{I}_{m-1}T^{1/2}, T)$, we have using Lemma 2 and the fact that $\boldsymbol{\psi}_{1x}^{\perp}$ and $\boldsymbol{\psi}_{1x}$ are orthogonals:

$$\mathbf{D}_T^{-1}\mathbf{C}'\mathbf{x}_{[Tr]}^d = \begin{bmatrix} T^{-1/2}(\boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\mathbf{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2})'\{\mathbf{u}_{x[Tr]}+o_p(1)\} \\ T^{-1}(\boldsymbol{\psi}_{1x}(\boldsymbol{\psi}_{1x}'\boldsymbol{\psi}_{1x})^{-1})'\{\mathbf{u}_{x[Tr]}+\boldsymbol{\psi}_{1x}t+o_p(1)\} \end{bmatrix} \Rightarrow \begin{bmatrix} \mathbf{W}_{1(m-1)}^d \\ r \end{bmatrix} \equiv \mathbf{W}_{1R}^d.$$

In the last step, we used the fact that $\mathbf{W}_{1(m-1)} = \mathbf{C}_1'\mathbf{B}_1(\mathbf{r}) = \boldsymbol{\psi}_{1x}^{\perp}(\boldsymbol{\psi}_{1x}^{\perp\prime}\boldsymbol{\Omega}_{11}\boldsymbol{\psi}_{1x}^{\perp})^{-1/2\prime}\mathbf{B}_1(\mathbf{r})$, see Hansen (1992, p. 93, eq. (12)). We have $u_t^d = -(\hat{\psi}_{0y} - \psi_{0y}) + \boldsymbol{\beta}'\hat{\boldsymbol{\psi}}_{0x}^* + u_t$ and $T^{-1/2}u_{[Tr]}^d \Rightarrow J_{2.1c}(r)$. Also, $T^{-1}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T \mathbf{x}_t^d\mathbf{x}_t^{d\prime}\mathbf{C}\mathbf{D}_T^{-1} \Rightarrow \int \mathbf{W}_{1R}^d\mathbf{W}_{1R}^{d\prime}$ and $T^{-3/2}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T \mathbf{x}_t^d\mathbf{u}_t^d \Rightarrow \omega_{2.1}^{1/2}\int \mathbf{W}_{1R}^d J_{2.1c}$ so that with $\mathbf{W}_{1R}^d = [\mathbf{W}_{1(m-1)}', r]'$:

$$
\begin{aligned}
T^{-1/2}\mathbf{D}_T\mathbf{C}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left[T^{-1}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T \mathbf{x}_t^d\mathbf{x}_t^{d\prime}\mathbf{C}\mathbf{D}_T^{-1}\right]^{-1}\left[T^{-3/2}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T \mathbf{x}_t^d u_t^d\right] \\
&\Rightarrow \omega_{2.1}^{1/2}\left[\int \mathbf{W}_{1R}^d\mathbf{W}_{1R}^{d\prime}\right]^{-1}\left[\int \mathbf{W}_{1R}^d J_{2.1c}\right] \equiv \omega_{2.1}^{1/2}\widetilde{\boldsymbol{\kappa}}_c^d.
\end{aligned}
$$

Before proving Theorem 2, we introduce some auxiliary results.

Lemma A.2. Let the data be generated according to (1) with $\rho = 1 + c/T$ and Assumptions 1-2 holding. Let $\hat{u}_t$ be the residuals from the cointegration regression (8) estimated using GLS detrended variables with a non-centrality parameter $\overline{\rho} = 1 + \overline{c}/T$, and s^2 be the estimate of the long-run variance. Then, as $T \to \infty$: 1) $T^{-2}\sum_{t=1}^T \hat{u}_t^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d$; 2) $T^{-1}\hat{u}_T^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(\mathbf{1})\boldsymbol{\eta}_c^d$; 3) $T^{-1}\sum_{t=1}^T \hat{u}_t\Delta\hat{u}_t \Rightarrow \omega_{2.1}[c\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d + \boldsymbol{\eta}_c^{d\prime}\int_0^1 \mathbf{W}_c^d d\widetilde{\mathbf{W}}'\boldsymbol{\eta}_c^d]$; 4) $s^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d$, where $\mathbf{W}_c^d = \left[\mathbf{W}_1^{d\prime}, J_{2.1c}^d\right]'$,

$$
\begin{aligned}
\boldsymbol{\eta}_c^{d\prime} &= \begin{bmatrix} -\left(\int_0^1 \mathbf{W}_1^d J_{2.1c}^d\right)\left(\int_0^1 \mathbf{W}_1^d\mathbf{W}_1^{d\prime}\right)^{-1} & 1 \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\kappa}_c^{d\prime} & 1 \end{bmatrix}, \\
\mathbf{A}_c^d &= \int_0^1 \mathbf{W}_c^d\mathbf{W}_c^{d\prime} = \begin{bmatrix} \int_0^1 \mathbf{W}_1^d\mathbf{W}_1^{d\prime} & \int_0^1 \mathbf{W}_1^d J_{2.1c}^{d\prime} \\ \int_0^1 \mathbf{W}_1^d J_{2.1c}^{d\prime} & \int_0^1 (J_{2.1c}^d)^2 \end{bmatrix}
\end{aligned}
$$

$$
\mathbf{A}_c^d(\mathbf{1}) = \begin{bmatrix} \mathbf{W}_1^d(\mathbf{1})\mathbf{W}_1^d(\mathbf{1})' & \mathbf{W}_1^d(\mathbf{1})J_{2.1c}^{d\prime}(1) \\ J_{2.1c}^d(1)\mathbf{W}_1^d(\mathbf{1})' & (J_{2.1c}^d(1))^2 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{I} & \overline{\boldsymbol{\delta}} \\ \overline{\boldsymbol{\delta}}' & 1 + \overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}} \end{bmatrix}
$$

$\widetilde{\mathbf{W}}' = [\mathbf{W}_1', W_{2.1}]'$, $W_{2.1} = [R^2/(1-R^2)]^{1/2}\overline{W}_1(r) + W_2$, $J_{2.1c}(r)$ is an Ornstein-Uhlenbeck process such that $J_{2.1c}(r) = W_{2.1}(r) + c\int_0^s e^{(s-r)c}W_{2.1}(r)dr$.

Proof of Lemma A.2. Note first that $\hat{u}_t = u_t^d - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{x}_t^d$. For part (1) and the cases with $p_x = p_y = 0$ and $p_x = p_y = 1$:

$$
\begin{aligned}
T^{-2}\sum_{t=1}^T \hat{u}_t^2 &= \begin{bmatrix} -(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' & 1 \end{bmatrix}\begin{bmatrix} T^{-2}\sum \mathbf{x}_t^d\mathbf{x}_t^{d\prime} & T^{-2}\sum \mathbf{x}_t^d u_t^{d\prime} \\ T^{-2}\sum u_t^d\mathbf{x}_t^{d\prime} & T^{-2}\sum u_t^d u_t^{d\prime} \end{bmatrix}\begin{bmatrix} -(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ 1 \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} -\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2} & 1 \end{bmatrix}
\end{aligned}
$$

$$
\times \begin{bmatrix} \mathbf{\Omega}_{11}^{1/2}\int_0^1 \mathbf{W}_1^d \mathbf{W}_1^{d\prime} \mathbf{\Omega}_{11}^{1/2} & \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{1/2}\int_0^1 \mathbf{W}_1^d J_{2.1c}^{d\prime} \\ \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{1/2}\int_0^1 J_{2.1c}^d \mathbf{W}_1^{d\prime} & \omega_{2.1}\int_0^1 (J_{2.1c}^d)^2 \end{bmatrix} \begin{bmatrix} -\omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}\boldsymbol{\kappa}_c^d \\ 1 \end{bmatrix}
$$

$$
\equiv \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d
$$

where $\boldsymbol{\eta}_c^{d\prime} = [\ -\boldsymbol{\kappa}_c^{d\prime} \quad 1\ ]$, and the other terms are as defined in the text and in Lemma A.2. For the case where $p_x = 1$ and $p_y = 0$, we have

$$
\begin{aligned}
& T^{-2}\sum_{t=1}^T \hat{u}_t^2 \\
= & \begin{bmatrix} -T^{-1/2}\mathbf{D}_T\mathbf{C}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' & 1 \end{bmatrix} \times \\
& \begin{bmatrix} T^{-1}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T \mathbf{x}_t^d\mathbf{x}_t^{d\prime}\mathbf{C}\mathbf{D}_T^{-1} & T^{-3/2}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T \mathbf{x}_t^d u_t^{d\prime} \\ T^{-3/2}\mathbf{D}_T^{-1}\mathbf{C}'\sum_{t=1}^T u_t^d\mathbf{x}_t^{d\prime} & T^{-2}\sum_{t=1}^T (u_t^d)^2 \end{bmatrix} \begin{bmatrix} -T^{-1/2}\mathbf{D}_T\mathbf{C}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ 1 \end{bmatrix} \\
\Rightarrow & \begin{bmatrix} -\widetilde{\boldsymbol{\kappa}}_c^{d\prime} & 1 \end{bmatrix} \begin{bmatrix} \int \mathbf{W}_1^d\mathbf{W}_1^{d\prime} & \int \mathbf{W}_1^d J_{2.1c}^{d\prime} \\ \int J_{2.1c}^d\mathbf{W}_1^{d\prime} & \int (J_{2.1c}^d)^2 \end{bmatrix} \begin{bmatrix} -\widetilde{\boldsymbol{\kappa}}_c^d \\ 1 \end{bmatrix} \equiv \omega_{2.1}^{1/2}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d
\end{aligned}
$$

where $\mathbf{W}_1^d = \mathbf{W}_{1R}^d = [\mathbf{W}_{1(m-1)}', r]'$ and $J_{2.1c}^d = J_{2.1c}$. For part (2),

$$
\begin{aligned}
T^{-1}\hat{u}_T^2 &= \begin{bmatrix} -(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' & 1 \end{bmatrix} \begin{bmatrix} T^{-1}\mathbf{x}_T^d\mathbf{x}_T^{d\prime} & T^{-1}\mathbf{x}_T^d u_T^{d\prime} \\ T^{-1}u_T^d\mathbf{x}_T^{d\prime} & T^{-1}u_T^d u_T^{d\prime} \end{bmatrix} \begin{bmatrix} -(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ 1 \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} -\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{\Omega}_{11}^{1/2}\mathbf{W}_1^d(\mathbf{1})\mathbf{W}_1^{d\prime}(\mathbf{1})\mathbf{\Omega}_{11}^{1/2} & \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{1/2}\mathbf{W}_1^d(\mathbf{1})J_{2.1c}^{d\prime}(1) \\ \omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{1/2}J_{2.1c}^d(1)\mathbf{W}_1^{d\prime}(\mathbf{1}) & \omega_{2.1}(J_{2.1c}^d(1))^2 \end{bmatrix} \\
&\quad \times \begin{bmatrix} -\omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}\boldsymbol{\kappa}_c^d \\ 1 \end{bmatrix} \equiv \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(\mathbf{1})\boldsymbol{\eta}_c^d
\end{aligned}
$$

where $\boldsymbol{\eta}_c^{d\prime} = [-\boldsymbol{\kappa}_c^{d\prime}, 1]$, and the other terms as defined in the text and in Lemma A.2. For the case where $p_x = 1$ and $p_y = 0$, the result follows using similar modifications as in part (1) with the limit result defined with $\mathbf{W}_1^d = \mathbf{W}_{1R}^d = [\mathbf{W}_{1(m-1)}', r]'$ and $J_{2.1c}^d = J_{2.1c}$.

For part (3), the proof follows using arguments in Phillips and Ouliaris (1990), see also Corollary A.1 of Pesavento (2004). For part (4), following the arguments of Phillips and Ouliaris (1990), s^2 is a zero frequency spectral density estimate based on $\Delta\hat{u}_t = [-(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})', 1]'\Delta\mathbf{z}_t^d = [-(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})', 1]'\mathbf{v}_t$. Since $[-(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})', 1]' \Rightarrow [-\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}, 1]'$, we can condition on the value of $[-\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}, 1]'$. Given the assumptions on $\mathbf{v}_t$, $[-\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\mathbf{\Omega}_{11}^{-1/2}, 1]'\mathbf{v}_t$ is 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 $[-\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2}, 1]'\mathbf{v}_t$ given by:

$$\begin{aligned} & \left[\begin{array}{cc} -\boldsymbol{\kappa}_c^{d\prime}\omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2} & 1 \end{array}\right]\left[\begin{array}{cc} \boldsymbol{\Omega}_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \omega_{22} \end{array}\right]\left[\begin{array}{c} -\omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2}\boldsymbol{\kappa}_c^{d} \\ 1 \end{array}\right] \\ = & (-\omega_{2.1}^{1/2}\boldsymbol{\kappa}_c^{d\prime}\boldsymbol{\Omega}_{11}^{1/2}+\boldsymbol{\omega}_{21})(-\omega_{2.1}^{1/2}\boldsymbol{\Omega}_{11}^{-1/2}\boldsymbol{\kappa}_c^{d})+(-\omega_{2.1}^{1/2}\boldsymbol{\kappa}_c^{d\prime}\boldsymbol{\Omega}_{11}^{-1/2}\boldsymbol{\omega}_{12}+\omega_{22}) \\ = & \omega_{2.1}\{\boldsymbol{\kappa}_c^{d\prime}\boldsymbol{\kappa}_c^{d}-\overline{\boldsymbol{\delta}}'\boldsymbol{\kappa}_c^{d\prime}-\boldsymbol{\kappa}_c^{d}\overline{\boldsymbol{\delta}}+\omega_{2.1}^{-1}\omega_{22}\}, \end{aligned}$$

after some algebra. Note that

$$\begin{aligned} \omega_{2.1}^{-1}\omega_{22} &= \omega_{22}[\omega_{22}-\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}]^{-1}=\omega_{22}\{\omega_{22}[1-\frac{\boldsymbol{\omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\omega}_{12}}{\omega_{22}}]\}^{-1} \\ &= \omega_{22}\{\omega_{22}[1-R^2]\}^{-1}=\omega_{22}[\frac{1}{\omega_{22}(1-R^2)}]=\frac{1}{1-R^2}=1+\overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}}. \end{aligned}$$

Therefore, collecting terms, we have:

$$s^2 \Rightarrow \omega_{2.1}\left[\begin{array}{cc} \boldsymbol{\kappa}_c^{d\prime} & 1 \end{array}\right]\left[\begin{array}{cc} \mathbf{I}_m & \overline{\boldsymbol{\delta}} \\ \overline{\boldsymbol{\delta}}' & 1+\overline{\boldsymbol{\delta}}'\overline{\boldsymbol{\delta}} \end{array}\right]\left[\begin{array}{c} -\boldsymbol{\kappa}_c^{d} \\ 1 \end{array}\right] \equiv \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^{d}.$$

For the case ($p_x = 1$, $p_y = 0$) the arguments are similar with appropriate modifications.

Proof of Theorem 2. Most proofs follow from Lemma A.2 and results in Phillips and Ouliaris (1990) and Pesavento (2004). We provide a brief outline. We have

$$MP_{T,\mu}^{GLS}=\frac{\overline{c}^2T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^2-\overline{c}T^{-1}\hat{u}_T^2}{s^2}.$$

From Lemma A.2, $\overline{c}^2T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^2 \Rightarrow \overline{c}^2\omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d$, $\overline{c}T^{-1}\hat{u}_T^2 \Rightarrow \overline{c}\omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(1)\boldsymbol{\eta}_c^d$ and $s^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d$, from which we readily obtain the limit distribution of $MP_{T,\mu}^{GLS}$. The limit distribution of $MP_{T,\tau}^{GLS}$ is obtained similarly. For $MZ_{\widehat{\rho}}^{GLS}=[T^{-1}\hat{u}_T^2-s^2]/[2T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^2]$, the limit distribution is obtained using the facts that $2T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^2 \Rightarrow 2\omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d$, $T^{-1}\hat{u}_T^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d(1)\boldsymbol{\eta}_c^d$ and $s^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d$. Similarly, the limit distribution of $MSB^{GLS} = [T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^2/s^2]^{1/2}$, follows using the facts that $T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{A}_c^d\boldsymbol{\eta}_c^d$ and $s^2 \Rightarrow \omega_{2.1}\boldsymbol{\eta}_c^{d\prime}\mathbf{D}\boldsymbol{\eta}_c^d$. The result for $MZ_{t_{\widehat{\rho}}}^{GLS}=MSB^{GLS}\times MZ_{\widehat{\rho}}^{GLS}$ then follows automatically. In the cases of the statistics ADF^{GLS}, $Z_{\widehat{\rho}}^{GLS}$ and $Z_{t_{\widehat{\rho}}}^{GLS}$, the results follow using arguments similar to those of Phillips and Ouliaris (1990) and Corollary A.1 of Pesavento (2004). Note that for the case $p_x = 1$ and $p_y = 0$, similar results are obtained using the appropriate modifications.

References

Banerjee, A., J.J. Dolado and R. Mestre (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis* 19, 267-283.

Berk, K.N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489-502.

Bhargava, A. (1986). On the theory of testing for unit root in observed time series. *Review of Economic Studies* 53, 369-384.

Boswijk, H.P. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics* 63, 37-60.

Boswijk, H.P., M. Jansson and M.O. Nielsen (2015). Improved likelihood ratio tests for cointegration rank in the VAR model. *Journal of Econometrics* 184, 97-110.

Campbell, J.Y. and P. Perron (1991). Pitfalls and opportunities: what macroeconomists should know about unit roots. In: O. Blanchard and S. Fischer (Eds.), *NBER Macroeconomics Annual.* MIT Press, Cambridge, MA, 141-201.

Chan, N.H. and C.Z. Wei (1988). Limiting distributions of least-squares estimates of unstable autoregressive process. *Annals of Statistics* 16, 367-401.

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.

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.

Elliott, G., M. Jansson and E. Pesavento (2005). Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. *Journal of Business & Economic Statistics* 23, 34-48.

Elliott, G., and E. Pesavento (2009). Testing the null of no cointegration when covariates are known to have a unit root. *Econometric Theory* 25, 1829-1850.

Elliott, G., T. Rothenberg and J.H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813-836.

Engle, R.F. and C.W.J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.

Granger, C.W.J. and P. Newbold (1974). Spurious regression in econometrics. *Journal of Econometrics* 2, 111-120.

Hansen, B.E. (1992). Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of Econometrics* 53, 87-121.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric Theory* 25, 587-636.

Jansson, M. and M.O. Nielsen (2012). Nearly efficient likelihood ratio tests of the unit root hypothesis. *Econometrica* 80, 2321-2332.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.

Kremers, J.J., N.R. Ericsson and J.J. Dolado (1992) The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* 54, 325-348.

Lütkepohl, H. and P. Saikkonen (2000). Testing for the cointegrating rank of a VAR process with a time trend. *Journal of Econometrics* 95, 177-198.

Müller U.K. and G. Elliott (2003). Tests for unit root and the initial condition. *Econometrica* 71, 1269-1286.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519-1554.

Park, J.Y. and P.C.B. Phillips (1988). Statistical inference in regressions with integrated processes: part 1. *Econometric Theory* 4, 468-497.

Pesavento, E. (2004). Analytical evaluation of the power of tests for absence of cointegration. *Journal of Econometrics* 122, 349-384.

Pesavento, E. (2007). Residuals-based tests for the null of no-cointegration: an analytical comparison. *Journal of Time Series Analysis* 28, 111-137.

Perron, P. and J.Y. Campbell (1992). Racines unitaires en macroéconomie: le cas multidimensionnel. *Annales D'Économie et de Statistique* 27, 1-50.

Perron, P. and J.Y. Campbell (1993). A note on Johansen's cointegration procedure when trends are present. *Empirical Economics* 18, 777-789

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.

Perron, P. and S. Ng (1998). An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* 14, 560-603.

Phillips, P.C.B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311-340.

Phillips, P.C.B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535-547.

Phillips, P.C.B. and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165-193.

Phillips, P.C.B. and P. Perron (1988). Testing for unit roots in time series regression. *Biometrika* 75, 335-346.

Said, S.E. and D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599-608.

Saikkonen, P. and H. Lütkepohl (2000a). Trend adjustment prior to testing for the cointegrating rank of a VAR process. *Journal of Time Series Analysis* 21, 435-456.

Saikkonen, P. and H. Lütkepohl (2000b). Testing for the cointegrating rank of a VAR process with an intercept. *Econometric Theory* 16, 373-406.

Stock, J.H. (1987). Asymptotic properties of least-squares estimators of cointegrating vectors. *Econometrica* 55, 1035-1056.

Stock, J.H. (1999). A class of tests for integration and cointegration. In Engle, R.F., & H. White (eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*. Oxford: Oxford University Press, 137-167.

Stock, J.H. and M.W. Watson (1988). Testing for common trends. *Journal of the American Statistical Association* 83, 1097-1107.

Xiao, Z. and P.C.B. Phillips (1999). Efficient detrending in cointegrating regression. *Econometric Theory* 15, 519-548.

Zivot, E. (2000). The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory* 16, 407-439.

Table 1. "Optimal Value of $\overline{c}$" obtained from the MP_T^{GLS} test with R^2= 0.4.

m	p_x= 0, p_y= 0	p_x= 1, p_y= 1	p_x= 1, p_y= 0
$x = 1$	-13.75	-20.50	-13.50
$x = 2$	-18.25	-23.75	-18.00
$x = 3$	-22.25	-27.25	-23.00
$x = 4$	-26.25	-30.75	-26.00
$x = 5$	-30.00	-33.75	-29.75

Table 2. Critical Values for Single-Equation Cointegration Tests with GLS Detrended Data
(Demeaned Case: $p_x = 0$, $p_y = 0$)

	MP_T					$Z_{\widehat{\rho}}$, $MZ_{\widehat{\rho}}$				
	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
1.0%	4.275	5.712	6.896	7.905	9.372	-23.633	-30.602	-37.266	-44.944	-49.568
2.5%	5.193	6.667	7.980	9.032	10.473	-19.143	-26.010	-32.252	-39.392	-44.141
5.0%	6.230	7.825	9.086	10.361	11.559	-15.984	-22.064	-28.164	-34.392	-40.040
7.5%	7.025	8.591	9.916	11.256	12.369	-14.169	-20.075	-25.798	-31.695	-37.197
10.0%	7.757	9.315	10.618	11.979	13.117	-12.708	-18.491	-24.113	-29.586	-35.224
15.0%	9.071	10.555	11.798	13.151	14.369	-10.857	-16.282	-21.632	-26.931	-32.047
20.0%	10.294	11.670	12.836	14.174	15.437	-9.466	-14.672	-19.889	-24.907	-29.850
	MSB					Z_t, MZ_t, ADF				
1.0%	0.144	0.126	0.115	0.105	0.100	-3.353	-3.849	-4.258	-4.641	-4.913
2.5%	0.159	0.137	0.123	0.112	0.105	-3.028	-3.531	-3.936	-4.345	-4.615
5.0%	0.172	0.148	0.131	0.119	0.111	-2.764	-3.279	-3.687	-4.055	-4.384
7.5%	0.182	0.155	0.137	0.124	0.115	-2.588	-3.104	-3.520	-3.898	-4.238
10.0%	0.191	0.160	0.141	0.128	0.118	-2.452	-2.975	-3.400	-3.783	-4.098
15.0%	0.206	0.171	0.149	0.134	0.123	-2.256	-2.780	-3.22	-3.598	-3.917
20.0%	0.219	0.179	0.155	0.139	0.128	-2.096	-2.630	-3.080	-3.453	-3.766

Table 3. Critical Values for Single-Equation Cointegration Tests with GLS Detrended Data
(Detrended Case: $p_x = 1$, $p_y = 1$)

	MP_T					$Z_{\widehat{\rho}}$, $MZ_{\widehat{\rho}}$				
	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
1.0%	7.014	7.638	8.778	9.588	10.592	-31.041	-38.102	-43.493	-50.662	-54.794
2.5%	8.166	8.824	9.890	10.906	11.759	-26.416	-33.099	-38.416	-44.482	-49.406
5.0%	9.242	10.121	11.160	12.156	12.944	-23.256	-28.474	-34.073	-39.851	-44.954
7.5%	10.243	11.075	12.083	13.079	13.868	-21.078	-26.111	-31.371	-36.811	-42.207
10.0%	11.093	11.940	12.905	13.861	14.523	-19.449	-24.336	-29.498	-34.822	-40.054
15.0%	12.660	13.204	14.175	15.180	15.723	-17.041	-21.863	-26.814	-31.724	-36.871
20.0%	13.929	14.372	15.370	16.134	16.833	-15.398	-20.065	-24.667	-29.506	-34.496
	MSB					Z_t, MZ_t, ADF				
1.0%	0.126	0.114	0.107	0.099	0.095	-3.913	-4.294	-4.627	-4.923	-5.179
2.5%	0.135	0.122	0.113	0.105	0.100	-3.635	-4.007	-4.340	-4.677	-4.910
5.0%	0.145	0.131	0.120	0.111	0.105	-3.401	-3.746	-4.064	-4.401	-4.668
7.5%	0.152	0.136	0.125	0.115	0.108	-3.229	-3.581	-3.907	-4.219	-4.525
10.0%	0.158	0.141	0.129	0.119	0.111	-3.085	-3.454	-3.787	-4.102	-4.402
15.0%	0.168	0.149	0.135	0.124	0.115	-2.879	-3.254	-3.606	-3.919	-4.222
20.0%	0.177	0.155	0.140	0.129	0.119	-2.721	-3.111	-3.455	-3.778	-4.069

Table 4. Critical Values for Single-Equation Cointegration Tests with GLS Detrended Data
(Case p_x= 1, p_y= 0)

	MP_T					$Z_{\widehat{\rho}}$, $MZ_{\widehat{\rho}}$				
	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
1.0%	4.015	5.499	7.244	7.854	9.169	-24.396	-30.815	-37.994	-44.330	-49.780
2.5%	4.874	6.422	8.235	9.090	10.305	-19.680	-26.300	-33.224	-38.407	-44.220
5.0%	5.837	7.522	9.424	10.248	11.443	-16.412	-22.477	-29.131	-34.099	-39.768
7.5%	6.640	8.272	10.236	11.068	12.242	-14.429	-20.374	-26.791	-31.597	-37.200
10.0%	7.240	8.936	10.995	11.797	12.958	-13.273	-18.861	-24.878	-29.542	-35.225
15.0%	8.334	10.138	2.359	12.957	14.177	-11.452	-16.578	-22.080	-26.840	-32.109
20.0%	9.367	11.217	13.642	14.025	15.292	-10.102	-14.971	-19.930	-24.885	-29.816
	MSB					Z_t, MZ_t, ADF				
1.0%	0.142	0.125	0.113	0.105	0.099	-3.427	-3.888	-4.307	-4.667	-4.914
2.5%	0.156	0.136	0.121	0.113	0.105	-3.112	-3.572	-4.020	-4.310	-4.630
5.0%	0.170	0.146	0.129	0.119	0.111	-2.833	-3.314	-3.762	-4.066	-4.387
7.5%	0.180	0.153	0.135	0.124	0.114	-2.665	-3.147	-3.582	-3.912	-4.229
10.0%	0.188	0.159	0.140	0.128	0.118	-2.540	-3.027	-3.465	-3.785	-4.122
15.0%	0.201	0.169	0.148	0.134	0.123	-2.349	-2.828	-3.269	-3.609	-3.924
20.0%	0.213	0.177	0.155	0.139	0.128	-2.201	-2.677	-3.116	-3.457	-3.766

Table 5. Small Sample Size and Power of Cointegration Tests; Demeaned Case

		$A=\begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix}$						$A=\begin{bmatrix} 0.5 & 0.0 \\ 0.2 & 0.3 \end{bmatrix}$						$A=\begin{bmatrix} 0.5 & 0.0 \\ 0.2 & 0.7 \end{bmatrix}$					
$-c$		0		10		20		0		10		20		0		10		20	
		OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
T=200																			
0.0	MP_T	0.037	0.049	0.269	0.394	0.750	0.891	0.055	0.067	0.340	0.446	0.754	0.871	0.057	0.070	0.285	0.382	0.627	0.768
	Z_ρ	0.049	0.057	0.279	0.419	0.759	0.911	0.061	0.067	0.319	0.451	0.738	0.877	0.062	0.070	0.264	0.377	0.606	0.763
	ADF	0.056	0.065	0.243	0.462	0.702	0.927	0.058	0.064	0.235	0.440	0.621	0.879	0.055	0.061	0.177	0.351	0.461	0.745
0.4	MP_T	0.037	0.049	0.194	0.305	0.643	0.834	0.056	0.066	0.231	0.326	0.615	0.783	0.081	0.100	0.254	0.340	0.548	0.702
	Z_ρ	0.049	0.057	0.200	0.329	0.651	0.860	0.061	0.068	0.216	0.328	0.587	0.786	0.085	0.102	0.239	0.336	0.523	0.697
	ADF	0.056	0.065	0.167	0.359	0.594	0.879	0.056	0.063	0.146	0.322	0.462	0.778	0.070	0.091	0.158	0.309	0.390	0.671
0.8	MP_T	0.037	0.049	0.102	0.188	0.478	0.740	0.056	0.066	0.126	0.216	0.484	0.699	0.200	0.218	0.179	0.266	0.456	0.645
	Z_ρ	0.049	0.057	0.109	0.207	0.483	0.773	0.060	0.069	0.118	0.217	0.454	0.699	0.207	0.222	0.172	0.273	0.432	0.643
	ADF	0.056	0.065	0.085	0.229	0.422	0.805	0.054	0.061	0.069	0.205	0.329	0.687	0.175	0.216	0.103	0.234	0.294	0.595
T=500																			
0.0	MP_T	0.040	0.050	0.284	0.354	0.776	0.884	0.049	0.053	0.322	0.388	0.792	0.893	0.047	0.054	0.256	0.316	0.682	0.799
	Z_ρ	0.049	0.050	0.268	0.362	0.757	0.894	0.054	0.057	0.297	0.384	0.765	0.896	0.050	0.052	0.235	0.309	0.642	0.794
	ADF	0.054	0.055	0.212	0.380	0.660	0.903	0.055	0.053	0.217	0.387	0.654	0.895	0.051	0.052	0.167	0.309	0.512	0.790
0.4	MP_T	0.040	0.050	0.207	0.294	0.675	0.833	0.046	0.051	0.210	0.288	0.650	0.802	0.081	0.079	0.235	0.310	0.631	0.787
	Z_ρ	0.049	0.050	0.200	0.302	0.654	0.845	0.050	0.050	0.194	0.285	0.617	0.804	0.084	0.080	0.221	0.307	0.597	0.783
	ADF	0.054	0.055	0.152	0.322	0.559	0.856	0.052	0.051	0.138	0.284	0.468	0.803	0.077	0.078	0.154	0.301	0.471	0.772
0.8	MP_T	0.040	0.050	0.113	0.201	0.519	0.760	0.045	0.050	0.115	0.199	0.501	0.732	0.233	0.218	0.164	0.263	0.527	0.744
	Z_ρ	0.049	0.050	0.109	0.208	0.502	0.776	0.049	0.048	0.104	0.202	0.470	0.729	0.244	0.224	0.154	0.266	0.499	0.742
	ADF	0.054	0.055	0.081	0.221	0.404	0.791	0.051	0.050	0.071	0.198	0.394	0.728	0.215	0.218	0.103	0.249	0.366	0.725

Table 5 (continued). Small Sample Size and Power of Cointegration Tests; Demeaned Case

		$A=\begin{bmatrix} 0.0 & 0.3 \\ 0.0 & 0.0 \end{bmatrix}$						$A=\begin{bmatrix} 0.0 & 0.7 \\ 0.0 & 0.0 \end{bmatrix}$						$A=\begin{bmatrix} 0.0 & 0.9 \\ 0.0 & 0.0 \end{bmatrix}$					
$-c$		0		10		20		0		10		20		0		10		20	
		OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
									T=200										
0.0	MP_T	0.050	0.062	0.280	0.400	0.735	0.881	0.066	0.078	0.303	0.413	0.723	0.867	0.068	0.082	0.311	0.419	0.726	0.860
	Z_ρ	0.061	0.070	0.287	0.428	0.745	0.902	0.084	0.091	0.302	0.436	0.736	0.889	0.090	0.096	0.320	0.440	0.739	0.884
	ADF	0.066	0.075	0.245	0.458	0.691	0.921	0.079	0.091	0.254	0.462	0.674	0.907	0.077	0.091	0.262	0.463	0.671	0.898
0.4	MP_T	0.079	0.090	0.250	0.383	0.706	0.875	0.095	0.107	0.306	0.423	0.747	0.885	0.105	0.114	0.310	0.433	0.751	0.885
	Z_ρ	0.099	0.102	0.270	0.411	0.721	0.899	0.140	0.137	0.328	0.464	0.765	0.908	0.169	0.161	0.341	0.479	0.772	0.906
	ADF	0.098	0.107	0.223	0.444	0.659	0.912	0.109	0.121	0.270	0.482	0.703	0.917	0.121	0.135	0.279	0.489	0.705	0.913
0.8	MP_T	0.096	0.105	0.161	0.283	0.601	0.825	0.165	0.170	0.211	0.343	0.687	0.860	0.200	0.200	0.242	0.369	0.702	0.866
	Z_ρ	0.130	0.128	0.178	0.312	0.621	0.855	0.298	0.275	0.258	0.393	0.719	0.894	0.396	0.362	0.282	0.432	0.730	0.900
	ADF	0.109	0.123	0.145	0.334	0.547	0.872	0.191	0.204	0.188	0.393	0.632	0.898	0.229	0.240	0.207	0.420	0.647	0.903
									T=500										
0.0	MP_T	0.051	0.057	0.296	0.361	0.768	0.885	0.060	0.061	0.313	0.390	0.773	0.884	0.060	0.065	0.324	0.396	0.776	0.884
	Z_ρ	0.059	0.059	0.283	0.368	0.755	0.893	0.072	0.066	0.301	0.395	0.752	0.890	0.070	0.073	0.225	0.408	0.760	0.892
	ADF	0.060	0.059	0.225	0.384	0.668	0.905	0.068	0.065	0.234	0.406	0.659	0.900	0.062	0.070	0.242	0.415	0.663	0.901
0.4	MP_T	0.061	0.062	0.260	0.345	0.727	0.868	0.081	0.084	0.272	0.371	0.735	0.877	0.089	0.086	0.282	0.372	0.742	0.879
	Z_ρ	0.074	0.067	0.249	0.355	0.711	0.878	0.108	0.095	0.276	0.383	0.728	0.889	0.125	0.112	0.288	0.390	0.739	0.891
	ADF	0.069	0.068	0.193	0.270	0.621	0.888	0.089	0.089	0.206	0.390	0.628	0.894	0.090	0.094	0.211	0.391	0.629	0.895
0.8	MP_T	0.077	0.074	0.152	0.265	0.613	0.819	0.114	0.110	0.177	0.286	0.634	0.836	0.138	0.126	0.183	0.288	0.650	0.834
	Z_ρ	0.095	0.082	0.154	0.276	0.601	0.835	0.199	0.163	0.192	0.310	0.635	0.851	0.290	0.231	0.206	0.327	0.653	0.855
	ADF	0.084	0.079	0.112	0.283	0.501	0.843	0.126	0.119	0.124	0.302	0.517	0.853	0.151	0.144	0.134	0.303	0.536	0.856

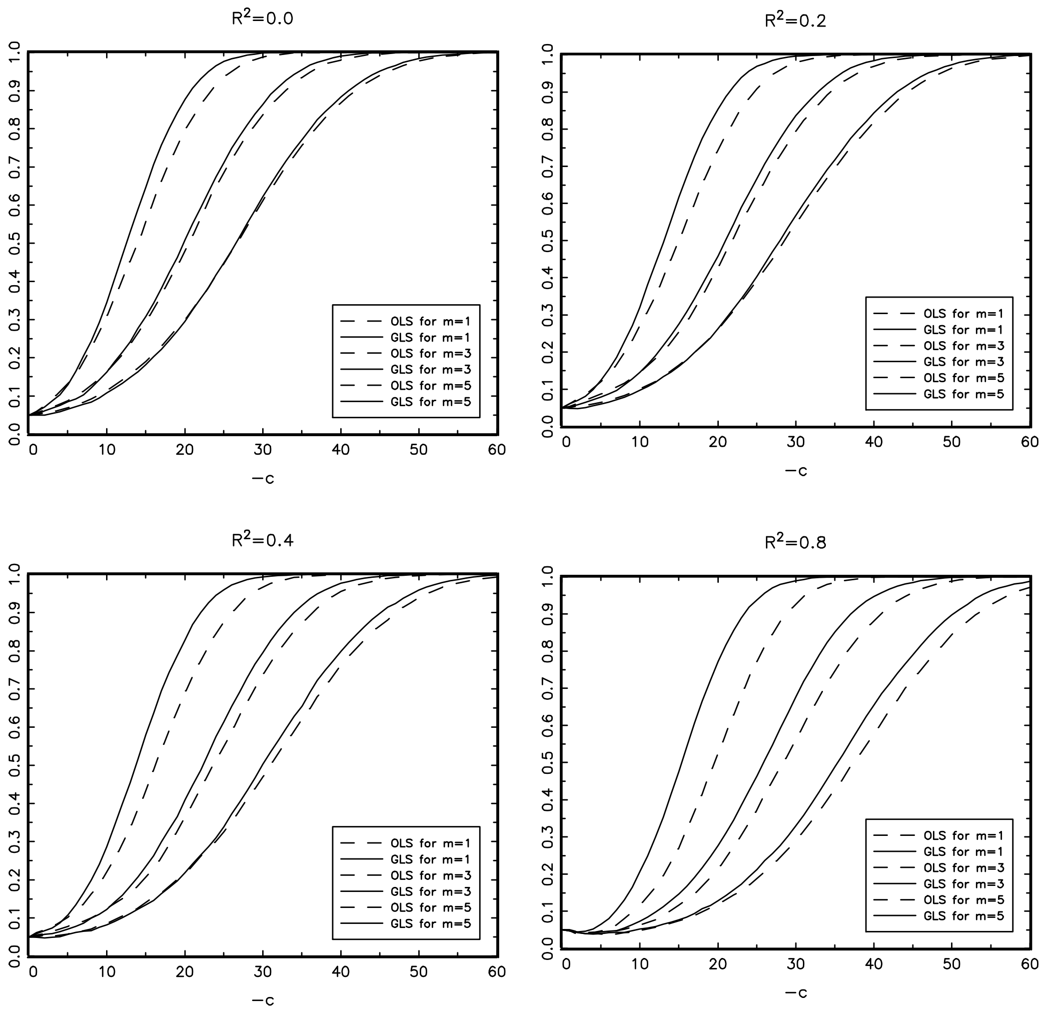

Figure 1a. Local Asymptotic Power Functions of MP_T^{OLS} and MP_T^{GLS}; Demeaned Case ($m = 1, 3, 5$).

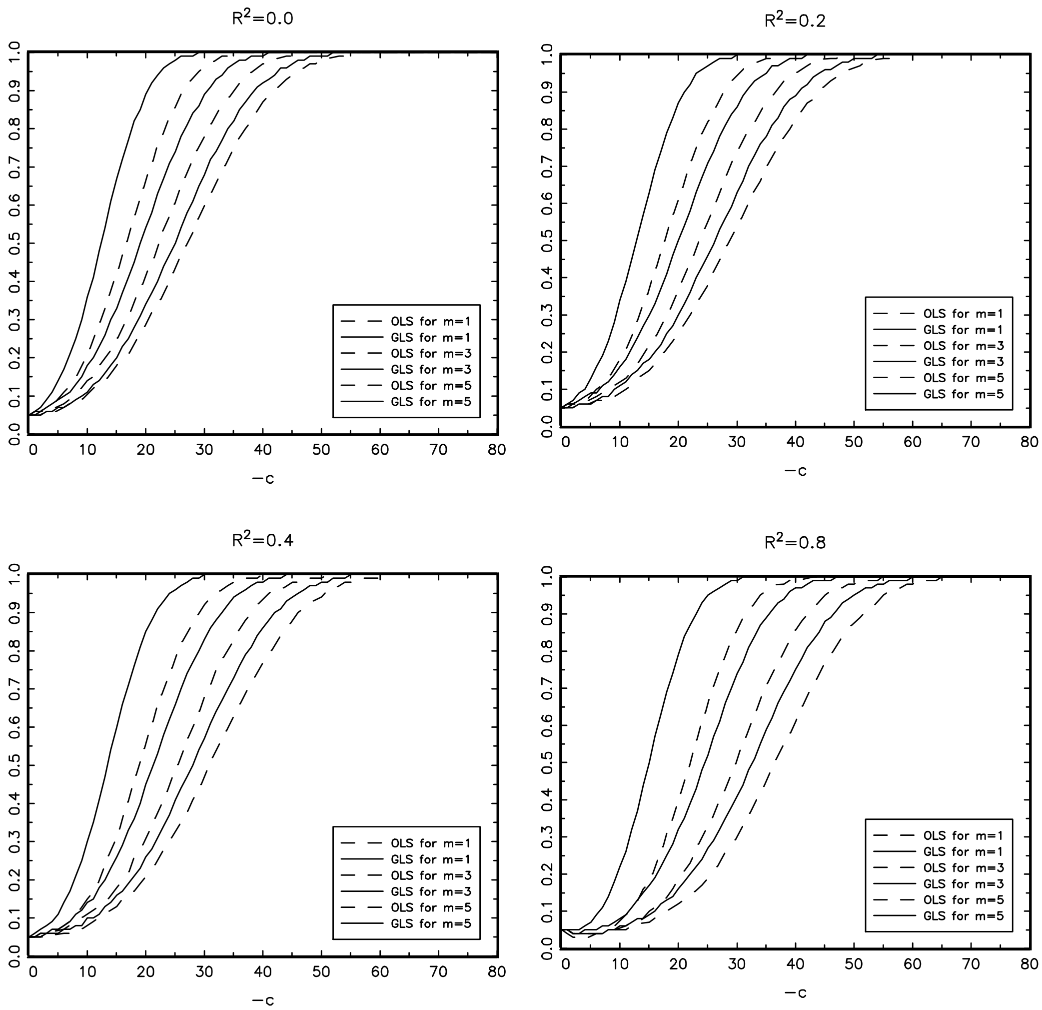

Figure 1b. Local Asymptotic Power Functions of ADF^{OLS} and ADF^{GLS}; Demeaned Case ($m = 1, 3, 5$).

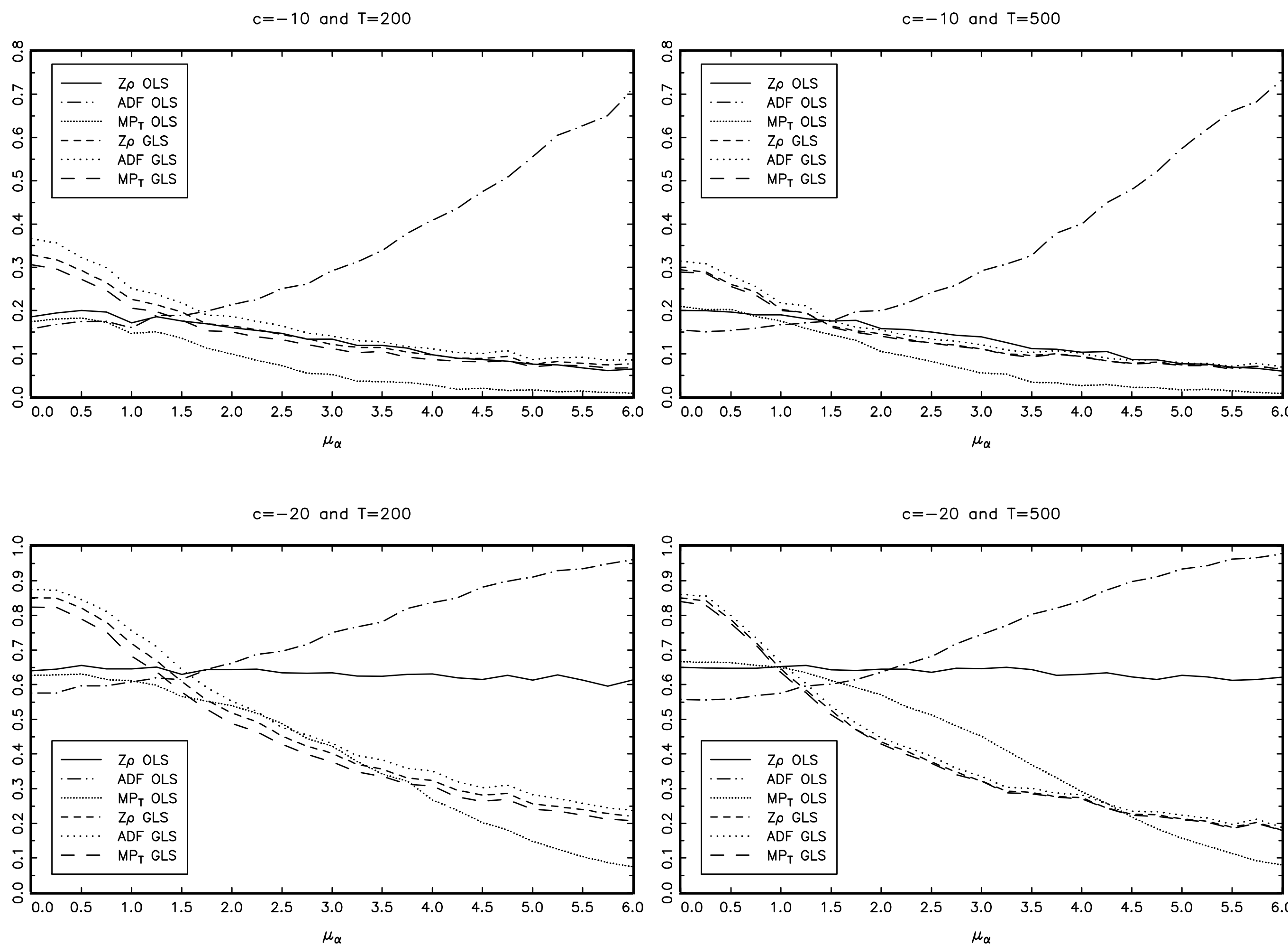

Figure 2. Power of the Residuals-based Tests with a Fixed Initial Condition

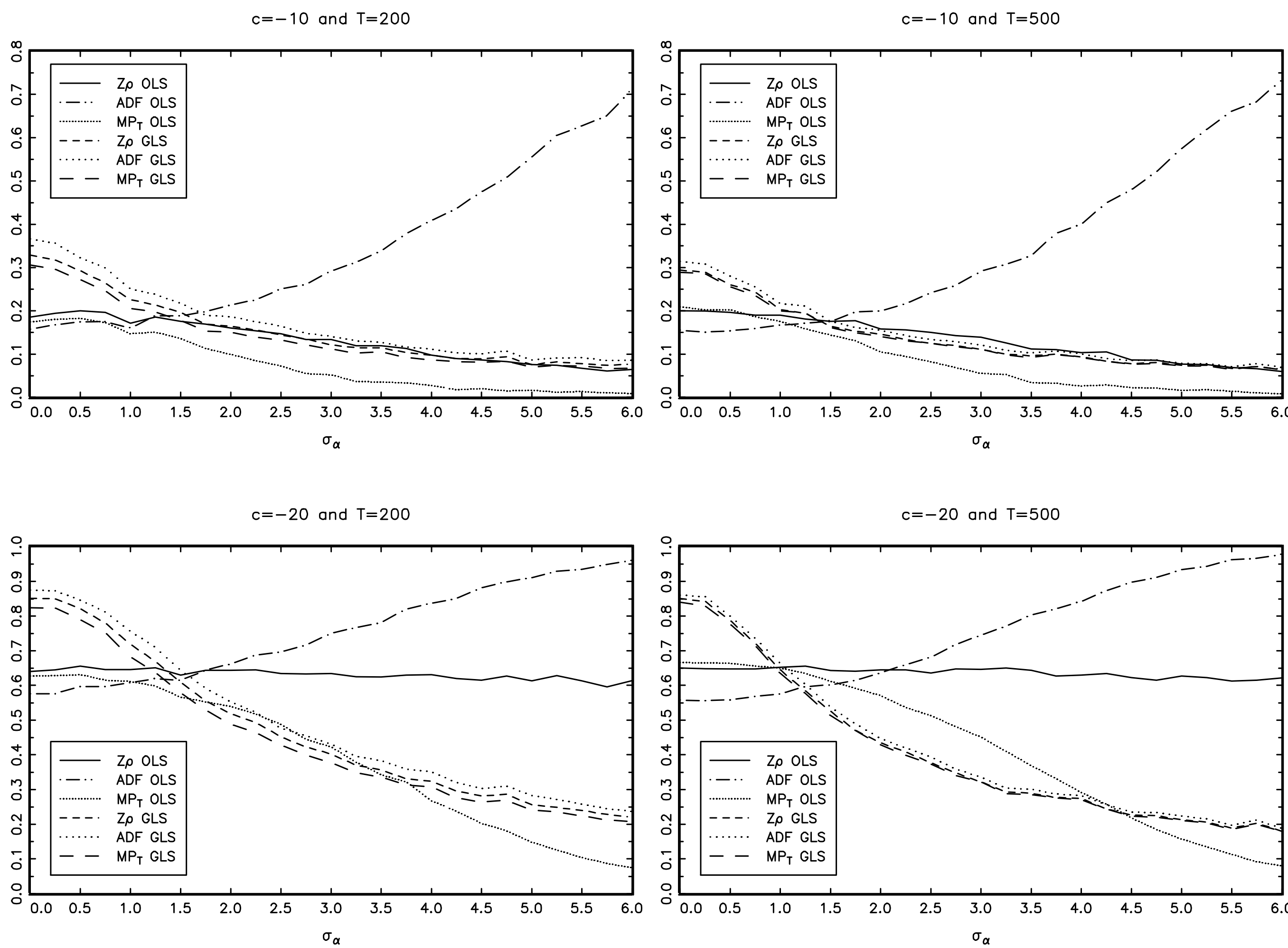

Figure 3. Power of the Residuals-based Tests with a Random Initial Condition