View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by European Online Journal of Natural and Social Sciences (ES)

European Online Journal of Natural and Social Sciences 2019; www.european-science.com
Vol.8, No 1 (s) Special Issue on Advancement of Business and Management Science
ISSN 1805-3602

A Monte Carlo Simulation Analysis of Panel Stationarity Tests under a Single Framework

Shahid Akbar[1], Ali Raza[2], Yasir Riaz[1]
[1]International Institute of Islamic Economics, International Islamic University, Islamabad, Pakistan;
[2]University of Haripur, Khyber Pakhtunkhwa, Pakistan
Email: shahidakb121@gmail.com
Tel.: +92-321-5772798

Abstract

A unified framework, stringency criterion have been used to compare the six panel unit root tests having the null hypothesis of stationary and to find the best performer test/tests. Simulated critical values, instead of asymptotic critical values, have been used to keep the size of all tests around nominal size of 5%. Our findings suggest HD and HL tests as better performing tests as compared to other panel stationarity tests.

Keywords: Better Performer, Power, Shortcomings, Simulations, Stringency Criterion.

Introduction

A variety of panel unit root tests has been developed in the last three decades based on different assumptions and mathematical structure to address different characteristics of the panel data. Panel unit root tests are preferred over time series unit root tests due to merging data across cross sections which make panel unit root tests more powerful as compared to time series unit root tests. Two types of panel unit root tests, one having the null hypothesis of panel unit root and second having the null hypothesis of panel stationary, have been used in the existing literature to observe whether a process has an infinite short or long memory to its past performance. Almost all comparative studies, (Maddala and Wu, 1999), (Hlouskova and Wagner, 2006) etc., based on Monte Carlo simulations have compared panel unit root tests having the null hypothesis of unit root by evaluating the size and power properties of these tests while a study of (Hlouskova and Wagner, 2006) and (Demetrescu et al., 2010) have considered two (i.e. (Hadri, 2000) and (Hadri and Larsson, 2005) tests) and three (i.e. (Hadri, 2000), (Hadri and Larsson, 2005) and (Demetrescu et al., 2010) tests) panel stationarity tests, respectively, to make comparison. (Hlouskova and Wagner, 2006) observed that (Hadri, 2000) and (Hadri and Larsson, 2005) tests have serious size distortion when the number of cross section and time series units are small and medium while at large time and cross section dimensions size distortion of these two tests become smaller and at last reaches to nominal size of 5%. Similarly, low power of these two tests have been observed at small time series and cross section levels due to large size distortion. However, a high power has observed at large combination of cross section and time series levels by showing that both of these tests reject stationarity most of the times. Also, they analyzed that both of these tests perform equally from power property point of view.

(Demetrescu et al., 2010) concluded that their test performs better as compared to (Hadri, 2000) and (Shin and Snell, 2006) tests at small time series and cross section dimensions by allowing cross-sectional dependence. These comparative studies have not considered all panel stationarity tests by taking a whole set of alternatives to comprehensively analyze size and power properties under a single framework resulting no definite results which have failed to provide a clear-cut guidelines to researchers. Moreover, all these comparative studies have taken into asymptotic critical values to evaluate the size and power performance which caused size distortion problem. In our study, we compare six panel stationarity tests (i.e. almost all) under a single framework using stringency criterion, a robust technique to compare tests (Zaman et al., 2017) and (Zaman, 1996), by considering a whole set of alternatives after stabilizing size equal to nominal size of 5%.

Panel Stationarity Tests

Basically, stationarity tests used to check the results of unit root tests in the empirical study but both of them have different structure and assumptions according to their layout. All stationarity tests are called residual based tests, as these tests have been derived from residuals of the model. We have considered (Hadri, 2000), (Hadri and Larsson, 2005), (Harris et al., 2005), (Shin and Snell, 2006), (Hadri and Kurozumi, 2009) and (Demetrescu et al., 2010) panel statioarity tests; these tests are abbreviated as HD, HL, HLM, SS, KK and DHT, respectively. All these tests have derived by using the following basic model:

$$y_{it} = r_{it} + \varepsilon_{it} \text{ ------------------------- } (1)$$

and

$$y_{it} = r_{it} + \beta_i t + \varepsilon_{it} \text{ ------------------- } (2)$$

$$r_{it} = r_{i,t-1} + u_{it} \text{ ---------------------- } (3)$$

where $t = 1,......,T$ and $i = 1,.......,N$, ε_{it} and u_{it} are mutually and independent normals and i.i.d across cross sections and over time with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$, $E(u_{it}) = 0$, and $E(u_{it}^2) = \sigma_u^2$. All these residual based tests are the panel extension of (Kwiatkowski et. al., 1992) time series test to test the null hypothesis of stationary around a deterministic level or around a deterministic trend against the alternative hypothesis of a unit root in all units of the panel data. HD and LH tests have same mathematical structure but the main difference is that HL test is constructed under fixed T which expands the finite sample performance by deriving an exact finite mean and variance with assumption of cross section independence. HLM test is developed as a nonparametric panel stationarity test which is robust in the presence of serial dependence and cross-sectional dependence across the panel. SS test is a panel-based mean group test in the occurrence of both serial correlation across time periods and heterogeneity across cross-section units. KK test is introduced for cross-sectional dependence in the form of a common factor in the error term of the model. While DHT test is proposed under an unbounded norm of the long run correlation matrix of the panel to allow for persistent cross correlation.

Methodology

Data generating process (DGP) is the key element to carry out any comparative simulation study. We take the following DGP for the null hypothesis of panel stationarity vs alternative hypothesis of panel unit root,

$$\Delta y_{it} = \alpha_i + \beta_i t - \varphi \varepsilon_{i,t-1} + \varepsilon_{it} \text{ ---------- } (4)$$

where $i = 1,.....,N$, $t = 1,.....,T$, $\alpha_i \sim U[0,10]$, $\beta_i \sim U[0,2]$, $\varepsilon_{it} \sim N\left(0, \delta_{\varepsilon_i}\right)$ and $\delta_{\varepsilon_i} \sim U[0.5,1.5]$. This panel version DGP has been adopted from (Hwang and Schmidt, 1993) in which they use this DGP for the null hypothesis of stationarity in time series. According to (Hwang and Schmidt, 1993), under the null hypothesis of stationarity $\varphi = 0.99999 \cong 1$ and under alterative hypothesis of unit root $0 \le \varphi < 1$. In our study we take $\varphi = \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0\}$ under alternative hypothesis.

Finding the most Stringent Panel Stationarity Test

Stringency criterion is discussed by (Zaman, 1996) which is a robust technique to compare and obtain a most stringent test in the identical field of study. This criterion is related to the powers of panel stationarity tests and point optimal test to find best test.

Point Optimal Test

By definition of Neyman Pearson Lemma (Neyman and Pearson, 1992) point optimal test is the ratio of log likelihood of null hypothesis and alternative hypothesis of interest, mathematically:

$$PO = L(1) - L(\varphi)$$

where log likelihood for given DGP in Equation 4 is,

$$L(\varphi) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\delta_{\varepsilon_i}^2\right) - \frac{1}{2}\log\left(1 + T(1-\varphi)^2\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - y_{t-1}\right)^2}{\left(\delta_{\varepsilon_i}^2\left(1 + T(1-\varphi)^2\right)\right)}$$

$$L(1) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\delta_{\varepsilon_i}^2\right) - \frac{1}{2}\sum_{t=1}^{T}\left(y_t - y_{t-1}\right)^2$$

$$PO = -\frac{1}{2}\sum_{t=1}^{T}\left(y_t - y_{t-1}\right)^2 + \frac{1}{2}\log\left(1 + T(1-\varphi)^2\right) + \frac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - y_{t-1}\right)^2}{\left(\delta_{\varepsilon_i}^2\left(1 + T(1-\varphi)^2\right)\right)}$$

Hence Panel Point Optimal (PPO) test for Null hypothesis of Stationarity is obtained as,

$$PPO = \frac{\sum_{i=1}^{N} PO}{N}$$

Let PE_j and PT_j^m denote the power of point optimal test and power of panel stationarity test "*m*" at a specific alternative "*j*". A difference between power of point optimal test and power of panel stationarity test "*m*" , called shortcomings, is obtained by using the expression $S_j^m = PE_j - PT_j^m$, $S_j^m \ge 0$ where $j = 1, 2, ..., l$ and $m = 1, 2, ..., k$ denotes total number of alternative and total number of tests, respectively, and S_j^m shows the shortcoming of test "*m*" at a specific alternative "*j*". This calculation of shortcomings is made for all the considered tests under the whole set of alternatives (i.e. $j = 1, 2, ..., l$). In the next step minimum of the maximum shortcomings is calculated for all the tests as $\Pi_{\min} = \min(\Pi)$ where $\Pi = \max\left(S_k^m\right)$ represents the maximum shortcomings of each panel stationarity test. Finally, a panel stationarity test having the minimum value among the maximum shortcomings is diagnosed as most stringent test.

Further, panel stationarity tests have been categorized into three categories; worst performer, mediocre performer and better performer having percentage maximum shortcomings in between 0 to 10%, 11% to 50%, and 51% to 100%, respectively.

Results and Discussion

A cross sectional dimension of N=4, 8, 16, 32 and time series dimension of T=10, 25, 50, 100 for Monet Carlo simulation size (MCSS) of 5000 have been taken to carry out our simulation study. Simulated critical values are calculated and used instead of asymptotic critical values to avoid size distortion problem and provide a stable size around nominal size of 5% of all tests to make comparison meaningful. In our study we have considered analysis with intercept term only in the deterministic part as other two situations of deterministic terms (i.e. without deterministic terms and, with intercept and trend terms) have the same results. In each figure of our simulation study, y-axis and x-axis indicate percentage maximum shortcomings and fixed time series size corresponding to different level of cross section units respectively.

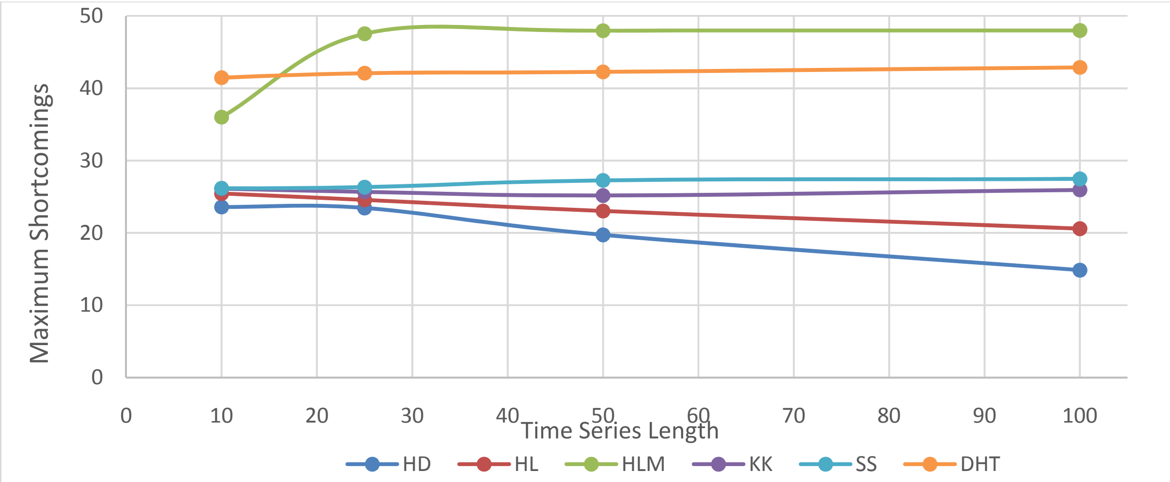

Figure 1: Percentage Maximum Shortcomings Assessments of Panel Stationarity Tests, N=4

Figure 1 shows the power performance of panel stationarity tests; HD, HL, KK, SS, DHT and HLM, corresponding to different time series level when the cross sectional dimension is small (i.e. N=4). Clearly, HD and HL tests with respect to their slow convergence pattern of maximum shortcomings have stood as better performing tests as compared to other four tests in the category of mediocre tests. Moreover, HD test with lowest maximum shortcomings of 23.6%, 23.4%, 19.7%, and 14.8% is considered as most stringent test as compared to maximum shortcomings of HL test over all-time series dimension in the better tests category.

Table 1: Percentage Maximum Shortcomings of Panel Stationarity Tests, N=4 and N=8

	Panel A: N=4				Panel B: N=8			
Tests/TL	10	25	50	100	10	25	50	100
HD	23.58*	23.44*	19.72*	14.84*	19.84*	19.78*	16.46*	14.38*
HL	25.44*	24.56*	23.02*	20.58*	19.84*	19.82*	17.82*	16.74*
HLM	36*	47.54*	47.96*	47.99*	36.04*	45.88*	46.57*	47.03*
KK	26.12*	25.68*	25.16*	25.94*	24.42*	24.38*	25.94*	26.02*
SS	26.17*	26.32*	27.26*	27.46*	25.44*	26.75*	27.24*	27.68*
DHT	41.46*	42.08*	42.26*	42.88*	42.54*	42.56*	42.94*	43.34*

Note: * indicates mediocre performer tests.

However, KK and SS tests with approximately constant pattern of maximum shortcomings over time dimension are categorized as mediocre tests beside with HLM and DHT tests showing a large distance between power curve and power envelope. Figure 1 and Table 1 (Panel A) also indicate that HLM and DHT tests have very bad performance in the category of mediocre performer tests. Moreover, no single test is ordered as best or worst test according to assigned maximum shortcomings at each level of time series when the number of cross section level is small.

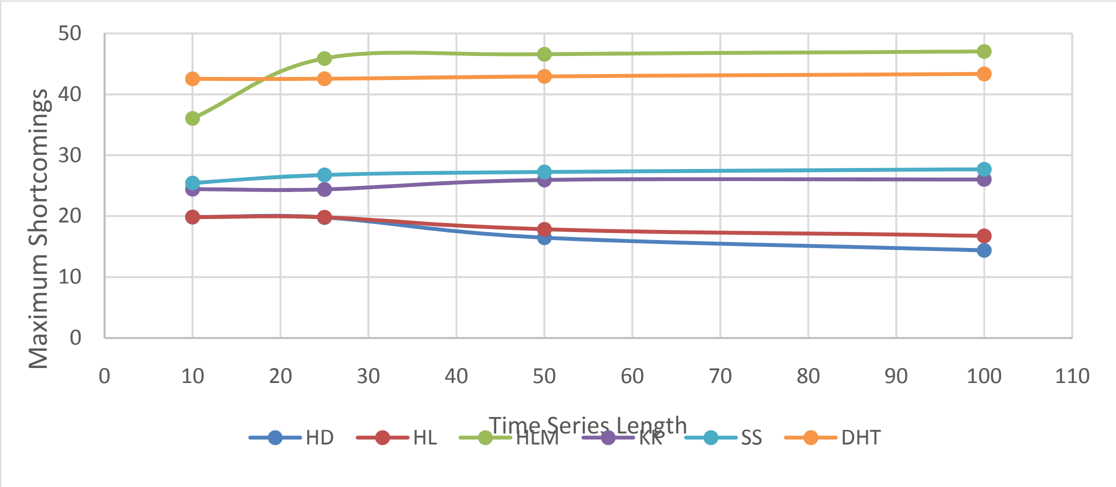

Figure 2: Percentage Maximum Shortcomings Assessments of Panel Stationarity Tests, N=8

Further, as the number of cross section increases from 4 to 8 a similar picture has been depicted from Figure 2 and Table 1 (Panel B) as has been inspected at N=4 over time series level of 10, 25, 50, and 100. Again, HD and DL tests in the category of mediocre tests are classified as better performer tests. Also, Figure 2 shows that HD test with less maximum shortcomings at each level of time series in the category of mediocre tests is detected as most stringent test. Similarly, HL test with maximum shortcomings 19.84%, 19.82%, 17.82%, and 16.74% corresponding to time series dimension 10, 25, 50, and 100 is ranked as the second better performer test among the mediocre tests. However, DHT and HLM tests with gain of maximum shortcomings in between 35% and 50% are identified as bad performer tests in the class of mediocre tests category. While, KK and SS tests remain in the middle of better performer and bad performer tests indicating their average level performance. Moreover, Figure 2 also indicates that none of the tests have fulfilled best test criteria with maximum shortcomings less than or equal to 10%.

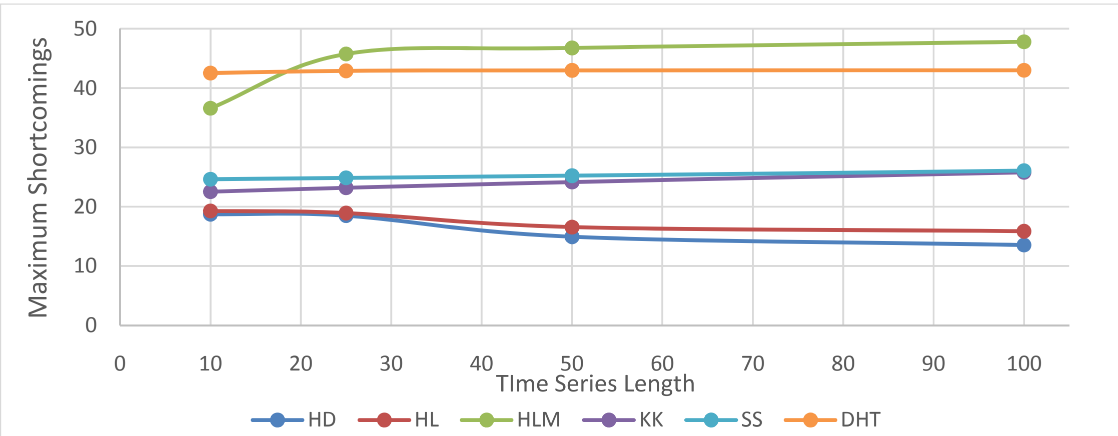

Figure 3: Percentage Maximum Shortcomings Assessments of Panel Stationarity Tests, N=16

Figure 3 and Table 2 (Panel A) demonstrate that almost all tests have same type of behavior when N=16 as has been observed for N=8 with little improvement for majority of the tests. It is analyzed that HD and HL tests have the same status of better performer tests corresponding to their gained maximum shortcomings as have been observed previously for small cross section levels (i.e. N=4 and N=8). Moreover, the maximum shortcomings of both of these tests are less at N=16 as compared to their performance at N=8 with convergence picture. However, HLM test with divergence behavior of maximum shortcomings of 36.6%, 45.7%, 46.8%, and 47.8% at time series level 10, 25, 50, and 100, respectively, is ranked as bad performer test among mediocre tests. This result is very similar to N=8 for HLM test at each time series level indicating its worst performance. While, KK and SS tests with a little improvement in their maximum shortcomings pattern as compared to previous cross section unit of 8 are stood as third and fourth mediocre tests with divergence behavior. While, DHT test with approximately constant divergence behavior over time series 10, 25, 50, and 100 is identified as second bad performer test beside HLM test in the mediocre tests category.

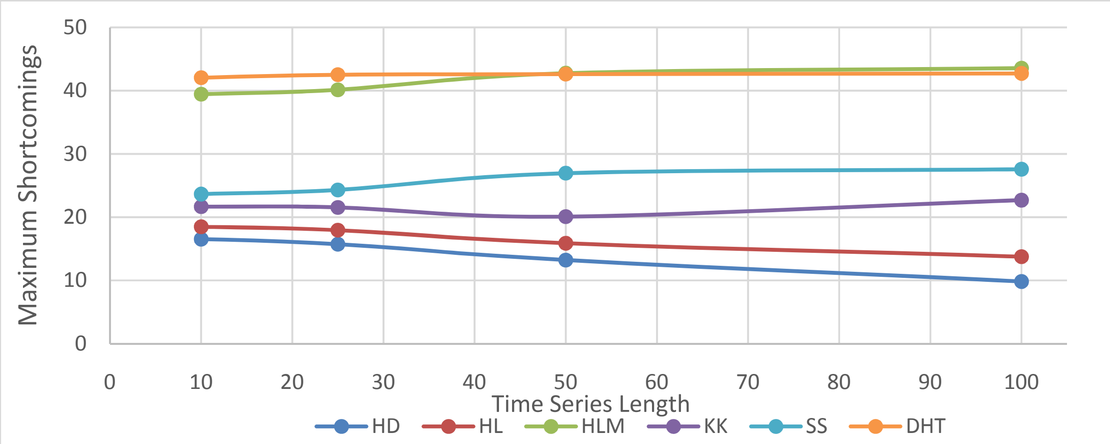

Figure 4: Percentage Maximum Shortcomings Assessments of Panel Stationarity Tests, N=32

When the number of cross section units are 32 then Figure 4 and Table 2 (Panel B) again indicate that all panel stationarity tests have categorized into the group of mediocre tests. In other words, no test is assigned as best and worst performing test excluding the best performing position of HD test with maximum shortcomings of 9% at T=10. At N=32, Figure 4 and Table 2 (Panel B) show that HD and HL tests again are identified as better performing tests in the category of mediocre tests over varying time series level of 10, 25, 50, and 100 corresponding to maximum shortcomings of 16.52%, 15.72%, 13.2%, 9.82% and 18.48%, 17.94%, 15.86%, 13.74%, respectively. At time series 100, HD test with its convergence pattern is ranked as best performing test having maximum shortcomings 9% showing that at large time series and cross section level this test will achieve its power curve equal to power envelop of point optimal test. However, HL test is not labeled in best performing test at large sample size of 100 as does HD test but at very large time series level this test will also finally have power equal to the power of point optimal test. Moreover, HLM test with maximum shortcomings of 39.44%, 40.14%, 42.76%, and 43.56% at time series level of 10, 25, 50, and 100, respectively, is ranked as the bad performing test among mediocre tests category. This result indicates a constant bad performance behavior of HLM test at each level of cross section unit whether this cross section unit is small, medium or large. Similarly, DHT test is also categorized as bad performer test among mediocre tests class with maximum shortcomings 42.06%, 42.52%, 42.62%, and 42.73% over time series level of 10, 25, 50, and 100, respectively. DHT test also shows a very similar constant behavioral picture that has been analyzed for HLM test. Finally, KK test with little improvement in their maximum shortcomings, which fluctuates in between 20% to22%, is categorized as third mediocre test. Similarly, SS test with divergence pattern of its maximum shortcomings is also remained in the mediocre tests class like its performance for the previous cross section units.

Overall, it is observed from Figure 1 to Figure 4 and Table 1 to Table 2 that all panel stationarity tests lie in the class of mediocre tests majority of the time when data generating process and test equation have drift term only. Among these six tests, HD and HL tests have maintained their convergence behavior and have ranked as better performing tests in the category of all mediocre tests. Also, majority of the time HD test is identified as most stringent test having minimum value of maximum shortcomings at each level of time series length as compared to other five tests. However, its shortcomings value is observed very close to the maximum value of HL test in each combination of time series and cross section levels. A similar results are observed if cross section effect is carried out over the power performance of panel stationarity tests.

Table 2: Percentage Maximum Shortcomings of Panel Stationarity Tests, N=16 and N=32

	Panel A: N=16				Panel B: N=32			
Tests/TL	10	25	50	100	10	25	50	100
HD	18.74*	18.5*	14.94*	13.54*	16.52*	15.72*	13.2*	9.82**
HL	19.28*	18.93*	16.56*	15.86*	18.48*	17.94*	15.86*	13.74*
HLM	36.6*	45.74*	46.76*	47.8*	39.44*	40.14*	42.76*	43.56*
KK	22.54*	23.18*	24.16*	25.78*	21.66*	21.54*	20.08*	22.68*
SS	24.63*	24.86*	25.22*	26.08*	23.64*	24.3*	26.95*	27.58*
DHT	42.52*	42.9*	42.98*	43*	42.06*	42.52*	42.62*	42.73*

Note: * and ** indicates mediocre and best performer tests respectively.

Conclusions and Recommendations

In this study we have compared six panel stationarity tests (almost all) under the unified framework to analyze the power performance by taking the whole set of alternatives. It is concluded that HD and HL tests are better performer tests according to their attained maximum shortcomings in the category of mediocre tests whether the time series and cross section lengths are small, medium or large. However, HLM and DHT tests with their attained maximum shortcomings in between 38% to 48% are identified as bad performer tests in the class of mediocre tests at each combination of time series and cross section length. We have discussed our simulation results in the presence of intercept term only in the DGP and test equations but a similar results have concluded if both of the deterministic parts are included or excluded. On the basis of our findings, it is recommend to use HD and HL tests in the class of residual based tests having the null hypothesis of panel stationarity if the number of time series and cross section units are small, medium or large. Also, HLM and DHT tests due to their bad performance are avoid to be used in empirical work.

References

Demetrescu, M., Hassler, U. and Tarcolea, A.I., (2010). Testing for Stationarity in Large Panels with Cross-Dependence, and US Evidence on Unit Labor Cost. Journal of Applied Statistics **37**(8): 1381-1397.

Hadri, K. (2000). Testing for Stationarity in Heterogeneous Panel Data. The Econometrics Journal **3**: 148-161.

Hadri, K., and Kurozumi, E. (2009). A Simple Panel Stationarity Test in the Presence of Cross-Sectional Dependence. Economics Working Papers 09-01, Queen's Management School, Queen's University Belfast.

Hadri, K., and Larsson, R. (2005). Testing for Stationarity in Heterogeneous Panel Data where the Time Dimension is Finite. The Econometrics Journal **8**(1): 55-69.

Harris, D., Leybourne, S., and McCabe, B. (2005). Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence. Journal of Business & Economic Statistics **23**(4): 395-409.

Hlouskova, J., and Wagner, M. (2006). The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. Econometric Reviews **25**(1): 85-116.

Hwang, J., and Schmidt, P. (1993). On the Power of Point Optimal Tests of the Trend Stationarity Hypothesis. Economics Letters **43**(2): 143-147.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are We that Economic Time Series have a Unit Root? Journal of Econometrics **54**(1-3): 159-178.

Maddala, G. S., and Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. Oxford Bulletin of Economics and Statistics **61**(S1): 631-652.

Neyman, J., and Pearson, E.S. (1992). On the Problem of the Most Efficient Tests of Statistical Hypotheses, Springer.

Shin, Y., and Snell, A. (2006). Mean Group Tests for Stationarity in Heterogeneous Panels. The Econometrics Journal **9**(1): 123-158.

Zaman, A. (1996). Statistical Foundations for Econometric Techniques, Academic Press.

Zaman, A., Zaman, A., and Rehman, A. U. (2017). The Concept of Stringency for Test Comparison: The Case of a Cauchy Location Parameter. International Econometric Review (IER), 9(1), 1-18.