ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

A TEST FOR THE NULL OF MULTPLE COINTEGRATING VECTORS

Javier Fernandez-Macho

Number 657
June 2013

Manor Road Building, Manor Road, Oxford OX1 3UQ

A TEST FOR THE NULL OF MULTIPLE COINTEGRATING VECTORS

JAVIER FERNÁNDEZ-MACHO

Abstract. This paper examines a test for the null of cointegration in a multivariate system based on the discrepancy between the OLS estimator of the full set of n cointegrating relationships in the $n + k$ system and the OLS estimator of the corresponding relationships among first differences without making specific assumptions about the short-run dynamics of the multivariate data generating process. It is shown that the proposed test statistics are asymptotically distributed as standard chi-square with $n + k$ degrees of freedom and are not affected by the inclusion 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 critical values for these statistics are tabulated using Monte Carlo simulation and it is shown that these non residual-based tests exhibit appropriate size and good power even for quite general error dynamics. In fact, simulation results suggest that they perform quite reasonably when compared to other tests of the null of cointegration.

Key words: Brownian motion; Cointegration; Econometric Methods; Integrated process; Multivariate Analysis; Time Series Models; Unit root.
JEL Classification: C22, C12.

1. Introduction

Since Granger (1981, 1983) introduced the notion of cointegrating relationships, testing for cointegration has become standard practice in the empirical analysis of economic time series. As a result, quite a number of tests have been proposed for this purpose. Some of the cointegration tests most widely used in practice are in fact already available unit-root tests (Dickey-Fuller and augmented Dickey-Fuller tests, Phillips' Z_α and Z_t statistics, Choi (1994)'s DHS, *etc.*) applied to the residuals from the cointegrating regression: the 'two-step' procedure suggested by Engle and Granger (1987). As these statistics are designed to test the presence of a unit root against a stationary alternative, when used on regression residuals cointegration appears as the alternative hypothesis rather than the null. Although testing the relevant hypothesis directly under the null certainly seems a more natural choice, so far there have been few attempts to provide proper statistics for a direct test of cointegration. Phillips and Ouliaris (1990) argue that the source of the difficulties lies in the failure of conventional asymptotic theory under the null of cointegration. However, some simple ways to overcome this problem have been proposed in the literature. In this connection, Hansen (1992) derived the large sample distribution of LM tests for parameter stability against several alternatives in the context of cointegrated regression models. In particular, testing for intercept stability against the alternative of a random-walk intercept (while the rest of the coefficients are held constant) would effectively be a test of the null of cointegration. However, Hansen (1992)'s test was not designed for this purpose and its actual alternative does not imply an $I(1)$ error process. Similarly Quintos and Phillips (1993)'s test does not take the alternative of noncointegration. Park (1990) proposed two statistics for cointegration testing (one of them for the null of cointegration) based on the addition of 'superfluous' regressors. If the variables in the underlying model are cointegrated, a standard testing procedure should be able to detect the superfluous nature of the added regressors, as compared to the 'true' ones; in contrast, we would not expect this when the variables under consideration are not

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.

cointegrated and the relationship is itself spurious. Shin (1994) says, however, that this test is rather *ad hoc* and indeed it remains unclear how the superfluous regressors should be selected.

Further attempts to provide for a direct test of cointegration have also been made within the family of residual-based tests itself. Leybourne and McCabe (1994)'s residual-based LBI test of cointegration is obtained as an extension of their previous LBI test for coefficient constancy (see Leybourne and McCabe (1989)) while Shin (1994)'s C test is a residual-based test obtained as an extension of an LM test of univariate stationarity (Kwiatkowski, Phillips, Schmidt, and Shin, 1992). Both proposals can be shown to be identical except for minor details including the choice of window in the estimation of the residual long-run variance (namely, Leybourne-McCabe's LBI uses a rectangular lag window while Shin's C uses Bartlett's triangular lag window, thus ensuring the nonnegativity of the long-run variance estimator). They both use the same stochastic components model which (apart from a deterministic trend and additional regressors) can be written as a regression whose disturbances would follow a random-walk-plus-error model. Then they test the null hypothesis that the residuals are just stationary around a constant. The limiting distribution of their statistic is a functional of Brownian motion and its critical values are calculated via Monte-Carlo simulation and tabulated in Shin (1994). In a related issue Shephard (1993) and Fernández-Macho (1996) investigate the small sample properties of respectively time domain and spectral estimates of the signal-to-noise ratio in a random-walk-plus-error model that could also be used to test cointegration if applied to the regression residuals. Choi and Ahn (1995) introduced tests for the null of cointegration that can be applied to a system of equations, which basically consist of a test of multivariate stationarity applied to regression residuals. More recently, Xiao (1999) has proposed a residual-based test based on Ploberger, Kramer, and Kontrus (1989)'s statistic for structural stability.

Adding to all this, the present paper investigates a class of cointegration tests (under the null) which are not based on the residuals but on the estimated regression coefficients themselves. Furthermore, the testing procedure can be applied to a system of equations as well as to a single equation. The statistics can be considered as based on the same simple principle as the Durbin-Hausman's specification test: here a regression in first differences is used as a benchmark for the (cointegrating) regression in levels. The Hausman-type test was originally proposed in Fernández-Macho (1994) for multivariate regression models with a single cointegrating relationship where short-run dynamics, in the shape of regressors' leads and lags, was prefiltered by auxiliary regressions (see also Fernández-Macho and Mariel 1994a,b). More recently, Choi, Hu, and Ogaki (2008) picked the same idea and proposed a modified version consisting in explicitly introducing the dynamic terms, as in Shin (1994) or Leybourne and McCabe (1989), into a single regression equation model.

Except for Choi and Ahn (1995) and, to some extent, Harris (1997), there have been no major suggestions in the literature that can be used for testing the null of cointegration against spurious alternatives in a system of equations[1]. If simultaneous relations among economic variables are of interest, then using a multiple equation testing procedure will be more convenient than applying tests several times in order to test the cointegrating relationships jointly. To be concise, the main advantages of the proposed testing procedure are that the suggested statistics are asymptotically distributed as a chi-square, they are not affected by the inclusion of deterministic terms or dynamics in the cointegrating equations so that no

[1]Recall that the alternatives in the two versions of Johansen (1991)'s cointegration test are either H_a : *'one fewer unit roots'* (λ_{max} test) or H_a : *'no unit roots'* (trace test).

special tables are needed and they can be applied to a system of equations as well as to single equations individually. Monte Carlo simulation suggests that in finite samples their nominal sizes are not much altered by changes in error dynamics and they have good power for a wide range of alternatives.

The paper is organised as follows. Section 2 presents statistics for the new Hausman-like test of the null of cointegration and Section 3 examines their asymptotic properties. Section 4 presents critical values in finite samples together with Monte Carlo evidence on size and power comparisons. Finally Section 5 summarises the main conclusions. 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 $\mathrm{vec}(A)$ 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, Δ denotes the difference operator such that $\Delta \equiv 1 - L$ where L is the lag operator, $\mathbf{1}(\cdot)$, 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. The test

Let $\zeta_t = (v_t', \eta_t')'$ follow an $(n + k)$-dimensional zero-mean stationary vector process and 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) (*c.f.* Phillips (1991)'s triangular representation)

$$\begin{aligned} \underset{(n)}{z_t} &= \underset{(n\times d)}{\alpha'}\ \underset{(d)}{g(t)} + \underset{(n\times k)}{\beta'}\ \underset{(k)}{x_t} + \underset{(n)}{u_t}, \\ x_t &= x_{t-1} + \eta_t, \end{aligned} \tag{1}$$

where the elements of vector $g(t)$ are deterministic functions of time (such as time trends) and u_t is a vector error process that may possibly have a unit root, *i.e.*

$$u_t = \frac{1}{1 - \delta L} v_t, \tag{1b}$$

where δ is a *residual unit root indicator* such that $\delta = 1 - \mathbf{1}(\mathcal{H}_0 : \text{cointegration})$, that is δ will take a value of 0 if there are no residual unit roots so that vector $(z_t', x_t')'$ is cointegrated in the sense of Engle and Granger (1987) (the null) and a value of 1 otherwise (the alternative). Methods of estimation and inference for equation systems like (1) are discussed in Phillips (1991) and Phillips and Hansen (1990) among others.

Assumption 1. *Let ζ_t have an autocovariance function (acvf) $\mathcal{E}(\zeta_t\zeta_{t-s}') = C(s)$, $(s = 0, \pm1, \pm2, \ldots)$ that is absolutely summable (*i.e. $\sum_{s=-\infty}^{\infty} \|C(s)\| < \infty$*) and a spectral density $f(\cdot)$ 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*

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

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

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

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

Following, say, Brillinger (1975, p.296), we may then start by conditioning v_t on $\{\eta_t\}$ so that

$$v_t = \sum_{s=-\infty}^{\infty} \gamma'_s \eta_{t-s} + \xi_t \tag{2}$$

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 $\mathcal{E}(\xi_t \eta'_{t-s}) = 0$, $(s = 0, \pm1, \pm2, \ldots)$, $\forall t$. Therefore, (*c.f.* Saikkonen (1991)) $\exists\, m$ large enough so that $\gamma_s \approx 0$ for $|s| > m$ and the sum in (2) 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, under the null of cointegration, z_t conditional on $\{x_{t-m-1}, \ldots, x_{t+m}\}$ can be written as

$$z_t = \alpha'\, g(t) + \beta'\, x_t + \sum_{s=-m}^{m} \gamma'_s\, \eta_{t-s} + \varepsilon_t \tag{3}$$

where ε_t is an n-dimensional zero-mean stationary process such that $\mathcal{E}(\varepsilon_t \eta'_{t-s}) = 0$, $\forall t, s$. OLS estimation of (3) 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 (Saikkonen, 1991). Furthermore, we observe that the asymptotic covariance matrix of x_t and $\{\eta_{t-m}, \ldots, \eta_{t+m}\}$ is block-diagonal (Phillips and Hansen, 1988) which suggests that the set of nuisance parameters $\{\gamma_{-s} \ldots \gamma_s\}$ may be estimated from a regression of the OLS residuals $\tilde{v}_t$ from (1) on $\{\eta_{t-s} = \Delta x_{t-s}, |s| \le m\}$. This will allow us to estimate β from $y_t = \beta x_t + \varepsilon_t$, where the regressand

$$y_t \overset{def}{=} z_t - \tilde{\alpha}'\, g(t) - \sum_{s=-m}^{m} \tilde{\gamma}'_s\, \Delta x_{t-s}$$

by construction, with $\tilde{\alpha}$ being the vector of OLS estimates of coefficients of deterministic components and $\{\tilde{\gamma}_s\}$ the estimates of the nuisance parameters in (3).

Model (1) may thus be rewritten as

$$y_t = \beta' x_t + \varepsilon_t + \delta\, u_{t-1}, \tag{4}$$

where, we recall, $\{\varepsilon_t\}$ is a zero-mean stationary process uncorrelated with the increments of $\{x_t\}$ at all leads and lags and $\delta = \mathbf{1}(\mathcal{H}_a)$. Under cointegration the error term is simply $\varepsilon_t \sim I(0)$, while under the alternative of no cointegration the error term becomes $u_t^\dagger = u_{t-1} + \varepsilon_t \sim I(1)$. As a result, the OLS estimator $\hat{\beta}_l$ from the levels regression (4) will be T-consistent under the null of cointegration (Phillips and Durlauf, 1986; Stock, 1987) but will have a non-degenerate distribution under the alternative. On the other hand, taking differences (*i.e.* imposing one unit root)

$$\Delta y_t = \beta' \eta_t + \varepsilon_t^\dagger, \tag{5}$$

where the error process $\varepsilon_t^\dagger = \Delta\varepsilon_t + \delta\, v_{t-1}$ is always stationary, so that standard asymptotics on stationary variables apply yielding an OLS estimator $\hat\beta_d$ that it is $\sqrt{T}$-consistent under the null while still being $O_p(T^{-1/2})$ under the alternative (note that $\hat\beta_d$ might be asymptotically biased since $\mathcal{E}(\eta_t v'_{t-1})$ need not be zero in general). This regression in differences may then be used as a benchmark for the regression in levels in order to test for cointegration. The fact that through use of a suitable time domain correction we are able to reformulate model (1) into regression (4) where the regressors x_t are strictly exogenous, has important consequences for the applicability of our testing procedure. Otherwise, the regression in differences would in general be inconsistent under the null and would not serve as a benchmark. Alternatively, instrumental variables could be used as in Phillips and Hansen (1990). However, their existence cannot be taken for granted and a more general setup is desirable.

The presence of $g(t)$ in equation (1) 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 our case since the regression in differences (5) is not only free from short-run-dynamics nuisance parameters but also free, by construction, from deterministic components. As a result, our testing procedure will not be affected by them.

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 $c = \mathrm{vec}(\hat\beta_d - \hat\beta_l)$ between the OLS estimators obtained from a regression in first differences and a regression in levels can be proposed. The test statistics are:

$$\mathrm{H1} = c'(\hat V_d + \hat V_l)^{-1} c, \qquad \mathrm{H2} = c' \hat V_d^{-1} c, \tag{6}$$

where $\hat V_d$ and $\hat V_l$ are consistent estimates of the covariance matrices of $\hat\beta_d$ and $\hat\beta_l$ respectively, given by

$$\hat V_l = \hat\Omega_\varepsilon \otimes (X'X)^{-1}, \qquad \hat V_d = \left(I_n \otimes (X'BX)^{-1}\right) \hat R \left(I_n \otimes (X'BX)^{-1}\right) \tag{7}$$

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

$$\hat R = (I_n \otimes X'B)\, \hat V_\varepsilon\, (I_n \otimes BX) \tag{7b}$$

where $B = D'D$ with D as the $(T-1\times T)$ matrix such that $D = \left[d_{ij} | d_{ii} = -1, d_{i,i+1} = 1, d_{ij} = 0 \text{ otherwise}\right]$ and $\hat V_\varepsilon$ is a consistent estimator of the covariance matrix of $\{\varepsilon_t\}$. Since $\hat V_l$ is $O_p(T^{-2})$ while $\hat V_d$ is $O_p(T^{-1})$ the two statistics are asymptotically equivalent. Table 1 summarises the steps involved in calculating the statistics.

One last comment on the interpretation of our H statistics from a different point of view: H2 may also be figured out as the typical chi-square statistic for testing whether $\hat\beta_d$ is significantly different from true β. In order to evaluate the statistic, the unknown β is replaced by $\hat\beta_l$ whose faster consistency rate ensures that the asymptotic distribution remains unaltered. The addition of $\hat V_l$ in the denominator of H1 intends to provide a better approximation in small samples (see lemma 6 in the appendix).

TABLE 1. Computation of the H test statistics.

(1) OLS regression: $z_t = \alpha' g(t) + \beta' x_t + v_t$ to obtain estimates $\tilde{\alpha}$ for deterministic components and residuals $\{\tilde{v}_t\}$
  (a) OLS regression: $\tilde{v}_t = \sum_{s=-m}^{m} \gamma_s' \Delta x_{t-s} + \varepsilon_t$ to obtain estimates of nuisance parameters $\{\tilde{\gamma}_{-m} \cdots \tilde{\gamma}_0 \cdots \tilde{\gamma}_m\}$
  (b) calculate $y_t = z_t - \tilde{\alpha}' g(t) - \sum_{s=-m}^{m} \tilde{\gamma}_s' \Delta x_{t-s}$
  (c) OLS regression of y_t on x_t (in levels): $y_t = \beta' x_t + \varepsilon_t$ to obtain $\hat{\beta}_l$ and $\hat{V}_l$ as from (7).
(2) OLS regression of Δy_t on Δx_t (in differences): $\Delta y_t = \beta' \eta_t + \varepsilon_t^{\dagger}$ to obtain $\hat{\beta}_d$ and $\hat{V}_d$ as from (7).
(3) Calculate the discrepancy $c = \mathrm{vec}(\hat{\beta}_d - \hat{\beta}_l)$ and the H statistics as from (6).

3. Asymptotic distributions

Let $C_\varepsilon(s) = \mathcal{E}(\varepsilon_t \varepsilon_{t-s}')$ be the acvf of the error process $\{\varepsilon_t\}$ in the levels regression, and let us define the covariance matrix V_ε as the $(nT \times nT)$ block-Toeplitz matrix formed by $(n \times n)$ blocks with $C_\varepsilon(|i-j|)$ in the (i, j)'th position. 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 $B_\varepsilon(r) = B_v - \Omega_{\eta v}' \Omega_\eta^{-1} B_\eta$ is a k-variate Brownian motion (uncorrelated by construction with B_η) with covariance matrix

$$\Omega_\varepsilon = \Omega_v - \Omega_{\eta v}' \Omega_\eta^{-1} \Omega_{\eta v} = 2\pi f_\varepsilon(0)$$

where $f_\varepsilon(\cdot)$ is the error spectral density in (3) which is related to the spectral density of ζ_t through the expression $f_\varepsilon(\cdot) = f_v(\cdot) - f_{\eta v}(\cdot)' f_\eta(\cdot)^{-1} f_{\eta v}(\cdot)$ (Brillinger, 1975, p.296). Since Ω is nonsingular, we also have that $\Omega_\varepsilon > 0$. Note that $\Omega_\eta > 0$ rules out cointegration between the regressors. Under these circumstances, we are able to state the following

Proposition 1. *In the multivariate regression model* (1) *under assumptions 1 and 2: under the null of cointegration, the Hausman-like statistics defined in* (6)

$$\mathrm{H1}, \mathrm{H2} \Rightarrow \chi^2(nk)$$

while under the alternative of no cointegration

$$\mathrm{H1} = O_p(T), \qquad \mathrm{H2} = O_p(T)$$

Proof: (see the appendix). □

Note that, although they are asymptotically equivalent under the null, under the alternative H1 and H2 would not have the same limit distribution. This different asymptotic behaviour under the alternative will have consequences for test power: indeed in the limit $T^{-1}\mathrm{H1} < T^{-1}\mathrm{H2}$ and test H2 will be asymptotically more powerful (see the appendix).

In short, it has been shown that the proposed Hausman-like 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 H 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.

4. Small sample evidence

Critical values: Tables 2 to 4 give critical values of the statistics H1 and H2 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* (1) *with* $\delta = 0$, $\alpha = 0$, $\beta_1 = \ldots = \beta_k = 1$ *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 $(1, -1, \ldots, -1)$. Using DGP 1 the quantiles of the small sample distribution of H1 and H2 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. Figs. 1 to 3 show density estimates of H1 and H2 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 as $T \to \infty$.

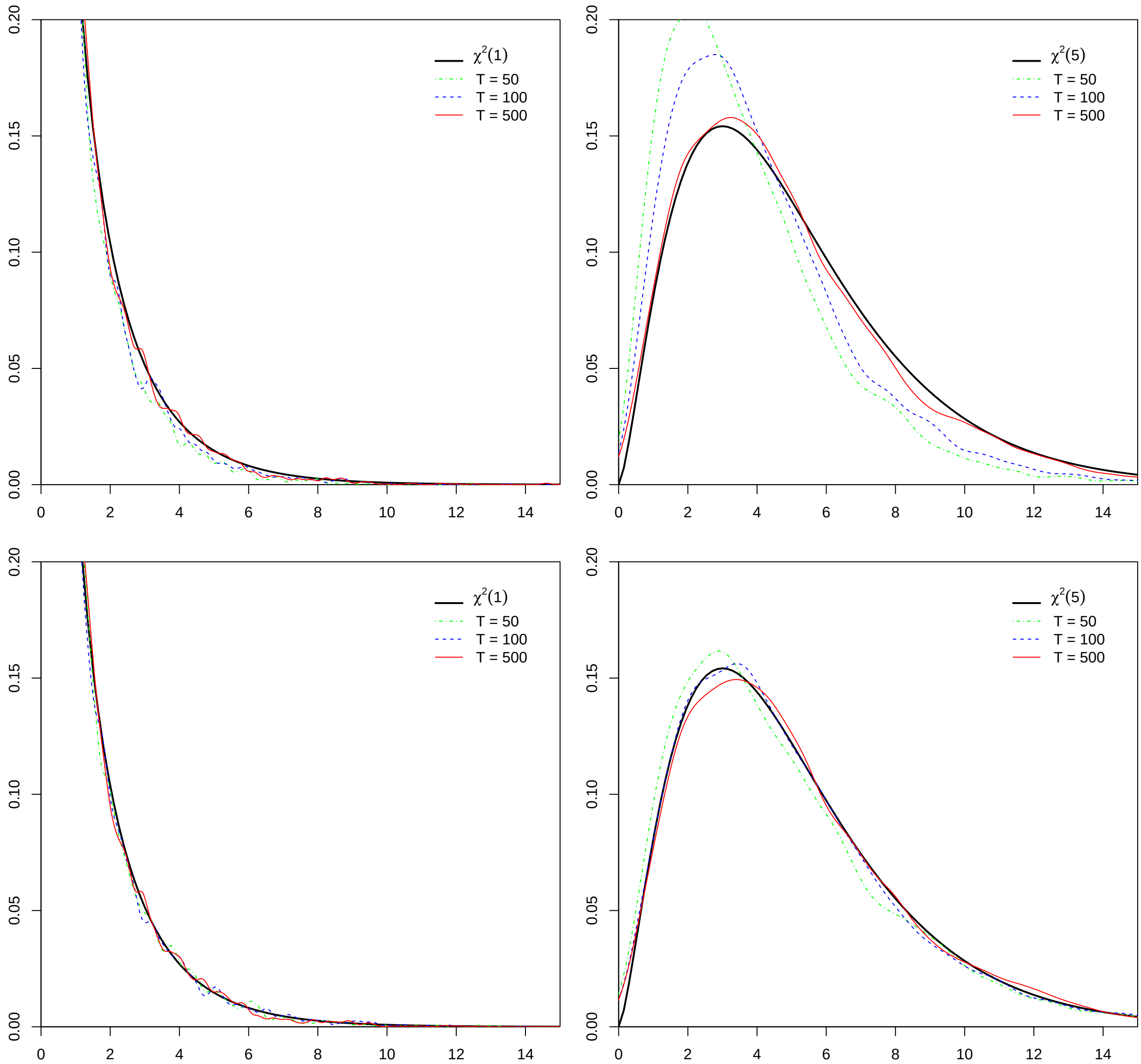
Size and power comparisons: As already mentioned, there are quite a few residual tests for the null of no-cointegration available in the literature. However, comparing the small sample performance of the H statistics with respect to them is not satisfactory given the reversal of the null and alternative hypothesis involved. Therefore, Shin (1994)'s C(ℓ), Leybourne and McCabe (1994)'s LBI and Hansen (1992)'s L_c residual-based tests of the null of cointegration were chosen for comparison purposes together with dynamic versions of the H statistics that include leads and lags of the regressors in the cointegrating regressions themselves.

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

$$w(s, \ell) = \frac{3}{\tau(\ell)^2} \left[\frac{\sin \tau(\ell)}{\tau(\ell)} - \cos \tau(\ell) \right],$$

with $\tau(\ell) = (6\pi/5)(s/\ell)$ and associated bandwidth parameter $\ell = 1.3221 \left[4T\hat{\rho}_1^2/(1 - \hat{\rho}_1)^4 \right]^{1/5}$, 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 Schwert (1989) or Kwiatkowski et al. (1992), the truncation lag length ℓ in these two cases was set as a function of T: $\ell_j = 1 + \lfloor j(T/100)^{1/4} \rfloor$, with $j = 0, 4, 12$. The results for $j = 0$ and 12 were less satisfactory than for $j = 4$ and, therefore, only the latter will be reported here. Also, the number m of additional regressors' leads and lags in the parametric correction for C and H statistics was set according to the BIC criterion. The last test of the null hypothesis of cointegration used in our comparison is Hansen's L_c test. This test is based on the fully-modified estimation method of Phillips and Hansen and uses semiparametric correction that leads to asymptotically median-unbiased estimates (Hansen, 1992; Haug, 1996).

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



Smoothed histogram: Gaussian kernel density estimates. H1 top, H2 bottom; $k = 1$ left, $k = 5$ right.

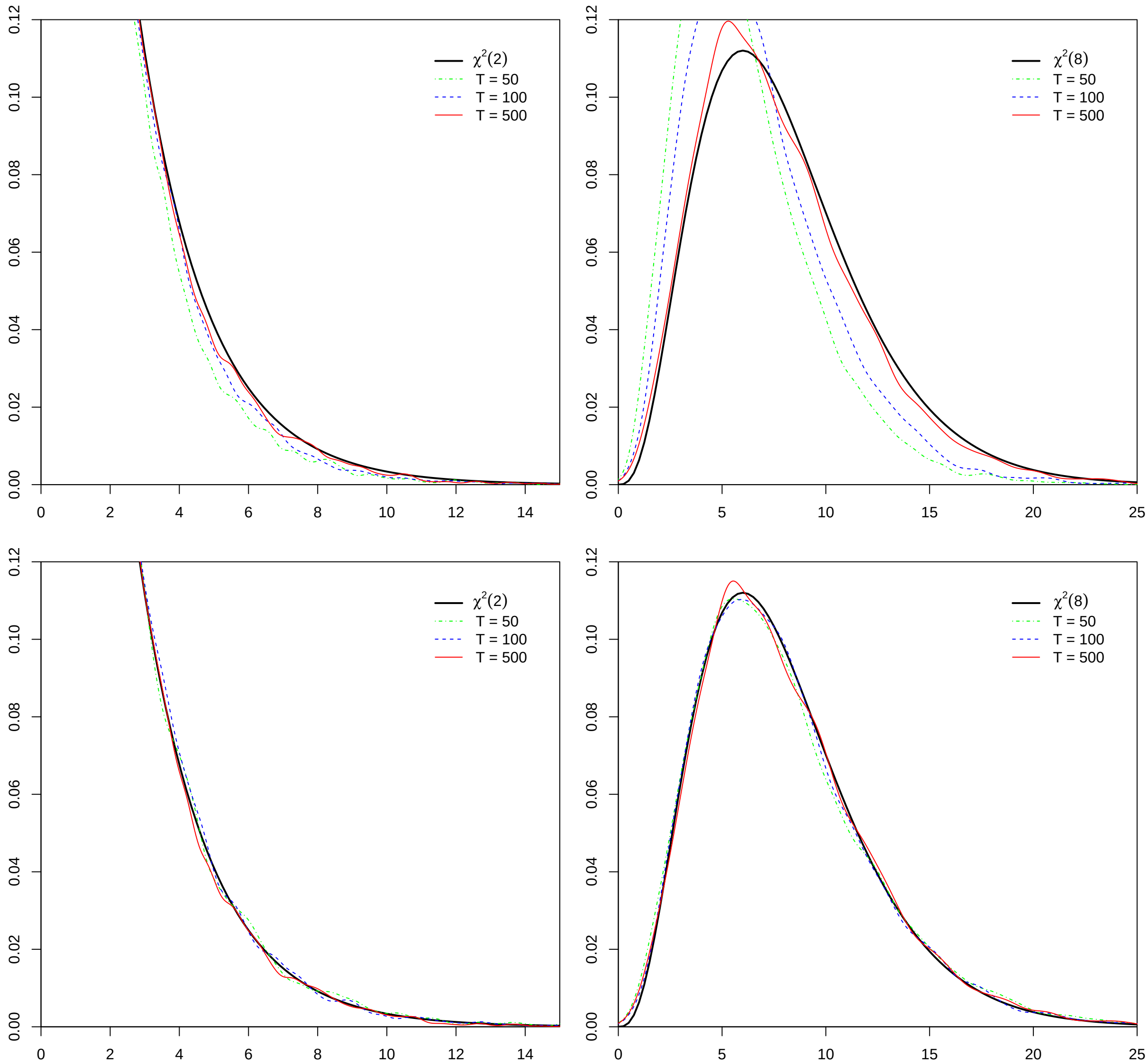
Tables 5 and 6 provide evidence on the size and power of tests of the null of cointegration in small samples. The first four rows refer to the proposed Hausman-like H1 and H2 test statistics and their dynamic versions using their asymptotic χ^2 critical values. Results for Shin's test for the quadratic kernel C(qs) and Barlett kernel C(ℓ_4) are reported next. The last two rows refer to Leybourne-McCabe's LBI and Hansen's L_c test statistics. The reported results were obtained through Monte Carlo simulation using 10 000 replications of size $T = 100$ from

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

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

[2]Note that δ in model (1) is simply an indicator of the existence of a unit root, while all other possible stationary AR roots are implicit in $\{v_t\}$. 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 $|\phi| < 1 \equiv \delta = 0$.

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

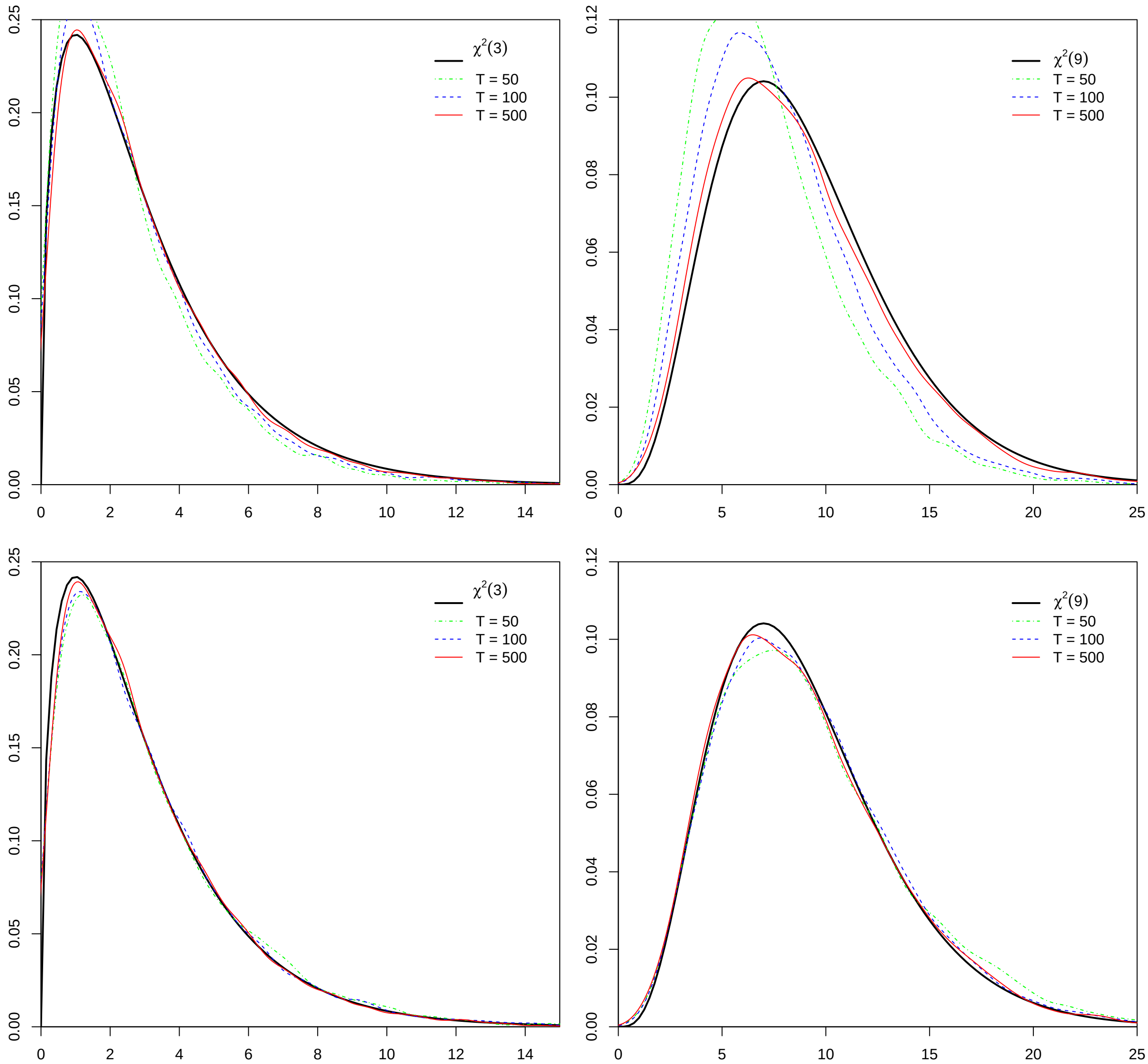


Smoothed histogram: Gaussian kernel density estimates. H1 top, H2 bottom; $k = 1$ left, $k = 4$ right.

This setup allows for quite general behaviour. It turns out that 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 v_t may exhibit nonzero correlations (at different lags). The lag-length m of the parametric correction in (3) was determined by the BIC method (Venables and Ripley, 2002). In the simulations, I 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:

$$
\begin{aligned}
\phi_\eta \quad &= \quad 0.45, \qquad \theta_\eta = -0.35, \qquad \rho = 0.5, \qquad \sigma = 1, \\
\theta = 0 \quad &, \quad \phi \in (0, 0.2, 0.4, 0.6, 0.8) \quad \text{under the null,} \\
\phi = 1 \quad &, \quad \theta \in (0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8, +1) \quad \text{under the alternative.}
\end{aligned}
\tag{8}
$$

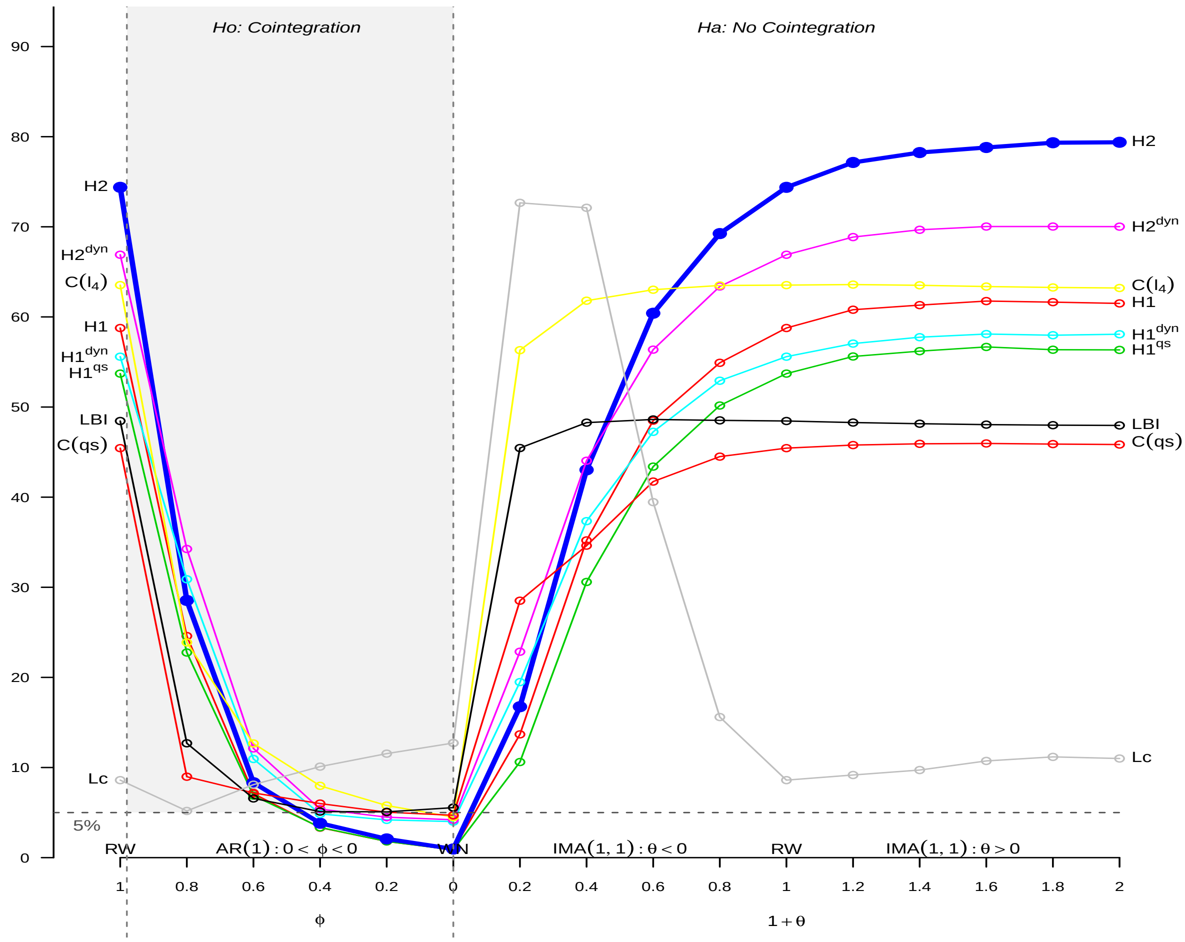
FIGURE 3. Empirical Distributions of H Statistics ($n = 3$).



Smoothed histogram: Gaussian kernel density estimates. H1 top, H2 bottom; $k = 1$ left, $k = 3$ right.

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) 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 5 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 6, θ was chosen so that changes in

FIGURE 4. Power of Cointegration Tests vs. distance from $\mathcal{H}_0$ (DGP 2; T = 100; 5% level).

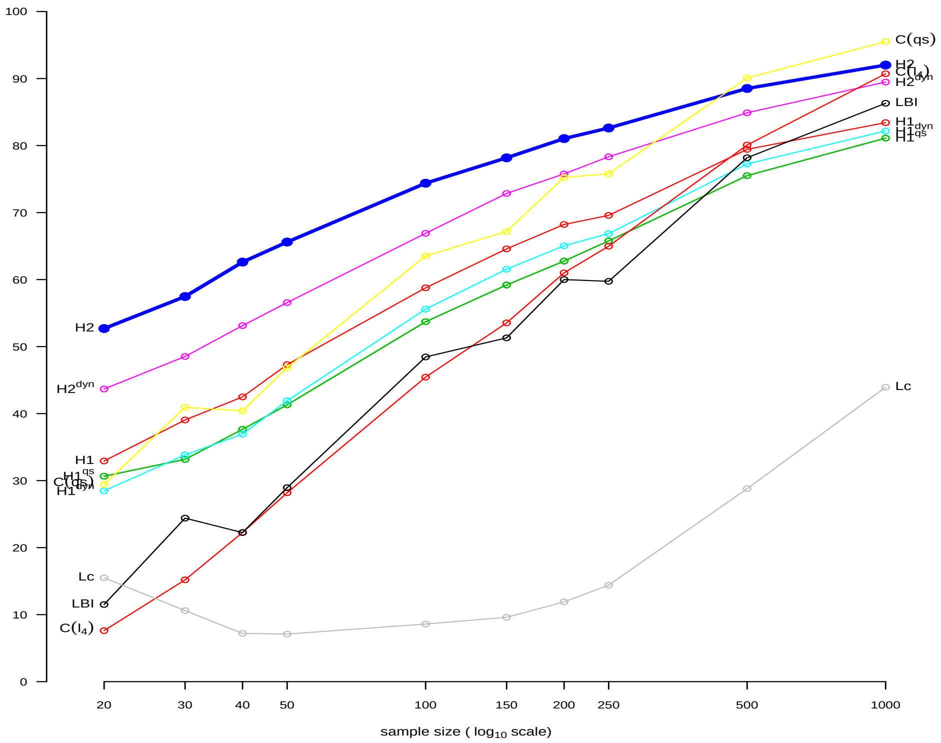


the tests' power can be observed as the null is being approached (negative values) and as it gets farther away from it (positive values). Fig. 4 shows the respective size and power of the tests involved in our comparison as functions of the said distances from $\mathcal{H}_a$ and $\mathcal{H}_0$ respectively.

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 except L_c maintain size distortions well within reasonable levels.

As far as test power is concerned, the last two columns of Table 5 correspond to the case $\phi = 1$ (which falls just outside the cointegration region: distance from $\mathcal{H}_a = 0$ in Fig. 4) and the case $\theta = 1$ (the farthest alternative considered: distance from $\mathcal{H}_0 = 2$ in Fig. 4). 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 H2 statistic in the sense that it appears more powerful. For example, for samples of size $T = 100$ at the 5% significance level, H2 will reject between 75% and 80% 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 H 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. 4). Such behaviour is of course very reasonable (although it is violated by the other test statistics, see

Figure 5. Power of Cointegration Tests vs. sample size (DGP 2; 5% level).



bold numbers in Table 6). On the other hand, the H tests appear to be less powerful for cases closer to the null. On the whole, the Hausman-like 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.

Table 7 (see also Fig. 5) 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 (8). That is, the table expands column $\phi = 1.0$ in Table 5 for sample sizes from $T = 20$ to 1 000. It shows that the H2 statistic performs quite well in small samples whilst showing reasonable power for larger sizes. In comparison, Shin's C statistic does not perform well in small samples. On the other hand, the very low power of the L_c statistic confirms the results obtained by Haug (1996).

5. Conclusions

This paper has introduced a new class of cointegration tests under the null which are not based on the residuals but on the estimated regression coefficients themselves. The statistics considered are based on the same simple principle as the Durbin-Hausman's specification test with a regression in first differences being used as a benchmark for the cointegrating regression in levels. The testing procedure can be applied to a system of equations as well as to a single equation, which can be of interest if simultaneous cointegrating relationships among economic variables need be considered jointly. An additional bonus is the simplicity of the calculations involved: the standard outcome of two OLS regressions is all we need in order to construct the H test statistics. It has been shown that, under the null hypothesis of cointegration, the proposed test statistics have a standard (χ^2) asymptotic distribution that is invariant to the presence of deterministic components, so that their asymptotic distribution does not show the shifts typical of residual-based tests. Asymptotic tests can thus be performed straight away without the need of special tables. Furthermore, it has also been shown that they diverge as $O_p(T)$ under the alternative, which ensures test consistency. Finally, Monte Carlo simulation suggests that in finite samples their nominal sizes remain fairly stable in the presence of changes in error dynamics, and they show good power against a wide range of alternatives. In this respect, it is also worth noticing that the power of the H statistics is directly related to the distance from the null, so that the farther we are from it the greater the power and *viceversa*. Therefore, we suggest that the H test is a useful tool for testing the null hypothesis of cointegration directly.

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 (3) while $\hat{\beta}_l$ and $\hat{\beta}_d$ are respectively the OLS estimator obtained from the level regression (4) and the OLS estimator obtained from the difference regression (5). Expressions for the corresponding estimators of their respective covariance matrices were given in (7).

First of all we want model (1) to be rewritten as (4) where the regressors are strictly exogenous. If $\gamma_s = 0$ for $|s| > m$ we have that $\varepsilon_t = \xi_t$, and the error term in (3) is uncorrelated with $\{\eta_t\}$ at all leads and lags so that the regressors in x_t 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 $(g(t), x_t, \varepsilon_t)$.

Lemma 1. $T^{-3/2}\sum x_t \Rightarrow \int B_\eta$, $T^{-5/2}\sum tx_t \Rightarrow \int rB_\eta$, $T^{-2}\sum x_t x_t' \Rightarrow \int B_\eta B_\eta'$, $T^{-3/2}\sum t\varepsilon_t \Rightarrow \int r\mathrm{d}B_\varepsilon$, $T^{-1}\sum x_t\varepsilon_t' \Rightarrow \int B_\eta \mathrm{d}B_\varepsilon'$.

The following lemma expresses in a concise form the limiting distributions of the least squares estimators in (3)

Lemma 2. *Let $\tilde\alpha$, $\tilde\beta$ and $\tilde\gamma_s$ be the OLS estimators obtained from* (3). *Then under the null*

$$\begin{array}{ll}
\text{(a)} & \operatorname{diag}(T^{1/2}, T^{3/2})(\tilde\alpha-\alpha) \Rightarrow h_0(B_\varepsilon, P(B_\eta)),\\
\text{(b)} & T(\tilde\beta-\beta) \Rightarrow h_0(B_\varepsilon, M(B_\eta)),\\
\text{(c)} & (T/m)^{1/2}\displaystyle\sum_{s=-m}^{m}(\tilde\gamma_s-\gamma_s) = O_p(1).
\end{array}$$

while under the alternative

$$\begin{array}{ll}
\text{(a}'\text{)} & \operatorname{diag}(T^{-1/2}, T^{1/2})(\tilde\alpha-\alpha) \Rightarrow h_a(B_v, P(B_\eta), 0),\\
 & \qquad = O_p(1)\\
\text{(b}'\text{)} & (\tilde\beta-\beta) \Rightarrow h_a(B_v, M(B_\eta), 0),\\
 & \qquad = O_p(1),\\
\text{(c}'\text{)} & m^{-1/2}\displaystyle\sum_{s=-m}^{m}(\tilde\gamma_s-\gamma_s) = O_p(1).
\end{array}$$

Results (a) and (b) were 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 (3) in compact form (assuming $g(t) = (1,t)'$ say):

$$\begin{array}{rrcl}
 & z_t &=& \beta^{*\prime} x_t^* + \varepsilon_t,\\
\text{where} & x_t^{*\prime} &=& (g(t)', x_t', \eta_{t+m}', \ldots, \eta_{t-m}')\\
\text{and} & \beta^{*\prime} &=& (\alpha', \beta', \gamma_{-m}', \ldots, \gamma_m').
\end{array} \tag{A-1}$$

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

$$\tilde Q_x = \begin{bmatrix} 1 & T^{-2}\sum t & T^{-3/2}\sum x_t' & 0\\ T^{-2}\sum t & T^{-3}\sum t^2 & T^{-5/2}\sum t x_t' & 0\\ T^{-3/2}\sum x_t & T^{-5/2}\sum t x_t & T^{-2}\sum x_t x_t' & 0\\ 0 & 0 & 0 & T^{-1}\sum \eta_t^*\eta_t^{*\prime}\end{bmatrix}$$

$$\Rightarrow \quad Q_x = \begin{bmatrix} \multicolumn{2}{c}{I_2} & \begin{matrix}\int B_\eta\\ \int rB_\eta\end{matrix} & \begin{matrix}0\\0\end{matrix}\\ \int B_\eta' & \int rB_\eta' & \int B_\eta B_\eta' & 0\\ 0 & 0 & 0 & V_\eta\end{bmatrix}$$

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

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

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 (A-1) gets an extra u_{t-1} term. We then define the scale matrices

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

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

$$\tilde Q_{xu} = \left[T^{-3/2}\sum u_t, T^{-5/2}\sum t u_t, T^{-2}\sum u_t x_t', T^{-1}\sum u_t\eta_t^{*\prime}\right]'$$

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

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* (3) *and in* (4) *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 $y_t \overset{def}{=} z_t - \tilde{\alpha} g(t) - \sum_{s=-m}^{m} \tilde{\gamma}_s \eta_{t-s} \xrightarrow{p} z_t - \alpha g(t) - \sum_{s=-m}^{m} \gamma_s \eta x_{t-s} = \beta x_t + \varepsilon_t$, 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 $y_t \overset{def}{=} z_t - \tilde{\alpha}' g(t) - \sum_{s=-m}^{m} \tilde{\gamma}_s' \eta_{t-s} = \beta' x_t + (u_t^* + \varepsilon_t)$ where

$$u_t^* = u_{t-1} - (\tilde{\alpha} - \alpha)' g(t) - \sum_{s=-m}^{m} (\tilde{\gamma}_s - \gamma_s)' \eta_{t-s}. \tag{A-2}$$

Then

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

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

$$\begin{aligned} T^{-2} \sum u_t^* u_t^{*\prime} &= T^{-2} \sum (u_{t-1} - (\tilde{\alpha} - \alpha)' g(t))(u_{t-1}' - g(t)'(\tilde{\alpha} - \alpha)) \\ &= T^{-2} \sum u_{t-1} u_{t-1}' - T^{-2} \Big[\sum u_{t-1} g(t)' \Big] (\tilde{\alpha} - \alpha) - (\tilde{\alpha} - \alpha)' T^{-2} \Big[\sum g(t) u_{t-1}' \Big] \\ &\quad + (\tilde{\alpha} - \alpha)' T^{-2} \Big[\sum g(t) g(t)' \Big] (\tilde{\alpha} - \alpha) \\ &\Rightarrow \int B_v B_v' + \pi_v. \end{aligned}$$

where $\pi_v = 4 - \int B_v - \int B_v' - \int r B_v - \int r B_v'$. □

The following lemma establishes the asymptotic distributions of the OLS levels regression estimator and its variance

Lemma 4. *Under the null of cointegration*

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

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

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

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

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

Proof: writing model (4) in matrix form

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

where $Y' = (y_1 \dots y_T)$, $X' = (x_1 \dots x_T)$, $U^{*\prime} = (u_1^* \dots u_T^*)$ (with u_t^* as in (A-2)) and $E' = (\varepsilon_1 \dots \varepsilon_T)$, we obtain under the null

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

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 (7)

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

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

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

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

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

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

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

which leads to the last result in the lemma. □

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

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

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

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

$$\begin{aligned}T^{1/2}\mathrm{vec}(\hat{\beta}_d - \beta) &\Rightarrow \mathcal{N}(0, V_d), \\ T\hat{V}_d &\xrightarrow{p} V_{d\varepsilon},\end{aligned}$$

where R_ε = plim $T^{-1}(I_n \otimes N'D)V_\varepsilon(I_n \otimes D'N)$ with $D = \left[d_{ij}|d_{ii} = -1, d_{i,i+1} = 1, d_{ij} = 0 \text{ otherwise}\right]$ and $N' = (\eta_1, \ldots, \eta_{T-1})$; while under the alternative

$$\begin{aligned}(\hat{\beta}_d - \beta) &\xrightarrow{p} C_{\eta v}(1)\Sigma_\eta^{-1} \\ &= O_p(1) \\ T\hat{V}_d &\xrightarrow{p} V_{dv},\end{aligned}$$

with R_v = plim $T^{-1}(I_n \otimes N')V_v(I_n \otimes N)$.

Therefore, $\hat{\beta}_d$ 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 levels regression.

Proof: from model (5) in matrix form

$$\Delta Y = N\beta + DE + \delta\,\Delta U^*$$

where $(\Delta Y)' = (\Delta y_2, \ldots, \Delta y_T)$ and $(\Delta U^*)' = (\Delta u_2^*, \ldots, \Delta u_T^*)$ with $\Delta u_t^* = v_{t-1} - (\tilde{\alpha}_1 - \alpha_1)' - \sum_{s=-m}^{m}(\tilde{\gamma}_s - \gamma_s)'\Delta\eta_{t-s}$. We can express the OLS estimator as

$$(\hat{\beta}_d - \beta) = (N'N)^{-1}N'\,[DE + \delta\,\Delta U^*] \tag{A-4}$$

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

$$\mathrm{vec}(T^{-1/2}N'DE) = T^{-1/2}(I_n \otimes N'D)\varepsilon \tag{A-5}$$

where $\varepsilon = \mathrm{vec}E' = (\varepsilon_1', \ldots, \varepsilon_T')'$. We note that $\mathcal{E}(\varepsilon\varepsilon') = V_\varepsilon$ defined in Section 3 as the $(nT \times nT)$ covariance matrix of process $\{\varepsilon_t\}$.

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

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

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

$$\mathrm{plim}\ T^{-1}\bar{N}'\bar{N} = R_\varepsilon = \mathrm{plim}\ T^{-1}(I_n \otimes N'D)V_\varepsilon(I_n \otimes D'N),$$

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

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

Substituting in (A-5) we get

$$\operatorname{vec}(T^{-1/2}N'DE) = T^{-1/2}\bar{N}'\bar{\varepsilon} \Rightarrow \mathcal{N}(0, R_\varepsilon).$$

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

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

On the other hand, from (7)

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

where $\hat{V}_\varepsilon \xrightarrow{p} V_\varepsilon$. Hence $T^{-1}\hat{R} \xrightarrow{p} R_\varepsilon$ and $T\hat{V}_d \xrightarrow{p} V_{d\varepsilon}$.

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

$$(\hat{\beta}_d - \beta) = (N'N)^{-1}N'\left[DE + \Delta U^*\right]$$

where

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

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

$$(\hat{\beta}_d - \beta) \xrightarrow{p} C_{\eta v}(1)\Sigma_\eta^{-1} = \Theta_d.$$

On the other hand

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

where $\hat{V}_v \xrightarrow{p} V_v$, the covariance matrix of process v_t. Hence $T^{-1}\hat{R} \xrightarrow{p} R_v$ and $T\hat{V}_d \xrightarrow{p} V_{dv}$. □

Lemma 6. *In the multivariate regression model* (1) *under the null of cointegration*

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

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

where V_l and V_d are as defined in the preceding lemmas.

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

Proof:

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

From lemmas 4 and 5 $\sqrt{T}\operatorname{vec}(\hat{\beta}_d - \beta) \Rightarrow \mathcal{N}(0, V_d)$, $T\operatorname{vec}(\hat{\beta}_l - \beta) \Rightarrow \mathcal{N}(0, V_l)$. Therefore, their joint distribution is also Gaussian. Let C_{dl} denote the asymptotic covariance matrix of $\sqrt{T}\operatorname{vec}(\hat{\beta}_d - \beta)$ and $T\operatorname{vec}(\hat{\beta}_l - \beta)$. Then we may write

$$\begin{bmatrix} \sqrt{T}\operatorname{vec}(\hat{\beta}_d - \beta) \\ T\operatorname{vec}(\hat{\beta}_l - \beta) \end{bmatrix} \Rightarrow \mathcal{N}\left[0, \begin{pmatrix} V_d & C_{dl}' \\ C_{dl} & V_l \end{pmatrix}\right]$$

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

$$\begin{pmatrix} 1 & -T^{-1/2} \end{pmatrix}\begin{pmatrix} V_d & C_{dl}' \\ C_{dl} & V_l \end{pmatrix}\begin{pmatrix} 1 \\ -T^{-1/2} \end{pmatrix} = V_d - T^{-1/2}(C_{dl} + C_{dl}') + T^{-1}V_l$$

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 $\Theta_c = \Theta_d - \Theta_l$

$$c = \operatorname{vec}(\hat{\beta}_d - \beta) - \operatorname{vec}(\hat{\beta}_l - \beta) \Rightarrow \operatorname{vec}\Theta_c.$$

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

$$\begin{aligned} T^{-1}\mathrm{H2} &= T^{-1}c'\hat{V}_d^{-1}c = c'(T\hat{V}_d)^{-1}c \Rightarrow [\mathrm{vec}(\Theta_c\Sigma_\eta^{-1})]'R[\mathrm{vec}(\Theta_c\Sigma_\eta^{-1})] \\ &= O_p(1), \end{aligned}$$

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

$$\begin{aligned} T^{-1}\mathrm{H1} &= T^{-1}c'(\hat{V}_d + \hat{V}_d)^{-1}c = c'(T\hat{V}_d + T\hat{V}_l)^{-1}c \\ &\Rightarrow (\mathrm{vec}\Theta_c)'\Big[(I_n \otimes \Sigma_\eta^{-1})R(I_n \otimes \Sigma_\eta^{-1}) + \Theta_u \otimes \Big(\int M(B_\eta)M(B_\eta)'\Big)^{-1}\Big]^{-1}\mathrm{vec}\Theta_c \\ &= O_p(1). \end{aligned}$$

As $\int MM' > 0$ and $\Theta_u > 0$ we also have that $\operatorname{plim} T^{-1}H1 < \operatorname{plim} T^{-1}H2$. □

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.

Choi, C.-Y., Hu, L., Ogaki, M., 2008. Robust estimation for structural spurious regressions and a Hausman-type cointegration test. Journal of Econometrics 142, 327–351.

Choi, I., 1994. Durbin-Hausman tests for cointegration. Journal of Economic Dynamics and Control 18, 467–480.

Choi, I., Ahn, B. C., 1995. Testing for cointegration in a system of equations. Econometric Theory 11, 952–983.

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.

Fernández-Macho, J., 1994. Hausman-like and variance-ratio tests statistics for the null of cointegration (with a detour on higher order and seasonal integration). BILTOKI DT 94.15, University of the Basque Country (UPV/EHU).

Fernández-Macho, J., 1996. Spectral maximum likelihood estimation of a signal to noise ratio lying in the vicinity of zero. Journal of Time Series Analysis 17 (5), 447–459.

Fernández-Macho, J., Mariel, P., 1994a. Hausman-like and variance-ratio tests for cointegrated regressions. In: MME94: Sborník Referátů. Vysoké Školy Ekonomické Praha, pp. 27–42.

Fernández-Macho, J., Mariel, P., 1994b. Testing the null of cointegration: Hausman-like tests for regressions with a unit root. BILTOKI DT 94.18, University of the Basque Country (UPV/EHU).

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

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.

Hansen, B. E., 1992. Tests for parameter instability in regressions with I(1) processes. Journal of Business and Economic Statistics 10, 321–335.

Harris, D., 1997. Principal components analysis of cointegrated time series. Econometric Theory 13, 529–557.

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.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59 (6), 1551–1580.

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.

Leybourne, S. J., McCabe, B. P. M., 1989. On the distribution of some test statistics for parameter constancy. Biometrika 76, 169–77.

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

Park, J. Y., 1990. Testing for Unit Roots and Cointegration by Variable Addition. Vol. 8 of Advances in Econometrics: Co-integration, Spurious Regressions and Unit Roots. Greenwich, JAI Press, pp. 107–133.

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

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., 1988. Estimation and inference in models of cointegration: A simulation study. Discussion Paper 881, Cowles Foundation.

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.

Phillips, P. C. B., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. Econometrica 58 (1), 165–193.

Ploberger, W., Kramer, W., Kontrus, K., 1989. A new test for structural stability in the linear regression model. Journal of Econometrics 40, 307–318.

Quintos, C. E., Phillips, P. C. B., 1993. Parameter constancy in cointegrating regressions. Empirical Economics 18 (4), 675–706.

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.

Schwert, G. W., 1989. Tests for unit roots: A Monte Carlo investigation. Journal of Business and Economic Statistics 7, 147–159.

Shephard, N., 1993. Maximum likelihood estimation of regression models with stochastic trend components. Journal of the American Statistical Association 88, 590–595.

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.

Xiao, Z., 1999. A residual based test for the null hypothesis of cointegration. Economics Letters 64, 133–141.

TABLE 2. Critical Values for the $H1$ and $H2$ Statistics ($n = 1$).

T	k	*H*1 statistic *mean*	*0.90*	*0.95*	*0.99*	*H*2 statistic *mean*	*0.90*	*0.95*	*0.99*
20	*1*	*0.75*	2.028	2.927	5.307	*1.00*	2.712	3.855	7.047
	2	*1.44*	3.359	4.506	7.644	*1.95*	4.547	6.078	10.180
	3	*2.10*	4.516	5.878	9.174	*2.86*	6.066	7.898	13.289
	4	*2.69*	5.471	7.059	10.870	*3.64*	7.425	9.502	14.924
	5	*3.22*	6.397	8.184	12.991	*4.31*	8.511	10.789	16.843
50	*1*	*0.83*	2.262	3.252	5.414	*0.97*	2.651	3.810	6.242
	2	*1.63*	3.753	4.914	7.944	*1.98*	4.590	6.091	9.544
	3	*2.36*	4.951	6.324	9.452	*2.95*	6.221	7.956	11.629
	4	*3.07*	6.093	7.525	10.778	*3.89*	7.708	9.574	13.418
	5	*3.70*	7.061	8.480	12.127	*4.77*	9.084	10.899	15.450
100	*1*	*0.89*	2.429	3.488	6.046	*0.98*	2.652	3.850	6.622
	2	*1.74*	4.072	5.180	7.821	*1.96*	4.544	5.908	8.993
	3	*2.59*	5.478	6.879	10.134	*2.98*	6.297	7.858	11.544
	4	*3.36*	6.541	7.978	11.368	*3.94*	7.663	9.346	13.151
	5	*4.10*	7.663	9.171	12.720	*4.91*	9.094	10.960	15.144
250	*1*	*0.97*	2.669	3.761	6.591	*1.01*	2.795	3.913	6.887
	2	*1.90*	4.400	5.859	8.797	*2.00*	4.623	6.166	9.285
	3	*2.80*	5.851	7.305	10.730	*3.00*	6.277	7.872	11.588
	4	*3.68*	7.164	8.778	12.250	*3.97*	7.734	9.428	13.171
	5	*4.49*	8.310	10.075	13.477	*4.91*	9.088	10.909	14.814
500	*1*	*0.96*	2.620	3.723	6.039	*0.98*	2.673	3.792	6.087
	2	*1.96*	4.520	5.929	9.156	*2.02*	4.629	6.075	9.407
	3	*2.89*	6.048	7.553	10.900	*2.99*	6.256	7.812	11.268
	4	*3.84*	7.488	9.127	12.580	*4.01*	7.813	9.521	13.164
	5	*4.80*	8.920	10.761	14.411	*5.04*	9.332	11.338	15.346
∞	*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	11.071	15.086	*5.00*	9.236	11.071	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 k d.f.

TABLE 3. Critical Values for the $H1$ and $H2$ Statistics ($n = 2$).

		$H1$ statistic				$H2$ statistic			
T	k	*mean*	*0.90*	*0.95*	*0.99*	*mean*	*0.90*	*0.95*	*0.99*
50	*1*	*1.70*	3.894	5.178	8.202	*2.03*	4.620	6.093	9.631
	2	*3.25*	6.365	7.904	11.481	*4.03*	7.848	9.798	14.075
	3	*4.70*	8.513	10.194	13.769	*6.02*	10.818	12.923	17.768
	4	*6.07*	10.361	12.225	16.146	*8.02*	13.655	16.033	21.301
100	*1*	*1.82*	4.153	5.436	8.265	*2.00*	4.600	5.973	9.091
	2	*3.56*	6.986	8.520	11.925	*4.02*	7.835	9.565	13.671
	3	*5.20*	9.268	11.133	15.046	*6.01*	10.762	12.718	17.110
	4	*6.74*	11.373	13.266	17.474	*7.99*	13.409	15.632	20.702
250	*1*	*1.94*	4.357	5.773	8.833	*2.02*	4.538	6.041	9.141
	2	*3.81*	7.451	9.028	12.614	*4.02*	7.862	9.593	13.412
	3	*5.61*	9.882	11.699	16.054	*5.99*	10.580	12.482	17.130
	4	*7.40*	12.318	14.553	18.513	*8.00*	13.416	15.707	19.897
500	*1*	*1.91*	4.378	5.748	8.711	*1.95*	4.468	5.868	8.893
	2	*3.88*	7.531	9.147	12.975	*3.99*	7.713	9.403	13.377
	3	*5.79*	10.305	12.157	16.472	*6.00*	10.649	12.615	17.006
	4	*7.73*	12.895	15.095	19.801	*8.05*	13.473	15.699	20.526
∞	*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 4. Critical Values for the $H1$ and $H2$ Statistics ($n = 3$).

		$H1$ statistic				$H2$ statistic			
T	k	*mean*	*0.90*	*0.95*	*0.99*	*mean*	*0.90*	*0.95*	*0.99*
50	*1*	*2.59*	5.435	6.801	10.021	*3.10*	6.519	8.123	11.719
	2	*4.99*	8.976	10.702	14.594	*6.23*	11.185	13.275	18.299
	3	*7.23*	12.108	14.007	18.346	*9.34*	15.656	18.159	23.283
100	*1*	*2.76*	5.730	7.216	10.927	*3.04*	6.331	7.964	11.983
	2	*5.36*	9.529	11.314	15.414	*6.06*	10.681	12.868	17.469
	3	*7.88*	12.922	14.734	19.224	*9.15*	14.943	17.202	22.264
250	*1*	*2.93*	6.115	7.651	11.054	*3.05*	6.395	7.981	11.484
	2	*5.73*	10.154	11.907	16.216	*6.05*	10.732	12.572	17.173
	3	*8.44*	13.935	16.074	20.289	*9.02*	14.903	17.092	21.719
500	*1*	*2.91*	6.012	7.518	10.853	*2.97*	6.147	7.678	11.030
	2	*5.84*	10.475	12.325	16.292	*6.01*	10.788	12.647	16.671
	3	*8.70*	14.367	16.537	21.631	*9.01*	14.850	17.161	22.226
∞	*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 5. Empirical size and power of cointegration tests.

	$\mathcal{H}_0$: *Cointegration*					$\mathcal{H}_a$: *No Coint.*	
	WN	AR(1)				RW	IMA(1,1)
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 1.0$	$\theta = 1.0$
H1	0.86	1.80	3.33	7.10	24.59	58.77	61.50
H1^{qs}	0.86	1.80	3.36	6.88	22.76	53.71	56.34
H2	0.96	2.08	3.82	8.32	28.54	**74.38**	**79.38**
H1^{dyn}	4.00	4.19	4.85	10.95	30.89	55.59	58.08
H2^{dyn}	4.20	4.49	5.39	12.12	**34.25**	66.90	70.02
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
L$_c$	12.72	11.54	10.10	8.10	5.18	8.60	11.00

% rejections at 5% significance level. Sample size $T = 100$. Column maxima in bold. H^{qs}: QS kernel with automatic bandwidth and elimination of leads and lags by BIC selection. H: *idem* with Bartlett kernel and T-dependent bandwidth. H^{dyn}: *idem* with leads and lags in CI equation. Simulations using 10 000 replications from DGP 2 (parameters values as in (8)).

TABLE 6. Empirical size and power of 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
H1	0.86	13.68	35.20	48.49	54.91	58.77	60.80	61.31	**61.76**	61.64	61.50
H1^{qs}	0.86	10.61	30.59	43.39	50.17	53.71	55.61	56.20	**56.67**	56.36	56.34
H2	0.96	16.75	43.03	60.41	69.25	74.38	77.14	78.24	78.80	79.33	**79.38**
H1^{dyn}	4.00	19.48	37.34	47.25	52.92	55.59	57.04	57.75	**58.10**	57.97	58.08
H2^{dyn}	4.20	22.85	44.04	56.37	63.38	66.90	68.86	69.67	70.03	70.03	**70.03**
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
L$_c$:	12.72	**72.66**	72.11	39.45	15.60	8.60	9.17	9.72	10.73	11.18	11.00

% 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 (8)).

TABLE 7. Power of cointegration tests as a function of sample size.

5% significance level:

T	*20*	*30*	*40*	*50*	*100*	*150*	*200*	*250*	*500*	*1 000*
H1	32.93	39.05	42.49	47.31	58.77	64.58	68.22	69.58	79.43	83.42
H1^{qs}	30.65	33.17	37.65	41.31	53.71	59.19	62.76	65.77	75.51	81.11
H2	**52.69**	**57.47**	**62.61**	**65.60**	**74.38**	**78.17**	**81.04**	**82.62**	88.52	92.01
H1^{dyn}	28.47	33.83	36.93	41.92	55.59	61.55	65.04	66.86	77.25	82.17
H2^{dyn}	43.67	48.52	53.13	56.56	66.90	72.85	75.77	78.33	84.89	89.48
C(qs)	29.42	40.94	40.41	46.84	63.53	67.18	75.25	75.78	**90.08**	**95.50**
C(ℓ_4)	7.61	15.19	22.26	28.18	45.44	53.54	61.00	64.98	80.08	90.72
LBI	11.51	24.40	22.24	28.95	48.45	51.30	60.00	59.74	78.18	86.31
L$_c$	15.50	10.60	7.20	7.10	8.60	9.60	11.90	14.40	28.80	43.94

1% significance level:

T	*20*	*30*	*40*	*50*	*100*	*150*	*200*	*250*	*500*	*1 000*
H1	19.51	25.75	29.66	34.23	47.33	54.27	58.82	60.98	72.71	78.56
H1^{qs}	18.29	20.11	24.79	28.45	41.60	48.56	52.97	56.11	67.81	75.84
H2	**41.42**	**46.71**	**52.55**	**56.45**	**66.92**	**72.33**	**75.31**	**78.07**	**85.01**	**89.40**
H1^{dyn}	16.06	20.20	23.74	29.16	43.65	50.06	54.59	57.32	69.87	76.36
H2^{dyn}	32.31	36.31	41.61	45.73	58.11	65.00	68.42	71.29	80.19	86.00
C(qs)	5.37	18.81	18.82	25.77	46.81	50.69	59.64	59.88	78.62	87.21
C(ℓ_4)	1.89	0.64	1.73	6.43	25.50	35.19	43.27	48.03	64.12	78.54
LBI	0.05	1.51	0.82	5.97	28.63	32.15	42.22	42.31	62.24	71.88

% rejections. Column maxima in bold. Simulations using 10 000 replications from DGP 2 with random walk errors ($\phi = 1$, $\theta = 0$; rest of parameters as in (8)).