ON THE RANKS OF TESTS HAVING NULL OF COINTEGRATION: A MONTE CARLO COMPARISON

Nicoleta ISAC
Istanbul Sabahattin Zaim University, Halkalı Cad. No: 281 Halkalı Küçükçekmece, Istanbul, Turkey.
nicoleta.isac@izu.edu.tr

Cosmin DOBRIN
Bucharest University of Economic Studies, Piata Romana 6, Bucharest, Romania
cdobrin@yahoo.com

Mehmood HUSSAN
International Islamic University Islamabad, New campus, Sector H-10, Islamabad, Pakistan.
mtilokar@yahoo.com

Asad ul Islam KHAN[1]
Ibn Haldun University, Kayabaşı Mahallesi Ulubatlı Hasan Caddesi No:2 Başakşehir, Istanbul, Turkey.
asad.khan@ihu.edu.tr, asadkhan1983@gmail.com

Alina- Andreea MARIN
Bucharest University of Economic Studies, Piata Romana 6, Bucharest, Romania
alinagsm2000!gmail.com

Abstract
The null of cointegration tests for testing the existence of cointegration are available in literature in great diversity. The selection of a particular test from all these available tests is the crucial problem and often researchers face this. This study is carried out to solve this problem by comparing eight tests on basis two properties of size and power using a new proposed methodology of rank scores. Three different specifications of deterministic component and four sample sizes are considered. It is concluded that if asymptotic critical values are used then it results into an uncontrolled empirical size. While, if the simulated critical values are used, then the empirical size is controlled around nominal size. Moreover, on basis of power, a simple test, which is based on KPSS statistic, is the sole better performer for all of the different 132 different cases of data generations considered in the study.
Keywords: Cointegration Tests; Comparison; Power; Rank Score; Size

1. INTRODUCTION

The concept of cointegration was firstly introduced by (Engle and Granger 1987). Presence of a long run relationship between two or more time series, having order of integration as one is termed as existence of cointegration. More technically, for two time series say X and Y, (both of them are integrated of order one) are termed as cointegrated time series if their linear combination is stationary i.e. integrated of order zero. After the introduction of the concept of cointegration, a variety of tests for testing the cointegration have been introduced ((P C B Phillips and Ouliaris 1990), (Johansen and Juselius 1990), (Pesaran, Shin, and Smith 2001), (B. Hansen 1990), (Choi 1994), (Sargan and Bhargava 1983), (Banerjee et al. 1986), (H. P. Boswijk 1989), (Akdi 1996) and many more). In the start, cointegration tests were developed to test the null hypothesis of no cointegration. The first test with null of cointegration was developed by (Leybourne and McCabe 1994). After this paper, again a variety of null of cointegration tests have been introduced: ((Shin 1994), (Fernández-Macho and Mariel 1994), (Leybourne and McCabe 1994), (McCabe, Leybourne, and Shin 1997), (B. E. Hansen 1992)

[1] Corresponding author

and many others). This variety of tests have been developed by assuming different mechanisms of cointegration.

For a same empirical problem or data set, two or more different tests of cointegration have contradicting results to each other ((Haug 1996) and (Mariel 1996)). Therefore, the practitioners and researchers were facing the problem of test selection. To answer the issue of selection of test, these cointegration tests have been compared to assess their performance in a number of studies in literature. Monte Carlo Simulations (MCS) have been frequently used for the said purpose of comparison as it has been widely used in the literature like (ul Islam Khan, Khan, and Hussan 2019), (W. M. Khan and Khan 2018), (Mantalos 2003), (Iglesias and Phillips 2017) and many others. However, there are a very few studies in the literature using real data or analytical comparisons like (Pesavento 2004) and (Gonzalo and Lee 1998). The tests have been compared on the basis of two properties: Size and Power in nearly all of the studies using MCS. The Size and Power of Test are defined as

$$Size\ of\ a\ Test = Probability\ (Rejecting\ H_0\ \ / H_0\ is\ True)$$

$$Power\ of\ a\ Test = Probability\ (Rejecting\ H_0\ \ / H_0\ is\ False)$$

Where H_0 represents the null hypothesis.

If a test has higher power as compared to another provided that both are having the same size, then that test is considered better. Also, the tests in comparison may not have size distortion i.e. having size higher or lesser than the nominal size assumed. These two properties of size and power have been used in a number of studies ((Banerjee et al. 1986), (Kremers, Ericsson, and Dolado 1992), (P. Boswijk and Franses 1992), (Haug 1996) and many more). (A. ul I. Khan and Zaman 2017) has provided a detailed survey of the literature of these comparisons. In most of these comparisons a fewer number of point alternative hypotheses have been considered, although the alternative space has infinite many point alternative hypotheses. From all of these comparative studies, two are broader than others and these are of (Haug 1996) and (Mariel 1996). These studies have used different data generation systems of cointegration, compared nine tests of two opposite null hypotheses each and covered almost all of the alternate space. However, (Haug 1996) and (Mariel 1996) concluded that there is not a single winner among all of these tests.

The studies assessing the performance of cointegration tests were comparing tests with null of no cointegration most of times. However, there are fewer studies who have either compared only the tests with null of cointegration or a mixture of tests covering both null hypotheses. The study of (Gabriel 2003) is the only one “to the best of our knowledge”, assessing the performance of six test of null of cointegration. A variety of data generating processes and a fewer number of point alternatives have been considered by (Gabriel 2003). Similar to (Haug 1996) and (Mariel 1996), (Gabriel 2003) also concluded that there is not a single winner.

As stated earlier, two or more tests are comparable on basis of power if they have the same size and especially when their size is stable around the nominal size assumed. However, asymptotic critical values have been used to draw conclusion about the rejection of null, in almost all of the comparative studies available in the literature. These asymptotic critical values lead to size distortion most of the times. Furthermore, again as stated earlier a fewer number of point alternative hypotheses have been considered in the literature which don’t lead to power curve. There are some studies in the literature that have focused only on one or two tests and tried to modify it by bootstrapping or fast bootstrapping or any other method to overcome the deficiencies of cointegration tests and then compared their results either using MCS or empirically. These studies include but are not restricted to (Kascha and Trenkler 2011), (Park, Ahn, and Cho 2011), (Ahlgren and Antell 2008) and (Jacobson and Larsson 1999).

Keeping in view, three gapes in the literature are the goals of this paper. One is to check how the tests perform in terms of power if their size is maintained around a nominal assumed level. Second is to investigate the size distortions of tests when asymptotic critical values are used and third is to compare the tests on a wider number of point alternative hypotheses enabling us to draw power curves by introducing a new criterion or methodology for comparison “the average rank score”. This new proposed criterion of Rank Scores and Average Rank Scores considers the whole alternative space and draws an overall conclusion about the performance of tests.

This study will recommend test or tests for empirical applications as well as it will provide the details of poor performing tests. In the light of study, the statisticians, applied econometricians, data scientists and researchers will be able to choose a better performing test both in terms of size and power. This study will also guide the researchers and practitioners in choosing between simulated critical values and asymptotic critical values.

The study is structured as: this section of "Introduction" is followed by the section "Tests of Null of Cointegration", specifying the details of tests compared in the study. Then the next section "Methodological Framework" gives the details of data generation and the new criterion of Rank Scores and Average Rank Scores. This section is followed by the section "Discussion of Results" and then "Conclusions and Recommendations". At the end "References" are listed.

2. TEST OF NULL OF COINTEGRATION

The performance of eight cointegration tests have been assessed in the study. Their details are:

For two I(1) time series X and Y of length T, with X having m components, (Leybourne and McCabe 1994) recommended the use of LM type test, denoted by "Lm", developed by (Kwiatkowski et al. 1992) for testing null of cointegration, i.e.

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

where

$$S_t = \sum_{t=1}^{i} \hat{\mu}_t \quad \text{and}$$

$$s^2(l) = \frac{\hat{\mu}_t{}'\hat{\mu}_t}{T}$$

$\hat{\mu}_t$ denotes the residuals of the Ordinary least Squares regression of Y on X i.e.

$$Y_t = \theta\psi_t + \beta X_t + \mu_t \qquad (1)$$

ψ_t contains both the intercept and linear time trend and θ is its respective coefficient matrix. Similarly, β is the respective coefficient vector of X.

In the same paper, (Leybourne and McCabe 1994) proposed the use of a modified LM stat, denoted by "Lb" i.e.

$$Lb = T^{-2}\frac{\sum_{t=1}^{T} S_t^2}{s^2(l)} \quad \text{where} \quad S_t = \sum_{t=1}^{i} \hat{\mu}_t$$

and now $s^2(l) = T^{-1}\sum_{t=1}^{T}\hat{\mu}_t^2 + 2T^{-1}\sum_{s=1}^{l}\sum_{t=s+1}^{T}\hat{\mu}_t\hat{\mu}_{t-s}$ is the long run variance.

According to (Mariel 1996), the lag truncation parameter l plays a vital role in the performance of test and $l = 4$ has been recommended by him.

Later on (Shin 1994) recommended the following OLS regression instead of (1).

$$Y_t = \theta\psi_t + \beta X_t + \sum_{j=-k}^{k}\pi_j \Delta X_{t-i} + \mu_t \qquad (2)$$

k denotes the maximum number of lags or leads. Again, the same test statistic is considered i.e.

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

Where

$$S_t = \sum_{t=1}^{i} \hat{\mu}_t \quad \text{and}$$

$$s^2(l) = T^{-1}\sum_{t=1}^{T}\hat{\mu}_t^2 + 2T^{-1}\sum_{s=1}^{l}\left(1 - s(l+1)^{-1}\right)\sum_{t=s+1}^{T}\hat{\mu}_t\hat{\mu}_{t-s}$$

According to the recommendations (Mariel 1996) and (Shin 1994), $l = 4$ and $k = 5$ have been used in the current study.

Continuing on to the same idea and philosophy, (McCabe, Leybourne, and Shin 1997) recommended the Maximum Likelihood Estimation (MLE) in place of OLS. According to (McCabe, Leybourne, and Shin 1997), estimate $\hat{\mu}_t$ from (2) using OLS and then use MLE to estimate $\hat{\eta}_t$ from

$$\hat{\mu}_t = \sum_{i=1}^{p}\lambda_i \Delta\hat{\mu}_{t-i} + \eta_t$$

The selection of p is based on the minimum Akaike information Criterion (AIC). The modified version of test is

$$Ls = \frac{\hat{\eta}_t{}'\Omega\hat{\eta}_t}{T^2 S^2(l)} \quad \text{where } S^2(l) = \frac{\hat{\eta}_t{}'\hat{\eta}_t}{T}$$

and $\Omega = \Lambda\Lambda'$ with Λ being lower triangular matrix of ones.

(Fernández-Macho and Mariel 1994) introduced two tests (H1 and H2), comparing two estimators , both of them consistent under null, whereas, under alternative, one is inconsistent. One is the OLS estimate of γ say $\hat{\gamma}_L$ from (2). Furthermore, defining

$$\upsilon_t = y_t - \sum_{j=-k}^{k}\hat{\pi}_j \Delta x_{t-j} \text{ ---------------- } (3)$$

$\hat{\pi}_j$ represent the same OLS estimates obtained from (2). The other estimator $\hat{\gamma}_D$ is obtained from the regression using OLS

$$\Delta\upsilon_t = \gamma_D \Delta x_t + \varepsilon_t \text{ ------------- } (4)$$

The test stats H1 and H2 are:

$$H_1 = (\hat{\gamma}_L - \hat{\gamma}_D)'\left(\hat{V}_D + \hat{V}_L\right)^{-1}(\hat{\gamma}_L - \hat{\gamma}_D)$$

$$H_2 = (\hat{\gamma}_L - \hat{\gamma}_D)'\hat{V}_D^{-1}(\hat{\gamma}_L - \hat{\gamma}_D)$$

$\hat{V}_D$ and $\hat{V}_L$ represents the estimated variance covariance matrices of $\hat{\gamma}_D$ and $\hat{\gamma}_L$ respectively.

On the basis of Fully Modified OLS (FMOLS) estimation method developed by (Peter C B Phillips and Hansen 1990), (B. E. Hansen 1992) introduced a test stat, denoted by "Lc", given as

$$L_c = trace\left[\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} \hat{S}_t \hat{\Omega}_{\mu\bullet\upsilon}^{-1} \hat{S}_t'\right] \text{ where } \hat{S}_t = \sum_{i=1}^{t}\left(x_i \hat{\mu}_i^{+} - \begin{bmatrix} 0 \\ \hat{\Lambda}_{\upsilon\mu}^{+} \end{bmatrix}\right)$$

and $\hat{\mu}_t^{+}$ are FMOLS residuals of (1). The technical details of the test can be found in (B. E. Hansen 1992). Moreover, (P. Boswijk and Franses 1992) gave the critical values of Lc.

(Xiao 1999) introduces another test denoted as "RT", using the fluctuations of FMOLS residuals $\hat{\mu}_t^{+}$, i.e.

$$R_T = \max_{i=1,\dots\dots\dots,T} \frac{i}{\sqrt{\Omega_{\mu g\upsilon} T}} \left|\frac{1}{i}\sum_{t=1}^{i}\hat{\mu}_t^{+} - \frac{1}{T}\sum_{t=1}^{T}\hat{\mu}_t^{+}\right|$$

The abbreviations of eight tests with their source paper are provided in Table 1. These abbreviations have been used in the rest of study for convenience.

TABLE 1 - ABBREVIATIONS OF TESTS

S. No	Source Paper and Name of Test	Abbreviation
1	Simple Lm test (Leybourne and McCabe 1994)	"LMKPSS"
2	Modified Lm i.e. Lb (Leybourne and McCabe 1994)	"LBI"
3	Shin's C i.e. Sc (Shin 1994)	"SC"
4	Modified Lm based on MLE (McCabe, Leybourne, and Shin 1997)	"MLS"
5	Hausman H_1 (Fernández-Macho and Mariel 1994)	"HH1"
6	Hausman H_2 (Fernández-Macho and Mariel 1994)	"HH2"
7	Hansen's Lc (B. E. Hansen 1992)	"HLC"
8	Xiao R_T (Xiao 1999)	"XFT"

3. METHODOLOGICAL FRAMEWORK

The details of equations used to generate the cointegrated system and the new proposed criterion of Rank Scores and Average Rank Scores have been laid out in the two subsections of this section. MATLAB has been used for analysis.

3.1. Data Generation (ADG):

Following set of equations have been used to generate the cointegrated system for two time series y and x of length T. This system of equations is a modified version of (Jansson 2005) for time trend.

$$y_t = D_t\delta' + x_t + \upsilon_t$$

$$x_t = x_{t-1} + \mu_t^{x}$$

$$\upsilon_t = \upsilon_{t-1} - \varphi\mu_{t-1}^{y} + \mu_t^{y}$$

$$\mu_t = \left(\mu_t^{y}, \mu_t^{x}\right) \text{ and } \mu_t \sim N\left(0, \Sigma\right),$$

Σ being an identity matrix. Under null hypothesis of cointegration and alternative hypothesis of no cointegration

$H_0 : \varphi = 1$ (Cointegration)
$H_A : 0 \leq \varphi < 1$ (No Cointegration)

Following set of φ under alternative hypothesis has been considered

$$\varphi = (0, 0.1, 0.2, 0.3, 0.4, ------- 0.9)$$

D_t comprises of intercept and linear time trend i.e.

$$D_t = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \bullet & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \\ 1 & T \end{bmatrix}$$

and δ is the respective coefficient vector. Three cases of are under consideration in the study:

Neither Intercept nor Linear Time Trend is present (denoted as D0LT0): $\delta = [0 \quad 0]$

Only Intercept is present (denoted as D1LT0): $\delta = [1 \quad 0]$

Both Intercept and Linear Time Trend are present (denoted as D1LT1): $\delta = [1 \quad 1]$

The performance of tests has been evaluated using 10 point alternatives, four sample sizes ($n = 30, n = 60, n = 120$ and $n = 240$) and three different cases of δ. So, it all sums up to 132 different data generations.

3.2. Rank Scores:

If $\pi_i^\tau(n)$ is the power of a test τ at sample size of n and point alternative hypothesis i for *l* number of tests i.e. $\tau = 1, 2, ..., l$ and κ number of point alternative hypotheses i.e. $i = 1, 2, ..., \kappa$, then rank of a test τ at a specific point alternative hypothesis i and at sample size n is denoted by $\Re_i^\tau(n)$ and defined as

$$\Re_i^\tau(n) = \underset{i=1,2,........\kappa}{Rankings}\left(\pi_i^\tau(n)\right)_{(Descending\ Order)}$$

The Rank Score $\Im^\tau(n)$ of a test τ at sample size n is obtained by averaging these rankings $\Re_i^\tau(n)$ over all the point alternate hypotheses i.e.

$$\Im^\tau(n) = \frac{\sum_{i=1}^{\kappa} \Re_i^\tau(n)}{\kappa}$$

Management Research and Practice
Volume 12, Issue 2 / June 2020

The reciprocals of sample size are used as weights to obtain a weighted average of all these rank scores to get a final statistic Υ^{τ}

$$\Upsilon^{\tau}=\frac{\sum_{n}\frac{1}{n}\times\Im^{\tau}\left(n\right)}{\sum_{n}\frac{1}{n}}$$

On basis of rank scores $\Im^{\tau}\left(n\right)$ or weighted average of rank scores Υ^{τ}, tests are categorized into three; better, mediocre or average and worst performers. From eight tests; tests possessing rank scores lesser or equal than 2 are regarded as better performers, tests possessing rank scores greater than 2 but lesser or equal than 4 are ranked as average or mediocre and tests possessing rank scores greater than 4 are regarded as poor or worst performers.

4. RESULTS AND DISCUSSIONS

Size and Power are obtained using $M=30{,}000$ for each test except the MLS test, as it uses the numerical optimization algorithms to find the maximum likelihood estimates, which takes huge amount of time. So, for MLS: $M=5{,}000$ is considered suitable. These sizes are given in Table 2. It is obvious from Table 2 that not a single test has empirical size around the considered significance level or nominal size of 0.05 for all three specifications of deterministic part and four sample sizes. For instance, LMKPSS has size around 0.05 for all four sample sizes only when deterministic part is D1LT0. The values of empirical size around the specified nominal size of 0.05 have been marked as BOLD in Table 2. It is observed that from out of total 96 different cases, only for 15 cases the empirical size is around 0.05.

TABLE 2 - ASYMPTOTIC CRITICAL VALUES' SIZE

Tests	D^0LT^0				D^1LT^0				D^1LT^1			
	n=30	n=60	n=120	n=240	n=30	n=60	n=120	n=240	n=30	n=60	n=120	n=240
LMKPSS	0.2271	0.2256	0.2194	0.2232	**0.0589**	**0.0534**	**0.0575**	**0.0533**	0.0002	0.0007	0.0004	0.0002
HH1	0.0076	0.0080	0.0065	0.0049	0.0023	0.0046	0.0051	0.0061	0.0033	0.0114	0.0148	0.0170
HH2	0.0111	0.0075	0.0064	0.0056	0.0055	0.0060	0.0063	0.0064	0.0193	0.0161	0.0166	0.0177
SC	0.0048	0.0186	0.0238	0.0252	**0.0638**	**0.0537**	**0.0498**	**0.0528**	0.2034	**0.0573**	**0.0521**	**0.0501**
LMLBI	0.3407	0.2840	0.2464	0.2384	0.2134	0.1299	0.0797	**0.0594**	0.1068	**0.0375**	0.0055	0.0005
XFT	0.0030	0.0220	0.0180	0.0180	0.1870	0.1740	0.1800	0.1870	0.4130	0.4370	0.4430	0.4090
HLC	0.5324	0.4352	0.2984	0.1552	0.5580	0.4074	0.3342	0.1772	0.5898	0.4208	0.3414	0.2188
MLS	0.6580	0.6090	0.5860	0.5400	0.5890	0.4120	0.3350	0.2980	0.3640	0.1170	**0.0650**	**0.0630**

TABLE 3 - SIMULATED CRITICAL VALUES' SIZE

Tests	D^0LT^0				D^1LT^0				D^1LT^1			
	n=30	n=60	n=120	n=240	n=30	n=60	n=120	n=240	n=30	n=60	n=120	n=240
LMKPSS	0.0476	0.0499	0.0510	0.0478	0.0514	0.0538	0.0482	0.0479	0.0454	0.0467	0.0468	0.0487
HH1	0.0523	0.0502	0.0491	0.0443	0.0506	0.0481	0.0502	0.0501	0.0480	0.0507	0.0511	0.0505
HH2	0.0509	0.0558	0.0484	0.0485	0.0518	0.0513	0.0442	0.0490	0.0561	0.0512	0.0472	0.0480
SC	0.0509	0.0529	0.0519	0.0473	0.0485	0.0548	0.0521	0.0501	0.0542	0.0498	0.0485	0.0481
LMLBI	0.0489	0.0505	0.0498	0.0511	0.0490	0.0513	0.0497	0.0513	0.0416	0.0517	0.0452	0.0473
XFT	0.0467	0.0441	0.0525	0.0472	0.0390	0.0517	0.0542	0.0482	0.0536	0.0542	0.0524	0.0459
HLC	0.0442	0.0478	0.0492	0.0500	0.0484	0.0492	0.0496	0.0490	0.0478	0.0500	0.0534	0.0540
MLS	0.0370	0.0430	0.0450	0.0640	0.0470	0.0460	0.0630	0.0380	0.0380	0.0530	0.0570	0.0570

To control empirical size around 0.05, simulated critical values for all tests at all four sample sizes using all three specifications of deterministic part have been found and then by using these simulated critical values, again empirical size is calculated and these sizes are displayed in

Table 3. It is evident from

Table 3 that all tests have empirical size around 0.05 now. It is very important to control empirical size around 0.05 in a statistical hypothesis testing, because a large size will enlarge the power. So, as compared to the considered nominal significance level, a decrease in empirical size leads to a decrease in power and an increase in empirical size leads to an increase in power resulting in a meaningless conclusion.

As the use of simulated critical values leads to controlled empirical size around 0.05, the nominal size considered in this paper, so, these have been used for the estimation of power.

The results of Monte Carlo experiment in finalized form of rank scores and weighted average of rank scores are displayed in Table 4, when presence of neither intercept nor linear time trend (D0LT0) is assumed in data generation and test implementation. An in-depth evaluation of Table 4 reveals that LMKPSS has a rank score of 1 at all sample sizes and LMLBI is worst performer at all sample sizes except $n = 60$, where it becomes just an average performer. Similarly, SC and HH2 are average performers at all sample sizes. However, HH1 is a mediocre performer at small sample sizes of $n = 30 \& 60$. XFT is a worst performer at all four sample sizes. While, HLC is worst performer at the small sample sizes of $n = 30 \& 60$, however, it improves and becomes an average one at larger sample sizes of $n = 120 \& 240$. The last test i.e. MLS is poor performer for all sample sizes.

TABLE 4 - RANK SCORES FOR D0LT0

Tests	$\Im^{\tau}(30)$	$\Im^{\tau}(60)$	$\Im^{\tau}(120)$	$\Im^{\tau}(240)$	Υ^{τ}
LMKPSS	1**	1**	1**	1**	1.00**
SC	3.1*	3.1*	2.6*	2**	2.96*
HH2	2.7*	3.2*	3.3*	4.1*	3.01*
HH1	3.9*	4.1*	4.7	5.3	4.15*
LMLBI	5.9	4.4*	5.2	4.9	5.34
XFT	4.6	8	7.3	7.3	6.05
HLC	7.8	5.4	4.3*	3.7*	6.42
MLS	7	6.8	7.6	7.6	7.07

Note: ** shows that test is a better performer and * shows that test is an average performer

The last column of Table 4 provides an overall summary about the performance of eight tests, as it is the weighted average of rank scores. According to this column only one test has a better performance, i.e. LMKPSS. Three tests have an overall average performance and all these three tests are based on DOLS estimation i.e. HH2, HH2 and HH1. The remaining four tests are overall worst performers.

TABLE 5 - RANK SCORES FOR D1LT0

Tests	$\Im^{\tau}(30)$	$\Im^{\tau}(60)$	$\Im^{\tau}(120)$	$\Im^{\tau}(240)$	Υ^{τ}
LMKPSS	1**	1**	1**	1**	1.00**
SC	3.3*	2.6*	2**	2**	2.85*
HH2	3.6*	3.4*	4.3*	4.7	3.71*
MLS	5	4*	5	5.9	4.79
XFT	3.4*	6.3	7.3	7.8	4.99
HH1	4.9	4.9	5.3	5.6	5.00
LMLBI	7.8	5.9	4*	3.1**	6.47
HLC	7	7.9	7.1	5.9	7.18

Note: ** shows that test is a better performer and * shows that test is an average performer

The results of Monte Carlo experiment in final form of rank scores and weighted average of rank scores are displayed in Table 5, when presence of only intercept (D1LT0) is believed in data generation and test implementation. A systematic inspection of Table 5 discloses that one test i.e. LMKPSS has a constant rank score of 1 throughout all four sample sizes and the other i.e. LMLBI improves with increase in sample size as it a worst performer up to sample size of 60, and it becomes a mediocre one at sample size of 120 and 240. While, SC exhibits a consistent behavior as it improves its rank score with increase in sample size and is an average performer at all four sample sizes. The rest of two (HH2 and HH1) demonstrate an inconsistent behavior as these two increase their rank scores with increase in sample size. However, HH2 is an average performer up to $n = 120$. Furthermore, SC is performing better for all four sample sizes. While, HH1 is a worst one for all four sample sizes. Similarly, MLS is a mediocre one at only $n = 60$ and for the rest of sample sizes it is the worst one. XFT and HLC are worst ones for all four sample sizes except XFT having mediocre performance for sample size of 30.

The last column of Table 5 illustrates the weighted average of rank scores and hence it provides an overall summary about the performance of tests. This last column reveals that only a single test i.e. LMKPSS is an overall better performer. Two DOLS estimation based tests i.e. SC and HH2 are overall performing average. The rest of five tests are overall worst performers.

The finalized results of Monte Carlo experiment in form of rank scores and weighted average of rank scores are displayed in Table 6, when it is believed that both intercept and linear time trend (D1LT1) are present in data generation and test application. An in-depth evaluation of Table 6 reveals that LMKPSS is the leading performer having the minimum constant rank score of 1 at all four sample sizes and the other LMLBI shows a consistent behavior as it decreases its rank score with increase in sample size. LMLBI is a poor performer for small sample sizes of 30 and 60, an average one at large sample size of 120 and 240. While, SC is an average one at the smallest sample size of 30, a better one for rest of all three sample sizes. Similarly, HH2 and HH1 are average ones at the smallest sample size of 30 and worst ones for rest of three sample sizes. XFT is a mediocre one at the smallest sample size of 30 and moderately small sample size of 60. However, it is worst one at large sample sizes of 120 and 240. Similarly, HLC is the worst one for all four sample sizes. While, MLS is the worst one at the smallest sample size of 30, an average one for rest of three sample sizes.

TABLE 6 - RANK SCORES FOR D1LT1

Tests	$\Im^{\tau}(30)$	$\Im^{\tau}(60)$	$\Im^{\tau}(120)$	$\Im^{\tau}(240)$	Υ^{τ}
LMKPSS	1**	1**	1**	1**	1.00**
SC	4.1*	2**	2.1**	2**	3.13*
HH2	2.7*	4.6	5.1	5.7	3.73*
XFT	3.4*	4.3*	6.6	6.8	4.29*
HH1	4*	6	5.9	6.1	4.93
MLS	7.7	3.1*	2.9*	4*	5.59
LMLBI	6.4	7.1	4.4*	3*	6.09
HLC	6.7	7.9	8	7.3	7.23

Note: ** shows that test is better performer and * shows that test is an average performer

The last column of Table 6 shows an overall performance of tests as it is the weighted average of rank scores. Single test is an overall better performer i.e. LMKPSS. Three tests are overall average performers, from these three; two are based on DOLS estimation i.e. SC and HH2 and one is based on FMOLS estimation i.e. XFT. Rest of four tests are overall worst performers.

5. CONCLUSIONS AND RECOMMENDATIONS

This study is aimed to evaluate the performance of null of cointegration tests on basis of a new proposed methodology of rank scores using Monte Carlo Simulations. Considering the discussion in Section 4, it is concluded that use of asymptotic critical values for these eight tests considered in this study, led to distortion in empirical size. For almost all of cases of deterministic component and sample sizes, these eight tests had empirical size way greater or way smaller than the nominal size considered. This means that these eight tests faced problems of over rejection as well as under rejection. In statistical and econometric analysis these two problems of over rejection and under rejection have worse implications, leading to useless conclusions and recommendations.

To solve these problems of over rejection and under rejection, simulated critical values were estimated and then again empirical sizes were calculated. This practice controlled the empirical size around nominal size. For further evaluation of performance on basis of Power, the same were used to ensure the empirical size around nominal size of 0.05 for all eight tests.

For 1st case of deterministic part i.e. D^0LT^0, LMKPSS, which is a test based on the residuals of cointegrating regression, is consistently performing better for all sample sizes. Three tests (SC, HH2 and HH1) which are based on DOLS estimation, have average performance. The rest of four tests (LMLBI, XFT, HLC and MLS) are worst performers. In general, only the presence of intercept (D^1LT^0) in both data generation and test implementation does not affect the performance of tests except slight changes, as the same LMKPSS is the better and leading performer. Similarly, the same two tests based on DOLS estimation i.e. SC and HH2 are average performers. However, HH1 joins the club of four worst performing tests. For 3rd case of deterministic part i.e. D^1LT^1, again there is not much of difference in performance of tests. The same LMKPSS is the winner in performance and three tests i.e. SC, HH2 and XFT are average performers. The rest of four tests are worst performers.

In the light of conclusions, it is recommended with great confidence that asymptotic critical values may not be used for decision of rejection of null, instead the simulated ones may be used. In modern age, the use of simulated critical values is easily possible due to the availability of fast computers. While testing for possible cointegration using tests with null of cointegration, the LM test may be preferred as compared to the others. The use of tests based on complex and complicated methods of estimation such as DOLS, FMOLS and MLE may be avoided.

AKNOWLEDGEMENT

This work was funded by TUBITAK Turkey under 2216-Research Fellowship Program for International Researchers.

REFERENCES

Ahlgren, Niklas, and Jan Antell. (2008). Bootstrap and Fast Double Bootstrap Tests of Cointegration Rank with Financial Time Series. *Computational Statistics & Data Analysis* 52 (10): 4754–67.

Akdi, Yilmaz. (1996). Periodogram Analysis for Unit Roots.

Banerjee, Anindya, Juan J Dolado, David F Hendry, and Gregor W Smith (1986). EXPLORING EQUILIBRIUM RELATIONSHIPS IN ECONOMETRICS THROUGH STATIC MODELS: SOME MONTE CARLO EVIDENCE*. *Oxford Bulletin of Economics and Statistics* 48: 253–77.

Boswijk, H Peter (1989). *Estimation and Testing for Cointegration with Trended Variables: A Comparison of a Static and a Dynamic Regression Procedure*.

Boswijk, Peter, and Philip Hans Franses (1992). DYNAMIC SPECIFICATION AND COINTEGRATION*. *Oxford Bulletin of Economics and Statistics* 54: 369–81.

Choi, In (1994). Durbin-Hausman Tests for Cointegration. *Journal of Economic Dynamics and Control* 18: 467–80.

Dobrin C., Dinulescu R., Costache R., Voicu L. (2015). One management method, two countries. Lean method applied in Romania and France, In *Proceedings of the 9th International Management Conference: Management and Innovation For Competitive Advantage*, November 5th-6th, Bucharest, Romania, pp. 950-957

Engle, Robert F, and Clive W J Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica: Journal of the Econometric Society* 55: 251–76.

https://doi.org/10.2307/1913236.

Fernández-Macho, Francisco Javier, and Petr Mariel (1994. *Testing the Null of Cointegration, Hausman-like Tests for Regressions with a Unit Root*. Universidad del País Vasco, Facultad de Ciencias Económicas.

Gabriel, Vasco J. (2003) Tests for the Null Hypothesis of Cointegration: A Monte Carlo Comparison. *Econometric Reviews* 22: 411–35. https://doi.org/10.1081/ETC-120025897.

Gonzalo, Jesus, and Tae-Hwy Lee (1998). Pitfalls in Testing for Long Run Relationships. *Journal of Econometrics* 86: 129–54.

Hansen, Bruce (1990). *A Powerful, Simple Test for Cointegration Using Cochrane-Orcutt*. Rochester Center for Economic Research, University of Rochester.

Hansen, Bruce E (1992). Tests for Parameter Instability in Regressions with I (1) Processes. *Journal of Business & Economic Statistics* 10: 321–35.

Haug, Alfred A (1996). Tests for Cointegration A Monte Carlo Comparison. *Journal of Econometrics* 71: 89–115.

Iglesias, Emma M, and Garry D A Phillips (2017). The Use of Bias Correction versus the Jackknife When Testing the Mean Reversion and Long Term Mean Parameters in Continuous Time Models. *Monte Carlo Methods and Applications* 23 (3): 159–64.

Jacobson, Tor, and Rolf Larsson (1999). Bartlett Corrections in Cointegration Testing. *Computational Statistics & Data Analysis* 31 (2): 203–25.

Jansson, Michael (2005). Point Optimal Tests of the Null Hypothesis of Cointegration. *Journal of Econometrics* 124: 187–201.

Johansen, Søren, and Katarina Juselius (1990). Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics* 52: 169–210.

Kascha, Christian, and Carsten Trenkler. 2011. Bootstrapping the Likelihood Ratio Cointegration Test in Error Correction Models with Unknown Lag Order. *Computational Statistics & Data Analysis* 55 (2): 1008–17.

Khan, Asad ul Islam, and Asad Zaman (2017). Theoretical and Empirical Comparisons of Cointegration Tests. Islamabad Pakistan: International Islamic University.

Khan, Waqar Muhammad, and Asad ul Islam Khan (2018). Most Stringent Test of Independence for Time Series. *Communications in Statistics-Simulation and Computation*, 1–19.

Kremers, Jeroen J M, Neil R Ericsson, and Juan J Dolado (1992). The Power of Cointegration Tests. *Oxford Bulletin of Economics and Statistics* 54: 325–48.

Kwiatkowski, Denis, Peter C B Phillips, Peter Schmidt, and Yongcheol Shin (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: 159–78.

Leybourne, Stephen James, and B P M McCabe (1994). Practitioners corner: A Simple Test for Cointegration. *Oxford Bulletin of Economics and Statistics* 56: 97–103.

Mantalos, Panagiotis (2003). Bootstrapping the Breusch-Godfrey Autocorrelation Test for a Single Equation Dynamic Model: Bootstrapping the Restricted vs. Unrestricted Model. *Monte Carlo Methods and Applications Mcma* 9 (3): 257–69.

Mariel, Petr (1996). A Comparison of Cointegration Tests. *Applications of Mathematics* 41: 411–31.

McCabe, B P M, S J Leybourne, and Y Shin (1997). A Parametric Approach to Testing the Null of Cointegration. *Journal of Time Series Analysis* 18: 395–413.

Park, Suk K, Sung K Ahn, and Sinsup Cho (2011). Generalized Method of Moments Estimation for Cointegrated Vector Autoregressive Models. *Computational Statistics & Data Analysis* 55 (9): 2605–18.

Pesaran, M Hashem, Yongcheol Shin, and Richard J Smith (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics* 16: 289–326.

Pesavento, Elena (2004). Analytical Evaluation of the Power of Tests for the Absence of Cointegration. *Journal of Econometrics* 122: 349–84.

Phillips, P C B, and S Ouliaris (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* 58: 165–93. https://doi.org/10.2307/2938339.

Phillips, Peter C B, and Bruce E Hansen (1990). Statistical Inference in Instrumental Variables Regression with I (1) Processes. *The Review of Economic Studies* 57: 99–125.

Sargan, J D, and Alok Bhargava (1983). Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors When the Root Lies on the Unit Circle. *Econometrica* 51: 799–820. https://doi.org/10.2307/1912159.

Shin, Yongcheol (1994). A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration. *Econometric Theory* 10: 91–115.

ul Islam Khan, Asad, Waqar Muhammad Khan, and Mehmood Hussan (2019). Most Stringent Test of Null of Cointegration: A Monte Carlo Comparison. *Communications in Statistics - Simulation and Computation* 0 (0): 1–19. https://doi.org/10.1080/03610918.2019.1691229.

Xiao, Zhijie (1999). A Residual Based Test for the Null Hypothesis of Cointegration. *Economics Letters* 64: 133–41.