Tests for the Null Hypothesis of Cointegration: A Monte Carlo Comparison

Vasco J. Gabriel*

Department of Economics, Birkbeck College, University of London, London, UK, and University of Minho, Braga, Portugal

ABSTRACT

The aim of this paper is to compare the relative performance of several tests for the null hypothesis of cointegration, in terms of size and power in finite samples. This is carried out using Monte Carlo simulations for a range of plausible data-generating processes. We also analyze the impact on size and power of choosing different procedures to estimate the long run variance of the errors. We found that the parametrically adjusted test of McCabe et al. (1997) is the most well-balanced test, displaying good power and relatively few size distortions.

Key Words: Cointegration; Tests; Monte Carlo.

JEL Classification: C12; C15; C22.

1. INTRODUCTION

The problem of testing for cointegration has been a central issue in the literature on economic time series. Usually, in the case where a single cointegrating vector is expected to exist, testing is carried out by means of residual based procedures that consist of

*Correspondence: Vasco J. Gabriel, University of Minho, 4710-057 Braga, Portugal; Fax: +351 253 676 375; E-mail: vjgabriel@eeg.uminho.pt.

extensions of unit root tests, i.e., one tests whether the residuals from the cointegrating regression contain a unit root or, by contrast, are $I(0)$. The null hypothesis is, thus, of no cointegration, against the alternative of cointegration. However, this approach seems unnatural, especially if one takes the existence of a long run equilibrium relationship among the variables as the hypothesis of interest, stemming from economic theory (e.g., the link between consumption and disposable income). With cointegration as the null hypothesis, one would reject it only if the data would provide strong evidence against the maintained hypothesis, unlike the situation where the hypotheses are reversed.

There have been some recent attempts to test directly for cointegration in single equation models. One route has been to extend existing univariate test procedures to test for an $I(0)$ null against an $I(1)$ alternative, such as the tests advocated by Kwiatkowski et al. (1992, henceforth KPSS), Leybourne and McCabe (1994) and Saikkonen and Lukkonen (1993), among others [see Stock (1994) for a survey]. These tests emerged from the apparently unrelated literature on testing for unit moving average roots [see Tanaka (1990), for example] and testing time-varying parameters [see Nabeya and Tanaka (1988), inter alia], and are one-sided LM tests with asymptotic locally optimal power properties. In this way, Harris and Inder (1994), Leybourne and McCabe (1993), McCabe et al. (1997) and Shin (1994) devised tests that generalize the so-called KPSS statistic to the context of cointegration. The main difference between these versions lies on the proposed estimation method to obtain the residuals and the variance, subsequently used to construct the test statistic.

A related test was suggested by Hansen (1992), although it was primarily conceived to test for parameter instability in cointegration models. Under the alternative hypothesis, each coefficient in the model is allowed to follow a random walk, so by testing the stability of the estimated parameters, one is also testing for cointegration. On the other hand, Park (1990) developed a test for the null of cointegration based on the addition of superfluous regressors to the cointegrating regression. More recently, Xiao (1999) proposed a residual based test that examines the fluctuation of the residuals from a regression.

The aim of this paper is to compare the relative performance of these testing approaches, in terms of size and power in finite samples. This is carried out using Monte Carlo simulations for a range of plausible data-generating processes. As of this writing, there is no study providing guidance on the use of this type of procedures in empirical situations, with the exception of the limited studies of Haug (1996) and McCabe et al. (1997). Moreover, it would be useful to know which tests are best suited for conducting confirmatory analysis, i.e., applying tests for the null of cointegration in conjunction with the standard tests for the null of no cointegration. If the two approaches give consistent results (i.e., there is an acceptance and a rejection of the nulls), one may conclude whether the series are cointegrated or not, whereas if both tests either reject or accept their respective null hypotheses, the results are inconclusive. See, for example, Maddala and Kim (1998) and Shin (1994) for a discussion, as well as Carrion-i-Silvestre et al. (2001) and Charemza and Syczewska (1998) for an application to the univariate case. Gabriel (2003) addresses the issue in a cointegration context, casting doubt on the use of this methodology.

The purpose of the paper is twofold. Besides the distinct performances due to the way each test is constructed, another important issue investigated is the impact on size and power of choosing different procedures to estimate the long run variance of the errors. Most of the tests analyzed here depend on the estimation of this nuisance parameter and it is well known that the use of semi-parametric estimators may lead to substantially oversized tests in samples of small size. Some results are known for stationarity tests [Hobijn et al., 1998; Lee, 1996], but there is little evidence concerning tests for cointegration, although one may expect similar conclusions.

We focus our attention on single equation methods, still very much used in empirical practice because of their simplicity and also when a single cointegrating vector is believed to exist, thus with no efficiency losses. Tests for the null of cointegration in systems of equations using more intricate methods have also been developed, such as those of Choi and Ahn (1995), using Park (1992) canonical cointegration regression (CCR), Harris (1997) and Snell (1999), based on Principal Components methods, and Nyblom and Harvey (2000).

The paper is organized as follows. Section 2 presents a general model of cointegration and briefly reviews the methods for estimating the long run variance. In Sec. 3, the tests for the null of cointegration are presented. The DGPs for the Monte Carlo experiments, as well as the simulation results, are analyzed in Sec. 4. Section 5 summarizes and concludes.

2. THE ISSUES

2.1. The Basic Model

Since each test considered here was derived under a specific model, it is difficult to present a common formulation for all tests. Nevertheless, we may write a general model as

$$y_t = \alpha + x_t'\beta + u_t \tag{1}$$

where y_t is a scalar $I(1)$ process and x_t is a $k \times 1$ vector $I(1)$ process, such that $\Delta x_t = v_t$, v_t being a k-vector stationary process and β is a vector of unknown coefficients. For simplicity, we concentrate on the single equation specification with an intercept, although more general specifications could be considered (e.g., containing time trends). The variables y_t and x_t are said to be cointegrated if u_t is $I(0)$, whereas if u_t is $I(1)$ there is no long run equilibrium relationship between y_t and x_t.

Some tests also differ on how the disturbance term is specified under the alternative hypothesis of no cointegration, as will be seen later. Under the null hypothesis of cointegration, $\zeta_t = (u_t, v_t')'$ follows a general stationary process obeying some mild regularity conditions and satisfies a multivariate invariance principle, such that

$$T^{-1/2}\sum_{t=1}^{[Tr]} \zeta_t \Rightarrow B(r), \quad \text{as } T \to \infty$$

Here, "$\Rightarrow$" denotes weak convergence, $[\,\cdot\,]$ is the integer part operator and $B(r)$ is a $k+1$ dimensional Brownian motion defined on $r \in [0, 1]$, with long run covariance matrix $\Omega = \lim_{T\to\infty} \mathrm{var}(T^{-1/2}\sum_{t=1}^{T} \zeta_t)$. These conditions allow for any stationary and invertible ARMA process, possibly with heterogeneous innovations.

We partition $B(r)$ conformably with $\zeta_t = (u_t, v_t')'$ as $B(r) = [B_1(r), B_2(r)']'$ and

$$\Omega = \begin{bmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

where ω_{11} is the long run variance of u_t. We will restrict our attention to the case where cointegration among the regressors is excluded, so that Ω_{22} is positive definite. If we allow for correlation between u_t and v_t, then an asymptotically efficient estimation method should be used to account for the endogeneity of the regressors, such as the fully-modified OLS method of Phillips and Hansen (1990) or the leads and lags OLS estimator of Saikkonen (1991) and Stock and Watson (1993), for example. In this case, we are also interested in the long run variance of u_t conditional on v_t, defined by $\omega_{1.2} = \omega_{11} - \Omega_{12}\Omega_{22}'\Omega_{21}$, which plays an important role in the construction of the test statistics studied in this paper.

2.2. Long Run Variance Estimation

The long run variance of u_t is usually estimated from

$$s^2(l) = \hat{\gamma}_0 + 2\sum_{j=1}^{l} w(j, l)\hat{\gamma}_j \tag{2}$$

where $\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{u}_t\hat{u}_{t-j}(j = 0, 1, \ldots, T)$ is the estimate of the j-th autocovariance and $w(j, l)$ is a kernel function depending on a bandwidth parameter (or truncation lag) l. Thus, the estimation depends on the chosen kernel and on the procedure to truncate the number of autocovariances.

Here, we compare three widely used kernels, namely the Parzen, the Bartlett and the Quadratic Spectral (QS) kernels [see Andrews (1991) for details]. Concerning the truncation lag selection, we study two procedures widely used in the literature, namely those of Andrews (1991) and Newey and West (1994),[a] see Den Haan and Levin (1997) for a survey. Both methods are data-based and nonparametric, consisting of obtaining, for a given kernel, the optimal bandwidth parameter sequence that minimizes asymptotic mean square errors. Andrews (1991) suggested the use of an automatic plug-in bandwidth estimator, which has the form

$$\begin{aligned} \hat{l}_T &= 1.1147[\hat{\alpha}(1)T]^{1/3} \quad \text{(Bartlett kernel)} \\ \hat{l}_T &= 2.6614[\hat{\alpha}(2)T]^{1/5} \quad \text{(Parzen kernel)} \\ \hat{l}_T &= 1.3221[\hat{\alpha}(2)T]^{1/5} \quad \text{(QS kernel)} \end{aligned} \tag{3}$$

where $\hat{\alpha}(1)$ and $\hat{\alpha}(2)$ are obtained from fitting AR(1) regressions to the residuals. This procedure has the advantage of avoiding an arbitrary choice of l, as practiced in previous literature [see Kwiatkowski et al. (1992), for example].

[a]In a previous version of this paper, available upon request, a deterministic selection rule as in Kwiatkowski et al. (1992) is also analyzed.

However, when testing for stationarity, this method requires truncation to assure consistency under the $I(1)$ hypothesis, since $\hat{l}_T$ becomes $O_p(T)$, as pointed out by Choi and Ahn (1995). To circumvent this problem, these authors suggest estimating the lag truncation parameter subject to a restriction, i.e., fixing $l = 2$ if $\hat{l}_T$ is larger than T^{ε}, where ε is a chosen number such that $0.5 < \varepsilon < 1$. Another possibility was proposed by Stock (1994), which involves the truncation of $\hat{l}_T$ after a given lag m, i.e., $\hat{l} = \min(\hat{l}_T, m(T/100)^{1/5})$. In this paper, we follow the suggestions of these authors by fixing $\varepsilon = 0.65$ [see Choi and Ahn (1995), p. 969] and $m = 12$ [see Stock (1994), p. 2803].

Note that with these methods the problem of the arbitrary choice of the bandwidth is reintroduced, as they depend on the choice of l, ε or m. This is especially relevant when the errors are strongly correlated (i.e., approaching the alternative hypothesis), since $\hat{l}_T$ will quickly become large, and, thus, the automatic bandwidth estimator will return the deterministic choice, therefore conditioning the performance of the tests. This will become apparent in our subsequent simulation study.

Another approach studied here is the procedure proposed by Newey and West (1994).[b] It is similar to the method of Andrews (1991), although they use nonparametric estimates of $\hat{\alpha}(1)$ and $\hat{\alpha}(2)$ in (3) to construct the automatic plug-in bandwidth estimator, instead of autoregressions. The nonparametric estimation of these parameters also depends on a parameter j which is selected with a non-stochastic rule, $j = \text{integer}[c(T/100)^d]$, where d depends on the kernel being used and c is fixed a priori. We choose $c = 4$, as recommended by Hobijn et al. (1998) and Newey and West (1994), which was also the value that produced better results, having also experimented $c = 2$, 8 and 12. According to these authors, the impact of this choice appears to be not very substantial for the final outcome, $s^2(l)$.

Further refinements have been proposed, namely a "pre-whitening" procedure (Andrews and Monahan, 1992) that filters the residuals with an AR regression in order to make them closer to white noise and then calculate the spectral density at the origin. However, this method renders null of cointegration tests inconsistent, since under the alternative of no cointegration (i.e., a unit root in the residuals) the AR estimate would be close to unity, the long run variance would tend to infinity and, thus, the tests statistics close to zero[c] [see discussion in Lee (1996) and Shin (1994), for example].

3. TESTS FOR THE NULL HYPOTHESIS OF COINTEGRATION

In this section, we will briefly describe the cointegration tests examined in the subsequent Monte Carlo study. As said earlier, we may group the tests into four different categories, according to the way the test is constructed.

[b]The application of the Newey–West method to univariate stationarity tests is documented in Hobijn et al. (1998).

[c]This fact helps to explain the poor performance of null of cointegration tests in Haug (1996), since the author prewhitens the residuals before constructing the test statistics.

3.1. Variable Addition Test

Park (1990) suggested an approach for testing the null hypothesis of cointegration, which consists of including a set of superfluous regressors z_t in the cointegration regression, so that

$$y_t = \alpha + x_t'\beta + z_t'\delta + e_t \tag{4}$$

If the variables in (1) are truly cointegrated, then the added regressors in (4) will not be significant, while the opposite holds if the regression is spurious. Standard significance tests (such as Wald) will be able to discriminate between the two situations. The test statistic may be written as

$$J_1 = \frac{\text{RSS}_1 - \text{RSS}_2}{\hat{\omega}_{1.2}} \tag{5}$$

where RSS_1 and RSS_2 are, respectively, the residual sum of squares from (1) and (4). The denominator is an estimate obtained with a consistent estimator of the (conditional) long run variance of u_t. A particular advantage of this test is that under the null hypothesis of cointegration (i.e., $\delta = 0$), J_1 has a limiting $\chi^2(q)$ distribution, where q is the dimension of the set of superfluous regressors, therefore avoiding extensive tabulations. If the alternative is true, J_1 diverges at a rate dependent on the chosen bandwidth for $\hat{\omega}_{1.2}$.

As recommended by Park (1990), we use a set of polynomial time trends $z_t = (t, t^2, t^3)$ as superfluous regressors in the Monte Carlo experiments.[d] Regarding the estimation method, the J_1 test may be implemented with any asymptotically efficient procedure, so we use FM-OLS, although the test was derived in the context of the CCR method[e] developed by Park (1992).

3.2. Fluctuation Test

A different approach is followed by Xiao (1999), by deriving a residual based test for the null of cointegration based on the fluctuation of the residuals $\hat{u}_t$ from the cointegrating regression. The idea is quite simple: if cointegration holds, the residuals will replicate the stationary behavior of the errors and will display a limited amount of fluctuation, whilst if the residuals fluctuate too much the converse should be true.

The fluctuation principle was originally proposed to study the stability of the estimated coefficients of a model [see Ploberger et al. (1989), for example]. Using the

[d]Computer generated random walks were also experimented as additional regressors, although with worse results in the subsequent simulations.

[e]Some preliminary simulations using CCR and FM-OLS revealed that the estimation method has no impact on the performance of the test.

FM-OLS method, Xiao (1999) constructs a statistic that is asymptotically free of nuisance parameters, based on the recursive estimates statistic

$$R_T = \max_{i=1,\dots,T} \frac{i}{\sqrt{\hat{\omega}_{1.2}T}} \left| \frac{1}{i}\sum_{t=1}^{i} \hat{u}_t^{+} - \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^{+} \right| \tag{6}$$

where $\hat{u}_t^{+}$ are the residuals resulting from FM-OLS estimation. The limiting distribution of R_T is non-standard and depends only on the dimension of the set of regressors.

3.3. L_c Test

Hansen (1992) proposed some LM-type structural change tests in cointegrated models, making use of the FM-OLS method. A versatile feature of those tests is the possibility of using them as cointegration tests. In fact, if the alternative hypothesis is the intercept following a random walk, then structural change testing becomes cointegration testing, albeit with the null hypothesis of cointegration. Decomposing u_t in (1) such that $u_t = w_t + v_t$, w_t being a random walk and v_t a stationary term, the model then becomes

$$y_t = \alpha_t + x_t'\beta + v_t \tag{7}$$

with $\alpha_t = \alpha + w_t$, i.e., the intercept "absorbs" the random walk w_t when there is no cointegration.

In view of this fact, Hansen (1992) suggested the use of the statistic

$$L_c = \frac{\sum_{t=1}^{T} \hat{s}_t'\hat{M}_t^{-1}\hat{s}_t}{\hat{\omega}_{1.2}T} \tag{8}$$

to test the null of cointegration, where $\hat{s}_t$ represents the scores of the FM-OLS estimates and the weighting matrix $\hat{M}$ is the moments matrix of the regressors. However, this statistic was designed to test the stability of the whole cointegration vector, so there are advantages in regarding a version that tests only (partial) structural change in the intercept. Hao has shown that such a test may be carried out by employing the KPSS statistic to test for the null of cointegration. This is considered next.

3.4. Tests Based on the KPSS Statistic

The tests previously presented are tests for a general null hypothesis of stationary errors [i.e., u_t in (1) is $I(0)$], against a general alternative of $I(1)$ errors. However, it is possible to formulate with more detail the behavior of u_t, both under the null and the alternative. Assume for the moment that y_t and x_t do not cointegrate and that u_t in (1) may be decomposed into the sum of a random walk and stationary component,

$$u_t = \gamma_t + \varepsilon_t \tag{9}$$

where the random walk is $\gamma_t = \gamma_{t-1} + \eta_t$, with $\gamma_0 = 0$[f] and η_t distributed as $i.i.d.(0, \sigma_\eta{}^2)$, while the stationary part ε_t is distributed as $i.i.d.(0, \sigma_\varepsilon^2)$[g] and is assumed independent of η_t [note the similarity with the model in (7)].

Cointegration stems from this formulation when $\sigma_\eta{}^2 = 0$, so that $\gamma_t = 0$ and no longer is a random walk. Therefore a test for cointegration has the null hypothesis H_0: $\sigma_\eta{}^2 = 0$ against the alternative $H_1{:}\sigma_\eta{}^2 > 0$. Following Kwiatkowski et al. (1992), an asymptotically equivalent test to the locally best invariant (LBI) test of H_0 against H_1 uses the LM-type statistic

$$L = T^{-2}\frac{\sum_{t=1}^{T} S_t{}^2}{s^2(l)} \tag{10}$$

where S_t is the partial sum process $S_t = \sum_{t=1}^{i} \hat{u}_t$ of the residuals from (1) and $s^2(l)$ is a consistent estimate of ω_{11}. Allowing for correlation between ε_t and v_t calls for the use of an efficient estimation method and the denominator should be replaced by $\hat{\omega}_{1.2}$, as discussed in Sec. 2. Different versions of this approach, using distinct estimation methods, have been proposed, as can be seen next.

3.4.1. Nonparametric Versions

Leybourne and McCabe (1993) suggested a version of (10) by considering an OLS regression of (1) and using the corresponding residuals $\hat{u}_t$ to construct the test statistic that we will denote as LM. These authors suggested estimating $\hat{\omega}_{11}$ with a simple truncated autocovariances estimator.[h]

The previous version, by using OLS, does not take into account the potential problems that arise from second-order biases in the estimation of the cointegrating regression. Harris and Inder (1994), using the FM-OLS method, suggested an extension of the KPSS test (denoted as *HI*) that uses an estimate of $\hat{\omega}_{1.2}$ rather than $\hat{\omega}_{11}$ as in LM, reflecting the fact that one is accounting for the possible endogeneity of x_t.

Another way of circumventing the problem of endogenous regressors is to use the estimation method advocated by Saikkonen (1991). The procedure consists of introducing past and future values of Δx_t, so that the regression becomes

$$y_t = \alpha + x_t'\beta + \sum_{j=-n}^{n} \Delta x_{t-j}'\pi_j + u_t^* \tag{11}$$

The truncation parameter n should increase with T at an appropriate rate and may be chosen with any model selection criterion such as AIC or BIC. We use the first criterion,[i] after fixing the maximum lag as $n = \text{integer}[T^{1/3}]$. To simplify, the same value of n is used for both leads and lags of Δx_t. Applying OLS to the modified regression in (11) will yield

[f] This implies no loss of generality, since (1) contains an intercept.

[g] The *i.i.d.* assumption of ε_t may be relaxed so that ε_t may follow a stationary process as discussed for u_t and v_t in Sec. 2.1.

[h] In our sampling experiments, we employ the kernel estimators mentioned in Sec. 2.2.

[i] Results were also obtained using the BIC, but were slightly worse in general.

efficient estimates, see Saikkonen (1991) for a more detailed discussion. A version of (10) may be constructed with the residuals $\hat{u}_t^*$ from (11), thus resulting the Shin test statistic S.

3.4.2. A Parametric Alternative

Unlike all the tests discussed previously, which used a nonparametric procedure to correct for excess correlation in the disturbances, McCabe et al. (1997) devised a parametric approach to test for the null of cointegration. They extend the parametric adjustment procedure of Leybourne and McCabe (1994) to the cointegration case, by considering a different formulation for the error component in (9). In fact, they assume that u_t follows

$$\Phi(L)u_t = \gamma_t + \varepsilon_t \tag{12}$$

where $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is a stable autoregressive polynomial of order p, with γ_t and ε_t as defined previously. Under the null hypothesis of cointegration u_t is a stationary AR(p) process, whereas if $\sigma_\eta{}^2 > 0$, u_t becomes non-stationary, with an ARIMA(p,1,1) representation (Leybourne and McCabe, 1994).

In order to implement the test, McCabe et al. (1997) advocate the use of Saikkonen's dynamic least squares method to estimate (1), but the autoregressive coefficients ϕ_j in (12) should be obtained by maximum likelihood by fitting an ARIMA(p,1,1) model to $\hat{u}_t^*$, the residuals from (11). To select the order p, we follow Leybourne and McCabe (1999), which propose a data-dependent selection criterion based on general-to-specific testing approach, from an initial value of $p = 4$.

The test statistic is then constructed with the "second stage" residuals $\hat{\varepsilon}_t = \hat{u}_t^* - \sum_{i=1}^{p} \hat{\phi}_i \hat{u}_{t-i}^*$ from (12) as

$$\mathrm{MLS} = T^{-2}\frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\hat{\varepsilon}_j\right)^2}{\hat{\sigma}^2} \tag{13}$$

with $\hat{\sigma}^2 = T^{-1}\sum_{i=t}^{T}\hat{\varepsilon}_t$ being a consistent estimator of $\sigma_\varepsilon{}^2$. However, we use an alternative estimator of $\hat{\sigma}^2$ suggested by Leybourne and McCabe (1999), $\hat{\sigma}^2 = \hat{\sigma}_\xi{}^2 \cdot \hat{\theta}$, where $\hat{\sigma}_\xi{}^2$ and $\hat{\theta}$ are the maximum likelihood estimates of the variance and MA coefficient of the ARIMA(p,1,1) auxiliary regression.

This test has, at least theoretically, some advantages comparatively to other KPSS versions. Indeed, the test is consistent at a faster rate, i.e., $O_p(T^2)$, and does not depend on any lag truncation parameter. As mentioned before, all previous statistics diverge at a rate dependent on the chosen bandwidth for $\hat{\omega}_{1.2}$, under the alternative hypothesis. This should be apparent even in terms of its finite sample performance. It also allows for cointegration among the regressors, unlike other tests. However, according to Hobijn et al. (1998), this test is not consistent for the alternative of a pure random walk, although this has recently been disputed by Lanne and Saikkonen (2000) for its univariate version. The following section will help to clarify this matter.

4. MONTE CARLO SIMULATIONS

To evaluate the finite sample performance of the null of cointegration tests discussed above, we develop a series of Monte Carlo experiments. A cautionary note should be overtly stated, however. Indeed, direct comparisons of the relative performances of the tests may be questionable, since they are affected by the choice of the DGPs. In this particular case, a major drawback is the fact that we are analyzing the behavior of several nonparametric tests in a parametric framework. As pointed out in Sec. 2.2, the advantage of a nonparametric formulation lies on the fact that no parametric assumptions on the error's structure are required. However, this robustness is likely to be overshadowed by a loss of efficiency in an explicit parametric context, as in our Monte Carlo experiment. Bearing this caveat in mind, we will try to provide some tentative remarks supported by the results of the simulation study.

4.1. Experimental Design

The general DGP is similar to Mc Cabe et al. (1997) and is based on the models previously presented:

$$\begin{aligned} y_t &= \alpha_t + x_t + \varepsilon_t \\ \alpha_t &= \alpha_{t-1} + \eta_t, \quad \alpha_0 = 1, \quad \eta_t \sim i.i.d.(0, \sigma_\eta{}^2) \\ x_t &= x_{t-1} + \zeta_t, \quad \zeta_t \sim i.i.d.(0, 1) \\ \varepsilon_t &= \rho\varepsilon_{t-1} + \omega_t - \theta\omega_{t-1}, \quad \omega_t \sim i.i.d.(0, 1) \end{aligned} \tag{14}$$

with η_t independent of ζ_t and ε_t. For size analysis, the model under the null hypothesis requires $\sigma_\eta{}^2 = 0$, while under the alternative we choose the values $\sigma_\eta{}^2 = \{0.01, 0.1, 0.5\}$ for power analysis. The other parameters are allowed to take values $\rho = \{0, 0.5, 0.9\}$ and $\theta = 0.5$ (with $\rho = 0$). Additionally to the random-walk-plus-noise model, we study the alternative of a pure random walk, with $\rho = 1$ and $\sigma_\eta{}^2 = 0$. We also considered the effects of correlation between the errors (endogeneity of regressors), with $\mathrm{corr}(\zeta_t, \omega_t) = \gamma = 0.7$ (with $\rho = 0$ and $\theta = 0$, for simplicity). The selected sample sizes are $T = 100$ and 200, most common in empirical studies, and the number of replications is 5000. All simulations were programmed in GAUSS.[j]

4.2. Simulation Results

The results are organized as follows. Tables 1–3 present the estimates of rejection frequencies of the different tests at the 5% level of significance, under the null and alternative hypotheses, for QS and Parzen kernels. For sake of economy, we do not show the results for the Bartlett kernel, since these were, in general, intermediate between the QS and Parzen results. Also, results for $\sigma_\eta{}^2 = 0.01$ are not presented, given that powers are very low for all DGPs, with very little differences in size-corrected power among the tests, which is not surprising, since it is an alternative very close to the null. Nevertheless, the full set of results is available upon request.

[j]Other DGPs were considered in initial exploratory simulations, namely cointegration with no constant and with trend, but the results were not qualitatively different from the case studied here.

Table 1. Rejection frequencies at the 5% significance level, $\sigma_\eta^2 = 0$.

	$T=100$					$T=200$				
	ρ			θ	γ	ρ			θ	γ
$\sigma_\eta^2=0$	0	0.5	0.9	0.5	0.7	0	0.5	0.9	0.5	0.7
MLS	0.054	0.079	0.391	0.098	0.055	0.049	0.059	0.316	0.08	0.051
QS										
S_{NW}	0.055	0.074	0.243	0.061	0.055	0.051	0.067	0.233	0.053	0.049
S_{ACA}	0.045	0.064	0.307	0.049	0.044	0.047	0.061	0.212	0.049	0.048
(S_{AS})			(0.118)					(0.125)		
HI_{NW}	0.045	0.065	0.182	0.048	0.032	0.047	0.07	0.22	0.051	0.033
HI_{ACA}	0.046	0.052	0.346	0.042	0.04	0.052	0.057	0.219	0.046	0.043
(HI_{AS})			(0.074)					(0.108)		
LM_{NW}	0.062	0.083	0.238	0.066	0.129	0.058	0.073	0.247	0.06	0.111
LM_{ACA}	0.051	0.072	0.387	0.055	0.102	0.053	0.065	0.242	0.054	0.105
(LM_{AS})			(0.133)					(0.135)		
$L_c{}^{NW}$	0.106	0.129	0.292	0.109	0.068	0.099	0.125	0.344	0.098	0.07
$L_c{}^{ACA}$	0.088	0.108	0.44	0.09	0.076	0.098	0.105	0.269	0.09	0.081
$(L_c{}^{AS})$			(0.14)					(0.169)		
$R_T{}^{NW}$	0.039	0.048	0.113	0.037	0.018	0.037	0.056	0.179	0.041	0.025
$R_T{}^{ACA}$	0.036	0.033	0.321	0.029	0.027	0.04	0.045	0.191	0.034	0.038
$(R_T{}^{AS})$			(0.021)					(0.062)		
$J_1{}^{NW}$	0.162	0.146	0.44	0.143	0.083	0.105	0.105	0.409	0.094	0.075
$J_1{}^{ACA}$	0.01	0.072	0.508	0.056	0.001	0.005	0.10	0.345	0.045	0.001
$(J_1{}^{AS})$			(0.402)					(0.375)		
Parzen										
S_{NW}	0.054	0.068	0.213	0.057	0.052	0.05	0.065	0.203	0.053	0.047
S_{ACA}	0.045	0.061	0.789	0.049	0.044	0.047	0.062	0.882	0.046	0.047
(S_{AS})			(0.192)					(0.22)		

(*continued*)

Table 1. Continued.

$\sigma_\eta^2=0$	$T=100$					$T=200$				
	ρ			θ	γ	ρ			θ	γ
	0	0.5	0.9	0.5	0.7	0	0.5	0.9	0.5	0.7
HI_{NW}	0.406	0.342	0.124	0.391	0.471	0.613	0.556	0.237	0.617	0.692
HI_{ACA}	0.475	0.684	0.167	0.631	0.517	0.603	0.79	0.298	0.737	0.625
(HI_{AS})	(0.411)	(0.36)	(0.131)	(0.403)	(0.45)	(0.651)	(0.627)	(0.247)	(0.664)	(0.691)
$\mathrm{LM}_{\mathrm{NW}}$	0.464	0.197	0.057	0.457	0.382	0.39	0.366	0.258	0.65	0.56
$\mathrm{LM}_{\mathrm{ACA}}$	0.553	0.612	0.093	0.666	0.494	0.526	0.767	0.301	0.765	0.598
$(\mathrm{LM}_{\mathrm{AS}})$	(0.511)	(0.221)	(0.066)	(0.471)	(0.403)	(0.426)	(0.43)	(0.266)	(0.703)	(0.599)
$L_{\mathrm{c}}^{\mathrm{NW}}$	0.308	0.281	0.10	0.316	0.441	0.606	0.544	0.24	0.603	0.69
$L_{\mathrm{c}}^{\mathrm{ACA}}$	0.471	0.684	0.161	0.621	0.518	0.597	0.793	0.329	0.73	0.637
$(L_{\mathrm{c}}^{\mathrm{AS}})$	(0.387)	(0.312)	(0.107)	(0.356)	(0.438)	(0.653)	(0.639)	(0.258)	(0.673)	(0.708)
$R_{\mathrm{T}}^{\mathrm{NW}}$	0.312	0.273	0.125	0.318	0.418	0.564	0.506	0.218	0.551	0.632
$R_{\mathrm{T}}^{\mathrm{ACA}}$	0.446	0.68	0.164	0.614	0.487	0.572	0.781	0.295	0.726	0.587
$(R_{\mathrm{T}}^{\mathrm{AS}})$	(0.349)	(0.318)	(0.131)	(0.352)	(0.409)	(0.599)	(0.582)	(0.237)	(0.618)	(0.63)
J_1^{NW}	0.258	0.308	0.085	0.283	0.36	0.727	0.719	0.136	0.753	0.759
J_1^{ACA}	0.586	0.70	0.208	0.671	0.675	0.638	0.806	0.438	0.757	0.678
(J_1^{AS})	(0.68)	(0.312)	(0.075)	(0.503)	(0.772)	(0.901)	(0.678)	(0.131)	(0.796)	(0.95)

Table 2. Rejection frequencies at the 5% significance level, $\sigma_\eta^2 = 0.1$.

	$T=100$					$T=200$				
	ρ			θ	γ	ρ			θ	γ
$\sigma_\eta^2=0.1$	0	0.5	0.9	0.5	0.7	0	0.5	0.9	0.5	0.7
MLS	0.40	0.153	0.055	0.19	0.391	0.735	0.407	0.064	0.517	0.738
QS										
S_{NW}	0.314	0.164	0.063	0.213	0.307	0.576	0.346	0.07	0.448	0.577
S_{ACA}	0.369	0.157	0.057	0.21	0.371	0.682	0.337	0.076	0.453	0.678
(S_{AS})			(0.064)					(0.07)		
HI_{NW}	0.309	0.165	0.063	0.22	0.344	0.582	0.341	0.07	0.459	0.637
HI_{ACA}	0.392	0.153	0.055	0.216	0.405	0.686	0.33	0.078	0.452	0.713
(HI_{AS})			(0.059)					(0.068)		
LM_{NW}	0.33	0.177	0.064	0.238	0.245	0.582	0.368	0.073	0.46	0.491
LM_{ACA}	0.403	0.171	0.058	0.24	0.306	0.696	0.355	0.078	0.463	0.605
(LM_{AS})			(0.064)					(0.071)		
$L_c{}^{NW}$	0.254	0.124	0.058	0.179	0.322	0.559	0.314	0.078	0.431	0.625
$L_c{}^{ACA}$	0.379	0.115	0.065	0.186	0.389	0.684	0.295	0.078	0.424	0.719
$(L_c{}^{AS})$			(0.055)					(0.067)		
$R_T{}^{NW}$	0.284	0.143	0.065	0.197	0.35	0.575	0.329	0.067	0.447	0.621
$R_T{}^{ACA}$	0.39	0.138	0.064	0.198	0.42	0.706	0.306	0.078	0.442	0.711
$(R_T{}^{AS})$			(0.061)					(0.061)		
$J_1{}^{NW}$	0.114	0.092	0.054	0.094	0.196	0.532	0.312	0.058	0.407	0.566
$J_1{}^{ACA}$	0.154	0.083	0.052	0.091	0.28	0.186	0.106	0.054	0.144	0.242
$(J_1{}^{AS})$			(0.054)					(0.055)		
Parzen										
S_{NW}	0.301	0.16	0.064	0.209	0.298	0.548	0.339	0.07	0.436	0.547
S_{ACA}	0.364	0.151	0.062	0.203	0.363	0.676	0.322	0.075	0.441	0.667
(S_{AS})		(0.155)	(0.063)	(0.204)			(0.341)	(0.07)	(0.445)	

(*continued*)

Table 2. Continued.

	$T=100$					$T=200$				
	ρ			θ	γ	ρ			θ	γ
$\sigma_\eta^2=0.1$	0	0.5	0.9	0.5	0.7	0	0.5	0.9	0.5	0.7
HI_{NW}	0.042	0.056	0.152	0.043	0.028	0.045	0.063	0.184	0.048	0.029
HI_{ACA}	0.044	0.049	0.836	0.04	0.038	0.05	0.053	0.906	0.045	0.041
(HI_{AS})			(0.152)					(0.204)		
$\mathrm{LM}_{\mathrm{NW}}$	0.062	0.079	0.214	0.064	0.127	0.057	0.071	0.218	0.058	0.109
$\mathrm{LM}_{\mathrm{ACA}}$	0.051	0.071	0.853	0.055	0.103	0.053	0.066	0.913	0.053	0.107
$(\mathrm{LM}_{\mathrm{AS}})$			(0.215)					(0.236)		
$L_{\mathrm{c}}{}^{\mathrm{NW}}$	0.103	0.118	0.249	0.103	0.06	0.095	0.115	0.29	0.095	0.064
$L_{\mathrm{c}}{}^{\mathrm{ACA}}$	0.086	0.10	0.919	0.085	0.072	0.098	0.10	0.952	0.087	0.078
$(L_{\mathrm{c}}{}^{\mathrm{AS}})$			(0.252)					(0.328)		
$R_{\mathrm{T}}{}^{\mathrm{NW}}$	0.033	0.039	0.077	0.032	0.014	0.036	0.05	0.14	0.036	0.022
$R_{\mathrm{T}}{}^{\mathrm{ACA}}$	0.033	0.026	0.833	0.025	0.026	0.041	0.039	0.916	0.033	0.036
$(R_{\mathrm{T}}{}^{\mathrm{AS}})$			(0.082)					(0.165)		
$J_1{}^{\mathrm{NW}}$	0.171	0.154	0.442	0.155	0.09	0.106	0.114	0.393	0.102	0.075
$J_1{}^{\mathrm{ACA}}$	0.019	0.139	0.91	0.092	0.003	0.014	0.118	0.924	0.074	0.001
$(J_1{}^{\mathrm{AS}})$			(0.487)					(0.442)		

Table 3. Rejection frequencies at the 5% significance level, $\sigma_\eta^2 = 0.5$.

	$T=100$					$T=200$				
	ρ			θ	γ	ρ			θ	γ
$\sigma_\eta^2=0.5$	0	0.5	0.9	0.5	0.7	0	0.5	0.9	0.5	0.7
MLS	0.82	0.566	0.135	0.548	0.826	0.965	0.845	0.223	0.901	0.957
QS										
S_{NW}	0.491	0.411	0.129	0.467	0.502	0.69	0.619	0.242	0.667	0.686
S_{ACA}	0.426	0.294	0.15	0.332	0.437	0.484	0.40	0.386	0.436	0.496
(S_{AS})	(0.428)	(0.278)	(0.116)	(0.336)	(0.44)	(0.529)	(0.438)	(0.228)	(0.496)	(0.542)
HI_{NW}	0.445	0.373	0.129	0.428	0.519	0.689	0.599	0.243	0.671	0.745
HI_{ACA}	0.339	0.115	0.096	0.278	0.374	0.389	0.357	0.399	0.365	0.432
(HI_{AS})	(0.333)	(0.21)	(0.107)	(0.257)	(0.376)	(0.459)	(0.396)	(0.213)	(0.445)	(0.507)
LM_{NW}	0.499	0.436	0.156	0.487	0.408	0.714	0.658	0.265	0.70	0.621
LM_{ACA}	0.469	0.341	0.18	0.384	0.34	0.509	0.442	0.401	0.462	0.395
(LM_{AS})	(0.471)	(0.311)	(0.147)	(0.381)	(0.345)	(0.558)	(0.48)	(0.247)	(0.528)	(0.448)
$L_c{}^{NW}$	0.38	0.32	0.108	0.376	0.502	0.671	0.607	0.256	0.668	0.755
$L_c{}^{ACA}$	0.299	0.235	0.176	0.245	0.335	0.343	0.331	0.393	0.326	0.412
$(L_c{}^{AS})$	(0.275)	(0.129)	(0.077)	(0.189)	(0.321)	(0.426)	(0.364)	(0.201)	(0.403)	(0.505)
$R_T{}^{NW}$	0.365	0.313	0.131	0.367	0.469	0.624	0.559	0.235	0.62	0.688
$R_T{}^{ACA}$	0.259	0.229	0.177	0.226	0.299	0.306	0.305	0.374	0.307	0.335
$(R_T{}^{AS})$	(0.234)	(0.126)	(0.093)	(0.174)	(0.287)	(0.382)	(0.326)	(0.196)	(0.371)	(0.423)
$J_1{}^{NW}$	0.263	0.352	0.084	0.313	0.429	0.779	0.761	0.154	0.785	0.811
$J_1{}^{ACA}$	0.353	0.35	0.04	0.34	0.445	0.355	0.414	0.029	0.393	0.417
$(J_1{}^{AS})$	(0.387)	(0.339)	(0.069)	(0.366)	(0.481)	(0.592)	(0.611)	(0.10)	(0.649)	(0.641)
Parzen										
S_{NW}	0.446	0.387	0.125	0.427	0.463	0.627	0.573	0.237	0.616	0.638
S_{ACA}	0.494	0.622	0.14	0.586	0.516	0.627	0.767	0.275	0.724	0.615
(S_{AS})	(0.467)	(0.375)	(0.125)	(0.416)	(0.483)	(0.686)	(0.617)	(0.244)	(0.667)	(0.675)

(continued)

Table 3. Continued.

	$T=100$					$T=200$				
	ρ			θ	γ	ρ			θ	γ
$\sigma_\eta^2=0.5$	0	0.5	0.9	0.5	0.7	0	0.5	0.9	0.5	0.7
HI_{NW}	0.30	0.16	0.061	0.212	0.332	0.552	0.332	0.069	0.441	0.612
HI_{ACA}	0.383	0.146	0.062	0.21	0.396	0.673	0.311	0.078	0.427	0.70
(HI_{AS})		(0.152)		(0.211)			(0.335)	(0.07)	(0.433)	
$\mathrm{LM}_{\mathrm{NW}}$	0.32	0.17	0.06	0.234	0.238	0.324	0.186	0.073	0.446	0.468
$\mathrm{LM}_{\mathrm{ACA}}$	0.398	0.165	0.059	0.234	0.297	0.452	0.176	0.078	0.449	0.586
$(\mathrm{LM}_{\mathrm{AS}})$		(0.166)	(0.057)	(0.235)			(0.199)	(0.074)	(0.453)	
$L_{\mathrm{c}}{}^{\mathrm{NW}}$	0.23	0.121	0.058	0.165	0.308	0.524	0.292	0.076	0.41	0.59
$L_{\mathrm{c}}{}^{\mathrm{ACA}}$	0.368	0.11	0.066	0.172	0.385	0.665	0.273	0.081	0.403	0.705
$(L_{\mathrm{c}}{}^{\mathrm{AS}})$		(0.115)	(0.058)	(0.175)			(0.309)	(0.076)	(0.412)	
$R_{\mathrm{T}}{}^{\mathrm{NW}}$	0.268	0.136	0.064	0.184	0.336	0.549	0.313	0.064	0.422	0.595
$R_{\mathrm{T}}{}^{\mathrm{ACA}}$	0.379	0.132	0.064	0.19	0.419	0.691	0.288	0.076	0.42	0.70
$(R_{\mathrm{T}}{}^{\mathrm{AS}})$		(0.138)		(0.192)			(0.322)	(0.067)	(0.426)	
$J_1{}^{\mathrm{NW}}$	0.113	0.089	0.056	0.094	0.17	0.496	0.304	0.056	0.409	0.525
$J_1{}^{\mathrm{ACA}}$	0.224	0.092	0.062	0.109	0.40	0.274	0.133	0.083	0.194	0.367
$(J_1{}^{\mathrm{AS}})$		(0.097)	(0.053)	(0.119)		(0.277)	(0.248)	(0.054)	(0.293)	

The bandwidth selection method is indicated in either subscript or superscript as NW for the Newey and West (1994) method, ACA (Andrews–Choi–Ahn) for the Choi and Ahn (1995) modification of Andrews (1991) and AS (Andrews–Stock) for the suggestion of Stock (1994), as discussed in Sec. 2.2. The latter is presented in parentheses only when it produced different results from the ACA procedure. Table 4 displays the results for the alternative hypothesis of a simple random walk.

By observing Table 1 ($\sigma_\eta{}^2 = 0$), we may analyze the size performance of the tests. As regards the first test (MLS), it is relatively well behaved, except for the case of very persistent errors. Note that the correction for endogeneity works quite well, although the test tends to overreject in the MA case ($\theta = 0.5$). This may have to do with the fact that one is essentially using AR components to filter the residuals and the p-lags approximation may not be long enough.

Concerning the remaining tests, we observe that in general the choice of kernel does not seem to play an important role in terms of finite sample size, considering the not very substantial differences between the results produced by each of them. However, the Parzen kernel, when used in conjunction with the ACA approach and with strongly correlated errors ($\rho = 0.9$), displays considerable size distortions that do not decrease with the sample size, which is not the case for the other kernels. In contrast, and for the other DGPs, this kernel appears to work slightly better in terms of delivering sizes closer to the nominal significance level.

On the other hand, the method for the determination of the bandwidth may induce distinct performances. In fact, the empirical sizes produced by the NW method tend to be slightly higher than those obtained with Andrews's procedure when ρ is less than 0.9. Note, however, that the Choi–Ahn modification is generally the most affected in the case of substantially correlated errors, while the AS approach presents less size distortions. An

Table 4. Random walk alternative.

	$\rho = 1,\ \sigma_\eta{}^2 = 0$					
Tests	$T = 100$			$T = 200$		
MLS		0.782			0.888	
QS	NW	ACA	AS	NW	ACA	AS
S	0.493	0.731	0.304	0.68	0.913	0.477
HI	0.447	0.794	0.214	0.679	0.933	0.421
LM	0.518	0.821	0.376	0.738	0.939	0.526
L_c	0.384	0.806	0.16	0.664	0.935	0.373
R_T	0.356	0.785	0.114	0.618	0.932	0.349
J_1	0.293	0.88	0.728	0.786	0.956	0.843
Parzen						
S	0.442	0.939	0.447	0.625	0.995	0.678
HI	0.403	0.974	0.406	0.62	0.997	0.658
LM	0.479	0.976	0.516	0.676	0.996	0.725
L_c	0.313	0.977	0.386	0.595	0.998	0.653
R_T	0.293	0.976	0.327	0.559	0.996	0.604
J_1	0.285	0.988	0.832	0.73	0.996	0.924

interesting result is that the size properties do not seem to automatically improve for larger sample sizes. In fact, in many situations, distortions are more severe with $T = 200$, especially when the NW procedure is used.

Considering the different types of tests, we observe that the KPSS-type generally have rejection frequencies closer to the nominal test level than other tests. Despite this, the Harris–Inder test slightly underrejects in some occasions, while the Shin test is more liberal than HI when $\rho = 0.9$. Regarding the LM statistic, its empirical rejection frequencies tend to deviate from the nominal size of 0.05 when computed with the NW method, and the disadvantages of using OLS residuals become evident when correlation between ζ_t and ω_t is introduced. On the other hand, Hansen's L_c test is always oversized, but not as badly as Park's J_1 test computed with the NW method, which presents large distortions even for $\rho = 0$ and more dramatically so for $\rho = 0.9$. In contrast, when using the Andrews procedure the J_1 statistic is far from rejecting as often as it should when $\rho = 0$. Equally, the R_T fluctuation test is generally too conservative, except for the case of $\rho = 0.9$.

In terms of power analysis, the rejection frequencies when the null is false at the 5% level of significance are shown in Tables 2 and 3 (for ${\sigma_\eta}^2 = 0.1$ and 0.5). To avoid misrepresenting power due to size distortions, size-adjusted powers are presented, based on size-corrected critical values obtained from the corresponding results with ${\sigma_\eta}^2 = 0$.[k] For alternatives farther away from the null, differences in the empirical rejection frequencies are clearer. Comparing the use of the different kernels in the nonparametric-based tests, the differences in the results are not substantial, at least up to ${\sigma_\eta}^2 = 0.1$. However, we observed that the Bartlett kernel clearly delivers higher power with ${\sigma_\eta}^2 = 0.5$ for almost all cases, while tests using the Parzen kernel perform worse. This is hardly explainable solely by the Type I error probability results of Table 1 and may be related with the intrinsic weighting scheme of each kernel. This will also be noted in further experiments later on.

Turning to the bandwidth estimation methods, the diversity of DGPs and respective results does not allow a clear pattern to emerge from the simulations. Nevertheless, it is possible to see that the NW method performs worse when ${\sigma_\eta}^2 = 0.1$ and $\rho = 0$, but is better for ${\sigma_\eta}^2 = 0.5$ with the QS and Bartlett kernels, except for $\rho = 0.9$. In this case, the Choi–Ahn modification is in general the best, more clearly so when the Parzen kernel is used.

Moreover, it is possible to see that, as expected, power declines with increasing ρ and that errors with the MA structure considered here do not imply a sizeable reduction in the power of all tests. The exception seems to be the situation where the ACA procedure is used and $\rho = 0.5$, since in this case the Parzen kernel performs clearly better, even when compared with the case of $\rho = 0$. By contrast, endogeneity of the regressor has no substantial implications, except for the case of the LM test. Curiously, power with $\gamma = 0.7$ is occasionally larger than in the simple case of no correlation in the errors.

Analyzing the results across the different types of tests, we find that as ${\sigma_\eta}^2$ increases the differences in performance become more apparent. Indeed, the MLS test tends to dominate all kernel-based tests, except for $\rho = 0.9$, in which case we concluded before that the test rejects too often. As for nonparametric-based tests, no clear ranking may be established, since their performance varies not only with the long run variance estimation method, but with the DGP as well. The observation of the simulation results reveals that the J_1 test,

[k]However, we must not forget that in empirical practice asymptotic critical values are used, rather than exact critical values obtained from simulations.

e.g., presents the greatest variation and depending on the method, its performance may rank first or last, well below the remaining tests. It is interesting to note that the LM test does very well and generally slightly better than the other tests, with the expected exception of the case of endogenous regressors. Therefore, an important conclusion that may be drawn is that for distinct versions of the same test, differences in performance are frequently larger than 20%, across kernels and bandwidth estimators.

In order to refine the previous analysis, we compute the power of the tests as a function of the ratio of the disturbances variances, $\lambda = \sigma_\eta{}^2/\sigma_\varepsilon{}^2$. As discussed by Kwiatkowski et al. (1992), if λ is close to zero, the process is stationary (which means cointegration) and it will tend to an $I(1)$ process (no cointegration) as λ approaches infinity. This allows us to examine the relative power of the tests in the continuum of the alternative hypothesis, rather than concentrating only on three possible points ($\sigma_\eta{}^2 = 0.01, 0.1, 0.5$), as in the previous experiments. We considered several values of λ ranging from 0.0001 to 10^6, and for simplicity we present nominal powers, setting ρ and θ equal to zero, with $T = 200$.

Figures 1–6 show the results for the tests computed with the QS and Parzen kernels and the different bandwidth estimation methods discussed in this paper. The results for the MLS test are shown in each figure so that the comparison is clearer.[1] In line with the results shown previously, the rejection frequencies are quite similar up to λ equal to 0.1–0.15, the distinct behavior becoming well defined for larger values of this parameter. It is also clear that power stabilizes after λ attains a given value. However, there are two striking features in this set of experiments. First, and unlike the case where the Newey and West (1994) estimator is used, the power of tests employing both modified methods of Andrews (1991) is non-monotonic (with the exception of the Shin test). In fact, when λ is larger than 0.2, power is reduced, quite substantially in some cases. This is to be expected when the AS approach is used, since for values of λ away from the null, the estimated bandwidth will correspond to the chosen upper bound, which is 12 in our study, and thus power will not increase. In the case of the ACA procedure, power rises again after 0.4–0.6.

Secondly, the MLS test does very well and is unequivocally better for λ greater than 0.1 than all other tests, with the exception of the tests computed with the Parzen–ACA procedure and only for λ larger than 0.9 in Fig. 6. Furthermore, the NW estimator seems to work better with the QS kernel, while the Parzen kernel suits better Andrews-type methods. These results, therefore, confirm our previous conclusions. Also note that the local power of the Shin test is in general lower than the other tests for near-cointegration alternatives. We should mention the fact that we were unable to closely replicate the results of McCabe et al. (1997), which report higher powers for the MLS and Shin test, although this could be explained by some differences in the computation of the test procedures and in the DGPs. Nevertheless, our conclusion is similar to theirs, in that the MLS test generally outperforms the Shin test.

Finally, we also consider the case of a pure random walk as the alternative hypothesis, when $\rho = 1$ and $\sigma_\eta{}^2 = 0$, which is the standard setting for the study of null of no cointegration tests. This will allow us to check the claim of Hobijn et al. (1998) regarding the inconsistency of the McCabe et al. (1997) test. Size-adjusted power is calculated from

[1]The graphs are truncated for larger values of λ since the results do not change significantly after unity.

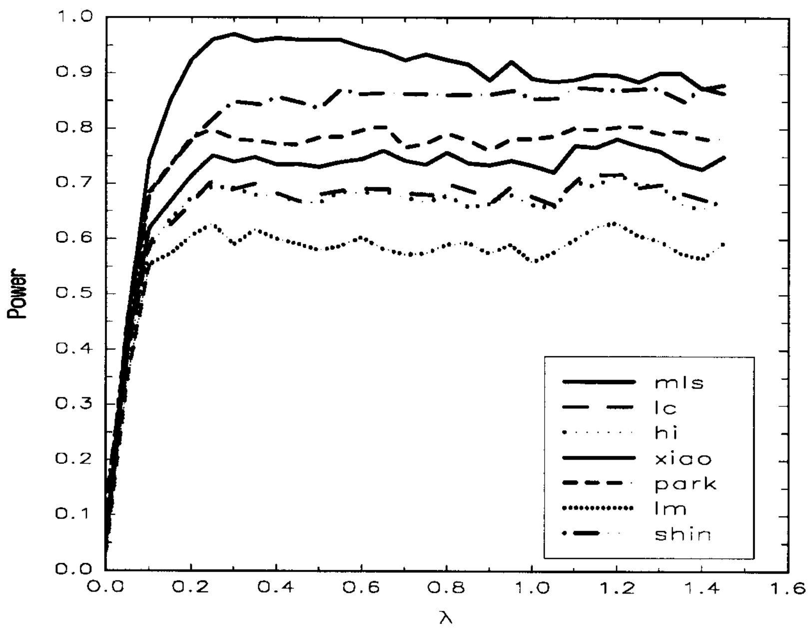

Figure 1. Power as a function of λ (NW, QS kernel, $T = 200$).

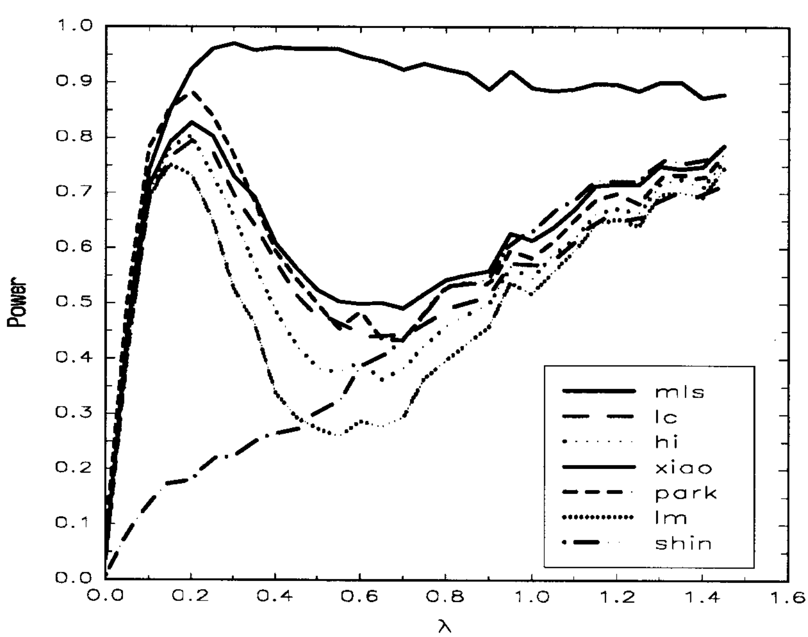

Figure 2. Power as a function of λ (ACA, QS kernel, $T = 200$).

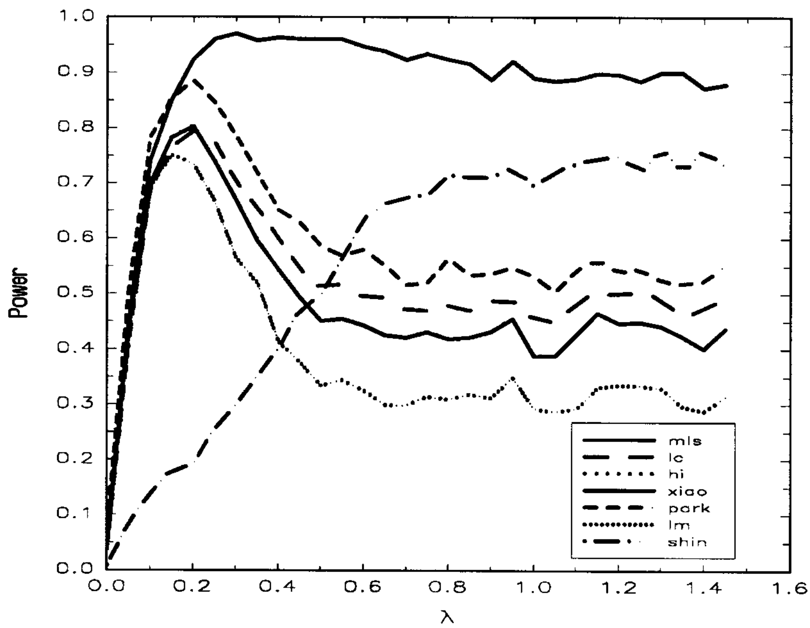

Figure 3. Power as a function of λ (AS, QS kernel, $T = 200$).

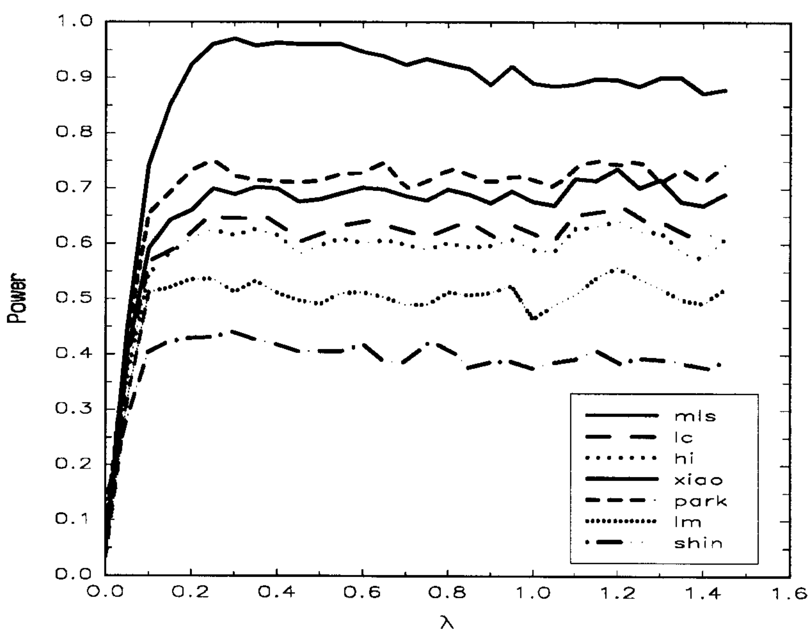

Figure 4. Power as a function of λ (NW, Parzen kernel, $T = 200$).

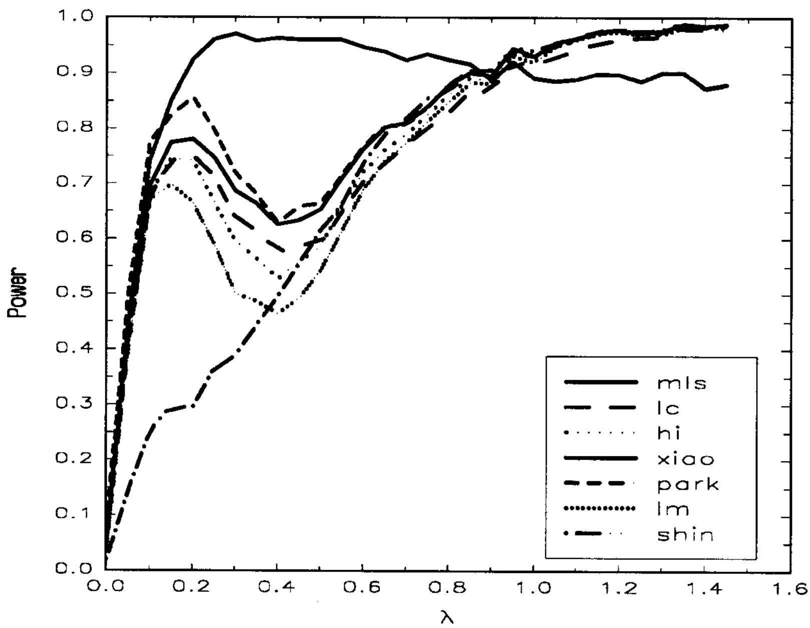

Figure 5. Power as a function of λ (ACA, Parzen kernel, $T = 200$).

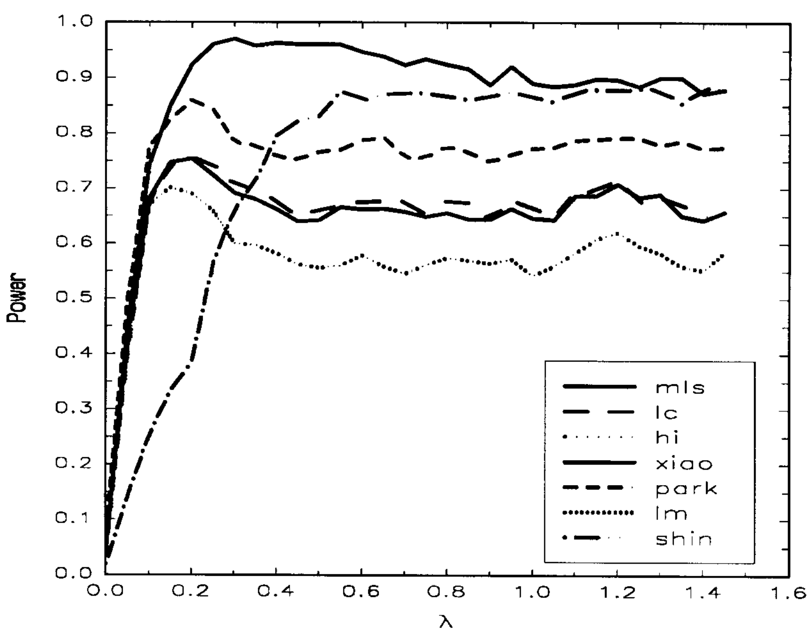

Figure 6. Power as a function of λ (AS, Parzen kernel, $T = 200$).

the results with $\rho = 0$ and $\sigma_\eta{}^2 = 0$. From Table 4 it is easily seen that the MLS does not appear to be inconsistent, with very reasonable power, growing with the sample size. Moreover, this test performs better than kernel-based tests when computed with the NW and AS procedures. However, if the ACA estimator is used, the power of these tests tends to be higher, especially if the Parzen kernel is used. It is also worth mentioning the performance of Park's J_1 test, which seems to be the best test in this set of simulations. Although this may suggest that this specific combination of methods would be useful in empirical practice, we must bear in mind the results from Table 3, where size distortions are very severe when $\rho = 0.9$. This means that the ability to distinguish no cointegration from cointegration with strongly correlated errors is virtually null.

5. CONCLUSION

Although less often used, tests for the null hypothesis of cointegration may be a useful instrument in the analysis of economic time series. Unlike tests for the null of no cointegration, there is no evidence on the properties of the different tests and their relative merits. This study tries to fill this gap by analyzing the performance of several tests that have been recently proposed. Conducting a series of Monte Carlo experiments, we found that no test dominates the others in every situation under analysis. However, some conclusions may be extracted from this study, despite the shortcomings pointed out earlier.

As could be seen from our experiments, the performance of nonparametric-based tests depends heavily on the method chosen for the estimation of the long run variance. Furthermore, and unlike previous studies, we found that the specific chosen kernel seems to matter in terms of the finite sample power performance of the tests, mainly for alternatives away from the null hypothesis. As pointed out earlier, this gives rise to substantial differences for distinct versions of the same test. Nevertheless, it is not possible to state which kernel-bandwidth estimator combination is preferable, given the irregular performances displayed in our simulation study. On the other hand, KPSS-type tests seem to have an overall better performance than the other approaches studied here.

In view of this, it seems that there are some advantages in the use of the MLS test, given its reasonable and well-balanced overall performance. Although our study is somewhat biased towards favoring this parametric version, as mentioned earlier, this test has, at least theoretically, some advantages, namely a faster rate of convergence. Moreover, its computation is free of the problems associated with all the kernel-based tests, i.e., the choice of method to obtain the scaling long run variance. Nevertheless, there is still room for improvements on the performance of this parametric version of the KPSS statistic, namely in terms of reducing size distortions, as the attempts of Lanne and Saikkonen (2000) in the univariate case indicate. Another issue that clearly merits future attention is the performance of tests for the null of cointegration in systems of equations.

ACKNOWLEDGMENTS

The author wishes to thank Luis Martins, Haris Psaradakis, Ron Smith, and three anonymous referees for their valuable comments. The usual disclaimer applies. Financial support from the Sub-Programa Ciência e Tecnologia do Segundo Quadro Comunitário de Apoio, grant number PRAXIS XXI/BD/16141/98 and from the Research Unit in Economic Policy – NIPE is gratefully acknowledged.

REFERENCES

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 59:817–858.

Andrews, D. W. K., Monahan, J. C. (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60:953–966.

Carrion-i-Silvestre, J. L., Sansó-i-Rosselló, A.; Ortuño, M. A. (2001). Unit root and stationarity tests' wedding. *Econ. Lett.* 70:1–8.

Charemza, W. W., Syczewska, E. M. (1998). Joint application of the Dickey–Fuller and KPSS tests. *Econ. Lett.*, 61:17–31.

Choi, I., Ahn, B. C. (1995). Testing for cointegration in a system of equations. *Economet. Theor.* 11:952–983.

Den Haan, W. J., Levin, A. (1997). A practitioner's guide to robust covariance matrix estimation. In: Maddala, G. S., Rao, C. R., eds. *Handbook of Statistics: Robust Inference.* Vol. 15. Amsterdam: Elsevier, pp. 291–341.

Gabriel, V. J. (2003). Cointegration and the joint confirmation hypothesis. *Econ. Lett.* 78:17–25.

Hansen, B. E. (1992). Tests for parameter instability in regressions with $I(1)$ processes. *J. Bus. Econ. Stat.* 10:321–335.

Hao, K. (1996). Testing for structural change in cointegrated regression models: some comparisons and generalizations. *Economet. Rev.* 15:401–429.

Leybourne, S. J., McCabe, B. P. M. (1999). Modified stationarity tests with data-dependent model-selection rules. *J. Bus. Econ. Stat.* 17:264–270.

Harris, D. (1997). Principal components analysis of cointegrated time series. *Economet. Theor.* 13:529–557.

Harris, D., Inder, B. (1994). A test of the null of cointegration. In: Hargreaves, C., ed. *Non-stationary Time Series Analysis and Cointegration.* Oxford: Oxford University Press, pp. 133–152.

Haug, A. A. (1996). Tests for cointegration: a Monte Carlo comparison. *J. Econometrics* 71:89–115.

Hobijn, B., Franses, P. H., Ooms, M. (1998). Generalizations of the KPSS-test for stationarity. Econometric Institute Report 9802/A.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics* 54:159–178.

Lanne, M., Saikkonen, P. (2000). Reducing size distortions of parametric stationarity tests. Manuscript.

Lee, J. (1996). On the power of stationarity tests using optimal bandwidth estimates. *Econ. Lett.* 51:131–137.

Leybourne, S. J., McCabe, B. P. M. (1993). A simple test for cointegration. *Oxford B. Econ. Stat.* 55:97–103.

Leybourne, S. J., McCabe, B. P. M. (1994). A consistent test for a unit root. *J. Bus. Econ. Stat.* 14:435–446.

Maddala, G. S., Kim, I. M. (1998). *Unit Roots, Cointegration and Structural Change.* Cambridge: Cambridge University Press.

McCabe, B. P. M., Leybourne, S. J., Shin, Y. (1997). A parametric approach to testing for the null of cointegration. *J. Time Ser. Anal.* 18(4):395–413.

Nabeya, S., Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Stat.* 16:218–235.

Newey, W. K., West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Rev. Econ. Stud.* 61:631–653.

Nyblom, J., Harvey, A. (2000). Tests of common stochastic trends. *Economet. Theor.* 16:176–199.

Park, J. Y. (1990). Testing for unit roots and cointegration by variable addition. In: Fomby, T., Rhodes, G. F., eds. *Advances in Econometrics.* Vol. 8. JAI Press, pp. 107–133.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica* 60:119–143.

Phillips, P. C. B., Hansen, B. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Rev. Econ. Stud.* 57:99–125.

Ploberger, W., Krämer, W., Kontrus, K. (1989). A new test for structural stability in the linear regression model. *J. Econometrics* 40:307–318.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegrating regressions. *Economet. Theor.* 7:1–21.

Saikkonen, P., Lukkonen, R. (1993). Testing for a moving average unit root in ARIMA models. *J. Am. Stat. Assoc.* 88:596–601.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Economet. Theor.* 10:91–115.

Snell, A. (1999). TESTING For r versus $r-1$ cointegrating vectors. *J. Econometrics* 88:151–191.

Stock, J. H. (1994). Unit roots, structural breaks and trends. In: Engle, R. F., McFadden, D. L., eds. *Handbook of Econometrics.* Vol. 4. Amsterdam: Elsevier, pp. 2740–2841.

Stock, J. H., Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61:783–820.

Tanaka, K. (1990). Testing for a moving average unit root. *Economet. Theor.* 6:433–444.

Xiao, Z. (1999). A residual based test for the null hypothesis of cointegration. *Econ. Lett.* 64:133–141.