ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

A WAVELET APPROACH TO MULTIPLE COINTEGRATION TESTING

Javier Fernandez-Macho

Number 668
July 2013

Manor Road Building, Manor Road, Oxford OX1 3UQ

A WAVELET APPROACH TO MULTIPLE COINTEGRATION TESTING

JAVIER FERNÁNDEZ-MACHO

ABSTRACT. This paper introduces a class of cointegration tests based on estimated low-pass and high-pass regression coefficients from the same wavelet transform of the original time series data. The procedure can be applied to test the null of cointegration in a $n + k$ multivariate system with n cointegrating relationships without the need of either detrending nor differencing. The proposed non residual-based wavelet statistics are asymptotically distributed as standard chi-square with nk degrees of freedom regardless of deterministic terms or dynamic regressors, thus offering a simple way of testing for cointegration under the null without the need of special tables. Small sample quantiles for these wavelet statistics are obtained using Monte Carlo simulation in different situations including I(1) and higher order cointegration cases and it is shown that these wavelet tests exhibit appropriate size and good power when compared to other tests of the null of cointegration.

KEY WORDS: Brownian motion; Cointegration; Econometric Methods; Integrated process; Multivariate Analysis; Spectral Analysis; Time Series Models; Unit roots; Wavelet Analysis.
JEL CLASSIFICATION: C22, C12.

1. INTRODUCTION

Wavelet-based spectral analysis has become popular in a wide range of scientific disciplines from physics and signal engineering to medical imaging and, more recently, macroeconomics and finance. With time series data wavelet analysis is used, more often than not, because of its ability to offer some degree of resolution in both dimensions of the time-frequency plane. Therefore, it is usually considered as an alternative to Fourier-based spectral analysis in situations where, in exchange for frequency resolution, the latter eradicates time resolution that may still be informative.

Economic and financial data are conspicuous for the presence of nonstationary behaviour, *i.e.* frequency behaviour that changes over time, often linked to the presence of unit roots (Granger, 1966; Nelson and Plosser, 1982). The notion of non-spurious relationships amongst such variables led to the introduction of cointegration analysis (Granger, 1981, 1983; Engle and Granger, 1987) and, since then, testing for cointegration has become standard practice in the empirical analysis of economic time series. As a result, quite a number of time-domain-based tests have been proposed for this purpose.

On the other hand, it can been argued that cointegration refers to the existence of a meaningful relationship at a specific range of frequencies and, therefore, there is a frequency-domain equivalent of the time-domain cointegration hypothesis (Levy, 2000). However, frequency domain counterparts for cointegration testing have been fairly absent from the literature because of the inability of the Fourier transform to resolve nonstationarity, and we feel that wavelet-based spectral analysis can be useful here. More specifically, of the four areas mentioned by Ramsey (1999) in which wavelets may be useful in time series economics and finance, scale decomposition is, after a fashion, the one of interest for the present purpose as it relates to relationships between economic variables that can possibly exist at some time scales but not at others. (For a review on wavelet analysis from a time series perspective see Ramsey

Nuffield College and INET@Oxford Martin School (University of Oxford) and Dpt. of Econometrics & Statistics (UPV/EHU). Current E-mail address: javier.fernandez-macho@economics.ox.ac.uk. (v1.1_July2013).

1999; Schleicher 2002; Crowley 2007; Percival and Walden 2000; Gençay, Selçuk, and Whitcher 2002 among others).

This paper sets out to introduce a class of wavelet cointegration tests based on estimated low-pass and high-pass regression coefficients from the same wavelet transform of the original time series data. Thus, the proposal exploits the outcome of both high frequency wavelet filters that can be viewed as generalised difference filters plus the scaling filter counterpart that gives a smooth approximation isolating the low frequencies at which cointegration may occur. The resulting procedure is not residual-based and provides a test of cointegration under the null.[1]

There are some advantages of this procedure. Namely, the ability of wavelet analysis to deal with both stationary and non-stationary data without the need of either detrending nor differencing, the resulting statistics being asymptotically distributed as chi-square regardless of deterministic terms so that no special tables are needed, the applicability of the procedure on a system of multiple equations as well as on a single equation and that it can be used to test for higher order cointegration (*cf.* traditional inference methods for single I(1) cases in Phillips (1991) and Phillips and Hansen (1990) among others, as well as Fernández-Macho (2013) that uses simple difference filters for multiple I(1) systems).

The paper is organised as follows. Section 2 provides a brief introduction to wavelet analysis as used in the paper. Sections 3 and 4 present the model and statistics for the new wavelet-based test of the null of cointegration whilst section 5 examines the asymptotic properties. Section 6 presents quantiles of the distributions in finite samples. Finally, sections 7 and 8 show Monte Carlo evidence on size and power comparisons in I(1) and I(2) systems respectively.

Throughout the paper the following notation is used: For a given sequence of n-dimensional vectors $\{a_1, \ldots, a_T\}$, A will denote the $(T \times n)$ matrix such that its transpose is $A' = (a_1, \ldots, a_T)$ and vec(A) will denote the nT column vector obtained by stacking the columns of matrix A one underneath the other. As usual, all limits apply as $T \to \infty$, the summation $\sum$ runs from $t = 1$ to T unless otherwise stated, and $\Rightarrow$ means weak convergence of the associated probability measures. $B \equiv B(r)$ denotes the multivariate Brownian motion with covariance Ω and the integral $\int B$ refers to the Lebesgue measure in the $(0, 1]$ interval, *i.e.* $\int_0^1 B(r)dr$. Furthermore, $\mathbf{1}(\cdot)$, is the indicator function that takes a value of one when the argument is true and zero otherwise, $\mathcal{E}(\cdot)$ stands for the expectation operator, $\lfloor\cdot\rfloor$ means the integer part of the argument, $\|\cdot\|$ denotes the Euclidean norm, and $\otimes$ denotes the Kronecker product.

2. Wavelet transforms

Let $\{z_t\}$ be an input series of length T and let $\{f_{z\,t}\}$ be its circular discrete convolution with some filter or weight function f of length L:

$$
\begin{aligned}
f_z &= f * z,\\
f_{z\,t} &= \sum_{s=0}^{T-1} f((s-t)_{\bmod T})\, z_s,
\end{aligned}
$$

with zero-padded values as necessary if $L < T$. In matrix form we would write

$$f_z = Fz,$$

[1] We may compare this with Fan and Gençay (2010)'s wavelet unit root test which could conceptually be extended to provide a residual-based test of no-cointegration.

where F is a circulant whose first row is of the form $f = [f(0), f(1), f(2), \ldots, f(T-1)]$. For example, a useful transformation of this kind is the Fourier transform obtained by convoluting the input series with a filter of root-of-unity weights $f(s) = e^{-is\lambda}$, $0 \leq \lambda < 2\pi$, $s = 0 \ldots T-1$.

A discrete wavelet transform (DWT) is a double linear transform and, therefore, it can be represented in a similar fashion as a $T \times T$ matrix with a two-block structure where each block is the top half of a (shift 2)-circulant[2]:

$$\begin{bmatrix} \ddot{w} \\ \ddot{v} \end{bmatrix} = \begin{bmatrix} \ddot{H} \\ \ddot{G} \end{bmatrix} z. \tag{1}$$

The $\ddot{H}$ upper block, when multiplied with an input vector of observations, is meant to produce a vector of wavelet or detail coefficients $\ddot{w}$ that measure changes on the original series over time. It is constructed from a list of L weights $\ddot{h}(s)$, $s = 0 \ldots L-1$, orthogonal to its even shifts $\sum \ddot{h}(s)\ddot{h}(s+2n) = 0, \forall n \neq 0$, and scaled so as to have unit energy $\sum \ddot{h}(s)^2 = 1$, such that it has zero sum $\sum \ddot{h}(s) = 0$. In other words, the rows of matrix $\ddot{H}$ are zero-sum orthonormal vectors. Conversely, the $\ddot{G}$ lower block obtains a vector of scaling coefficients $\ddot{v}$ as a smooth approximation of the original series. It is constructed in a similar manner but with a different filter $\ddot{g}(s)$, $s = 0 \ldots L-1$ that is a quadrature mirror of $\ddot{h}(s)$, *i.e.* $\ddot{g}(s) = (-1)^{s+1}\ddot{h}(L-1-s)$, $s = 0 \ldots L-1$. Therefore, the rows of matrix $\ddot{G}$ are also orthonormal vectors but with sum $\sum \ddot{g}(s) = \sqrt{2}$.

From the frequency response functions corresponding to the high and low-pass filters, $\ddot{\mathfrak{H}}(\lambda) = \sum \ddot{h}(s)e^{-is\lambda}$ and $\ddot{\mathfrak{G}}(\lambda) = \sum \ddot{g}(s)e^{-is\lambda}$, we can see that the respective squared gains at the origin are $|\ddot{\mathfrak{H}}(0)|^2 = 0$ and $|\ddot{\mathfrak{G}}(0)|^2 = 2$. Furthermore, it can be shown (see Percival and Walden, 2000, p. 69,76) that the properties of the wavelet filters imply that their associated squared gain functions are such that

$$|\ddot{\mathfrak{H}}(\lambda)|^2 + |\ddot{\mathfrak{G}}(\lambda)|^2 = 2.$$

Common examples are members of Daubechies (1988, 1992) family of filters that are best described through their low-pass squared gain functions

$$|\ddot{\mathfrak{G}}(\lambda)|^2 = 2^{1-L/2}(1+\cos\lambda)^{L/2} \sum_{s=0}^{L/2-1} 2^{-s} \binom{L/2-1+s}{s} (1-\cos\lambda)^s,$$

(*cf.* Percival and Walden, 2000, p. 105). Furthermore, different spectral factorisations of the frequency response function $\ddot{\mathfrak{G}}(\lambda)$ give rise to D(L) (extremal phase or minimum delay) and LA(L) (least asymmetric) groups of Daubechies filters, of which the D(4) and the LA(8) are probably the most popular and the ones used in the empirical results section of this paper.

In short, the DWT can be thought of as an energy (variance) preserving transform that uses a pair of high-pass and low-pass filters to decompose an input series into the detail achieved by the wavelet filter and the corresponding smooth approximation. The level of decomposition can be escalated by applying the transformation to the previous approximation recursively so that a new vector of wavelet coefficients associated with changes at a higher scale (lower frequency) is produced alongside a remainder associated with an even smoother approximation of the original series at the new scale.

Let the series length be a power of two, *i.e.* $T = 2^J$. Equation (1) represents the first level of a wavelet transform. The full transform consists of a cascade of filters:

$$\begin{bmatrix} \ddot{w}_j \\ \ddot{v}_j \end{bmatrix} = \begin{bmatrix} \ddot{H}_j \\ \ddot{G}_j \end{bmatrix} \ddot{v}_{j-1}, \quad \ddot{v}_0 \equiv z, \qquad \text{for level } j = 1, \ldots, J,$$

[2] In a (shift k)-circulant the elements of the first row are circularly shifted k columns to the right.

where $\ddot{H}_j$ and $\ddot{G}_j$ are the $(T/2^j \times T/2^{j-1})$ top half of the (shift 2)-circulants with $[\ddot{h}(0)\ldots\ddot{h}(T/2^{j-1}-1)]$ and $[\ddot{g}(0)\ldots\ddot{g}(T/2^{j-1}-1)]$ as their respective first rows, so that $\ddot{w}_j$ represents the detail achieved by the wavelet transform at level j and $\ddot{v}_j$ the corresponding approximation. The maximum decomposition level J is given by $\lfloor \log_2(T) \rfloor$. However, the analysis is usually stopped short of that value as the number of wavelet coefficients $T/2^j$ gets critically small for high values of the decomposition level j. This notwithstanding, there is no redundancy no matter the level of decomposition since it is easy to see that the number of total coefficients obtained remains equal to the original length T regardless of j. Figure 1 shows an example of a length $L = 4$ DWT matrix for processing a sample of size $T = 8$.

The DWT succinctly described above uses the least number of coefficients conveying the same amount of information as the original series. However, it can only be applied to dyadic samples whilst, in certain applications, it is convenient that the number of coefficients remains the same as the original length T at each level of decomposition. An alternative transform, the maximal overlap DWT (MODWT), is obtained by not subsampling the filtered series. That is, the rows of the two blocks of the MODWT matrix W are obtained by circularly right-shifting the previous row one unit only until T rows are completed and rescaling by $2^{-1/2}$ so that the variance preserving property is maintained, *i.e.* $\sum h(s)^2 = \sum g(s)^2 = 1/2$. Also, we have that $\sum h(s) = 0$ and $\sum g(s) = 1$.

Therefore, each block of the wavelet matrix has T rows (instead of $T/2^j$) and the number of total coefficients obtained up to level j is $(j+1)T$ so that there is a high degree of redundancy involved and the transform is no longer orthogonal. The full MODWT consists of the following cascade of filters:

$$\begin{bmatrix} w_j \\ v_j \end{bmatrix} = \begin{bmatrix} H_j \\ G_j \end{bmatrix} v_{j-1}, \quad v_0 \equiv z, \qquad \text{for level } j = 1, \ldots, J, \tag{2}$$

where H_j and G_j are $(T \times T)$ circulants with $[\ddot{h}_s/\sqrt{2}, s = 0, \ldots, T-1]$ and $[\ddot{g}_s/\sqrt{2}, s = 0, \ldots, T-1]$ as their respective first rows, so that w_j represents the detail achieved by the wavelet transform at level j and v_j the corresponding approximation. The right hand side of fig. 1 shows an example of a MODWT matrix of length $L = 4$ for processing samples of size $T = 8$ where $h_s = \ddot{h}_s/\sqrt{2}$ and $g_s = \ddot{g}_s/\sqrt{2}$, $s = 0\ldots3$.

For the MODWT, it can be shown (see Percival and Walden, 2000, p. 163) that these properties imply that the associated squared gain functions satisfy

$$|\mathfrak{H}(\lambda)|^2 + |\mathfrak{G}(\lambda)|^2 = 1.$$

Indeed, from the frequency response functions corresponding to the high and low-pass filters, $\mathfrak{H}(\lambda) = \sum h(s)e^{-is\lambda}$ and $\mathfrak{G}(\lambda) = \sum g(s)e^{-is\lambda}$, it is easy to see that at the origin $|\mathfrak{H}(0)|^2 = 0$ and $|\mathfrak{G}(0)|^2 = 1$. In particular, this means that a MODWT using a wavelet filter of sufficient length will decompose an $I(d)$ series into two components associated respectively with high and low frequencies: the first one will effectively be stationary while the second one will retain the $I(d)$ behaviour. This characteristic can be used to advantage in testing the cointegration hypothesis.

3. The model

Let us consider the $(n+k)$-dimensional time series $(z_t', x_t')'$ $(t = 1\ldots T)$ generated by the following data generating process (DGP) (*cf.* Phillips 1991's triangular representation)

$$\underset{(n)}{z_t} = \underset{(n\times d)}{\alpha'}\ \underset{(d)}{g(t)} + \underset{(n\times k)}{\beta'}\ \underset{(k)}{x_t} + \underset{(n)}{u_t}, \tag{3a}$$

$$\ddot{W} = \begin{bmatrix} \ddot{H} \\ \ddot{G} \end{bmatrix} = \begin{bmatrix} \ddot{h}_0 & \ddot{h}_1 & \ddot{h}_2 & \ddot{h}_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddot{h}_0 & \ddot{h}_1 & \ddot{h}_2 & \ddot{h}_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddot{h}_0 & \ddot{h}_1 & \ddot{h}_2 & \ddot{h}_3 \\ \ddot{h}_2 & \ddot{h}_3 & 0 & 0 & 0 & 0 & \ddot{h}_0 & \ddot{h}_1 \\ \ddot{g}_0 & \ddot{g}_1 & \ddot{g}_2 & \ddot{g}_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddot{g}_0 & \ddot{g}_1 & \ddot{g}_2 & \ddot{g}_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddot{g}_0 & \ddot{g}_1 & \ddot{g}_2 & \ddot{g}_3 \\ \ddot{g}_2 & \ddot{g}_3 & 0 & 0 & 0 & 0 & \ddot{g}_0 & \ddot{g}_1 \end{bmatrix}$$

DWT

$$W = \begin{bmatrix} H \\ G \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 \\ g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 \\ g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 \\ g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 \end{bmatrix}$$

MODWT

Figure 1. Length $L = 4$ DWT and MODWT matrices for samples of size $T = 8$

where the elements of vector $g(t)$ are deterministic functions of time (*e.g.* a polynomial time trend of order r), $x_t \sim I(d)$ and u_t is a vector error process whose order of integration may possibly be less than that of z_t, *i.e.* $u_t \sim I(d_u)$ with $\mathrm{d}_u < d$ so that vector $(z_t', x_t')'$ is cointegrated in the sense of Engle and Granger (1987) and $\mathrm{d}_u = d$ otherwise. In what follows we will assume $d_u = 0$ for simplicity, but we will consider $d_u > 0$ in section 8.2.

Let us apply a MODWT based on a Daubechies wavelet filter of length $L \geq 2\max(r+1, d)$ to the right-hand side of eq. (3a). From Percival and Walden (2000, p.370) we have that the unit level high-pass filter h will cancel $g(t)$ out, whilst, on the other hand, it will transform the elements of x_t and u_t into stationary processes.

More formally, let $\zeta_t = (\mu_t', \eta_t')'$ follow an $(n+k)$-dimensional zero-mean stationary vector process and let h, g be a quadrature mirror pair such that h is a MODWT high-pass filter of length L and g the corresponding low-pass filter. We write

$$(h * x)_t = \eta_t, \tag{3b}$$

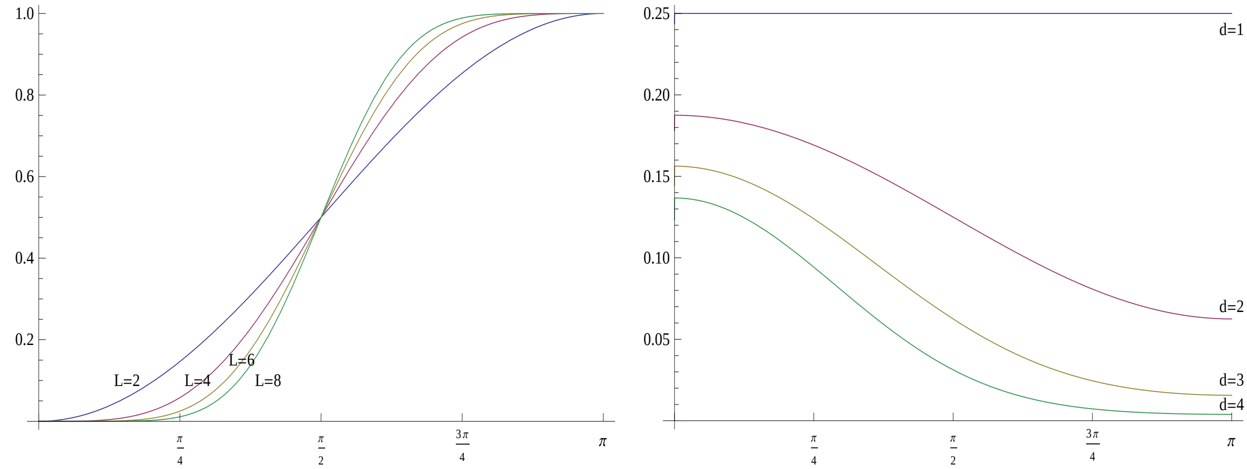
$$u_t - \delta[u_t - (h * u)_t] = \mu_t, \tag{3c}$$

where $\delta = 1 - \mathbf{1}(\mathcal{H}_0 : \text{cointegration})$, that is δ will take a value of 0 if vector $(z_t', x_t')'$ is cointegrated (the null) and a value of 1 otherwise (the alternative).

Note that the spectral density function (sdf) of $\{\eta_t\}$ is related to the sdf of $\Delta^d x_t$ through the square gain of the band-pass filter $h * \Delta^{-d}$ which, for h of sufficient length as mentioned, is finite everywhere in $[-\pi, \pi]$. Figure 2 shows the corresponding square gains for Daubechies filters of length $L = 2d$ and the corresponding band-pass filters for $d = 1 \ldots 4$. Therefore, the sdf of $\{\eta_t\}$, and hence its autocovariance function (acvf), can then be easily deduced from $\Delta^d x_t$. Likewise, the sdf and acvf of $\{\mu_t\}$ are related to $\{u_t\}$ in a similar manner.

Assumption 1. *Let* $\{\zeta_t\}$ *have an acvf* $\mathcal{E}(\zeta_t \zeta_{t-s}') = C(s)$, *(*$s = 0, \pm 1, \pm 2, \ldots$*) that is absolutely summable (*i.e. $\sum_{s=-\infty}^{\infty} \|C(s)\| < \infty$*) and sdf* $f_{\rho(\zeta)}(\cdot)$ *that is nowhere singular in* $[-\pi, \pi]$*. As in Park and Phillips (1988) we require that the partial sums of* $\{\zeta_t\}$ *satisfy a multivariate invariance principle*

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \zeta_t \Rightarrow B(r), \qquad r \in (0, 1]$$



h = Daubechies MODWT $\qquad$ $h * \Delta^{-d}$

FIGURE 2. Square gain of the band-pass filter $h * \Delta^{-d}$.

where $B(r) = (B_\mu(r)', B_\eta(r)')'$ denotes an $(n+k)$-variate Brownian motion with covariance matrix

$$\Omega = \lim_{T\to\infty} Var\left(T^{-1/2} \sum_{t=1}^{T} \zeta_t\right) = \begin{bmatrix} \Omega_\mu & \Omega'_{\eta\mu} \\ \Omega_{\eta\mu} & \Omega_\eta \end{bmatrix} = 2\pi f_{\rho(\zeta)}(0)$$

where $B(r)$ and Ω have been partitioned conformably with ζ_t. Note that $\Omega > 0$, since the sdf is nonsingular within $(-\pi, \pi)$. □

In this respect, note that, since $|\mathfrak{H}(0)|^2 = 0$ and $f_{\rho(\eta)}(0) > 0$, *i.e.* the convolution $(h * x)$ has non-zero sdf at the origin, then $f_{\rho(x)}(0)$ cannot be finite and $x \sim I(d)$ with $d > 0$. Similarly, under $\mathcal{H}_a$, the convolution $(h * u)$ has non-zero sdf at the origin, then $f_{\rho(u)}(0)$ cannot be finite and $u \sim I(d)$.

4. THE TEST

Applying a MODWT (2) of level $j < J = \lfloor \log_2(T) \rfloor$ to model (3) the low-pass filter will leave the time polynomial $g(t)$ unaffected (Percival and Walden, 2000, p.370) and we write

$$v_{z,jt} = \alpha' g(t) + \beta' v_{x,jt} + v_{\mu,jt} + \delta(v_{u,jt} - w_{u,j+1,t}), \tag{4}$$

where $v'_{u,jt}$ and $w'_{u,jt}$ are the $(t+1)$-th rows of the $T \times n$ matrices $v_{u,j} = G_j G_{j-1} \dots G_1 u$ and $w_{u,j} = H_j G_{j-1} \dots G_1 u$ respectively, and similarly $v'_{z,jt}$, $v'_{\mu,jt}$ and $v'_{x,jt}$ are the $(t+1)$-th rows of the $T \times n$ matrices $v_{z,j} = G_j G_{j-1} \dots G_1 z$, and $v_{\mu,j} = G_j G_{j-1} \dots G_1 \mu$ and the $T \times k$ matrix $v_{x,j} = G_j G_{j-1} \dots G_1 x$ respectively.

Following, say, Brillinger (1975, p.296), we start by conditioning $v_{\mu,jt}$ on $\{\eta_t\}$ so that

$$v_{\mu,jt} = \sum_{s=-\infty}^{\infty} \gamma'_s \eta_{t-s} + \xi_t \tag{5}$$

where the $(k \times n)$ filter $\{\gamma_s\}$ is absolutely summable, *i.e.* $\sum_{s=-\infty}^{\infty} \|\gamma_s\| < \infty$, and ξ_t is an n-dimensional zero-mean stationary process such that $\mathcal{E}(\xi_t \eta'_{t-s}) = 0$, $(s = 0, \pm1, \pm2, \dots)$, $\forall t$. Therefore, (*cf.* Saikkonen, 1991) $\exists\, m$ large enough so that $\gamma_s \approx 0$ for $|s| > m$ and the sum in eq. (5) may be truncated at $|s| = m$. More specifically, we require

Assumption 2. *Let $m \to \infty$ with T at a suitable rate such that $m^3/T \to 0$ and $T^{1/2} \sum_{|s|>m} \|\gamma_s\| \to 0$ specify upper and lower rate bounds for m.* □

Then, the low-pass filtered part of z_t conditional on $\{x_{t-m-1}, \dots, x_{t+m}\}$ can be written as

$$v_{z,jt} = \alpha' \, g(t) + \beta' \, v_{x,jt} + \sum_{s=-m}^{m} \gamma_s' \, w_{x,t-s} + \varepsilon_t + \delta u_t^{\dagger}, \tag{6a}$$

where $\{(h * x)_t \equiv w_{xt} = \eta_t\}$ are the MODWT wavelet coefficients of the x regressors, ε_t is an n-dimensional zero-mean stationary process such that $\mathcal{E}(\varepsilon_t \eta_{t-s}') = 0$, $\forall t, s$, $u_t^{\dagger} = v_{u,jt} - w_{u,j+1,t}$ and $\delta = \mathbf{1}(\mathcal{H}_a)$. Under the null of cointegration, $\delta = 0$ and OLS estimation will produce an efficient (and superconsistent) estimator of the cointegrating vectors whose limiting distribution is free of the nuisance parameters γ_j arising from the short run dynamics of the DGP.

Alternatively, we may define

$$y_t \stackrel{def}{=} v_{z,jt} - \tilde{\alpha}' \, g(t) - \sum_{s=-m}^{m} \tilde{\gamma}_s' \, w_{x,t-s}$$

with $\tilde{\alpha}$ being the vector of OLS estimates of coefficients of deterministic components and $\{\tilde{\gamma}_s\}$ the estimates of the nuisance parameters, and write the low-pass regression as

$$y_t = \beta' v_{x,jt} + \varepsilon_t + \delta u_t^{\dagger}. \tag{6b}$$

In summary, under the null of cointegration the error term in either eq. (6) is simply $\varepsilon_t \sim I(0)$, whilst under the alternative of no cointegration the error term contains $v_{u,jt} \sim I(d)$ in it. As a result, the OLS estimator $\hat{\beta}_g$ from the low-pass regression will be T-consistent under the null of cointegration but will have a non-degenerate distribution under the alternative (Phillips and Durlauf, 1986; Stock, 1987).

On the other hand, the unit level high-pass filter from the same MODWT gives either

$$w_{zt} = \beta' \eta_t + \sum_{s=-m}^{m} \gamma_s' \, w_{\eta,t-s} + \varepsilon_t^{\dagger}, \tag{7a}$$

or

$$w_{yt} = \beta' \eta_t + \varepsilon_t^{\dagger}, \tag{7b}$$

where the error process $\varepsilon_t^{\dagger} = w_{\varepsilon,t} + \delta(\mu_t - w_{\mu,t})$ is always stationary, so that standard asymptotics on stationary variables apply yielding an OLS estimator $\hat{\beta}_h$ that it is $\sqrt{T}$-consistent under the null while still being $O_p(T^{-1/2})$ under the alternative (note that in this case $\hat{\beta}_h$ might be asymptotically biased since $\mathcal{E}(\eta_t w_{\mu,t}')$ need not be zero in general). This high-pass regression may then be used as a benchmark for the low-pass regression in order to test for cointegration.

The presence of $g(t)$ in equation eq. (3) implies 'stochastic' cointegration around some deterministic function of time. On the other hand, absence of any deterministic component means that there is 'deterministic' cointegration in the sense that deterministic components are eliminated together with the stochastic components. However, it is well known that the inclusion of deterministic components in a cointegrating regression causes shifts in the asymptotic distributions of residual-based tests. This will not be so in the present case since the high-pass regression in eq. (7) is not only free from short-run-dynamics nuisance parameters but also free, by construction, from deterministic components.

The so called Hausman test statistic (Hausman, 1978; Durbin, 1954), rests on the comparison between two estimators, both of them consistent under the null but with different probability limits under the alternative. The standardised difference between the two estimates will then have zero probability limit under the null but will diverge under the alternative (for test consistency). Accordingly, a testing procedure based on the discrepancy $c = \mathrm{vec}(\hat\beta_h - \hat\beta_g)$ between the OLS estimators obtained from a high-pass regression and a low-pass regression can be proposed. The test statistics are:

$$\mathrm{W1} = c'(\hat V_h + \hat V_g)^{-1} c, \qquad \mathrm{W2} = c' \hat V_h^{-1} c, \tag{8}$$

where $\hat V_h$ and $\hat V_g$ are consistent estimates of the covariance matrices of $\hat\beta_h$ and $\hat\beta_g$ respectively, given by

$$\hat V_g = \hat\Omega_\varepsilon \otimes (v'_{x,j} v_{x,j})^{-1}, \qquad \hat V_h = \left(I_n \otimes (w'_x w_x)^{-1}\right) \hat R \left(I_n \otimes (w'_x w_x)^{-1}\right) \tag{9}$$

where $\hat\Omega_\varepsilon$ is a consistent estimator of the 'long run variance' matrix $\Omega_\varepsilon = 2\pi f_\varepsilon(0)$ and

$$\hat R = (I_n \otimes w'_x)\, \hat V_{w\varepsilon}\, (I_n \otimes w_x) \tag{9b}$$

where $\hat V_{w\varepsilon}$ is a consistent estimator of the covariance matrix of $\{w_{\varepsilon,t}\}$. Since $\hat V_g$ is $O_p(T^{-2})$ while $\hat V_h$ is $O_p(T^{-1})$ the two statistics are asymptotically equivalent.

One last comment on the interpretation of our W statistics from a different point of view: W2 may also be figured out as the typical chi-square statistic for testing whether $\hat\beta_h$ is significantly different from true β. In order to evaluate the statistic, the unknown β is replaced by $\hat\beta_g$ whose faster consistency rate ensures that the asymptotic distribution remains unaltered. The addition of $\hat V_g$ in the denominator of W1 intends to provide a better approximation in small samples (see lemma 6 in the appendix).

5. Asymptotic distributions

Let $f_{\rho(\varepsilon)}(\cdot)$ be the error sdf of $\{\varepsilon_t\}$ in eq. (6). Under assumption 1 (Saikkonen, 1991, p.11) the partial sums of $\{\varepsilon_t\}$ will be such that

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \Rightarrow B_\varepsilon(r), \qquad r \in (0,1]$$

where $B_\varepsilon(r) = B_\mu - \Omega'_{\eta\mu}\Omega_\eta^{-1} B_\eta$ is a k-variate Brownian motion (uncorrelated by construction with B_η) with covariance matrix

$$\Omega_\varepsilon = \Omega_\mu - \Omega'_{\eta\mu}\Omega_\eta^{-1}\Omega_{\eta\mu} = 2\pi f_{\rho(\varepsilon)}(0).$$

Note that $f_{\rho(\varepsilon)}(\cdot)$ is related to the sdf of ζ_t through the expression $f_{\rho(\varepsilon)}(\cdot) = f_{\rho(\mu)}(\cdot) - f_{\rho(\eta\mu)}(\cdot)' f_{\rho(\varepsilon)}(\cdot)^{-1} f_{\rho(\eta\mu)}(\cdot)$ (Brillinger, 1975, p.296). Since Ω is nonsingular, we also have that $\Omega_\varepsilon > 0$ whilst $\Omega_\eta > 0$ rules out cointegration between the regressors. Under these circumstances, we are able to state the following

Proposition 1. *In the multivariate regression model* (3) *with assumptions 1 and 2: under the null of cointegration, the wavelet statistics defined in eq.* (8)

$$\mathrm{W1}, \mathrm{W2} \Rightarrow \chi^2(nk)$$

while under the alternative of no cointegration

$$\mathrm{W1} = O_p(T), \qquad \mathrm{W2} = O_p(T)$$

Proof: (see the appendix). □

Note that, although they are asymptotically equivalent under the null, under the alternative W1 and W2 would not have the same limit distribution. This different asymptotic behaviour under the alternative will have consequences for test power: indeed in the limit $T^{-1}\mathrm{W1} < T^{-1}\mathrm{W2}$ and test W2 will be asymptotically more powerful (see the appendix).

In short, it has been shown that the proposed wavelet test statistics are bounded in probability under the null hypothesis of cointegration but they are $O_p(T)$ under the alternative, which ensures test consistency.

Furthermore it has been shown that under cointegration the W statistics tend asymptotically towards the standard chi-square distribution. Asymptotic tests can thus be performed straight away. All this means that the test statistics proposed may constitute a useful procedure for testing directly the hypothesis of cointegration under the null.

6. Small sample quantiles

Tables 1 to 3 give critical values of the W statistics for cointegrated systems with $n = 1, 2, 3$ and up to $n + k = 6$ cointegrating equations calculated *via* Monte Carlo simulation. The data generating process was

DGP 1. *Model* (3) *with* $d = 1$, $\delta = 0$, $\alpha = 0$, $\beta_1 = \ldots = \beta_k = 1$ *and* $\zeta_t \sim \mathrm{iid}\mathcal{N}(0, I_{n+k})$.

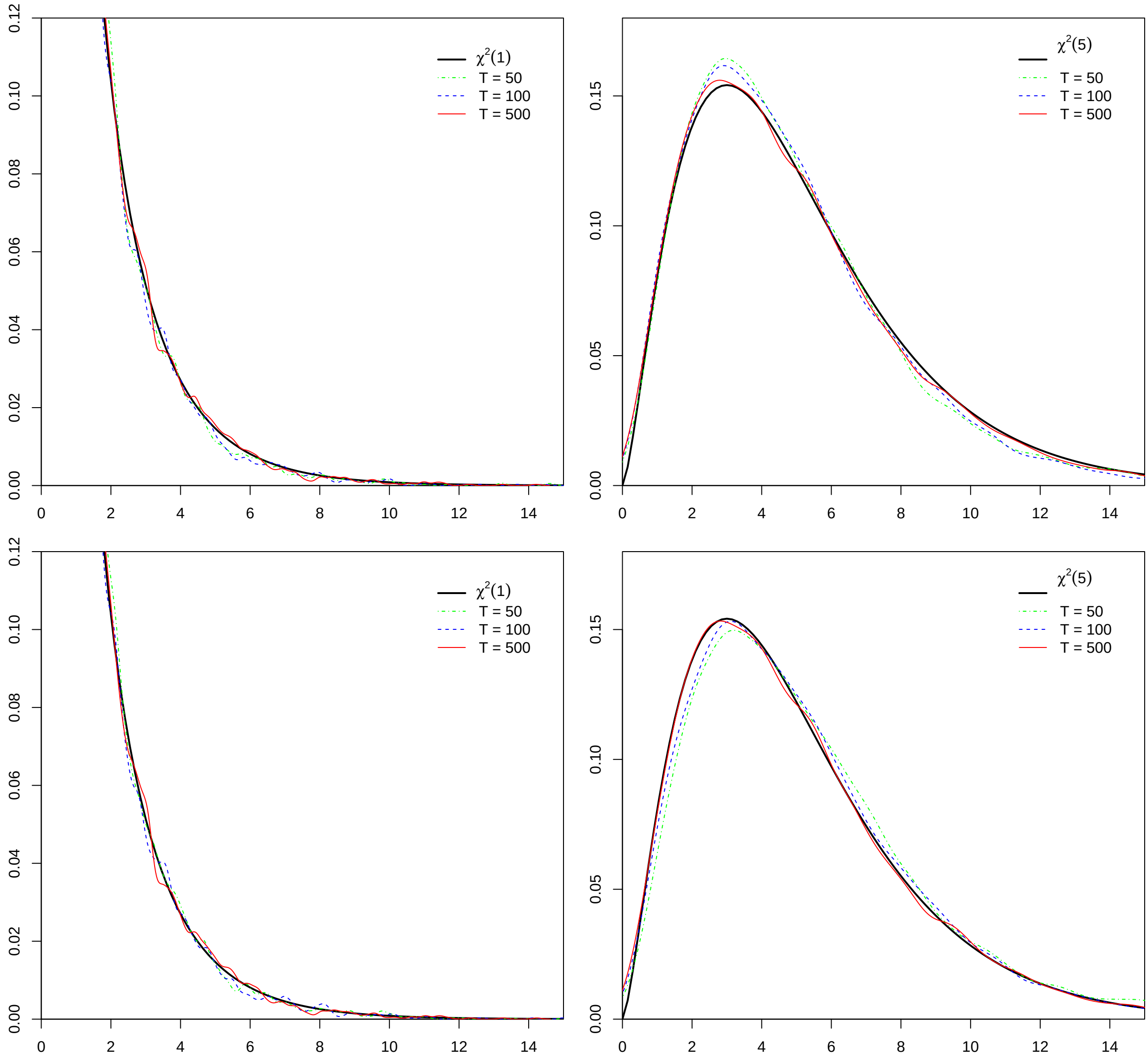
All $z_{1t} \ldots z_{nt}$, $x_{1t} \ldots x_{kt}$ thus generated are $I(1)$ cointegrated series with n cointegrating vectors $(1, -1, \ldots, -1)$. A unit-level MODWT based on the D(4) filter was used and, therefore, the statistics' distributions would be unaffected by the presence of a constant or polynomial linear trend. Using DGP 1 the quantiles of the small sample distribution of W1 and W2 for different sample sizes from $T = 20$ to 500 were approximated out of 10 000 replications each using the random number generator available with the R statistical package, version 3.0.0. Figures 3 to 5 show density estimates of W1 and W2 for some combinations of cointegrating equations (n) and unit roots (k). It may be worth noting how the finite sample distributions approach their corresponding asymptotic χ^2 distribution even for sample sizes as small as $T = 100$.

7. Size and power comparisons in I(1) systems

Comparing the small sample performance of the wavelet W statistics with respect to standard CI tests would not be satisfactory given the reversal of the null and alternative hypothesis involved. Therefore, Shin (1994)'s C(ℓ) and Leybourne and McCabe (1994)'s LBI residual-based tests of the null of cointegration as well as wavelet statistics based on the Haar filter[3] were chosen for comparison purposes against both dynamic (6a) and short-run dynamics free (6b) versions of the W statistics.

Shin's statistic is given by $C(\ell) = \sum_{t=1}^{T} S_t^2 / T^2 s^2(\ell)$ where $\{S_t\}$ denotes the partial sum process of the OLS residuals from the cointegrating regression and $s^2(\ell) = \sum_{|s|<T} w(s, \ell)\, \hat{r}_s$, with $\hat{r}_s$ as the s-th residual autocovariance, is a consistent semiparametric estimator of the long-run variance of the regression errors. The first choice for $w(\cdot)$ was a lag window corresponding to the Bartlett-Priestley quadratic spectral kernel (qs), which, according to Andrews (1991), is optimal in terms of bias, mean squared error and true confidence levels $w(s, \ell) = 3[\sin \tau(\ell)/\tau(\ell) - \cos \tau(\ell)]/\tau(\ell)^2$, with $\tau(\ell) = (6\pi/5)(s/\ell)$ and associated bandwidth parameter $\ell = 1.3221[4T\hat{\rho}_1^2/(1 - \hat{\rho}_1)^4]^{1/5}$, where $\hat{\rho}_1$ is the first-order autoregressive coefficient estimate of the regression errors, so that the lag length is somehow automatically chosen based on the data. The C statistic with Bartlett's triangular lag window $w(s, \ell) = (1 - |s|/\ell)\, \mathbf{1}(|s| < \ell)$ was also calculated. The LBI statistic is obtained in a similar fashion as Shin's statistic but for the use of a rectangular lag window: $w(s, \ell) = \mathbf{1}(|s| < \ell)$. Following Kwiatkowski, Phillips, Schmidt, and Shin (1992), the truncation lag length ℓ in these two cases was set as a function of T: $\ell_4 = 1 + \lfloor 4(T/100)^{1/4} \rfloor$. Also, the number m of additional regressors' leads and lags in the parametric correction for C and W statistics was set according to the BIC criterion as in Venables and Ripley (2002).

[3] The use of a Haar filter (length $L = 2$) will make the wavelet statistics roughly equivalent to Fernández-Macho (2013)'s Hausman-like test statistics of the null of cointegration.



Smoothed histograms: Gaussian kernel density estimates. W1 top, W2 bottom; $k = 1$ left, $k = 5$ right.

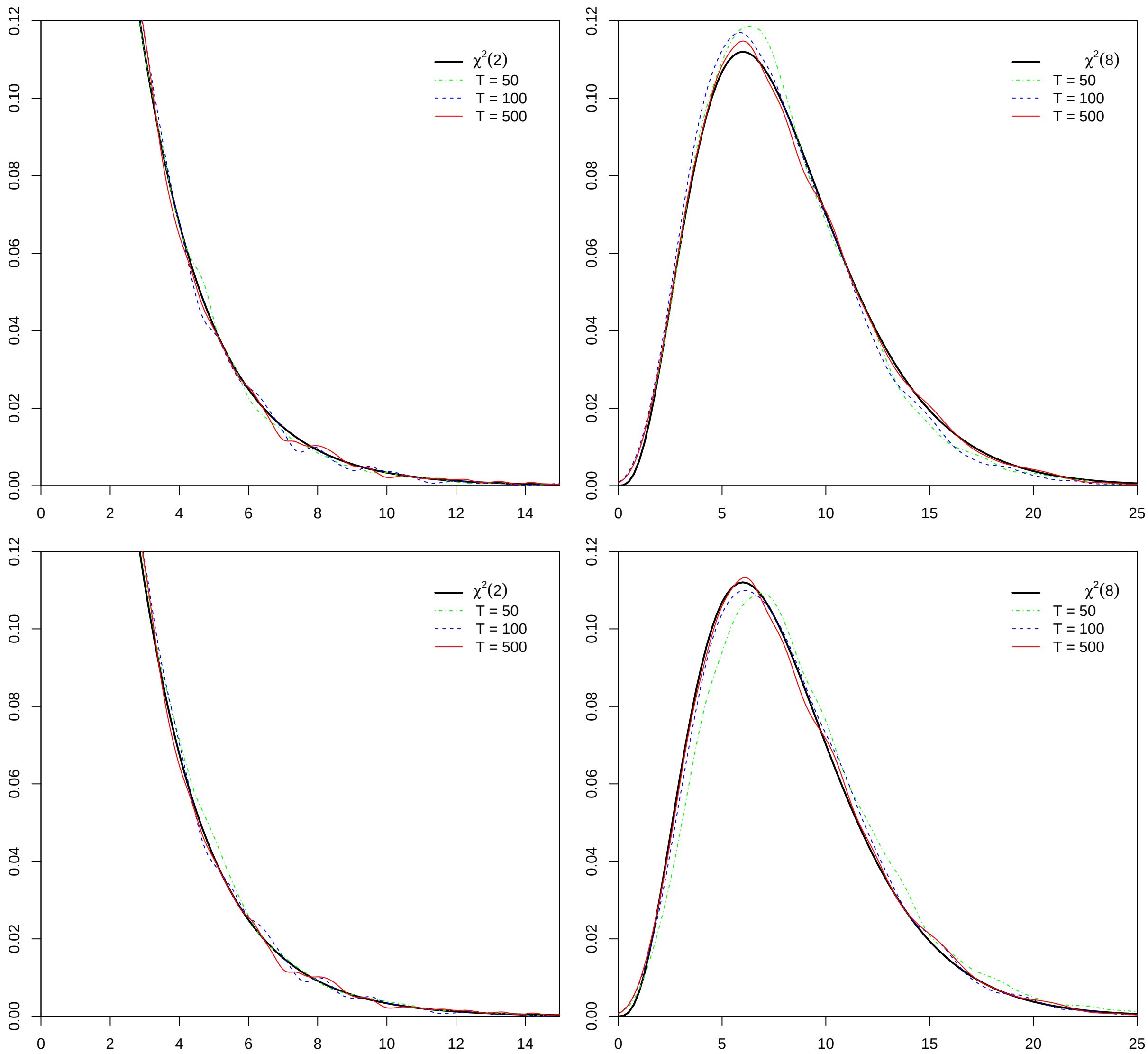
FIGURE 3. Empirical Distributions of W Statistics ($n = 1$).

Figure 6 summarises the evidence provided by tables 4 and 5 on the size and power of tests of the null of cointegration in small I(1) samples. The first five rows in the tables refer to the proposed wavelet W1 and W2 test statistics using a unit-level MODWT based on the LA(8) Daubechies filter and their asymptotic χ^2 critical values. The next two rows (H1 and H2) refer to the same wavelet test statistics but based on the Haar filter. Results for Shin's test for the quadratic kernel C(qs) and Barlett kernel C(ℓ_4) are reported next. The last row refer to Leybourne-McCabe's LBI test statistic. The reported results were obtained through Monte Carlo simulation using 10 000 replications of size $T = 100$ from

DGP 2. *A bivariate regression model* (3a) *with* $d = 1$, $\alpha = 0$, $\beta = 1$ *and* [4]

$$\begin{aligned} \mu_t &= \phi\,\mu_{t-1} + a_{1t} + \theta\,a_{1,t-1} \\ \eta_t &= \phi_\eta\,\eta_{t-1} + a_{2t} + \theta_\eta\,a_{2,t-1} \end{aligned} \qquad \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} \sim \text{iid}\,\mathcal{N}\left[0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right].$$

[4] Note that δ in model (3) is simply an indicator of the existence of a unit root, while all other possible stationary AR roots are implicit in $\{\mu_t\}$. In this section, however, ϕ makes explicit the value of one AR root. In other words, these statements are equivalent: $\phi = 1 \equiv \delta = 1$ and $|\phi| < 1 \equiv \delta = 0$.



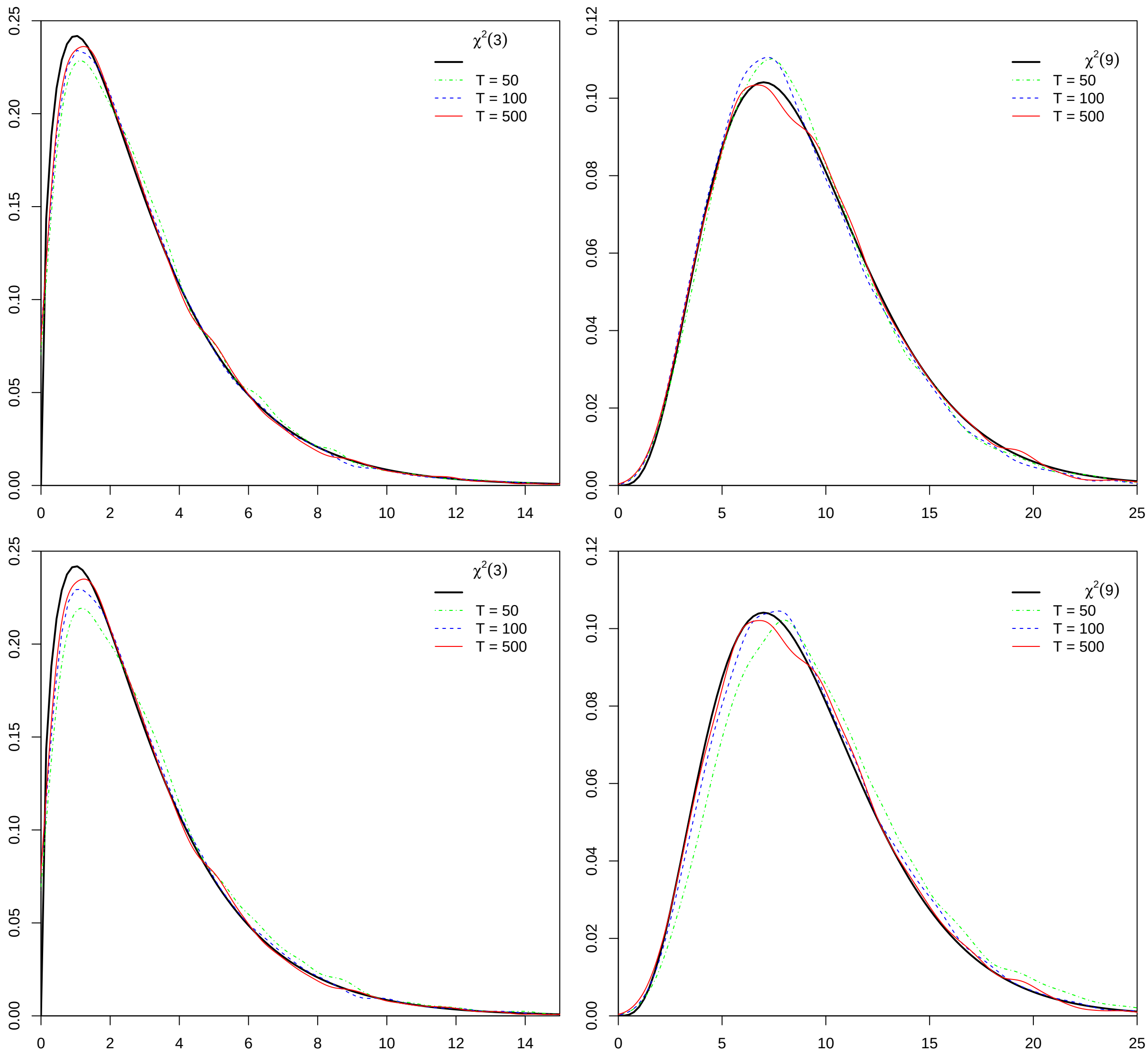
Smoothed histograms: Gaussian kernel density estimates. W1 top, W2 bottom; $k = 1$ left, $k = 4$ right.

FIGURE 4. Empirical Distributions of W Statistics ($n = 2$).

This setup allows for quite general behaviour. The regressor $\{x_t\}$ follows an ARIMA$(1, 1, 1)$ process while the error term $\{u_t\}$ follows an ARIMA$(1, 0, 1)$ process under the null but an ARIMA$(1, 1, 1)$ under the alternative. Besides, x_t is not strongly exogenous since η_t and μ_t may exhibit nonzero correlations (at different lags). The lag length m of the parametric correction in (6) was again determined by the BIC method. In the simulations, we considered either stationary AR(1) error processes under the null or IMA$(1, 1)$ error processes under the alternative, with the following values for the parameters:

$$\begin{aligned} &\phi_\eta = 0.45, \qquad \theta_\eta = -0.35, \qquad \rho = 0.5, \qquad \sigma = 1, \\ &\theta = 0, \phi \in (0, 0.2, 0.4, 0.6, 0.8) \quad \text{under the null,} \\ &\phi = 1, \theta \in (0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8, +1) \quad \text{under the alternative.} \end{aligned} \tag{10}$$

The parameter values chosen for the regressor generating process and its correlation with the error term are high enough to offer a clear departure from both pure random walks and strongly exogenous regressors. Haug (1996) uses essentially the same DGP 2 (see also Hansen and Phillips (1990) and Gonzalo (1994)



Smoothed histograms: Gaussian kernel density estimates. W1 top, W2 bottom; $k = 1$ left, $k = 3$ right.

FIGURE 5. Empirical Distributions of W Statistics ($n = 3$).

amongst others) although he allows for more diversity in the regressor dynamics than here; in turn, our simulations contemplate more variety in the dynamics of the error term both under the null and, especially, under the alternative. We note that the two $I(1)$ series z_t and x_t generated by DGP 2 are cointegrated as long as $|\phi| < 1$; however, the regression error term will become a random walk as $\phi \to 1$. Therefore $(1-\phi)$ can be taken as a measure of distance from $\mathcal{H}_a$: non-cointegration. The values of ϕ in table 4 were chosen so that changes in the tests' size can be observed as the alternative is being approached. On the other hand, with $\phi = 1$, both z_t, x_t generated by DGP 2 are not cointegrated as long as $\theta \neq -1$; however, the regression disturbance will collapse to white noise as $\theta \to -1$. Therefore $(1+\theta)$ can be taken as a measure of distance from $\mathcal{H}_0$: cointegration. In table 5, θ was chosen so that changes in the tests' power can be observed as the null is being approached (negative values) and as it gets farther away from it (positive values). Figure 6 shows the respective size and power of the tests involved in our comparison as functions of the said distances from $\mathcal{H}_a$ and $\mathcal{H}_0$ respectively.

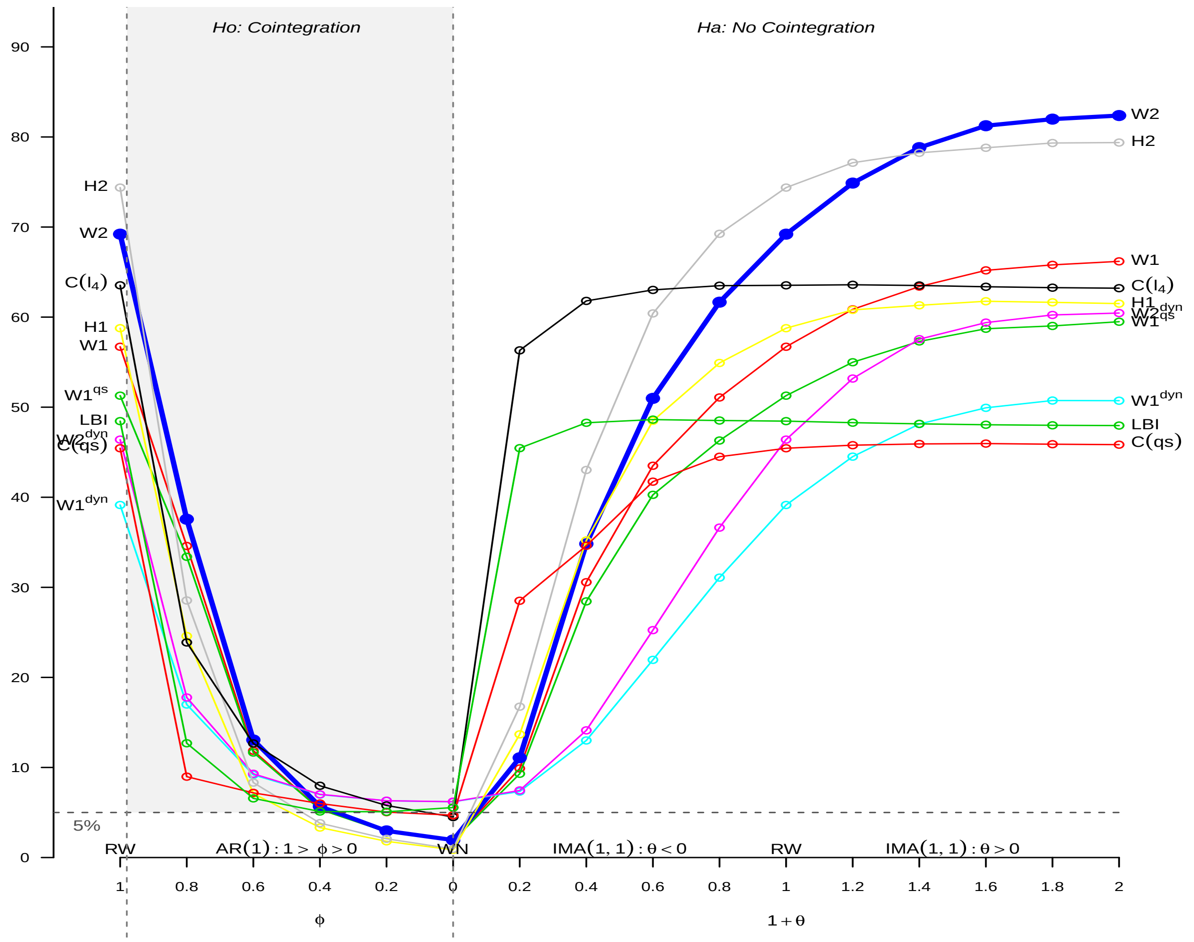


FIGURE 6. CI(1,1): CI Tests Power vs. distance from $\mathcal{H}_0$ (DGP 2; $T = 100$; 5% level).

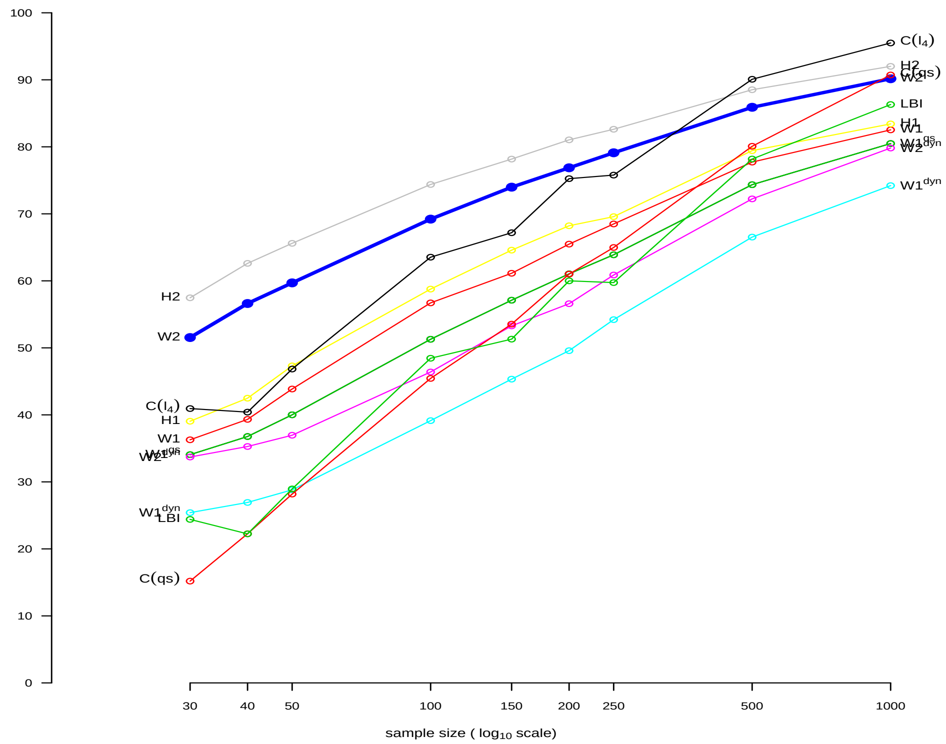


FIGURE 7. CI(1,0): Power of Cointegration Tests vs. sample size (DGP 2; 5% level).

As far as test size is concerned, the results reported indicate that in the presence of moderately autocorrelated errors ($0 < \phi \leq 0.6$ say) all tests maintain size distortions well within reasonable levels.

As far as test power is concerned, the last two columns of table 4 correspond to the case $\phi = 1$ (which falls just outside the cointegration region: distance from $\mathcal{H}_a = 0$ in fig. 6) and the case $\theta = 1$ (the farthest alternative considered: distance from $\mathcal{H}_0 = 2$ in fig. 6). In fact $\{z_t\}$ and $\{x_t\}$ are two I(1) processes that are related through their changes, although no meaningful relationship in levels exists. The results presented clearly favour the W2 statistic in the sense that it appears more powerful. For example, for samples of size $T = 100$ at the 5% significance level, W2 will reject up to 82% of the times the (wrong) null hypothesis of a relationship in levels in favour of a (true) relationship in changes. In the same circumstances C(qs) managed just under 46% rejections. It is interesting to note that the power of the W statistics is directly related to the 'distance' from $\mathcal{H}_0$: the farther we are from it (as $\theta \to +1$) the greater the power and *viceversa*, the closer we are from the null (as $\theta \to -1$) the smaller the power (see fig. 6). Such behaviour is of course very reasonable although it is violated by the other test statistics (see bold numbers in table 5); for example, C(ℓ_4) attains a maximum at $\theta = 0.2$ and LBI at $\theta = -0.4$. On the other hand, the W tests appear to be less powerful for cases closer to the null. On the whole, the wavelet test statistic exhibits a very satisfactory behaviour: it is more powerful for a wide range of the alternatives considered, and does not reject more often than it should under the null.

Figure 7 (table 6) presents evidence on the evolution of the tests' power as the sample size increases. It has been obtained from simulations using 10 000 replications from DGP 2 with random walk errors, (*i.e.* $\phi = 1, \theta = 0$) and the rest of parameter values as in eq. (10). That is, it expands the RW case $\phi = 1.0$ in fig. 6 for sample sizes from $T = 30$ to 1 000. It shows that the W2 statistic performs reasonably well both in small as well as in larger sizes. In comparison, the H test statistic (based on the Haar wavelet filter) performs better for smaller sizes whilst Shin's C statistics dominates for larger sizes (about 5% higher power than the wavelet W test statistic in both cases). This is not surprising however, as these tests are designed with I(1) regressions in mind whilst, we may recall, the reported simulations for the wavelet W test statistics were carried out based on a wavelet filter of larger length which may be more suitable for testing higher order cointegration.

8. Size and power comparisons in I(2) systems

Figures 8 and 9 provide evidence on the size and power of tests of the null of cointegration in finite I(2) samples. Again the W statistics were calculated from a unit-level MODWT based on the LA(8) Daubechies filter and asymptotic χ^2 critical values were used to obtain the rejection levels. The reported results were obtained through Monte Carlo simulation using 10 000 replications of size $T = 100$ from

DGP 3. *A bivariate regression model like DGP 2 with $d = 2$.*

The generated $\{z_t\}$ and $\{x_t\}$ are two I(2) processes that, under the null, are cointegrated and, therefore, there exist meaningful relationship in levels (Park and Phillips, 1989). On the other hand, under the alternative, in spite of being related through their second derivatives, any relationship in levels would be totally spurious.

8.1. CI(2,2) vs CI(2,0) systems.

Figure 8 (tables 7 and 9) illustrates the case of a stationary CI relationship between I(2) variables. As far as test size is concerned, the results reported indicate that even

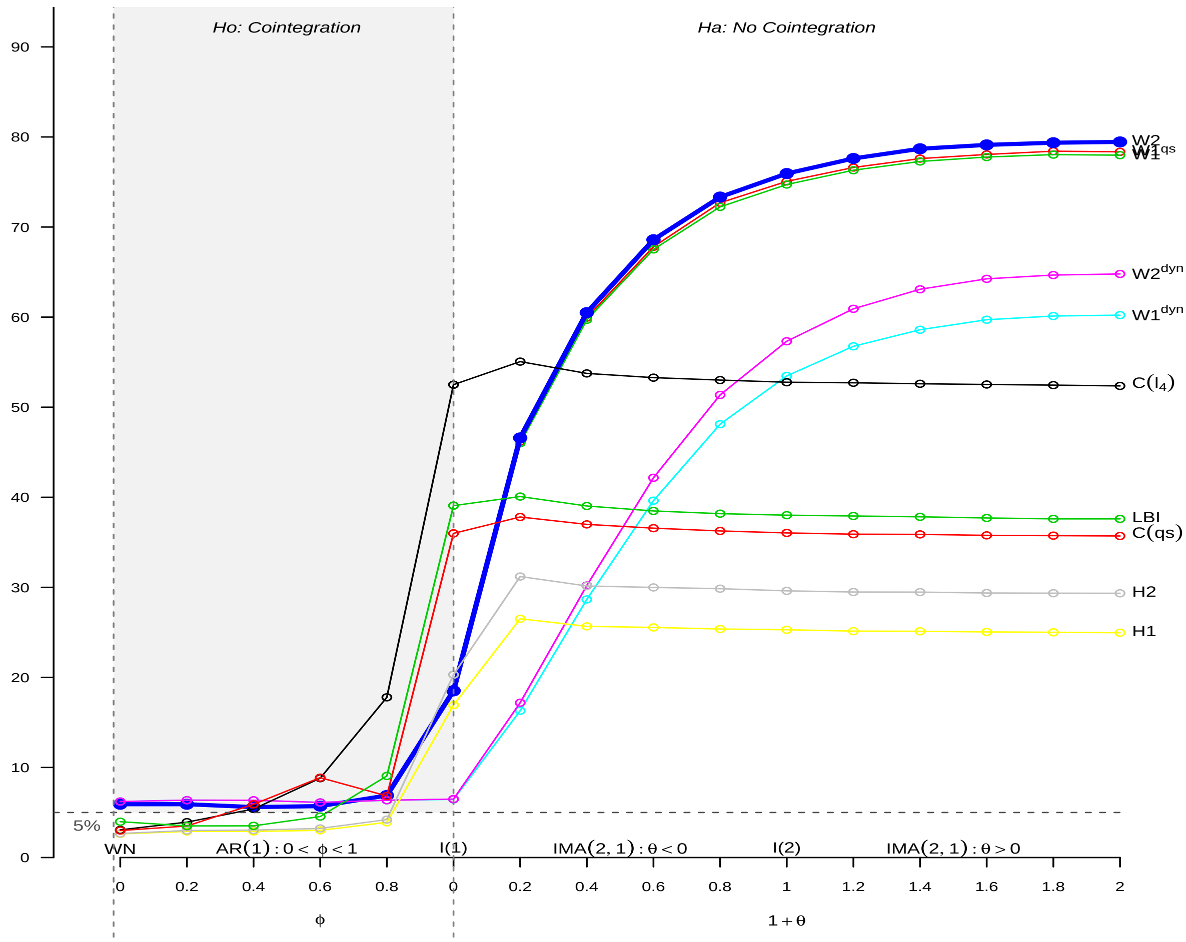


FIGURE 8. CI(2,2): CI Tests Power vs. distance from $\mathcal{H}_0$ (DGP 3; $T = 100$; 5% level).

in the presence of highly autocorrelated errors ($0 < \phi < 0.8$) all tests, except perhaps C(ℓ_4), maintain size distortions well within reasonable levels. However, for values of $\phi > .8$ (note the last but one column of table 7 that corresponds to the case $\phi = 1$ which is just within the cointegration region in fig. 8) the non-wavelet statistics have very large size distortions, rejecting up to more than 52% the (true) cointegration hypothesis in the case of C(ℓ_4).

As far as test power is concerned, the results presented clearly favour the W statistics in the sense that they appear more powerful. For example, for samples of size $T = 100$ at the 5% significance level, W2 will reject up to 79% of the times the (wrong) null hypothesis of a relationship in levels in favour of a (true) relationship in changes. In the same circumstances C(qs) managed just under 36% rejections. As before, we observe that the power of the W statistics is directly related to the 'distance' from $\mathcal{H}_0$, but not so for the other test statistics which actually decrease in power as they get farther away from $\mathcal{H}_0$ (see bold numbers in table 9). On the other hand, W2 appear to be less powerful than just C(ℓ_4) for cases very close to the null ($0 < 1 + \theta < 0.2$) whilst always being more powerful than all the rest. On the whole, as for the I(1) cases, for the I(2) cases the wavelet test statistics exhibit a very satisfactory behaviour: they appear more powerful for a wide range of the alternatives considered and do not reject more often than they should under the null.

8.2. **CI(2,1) vs CI(2,0) systems.** Figure 9 (tables 8 and 9) provides evidence on the size and power of tests of the null of cointegration in finite I(2) samples with I(1) errors. It shows how the residual-based

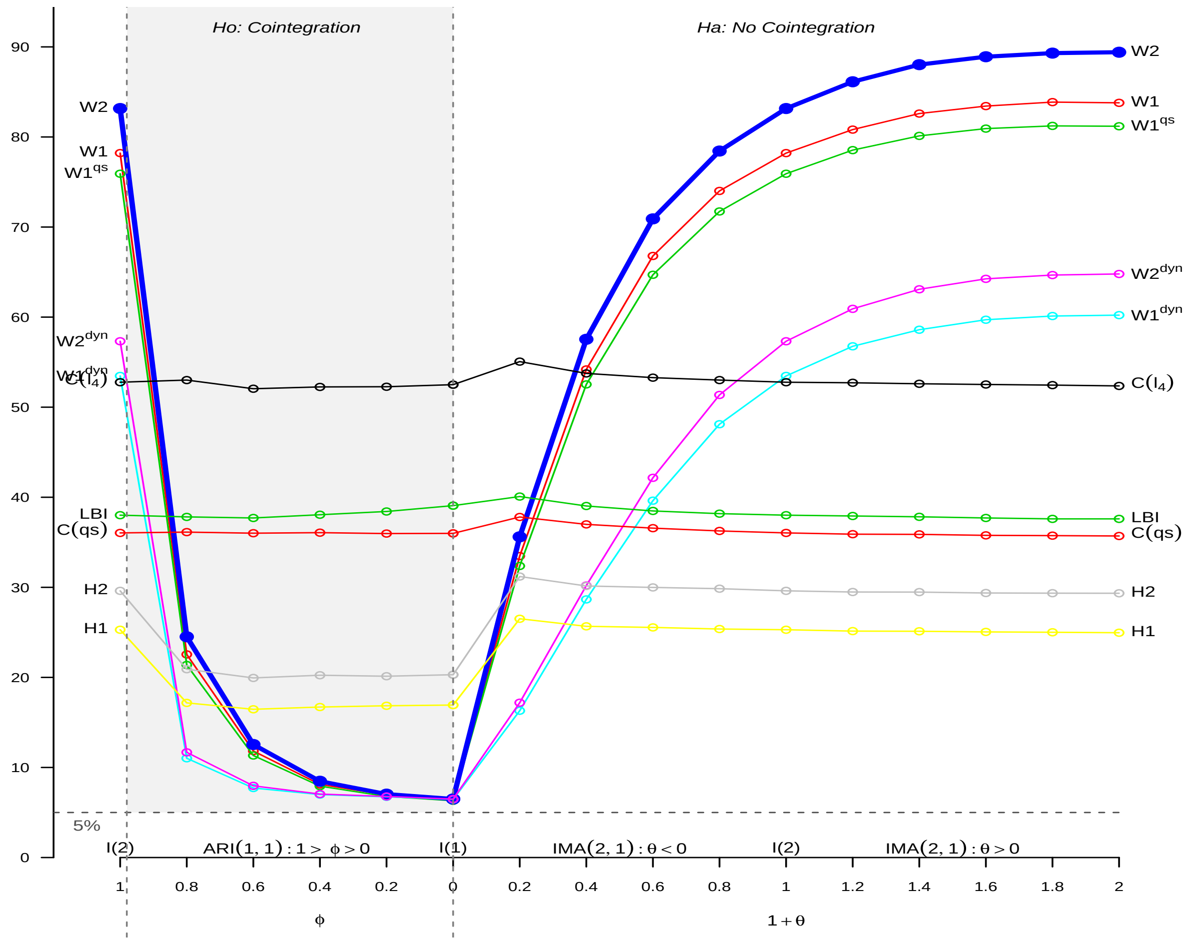
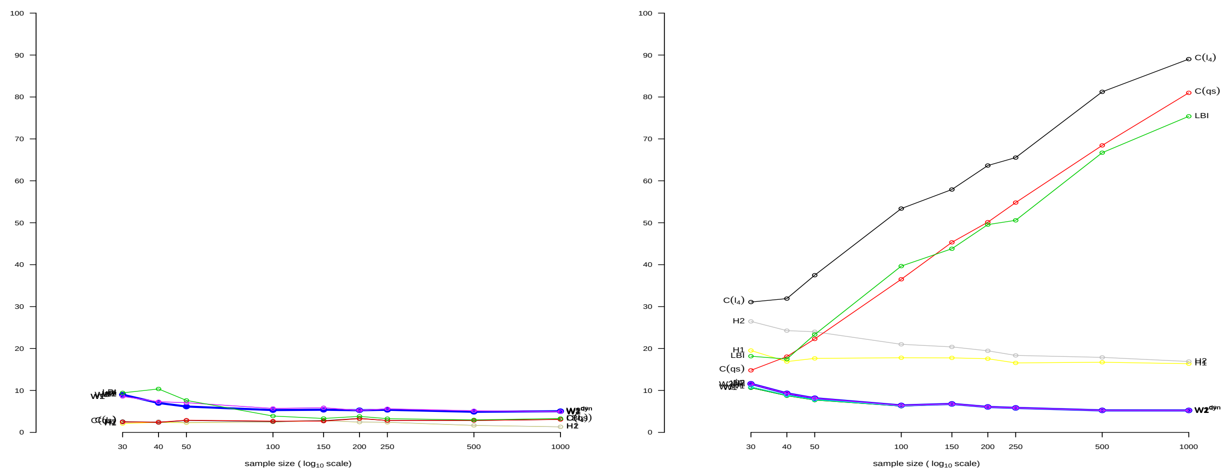


FIGURE 9. CI(2,1): CI Tests Power vs. distance from $\mathcal{H}_0$ (DGP 3; T = 100; 5% level).



CI(2,2) CI(2,1)

FIGURE 10. CI(2,2) & CI(2,1): Size of Cointegration Tests vs. sample size (DGP 3; 5% level).

tests completely fail to discriminate between cointegrated and spurious cases and, therefore, are rather useless in this situation. This is to be expected since these tests are designed with I(0) errors in mind. On the other hand, we may recall, the results for the wavelet W test statistics were based on a wavelet filter of larger length and it seems to be better suited for testing under these conditions.

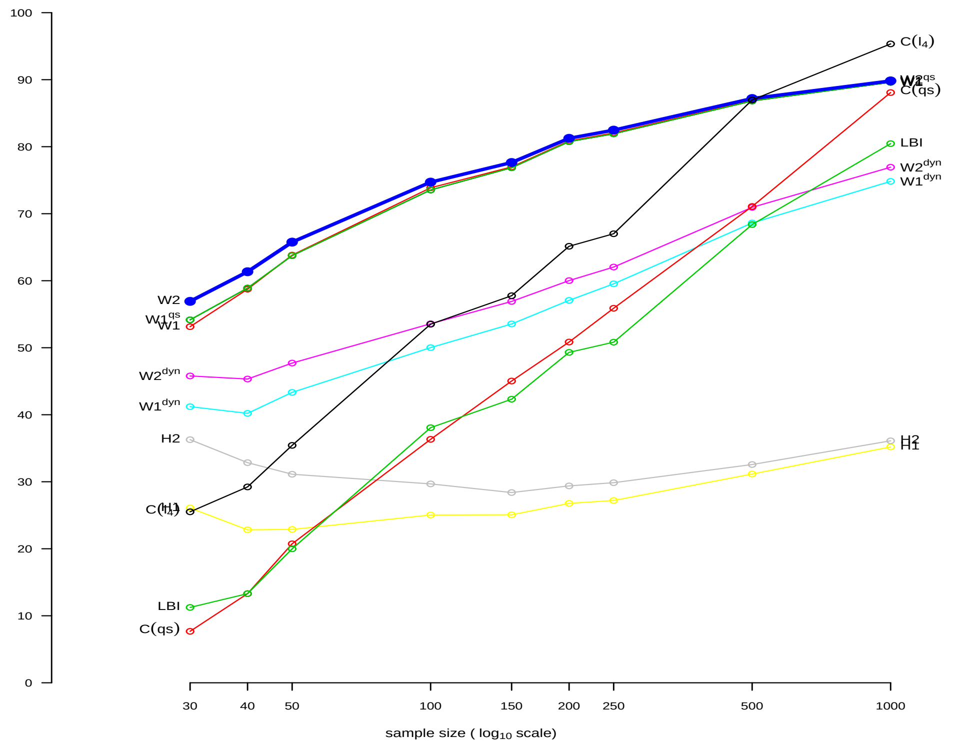


FIGURE 11. CI(2,0): Power of Cointegration Tests vs. sample size (DGP 3; 5% level).

8.3. **Size as power as sample size increases.** Figure 10 (tables 10 and 11) presents evidence on the evolution of the tests' size as the sample size increases. That is, it expands the I(0) case $\phi = 0.0$ and the I(1) case $\phi = 1.0$ in fig. 8 for sample sizes from $T = 30$ to 1 000.

It shows that, while all tests behave well for I(2) cointegrated systems with stationary errors, the size of the residual-based tests tests increases without bound in the CI(2,1) case. In fact we could generalise this tho higher order cases and infer that the residual-based tests are not to be used in I($d \geq 2$) systems if I($d \geq 1$) cointegrating relationships are suspected. On the other hand, the wavelet W statistics behave correctly with empirical sizes tending to the nominal size of the test as $T \to \infty$ in both situations.

Figure 11 (table 12) presents evidence on the evolution of the tests' power as the sample size increases. That is, it expands the I(2) case $\theta = 0.0$ in fig. 8 for sample sizes from $T = 30$ to 1 000.

The results show that the W statistics perform quite well, particularly in small to moderate samples. In fact, they show reasonable power for larger sizes and appear to dominate with samples are large as $T = 500$. In comparison, the residual-based tests do not perform well in small samples. An this, notwithstanding the fact that, as already shown, the residual-based tests are not to be used with I($d \geq 2$) systems if I($d \geq 1$) cointegrating relationships are suspected.

ACKNOWLEDGEMENTS

Research carried out while the author was Associate Member, Nuffield College and Academic Visitor, Programme on Economic Modelling, INET@Oxford, (University of Oxford), on leave from the Dpt. of Econometrics & Statistics (University of the Basque Country — UPV/EHU). Financial support from Basque Government's Dpt. of Education and *UPV/EHU Econometrics Research Group* (Basque Government grant IT-642-13) is gratefully acknowledged.

Appendix

Let us recall that $\tilde{\alpha}$ and $\tilde{\gamma}_j$ denote the OLS estimators of deterministic component coefficients and nuisance parameters obtained from eq. (6a) while $\hat{\beta}_g$ and $\hat{\beta}_h$ are respectively the OLS estimator obtained from the low-pass regression (6) and the OLS estimator obtained from either of the high-pass regressions (7). Expressions for the corresponding estimators of their respective covariance matrices were given in eq. (9).

First of all we want the low-pass model (4) to be rewritten as eq. (6a) where the regressors are strictly exogenous under the null. If $\gamma_s = 0$ for $|s| > m$ we have that $\varepsilon_t = \xi_t$, and the error term in eq. (6a) is uncorrelated with $\{\eta_t\}$ at all leads and lags so that the regressors in $v_{x,jt}$ are strictly exogenous. In general, of course, we cannot assume that $\gamma_s = 0$ for $|s| > m$ fixed; so that, following Saikkonen (1991), in order to work out the asymptotic distribution of the test statistics we require assumption 2 (see also Said and Dickey (1984)). □

Following the convention established by Park and Phillips (1988), it is convenient to define functionals of Brownian motion such as

$$h_0(B, M) = \left(\int MM'\right)^{-1}\left(\int M\mathrm{d}B'\right), \qquad h_a(B, M, \pi) = \left(\int MM'\right)^{-1}\left(\int MB' + \pi\right),$$

$$M(B) = \begin{cases} B(r), & \text{if } g(t) = 0, \\ B^*(r) = B(r) - \int B, & \text{if } g(t) = 1, \\ B^{**}(r) = B(r) + (6r-4)\int B + (6-12r)\int sB, & \text{if } g(t) = (1,t)', \end{cases}$$

that is, $M(B)$ stands for 'standard', 'demeaned' or 'detrended' Brownian motion depending on whether there are no deterministic components in the cointegrating regression ($g(t) = 0$), it is a regression with a constant ($g(t) = 1$) or it is a regression with a linear trend ($g(t) = (1,t)'$), and

$$P(B) = \begin{cases} 1 - [\int B'(\int BB')^{-1}]B(r) & \text{if } g(t) = 1, \\ \begin{bmatrix} 1 - \frac{3}{2}r - [(B' - \frac{3}{2}\int sB')(\int BB' - 3\int sB\int sB')^{-1}](B(r) - 3r\int sB) \\ r - \frac{1}{2} - [(\int sB' - \frac{1}{2}\int B')(\int BB' - \int B\int B')^{-1}](B(r) - \int B) \end{bmatrix} & \text{if } g(t) = (1,t)'. \end{cases}$$

Also we will make use of the following lemma, adapted from (Park and Phillips, 1988, lemma 2.1), about weak convergence results of sample moments of $(g(t), v_{x,jt}, \varepsilon_t)$.

Lemma 1. $T^{-3/2}\sum v_{x,jt} \Rightarrow \int B_\eta$, $T^{-5/2}\sum t v_{x,jt} \Rightarrow \int rB_\eta$, $T^{-2}\sum v_{x,jt}v'_{x,jt} \Rightarrow \int B_\eta B'_\eta$, $T^{-3/2}\sum t\varepsilon_t \Rightarrow \int r\mathrm{d}B_\varepsilon$, $T^{-1}\sum v_{x,jt}\varepsilon'_t \Rightarrow \int B_\eta \mathrm{d}B'_\varepsilon$.

The following lemma expresses in a concise form the limiting distributions of the least squares estimators in eq. (6a)

Lemma 2. *Let $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\gamma}_s$ be the OLS estimators obtained from eq.* (6a). *Then under the null*

$$\text{(a)} \quad \mathrm{diag}(T^{1/2}, T^{3/2})(\tilde{\alpha} - \alpha) \Rightarrow h_0(B_\varepsilon, P(B_\eta)),$$

$$\text{(b)} \quad T(\tilde{\beta} - \beta) \Rightarrow h_0(B_\varepsilon, M(B_\eta)),$$

$$\text{(c)} \quad (T/m)^{1/2}\sum_{s=-m}^{m}(\tilde{\gamma}_s - \gamma_s) = O_p(1).$$

while under the alternative

$$\text{(a}'\text{)} \quad \mathrm{diag}(T^{-1/2}, T^{1/2})(\tilde{\alpha} - \alpha) \Rightarrow h_a(B_\mu, P(B_\eta), 0), \\ = O_p(1)$$

$$\text{(b}'\text{)} \quad (\tilde{\beta} - \beta) \Rightarrow h_a(B_\mu, M(B_\eta), 0), \\ = O_p(1),$$

$$\text{(c}'\text{)} \quad m^{-1/2}\sum_{s=-m}^{m}(\tilde{\gamma}_s - \gamma_s) = O_p(1).$$

Results (a) and (b) are similar to those first obtained by Phillips and Durlauf (1986) and Park and Phillips (1988) (See also (Shin, 1994, lemma 1)). The probabilistic order of magnitude in (c) was obtained by Saikkonen (1991).

Proof: Write eq. (6a) in compact form (assuming $g(t) = (1,t)'$ say):

$$v_{z,jt} = \beta^{*\prime}\, x^*_t + \varepsilon_t, \tag{A-1}$$

$$\text{where} \quad x^{*\prime}_t = (g(t)', v'_{x,jt}, \eta'_{t+m}, \ldots, \eta'_{t-m})$$

$$\text{and} \quad \beta^{*\prime} = (\alpha', \beta', \gamma'_{-m}, \ldots, \gamma'_m).$$

Under the null, we define the scale matrix

$$L = \operatorname{diag}(T^{-1/2}, T^{-3/2}, T^{-1}I_k, T^{-1/2}I_k, \ldots, T^{-1/2}I_k)$$

and, since $(\varepsilon_t - \xi_t) = o_p(T^{-1/2})$ (Saikkonen, 1991, lemma A5), applying lemma 1 obtains

$$L^{-1}(\tilde{\beta}^* - \beta^*) = (L \sum x_t^* x_t^{*\prime} L)^{-1}(L \sum x_t^* \varepsilon_t') \Rightarrow Q_x^{-1} Q_{x\varepsilon}$$

where

$$\tilde{Q}_x = \begin{bmatrix} 1 & T^{-2}\sum t & T^{-3/2}\sum v_{x,jt}' & 0 \\ T^{-2}\sum t & T^{-3}\sum t^2 & T^{-5/2}\sum t v_{x,jt}' & 0 \\ T^{-3/2}\sum v_{x,jt} & T^{-5/2}\sum t v_{x,jt} & T^{-2}\sum v_{x,jt} v_{x,jt}' & 0 \\ 0 & 0 & 0 & T^{-1}\sum \eta_t^* \eta_t^{*\prime} \end{bmatrix}$$

$$\Rightarrow Q_x = \begin{bmatrix} \multicolumn{2}{c}{\multirow{2}{*}{I_2}} & \int B_\eta & 0 \\ & & \int r B_\eta & 0 \\ \int B_\eta' & \int r B_\eta' & \int B_\eta B_\eta' & 0 \\ 0 & 0 & 0 & V_\eta \end{bmatrix}$$

$$\tilde{Q}_{x\varepsilon} = \left[T^{-1/2}\sum \xi_t, T^{-3/2}\sum t\xi_t, T^{-1}\sum \xi_t v_{x,jt}', T^{-1/2}\sum \xi_t \eta_t^{*\prime}\right]'$$

$$\Rightarrow Q_{x\varepsilon} = \left[B_\varepsilon(1), \int r \mathrm{d}B_\varepsilon, \int \mathrm{d}B_\varepsilon B_\eta', C_{\varepsilon\eta}\right]'$$

with $\eta_t^{*\prime} = (\eta_{t+m}', \ldots, \eta_{t-m}')$, $V_\eta = \mathcal{E}(\eta_t^* \eta_t^{*\prime})$ and $C_{\varepsilon\eta} = \operatorname{plim} T^{-1/2}\sum \xi_t \eta_t^{*\prime}$. It is understood that the terms corresponding to $g(t)$ are deleted when not applicable. After inverting matrix Q_x and rearranging we obtain the results given.

Under the alternative the error term in eq. (A-1) gets an extra $v_{u,jt}$ term. We then define the scale matrices

$$L_1 = \operatorname{diag}(T^{1/2}, T^{-1/2}, I_k, I_k, \ldots, I_k)$$
$$L_2 = \operatorname{diag}(T^{-3/2}, T^{-5/2}, T^{-2}I_k, T^{-1}I_k, \ldots, T^{-1}I_k),$$

and applying lemma 1

$$L_1^{-1}(\tilde{\beta}^* - \beta^*) = (L_2 \sum x_t^* x_t^{*\prime} L_1)^{-1}(L_2 \sum x_t^* u_t') \Rightarrow Q_x^{-1} Q_{xu}$$

where Q_x is as before and

$$\tilde{Q}_{xu} = \left[T^{-3/2}\sum v_{u,jt}, T^{-5/2}\sum t v_{u,jt}, T^{-2}\sum v_{u,jt} v_{x,jt}', T^{-1}\sum v_{u,jt}\eta_t^{*\prime}\right]'$$

$$\Rightarrow Q_{xu} = \left[\int B_\mu, \int r B_\mu, \int B_\mu B_\eta', \int B_\mu \mathrm{d}B_\eta', \ldots, \int B_\mu \mathrm{d}B_\eta'\right]'$$

again in the understanding that the terms corresponding to $g(t)$ are deleted when not applicable. Inverting matrix Q_x in each case and rearranging leads us to the results given. □

Lemma 3. *the OLS estimators of β in eqs.* (6a) *and* (6b) *are asymptotically equivalent under the null, but differ under the alternative by an amount that is bounded in probability.*

Proof: Trivially under the null since from $\tilde{\alpha} \xrightarrow{p} \alpha$, and $\tilde{\gamma}_s \xrightarrow{p} \gamma_s$, we have that $y_t \overset{def}{=} v_{z,jt} - \tilde{\alpha}' g(t) - \sum_{s=-m}^{m} \tilde{\gamma}_s' \eta_{t-s} \xrightarrow{p} v_{z,jt} - \alpha' g(t) - \sum_{s=-m}^{m} \gamma_s' \eta_{t-s} = \beta' v_{x,jt} + \varepsilon_t$, and the first part of the lemma follows.

Under the alternative, from lemma 2, $T^{-1/2}(\tilde{\alpha}_0 - \alpha_0) = O_p(1)$, $T^{1/2}(\tilde{\alpha}_1 - \alpha_1) = O_p(1)$, and $\sum_{s=-m}^{m}(\tilde{\gamma}_s - \gamma_s) = O_p(m^{1/2})$, and we have that $y_t \overset{def}{=} v_{z,jt} - \tilde{\alpha}' g(t) - \sum_{s=-m}^{m} \tilde{\gamma}_s' \eta_{t-s} = \beta' v_{x,jt} + (u_t^* + \varepsilon_t)$ where

$$u_t^* = v_{u,jt} - (\tilde{\alpha} - \alpha)' g(t) - \sum_{s=-m}^{m} (\tilde{\gamma}_s - \gamma_s)' \eta_{t-s}. \tag{A-2}$$

Then

$$\begin{aligned} T^{-2}\sum v_{x,jt} u_t^{*\prime} &= T^{-2}\sum v_{x,jt} v_{u,jt}' - T^{-3/2}\sum v_{x,jt} T^{-1/2}(\tilde{\alpha}_0 - \alpha_0) - T^{-2}\sum t v_{x,jt} T^{1/2}(\tilde{\alpha}_1 - \alpha_1) \\ &\quad - \sum_{s=-m}^{m} (T^{-2}\sum v_{x,jt}\eta_{t-s}')(\tilde{\gamma}_s - \gamma_s) \\ &\Rightarrow \int M(B_\eta) B_\mu' + \pi_{\eta\mu}', \end{aligned} \tag{A-3}$$

which, furthermore, implicitly defines $\pi_{\eta\mu}$ needed later, and the lemma follows. Also □

$$
\begin{aligned}
T^{-2}\sum u_t^* u_t^{*\prime} &= T^{-2}\sum (v_{u,jt} - (\tilde{\alpha}-\alpha)' g(t))(v_{u,jt}' - g(t)'(\tilde{\alpha}-\alpha)) \\
&= T^{-2}\sum v_{u,jt} v_{u,jt}' - T^{-2}\Big[\sum v_{u,jt} g(t)'\Big](\tilde{\alpha}-\alpha) - (\tilde{\alpha}-\alpha)' T^{-2}\Big[\sum g(t) v_{u,jt}'\Big] \\
&+ (\tilde{\alpha}-\alpha)' T^{-2}\Big[\sum g(t) g(t)'\Big](\tilde{\alpha}-\alpha) \\
&\Rightarrow \int B_v B_\mu' + \pi_\mu.
\end{aligned}
$$

where $\pi_\mu = 4 - \int B_\mu - \int B_\mu' - \int r B_\mu - \int r B_\mu'$. □

The following lemma establishes the asymptotic distributions of the OLS low-pass regression estimator and its variance

Lemma 4. *Under the null of cointegration*

$$
\begin{aligned}
T\operatorname{vec}(\hat{\beta}_g - \beta) &\Rightarrow \mathcal{N}(0, V_g), \\
T^2 \hat{V}_g &\Rightarrow V_g
\end{aligned}
$$

where $V_g = \Omega_\varepsilon \otimes \Big(\int M(B_\eta) M(B_\eta)'\Big)^{-1}$ *; while under the alternative*

$$
\begin{aligned}
(\hat{\beta}_g - \beta) &\Rightarrow h_a(B_\mu, M(B_\eta), \pi_{\eta\mu}) \\
&= O_p(1) \\
T\hat{V}_g &\Rightarrow \Theta_u \otimes \Big(\int M(B_\eta) M(B_\eta)'\Big)^{-1} \\
&= O_p(1)
\end{aligned}
$$

where $\Theta_u = (\int B_v B_\mu' + \pi_\mu) - h_a(B_\mu, M(B_\eta), \pi_{\eta\mu})(\int M(B_\eta) B_\mu' + \pi_{\eta\mu}')$,

so that $\hat{\beta}_g$ is a T-consistent estimator under the null but there is a stochastic asymptotic bias under the alternative.

Proof: writing model (6b) in matrix form

$$Y = X\beta + E + \delta\, U^*$$

where $Y' = (y_1 \dots y_T)$, $X' = (v_{x,j1} \dots v_{x,jT})$, $U^{*\prime} = (u_1^* \dots u_T^*)$ (with u_t^* as in eq. (A-2)) and $E' = (\varepsilon_1 \dots \varepsilon_T)$, we obtain under the null

$$
\begin{aligned}
T(\hat{\beta}_g - \beta) = (T^{-2}X'X)^{-1}(T^{-1}X'E) &\Rightarrow \Big(\int M(B_\eta) M(B_\eta)'\Big)^{-1}\Big(\int M(B_\eta)\mathrm{d}B_\varepsilon'\Big) = h_0(B_\varepsilon, M(B_\eta)) \\
T\operatorname{vec}(\hat{\beta}_g - \beta) &\sim \mathcal{N}\Big(0, \Omega_\varepsilon \otimes \Big(\int M(B_\eta) M(B_\eta)'\Big)^{-1}\Big),
\end{aligned}
$$

where the normal distribution is obtained because $M(B_\eta)$ is a vector process independent of B_ε (Park and Phillips, 1988, lemma 5.1).

On the other hand, from eq. (9)

$$T^2\hat{V}_g = \hat{\Omega}_\varepsilon \otimes (T^{-2}X'X)^{-1} \Rightarrow \Omega_\varepsilon \otimes \Big(\int M(B_\eta) M(B_\eta)'\Big)^{-1}$$

so that $\hat{V}_g$ converges to V_g at a very fast rate.

Under the alternative $T(\hat{\beta}_g - \beta)$ clearly diverges but using eq. (A-3)

$$
\begin{aligned}
(\hat{\beta}_g - \beta) = (T^{-2}X'X)^{-1}[T^{-2}X'(E + U^*)] &\Rightarrow h_a(B_\mu, M(B_\eta), \pi_{\eta\mu}) = \Theta_g \\
&= O_p(1)
\end{aligned}
$$

so that $\hat{\beta}_g$ is inconsistent, as well as its variance estimator

$$T\hat{V}_g = T^{-1}\hat{\Omega}_{\hat{u}} \otimes (T^{-2}X'X)^{-1}$$

$$
\begin{aligned}
T^{-1}\hat{\Omega}_{\hat{u}} &\to T^{-2}\hat{U}^{*\prime}\hat{U}^* \\
&= T^{-2}U'U - (\hat{\beta}_g - \beta)' T^{-2}X'X(\hat{\beta}_g - \beta) \\
&\Rightarrow (\int B_v B_\mu' + \pi_\mu) - h_a(B_\mu, M(B_\eta), \pi_{\eta\mu})(\int M(B_\eta) B_\mu' + \pi_{\eta\mu}') = \Theta_u
\end{aligned}
$$

which leads to the last result in the lemma. □

The next lemma establishes the asymptotic distributions of the high-pass regression OLS estimator and its variance. For convenience, the respective variances (at $s = 0$) will be written as

$$\Sigma = \begin{bmatrix} \Sigma_\mu & \Sigma'_{\eta\mu} \\ \Sigma_{\eta\mu} & \Sigma_\eta \end{bmatrix} \equiv C(0),$$

with $C(s)$ as defined in assumption 1.

Lemma 5. *Let $V_h = (I_n \otimes \Sigma_\eta^{-1})R(I_n \otimes \Sigma_\eta^{-1})$. Under the null of cointegration*

$$T^{1/2}\text{vec}(\hat{\beta}_h - \beta) \Rightarrow \mathcal{N}(0, V_h),$$
$$T\hat{V}_h \xrightarrow{p} V_h,$$

with $R = \text{plim}\, T^{-1}(I_n \otimes N')V_{w\varepsilon}(I_n \otimes N)$ and $N' = (\eta_1, \ldots, \eta_{T-1})$; while under the alternative

$$\begin{aligned}(\hat{\beta}_h - \beta) &\xrightarrow{p} C_{\eta\mu}(1)\Sigma_\eta^{-1} \\ &= O_p(1) \\ T\hat{V}_h &\xrightarrow{p} V_h,\end{aligned}$$

with $R = \text{plim}\, T^{-1}(I_n \otimes N')V_\mu(I_n \otimes N)$.

Therefore, $\hat{\beta}_h$ is a $\sqrt{T}$-consistent estimator under the null although, in general, there may be a nonstochastic asymptotic bias under the alternative, in contrast with the stochastic nature of the bias in the low-pass regression.

Proof: from model (7) in matrix form

$$W_y = N\beta + W_\varepsilon + \delta\, W_u^*$$

where $(W_y)' = (w_{y1}, \ldots, w_{yT})$ and $(W_u^*)' = (w_{u1}^*, \ldots, w_{uT}^*)$ with $w_{ut}^* = \mu_t - (\tilde{\alpha}_1 - \alpha_1)' - \sum_{s=-m}^{m}(\tilde{\gamma}_s - \gamma_s)'\Delta\eta_{t-s}$. We can express the OLS estimator as

$$(\hat{\beta}_h - \beta) = (N'N)^{-1}N'\left[W_\varepsilon + \delta\, W_u^*\right] \tag{A-4}$$

Under the null ($\delta = 0$) we are interested in the form of the limiting distribution of

$$\text{vec}(T^{-1/2}N'W_\varepsilon) = T^{-1/2}(I_n \otimes N')w_\varepsilon \tag{A-5}$$

where $w_\varepsilon = \text{vec}W'_\varepsilon = (w'_{\varepsilon 1}, \ldots, w'_{\varepsilon T})'$ and $\mathcal{E}(w_\varepsilon w'_\varepsilon) = V_{w\varepsilon}$ is the $(nT \times nT)$ covariance matrix of process $\{w_{\varepsilon t}\}$.

Let us define $\bar{\varepsilon} = V_{w\varepsilon}^{-1/2}w_\varepsilon$, where $V_{w\varepsilon} = V_{w\varepsilon}^{1/2}(V_{w\varepsilon}^{1/2})'$. Then

$$\mathcal{E}(\bar{\varepsilon}\bar{\varepsilon}') = V_{w\varepsilon}^{1/2}\mathcal{E}(w_\varepsilon w'_\varepsilon)(V_{w\varepsilon}^{1/2})' = I_{nT}$$

so that $\bar{\varepsilon}_t \sim \text{iid}(0, 1)$. Let us further define the $(kn \times Tn)$ matrix $\bar{N}' = (I_n \otimes N')V_{w\varepsilon}^{1/2}$ with

$$\text{plim}\, T^{-1}\bar{N}'\bar{N} = R = \text{plim}\, T^{-1}(I_n \otimes N')V_{w\varepsilon}(I_n \otimes N),$$

and $\mathcal{E}(\bar{N}_t\bar{\varepsilon}_t) = 0, \forall t$. Therefore, applying the Mann-Wald theorem

$$\text{plim}\, T^{-1}\bar{N}'\bar{\varepsilon} = 0, \qquad T^{-1/2}\bar{N}'\bar{\varepsilon} \Rightarrow \mathcal{N}(0, R).$$

Substituting in eq. (A-5) we get

$$\text{vec}(T^{-1/2}N'W_\varepsilon) = T^{-1/2}\bar{N}'\bar{\varepsilon} \Rightarrow \mathcal{N}(0, R).$$

Finally, let $\text{plim}\, T^{-1}N'N = \Sigma_\eta > 0$ in eq. (A-4). We may then write

$$\sqrt{T}\text{vec}(\hat{\beta}_h - \beta) \Rightarrow \mathcal{N}(0, (I_n \otimes \Sigma_\eta^{-1})R(I_n \otimes \Sigma_\eta^{-1})).$$

On the other hand, from eq. (9)

$$\begin{aligned}T\hat{V}_h &= [I_n \otimes (T^{-1}N'N)^{-1}]T^{-1}\hat{R}[I_n \otimes (T^{-1}N'N)^{-1}], \\ T^{-1}\hat{R} &= (I_n \otimes T^{-1/2}N')\hat{V}_{w\varepsilon}(I_n \otimes T^{-1/2}N),\end{aligned}$$

where $\hat{V}_{w\varepsilon} \xrightarrow{p} V_{w\varepsilon}$. Hence $T^{-1}\hat{R} \xrightarrow{p} R$ and $T\hat{V}_h \xrightarrow{p} V_h$.

Under the alternative ($\delta = 1$) and eq. (A-4) becomes

$$(\hat{\beta}_h - \beta) = (N'N)^{-1}N'\left[W_\varepsilon + W_u^*\right]$$

where

$$\begin{aligned}T^{-1}\sum \eta_t(w^{*\prime}_{ut}+w'_{\varepsilon t}) &= T^{-1}\sum \eta_t\mu'_t - T^{-3/2}\sum \eta_t T^{1/2}(\tilde{\alpha}_1-\alpha_1)\\ &\quad -\sum_{s=-m}^{m}(T^{-1}\sum \eta_t\Delta\eta_{t-s})(\tilde{\gamma}_s-\gamma_s)+T^{-1}\sum \eta_t w'_{\varepsilon t}\\ &\xrightarrow{p} \mathcal{E}(\eta_t\mu'_t)=C_{\eta\mu}(1).\end{aligned}$$

(Note that the bias from the inconsistency of $\tilde{\alpha}_1$ or $\tilde{\gamma}_s$ under the alternative disappears asymptotically). Therefore

$$(\hat{\beta}_h-\beta)\xrightarrow{p} C_{\eta\mu}(1)\Sigma_\eta^{-1}=\Theta_h.$$

On the other hand

$$\begin{aligned}T\hat{V}_h &= [I_n\otimes(T^{-1}N'N)^{-1}]T^{-1}\hat{R}[I_n\otimes(T^{-1}N'N)^{-1}]\\ T^{-1}\hat{R} &= (I_n\otimes T^{-1/2}N')\hat{V}_\mu(I_n\otimes T^{-1/2}N')\end{aligned}$$

where $\hat{V}_\mu \xrightarrow{p} V_\mu$, the covariance matrix of process μ_t. Hence $T^{-1}\hat{R}\xrightarrow{p} R_\mu$ and $T\hat{V}_h\xrightarrow{p} V_h$. □

Lemma 6. *In the multivariate regression model* (3) *under the null of cointegration*

$$\sqrt{T}c \Rightarrow \mathcal{N}(0,V_h),$$

$$Tc'V_h^{-1}c\Rightarrow\chi^2(nk),\qquad Tc'(V_h+T^{-1}V_g)^{-1}c\Rightarrow\chi^2(nk),$$

where V_g and V_h are as defined in the preceding lemmas.

Lemma 6 justifies the two suggested wavelet test statistics. Obviously, the effect of correcting the asymptotic variance of c adding the $T^{-1}V_g$ term disappears asymptotically but it nevertheless may provide a better approximation in small samples.

Proof:

$$\begin{aligned}\sqrt{T}c &= \mathrm{vec}\,\sqrt{T}(\hat{\beta}_h-\beta)-T^{-1/2}\mathrm{vec}T(\hat{\beta}_g-\beta)\\ &= (1,-T^{-1/2})\begin{bmatrix}\sqrt{T}\mathrm{vec}(\hat{\beta}_h-\beta)\\ T(\hat{\beta}_g-\beta)\end{bmatrix}.\end{aligned}$$

From lemmas 4 and 5 $\sqrt{T}\mathrm{vec}(\hat{\beta}_h-\beta)\Rightarrow\mathcal{N}(0,V_h)$, $T\mathrm{vec}(\hat{\beta}_g-\beta)\Rightarrow\mathcal{N}(0,V_g)$. Therefore, their joint distribution is also Gaussian. Let C_{hg} denote the asymptotic covariance matrix of $\sqrt{T}\mathrm{vec}(\hat{\beta}_h-\beta)$ and $T\mathrm{vec}(\hat{\beta}_g-\beta)$. Then we may write

$$\begin{pmatrix}\sqrt{T}\mathrm{vec}(\hat{\beta}_h-\beta)\\ T\mathrm{vec}(\hat{\beta}_g-\beta)\end{pmatrix}\Rightarrow\mathcal{N}\left[0,\begin{pmatrix}V_h & C'_{hg}\\ C_{hg} & V_g\end{pmatrix}\right]$$

and hence the limiting distribution of $\sqrt{T}c$ has zero mean and variance

$$\begin{pmatrix}1 & -T^{-1/2}\end{pmatrix}\begin{pmatrix}V_h & C'_{hg}\\ C_{hg} & V_g\end{pmatrix}\begin{pmatrix}1\\ -T^{-1/2}\end{pmatrix}=V_h-T^{-1/2}(C_{hg}+C'_{hg})+T^{-1}V_g$$

where the last two terms disappear asymptotically and the proposition follows. □

Proof of proposition 1: Under the null, from lemmas 4 to 6 the result follows trivially.

Under the alternative, from lemmas 4 and 5 and defining $\Theta_c=\Theta_h-\Theta_g$

$$c=\mathrm{vec}(\hat{\beta}_h-\beta)-\mathrm{vec}(\hat{\beta}_g-\beta)\Rightarrow\mathrm{vec}\Theta_c.$$

Since Θ_g is a stochastic matrix but Θ_h is a matrix of constants, we have that $\Pr(\Theta_g=\Theta_h)=0$ and $c\neq 0$ a.s. In fact, it also follows that under the alternative c will have the same asymptotic distribution as $-\hat{\beta}_g$ except for a shift in the mean of value $\mathrm{vec}\Theta_h$. Then from lemma 5, $T\hat{V}_h\xrightarrow{p}(I_n\otimes\Sigma_\eta^{-1})R(I_n\otimes\Sigma_\eta^{-1})$; therefore

$$\begin{aligned}T^{-1}\mathrm{W2} &= T^{-1}c'\hat{V}_h^{-1}c=c'(T\hat{V}_h)^{-1}c\Rightarrow[\mathrm{vec}(\Theta_c\Sigma_\eta^{-1})]'R[\mathrm{vec}(\Theta_c\Sigma_\eta^{-1})]\\ &= O_p(1),\end{aligned}$$

and from lemma 4, $T\hat{V}_g\xrightarrow{p}\Theta_u\otimes\left(\int M(B_\eta)M(B_\eta)'\right)^{-1}$; therefore

$$\begin{aligned}T^{-1}\mathrm{W1} &= T^{-1}c'(\hat{V}_h+\hat{V}_h)^{-1}c=c'(T\hat{V}_h+T\hat{V}_g)^{-1}c\\ &\Rightarrow(\mathrm{vec}\Theta_c)'\Big[(I_n\otimes\Sigma_\eta^{-1})R(I_n\otimes\Sigma_\eta^{-1})+\Theta_u\otimes\Big(\int M(B_\eta)M(B_\eta)'\Big)^{-1}\Big]^{-1}\mathrm{vec}\Theta_c\\ &= O_p(1).\end{aligned}$$

As $\int MM'>0$ and $\Theta_u>0$ we also have that $\mathrm{plim}\,T^{-1}W1<\mathrm{plim}\,T^{-1}W2$. □

References

Andrews, D. W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Brillinger, D. R., 1975. Time Series: Data Analysis and Theory. Holt, Rinehart and Wilson, New York.

Crowley, P. M., 2007. A guide to wavelets for economists. Journal of Economic Surveys 21 (2), 207–267.

Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets. Comm. Pure & Appl. Math 41 (7), 909–996.

Daubechies, I., 1992. Ten lectures on wavelets. Society for Industrial Mathematics.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

Engle, R. F., Granger, C. W. J., 1987. Cointegration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Fan, Y., Gençay, R., 2010. Unit root tests with wavelets. Econometric Theory 26, 1305–1331.

Fernández-Macho, J., 2013. A test for the null of multiple cointegrating vectors. Discussion Paper Series 657, Dpt. of Economics, University of Oxford.

Gençay, R., Selçuk, F., Whitcher, B., 2002. An introduction to wavelets and other filtering methods in finance and economics. Academic Press.

Gonzalo, J., 1994. Comparison of five alternative methods of estimating long-run equilibrium relationship. Journal of Econometrics 60, 203–233.

Granger, C. W., 1966. The typical spectral shape of an economic variable. Econometrica 34 (1), 150–161.

Granger, C. W. J., 1981. Some properties of time series data and their use in econometric model specification. Journal of Econometrics 16, 121–130.

Granger, C. W. J., 1983. Co-integrated variables and error-correcting models. Discussion Paper 83-13, UCSD.

Hansen, B., Phillips, P. C. B., 1990. Estimation and inference in models of cointegration: A simulation study. Vol. 8 of Advances in Econometrics: Co-integration, Spurious Regressions and Unit Roots. Greenwich, JAI Press, pp. 225–248.

Haug, A. A., 1996. Tests for cointegration. A Monte Carlo comparison. Journal of Econometrics 71, 89–115.

Hausman, J. A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics 54, 159–178.

Levy, D., 2000. Cointegration in frequency domain. Journal of Time Series Analysis 23 (3), 333–339.

Leybourne, S. J., McCabe, B. P. M., 1994. A simple test for cointegration. Oxford Bulletin of Economics and Statistics 56, 97–103.

Nelson, C. R., Plosser, C. I., 1982. Trends versus random walks in macroeconomic time series. Journal of Monetary Economics 10, 139–62.

Park, J. Y., Phillips, P. C. B., 1988. Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4, 468–497.

Park, J. Y., Phillips, P. C. B., 1989. Statistical inference in regressions with integrated processes: Part 2. Econometric Theory 5, 95–131.

Percival, D., Walden, A., 2000. Wavelet methods for time series analysis. Cambridge University Press, reprinted 2006.

Phillips, P. C. B., 1991. Optimal inference in cointegrated systems. Econometrica 59 (2), 283–306.

Phillips, P. C. B., Durlauf, S. N., 1986. Multiple time series regression with integrated processes. Review of Economics and Statistics 53, 473–496.

Phillips, P. C. B., Hansen, B. E., 1990. Statistical inference in instrumental variables regression with I(1) processes. Review of Economics and Statistics 57, 99–125.

Ramsey, J. B., 1999. The contribution of wavelets to the analysis of economic and financial data. Phil. Trans. R. Soc. A. Math. Phys. Eng. Sci. 357 (1760), 2593–2606.

Said, S. E., Dickey, D. A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, 599–607.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21.

Schleicher, C., 2002. An introduction to wavelets for economists. Working Papers 02-3, Bank of Canada. URL http://ideas.repec.org/p/bca/bocawp/02-3.html

Shin, Y., 1994. A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.

Stock, J., 1987. Asymptotic properties of least-squares estimators of cointegrating vectors. Econometrica 55, 1035–56.

Venables, W. N., Ripley, B. D., 2002. Modern Applied Statistics with S, 4th Edition. Springer, New York.

Table 1. Quantiles for $W1$ and $W2$ Statistics ($n = 1$).

		$W1$ statistic				$W2$ statistic			
T	k	*mean*	*0.90*	*0.95*	*0.99*	*mean*	*0.90*	*0.95*	*0.99*
50	*1*	*0.96*	2.585	3.683	6.348	*1.00*	2.682	3.807	6.521
	2	*1.93*	4.425	5.708	8.808	*2.04*	4.670	6.009	9.398
	3	*2.86*	5.863	7.291	10.569	*3.08*	6.311	7.882	11.489
	4	*3.81*	7.302	8.870	12.145	*4.16*	7.925	9.648	13.544
	5	*4.79*	8.696	10.522	14.374	*5.27*	9.532	11.519	15.876
100	*1*	*0.96*	2.601	3.652	6.453	*0.98*	2.638	3.746	6.601
	2	*1.93*	4.404	5.706	8.686	*1.99*	4.561	5.899	9.023
	3	*2.86*	5.916	7.447	10.358	*2.99*	6.200	7.764	10.786
	4	*3.80*	7.288	8.877	12.313	*4.02*	7.674	9.338	13.124
	5	*4.73*	8.686	10.281	13.744	*5.05*	9.189	10.883	14.659
150	*1*	*1.00*	2.674	3.840	6.706	*1.01*	2.712	3.887	6.789
	2	*1.95*	4.455	5.789	9.016	*1.99*	4.553	5.960	9.210
	3	*2.91*	6.104	7.509	10.790	*3.01*	6.303	7.752	11.106
	4	*3.88*	7.522	9.192	12.553	*4.04*	7.819	9.601	13.066
	5	*4.77*	8.699	10.280	13.748	*5.00*	9.181	10.813	14.346
250	*1*	*0.99*	2.720	3.745	6.652	*1.00*	2.739	3.775	6.688
	2	*1.95*	4.526	5.958	8.843	*1.98*	4.603	6.041	8.948
	3	*2.94*	6.067	7.773	11.182	*3.01*	6.209	7.948	11.426
	4	*3.88*	7.659	9.346	13.206	*3.99*	7.878	9.605	13.515
	5	*4.80*	8.828	10.569	14.698	*4.95*	9.109	10.948	15.063
500	*1*	*0.99*	2.685	3.812	6.280	*0.99*	2.698	3.826	6.289
	2	*1.99*	4.581	5.931	8.942	*2.01*	4.620	5.982	8.995
	3	*2.95*	6.172	7.648	11.140	*2.98*	6.228	7.754	11.238
	4	*3.91*	7.634	9.387	12.756	*3.96*	7.753	9.534	12.955
	5	*4.87*	9.080	10.831	14.547	*4.96*	9.271	11.010	14.835
∞	*1*	*1.00*	2.706	3.841	6.635	*1.00*	2.706	3.841	6.635
	2	*2.00*	4.605	5.991	9.210	*2.00*	4.605	5.991	9.210
	3	*3.00*	6.251	7.815	11.345	*3.00*	6.251	7.815	11.345
	4	*4.00*	7.779	9.488	13.277	*4.00*	7.779	9.488	13.277
	5	*5.00*	9.236	11.070	15.086	*5.00*	9.236	11.070	15.086

Simulations from DGP 1 using 10 000 replications each. The values for $T = \infty$ correspond to quantiles of the chi squared distribution with $1k$ d.f.

Table 2. Quantiles for $W1$ and $W2$ Statistics ($n = 2$).

		$W1$ statistic				$W2$ statistic			
T	k	*mean*	*0.90*	*0.95*	*0.99*	*mean*	*0.90*	*0.95*	*0.99*
50	*1*	*1.99*	4.544	5.791	9.047	*2.09*	4.739	6.028	9.539
	2	*3.94*	7.585	9.274	12.781	*4.20*	8.089	9.876	13.816
	3	*5.84*	10.174	12.005	16.267	*6.32*	11.049	13.157	17.572
	4	*7.77*	12.666	14.710	19.695	*8.53*	13.987	16.338	21.923
100	*1*	*1.96*	4.488	5.883	8.984	*2.01*	4.569	5.989	9.168
	2	*3.90*	7.604	9.323	12.695	*4.04*	7.897	9.614	13.189
	3	*5.77*	10.135	11.892	15.807	*6.05*	10.616	12.411	16.467
	4	*7.65*	12.635	14.624	18.797	*8.10*	13.362	15.429	19.809
150	*1*	*1.97*	4.560	5.936	9.164	*2.00*	4.612	6.013	9.358
	2	*3.93*	7.617	9.414	12.727	*4.03*	7.825	9.630	13.141
	3	*5.87*	10.387	12.110	15.814	*6.07*	10.789	12.508	16.247
	4	*7.70*	12.789	14.793	18.567	*8.03*	13.339	15.416	19.381
250	*1*	*2.00*	4.570	5.968	9.379	*2.01*	4.603	6.004	9.416
	2	*3.99*	7.764	9.422	13.034	*4.06*	7.915	9.569	13.274
	3	*5.91*	10.457	12.435	16.943	*6.04*	10.702	12.673	17.188
	4	*7.87*	13.111	15.290	20.117	*8.09*	13.444	15.689	20.570
500	*1*	*1.99*	4.551	5.951	9.284	*2.00*	4.574	5.975	9.297
	2	*3.99*	7.757	9.350	12.899	*4.02*	7.822	9.453	13.022
	3	*5.85*	10.319	12.255	16.142	*5.92*	10.421	12.391	16.364
	4	*7.91*	13.267	15.347	19.611	*8.03*	13.462	15.589	19.996
∞	*1*	*2.00*	4.605	5.991	9.210	*2.00*	4.605	5.991	9.210
	2	*4.00*	7.779	9.488	13.277	*4.00*	7.779	9.488	13.277
	3	*6.00*	10.645	12.592	16.812	*6.00*	10.645	12.592	16.812
	4	*8.00*	13.362	15.507	20.090	*8.00*	13.362	15.507	20.090

Simulations from DGP 1 using 10 000 replications each. The values for $T = \infty$ correspond to quantiles of the chi squared distribution with $2k$ d.f.

TABLE 3. Quantiles for $W1$ and $W2$ Statistics ($n = 3$).

		$W1$ statistic				$W2$ statistic			
T	k	*mean*	*0.90*	*0.95*	*0.99*	*mean*	*0.90*	*0.95*	*0.99*
50	*1*	*3.06*	6.309	7.811	11.315	*3.19*	6.587	8.184	11.952
	2	*5.95*	10.383	12.299	17.029	*6.36*	11.112	13.062	18.174
	3	*8.88*	14.366	16.463	21.627	*9.70*	15.691	18.246	23.334
100	*1*	*2.94*	6.156	7.654	11.215	*3.01*	6.297	7.808	11.531
	2	*5.88*	10.335	12.134	16.055	*6.09*	10.704	12.603	16.596
	3	*8.74*	14.183	16.162	20.449	*9.17*	14.857	17.029	21.552
150	*1*	*2.99*	6.266	7.845	11.083	*3.04*	6.366	7.981	11.294
	2	*5.93*	10.559	12.361	16.480	*6.08*	10.846	12.671	16.864
	3	*8.80*	14.276	16.250	20.419	*9.11*	14.763	16.813	21.086
250	*1*	*3.03*	6.296	7.965	11.275	*3.05*	6.355	8.021	11.399
	2	*5.94*	10.552	12.367	16.348	*6.04*	10.731	12.569	16.581
	3	*8.89*	14.482	16.585	21.323	*9.09*	14.773	17.062	21.770
500	*1*	*2.97*	6.113	7.681	11.377	*2.98*	6.143	7.748	11.387
	2	*5.87*	10.425	12.231	15.990	*5.92*	10.519	12.308	16.127
	3	*8.92*	14.550	16.696	20.640	*9.02*	14.695	16.905	20.814
∞	*1*	*3.00*	6.251	7.815	11.345	*3.00*	6.251	7.815	11.345
	2	*6.00*	10.645	12.592	16.812	*6.00*	10.645	12.592	16.812
	3	*9.00*	14.684	16.919	21.666	*9.00*	14.684	16.919	21.666

Simulations from DGP 1 using 10 000 replications each. The values for $T = \infty$ correspond to quantiles of the chi squared distribution with $3k$ d.f.

TABLE 4. CI(1,1) vs CI(1,0): Empirical size and power of wavelet cointegration tests.

	$\mathcal{H}_0$: *Cointegration*					$\mathcal{H}_a$: *No Coint.*	
	WN	AR(1)				RW	IMA(1,1)
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 1.0$	$\theta = 1.0$
W1^{dyn}	6.19	6.30	6.99	9.18	16.99	39.14	50.72
W2^{dyn}	**6.20**	**6.31**	7.02	9.30	17.77	46.40	60.45
W1	1.89	2.86	5.36	11.86	34.59	56.71	66.19
W1^{qs}	1.88	2.84	5.34	11.66	33.39	51.28	59.49
W2	1.94	2.96	5.68	**13.05**	**37.56**	69.21	**82.38**
H1	0.86	1.80	3.33	7.10	24.59	58.77	61.50
H2	0.96	2.08	3.82	8.32	28.54	**74.38**	79.38
C(ℓ_4)	4.50	**5.78**	**7.97**	12.67	23.88	63.53	63.21
C(qs)	4.70	5.04	6.00	7.18	8.98	45.44	45.84
LBI	5.55	5.08	5.11	6.57	12.69	48.45	47.96

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Sample size $T = 100$. Column maxima in bold. W^{qs}: QS kernel with automatic bandwidth and elimination of leads and lags by BIC selection. W: *idem* with Bartlett kernel and T-dependent bandwidth. H: *idem* with Haar wavelet. W^{dyn}: with leads and lags in CI equation. Simulations using 10 000 replications from DGP 2 (parameters values as in (10)).

TABLE 5. CI(1,1) vs CI(1,0): Empirical size and power of wavelet cointegration tests (2).

	$\mathcal{H}_0$	$\mathcal{H}_a$: *No Cointegration*									
	WN	IMA(1,1): $\theta < 0$				RW	IMA(1,1): $\theta > 0$				
	−1.0	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
W1^{dyn}	6.19	7.33	13.00	21.95	31.08	39.14	44.51	48.13	49.93	**50.74**	50.72
W2^{dyn}	6.20	7.45	14.12	25.25	36.63	46.40	53.18	57.55	59.39	60.24	**60.45**
W1	1.89	9.89	30.57	43.50	51.08	56.71	60.86	63.39	65.19	65.80	**66.19**
W1^{qs}	1.88	9.30	28.44	40.27	46.30	51.28	54.99	57.30	58.71	59.02	**59.49**
W2	1.94	11.08	34.83	50.98	61.66	69.21	74.88	78.82	81.25	81.98	**82.38**
H1	0.86	13.68	35.20	48.49	54.91	58.77	60.80	61.31	**61.76**	61.64	61.50
H2	0.96	16.75	43.03	60.41	69.25	74.38	77.14	78.24	78.80	79.33	**79.38**
C(ℓ_4)	4.50	56.31	61.80	63.02	63.49	63.53	**63.59**	63.51	63.37	63.27	63.21
C(qs)	4.70	28.51	34.62	41.73	44.50	45.44	45.78	45.92	**45.96**	45.89	45.84
LBI	5.55	45.45	48.27	**48.62**	48.52	48.45	48.28	48.15	48.05	47.99	47.96

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Sample size $T = 100$. Row maxima in bold. Simulations using 10 000 replications from DGP 2 with $\phi = 1$ and $|\theta| \leq 1$ (rest of parameter values as in (10)).

TABLE 6. CI(1,0): Power of cointegration tests as a function of sample size.

T	*30*	*40*	*50*	*100*	*150*	*200*	*250*	*500*	*1 000*
W1^{dyn}	25.41	26.93	28.81	39.14	45.33	49.58	54.21	66.53	74.21
W2^{dyn}	33.71	35.28	36.96	46.40	53.28	56.59	60.87	72.23	79.82
W1	36.28	39.33	43.86	56.71	61.13	65.48	68.49	77.72	82.53
W1^{qs}	34.06	36.77	40.01	51.28	57.11	61.04	63.89	74.35	80.49
W2	51.54	56.62	59.70	69.21	73.98	76.87	79.11	85.90	90.16
H1	39.05	42.49	47.31	58.77	64.58	68.22	69.58	79.43	83.42
H2	**57.47**	**62.61**	**65.60**	**74.38**	**78.17**	**81.04**	**82.62**	88.52	92.01
C(ℓ_4)	40.94	40.41	46.84	63.53	67.18	75.25	75.78	**90.08**	**95.50**
C(qs)	15.19	22.26	28.18	45.44	53.54	61.00	64.98	80.08	90.72
LBI	24.40	22.24	28.95	48.45	51.30	60.00	59.74	78.18	86.31

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Column maxima in bold. Simulations using 10 000 replications from DGP 2 with random walk errors ($\phi = 1$, $\theta = 0$; rest of parameter values as in (10)).

TABLE 7. CI(2,2) vs CI(2,0): Empirical size and power of wavelet cointegration tests.

	$\mathcal{H}_0$: *Cointegration*						$\mathcal{H}_a$: *No Coint.*
	WN	AR(1)				I(1)	IMA(2,1)
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 1.0$	$\theta = 1.0$
W1^{dyn}	6.22	6.37	6.35	6.12	6.36	6.46	60.22
W2^{dyn}	**6.22**	**6.37**	**6.36**	6.12	6.36	6.49	64.79
W1	5.92	5.92	5.60	5.71	6.88	6.36	83.79
W1^{qs}	5.92	5.92	5.60	5.71	6.88	6.31	81.19
W2	5.92	5.92	5.60	5.71	6.89	6.50	**89.41**
H1	2.63	2.90	2.90	3.02	3.91	16.93	24.96
H2	2.67	3.00	3.05	3.23	4.21	20.30	29.34
C(ℓ_4)	3.07	3.93	5.39	**8.81**	**17.79**	**52.50**	52.36
C(qs)	3.00	3.50	5.91	8.87	6.84	35.99	35.70
LBI	3.98	3.51	3.52	4.55	9.06	39.07	37.60

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Sample size $T = 100$. Column maxima in bold. W^{qs}: QS kernel with automatic bandwidth and elimination of leads and lags by BIC selection. W: *idem* with Bartlett kernel and T-dependent bandwidth. H: *idem* with Haar wavelet. W^{dyn}: with leads and lags in CI equation. Simulations using 10 000 replications from DGP 2 (parameters values as in (10)).

TABLE 8. CI(2,1) vs CI(2,0): Empirical size and power of wavelet cointegration tests.

	$\mathcal{H}_0$: *Cointegration*					$\mathcal{H}_a$: *No Coint.*	
	I(1)	ARI(1,1)				I(2)	IMA(2,1)
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 1.0$	$\theta = 1.0$
$W1^{dyn}$	6.46	6.74	6.99	7.72	11.02	53.47	60.22
$W2^{dyn}$	6.49	6.78	7.05	7.96	11.68	57.32	64.79
W1	6.36	6.95	8.12	11.83	22.55	78.21	83.79
$W1^{qs}$	6.31	6.78	7.93	11.33	21.38	75.92	81.19
W2	6.50	7.04	8.47	12.55	24.51	83.16	89.41
H1	16.93	16.85	16.71	16.46	17.17	25.29	24.96
H2	20.30	20.13	20.24	19.94	20.91	29.61	29.34
$C(\ell_4)$	**52.50**	**52.27**	**52.25**	**52.05**	**53.01**	52.77	52.36
C(qs)	35.99	35.97	36.07	36.01	36.13	36.04	35.70
LBI	39.07	38.42	38.06	37.70	37.82	38.01	37.60

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Sample size $T = 100$. Column maxima in bold. W^{qs}: QS kernel with automatic bandwidth and elimination of leads and lags by BIC selection. W: *idem* with Bartlett kernel and T-dependent bandwidth. H: *idem* with Haar wavelet. W^{dyn}: with leads and lags in CI equation. Simulations using 10 000 replications from DGP 2 (parameters values as in (10)).

TABLE 9. CI(2,2) & CI(2,1) vs CI(2,0): Empirical size and power of wavelet cointegration tests.

	$\mathcal{H}_0$		$\mathcal{H}_a$: *No Cointegration*									
	CI(2,2)	CI(2,1)	IMA(2,1): $\theta < 0$				I(2)	IMA(2,1): $\theta > 0$				
			−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
$W1^{dyn}$	6.22	6.46	16.30	28.66	39.62	48.11	53.47	56.76	58.60	59.71	60.12	**60.22**
$W2^{dyn}$	6.22	6.49	17.19	30.20	42.16	51.36	57.32	60.92	63.09	64.25	64.67	**64.79**
W1	5.92	6.36	33.47	54.19	66.79	74.01	78.21	80.82	82.60	83.43	83.87	83.79
$W1^{qs}$	5.92	6.31	32.37	52.53	64.70	71.73	75.92	78.54	80.12	80.93	81.23	81.19
W2	5.92	6.50	35.61	57.54	70.91	78.44	83.16	86.13	88.03	88.91	89.31	89.41
H1	2.63	16.93	**26.51**	25.67	25.55	25.37	25.29	25.14	25.12	25.05	25.01	24.96
H2	2.67	20.30	**31.20**	30.15	29.99	29.85	29.61	29.48	29.47	29.37	29.35	29.34
$C(\ell_4)$	3.07	52.50	**55.06**	53.75	53.28	53.01	52.77	52.71	52.60	52.52	52.45	52.36
C(qs)	3.00	35.99	**37.81**	36.99	36.57	36.26	36.04	35.90	35.88	35.77	35.74	35.70
LBI	3.98	39.07	**40.08**	39.03	38.48	38.18	38.01	37.92	37.83	37.70	37.60	37.60

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Sample size $T = 100$. Row maxima in bold. Simulations using 10 000 replications from DGP 2 with $\phi = 1$ and $|\theta| \leq 1$ (rest of parameter values as in (10)).

TABLE 10. CI(2,2): Size of cointegration tests as a function of sample size.

T	*30*	*40*	*50*	*100*	*150*	*200*	*250*	*500*	*1 000*
$W1^{dyn}$	8.57	7.29	7.07	5.71	5.88	5.26	5.63	5.16	**4.94**
$W2^{dyn}$	8.57	7.32	7.08	5.71	5.88	5.26	5.63	5.16	**4.94**
W1	8.99	7.00	6.17	5.32	5.40	5.28	5.39	4.91	**5.05**
$W1^{qs}$	8.99	7.00	6.17	5.32	5.40	5.28	5.39	4.91	**5.05**
W2	8.99	7.01	6.17	5.32	5.40	5.28	5.39	4.91	**5.05**
H1	2.11	2.42	2.30	2.48	2.81	2.45	2.40	1.65	1.33
H2	2.35	2.62	2.39	2.49	2.80	2.48	2.43	1.68	1.32
$C(\ell_4)$	2.57	2.36	2.86	2.60	2.78	3.33	2.83	2.84	3.09
C(qs)	2.49	2.42	2.91	2.69	2.68	3.28	2.84	2.85	3.08
LBI	9.43	10.36	7.63	3.90	3.32	3.82	3.27	2.98	3.31

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Simulations using 10 000 replications from DGP 2 with white noise errors ($\phi = \theta = 0$; rest of parameter values as in (10)).

TABLE 11. CI(2,1): Size of cointegration tests as a function of sample size.

T	*30*	*40*	*50*	*100*	*150*	*200*	*250*	*500*	*1 000*
$W1^{dyn}$	10.88	8.90	7.85	6.35	6.75	6.04	5.81	5.23	**5.24**
$W2^{dyn}$	11.59	9.31	8.14	6.43	6.79	6.07	5.84	5.25	**5.24**
W1	10.88	8.90	7.85	6.35	6.75	6.04	5.81	5.23	**5.24**
$W1^{qs}$	10.66	8.76	7.72	6.27	6.75	6.02	5.80	5.23	**5.24**
W2	11.59	9.31	8.14	6.43	6.79	6.07	5.84	5.25	**5.24**
H1	19.53	16.89	17.65	17.80	17.77	17.57	16.57	16.73	16.38
H2	26.47	24.27	24.00	21.01	20.39	19.46	18.36	17.89	16.91
$C(\ell_4)$	31.09	31.91	37.49	53.38	57.93	63.65	65.54	81.23	89.03
C(qs)	14.80	18.08	22.29	36.52	45.33	50.14	54.81	68.45	80.99
LBI	18.17	17.49	23.34	39.66	43.80	49.56	50.58	66.71	75.39

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Simulations using 10 000 replications from DGP 2 with white noise errors ($\phi = \theta = 0$; rest of parameter values as in (10)).

TABLE 12. CI(2,0): Power of cointegration tests as a function of sample size.

T	*30*	*40*	*50*	*100*	*150*	*200*	*250*	*500*	*1 000*
$W1^{dyn}$	41.21	40.21	43.34	50.02	53.55	57.07	59.54	68.62	74.83
$W2^{dyn}$	45.79	45.33	47.71	53.58	56.91	60.02	62.04	70.95	76.94
W1	53.14	58.74	63.82	73.87	76.98	80.85	82.05	86.91	89.61
$W1^{qs}$	54.16	58.90	63.75	73.54	76.86	80.75	81.92	86.82	89.60
W2	**56.92**	**61.34**	**65.76**	**74.71**	**77.64**	**81.25**	**82.47**	**87.19**	89.82
H1	26.09	22.83	22.89	25.04	25.07	26.78	27.21	31.16	35.20
H2	36.28	32.85	31.12	29.68	28.39	29.38	29.86	32.57	36.11
$C(\ell_4)$	25.51	29.23	35.43	53.51	57.76	65.14	67.01	86.97	**95.34**
C(qs)	7.68	13.30	20.74	36.33	45.03	50.85	55.89	71.08	88.08
LBI	11.25	13.30	19.99	38.07	42.32	49.31	50.83	68.38	80.46

Wavelet filter: LA8, $J = 1$. % rejections at 5% significance level. Column maxima in bold. Simulations using 10 000 replications from DGP 2 with random walk errors ($\phi = 1$, $\theta = 0$; rest of parameter values as in (10)).